# ZERO-SHOT LOW-LIGHT IMAGE ENHANCEMENT VIA DIFFUSION-LEARNED PHOTOMETRIC PRIORS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Zero-shot low-light image enhancement (LLIE) is remarkably challenging due to that it has to jointly address spatially varying illumination, color shifts, and noise from a single low-light observation without any paired supervision. Motivated by the observation that pre-trained diffusion models can offer rich natural-image statistics as photometric priors to guide this task, we introduce a structure–photometry decoupled LLIE framework that treats diffusion outputs as photometric references while anchoring scene structures to the low-light observation. To be more specific, We first empoly a noise-aware internal image prior to estimate the observation-derived structure. Then a frozen Stable Diffusion model is applied to produce a stochastic normal-light reference using only two reverse updates. From the photometric relationship between the obtained reference and the estimated structure, we infer global and spatially varying RGB gain–bias transformations. The global transformation provides stable image-level correction, whereas the spatial transformations accommodate non-uniform illumination. A global-anchored spatial trust region further regularizes unreliable local estimates toward the global transformation and bounds excessive local modulation. By applying these constrained transformations to the observation-derived structures, we obtain an enhanced image that transfers the illumination and color statistics learned from the diffusion model while minimizing content drift. Extensive experiments on both paired and unpaired LLIE benchmarks demonstrate competitive zero-shot enhancement performance, efficient inference, and substantially improved structural fidelity over direct diffusion-based reconstruction. Additional automatic white-balance results further demonstrate the potential of converting generative priors into constrained photometric priors for restoration tasks beyond low-light enhancement.

## 1 INTRODUCTION

Low-light image enhancement (LLIE) aims to recover a high-quality image with visible structures, vivid color, and fine details from a poorly illuminated image. Early deep learning-based methods typically learn task-specific mappings from paired low- and normal-light images, often with explicit illumination modeling Wei et al. (2018); Zhang et al. (2019b); Xu et al. (2020; 2022). More recent approaches have reduced paired supervision through reference-free learning, unfolded optimization, and adaptive illumination estimation Guo et al. (2020); Liu et al. (2021); Ma et al. (2022). Despite their effectiveness, these methods still depend on LLIE-specific training data, objectives, or optimization procedures. Therefore, it is more applicable to develop zero-shot LLIE from a single low-light observation. The challenge of zero-shot LLIE is not merely to recover a plausible normal-light appearance, but also to preserve the scene evidence contained in the original input.

Pre-trained diffusion models Ho et al. (2020); Rombach et al. (2022) provide rich natural-image priors and have been widely explored for zero-shot inverse problems through measurement-constrained restoration, including DDRM Kawar et al. (2022), DPS Chung et al. (2023), and DDNM Wang et al. (2023b). Diffusion priors have also been introduced into LLIE through Retinex decomposition Yi et al. (2023), exposure-aware degradation modeling and latent regularization Huang et al. (2025), attribute-guided diffusion sampling Lin et al. (2025), and diffusion-based modulation and refinement Cho et al. (2026). Despite these advances, existing methods commonly allow diffusion-generated or diffusion-optimized high-dimensional representations to directly influence the restora-



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

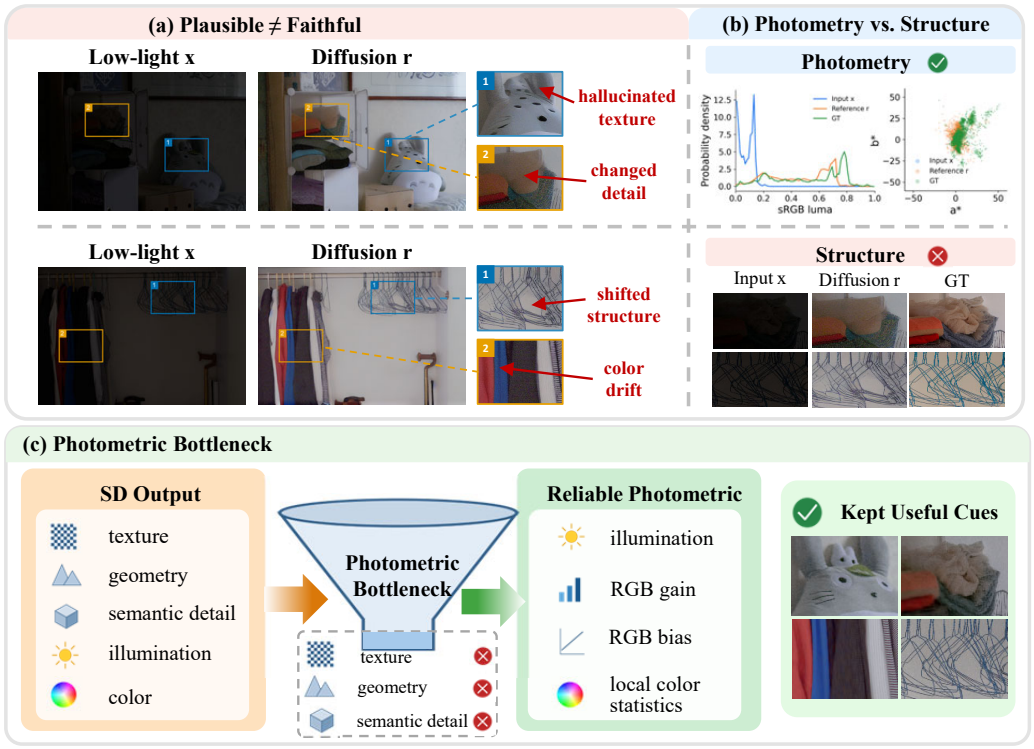



Figure 1: **Motivation of our diffusion-learned photometric priors.** (a) Direct diffusion restoration can produce visually plausible results but unfaithful appearance with hallucinated textures, unpleasing details, unexpected structural shifts, and color drift. (b) Despite unreliable structures, diffusion retains consistent photometric statistics with normal-light references, providing reliable diffusion-learned priors for zero-shot LLIE. (c) This motivates us to design a photometric bottleneck to filter out unreliable structures and preserve illumination and color cues for structure-preserving zero-shot LLIE.

tion process. This generative paradigm raises an inherent concern: direct diffusion restoration may produce visually plausible yet unfaithful results with respect to the input scene (see Fig. 1(a)). Importantly, however, such outputs can still retain useful normal-light photometric statistics, even though their local structural cues remain unreliable (see Fig. 1(b)). This observation raises a more fundamental question: *how can we selectively harness reliable information from diffusion for restoration, rather than struggling to make the diffusion outputs more faithful to the input?*

We therefore rethink diffusion as a photometric prior rather than an image reconstructor. Our key idea is to decouple observation-derived structure from diffusion-learned photometry, enabling the enhanced image to preserve scene fidelity while exploiting diffusion priors for reliable illumination and color correction. Instead of transferring high-dimensional generated content into the restored image, we introduce a constrained photometric bottleneck that retains illumination and color cues while suppressing generated texture, geometry, and semantic content. As shown in Fig. 1(c), the diffusion reference is compressed into low-capacity photometric transformations, allowing diffusion to guide how the observed scene should be corrected without determining what appears in the scene.

Following this principle, we develop a novel structure–photometry decoupled framework for zero-shot LLIE. We first employ a noise-aware internal image prior to recover a structure-preserving estimate directly from the observed low-light image. Next, a frozen Stable Diffusion model provides stochastic normal-light references as photometric evidence. We then exploit the photometric relationship between the observation-derived estimate and the diffusion references to estimate global and spatial RGB gain–bias transformations, while a globally anchored spatial constraint suppresses unreliable local corrections. The final enhanced image is rendered exclusively from the observation-derived structure. Unlike existing diffusion-based restoration methods, our approach uses the pre-trained diffusion model only to derive a constrained photometric prior, rather than for pixel-level image reconstruction. Experiments on both paired and unpaired LLIE benchmarks demonstrate

competitive zero-shot enhancement performance, improved structural fidelity, and efficient inference. Results on automatic white balance further demonstrate that the proposed principle can extend beyond low-light enhancement.

To sum up, our contributions are threefold as follows:

- We propose a novel zero-shot LLIE framework that learns photometric priors from a pre-trained diffusion model, which decouples diffusion-learned photometry from scene content for fidelity-sensitive restoration.
- We introduce a constrained photometric bottleneck that converts diffusion-learned photometric priors into RGB affine transformations for better recovering normal-light illumination and color while preserving scene content.
- Extensive experiments on several benchmarks demonstrate that the newly proposed approach harvests competitive zero-shot performance, improved structural fidelity, and efficient inference. Moreover, our method can easily be extended to auto white balance.

## 2 Related Work

### 2.1 Low-Light Image Enhancement

LLIE has been studied through Retinex modeling, supervised restoration, and learning with reduced supervision Li et al. (2021a). Early deep methods rely on paired data through raw sensor restoration Chen et al. (2018) and Retinex decomposition Wei et al. (2018). Later approaches introduce SNR-aware modeling Xu et al. (2022), conditional distributions Wang et al. (2022), and transformers Wang et al. (2023a); Cai et al. (2023). To reduce paired supervision, other methods explore zero-reference learning Guo et al. (2020), self-calibrated illumination Ma et al. (2022), paired low-light instances Fu et al. (2023), and Retinex-inspired unfolding Liu et al. (2021); Wu et al. (2022). Recent studies further incorporate structural and semantic priors Xu et al. (2023); Wu et al. (2023); Zheng et al. (2023), dedicated color spaces Yan et al. (2025), and efficient architectures Feijoo et al. (2025). Our work instead explores LLIE with a frozen generative prior and no task-specific training.

