

When Extreme Darkness Meets Motion Blur: MeanFlow for Unified RAW Restoration

Zepu Wang^{✉*}, Jingze Liang^{✉*}, Weijie Xiao^{✉*}, Kexin Chen[✉]

University of Electronic Science and Technology of China

Abstract

Extremely low-light RAW enhancement aims to recover severely attenuated sensor signals, yet existing methods often focus on illumination and noise while overlooking the motion-induced degradations inherent in practical low-light imaging. We present a framework for robust extremely low-light RAW enhancement under realistic acquisition degradations. First, we introduce See in the Degraded Extremely Dark (SIDED), a new dataset that applies controlled motion degradation to extremely low-light RAW pairs while retaining their original sensor noise. Second, we propose a unified RAW tokenizer equipped with explicit domain-conditioned representation calibration to align extremely low-light and well-exposed RAW data, followed by a MeanFlow that performs enhancement in a single function evaluation. To our knowledge, this is the first work to formulate extremely low-light RAW enhancement under realistic motion-degraded acquisition and address it with MeanFlow. We further introduce a physics-guided refinement model to strengthen illumination–reflectance consistency, pixel fidelity, and color preservation without incurring additional inference cost. Extensive experiments demonstrate that our framework achieves state-of-the-art performance in extremely low-light RAW enhancement, and robustly handles coupled motion and noise degradations. Code is available at https://github.com/JingZe-Liang/Low_Light_Enhancement.

Introduction

Extreme low-light RAW enhancement aims to recover visible and faithful images from sensor measurements captured under illumination far below conventional low-light conditions (Chen et al. 2018; Wei et al. 2020; Jiang et al. 2025a). It is important for computational photography, night-time imaging, and vision systems operating in poorly illuminated environments. Compared with post-ISP RGB images, RAW measurements retain weak photon responses and sensor-level information that may otherwise be suppressed during image signal processing (Chen et al. 2018; Jin et al. 2023; Chen et al. 2026). This information is valuable for restoring severely under-exposed scenes, but its recovery remains difficult because extreme darkness produces substantial intensity attenuation, low signal-to-noise ratios, color ambiguity, and complex noise (Wei et al. 2020; Jiang et al. 2025a).

Existing learning-based methods have made substantial progress in exposure correction, denoising, and color restoration in both RGB and RAW domains (Chen et al. 2018; Jin et al. 2023; Chen et al. 2026; Jiang et al. 2025a). Nevertheless, most extremely low-light enhancement studies concentrate on photometric degradation and sensor noise, while the blur accompanying handheld capture is often overlooked. In practical handheld photography, insufficient illumination is also frequently accompanied by unintended camera motion, which smears edges and textures while severe sensor noise further obscures them (Zhou, Li, and Loy 2022; Zhao et al. 2022). Enhancing brightness alone can make these blurred structures more visible, but cannot recover the spatial details already lost to motion (Zhou, Li, and Loy 2022; Lv et al. 2024). This gap limits the applicability of existing approaches to practical extreme-darkness scenarios, where illumination loss, noise, and motion degradation frequently coexist.

To directly confront this overlooked setting, we formalize extreme low-light RAW restoration with accompanying handheld motion as a practical restoration problem, rather than treating blur as an incidental corruption. We instantiate this setting through See in the Degraded Extremely Dark (SIDED), constructed from SIED (Jiang et al. 2025a) by introducing controlled motion degradations that simulate real-world handheld capture while retaining paired clean references. To our knowledge, SIDED is the first controlled RAW benchmark for accompanying motion degradation under extremely low illumination. We further propose a latent RAW restoration framework centered on a task-aware MeanFlow group (Geng et al. 2025). A unified RAW tokenizer embeds extremely low-light and well-exposed measurements into a shared latent space. Building on this representation, the MeanFlow group performs the core dark-to-light transformation in one function evaluation and incorporates motion-aware latent recovery when camera-shake degradation is present. A training-only physics-guided strategy further encourages illumination–reflectance consistency and faithful image recovery (Cai et al. 2023).

Experiments on the Sony subsets at three extreme illumination levels demonstrate the effectiveness of the proposed framework. Our method achieves the highest PSNR and SSIM on SIED across all three levels, and ranks first in PSNR, SSIM, and Spectral Angle Mapper (SAM) on SIDED.

*These authors contributed equally.

In the motion-degraded setting, it improves PSNR over the strongest competing results by 6.06, 6.84, and 7.42 dB at 0.01, 0.001, and 0.0001 lux, respectively.

The main contributions of this paper are summarized as follows:

- We identify and formalize accompanying motion degradation as a central yet overlooked challenge in extreme low-light RAW imaging, and introduce SIDED as a controlled RAW benchmark for this practical setting.
- We propose a unified RAW tokenizer and a task-aware MeanFlow group that performs extreme low-light enhancement in one function evaluation and incorporates motion-aware latent recovery when required. To the best of our knowledge, this is the first successful application of MeanFlow to extremely low-light RAW enhancement.
- We develop a training-only physics-guided refinement strategy for illumination–reflectance consistency and faithful recovery. Extensive comparisons demonstrate state-of-the-art restoration quality on both low-light-only and motion-degraded benchmarks.

