# LUCID: Learning Unified Control for Image Deflaring and Exposure Mastery in Nighttime Photography

ANONYMOUS AUTHOR(S)

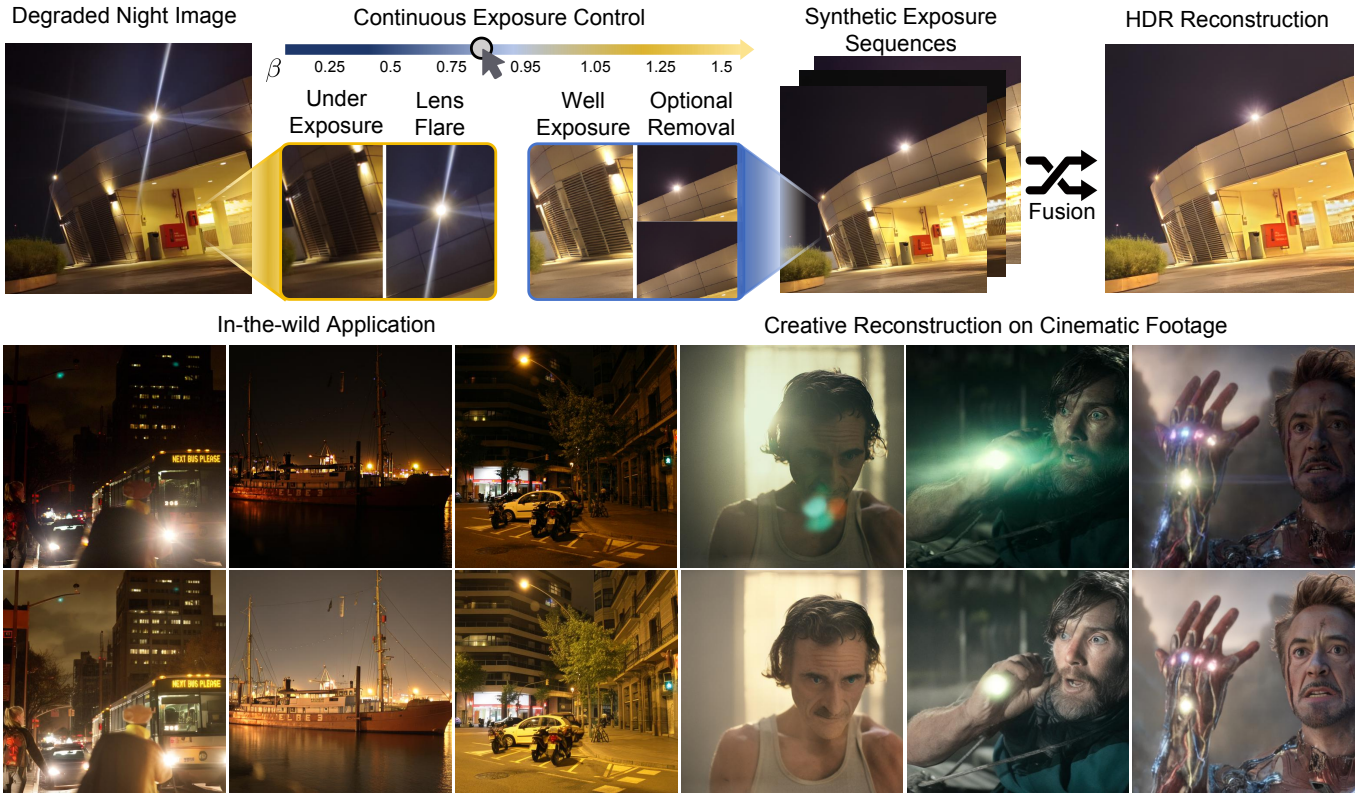


Fig. 1. Unveiling the night with LUCID. We present a unified diffusion framework that jointly addresses severe underexposure and intense lens flares. By enabling continuous exposure modulation from a single view, LUCID facilitates the synthesis of pseudo-exposure sequences for HDR reconstruction (top). Bottom examples demonstrate robust generalization across complex in-the-wild illumination and cinematic footage (images courtesy of ShotDeck).

Photography is the art of painting with light, yet nighttime scenes are shaped by competing degradations: intense flares obscure scene structure, while photon-limited regions collapse into noise. Conventional approaches address these factors in isolation, overlooking the fact that these degradations are fundamentally entangled. To bridge this gap, we introduce LUCID, a unified framework that reframes nighttime restoration as a continuous and controllable process rather than a fixed correction. We decompose nighttime restoration into two cooperative components: a flare disentanglement module that lifts the 'curtain' of optical artifacts to provide reliable structural guidance, and a diffusion-driven module that leverages generative priors to reconstruct clean and well-exposed imagery. Crucially, LUCID introduces explicit controllability through a novel four-mode training strategy, enabling users to steer the restoration process via classifier-free guidance (CFG) and allowing selective control over light sources and their associated flare and ghosting artifacts, while also supporting high dynamic range (HDR) reconstruction through continuous exposure control. Extensive experiments demonstrate that LUCID consistently outperforms state-of-the-art methods across diverse real-world nighttime scenarios.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Do, Not, Use, This, Code, Put, the, Correct, Terms, for, Your, Paper

**ACM Reference Format:**
Anonymous Author(s). 2018. LUCID: Learning Unified Control for Image Deflaring and Exposure Mastery in Nighttime Photography. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation*


*Conference acronym 'XX, Woodstock, NY*

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

*email (Conference acronym 'XX).* ACM, New York, NY, USA, Article 111, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 Introduction

Photography is the art of painting with light. Yet, at night, this canvas becomes notoriously difficult to control. While the human eye effortlessly adapts to the dance between deep shadows and piercing artificial lights, physical sensors struggle profoundly. Limited dynamic range swallows details in the dark, while bright sources spill into overwhelming flare, veiling the scene's true geometry. Admittedly, such artifacts can be stylistically intentional; in nighttime cinematography, deliberate flares often define the mood.

Admittedly, such artifacts can be intentional stylistic choices. In J.J. Abrams' Star Trek (2009), for instance, horizontal anamorphic flares are deliberately embraced to evoke kinetic energy and a futuristic atmosphere. Conversely, maintaining optical purity requires immense physical effort. In the production of The Batman (2022), the cinematography team employed massive physical barriers and custom-built lenses specifically to shield the sensor from stray city lights, fighting to preserve the deep, immersive blacks essential to its noir aesthetic. Fortunately, digital tools have fundamentally reshaped this landscape. Sculpting light in post-production is now routine, granting artists unprecedented creative freedom. Yet, a crucial asymmetry persists: while adding stylized flare to a clean image is straightforward, excavating a pristine signal from a glare-compromised capture is a formidable challenge, as the artifacts overwrite the very information needed for restoration. This motivates our goal: to computationally recover a clean, high-dynamic-range baseline that decouples creative intent from optical accidents, providing a reliable foundation for artistic night photography.

The challenges inherent to this domain, however, are fundamentally structural and intricate. Nighttime degradations are not mere suboptimal exposures, but a profound eclipse of information: intense flare bleeds across the sensor, erasing underlying geometry, while photon-starved regions dissolve into noise and quantization. Critically, restoration becomes a battle between opposing forces. Taming the flare risks extinguishing genuine highlights, while pulling details from the dark inevitably amplifies artifacts and residual ghosting. Thus, nighttime imaging is less a single restoration task and more a delicate balancing act among competing physical constraints.