### 2.2 Diffusion Priors for Image Restoration

Diffusion models Ho et al. (2020); Rombach et al. (2022) provide strong natural image priors for restoration. DDRM, DPS, and DDNM introduce measurement consistency through spectral decomposition, posterior guidance, and null-space constraints Kawar et al. (2022); Chung et al. (2023); Wang et al. (2023b), while later methods improve constrained restoration, equilibrium modeling, residual prediction, and sampling guidance Ding et al. (2024); Cao et al. (2024); Liu et al. (2024a); Garber & Tirer (2024). DiffIR Xia et al. (2023) predicts a compact diffusion prior to guide a learned restoration transformer. In contrast, our method keeps the diffusion model frozen and converts stochastic references into explicit photometric transformations applied directly to observation-derived structure.

### 2.3 Diffusion-Based Low-Light Enhancement

Diffusion-based LLIE includes Retinex component restoration Yi et al. (2023), exposure-guided diffusion Wang et al. (2023c), multi-condition nighttime enhancement Li et al. (2024), and methods using physical and frequency priors Wang et al. (2024); Lv et al. (2024). ReDDiT Lan et al. (2025) improves efficiency through trajectory refinement and distillation. Recent approaches further explore test-time use of pre-trained diffusion priors for LLIE, including training-free and zero-shot formulations Huang et al. (2025); Lin et al. (2025); Cho et al. (2026). Despite different designs, diffusion remains involved in image reconstruction or refinement. Our method instead uses diffusion outputs only as stochastic photometric references, which are projected into global and spatial RGB gain-bias transformations applied to structure recovered from the observation.

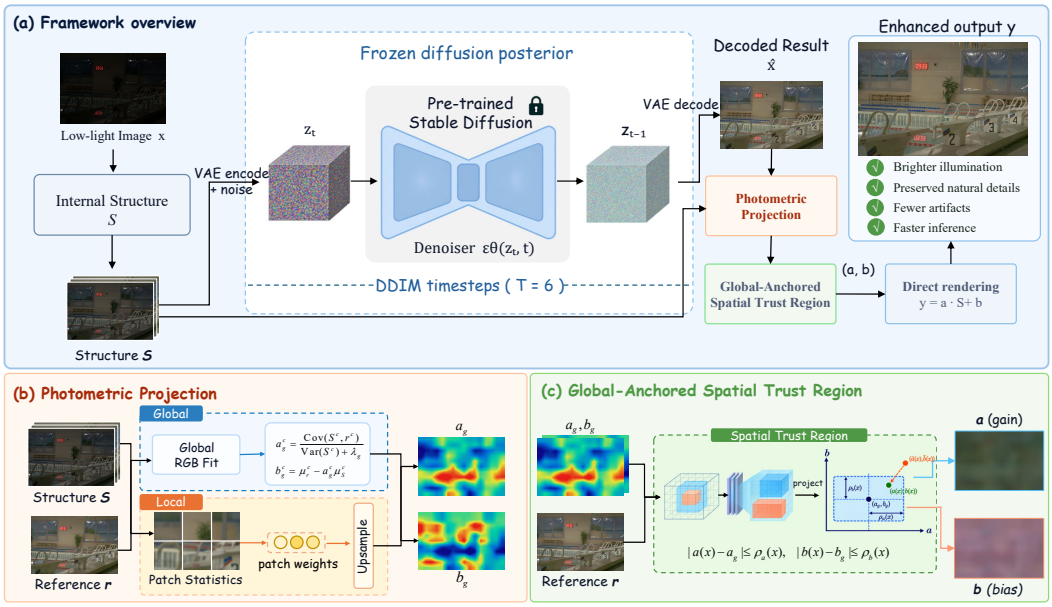



Figure 2: **Overview of our zero-shot structure–photometry decoupled framework.** (a) From a low-light input $\mathbf{x}$, we recover an observation-derived structure $\mathbf{S}$ and generate a stochastic diffusion reference $\mathbf{r}$ from a photometric anchor $\mathbf{A}$. (b) Their photometric relation is projected into global and spatial RGB gain–bias transformations. (c) A global-anchored spatial trust region constrains local corrections, and the final image is rendered directly from $\mathbf{S}$.

# 3 METHOD

Our framework separates observation-derived structure from diffusion-based photometric guidance, as shown in Fig. 2. We first estimate a conservative photometric anchor and scene structure from the low-light input (Sec. 3.1). The anchor initializes a short frozen diffusion process to generate normal-light references (Sec. 3.2), whose photometric cues are projected into constrained RGB gain–bias transformations rather than directly used as restored content (Sec. 3.3). A globally anchored spatial trust region further suppresses unreliable local corrections, allowing diffusion to guide illumination and color while preserving the observed structure.

## 3.1 OBSERVATION-DERIVED STRUCTURE PRIOR

Given a low-light image $\mathbf{x} \in [0,1]^{H\times W\times 3}$, we first recover an observation-derived structure estimate $\mathbf{S}$ that preserves scene content while providing a stable basis for subsequent photometric correction.

We convert $\mathbf{x}$ to linear RGB, denoted by $\mathbf{x}_{\text{lin}}$, and estimate a coarse illumination map from the maximum channel response:

$$\mathbf{L} = \mathcal{G}\left(\max_{c\in\{1,2,3\}} \mathbf{x}_{\text{lin},c}\right), \tag{1}$$

where $\mathcal{G}(\cdot)$ denotes luminance-guided filtering. Following a Retinex-style decomposition, we obtain $\mathbf{R} = \mathbf{x}_{\text{lin}}/(\mathbf{L}+\epsilon)$ and construct a conservative illumination anchor:

$$\mathbf{A}_0 = \mathbf{R} \odot \mathbf{L}^{\gamma_a}. \tag{2}$$

The exposure exponent $\gamma_a$ is adaptively determined from the input luminance, such that darker observations receive stronger but bounded illumination lifting. A global monotonic tone mapping further adjusts exposure while suppressing clipping and excessive chroma amplification:

$$\mathbf{A} = \mathcal{T}(\mathbf{A}_0; \gamma_t, g_t). \tag{3}$$

Since illumination lifting may amplify sensor noise together with weak structures, we further refine $\mathbf{A}$ using noise-aware internal patch statistics. Specifically, the noise level is estimated from smooth

regions of the observation and propagated to the anchor domain according to the local photometric response. Components that cannot be reliably distinguished from noise are suppressed through patch-domain spectral shrinkage, and overlapping patches are aggregated to obtain:

$$\mathbf{S} = \mathcal{P}(\mathbf{A}), \tag{4}$$

where $\mathcal{P}(\cdot)$ denotes the noise-aware internal reconstruction.

The resulting $\mathbf{S}$ serves as the structure carrier for final rendering, whereas $\mathbf{A}$ is used only to initialize the frozen diffusion prior.

### 3.2 Diffusion as a Stochastic Photometric Prior

We next introduce the generative prior while deliberately restricting its role to photometric guidance rather than image reconstruction. Starting from the photometric anchor $\mathbf{A}$, a frozen pre-trained Stable Diffusion model produces a stochastic reference for plausible normal-light appearance.

We first encode the anchor into the latent space,

$$\mathbf{z}_0 = \mathcal{E}(\mathbf{A}), \tag{5}$$

where $\mathcal{E}(\cdot)$ denotes the frozen VAE encoder. Instead of starting from pure Gaussian noise, we perturb $\mathbf{z}_0$ at an intermediate timestep $t_s$, and perform a short DDIM reverse trajectory:

$$\mathbf{z}_{t_s} = \sqrt{\bar{\alpha}_{t_s}}\mathbf{z}_0 + \sqrt{1-\bar{\alpha}_{t_s}}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{6}$$

This image-conditioned initialization allows the diffusion prior to adjust normal-light appearance while retaining information already contained in the observation-derived anchor.

To further limit sampling drift, we impose measurement and structural consistency during reverse diffusion. For the decoded clean prediction $\tilde{\mathbf{r}}_t$, we model its corresponding low-light observation as:

$$\tilde{\mathbf{x}}_t = \mathbf{L} \odot \tilde{\mathbf{r}}_{t,\mathrm{lin}}, \tag{7}$$

where $\mathbf{L}$ is the illumination estimate in Eq. 1. Under a heteroscedastic read–shot noise model, the robust measurement objective is:

$$\mathcal{L}^t_{\mathrm{mean}} = \frac{1}{|\Omega|}\sum_{p\in\Omega}\sqrt{\left(\frac{\tilde{\mathbf{x}}_t(p) - \mathbf{x}_{\mathrm{lin}}(p)}{\sigma_t(p)+\epsilon}\right)^2 + \epsilon^2}, \tag{8}$$

with

$$\sigma_t = \sqrt{\sigma^2_{\mathrm{read}} + \sigma^2_{\mathrm{shot}}[\tilde{\mathbf{x}}_t]_+}. \tag{9}$$

We additionally discourage structural deviation from the observation-derived anchor using:

$$\mathcal{L}^t_{\mathrm{str}} = \left\|\nabla\tilde{\mathbf{r}}_t - \nabla\mathbf{A}\right\|_1. \tag{10}$$

The two guidance terms are independently RMS-normalized and combined as

$$\mathbf{g}_t = \alpha_t\mathcal{N}\left(\nabla_{\mathbf{z}_t}\mathcal{L}^t_{\mathrm{mean}}\right) + \lambda_s\mathcal{N}\left(\nabla_{\mathbf{z}_t}\mathcal{L}^t_{\mathrm{str}}\right), \tag{11}$$

where $\alpha_t$ adaptively increases with the current measurement residual. The standard DDIM update is then corrected by

$$\mathbf{z}_{t-1} = \mathbf{z}^{\mathrm{DDIM}}_{t-1} - \eta\mathbf{g}_t. \tag{12}$$

After the short reverse trajectory, the final latent is decoded as

$$\mathbf{r} = \mathcal{D}(\mathbf{z}^\star_0), \tag{13}$$

where $\mathcal{D}(\cdot)$ is the frozen VAE decoder. The resulting $\mathbf{r}$ is used solely as a stochastic photometric reference for estimating the constrained gain–bias transformations in Sec. 3.3.