Related Work

Extremely Low-Light RAW Enhancement

Low-light image enhancement (LLIE) has been widely studied in both RGB and RAW domains. RGB-based methods (Guo et al. 2020; Xu et al. 2022; Wang et al. 2023; Wei et al. 2018; Zhang, Zhang, and Guo 2019) operate on post-ISP images and improve brightness, contrast, color rendition, and perceptual quality through curve estimation, transformer-based restoration, learned color transformations, or decomposition-based modeling. Although RGB images are readily available and visually interpretable, the ISP has already altered sensor measurements through demosaicing, denoising, white balance, color correction, and tone mapping. Weak signals in extremely dark regions may therefore be suppressed or entangled with ISP artifacts. RAW-based methods (Chen et al. 2018; Jin et al. 2023; Chen et al. 2026; Dong et al. 2022; Lamba, Balaji, and Mitra 2020) instead restore images closer to the sensor domain, where low-intensity responses and camera-dependent noise statistics remain accessible. A central challenge is the large distribution gap between extremely dark and well-exposed RAW observations.

Paired RAW benchmarks have advanced the study of extreme exposure correction, denoising, and color restoration (Chen et al. 2018; Wei et al. 2020), while SIED further targets scenes under extremely low illumination (Jiang et al. 2025a). However, these benchmarks primarily characterize exposure and noise degradation, leaving the blur that frequently accompanies handheld capture underrepresented. This omission limits their coverage of practical photography, in which camera shake may occur together with extreme darkness. SIDED represents each scene through the quadruplet (x_l, x_g, x_l^m, x_g^m) , pairing low-light and normal-exposure RAW measurements with their motion-degraded counterparts while preserving scene correspondence.

Joint LLIE and Motion Deblurring

Motion blur is particularly detrimental under low illumination because longer exposures increase the likelihood of camera shake, while weak signals and severe noise obscure the edges and textures required for reliable restoration. Recent image deblurring methods recover sharp structures through multi-scale networks, transformer-based restoration, and frequency- or motion-adaptive filtering (Cho et al. 2021; Zamir et al. 2022; Kong et al. 2023; Liu et al. 2024). More recent approaches jointly address low-light enhancement and motion deblurring, showing that exposure correction and motion restoration should be considered together in practical nighttime imaging (Zhou, Li, and Loy 2022; Zhao et al. 2022; Li et al. 2024; Lv et al. 2024; Liu et al. 2025).

Despite this progress, two limitations remain. First, existing joint restoration methods are developed mainly for rendered RGB or sRGB images (Zhou, Li, and Loy 2022; Li et al. 2024; Lv et al. 2024; Liu et al. 2025), where demosaicing, denoising, color processing, tone mapping, and clipping may alter or suppress weak sensor measurements. Second, current RAW benchmarks primarily study exposure correction and noise removal without explicitly modeling accompanying camera shake (Chen et al. 2018; Wei et al. 2020; Jiang et al. 2025a). Under extremely low illumination, the narrow signal range and severe sensor noise further obscure motion-affected structures, making direct transfer from conventional RGB-domain deblurring unreliable. SIDED targets this underexplored regime with paired extremely dark RAW observations and corresponding motion-degraded counterparts. Our framework performs enhancement and motion-aware latent restoration directly from sensor measurements, without requiring a blur kernel at inference.

Latent Restoration and Physical Priors

Generative models capture clean-image distributions and recover perceptually plausible details. Diffusion restoration progressively denoises corrupted representations for low-light enhancement, denoising, and deblurring (Yi et al. 2023; Kavar et al. 2022; Zhu et al. 2023; Lin et al. 2025). Flow matching learns continuous transport between distributions, with high-order formulations extending its dynamics and statistical analysis (Lipman et al. 2023; Su et al. 2025), whereas MeanFlow uses an average velocity field for one-step generation (Geng et al. 2025).

Latent restoration moves this transformation to an encoder–decoder representation (Rombach et al. 2022; Jiang et al. 2025b). For RAW enhancement, a shared latent space must preserve scene content despite the exposure and noise differences between low-light and well-exposed domains. Feature-wise affine modulation, domain-specific normalization, and feature-reconstruction constraints provide mechanisms for conditioning shared representations (Perez et al. 2018; Chang et al. 2019; Zhu et al. 2024). Our unified tokenizer therefore uses shared encoding and decoding with domain-adaptive affine modulation. Its latent space supports both one-step scaled-low-to-GT transformation and motion-aware restoration matched to the enhancement output.

Retinex-inspired methods decompose illumination and reflectance to provide a physical prior for low-light enhance-