Most existing methods, however, treat low-light enhancement (LLIE) and flare mitigation as independent problems [Cai et al. 2023; Dai et al. 2022; Feijoo et al. 2025; Jiang et al. 2024], ignoring their physically coupled nature. This decoupling is perilous: simply cascading an LLIE model with a flare-removal method often yields unstable behaviors and severe artifacts (Fig. 2), as the errors from one stage are amplified by the other. Compounding this difficulty is the scarcity of dataset. Acquiring paired nighttime samples in the wild is hindered by the prohibitive difficulty of capturing clean references and the intrinsic ambiguity of ground-truth illumination. Consequently, current datasets largely rely on staged exposure bracketing[Li et al. 2021; Wei et al. 2018], which restricts scene diversity and fails to capture the chaotic illumination patterns of real environments. Finally, most frameworks operate as rigid black boxes, lacking mechanisms for user agency. This prevents users from tailoring brightness or visibility to their needs.

To navigate this complex landscape, we draw inspiration from two pivotal advancements. First, the emergence of synthetic flare datasets, such as Flare7K [Dai et al. 2022], has offered a modular way to represent optical artifacts. By decomposing flare into additive components, it enables the flexible synthesis of training data that covers diverse scattering patterns. Simultaneously, diffusion models have reshaped image restoration [Lin et al. 2024; Wu et al. 2024a,b; Zhang et al. 2024] by providing powerful generative priors. Their innate ability to generate natural details and reconstruct missing structures makes them uniquely suited for creating a clean image from heavily degraded inputs. Nevertheless, how to integrate these generic priors to handle the specific, adversarial constraints of nighttime photography remains an open question.

Building on these insights, we introduce LUCID: a unified framework for continuous flare mitigation and exposure adjustment in nighttime photography (Fig. 3). Instead of treating restoration as a static regression, LUCID decouples the problem into two distinct yet cooperative stages. First, a Flare Disentanglement Module isolates the "curtain" of light artifacts, extracting a structural flare map that serves as a precise guide for the subsequent process. Second, a Diffusion-Driven Restoration Module incorporates reference mixing layers to reorganize this disentangled information, reconstructing a clean, well-exposed image. Crucially, to empower the creator, we introduce a novel four-mode training strategy. By supervising the model with stratified pairings of exposure and light source intensity, we enable continuous control during inference. Through classifier-free guidance (CFG), users can smoothly control the restoration process, adjusting the output from a flare-dominated input to a clean, well-balanced image within a single unified model.

Finally, LUCID redefines the workflow of nighttime photography, supporting fine-grained control ranging from fully removing flare to preserving the natural structure of light sources. Extensive experiments demonstrate that LUCID not only produces visually superior results compared to state-of-the-art (SOTA) baselines but also exhibits robust generalization to diverse real-world scenes. Beyond standard restoration, LUCID naturally extends to High-Dynamic-Range (HDR) reconstruction, recovering faithful luminance transitions from single exposures. By bridging the gap between physical limitations and creative intent, LUCID offers a versatile instrument for both automated enhancement and artistic expression.

## 2 Related Works

### 2.1 Low-Light and Nighttime Image Enhancement

Deep learning has significantly advanced low-light image enhancement (LLIE), evolving from early CNN-based decomposition [Chen Wei 2018; Zhang et al. 2019] to sophisticated restoration frameworks. To bypass reliance on paired data, Zero-DCE[Guo et al. 2020] introduces a zero-reference strategy, learning pixel-wise curves for efficient dynamic range adjustment. Subsequent architectures focus on global context and degradation coupling: Retinexformer[Cai et al. 2023] leverages illumination-Guided transformers to capture long-range dependencies often missed by CNNs, while DarkIR[Feijoo et al. 2025] addresses the practical coupling of low-light and blur, utilizing a specialized structure to disentangle illumination correction from spatial restoration. Pushing perceptual boundaries, Reti-Diff[He et al. 2025] integrates Retinex priors with diffusion models,

synthesizing plausible high-frequency details through generative processes. While these methods excel in recovering visibility, their application to real-world scenes is often undermined by the extreme high dynamic range typical of nighttime photography. Standard approaches tend to amplify pixel intensities, recovering dark details at the cost of clipping high-intensity light sources[Sharma and Tan 2021]. Even methods attempting to balance glare suppression, such as Jin et al.[Jin et al. 2022], lack explicit mechanisms to model the intricate geometric structures of lens flare. Consequently, they struggle to distinguish between scene details and optical artifacts, highlighting the necessity for specialized flare removal techniques.

## 2.2 Nighttime Flare Mitigation

Nighttime flare mitigation presents a formidable challenge due to the high dynamic range of artificial lights and the intricate scattering patterns they produce. While hardware solutions like anti-reflective coatings [Macleod 2010] and fluid-filled lenses [Boynton and Kelley 2003] offer physical suppression, they struggle to fully eliminate artifacts under extreme contrast. Data-driven approaches have thus emerged to address residual artifacts. Wu et al.[Wu et al. 2021] pioneered the use of optical Point Spread Functions (PSF) for semi-synthetic training, a strategy refined by Flare7K[Dai et al. 2022] using real-world statistical priors. To enhance physical realism, Zhou et al.[Zhou et al. 2023] further model ISP and automatic exposure physics to better recover saturated sources. MfdNet[Jiang et al. 2024] advances the field by leveraging frequency domain analysis to disentangle low-frequency glow from scene details. However, intense flares often cause irreversible information loss by completely occluding underlying scene structures. This necessitates the use of generative models capable of inferring missing content.

## 2.3 Diffusion Models

Recent advances have effectively repurposed the rich generative manifold of pre-trained diffusion models for image restoration. Frameworks like DiffBir[Lin et al. 2024] and SeeSR[Wu et al. 2024b] integrate structural and semantic guidance to steer the denoising trajectory, while SUPIR[Yu et al. 2024] introduces partial controllability via Classifier-Free Guidance (CFG). To accelerate inference, methods such as OSEDiff[Wu et al. 2024a], Difix3D+[Wu et al. 2025], and DMDiff[Jianing Zhang 2025] leverage one-step diffusion backbones, enabling efficient correction of complex degradations ranging from blur to metalens aberrations and 3D rendering errors. Furthermore, diffusion-based frameworks are also proposed to address the physical constraints of light and dynamic range. For instance, in the domain of HDR imaging, LEDiff[Chao Wang 2025] and GaSLight[Bolduc et al. 2025] perform latent space manipulation to estimate spatially-varying illuminance, plausibly reconstructing clipped highlights and shadows. Specifically for lens flare, Difflare [Zhou et al. 2024] exploits diffusion priors to hallucinate scene content occluded by saturated scattering artifacts.

However, current approaches overlook the inherently subjective nature of nighttime enhancement. Both low-light restoration and flare mitigation involve a perceptual trade-off between clarity and atmosphere, where the optimal result depends on artistic intent. Existing methods, constrained to rigid deterministic mappings, fail to accommodate these diverse aesthetic needs. This highlights an urgent necessity for continuous control mechanisms.

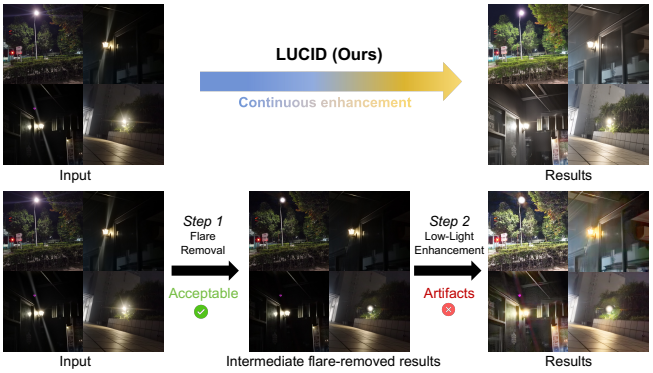


Fig. 2. Comparison with a disjoint baseline. The bottom row illustrates the error accumulation inherent in sequential processing: while the intermediate flare-removed results (Step 1) appear acceptable, latent flare residuals are masked by darkness and subsequently amplified into severe artifacts during enhancement (Step 2). In contrast, our unified LUCID (top) jointly addresses both degradations, yielding clean and coherent reconstruction.