### 3.3 CONSTRAINED PHOTOMETRIC BOTTLENECK

The final stage converts the diffusion prior from pixel-level generation into an explicit photometric correction. Given the observation-derived structure $\mathbf{S}$ and the diffusion reference $\mathbf{r}$, we next transfer the diffusion prior from pixel space to a compact photometric parameter space.

We first estimate a global photometric transformation:

$$\mathbf{r} \approx \mathbf{a}_g \odot \mathbf{S} + \mathbf{b}_g, \tag{14}$$

where $\mathbf{a}_g, \mathbf{b}_g \in \mathbb{R}^3$ denote channel-wise gain and bias. For each channel $c$, the regularized least-squares solution can be written as:

$$a_g^c = \frac{\mathrm{Cov}(S^c, r^c)}{\mathrm{Var}(S^c) + \lambda_g}, \qquad b_g^c = \mu_r^c - a_g^c \mu_S^c, \tag{15}$$

where $\mu_S^c$ and $\mu_r^c$ denote the corresponding channel means.

Nevertheless, a global-only transformation fails to model spatially non-uniform illumination. To address this limitation, we derive spatially-varying gain and bias fields from the local statistics of $\mathbf{S}$ and $\mathbf{r}$. Consider a local neighborhood $\mathcal{W}_p$ centered at pixel p; the unconstrained local parameters are defined as:

$$\begin{aligned} \tilde{a}^c(p) &= \frac{\mathrm{Cov}_{\mathcal{W}_p}(S^c, r^c)}{\mathrm{Var}_{\mathcal{W}_p}(S^c) + \lambda_l}, \\ \tilde{b}^c(p) &= \mu_r^c(p) - \tilde{a}^c(p)\mu_S^c(p). \end{aligned} \tag{16}$$

Since local statistics become unreliable in weak-texture or poorly exposed regions, we shrink the local estimate toward the global solution according to a confidence weight $w^c(p)$:

$$\begin{aligned} \bar{a}^c(p) &= a_g^c + w^c(p)\big(\tilde{a}^c(p) - a_g^c\big), \\ \bar{b}^c(p) &= b_g^c + w^c(p)\big(\tilde{b}^c(p) - b_g^c\big). \end{aligned} \tag{17}$$

To further prevent stochastic diffusion variations from inducing excessive local modulation, we introduce a global-anchored spatial trust region. Specifically, the spatial parameters are explicitly restricted around the global photometric solution:

$$a^c(p) = \mathrm{proj}_{\left[(1-\delta_a)a_g^c,\,(1+\delta_a)a_g^c\right]}\big(\bar{a}^c(p)\big), \tag{18}$$

$$b^c(p) = \mathrm{proj}_{\left[b_g^c-\delta_b,\,b_g^c+\delta_b\right]}\big(\bar{b}^c(p)\big), \tag{19}$$

Thus, the global transformation serves as a reliable photometric anchor, while the spatial fields are allowed to model only bounded residual variations.

Finally, the enhanced image is rendered directly from the structure estimate:

$$\hat{\mathbf{y}} = \mathbf{a} \odot \mathbf{S} + \mathbf{b}. \tag{20}$$

This projection converts the high-dimensional diffusion prior into a constrained photometric prior, preserving the input-derived structure while retaining the natural-light statistics provided by the frozen generative model.

## 4 EXPERIMENTS

**Datasets.** We evaluate our method on both paired and unpaired low-light image enhancement benchmarks. For paired evaluation, we employ LOLv1 Wei et al. (2018), LOLv2 Yang et al. (2021), and LSRW Hai et al. (2023). Specifically, LOLv1 contains 15 test pairs, LOLv2 provides 100 real-world test pairs, and we use 50 test pairs from LSRW. For unpaired evaluation, we follow the commonly adopted protocol and evaluate on five real-world benchmarks: DICM Lee et al. (2013), LIME Guo et al. (2016), NPE Wang et al. (2013), MEF Ma et al. (2015), and VV Vonikakis et al. (2018), which are collectively denoted as *Unpaired* in Table 1. The evaluation set contains 44 images from DICM, 10 from LIME, 75 from NPE, 79 from MEF, and 24 from VV.

Table 1: Quantitative comparison with state-of-the-art methods on paired and unpaired low-light image enhancement benchmarks. The average runtime was evaluated on the LOL dataset using a single NVIDIA L20 GPU. "S", "U", and "Z" denote supervised, unsupervised, and zero-shot methods, respectively. ↑ (↓) indicates that higher (lower) is better. "–" denotes unavailable results.

| | Method | Venue | Time (min) | LOL | | | | | | LSRW | | | Unpaired | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | V1 | | | V2-Real | | | | | | | | |
| | | | | PSNR ↑ | SSIM ↑ | LPIPS ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ | ILNIQE ↓ | BRISQUE ↓ | NL ↓ |
| S | SNR-Aware Xu et al. (2022) | CVPR'22 | 0.01 | 21.888 | 0.848 | 0.156 | 21.480 | 0.849 | 0.163 | 17.840 | 0.564 | 0.477 | 28.419 | 35.248 | 0.404 |
| S | GSAD Hou et al. (2023) | NeurIPS'23 | 0.03 | 20.605 | 0.847 | 0.112 | 20.153 | 0.846 | 0.113 | 17.584 | 0.551 | 0.327 | 26.361 | 29.020 | 0.750 |
| S | Retinexformer Cai et al. (2023) | ICCV'23 | 0.01 | 13.026 | 0.426 | 0.365 | 22.794 | 0.840 | 0.171 | 11.412 | 0.269 | 0.506 | 36.318 | 38.295 | 0.399 |
| S | Diff-Plugin Liu et al. (2024b) | CVPR'24 | 0.04 | 18.827 | 0.704 | 0.183 | – | – | – | 17.962 | 0.522 | 0.303 | 22.673 | 18.144 | 0.853 |
| S | CIDNet Yan et al. (2025) | CVPR'25 | 0.02 | 28.920 | 0.889 | 0.067 | 24.111 | 0.871 | 0.108 | 17.402 | 0.535 | 0.286 | 24.987 | 24.437 | 0.682 |
| S | GT-mean loss Liao et al. (2025) | ICCV'25 | 0.01 | 27.709 | 0.875 | 0.123 | 21.810 | 0.852 | 0.143 | 17.534 | 0.549 | 0.391 | 26.681 | 29.912 | 0.609 |
| U | Zero-DCE Guo et al. (2020) | CVPR'20 | 0.02 | 17.642 | 0.572 | 0.315 | 18.060 | 0.574 | 0.313 | 15.868 | 0.451 | 0.315 | 26.834 | 23.164 | 1.534 |
| U | Zero-DCE++ Li et al. (2021b) | TPAMI'21 | 0.02 | 17.018 | 0.444 | 0.315 | 17.370 | 0.437 | 0.312 | 16.245 | 0.457 | 0.327 | 25.950 | 17.118 | 1.257 |
| U | EnlightenGAN Jiang et al. (2021) | TIP'21 | 0.01 | 18.489 | 0.672 | 0.311 | 18.680 | 0.673 | 0.301 | 17.081 | 0.471 | 0.327 | 24.714 | 18.976 | 0.967 |
| U | CLIP-LIT Liang et al. (2023) | ICCV'23 | 0.01 | 14.818 | 0.524 | 0.371 | 15.180 | 0.529 | 0.369 | 13.484 | 0.405 | 0.353 | 28.140 | 26.042 | 1.864 |
| U | NeRCo Yang et al. (2023) | ICCV'23 | 0.02 | 21.073 | 0.725 | 0.259 | 19.660 | 0.717 | 0.271 | 17.381 | 0.530 | 0.518 | 28.131 | 30.017 | 0.836 |
| U | PairLIE Fu et al. (2023) | CVPR'23 | 0.01 | 19.700 | 0.774 | 0.235 | 19.880 | 0.778 | 0.234 | 17.610 | 0.519 | 0.331 | 26.796 | 31.042 | 1.641 |
| U | ZeroIG$_{\text{LOL}}$ Shi et al. (2024) | CVPR'24 | 0.01 | 18.659 | 0.750 | 0.242 | 18.130 | 0.746 | 0.248 | 16.443 | 0.509 | 0.376 | 27.578 | 26.223 | 1.486 |
| U | QuadPrior Wang et al. (2024) | CVPR'24 | 0.05 | 20.302 | 0.810 | 0.203 | 20.300 | 0.791 | 0.186 | 16.947 | 0.560 | 0.382 | 24.437 | 18.174 | 0.422 |
| U | GEFU Wang et al. (2025) | ICCV'25 | 0.01 | 20.831 | 0.790 | 0.208 | 21.320 | 0.807 | 0.193 | 18.750 | 0.542 | 0.281 | 24.118 | 17.806 | 0.635 |
| Z | ExCNet Zhang et al. (2019a) | ACM MM'19 | 0.17 | 16.297 | 0.459 | 0.375 | 15.500 | 0.574 | 0.410 | 15.702 | 0.410 | 0.338 | 27.393 | 19.242 | 1.823 |
| Z | RRDNet Zhu et al. (2020) | ICME'20 | 0.23 | 13.572 | 0.479 | 0.324 | – | – | – | 13.427 | 0.392 | 0.336 | 26.657 | 17.958 | 1.157 |
| Z | GDP Fei et al. (2023) | CVPR'23 | 8.56 | 14.663 | 0.504 | 0.356 | 14.360 | 0.630 | 0.364 | 13.068 | 0.392 | 0.448 | 28.644 | 27.014 | 0.528 |
| Z | TAO Gou et al. (2024) | ICML'24 | 1.57 | 19.181 | 0.607 | 0.390 | – | – | – | 15.689 | 0.431 | 0.702 | 42.085 | 42.138 | 0.384 |
| Z | CoLIE Chobola et al. (2024) | ECCV'24 | 0.02 | 14.900 | 0.499 | 0.327 | 15.787 | 0.497 | 0.340 | 14.001 | 0.405 | 0.342 | 26.819 | 18.968 | 0.964 |
| Z | FourierDiff Lv et al. (2024) | CVPR'24 | 0.82 | 16.953 | 0.604 | 0.293 | 18.708 | 0.783 | 0.303 | 15.625 | 0.461 | 0.321 | 25.927 | 26.569 | 1.221 |
| Z | LLIEDiff Huang et al. (2025) | AAAI'25 | 10.12 | 19.680 | 0.750 | 0.263 | 18.049 | 0.723 | 0.343 | 16.980 | 0.564 | 0.420 | 36.229 | 46.079 | 0.487 |
| Z | MR. Illuminate* Cho et al. (2026) | CVPR'26 | 0.12 | 19.739 | 0.715 | 0.177 | 20.338 | 0.610 | 0.247 | 17.211 | 0.473 | 0.307 | 25.421 | 15.417 | 0.434 |
| | **Ours** | – | 0.01 | **19.747** | **0.793** | 0.201 | **20.951** | **0.825** | **0.224** | **18.107** | **0.525** | **0.303** | **25.404** | 20.703 | **0.403** |