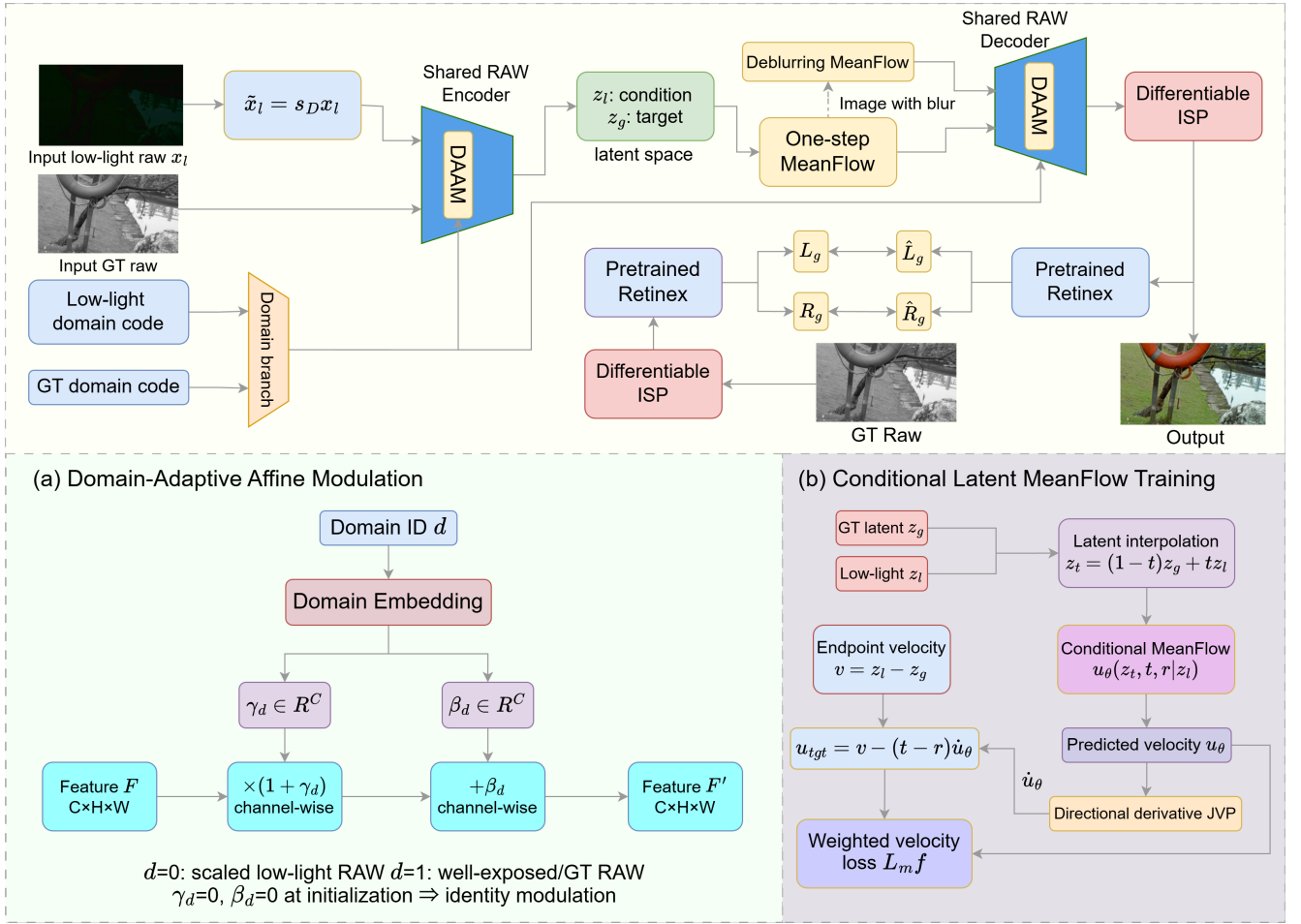


Figure 1: Overview of the proposed framework. A unified RAW tokenizer establishes a shared latent space, in which a task-aware MeanFlow group performs one-step extreme low-light enhancement and motion-aware restoration. Physics-guided refinement is used only during training.

ment (Wei et al. 2018; Wu et al. 2022), and recent work combines this prior with transformers or diffusion (Cai et al. 2023; Yi et al. 2023). Because inference-time decomposition may propagate errors, we use a frozen Retinex network only to guide training.

Methodology

Overview

As illustrated in Fig. 1, our framework consists of a unified RAW tokenizer, a task-aware MeanFlow group, and training-only physics guidance. SIDED provides extremely low-light observations and normal-exposure references together with their motion-degraded counterparts, representing both low-light-only and motion-affected capture conditions. The unified tokenizer embeds these RAW measurements into a shared compact space while preserving the information required for restoration.

Within this latent space, MeanFlow performs the dark-to-light transformation in one function evaluation and incorporates motion-aware recovery for observations affected by

camera shake. The resulting latent representation is decoded into a restored RAW image. Physics guidance further promotes illumination–reflectance consistency during learning and is removed at inference. This overview defines the roles of the main components; their optimization objectives and training procedures are presented in the following subsections.

Stage I: Unified Domain-Adaptive RAW Tokenizer

Extremely low-light and GT RAW images have substantially different numerical distributions, making direct joint tokenization unstable. We therefore train a unified tokenizer with a shared encoder E_ϕ and decoder D_ψ through three phases: dataset-adaptive scaling, scale-only pretraining, and domain-adaptive joint training. This design establishes a stable common latent space while retaining domain-specific RAW statistics for subsequent MeanFlow learning.