# 3 Methodology

We view nighttime photography not merely as a pixel-level restoration problem, but as a re-lighting process that seeks to recover visual information jointly obscured by darkness and optical flare. From a physical perspective, nighttime degradation arises from the co-existence of insufficient photon exposure and strong stray light, resulting in both signal collapse and structured artifacts.

Following the Retinex formulation [Land and McCann 1971], which is widely adopted in nighttime imaging [Cai et al. 2023; Yi et al. 2023], image formation is commonly modeled as $I = R \cdot L$, where $R$ and $L$ denote the intrinsic scene reflectance and ambient illumination, respectively. In the presence of intense light sources, this model is naturally extended to $I = R \cdot L + F$, where $F$ represents additive stray light caused by flare and ghosting.

This formulation reveals two mathematically distinct restoration objectives: recovering multiplicative illumination and suppressing additive flare. However, in complex real nighttime scenes, these degradations are physically entangled. Naively increasing exposure amplifies flare artifacts, while aggressively suppressing flare often destroys legitimate scene structure. This inherent tension motivates a framework that decomposes these factors explicitly, yet resolves them cooperatively (Fig. 3).

## 3.1 Flare Disentanglement

We begin by explicitly disentangling additive optical artifacts from the underlying scene content. Observing that flare exhibits strong spatial coherence, such as smooth halos and structured ghosting patterns, we employ a lightweight U-Net with a shared encoder and two parallel decoders to decompose the input nighttime image $\mathbf{I}_{\text{in}}$ into a flare component $\mathbf{I}_{\text{flare}}$ and a background component $\mathbf{I}_{\text{bg}}$:

$$\begin{aligned} \mathbf{I}_{\text{flare}} &= \mathcal{D}_{\text{flare}}\left(\mathcal{E}_{\text{decomp}}(\mathbf{I}_{\text{in}})\right), \\ \mathbf{I}_{\text{bg}} &= \mathcal{D}_{\text{bg}}\left(\mathcal{E}_{\text{decomp}}(\mathbf{I}_{\text{in}})\right). \end{aligned} \tag{1}$$

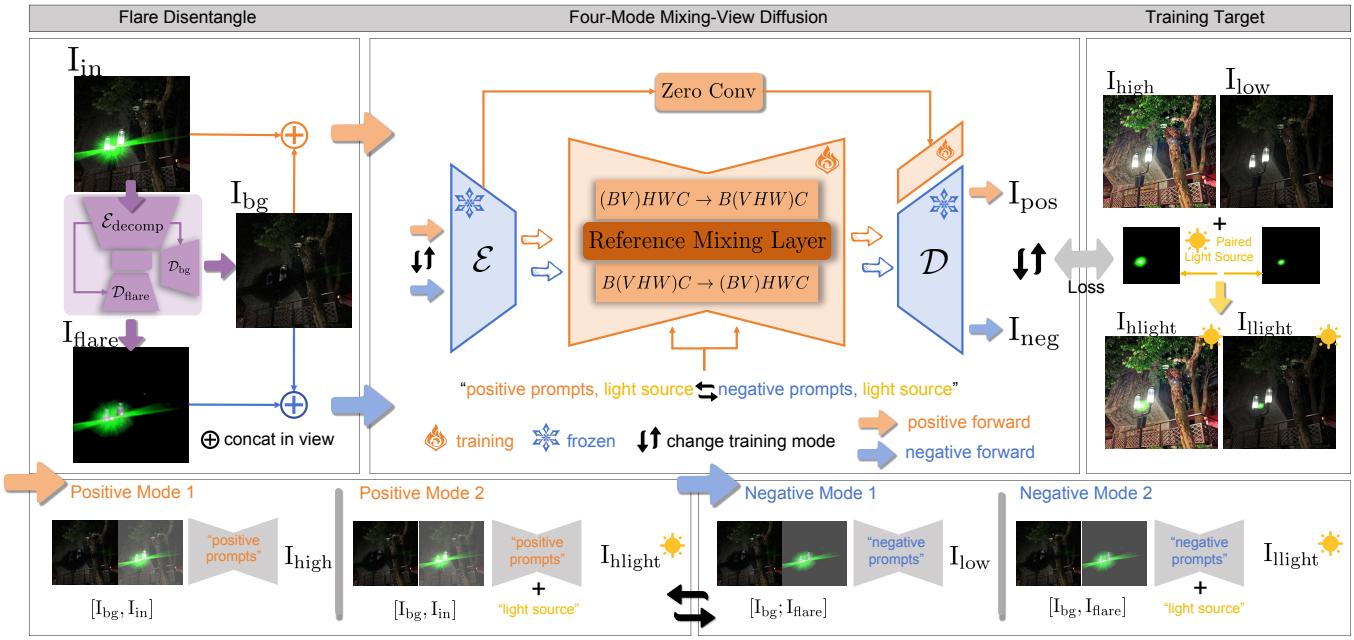


Fig. 3. Training framework of LUCID: The pipeline begins with the Flare Disentangle stage (left), where the degraded input $\mathbf{I}_{in}$ is decomposed into a background estimate $\mathbf{I}_{\text{bg}}$ and a flare component $\mathbf{I}_{\text{flare}}$. These components are selectively concatenated along the view dimension to form multi-view inputs for the Mixing-View Diffusion (center). Inside the network, cross-view interactions guide the diffusion process to achieve fine-grained restoration. The bottom panel illustrates our four distinct Training Regimes: we alternate between Positive Modes (using $\mathbf{I}_{\text{in}}$ as reference for enhancement) and Negative Modes (using $\mathbf{I}_{\text{flare}}$ as reference for suppression), with optional "light source" prompts to enforce semantic controllability over high-intensity regions.

Here, $\mathcal{E}_{\text{decomp}}$ denotes the shared encoder, while $\mathcal{D}_{\text{flare}}$ and $\mathcal{D}_{\text{bg}}$ are the flare and background decoders, respectively. Rather than explicitly separating reflectance $R$ and illumination $L$, we use the term "background" to represent their product $R \cdot L$, which provides a structurally faithful yet flare-free representation. This decomposition serves two complementary purposes: isolating additive flare artifacts for targeted suppression and producing a reliable structural prior to guide the subsequent restoration.

To encourage meaningful disentanglement, the two decoders share weights for the first $m$ layers and diverge thereafter. We impose an orthogonality constraint on the feature maps at the $k$th layer (first divergent layer) to enforce mutual exclusivity:

$$\mathcal{L}_{\text{ortho}} = \left\| \mathbf{F}_{\text{bg}}^{k} \odot \mathbf{F}_{\text{flare}}^{k} \right\|_2^2, \tag{2}$$

where $\mathbf{F}_{\text{bg}}^{k}$ and $\mathbf{F}_{\text{flare}}^{k}$ denote the feature maps at the $k$th layer of the respective decoders. In addition, we impose a reconstruction constraint to ensure self-consistency:

$$\mathcal{L}_{\text{recon}} = \left\| (\mathbf{I}_{\text{flare}} + \mathbf{I}_{\text{bg}}) - \mathbf{I}_{\text{in}} \right\|_2^2. \tag{3}$$

Combined with component-level supervision, the flare disentanglement network is trained using reconstruction, orthogonality, and standard $\ell_2$ losses on the flare and background components:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{ortho}} + \mathcal{L}_{\text{comp}}, \tag{4}$$

where $\mathcal{L}_{\text{comp}}$ denotes the $\ell_2$ supervision on the predicted flare and background against their ground-truth counterparts.