∗ denotes our reproduction using the original Stable Diffusion VAE.

**Implementation Details.** We implement our method using PyTorch and conduct all experiments on a single NVIDIA L20 GPU. For inference, we utilize a pre-trained Stable Diffusion v1.5 model with all parameters fixed. To generate diffusion references, we employ a six-step DDIM sampling schedule with an image-to-image strength of 0.35, which yields only two valid reverse diffusion updates. We adopt an empty text prompt and produce one diffusion reference for each input image. Specifically, the step size for measurement guidance and the coefficient for structural guidance are set to 0.12 and 0.35, respectively. For local photometric estimation, the window radius is adaptively configured to 0.08 times the shorter dimension of the input image. Implementation details for the auto white balance evaluation are provided in the Appendix.

**Quantitative Results.** Table 1 compares our method with supervised, unsupervised, and zero-shot LLIE approaches. Among zero-shot methods, our approach achieves a favorable fidelity and perception trade-off across the paired benchmarks while maintaining high inference efficiency.

**Qualitative Results.** As shown in Fig. 3, existing methods may suffer from over-enhancement, residual exposure bias, or color drift. In contrast, our results preserve input-consistent structures while producing illumination and color closer to the ground truth.



(a) Input (b) SNR-Aware (c) QuadPrior (d) FourierDiff (e) LLIEDiff (f) Mr.Illuminate (g) Ours (h) GT

Figure 3: **Qualitative comparison on the LOL benchmark.** Existing approaches may suffer from residual under-/over-enhancement or noticeable color shifts, whereas our method produces more balanced illumination and color while preserving input-consistent scene structures.

Table 2: Quantitative comparison on the CUBE+ dataset. We report the average CIE76 color difference ($\Delta E$), mean angular error (MAE), and mean squared error (MSE). Lower values indicate better performance.

| | Method | Train Data | $\Delta E \downarrow$ | MAE $\downarrow$ | MSE $\downarrow$ |
|---|---|---|---|---|---|
| S | mixedillWB Afifi et al. (2022) | RenderedWB | 7.24 | 4.16° | 115.25 |
| | WB-sRGB Afifi et al. (2019) | NUS, Gehler | 7.81 | 4.04° | 284.66 |
| | Quasi-CC Bianco & Cusano (2019) | Flickr100k | 24.26 | 6.24° | 3185.06 |
| | Quasi-CC Bianco & Cusano (2019) | ILSVRC12 | 24.12 | 6.13° | 3179.82 |
| | Quasi-CC Bianco & Cusano (2019) | Places365 | 24.26 | 6.21° | 3194.27 |
| Z | DDNM Wang et al. (2023b) | n/a | 45.73 | 18.97° | 6146.39 |
| | TAO Gou et al. (2024) | n/a | 16.84 | 7.48° | 1133.40 |
| | MR. Illuminate Cho et al. (2026) | n/a | 13.41 | **5.75**° | 593.46 |
| | **Ours** | n/a | **12.97** | 7.24° | **396.60** |

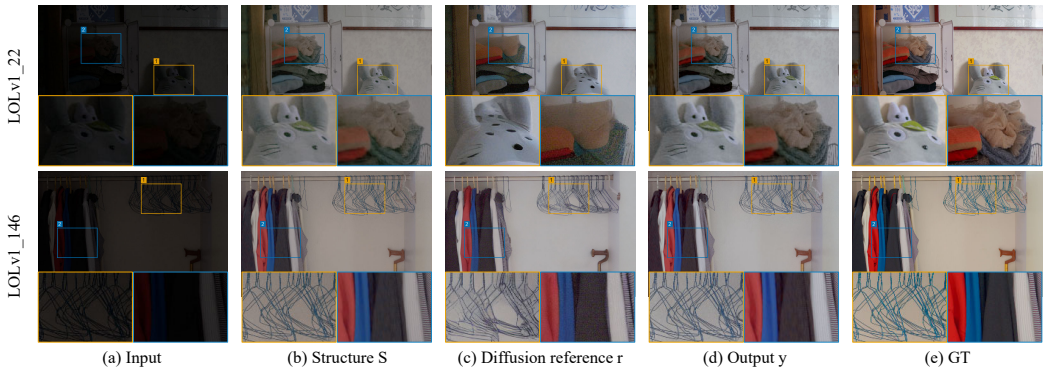



Figure 4: **Visualization of intermediate representations.** The zoomed regions show that **S** preserves input-consistent geometry and fine structures, whereas **r** provides useful normal-light photometric cues but may introduce hallucinated textures, colors, and local structures.

**Analysis on Challenging Cases.** Notably, our method still faces challenges under such extreme underexposure conditions. As shown in Fig. 5, while it recovers the main scene structure and improves illumination and color, some discrepancy from the ground truth remains. This is mainly because severely degraded observations provide limited structural and color cues, which fundamentally affect both the observation-derived structure estimate and the photometric statistics inferred from the diffusion reference.




Figure 5: Qualitative analysis of a severely underexposed LOLv1 example. Our constrained photometric projection exploits diffusion-derived illumination and color cues while preserving input-consistent structure, resulting in conservative enhancement in extremely dark regions.

## 5 Ablation Study

**Component Ablation.** We evaluate the contribution of each component in Table 3. Directly using the diffusion output causes clear fidelity degradation, while projecting it into a global photometric transformation substantially improves restoration quality. Unconstrained spatial modulation introduces instability, which is progressively alleviated by confidence-based shrinkage and the global-anchored trust region. The full model performs best overall, supporting our design of restricting diffusion to a compact, constrained photometric space.

**Sensitivity Analysis.** As shown in Table 4, performance is largely insensitive to the number of diffusion references. Adding more references brings little benefit, while overly fine spatial grids slightly degrade the results, likely due to less reliable local statistics. Accordingly, we use one reference and a grid size of 8 throughout the experiments.

Table 3: Component ablation on LOLv1. ↑ (↓) indicates higher (lower) is better.

| Variant | D | G | L | Sh | TR | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|---|---|---|---|---|
| Structure $\mathbf{S}$ | × | × | × | × | × | 17.943 | 0.779 | 0.213 |
| Direct Diff. | ✓ | × | × | × | × | 17.729 | 0.543 | 0.397 |
| Global | ✓ | ✓ | × | × | × | 18.853 | 0.779 | 0.207 |
| Spatial | ✓ | ✓ | ✓ | × | × | 18.114 | 0.771 | 0.247 |
| Spatial+Shrink | ✓ | ✓ | ✓ | ✓ | × | 18.953 | 0.780 | 0.208 |
| **Full** | ✓ | ✓ | ✓ | ✓ | ✓ | **19.747** | **0.793** | **0.201** |

*D*: diffusion reference; *G*: global photometry; *L*: local photometry; *Sh*: confidence shrinkage; *TR*: global-anchored trust region.