Phase A: Dataset-Adaptive Scale Estimation For each dataset $\mathcal{D}$, we estimate a reversible low-light scale from sampled normalized training RAW patches rather than using a

universal exposure ratio. Let $q_{0.99}^{(i)}$ denote the 99th percentile of the i -th sampled low-light patch. Using their median as a robust training-set statistic and a target intensity τ , we define

$$s_{\mathcal{D}} = \frac{\tau}{\text{median}_i q_{0.99}^{(i)}}, \quad \tilde{x}_l = s_{\mathcal{D}} x_l.$$

The low-light input and its reconstruction target share this scale, whereas the GT branch remains unchanged. We avoid hard clipping and use one scale per dataset: it adapts across datasets but remains fixed across samples within each dataset, preserving reversibility and a stable latent distribution.

Phase B: Scale-Only Tokenizer Pretraining We first disable domain conditioning and pretrain the shared encoder-decoder to self-reconstruct $\tilde{x}_l$ and x_g . This scale-only baseline isolates the effect of reversible scaling and provides a stable initialization for DAAM. It uses an L1 reconstruction loss in both domains and an additional low-light gradient loss to preserve local structures.

Phase C: Domain-Adaptive Joint Training Starting from the scale-only checkpoint, we enable Domain-Adaptive Affine Modulation (DAAM) at every encoder and decoder stage. Given the scaled-low and GT domain codes m_l and m_g , DAAM modulates a feature map F using domain-specific parameters:

$$F' = F \odot (1 + \gamma_d) + \beta_d,$$

where $d \in \{0, 1\}$ indexes m_l and m_g , and the zero-initialized γ_d, β_d preserve the pretrained mapping at activation. With balanced scaled-low/GT mini-batches, the tokenizer is optimized by

$$\mathcal{L}_{tok} = \lambda_l (\|\hat{x}_l - \tilde{x}_l\|_1 + \lambda_{\nabla} \mathcal{L}_{grad}(\hat{x}_l, \tilde{x}_l)) + \lambda_g \|\hat{x}_g - x_g\|_1.$$

DAAM uses a separate learning rate $\eta_{\text{DAAM}} = \rho \eta_{\text{base}}$ with $\rho > 1$. We first warm up DAAM with the shared backbone frozen and then jointly optimize both parameter groups.

Stage II: One-Step Enhancement MeanFlow

With the Stage I tokenizer frozen, we train the Enhancement MeanFlow to transport scaled low-light latents to the GT latent distribution in one function evaluation. For a paired sample (x_l, x_g) from dataset $\mathcal{D}$, the corresponding DAAM branches produce the condition z_l and target z_g . Following MeanFlow, we interpolate the endpoints and construct the velocity target as

$$\begin{aligned} z_l &= E_{\phi}(s_{\mathcal{D}} x_l; m_l), \quad z_g = E_{\phi}(x_g; m_g), \\ z_t &= (1-t)z_g + tz_l, \quad v = z_l - z_g, \\ u_{\text{tgt}} &= v - (t-r)u_{\theta}, \end{aligned}$$

where $t, r \in [0, 1]$, and u_{θ} is the directional derivative of $u_{\theta}(z_t, t, r, z_l)$ computed by a Jacobian-vector product. Only the Enhancement MeanFlow is optimized using

$$\mathcal{L}_{mf} = \mathbb{E}_{t,r} \left[w \|u_{\theta}(z_t, t, r, z_l) - u_{\text{tgt}}\|_2^2 \right],$$

where w is an adaptive normalization weight.

Stage III: Retinex-Guided MeanFlow Refinement

Unlike other methods which enforce illumination-reflectance consistency from the very beginning, we initialize from the Stage II checkpoint and refine the Enhancement MeanFlow with training-only physical guidance after its predictions have become sufficiently reliable for Retinex decomposition. The tokenizer, differentiable ISP, and pre-trained Retinex network remain frozen. The restored RGB output and its decomposition are

$$\begin{aligned} \hat{z}_g &= \text{MeanFlow}_{\theta}(E_{\phi}(\tilde{x}_l; m_l)), \\ \hat{y}_g &= \text{ISP}(D_{\psi}(\hat{z}_g; m_g)), \quad y_g = \text{ISP}(x_g), \\ (\hat{L}_g, \hat{R}_g) &= \mathcal{R}(\hat{y}_g), \quad (L_g, R_g) = \mathcal{R}(y_g). \end{aligned}$$

Retinex consistency is measured by $\mathcal{L}_{ret} = \|\hat{L}_g - L_g\|_1 + \|\hat{R}_g - R_g\|_1$. We additionally use MSE for pixel fidelity and SAM for RGB color-direction consistency. The Stage II latent loss is disabled, and the refinement objective contains only these three terms:

$$\mathcal{L}_{stage3} = \lambda_{ret} \mathcal{L}_{ret} + \lambda_{mse} \mathcal{L}_{mse} + \lambda_{sam} \mathcal{L}_{sam}.$$

Only the Enhancement MeanFlow parameters are updated.

Blind Latent Deblurring MeanFlow

Motion-degraded inputs require structural recovery in addition to exposure correction. We therefore include Blind Latent Deblurring MeanFlow as the optional motion-restoration component of the MeanFlow group. It receives neither a blur kernel nor a forward operator and is trained through GT-domain initialization followed by cascade adaptation.