## 3.2 Four-Mode Mixing-View Diffusion

Although flare mitigation and illumination recovery are decomposed at the representation level, they must ultimately be resolved jointly to ensure holistic visual consistency. To this end, we introduce a diffusion-based restoration module that explicitly models their interaction. Drawing architectural inspiration from the paradigm of multi-view diffusion [Shi et al. 2023; Wu et al. 2025], we cast the restoration as a reconstruction process driven by dual priors: $\{\mathbf{I}_{\text{in}}, \mathbf{I}_{\text{ref}}\} \rightarrow \mathbf{I}_{\text{out}}$, where the reference view provides auxiliary cues for structure or flare characteristics. Both inputs are encoded into the latent space via a shared VAE encoder $\mathcal{E}$:

$$\mathbf{z}_{\text{in}} = \mathcal{E}(\mathbf{I}_{\text{in}}), \quad \mathbf{z}_{\text{ref}} = \mathcal{E}(\mathbf{I}_{\text{ref}}). \tag{5}$$

Cross-view interaction is realized through mixing-view self-attention layers embedded within the denoising U-Net. Prior to each attention operation, the latent representations are concatenated along a view dimension, $\mathbf{z} \in \mathbb{R}^{B\times 2\times C\times H\times W}$, then rearranged to collapse the view dimension into the spatial domain, enabling attention across both inputs:

$$\begin{aligned} \mathbf{z}' &= \text{Rearrange}(\mathbf{z}, B \times (2HW) \times C), \\ \mathbf{z}' &= \text{Self-Attention}(\mathbf{z}', \mathbf{z}'), \\ \mathbf{z}' &= \text{Rearrange}(\mathbf{z}', B \times 2 \times C \times H \times W). \end{aligned} \tag{6}$$

This mechanism compels the network to explicitly attend to cues provided by the reference view, enabling robust extraction of structure- and flare-related priors under a unified denoising process.

To further constrain illumination fidelity, we reuse the encoder of the flare disentanglement network as a semantic feature extractor

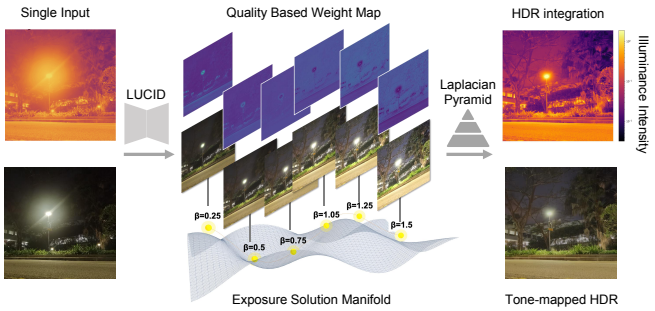


Fig. 4. Single-image HDR reconstruction.

and define an intrinsic feature loss:

$$\mathcal{L}_{\text{intri}} = \sum_{\ell} w_{\ell} \left\| \mathbf{f}^{\ell}_{\text{out}} - \mathbf{f}^{\ell}_{\text{tgt}} \right\|_2^2, \tag{7}$$

where $\mathbf{f}^{\ell}_{\text{out}}$ and $\mathbf{f}^{\ell}_{\text{tgt}}$ denote the $\ell$-th layer features extracted from the network output $\mathbf{I}_{out}$ and the corresponding training target $\mathbf{I}_{tgt}$, respectively. The final diffusion loss combines pixel-level and perceptual objectives:

$$\mathcal{L}_{\text{diff}} = \mathcal{L}_{\text{intri}} + \mathcal{L}_2 + \mathcal{L}_{\text{LPIPS}}. \tag{8}$$

### 3.3 Classifier-Free Guidance-based Control

Classifier-Free Guidance (CFG) [Ho 2022] provides a principled mechanism for conditional control in diffusion models. While prior restoration methods [Jianing Zhang 2025; Zhang et al. 2024] primarily use CFG to trade off fidelity and realism, we extend this paradigm to enable continuous control over exposure and light source appearance. As illustrated in Fig. 3, we strategically design four training modes defined by different combinations of inputs, references, and targets, to support continuous control along two orthogonal dimensions: exposure and light-source presence.

*Continuous Exposure Control.* To model exposure as a controllable dimension, we define two complementary training modes. In the positive mode, the main input is the entangled background $\mathbf{I}_{\text{bg}}$, the reference is the original nighttime image $\mathbf{I}_{\text{in}}$, and the target is a flare-free, well-exposed image $\mathbf{I}_{\text{high}}$. This configuration encourages the model to exploit reference-derived structural cues for both illumination recovery and flare suppression. In contrast, the negative mode replaces the reference with the flare $\mathbf{I}_{\text{flare}}$ and sets the target to an under-exposed image $\mathbf{I}_{\text{low}}$, guiding the model to associate the reference view with flare characteristics alone. These two modes jointly define the endpoints of the exposure control spectrum.

*Continuous Light-Source Control.* To further empower selective manipulation of scene illuminants, we augment each exposure mode with specialized sub-configurations designed to either preserve or suppress explicit light-source appearances. We synthesize pseudo-illuminant maps, $\mathbf{I}_{\text{hlight}}$ and $\mathbf{I}_{\text{llight}}$, adapting the synthesis protocols of Flare7K. Crucially, these components are engineered to simulate the pristine, intrinsic radiance of light sources devoid of optical scattering defects. For the supervision of these light-preserving sub-modes, we designate the composite states as the GT targets: integrating $\mathbf{I}_{\text{hlight}}$ into $\mathbf{I}_{\text{high}}$ for positive regime, and $\mathbf{I}_{\text{llight}}$ into $\mathbf{I}_{\text{low}}$ for negative counterpart, respectively. To explicitly trigger this modal

(a) Input (b) "Light Source" (c) w/o "Light Source"

Fig. 5. Prompt-driven light source preservation. By supervising the network with distinct GT targets conditioned on specific textual prompts, LUCID learns to selectively retain or fully remove the light source.

control, we append the textual descriptor "light source" to the base prompts. This instructs the model to semantically modulate the restoration behavior, ensuring the faithful retention of illuminants aligned with the constructed targets.

At inference time, CFG enables continuous traversal between negative and positive solutions by modulating a scalar $\beta$:

$$\hat{\mathbf{z}} = \mathbf{z}_{\text{neg}} + \beta(\mathbf{z}_{\text{pos}} - \mathbf{z}_{\text{neg}}). \tag{9}$$

After latent interpolation, a VAE decoder augmented with encoder skip connections reconstructs the final output. By adjusting $\beta$ and toggling the "light source" prompt, users can seamlessly transition from aggressive flare suppression to selective preservation of light sources and their associated artifacts.

### 3.4 Nighttime Single-Image HDR Reconstruction

The proposed framework naturally extends to single-image High Dynamic Range (HDR) reconstruction. By leveraging the continuous exposure control enabled by CFG, a sequence of exposure-consistent outputs can be synthesized from a single nighttime input. These outputs are fused to produce the final HDR result. We employ Laplacian pyramid blending with quality-aware weighting to aggregate the synthesized exposures into an HDR representation with balanced luminance distribution and light-source appearance. Detailed algorithmic procedures are provided in the Supplementary Material.