Table 4: Sensitivity analysis on LOLv1. ↑ (↓) indicates higher (lower) is better.

| Setting | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| *Number of References* | | | |
| $K = 1$ | 19.747 | **0.793** | **0.201** |
| $K = 2$ | **19.750** | 0.792 | 0.202 |
| $K = 4$ | 19.746 | 0.792 | 0.201 |
| *Spatial Grid Size* | | | |
| Size = 8 | **19.747** | **0.793** | **0.201** |
| Size = 16 | 19.744 | 0.792 | 0.202 |
| Size = 32 | 19.725 | 0.792 | 0.203 |

Table 5: Ablation of photometric reference sources on LOL dataset. The projection pipeline is fixed across variants.

| Reference Source | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| Anchor $A$ | 17.958 | 0.778 | 0.213 |
| VAE $D(E(A))$ | 17.922 | 0.779 | 0.210 |
| Classical | 18.020 | 0.778 | 0.215 |
| Random Stats. | 18.169 | 0.778 | 0.219 |
| SD Reference | **19.747** | **0.793** | **0.201** |

Table 6: Structural fidelity on LOL dataset. Higher values indicate better consistency.

| Metric ↑ | Direct | Projection |
|---|---|---|
| Multi-scale HF Corr. | 0.388 | **0.827** |
| Gradient Dir. Cons. | 0.459 | **0.848** |
| Sobel-GMS | 0.934 | **0.976** |
| SSIM | 0.543 | **0.793** |

**Reference Source.** To examine whether the gain comes from diffusion itself or simply from the projection process, we vary only the reference source while keeping the remaining pipeline fixed. The anchor, VAE reconstruction, and hand-crafted or randomly sampled photometric statistics perform similarly, whereas the SD reference gives consistently better results (Table 5). This supports the use of diffusion as a source of normal-light photometric cues. Implementation details of these reference variants are provided in Appendix B.2.

**Structural Fidelity.** We further assess whether the photometric bottleneck better preserves observation-consistent structure than direct diffusion reconstruction. We report multi-scale high-frequency correlation, gradient-direction consistency, Sobel-GMS, and SSIM, covering both local structural details and overall similarity. As shown in Table 6, our projection consistently improves all four metrics, indicating reduced structural drift.

## 6 Conclusion

In this paper, we have presented a novel structure-photometry decoupled framework for zero-shot LLIE. The proposed method rethinks diffusion in fidelity-sensitive restoration by using a pretrained diffusion model as a constrained photometric prior rather than a pixel-level image reconstructor. By anchoring scene structures to the observed image and selectively transferring reliable illumination and color statistics from diffusion, our approach improves low-light visibility while minimizing content drift. Extensive experiments on paired and unpaired benchmarks demonstrate competitive enhancement performance, improved structural fidelity, and efficient inference. Results on automatic white balance further support the broader potential of converting generative priors into constrained photometric priors for image restoration.

## AI USE STATEMENT

In this study, generative AI tools were used only as auxiliary aids for improving the presentation of the manuscript and supporting software development. Specifically, they were employed to refine wording, improve grammatical correctness and readability, and assist with debugging portions of the experimental code. Generative AI was not used for any activities that fall under the mandatory disclosure categories defined by the ICLR 2027 AI Policy for Authors.

All AI-assisted revisions were carefully checked by the authors before being incorporated into the final manuscript. Textual changes were verified to ensure consistency with the intended technical content and to avoid introducing inaccurate statements or unsupported conclusions. Any code changes suggested with AI assistance were independently reviewed, tested, and validated through rerunning the associated experiments. The authors remain fully responsible for the manuscript, methodology, implementation, experimental results, and all other materials included in this submission.

## ETHICS STATEMENT

This work focuses on low-light image enhancement and does not involve human-subject studies, the collection of personal data, or the release of new datasets. All experiments are conducted on publicly available image benchmarks and pre-trained models in accordance with their respective licenses and usage terms. The proposed method is intended for general-purpose image restoration and enhancement. While improved visibility in low-light imagery may benefit applications such as photography and visual perception, similar techniques could potentially be used in privacy-sensitive or surveillance-related settings. We therefore encourage responsible use of the method in compliance with applicable privacy, security, and legal requirements. We are not aware of any additional ethical concerns specific to the methodology presented in this work.

## REPRODUCIBILITY STATEMENT

We detail the model architecture, dataset splits, and preprocessing pipelines. All experiments use standard public datasets, and we provide the exact hyperparameters and complete experimental configurations in the Appendix.

## REFERENCES


Mahmoud Afifi, Brian Price, Scott Cohen, and Michael S Brown. When color constancy goes wrong: Correcting improperly white-balanced images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1535–1544, 2019.

Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. Auto white-balance correction for mixed-illuminant scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 934–943, 2022.

Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12204–12213, 2019.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12470–12479, 2023.

Jiezhang Cao, Yue Shi, Kai Zhang, Yulun Zhang, Radu Timofte, and Luc Van Gool. Deep equilibrium diffusion restoration with parallel sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2824–2834, 2024.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3291–3300, 2018.

Joshua Cho, Sara Aghajanzadeh, Zhen Zhu, and David Forsyth. MR. Illuminate: Zero-shot low-light image enhancement with diffusion prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8460–8470, 2026.

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

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Zheng Ding, Xuaner Zhang, Zhuowen Tu, and Zhihao Xia. Restoration by generation with constrained priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2567–2577, 2024.

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9935–9946, 2023.

Daniel Feijoo, Juan C Benito, Alvaro Garcia, and Marcos V Conde. DarkIR: Robust low-light image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10879–10889, 2025.

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 (CVPR)*, pp. 22252–22261, 2023.

Tomer Garber and Tom Tirer. Image restoration by denoising diffusion models with iteratively preconditioned guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25245–25254, 2024.

Yuanbiao Gou, Haiyu Zhao, Boyun Li, Xinyan Xiao, and Xi Peng. Test-time degradation adaptation for open-set image restoration. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 16167–16177. PMLR, 2024.

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 (CVPR)*, pp. 1777–1786, 2020.

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.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851, 2020.

Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 79734–79747, 2023.

Yan Huang, Xiaoshan Liao, Jinxiu Liang, Yuhui Quan, Boxin Shi, and Yong Xu. Zero-shot low-light image enhancement via latent diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 3815–3823, 2025.

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

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:23593–23606, 2022.

Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21277–21286, 2025.

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.

Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9396–9416, 2021a.

Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2021b.

Jinlong Li, Baolu Li, Zhengzhong Tu, Xinyu Liu, Qing Guo, Felix Juefei-Xu, Runsheng Xu, and Hongkai Yu. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15205–15215, 2024.

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 (ICCV)*, pp. 8060–8069, 2023.

Jingxi Liao, Shijie Hao, Richang Hong, and Meng Wang. GT-Mean loss: A simple yet effective solution for brightness mismatch in low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6112–6121, 2025.

Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Wenxue Li, Lei Zhu, and Xinghao Ding. AGLLDiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 5307–5315, 2025.

Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, and Liangqiong Qu. Residual denoising diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2773–2783, 2024a.

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 (CVPR)*, pp. 10556–10565, 2021.

Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-Plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4197–4208, 2024b.

Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25378–25388, 2024.

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 (CVPR)*, pp. 5627–5636, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, 2022.

Yiqi Shi, Duo Liu, Liguo Zhang, Ye Tian, Xuezhi Xia, and Xiaojing Fu. ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3015–3024, 2024.

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

Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, et al. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13804–13814, 2025.

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, Bjorn 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 (AAAI)*, pp. 2654–2662, 2023a.

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 (CVPR)*, pp. 26057–26066, 2024.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations (ICLR)*, 2023b.

Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2604–2612, 2022.

Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. Exposurediffusion: Learning to expose for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12404–12414, 2023c.

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

Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5891–5900, 2022.

Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1662–1671, 2023.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. DiffIR: Efficient diffusion model for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13049–13059, 2023.

Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2278–2287, 2020.

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 (CVPR)*, pp. 17693–17703, 2022.

Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9893–9903, 2023.

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 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 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 (ICCV)*, pp. 12872–12881, 2023.

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.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-Retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12268–12277, 2023.

Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM)*, pp. 1623–1631, 2019a.

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 (ACM MM)*, pp. 1632–1640, 2019b.

Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. Empowering low-light image enhancer through customized learnable priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12525–12535, 2023.

Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. In *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2020.

# Contents


**A Additional Implementation Details** 15

A.1 Internal Structure Reconstruction . . . . . 15

A.2 Local Confidence for Spatial Photometry . . . . . 16

A.3 Spatial Trust Region . . . . . 16

**B Further Ablation Study** 16

B.1 Component Variants . . . . . 16

B.2 Photometric Reference Sources . . . . . 16

B.3 Photometric Error Analysis . . . . . 17

**C Auto White Balance** 17

C.1 Extension to Auto White Balance . . . . . 17

C.2 Additional Results on Auto White Balance . . . . . 18

**D Additional LLIE Qualitative Comparisons** 19


## A Additional Implementation Details

### A.1 Internal Structure Reconstruction

The internal reconstruction $P(\cdot)$ suppresses noise amplified by the initial photometric correction while preserving structures supported by the observation. It consists of noise estimation, noise propagation, and patch-domain reconstruction.

**Noise Estimation.** We estimate noise statistics from locally smooth regions of the input image. The image is divided into non-overlapping patches, from which constant and planar components are removed. Patches with the lowest residual energy are retained as candidate noise regions. For each color channel $c$, a Hann window is applied before computing the Fourier spectrum, and the channel-wise noise variance is estimated from the averaged spectral power:

$$\sigma_c^2 = \mathrm{mean}_\omega \left[ \frac{1}{N} \sum_{i=1}^{N} \left| \mathcal{F}\left(w \odot e_{i,c}\right)(\omega) \right|^2 \right], \tag{21}$$

where $e_{i,c}$ denotes the residual of the $i$-th selected patch and $w$ is a Hann window. The spectrum is lightly smoothed before averaging to reduce estimation variance.