For initialization, paired motion-degraded and sharp GT RAW images (x_g^m, x_g) are encoded by the frozen tokenizer as condition z_g^m and target z_g . Unlike enhancement, the deblurring transport starts from independent Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$:

$$z_t = (1-t)z_g + t\epsilon, \quad u_{\theta_d} = u_{\theta_d}(z_t, t, r \mid z_g^m).$$

Only the MeanFlow objective is used in this initialization. We then cascade-adapt the deblurring model behind each frozen Enhancement MeanFlow using its actual motion-affected output as the condition; only θ_d is updated. Full objectives and settings are provided in the supplementary material.

Inference

At inference, all components are frozen and the Retinex network is discarded. For $x_{in} \in \{x_l, x_l^m\}$, the low-light DAAM branch and Enhancement MeanFlow first perform the one-step dark-to-light transformation. The output route is

$$\begin{aligned} z_{in} &= E_{\phi}(s_{\mathcal{D}} x_{in}; m_l), \quad \hat{z}_g^{\text{enh}} = \text{MeanFlow}_{\theta_e}^{\text{enh}}(z_{in}), \\ \hat{z}_g^{\text{out}} &= \begin{cases} \hat{z}_g^{\text{enh}}, & x_{in} = x_l, \\ \text{DeblurMeanFlow}_{\theta_d}(\epsilon \mid \hat{z}_g^{\text{enh}}), & x_{in} = x_l^m, \end{cases} \\ \hat{x}_g &= D_{\psi}(\hat{z}_g^{\text{out}}; m_g), \quad \hat{y}_g = \text{ISP}(\hat{x}_g), \quad \epsilon \sim \mathcal{N}(0, I). \end{aligned}$$

Thus, low-light-only inputs bypass motion restoration, whereas motion-degraded inputs activate the blind deblurring component before decoding. Retinex supervision is absent from both routes.

Method	0.01 lux			0.001 lux			0.0001 lux		
	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$
SID	20.73	0.745	0.094	19.21	0.708	0.095	19.60	0.695	0.103
LLPackNet	19.93	0.735	0.118	18.62	0.712	0.121	18.96	0.696	0.128
RRT	18.16	0.569	0.152	17.79	0.476	0.146	17.24	0.493	0.155
MCR	18.34	0.638	0.137	17.57	0.592	0.130	17.35	0.579	0.138
DNF	21.89	0.719	0.094	21.26	0.677	0.098	20.49	0.664	0.106
RAWMamba	20.11	0.660	0.127	19.96	0.617	0.124	18.77	0.596	0.133
SIED	<u>24.94</u>	<u>0.812</u>	0.061	<u>24.41</u>	<u>0.802</u>	<u>0.062</u>	<u>23.15</u>	<u>0.773</u>	0.068
Ours	26.03	0.816	<u>0.073</u>	25.73	0.814	0.061	25.04	0.774	<u>0.074</u>

Table 1: RGB-space quantitative comparison on the Sony subset of SIED at three illumination levels. Bold and underlined denote the best and second-best results, respectively; $\uparrow/\downarrow$ indicate higher/lower is better.

Method	0.01 lux			0.001 lux			0.0001 lux		
	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$
SID	17.62	0.614	0.122	15.61	0.548	0.122	15.90	0.569	0.134
LLPackNet	17.45	0.549	0.121	14.44	0.521	0.124	15.5	0.555	0.129
RRT	16.46	0.569	0.130	14.89	0.496	0.129	14.80	0.519	0.129
MCR	16.07	0.569	0.128	13.87	0.502	0.127	14.50	0.532	0.127
DNF	18.20	0.622	<u>0.113</u>	17.36	0.565	<u>0.111</u>	16.11	0.571	0.128
RAWMamba	<u>18.56</u>	0.611	0.114	<u>17.81</u>	0.549	0.114	<u>16.81</u>	0.565	<u>0.118</u>
SIED	17.77	<u>0.627</u>	0.135	16.80	<u>0.615</u>	0.148	16.12	<u>0.605</u>	0.162
Ours	24.62	0.662	0.111	24.65	0.681	0.081	24.23	0.654	0.100

Table 2: RGB-space quantitative comparison on the Sony subset of SIED at three illumination levels. Bold and underlined denote the best and second-best results, respectively; $\uparrow/\downarrow$ indicate higher/lower is better.

Experiments and Discussion

Experimental Setup

Datasets and Evaluation Protocols We evaluate all methods on the Sony subsets of SIED and SIED at 0.01, 0.001, and 0.0001 lux using the corresponding per-level test data.

SIED construction. For every selected split and illumination level, SIED retains each paired low-light/GT RAW sample and applies channel-wise motion degradation to both Bayer measurements, forming (x_l, x_g, x_l^m, x_g^m) . No independent Gaussian noise is added; the original sensor noise inherited from SIED is preserved. The original pair supports low-light-only evaluation, while the motion-degraded counterparts support GT-domain deblurring initialization and motion-aware testing. Complete generation details are provided in the supplementary material.

Implementation Details SIED uses the unified tokenizer and Enhancement MeanFlow, whereas SIED additionally activates the motion-restoration component for camera-shake degradation. All experiments are conducted on an NVIDIA GeForce RTX 4090, and every compared method is trained for the same number of optimization steps.