## 4 Experiments and Results

### 4.1 Implementation

We adapt SD-Turbo [Sauer et al. 2023] as the generative backbone of our diffusion-based restoration module. Training data are curated from several real-world low-light datasets, including RELLISUR [Aakerberg et al. 2021], LSRW [Hai et al. 2023], SICE [Cai et al. 2018], and SID [Chen et al. 2018]. To synthesize degraded nighttime inputs $\mathbf{I}_{\text{in}}$, we follow the physically grounded flare generation pipeline of Flare7K [Dai et al. 2022], superimposing optical flare artifacts onto under-exposed images $\mathbf{I}_{\text{low}}$. During training, negative conditioning in CFG is applied with a probability of 0.2. Independently, the presence of the light-source semantic prompt is sampled with a probability of 0.5. Training is conducted on a single NVIDIA A800 GPU at a spatial resolution of $512 \times 512$ with a batch size of 4.

### 4.2 Evaluation Protocol

*Evaluation Benchmark.* To evaluate LUCID under realistic nighttime photography conditions, we adopt three task-oriented complementary benchmarks:

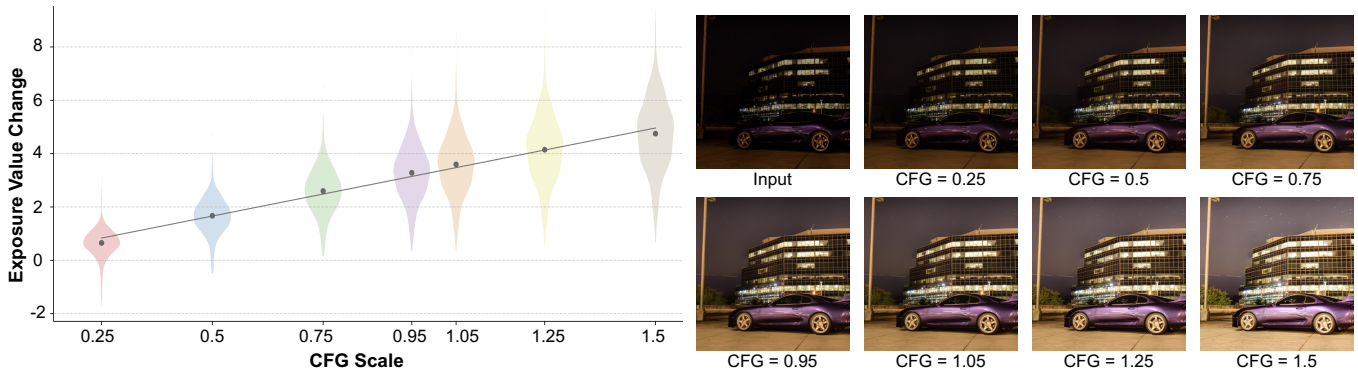


Fig. 6. Relationship between CFG scales ($\beta$) and exposure value changes ($\Delta$EV). Left: Violin plots demonstrates a predictable linear response under a standard sRGB gamma mapping ($\gamma = 2.2$). Right: Visual examples verify that adjusting $\beta$ achieves perceptually smooth exposure ramping.

(1) General nighttime enhancement: We build a diverse benchmark based on the Exclusively Dark (ExDark) dataset [Loh and Chan 2019], which contains 7,363 real-world nighttime images ranging from extreme darkness to twilight. Unlike controlled low-light datasets, ExDark exhibits strong illumination imbalance and complex light-source interference that better reflect real nighttime photography. To exclude samples compromised by severe degradations (e.g., heavy blur or statistical distortions), we applied a quality screening process based on Laplacian variance and BRISQUE [Mittal et al. 2012] scores. Consequently, we select a subset of 1,271 images that feature pronounced dynamic range, visible light sources, and minimal ambient illumination.
(2) Flare mitigation: We use the Flare7K dataset [Dai et al. 2022], which provides real-world nighttime images containing prominent lens flare and ghosting artifacts.
(3) Single-image HDR: We include the SiHDR [Hanji et al. 2022] dataset to assess the effectiveness of LUCID in recovering extended dynamic range from a single nighttime exposure.

*Comparison Methods.* We compare LUCID against recent State-Of-The-Art (SOTA) low-light image enhancement (LLIE) methods. To comprehensively benchmark general enhancement performance, we select representative methods spanning diverse technical paradigms, including Zero-DCE [Guo et al. 2020], RetinexFormer [Cai et al. 2023], RetiDiff [He et al. 2025], DarkIR [Feijoo et al. 2025], and the unsupervised approach of Jin et al. [Jin et al. 2022]. For all competing approaches, we use the official pre-trained models recommended by the authors for in-the-wild inference to ensure a fair comparison.

To evaluate flare mitigation, we additionally include representative flare removal methods, namely Flare7K [Dai et al. 2022], MFDNet [Jiang et al. 2024], and the method of Zhou *et al.* [Zhou et al. 2023]. For single-image HDR reconstruction, we compare against established HDR methods, including IntrinsicHDR [Dille et al. 2024], LEDiff [Chao Wang 2025], and GasLight [Bolduc et al. 2025].

Unlike controlled low-light benchmarks, which are constructed by varying camera exposure and therefore exhibit limited dynamic range, real nighttime scenes involve highly heterogeneous illumination with no unique ground-truth brightness. Accordingly, we primarily rely on no-reference image quality metrics [Ke et al. 2021; Talebi and Milanfar 2018; Wang et al. 2023; Yang et al. 2022; Zhang et al. 2023] for quantitative evaluation.

| Method | CLIPIQA | MANIQA | MUSIQ | LIQE | NIMA |
|---|---|---|---|---|---|
| ExDARK | 0.4281 | 0.2939 | 50.99 | 2.267 | 5.227 |
| Zero-DCE | 0.4243 | 0.2899 | 50.98 | 2.078 | 4.999 |
| Retinexformer | 0.3749 | 0.2588 | 52.26 | 2.058 | 5.121 |
| Jin et al. | 0.3664 | 0.2585 | 51.89 | 2.087 | 5.078 |
| Reti-Diff | 0.4159 | 0.2867 | 50.91 | 1.967 | 5.083 |
| DarkIR | 0.4107 | 0.3078 | 52.03 | 2.076 | 5.046 |
| **Ours ($\beta = 1.05$)** | **0.4774** | **0.3264** | **61.45** | **3.019** | **5.390** |

Table 1. Quantitative comparison on IQA metrics. All metrics are reference-free, higher is better (↑). Best results are **bolded**.

## 4.3 Results

*Performance on Holistic Nighttime Restoration.* Fig. 9 validates the robustness of LUCID across four diverse nighttime regimes (top to bottom): extreme photon starvation, high-contrast dynamic range, strong backlighting with veiling glare, and severe illuminant color casts. Baseline methods exhibit characteristic failures: they linearly amplify sensor noise in signal-starved regions (Row 1), over-expose localized headlights (Row 2), struggle to penetrate optical artifacts (Row 3), or blindly enhance the dominant hue, causing unnatural chromatic distortion. In contrast, LUCID demonstrates superior semantic consistency. It effectively suppresses noise in ultra-dark limits and compresses dynamic range to recover highlight textures. Simultaneously, it rectifies color deviations and restores clarity in washed-out backlit scenarios. This confirms its versatility as a robust solution for uncontrolled real-world nighttime imaging.

Quantitative results based on no-reference perceptual metrics are reported in Tab. 1. These results statistically validate that our method not only enhances visibility but also aligns better with human perceptual preferences, demonstrating superior generalization to diverse authentic nighttime imagery.