Because the anchor $\mathbf{A}$ is obtained through a nonlinear photometric transform, its noise statistics differ from those of the input. We therefore propagate the estimated noise through the local response of the fixed transform:

$$\widetilde{\sigma}_c^2(p) \approx J_c^2(p)\sigma_c^2, \qquad J_c(p) = \frac{\partial A_c(p)}{\partial x_c(p)}. \tag{22}$$

The illumination and tone parameters are held fixed when computing $J_c(p)$, such that the Jacobian captures only the local noise amplification induced by the photometric correction.

**Patch Reconstruction.** Overlapping patches are then extracted from the anchor and reconstructed in the frequency domain using the propagated noise statistics. Spectral components that are weak relative to the estimated noise are attenuated, while reliable components are preserved:

$$\widehat{q}_{i,c} = \mathcal{F}^{-1}\left[\mathcal{T}\left(\mathcal{F}(q_{i,c}); \widetilde{\sigma}_{i,c}^2\right)\right], \tag{23}$$

where $\mathcal{T}(\cdot)$ denotes the noise-adaptive spectral shrinkage operator,

$$\mathcal{T}\left(Z_{i,c}(\omega); \widetilde{\sigma}_{i,c}^2\right) = \left[1 - \frac{\widetilde{\sigma}_{i,c}^2}{|Z_{i,c}(\omega)|^2 + \epsilon}\right]_+ Z_{i,c}(\omega), \tag{24}$$

with $Z_{i,c} = \mathcal{F}(q_{i,c})$, $[\cdot]_+ = \max(\cdot, 0)$, and a small $\epsilon$ for numerical stability. The reconstructed overlapping patches are finally aggregated by weighted averaging to obtain $\mathbf{S} = P(\mathbf{A})$.

### A.2 Confidence Weight for Spatial Photometry

The confidence weight $w^c(p)$ is determined solely from the local variance of the structure estimate $\mathbf{S}$ and does not require semantic features or an additional learned predictor. Before computing local statistics, $\mathbf{S}$ and the diffusion reference $\mathbf{r}$ are mildly smoothed, and the statistics window is scaled with the input resolution.

Let $v^c(p)$ denote the local variance of channel $c$ in the smoothed structure estimate. The confidence weight is defined as

$$w^c(p) = \frac{v^c(p)}{v^c(p) + \kappa}, \tag{25}$$

where $\kappa$ is a fixed regularization constant.

Low-texture or weak-signal regions therefore receive smaller confidence and rely more strongly on the global photometric solution, whereas textured regions permit greater spatial adaptation:

$$\bar{a}^c(p) = a_g^c + w^c(p)\left(\widetilde{a}^c(p) - a_g^c\right), \tag{26}$$

$$\bar{b}^c(p) = b_g^c + w^c(p)\left(\widetilde{b}^c(p) - b_g^c\right). \tag{27}$$

### A.3 Trust-Region Width

Confidence shrinkage reduces the influence of unreliable local estimates but does not explicitly bound their magnitude. We therefore constrain the spatial gain and bias around the global photometric solution:

$$a^c(p) = \mathrm{proj}_{[(1-\delta_a)a_g^c,\,(1+\delta_a)a_g^c]}\left(\bar{a}^c(p)\right), \tag{28}$$

$$b^c(p) = \mathrm{proj}_{[b_g^c-\delta_b,\,b_g^c+\delta_b]}\left(\bar{b}^c(p)\right). \tag{29}$$

Here, $\delta_a$ and $\delta_b$ are fixed trust-region widths shared across all images. They are not adapted according to image brightness or content, ensuring that local photometric corrections remain bounded around the global solution.

## B Further Ablation Study

In this section, we provide additional details for the ablation studies reported in the main paper. Unless otherwise noted, the input image, structure branch, diffusion backbone, photometric projection, and evaluation protocol are kept unchanged.

### B.1 Component Variants

The component ablation in Table 3 of the main paper follows the same inference pipeline and changes only the photometric readout. *Structure* $\mathbf{S}$ returns the observation-derived structure without diffusion-based photometric correction. *Direct Diff.* directly uses the decoded diffusion result as the output. *Global* uses only the image-level RGB correction. *Spatial* adds local photometric estimation without confidence shrinkage. *Spatial+Shrink* further reduces unreliable local deviations toward the global solution. *Full* additionally applies the global-anchored spatial trust region.

### B.2 Photometric Reference Sources

Table 5 of the main paper keeps the projection pipeline fixed and changes only the source of photometric information. *Anchor* directly uses the photometric anchor from the structure branch. *VAE* uses the reconstruction obtained by passing the anchor through the frozen VAE without diffusion sampling. *SD Reference* uses the short guided diffusion trajectory described in Section A.2.

*Classical* applies a deterministic exposure and white-balance correction to the anchor. Specifically, it performs global gamma correction toward a target luminance, with the gamma constrained to a fixed range, followed by gray-world white balancing with bounded per-channel gains.

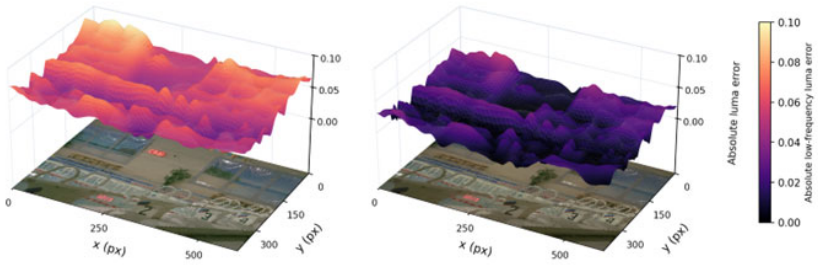
*Random Stats.* samples a normal-light image $\mathbf{D}$ from the disjoint LOLv1 training split using a fixed seed and transfers only its channel-wise first- and second-order statistics to the anchor:

$$\mathbf{R}_c = \big(\mathbf{A}_c - \mu_c(\mathbf{A})\big)s_c + \mu_c(\mathbf{D}), \qquad s_c = \frac{\sigma_c(\mathbf{D})}{\sigma_c(\mathbf{A})}. \tag{30}$$

where $c$ indexes the RGB channels and $s_{\min}$ and $s_{\max}$ denote fixed bounds on the statistic-transfer scale.

### B.3 Photometric Error Analysis.

Low-frequency luminance errors reveal the effect of photometric projection. As shown in Fig. 6, the structure estimate $\mathbf{S}$ retains noticeable illumination bias, whereas the final output shows a flatter error surface. This suggests that the diffusion prior mainly corrects large-scale photometric deviations while leaving fine structures to $\mathbf{S}$.




(a) Structure S (b) Output y

Figure 6: **Low-frequency luminance error before and after photometric projection.** (a) Structure $\mathbf{S}$ and (b) final output $\mathbf{y}$. Surface height and color denote absolute low-frequency luminance error. The final output shows lower and more uniform error, indicating effective correction of large-scale illumination deviations.

## C Auto White Balance

### C.1 Extension to Auto White Balance

For auto white balance (AWB), we preserve the structure–photometry decomposition and replace the low-light anchor with a deterministic white-balance anchor.

**White-Balance Anchor.** The input is first converted from sRGB to linear RGB. To avoid unreliable illuminant estimates, very dark and near-saturated pixels are excluded using fixed lower and upper intensity thresholds:

$$\Omega = \left\{ p \,\middle|\, \tau_{\text{dark}} < \max_c x_c(p) < \tau_{\text{sat}} \right\}, \tag{31}$$

We estimate the scene illuminant using Shades-of-Gray with Minkowski norm $p = 5$:

$$e_c = \left( \frac{1}{|\Omega|} \sum_{p \in \Omega} x_{\text{lin},c}^5(p) \right)^{1/5}. \tag{32}$$

The illuminant is normalized to unit channel mean and bounded to a fixed range. The white-balance anchor is then obtained by channel-wise correction, followed by conversion back to sRGB:

$$A_{\text{lin},c}(p) = \frac{x_{\text{lin},c}(p)}{e_c}, \tag{33}$$

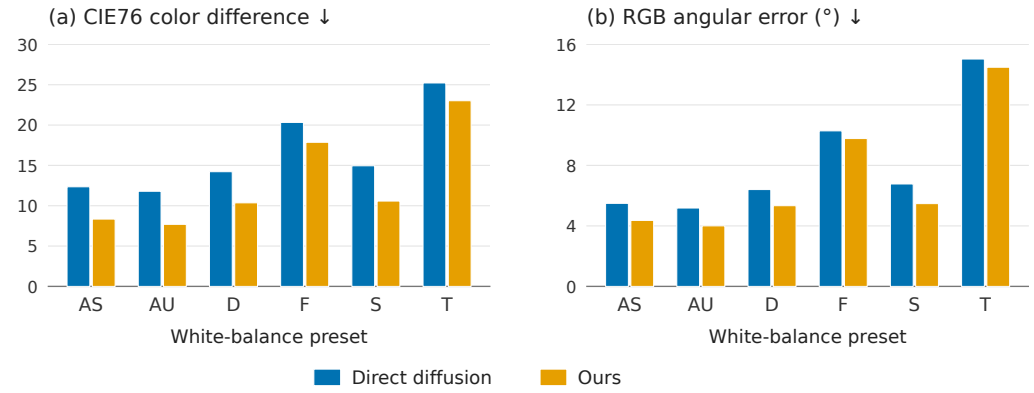



Figure 7: Comparison of direct diffusion and our global photometric readout across six white-balance presets on rendered CUBE+. Each preset contains the same 1,707 scenes. We report mean CIE76 color difference (left) and mean pixel-wise RGB angular error in degrees (right); lower is better for both metrics. Our readout reduces both errors relative to direct diffusion in every preset.