Compared Methods and Metrics Existing joint enhancement and deblurring methods operate on rendered RGB/sRGB inputs (Zhou, Li, and Loy 2022; Zhao et al. 2022; Li et al. 2024; Lv et al. 2024; Liu et al. 2025) and

are therefore not directly comparable to SIED’s extremely low-light RAW setting. Applying an ISP before restoration would alter the sensor noise, dynamic range, and motion characteristics, effectively defining a different task.

We compare with SID (Chen et al. 2018), LLPackNet (Lamba, Balaji, and Mitra 2020), RRT (Lamba and Mitra 2021), MCR (Dong et al. 2022), DNF (Jin et al. 2023), RAW-Mamba (Chen et al. 2026), and SIED (Jiang et al. 2025a). All outputs are evaluated in RGB space using PSNR, SSIM, and SAM; higher PSNR/SSIM and lower SAM are better.

Results on SIED

Quantitative Comparison Table 1 reports RGB-space PSNR, SSIM, and SAM at all illumination levels. Our method obtains PSNR values of 26.03, 25.73, and 25.04 dB and SSIM values of 0.816, 0.814, and 0.774 from 0.01 to 0.0001 lux. The limited PSNR decrease indicates stable enhancement as the signal weakens. Reporting each level avoids concealing illumination-dependent behavior in a single average.

Qualitative Comparison Figure 2 compares results under the same RAW-to-RGB pipeline. Our result preserves the red bicycle boundary and color at 0.01 lux, thin mechanical structures at 0.001 lux, and railing continuity at 0.0001 lux. These regions expose differences in exposure, artifacts, and fine-structure preservation not fully represented by global metrics.

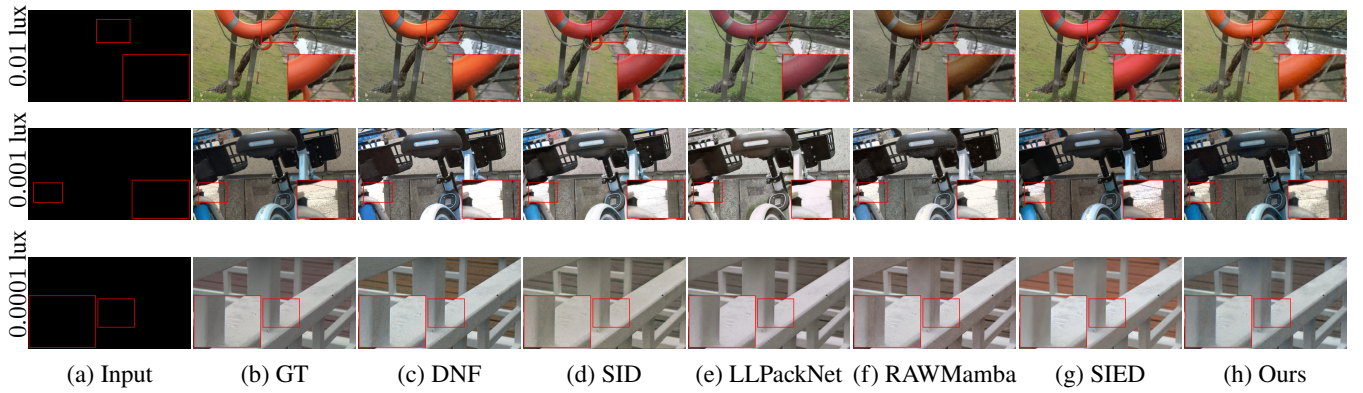


Figure 2: Qualitative comparison on the Sony subset of SIED without motion blur at 0.01, 0.001, and 0.0001 lux. From left to right: Input, GT, DNF, SID, LLPackNet, RAWMamba, SIED, and Ours. All results are visualized in RGB space under the same evaluation protocol.

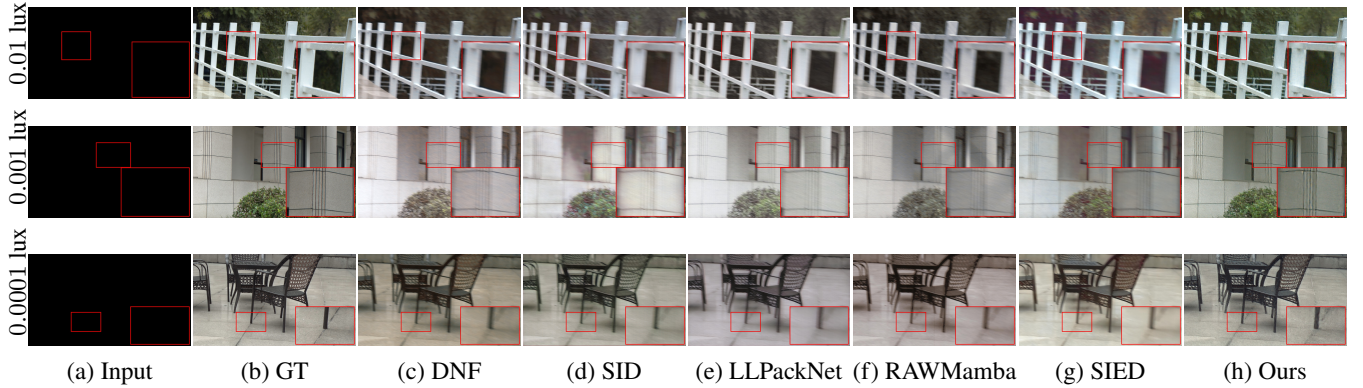


Figure 3: Qualitative comparison on the Sony subset of SIED with motion blur at 0.01, 0.001, and 0.0001 lux. From left to right: Input, GT, DNF, SID, LLPackNet, RAWMamba, SIED, and Ours. All results are visualized in RGB space under the same evaluation protocol. Zoom in to see more details.