*Flare Mitigation.* To isolate and evaluate flare mitigation performance, we conduct experiments on the Flare7K dataset. Crucially, the comparative baselines are specialized strictly for artifact removal, lacking the low-light enhancement capabilities integral to our method. Since the dataset is constructed by introducing physical smudges, the Ground-truth (GT) may inherently retain slight residual flares from the lens optics. As shown in Fig. 11, baseline methods struggle to achieve a balance between removal and preservation.


Fig. 7. Visual comparison of controllability across different workflows.

They often exhibit incomplete removal, leaving residual streaks and hazy artifacts (1st and 2nd rows). Conversely, some methods tend to over-subtract the signal around light sources, resulting in unnatural sharp boundaries (e.g., Zhou et al. in the 3rd row). In contrast, LUCID successfully disentangles and removes the scattering artifacts. The proposed method leverages generative priors to synthesize reliable background textures while reconstructing the natural optical fall-off of light sources. The resulting images are visually clean and physically plausible, occasionally exceeding the quality of GT.

*Continuous Control.* A core contribution of LUCID is empowering photographers with precise, continuous control over the restoration process. To validate the predictability of our control mechanism, we first analyze the statistical relationship between the CFG scale ($\beta$) and the relative exposure value change ($\Delta$EV). As visualized in Fig. 6, the mean exposure value increment exhibits a monotonic and quasi-linear response to the guidance scale, allowing users to intuitively dial in the desired brightness. The visual examples visualize the continuous modulation effects. Spanning $\beta \in [0.25, 1.5]$, LUCID maintains robust structural consistency across all intervals while rendering smooth, natural transitions in illuminance. This progression effectively mimics the physical behavior of gradually intensifying a dimmer-controlled light source. We contrast our controllable paradigm against three alternatives (Fig.7).
(1) Naive Baseline: Global exposure adjustment is physically flawed; it amplifies sensor noise and expands the radius of veiling glare, severely washing out scene contrast.
(2) Cascaded Pipeline: We sequentially chain a deflare network[Dai et al. 2022] with a low-light enhancer [Feijoo et al. 2025], followed by manual dimming (Photoshop). This suffers from compound error accumulation, where residual artifacts missed by the deflare stage are aggressively amplified by the enhancer into unnatural residues.
(3) Commercial GenAI (NanoBanana Pro): Despite impressive resolution, relying solely on text prompts proves insufficient: such control is too coarse for precise illuminance tuning and prone to semantic drift (e.g., hallucinating daylight structures).

Ultimately, LUCID transcends these limitations by reconciling generative fidelity with user intent. It provides a transparent, controllable framework that eliminates optical degradations. This capability offers a level of interaction and consistency that is typically absent in unconstrained, purely text-guided generative architectures.

(a) Input (b) $\beta = 1.05$ (c) HDR

Fig. 8. Dual restoration aesthetics. LUCID yields both perceptual realism at $\beta = 1.05$ and an alternative HDR aesthetic, maximizing detail visibility across shadows and highlights.

*Single-Image HDR Reconstruction.* By fusing "virtual exposure brackets" synthesized via CFG control, LUCID extends naturally to HDR reconstruction. Fig. 11 compares our results against specialized HDR methods. Baselines exhibit distinct failures: IntrinsicHDR leaves deep shadows crushed (Row 1); LEDiff and Gaslight suffer from generative hallucinations (e.g., distorted roof tiles in Row 1), unnatural color shifts (e.g., the artificial orange cast in Row 2), and or erroneously amplifying lens flare (Row 3). In contrast, LUCID achieves superior photorealism, maintaining structural fidelity and natural color balance across all exposure levels. This introduces a flexible aesthetic dimension, enabling idealized scene rendering that transcends the imperfections of raw physical capture (Fig. 8).

*Downstream Applications.* LUCID serves as a plug-and-play module for diverse workflows (Fig. 12). As a pre-processor (top), it suppresses pre-existing flares to provide a clean, artifact-free input, ensuring subsequent prompt-driven editing remains free from original lighting interference. As a post-processor (bottom), it introduces post-hoc tunability to generative models, enabling flexible adjustment of exposure and flare on otherwise static AI-generated outputs.

## 5 Conclusion

In this work, we introduced LUCID to address the entangled challenges of nighttime photography. Diverging from fixed-solution baselines, our framework synergizes a flare disentanglement module with a controllable diffusion backbone. This integration facilitates continuous modulation of image luminance, affording users precise control over the exposure recovery process. Extensive experiments demonstrate the superior generalization of LUCID in rendering night scenes with clear visibility and pristine aesthetics. By constructing a disentangled lighting manifold within the latent space, LUCID effectively circumvents the semantic ambiguity inherent in purely generative approaches. We believe this paradigm of controllable, physics-aware generation paves the way for future intelligent computational photography tools.

## References


Andreas Aakerberg, Thomas B Moeslund, and Kamal Nasrollahi. 2021. RELLISUR: A Real Low-Light Image Super-Resolution Dataset. In *Adv. Neural Inform. Process. Syst.*

Christophe Bolduc, Yannick Hold-Geoffroy, Zhixin Shu, and Jean-François Lalonde. 2025. GaSLight: Gaussian Splats for Spatially-Varying Lighting in HDR. In *Int. Conf. Comput. Vis.*

Paul A Boynton and Edward F Kelley. 2003. Liquid-filled camera for the measurement of high-contrast images. In *Cockpit Displays X*, Vol. 5080. SPIE, 370–378.

Jianrui Cai, Shuhang Gu, and Lei Zhang. 2018. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Transactions on Image Processing* 27, 4 (2018), 2049–2062.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *Int. Conf. Comput. Vis.*

Thomas Leimkuhler Karol Myszkowski Xuaner Zhang Chao Wang, Zhihao Xia. 2025. LEDiff: Latent Exposure Diffusion for HDR Generation. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to See in the Dark. In *IEEE Conf. Comput. Vis. Pattern Recog.* 3291–3300.

Wenhan Yang Jiaying Liu Chen Wei, Wenjing Wang. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference.*

Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. 2022. Flare7K: A Phenomenological Nighttime Flare Removal Dataset. In *Adv. Neural Inform. Process. Syst.*

Sebastian Dille, Chris Careaga, and Yağız Aksoy. 2024. Intrinsic Single-Image HDR Reconstruction. In *Eur. Conf. Comput. Vis.*

Daniel Feijoo, Juan C. Benito, Alvaro Garcia, and Marcos V. Conde. 2025. DarkIR: Robust Low-Light Image Restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.* 10879–10889.

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

Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation* 90 (2023), 103712.

Param Hanji, Rafal Mantiuk, Gabriel Eilertsen, Saghi Hajisharif, and Jonas Unger. 2022. SI-HDR - dataset for comparison of single-image high dynamic range reconstruction methods. (2022). doi:10.17863/CAM.87333

Chunming He, Chengyu Fang, Yulun Zhang, Kai Li, Longxiang Tang, Chenyu You, Fengyang Xiao, Zhenhua Guo, and Xiu Li. 2025. Reti-Diff: Illumination Degradation Image Restoration with Retinex-based Latent Diffusion Model. (2025).

Jonathan Ho. 2022. Classifier-Free Diffusion Guidance. *ArXiv* abs/2207.12598 (2022).

Yiguo Jiang, Xuhang Chen, Chi-Man Pun, Shuqiang Wang, and Wei Feng. 2024. MFDNet: Multi-Frequency Deflare Network for Efficient Nighttime Flare Removal. *The Vis. Comput.* 40 (2024), 7575–7588.