**Photometric Readout.** The AWB anchor is processed by the same noise-aware internal reconstruction $P(\cdot)$ to obtain the structure estimate $\mathbf{S}$. A single photometric reference $\mathbf{r}$ is generated using the frozen Stable Diffusion backbone.

As AWB mainly requires global chromatic correction, we use the global RGB gain–bias model:

$$\mathbf{r} \approx \mathbf{a}_g \odot \mathbf{S} + \mathbf{b}_g. \tag{34}$$

For each channel,

$$a_g^c = \frac{\mathrm{Cov}(S^c, r^c)}{\mathrm{Var}(S^c) + \lambda_g}, \qquad b_g^c = \mu_r^c - a_g^c \mu_S^c. \tag{35}$$

The final output is rendered directly from the structure estimate:

$$\widehat{\mathbf{y}} = \mathbf{a}_g \odot \mathbf{S} + \mathbf{b}_g. \tag{36}$$

Therefore, for AWB, only the photometric anchor and readout are changed, with diffusion still used solely as a photometric reference.

### C.2 Additional Results on Auto White Balance

Table 7: Comparison of intermediate and final outputs on rendered CUBE+, averaged over 10,242 inputs from 1,707 scenes. The spatial variant removes the relative bounds around the global fit while retaining the absolute parameter limits. Lower is better for all metrics.

| Output | $\Delta E_{76} \downarrow$ | MAE (°)↓ | $\mathrm{MSE}_{\mathrm{WB}} \downarrow$ |
|---|---|---|---|
| White-balanced input | 16.800 | 9.788 | 609.205 |
| Direct diffusion reference | 16.475 | 8.193 | 880.286 |
| Spatial readout without relative bounds | 13.351 | 7.294 | 426.314 |
| **Ours (global photometric readout)** | **12.975** | **7.239** | **396.601** |

Table 7 compares different readout strategies on rendered CUBE+. Directly using the diffusion reference reduces color difference and angular error relative to the input, but increases MSE, indicating that improved color statistics do not necessarily translate into pixel-level fidelity. Projecting the reference into photometric parameters substantially reduces all three errors. The global readout achieves the lowest errors and slightly improves over the spatial variant without relative bounds, supporting the use of global correction for AWB, where the dominant variation is chromatic rather than spatial.

Figure 7 compares the diffusion reference with our final photometric readout across six white-balance presets. The improvement is consistent across all presets, suggesting that the gain does not arise from a particular input color cast. Instead, the photometric readout systematically regularizes the reference toward a more faithful global chromatic correction. Preset T remains the most difficult case, indicating that when the reference itself carries a stronger color bias, the global readout can reduce but not completely remove that error. Overall, the comparison shows that diffusion is more reliable as a source of photometric statistics than as the final image prediction for AWB.

Figure 8 further illustrates how the photometric readout operates across different input color casts. The diffusion reference generally moves the image toward a plausible normal-light appearance, but its local colors and fine structures are not always consistent with the target. In contrast, the final output remains visually aligned with the structure estimate $\mathbf{S}$ while correcting its dominant chromatic bias. This behavior is particularly clear in text, object boundaries, and small foreground details, which are preserved in the output even when the diffusion reference exhibits noticeable appearance changes.

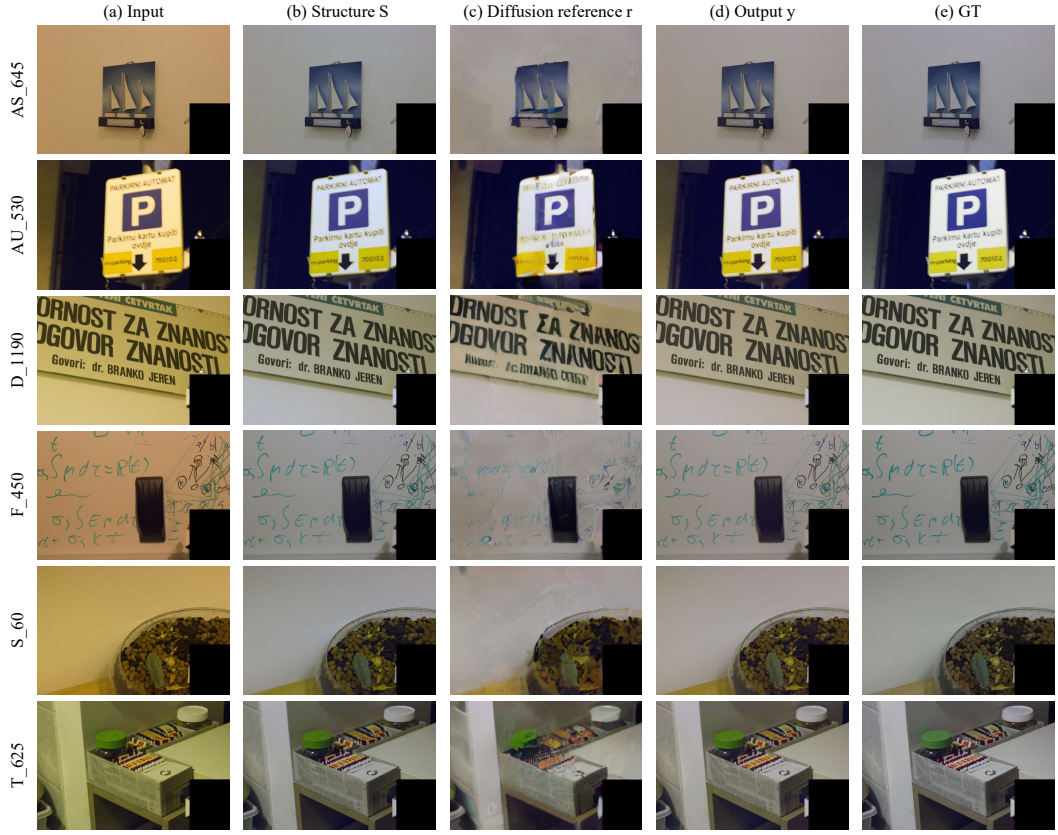



Figure 8: Qualitative AWB results on CUBE+ across six white-balance presets. From left to right: input, structure estimate $\mathbf{S}$, diffusion reference $\mathbf{r}$, final output $\mathbf{y}$, and ground truth. The diffusion reference provides useful normal-light color cues but may introduce local appearance deviations. Our photometric readout retains the structure of $\mathbf{S}$ while transferring only the global chromatic correction, producing results that more closely follow the ground-truth appearance.

## D Additional LLIE Qualitative Comparisons

To further evaluate the generalization ability of our method, we present qualitative comparisons on the paired LSRW dataset and the unpaired DICM, NPE, and VV datasets. Across diverse scenes and illumination conditions, our method achieves more balanced exposure, natural color rendition, and clearer structural details than LLIEDiff and MR.Illuminate. In comparison, LLIEDiff frequently introduces overexposure or color shifts, whereas MR. Illuminate tends to leave dark regions insuffi-

ciently enhanced, demonstrating the robustness of our method under challenging real-world lighting conditions.

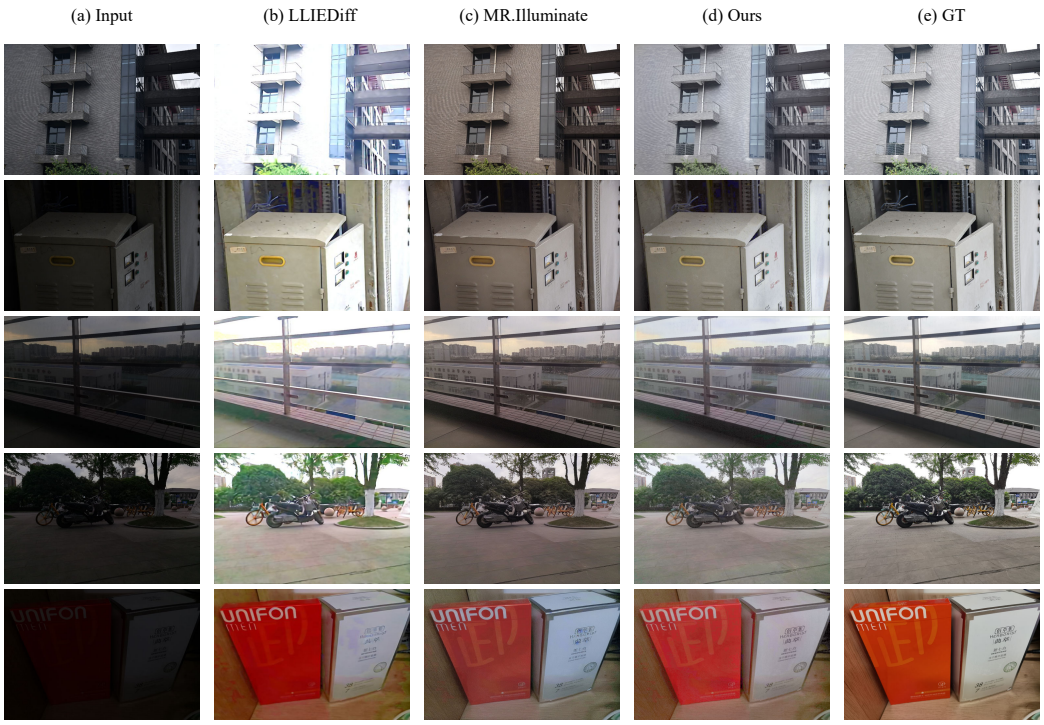



Figure 9: Additional qualitative comparisons on the LSRW dataset.