Results on SIED

Quantitative Comparison Table 2 evaluates motion-aware restoration under the same RGB protocol. Our method records 24.62, 24.65, and 24.23 dB PSNR, SSIM between 0.654 and 0.681, and SAM at or below 0.111. This stability contrasts with the larger degradation of competing methods and indicates effective restoration when motion accompanies the weakening signal.

Qualitative Comparison Figure 3 shows the motion-degraded input x_i^m followed by restored results. Across the three levels, vertical fence bars, wall boundaries, shrub contours, chair frames, and floor seams expose residual blur. Competing results often retain directional smearing or merge structures, whereas ours restores more coherent boundaries, exposure, and color.

Discussion

On SIED, our method achieves the highest PSNR and SSIM at every illumination level, exceeding SIED by 1.09, 1.32, and 1.89 dB as illumination decreases. It obtains the best SAM only at 0.001 lux; SIED remains better at 0.01 and

0.0001 lux. Thus, improved pixel fidelity and structural similarity do not uniformly yield the best spectral-angle consistency, leaving color-direction preservation as a remaining challenge. Figure 2 nevertheless confirms stronger preservation of thin bicycle structures, railing details, and local contrast with plausible exposure.

On SIED, our method ranks first in all metrics at every level, improving the strongest competing PSNR by 6.06, 6.84, and 7.42 dB. Its PSNR also remains stable at 24.23–24.65 dB, while competing methods degrade more substantially. Figure 3 shows sharper fence bars, building boundaries, chair frames, and floor seams, while competing restorations often retain directional blur. These results support activating motion restoration only for motion-degraded inputs.

The three metrics provide complementary evidence in this setting. PSNR measures pixel-wise agreement after the common RAW-to-RGB conversion, SSIM emphasizes structural organization, and SAM reflects color-direction consistency. Their joint improvement on SIED indicates that the gain is not explained solely by brighter outputs or local contrast adjustment. Moreover, the narrow PSNR range across the three illumination levels suggests that the motion-aware route re-

Variant	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$
Tokenizer w/o DAAM	<u>47.62</u>	<u>0.900</u>	<u>0.086</u>
Tokenizer w/ DAAM	50.11	0.923	0.060

Table 3: Standalone tokenizer recovery with and without DAAM, evaluated in rendered RGB space after the frozen ISP; MeanFlow and Retinex-guided refinement are excluded.

mains stable as photon information becomes increasingly limited. Conversely, the SIED SAM results show that agreement should be evaluated per degradation regime instead of inferred from PSNR.

Across SIED, our PSNR decreases by only 0.99 dB from 0.01 to 0.0001 lux, while its advantage over SIED grows from 1.09 to 1.89 dB. Under motion degradation, the improvement over the strongest competing result increases from 6.06 to 7.42 dB. This widening gap suggests that exposure-and-noise methods are vulnerable when blur removes scarce structural cues. It is not an isolated deblurring gain because illumination, noise, and motion are evaluated jointly on SIED.

Together, SIED evaluates one-step dark-to-light transformation, while SIED additionally stresses structures damaged by camera motion. Because their degradation distributions differ, the larger SIED margin is not a component ablation; it instead shows that exposure-and-noise methods transfer poorly to realistic accompanying motion, while the proposed motion-aware route remains robust.

The MeanFlow group uses degradation-specific restoration routes. SIED inputs are decoded after enhancement, whereas motion-degraded SIED inputs additionally pass through Blind Latent Deblurring MeanFlow. Initialized on GT-domain motion pairs, the deblurring component is cascade-adapted to actual outputs of the frozen Enhancement MeanFlow. Updating only its parameters preserves the dark-to-light mapping while aligning its condition with enhancement-generated latents. The shared backbone thereby supports both low-light-only and motion-degraded restoration.

Ablation Studies

Effect of DAAM on Tokenizer Recovery We evaluate standalone tokenizer recovery before MeanFlow and Retinex refinement. Each input is reconstructed through its domain branch. The ablation removes DAAM from the encoder and decoder while retaining dataset-adaptive scaling, the shared backbone, and the training/evaluation settings. This comparison isolates DAAM’s effect on scaled-low and GT RAW recovery using PSNR, SSIM, and SAM.

Table 3 shows that DAAM raises PSNR by 2.49 dB and SSIM from 0.900 to 0.923, while reducing SAM from 0.086 to 0.060 (30.2%). Thus, domain-specific modulation helps the shared tokenizer accommodate distinct scaled-low and GT statistics while improving pixel, structural, and color-direction fidelity.

Effect of Retinex-Guided Stage III Refinement We evaluate the effect of Stage III refinement by comparing the Stage

Variant	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$
Stage II only	26.80	<u>0.781</u>	<u>0.198</u>
Stage II + Stage III	<u>25.86</u>	0.813	0.074

Table 4: Effect of Retinex-guided Stage III refinement. Retinex is training-only; bold and underlined denote the best and second-best results.