Feiyu Ji Xiaokang Yang Xiaoyun Yuan Jianing Zhang, Jiayi Zhu. 2025. Degradation-Modeled Multipath Diffusion for Tunable Metalens Photography. In *Int. Conf. Comput. Vis.*

Yeying Jin, Wenhan Yang, and Robby T Tan. 2022. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *Eur. Conf. Comput. Vis.* Springer, 404–421.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. 2021. MUSIQ: Multi-scale Image Quality Transformer. In *Int. Conf. Comput. Vis.* 5148–5157.

Edwin H Land and John J McCann. 1971. Lightness and retinex theory. *Journal of the Optical Society of America* 61, 1 (1971), 1–11. doi:10.1364/JOSA.61.000001

Chongyi Li, Chunle Guo, Ruixing Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chetan C Gupta. 2021. Benchmarking Low-Light Image Enhancement and Beyond. *Int. J. Comput. Vis.* 129, 4 (2021), 1153–1174.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. 2024. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior. In *Eur. Conf. Comput. Vis.*

Yuen Peng Loh and Chee Seng Chan. 2019. Getting to Know Low-light Images with The Exclusively Dark Dataset. *Computer Vision and Image Understanding* 178 (2019), 30–42. doi:10.1016/j.cviu.2018.10.010

H Angus Macleod. 2010. *Thin-film optical filters.* CRC press.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21, 12 (2012), 4695–4708.

Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. 2023. Adversarial Diffusion Distillation. *arXiv preprint arXiv:2311.17042* (2023).

Aashish Sharma and Robby T Tan. 2021. Nighttime Visibility Enhancement by Increasing the Dynamic Range and Suppression of Light Effects. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Yichun Shi, Peng Wang, Jianglong Ye, Long Mai, Kejie Li, and Xiao Yang. 2023. MV-Dream: Multi-view Diffusion for 3D Generation. *arXiv:2308.16512* (2023).

Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing* 27, 8 (2018), 3998–4011. doi:10.1109/TIP.2018.2831899

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. 2023. Exploring CLIP for Assessing the Look and Feel of Images. In *AAAI.*

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *Brit. Mach. Vis. Conf.*

Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, and Huan Ling. 2025. DIFIX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models. In *Proceedings of the Computer Vision and Pattern Recognition Conference.* 26024–26035.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. 2024a. One-Step Effective Diffusion Network for Real-World Image Super-Resolution. In *Adv. Neural Inform. Process. Syst.*

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. 2024b. SeeSR: Towards semantics-aware real-world image super-resolution. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T Barron. 2021. How to Train Neural Networks for Flare Removal. In *Int. Conf. Comput. Vis.* 2239–2247.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. 2022. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*

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

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. 2024. Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. 2024. Degradation-Guided One-Step Image Super-Resolution with Diffusion Priors. *arxiv* (2024).

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. 2023. Blind Image Quality Assessment via Vision-Language Correspondence: A Multitask Learning Perspective. In *Int. Conf. Comput. Vis.*

Yu Zhang, Xiaojie Guo, Chongyi Li, Jun Ma, Jinjian Su, and Wenhao Li. 2019. Kindling the Darkness: a Practical Low-light Image Enhancer.

Tianwen Zhou, Qihao Duan, and Zitong Yu. 2024. Difflare: Removing Image Lens Flare with Latent Diffusion Model. In *Brit. Mach. Vis. Conf.*

Yuyan Zhou, Dong Liang, Songcan Chen, Sheng-Jun Huang, Shuo Yang, and Chongyi Li. 2023. Improving Lens Flare Removal with General Purpose Pipeline and Multiple Light Sources Recovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).* 12586–12596.

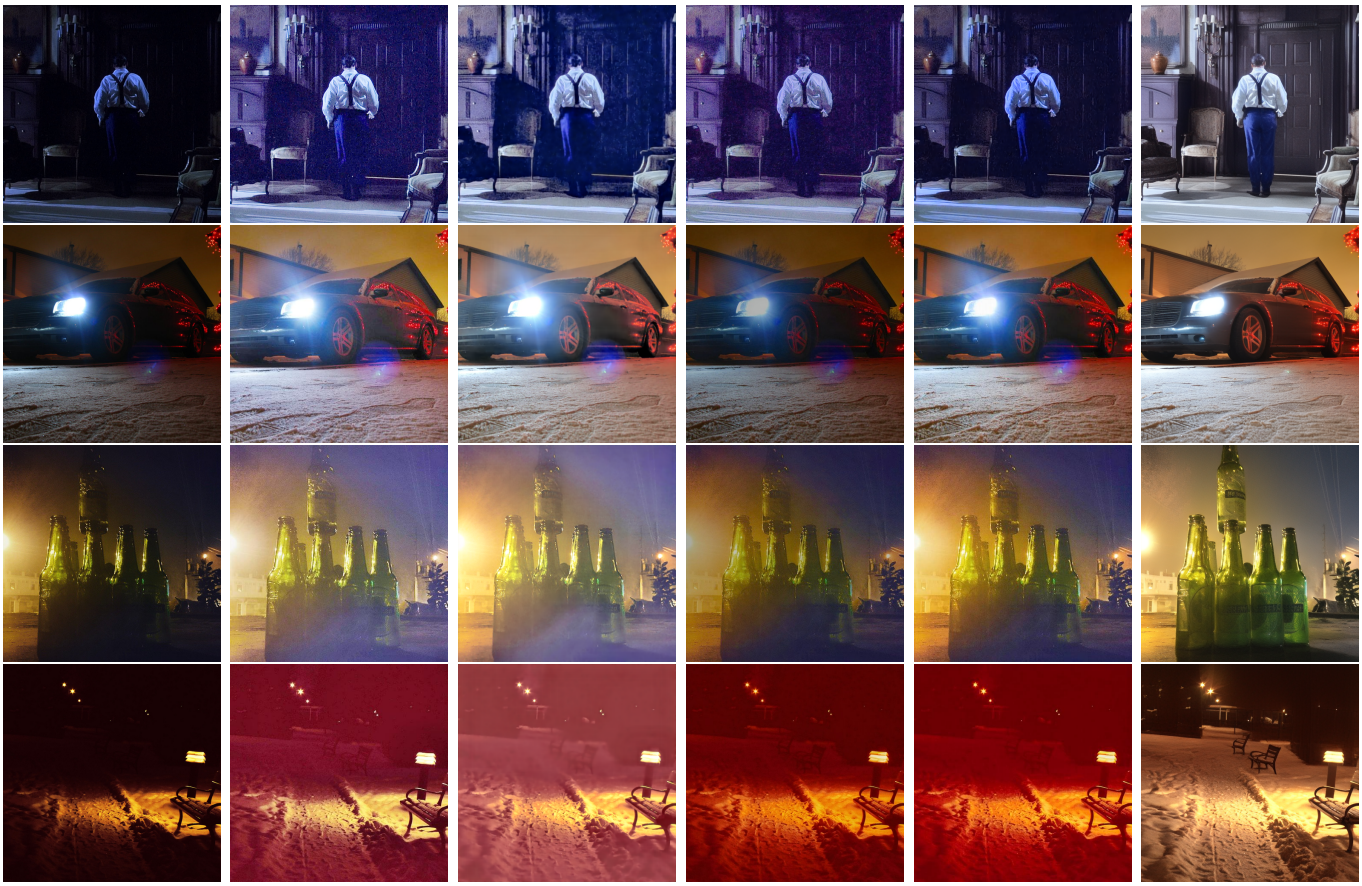


Fig. 9. Visual comparison on the ExDark dataset. This comparison focuses on evaluating the enhancement performance on authentic night scenes.