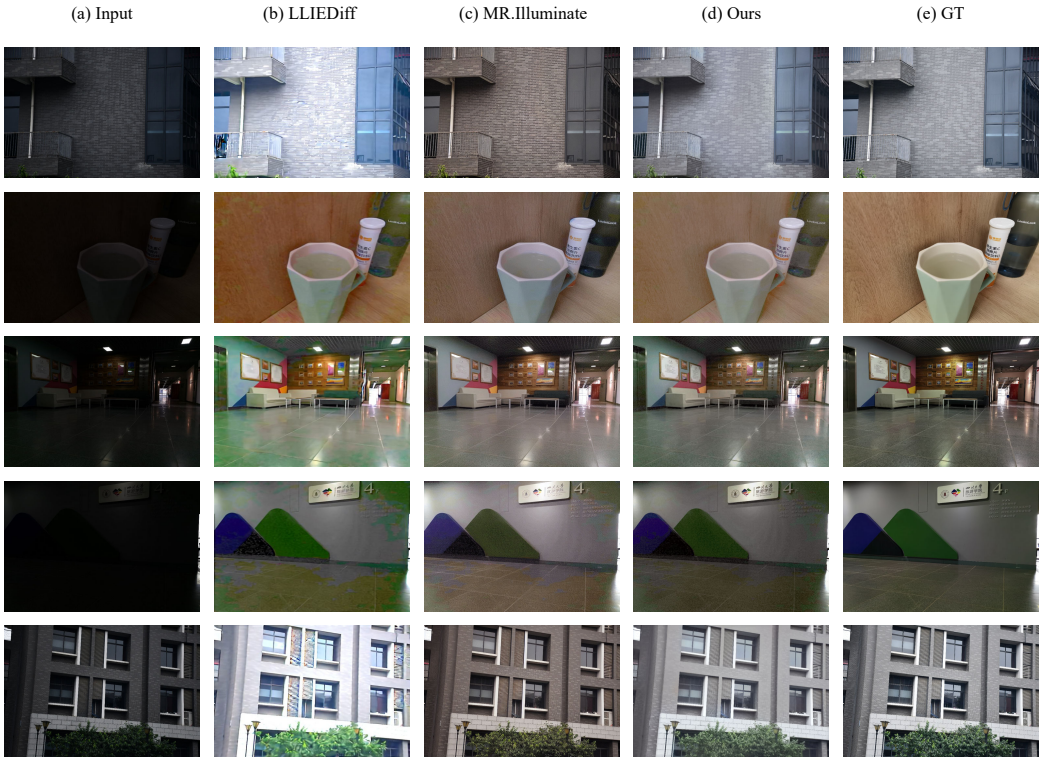



Figure 10: Additional qualitative comparisons on the LSRW dataset.

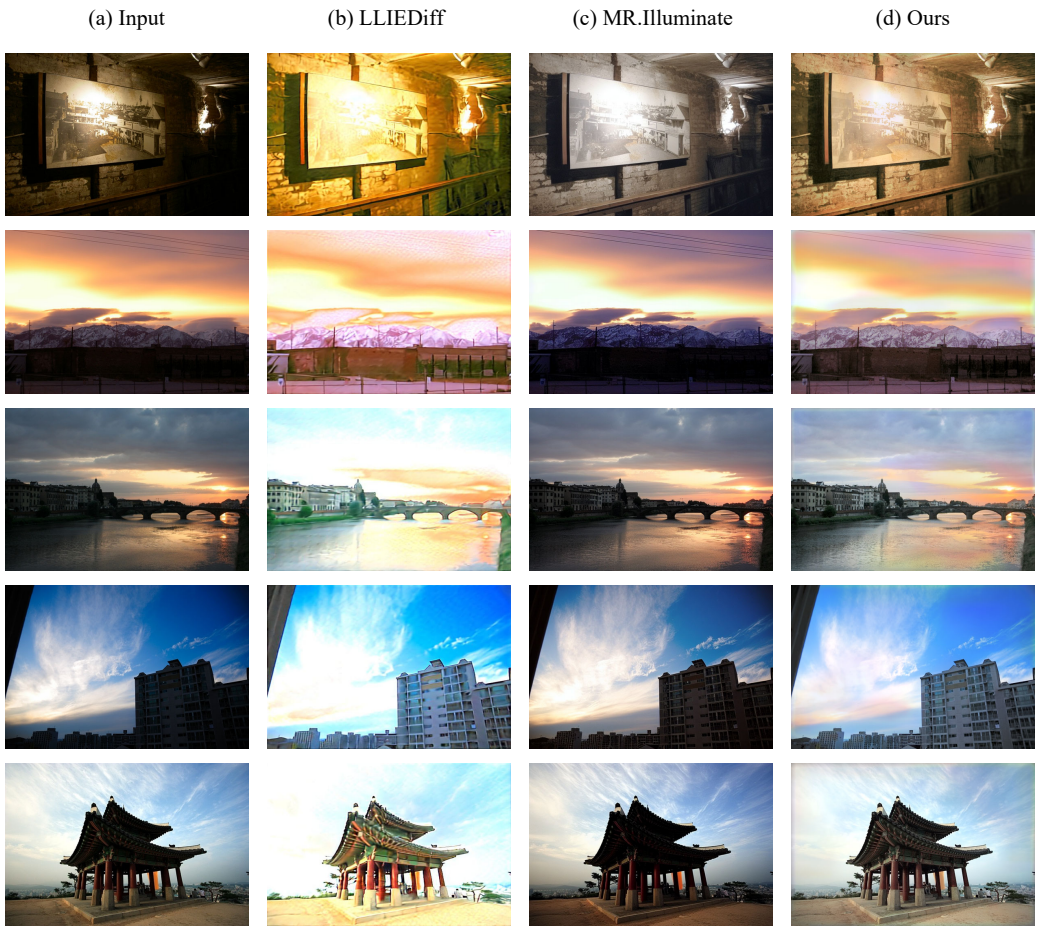



Figure 11: Additional qualitative comparisons on the Unpaired dataset.

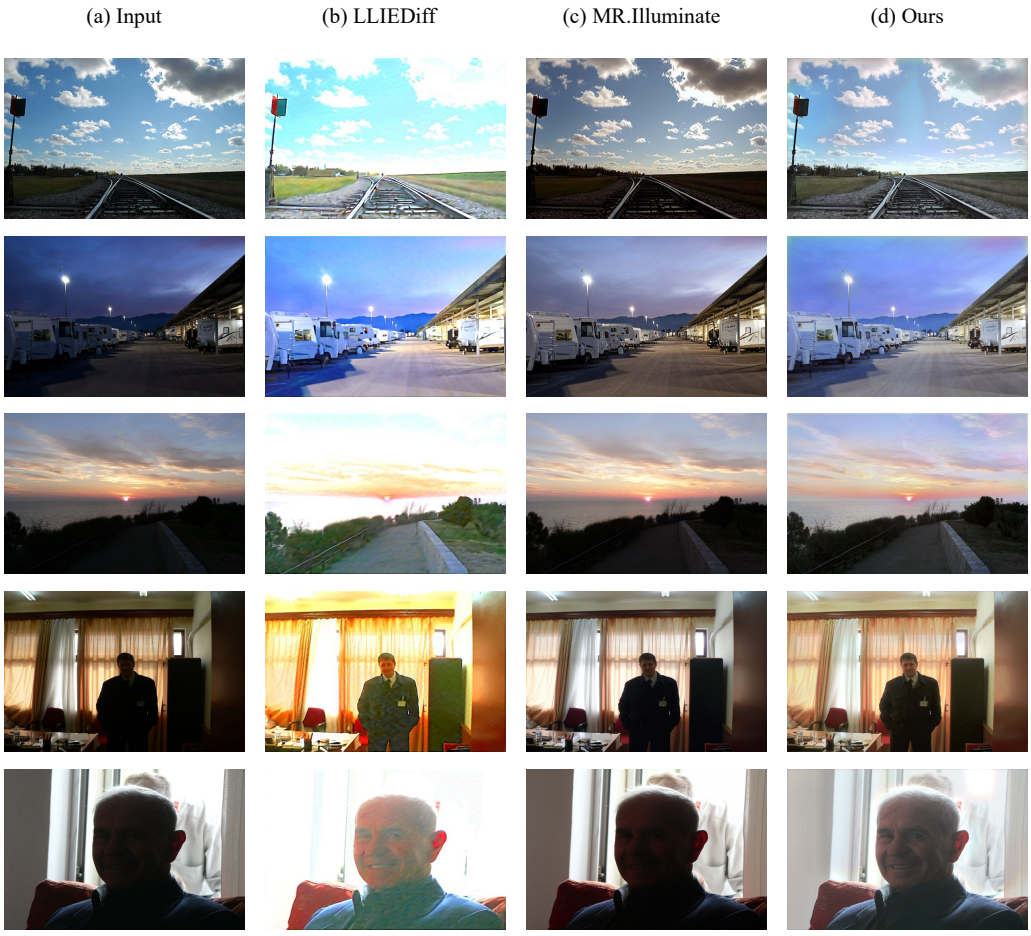



Figure 12: Additional qualitative comparisons on the Unpaired dataset.