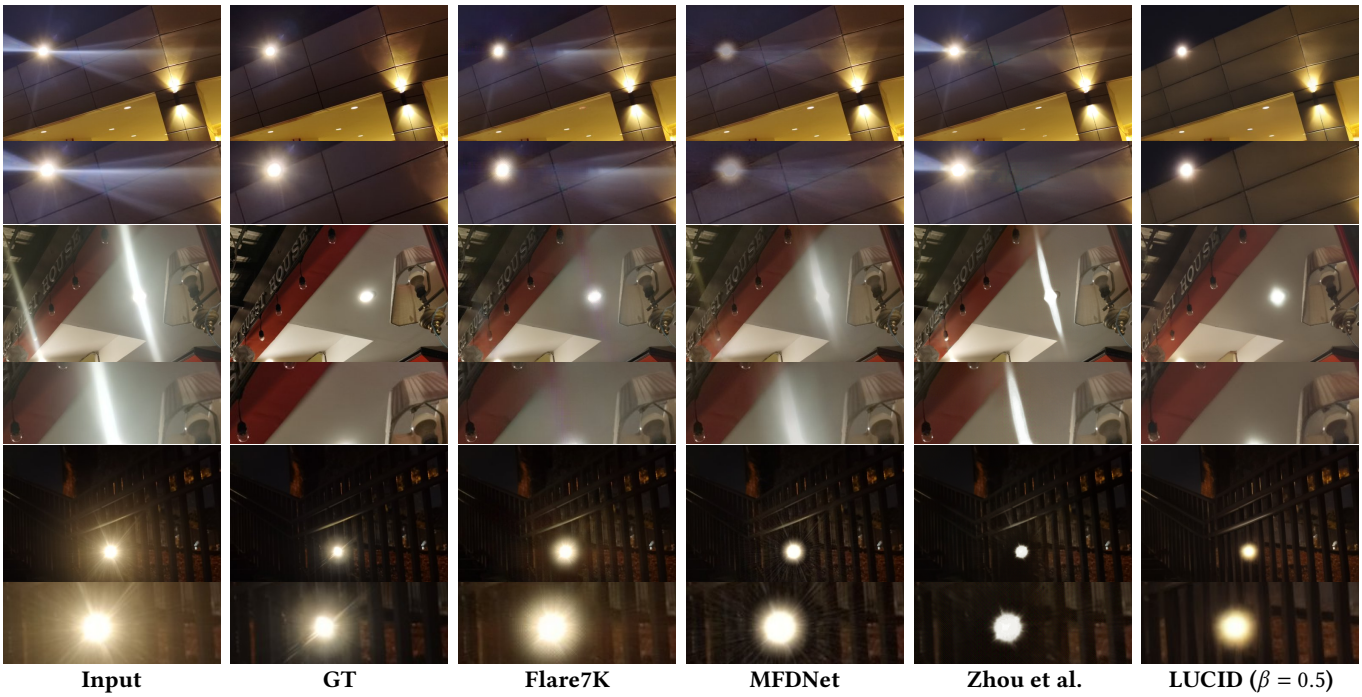


Fig. 10. Visual comparison on the Flare7K dataset. The comparison centers on the effectiveness of flare mitigation and preservation of light source.

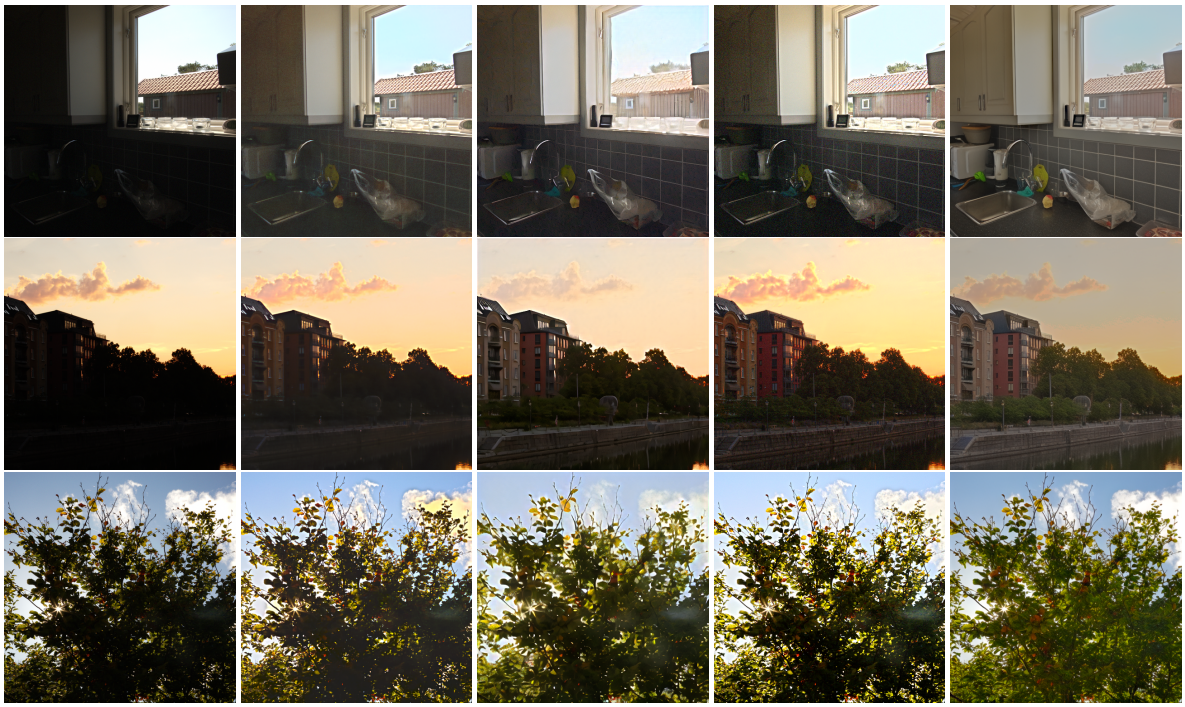


Fig. 11. Visual comparison on the SiHDR dataset. The comparison centers on HDR reconstruction.

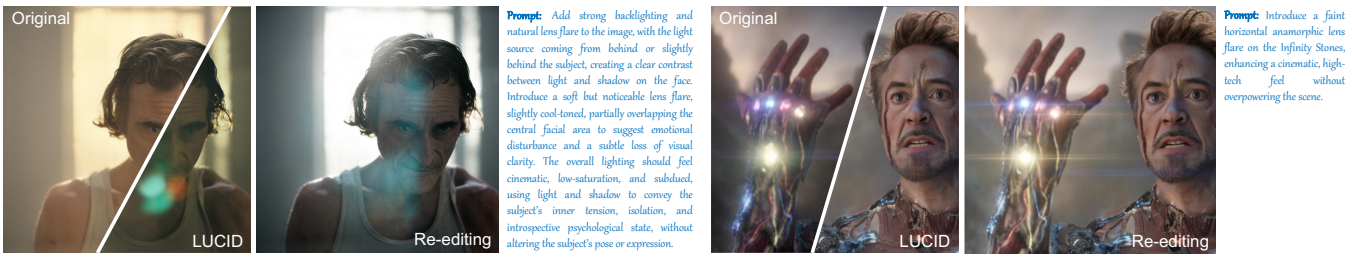


Fig. 12. Lucid enables creative image editing. Top: Lucid suppresses existing lens flare, effectively decoupling flare artifacts from the underlying image and preparing the result for subsequent editing (using Nano-Banana Pro). Bottom: Combined with commercial AI image generators (Doubao AI and Nano-Banana Pro), Lucid serves as a post-processing module that allows flexible adjustment of exposure and lens flare.