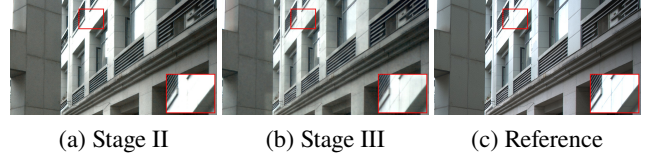


Figure 4: Local Stage III ablation. Stage III recovers the reference’s cross-shaped facade seam suppressed by Stage II despite its lower PSNR. Zoom in to see more details.

II checkpoint against its Retinex-guided counterpart. As Table 4 shows, Stage III significantly improves structural and color fidelity, raising SSIM from 0.781 to 0.813 and reducing SAM by 62.6%. Although PSNR drops by 0.94 dB, this MSE-derived metric intrinsically favors overly smoothed estimates and penalizes slight high-frequency spatial shifts. By explicitly enforcing physical illumination-reflectance consistency, Stage III prioritizes perceptually faithful recovery over pixel-wise smoothing. For instance, Figure 4 demonstrates that Stage III successfully restores the cross-shaped facade seam flattened by Stage II. Together, the quantitative metrics and local visual evidence confirm genuine structural and color-direction improvements. Finally, since the Retinex guidance is training-only, these robust gains incur zero additional inference cost.

Conclusion

We presented a RAW restoration framework for extremely dark handheld imaging, where severe illumination loss may be accompanied by motion degradation. SIED extends paired extremely low-light evaluation to this practical setting, enabling assessment of both low-light-only and motion-degraded inputs. Our unified domain-adaptive tokenizer maps scaled low-light and well-exposed RAW measurements into a shared latent space, while a task-aware MeanFlow group performs one-step enhancement and activates blind latent motion restoration only when required. Training-only physics guidance further improves illumination-reflectance consistency, structural preservation, and color fidelity without altering inference. Experiments on SIED show leading PSNR and SSIM across three illumination levels, while SIED results demonstrate consistent gains in PSNR, SSIM, and SAM under motion degradation. Ablations validate DAAM and Stage III refinement while revealing the trade-off between pixel-wise PSNR and local structural recovery. Future work will extend this framework to unconstrained real-world acquisitions with diverse sensors and dynamic scenes.

References

- Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12504–12513.
- Chang, W.-G.; You, T.; Seo, S.; Kwak, S.; and Han, B. 2019. Domain-Specific Batch Normalization for Unsupervised Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to See in the Dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Chen, X.; Han, L.; Huang, P.; Feng, X.; Zhang, D.; and Han, J. 2026. Retinex-RAWMamba: Bridging Demosaicing and Denoising for Low-Light RAW Image Enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(1): 406–420.
- Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; and Ko, S.-J. 2021. Rethinking Coarse-To-Fine Approach in Single Image Deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4641–4650.
- Dong, X.; Xu, W.; Miao, Z.; Ma, L.; Zhang, C.; Yang, J.; Jin, Z.; Teoh, A. B. J.; and Shen, J. 2022. Abandoning the Bayer-Filter To See in the Dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17431–17440.
- Geng, Z.; Deng, M.; Bai, X.; Kolter, Z.; and He, K. 2025. Mean Flows for One-Step Generative Modeling. In *Advances in Neural Information Processing Systems*, volume 38, 75460–75482. Curran Associates, Inc.
- Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Jiang, H.; Guan, B.; Liu, Z.; Liu, X.; Yu, J.; Liu, Z.; Han, S.; and Liu, S. 2025a. Learning to See in the Extremely Dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7676–7685.
- Jiang, Y.; Zhang, Z.; Xue, T.; and Gu, J. 2025b. AutoDIR: Automatic All-in-One Image Restoration with Latent Diffusion. In *Computer Vision – ECCV 2024*, 340–359. Cham: Springer Nature Switzerland.
- Jin, X.; Han, L.-H.; Li, Z.; Guo, C.-L.; Chai, Z.; and Li, C. 2023. DNF: Decouple and Feedback Network for Seeing in the Dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18135–18144.
- Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising Diffusion Restoration Models. In *Advances in Neural Information Processing Systems*, volume 35, 23593–23606. Curran Associates, Inc.
- Kong, L.; Dong, J.; Ge, J.; Li, M.; and Pan, J. 2023. Efficient Frequency Domain-Based Transformers for High-Quality Image Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5886–5895.
- Lamba, M.; Balaji, A.; and Mitra, K. 2020. Towards Fast and Light-Weight Restoration of Dark Images. arXiv:2011.14133.
- Lamba, M.; and Mitra, K. 2021. Restoring Extremely Dark Images in Real Time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3487–3497.
- Li, Y.; Xu, R.; Niu, Y.; Guo, W.; and Zhao, T. 2024. Perceptual Decoupling With Heterogeneous Auxiliary Tasks for Joint Low-Light Image Enhancement and Deblurring. *IEEE Transactions on Multimedia*, 26: 6663–6675.
- Lin, X.; He, J.; Chen, Z.; Lyu, Z.; Dai, B.; Yu, F.; Qiao, Y.; Ouyang, W.; and Dong, C. 2025. DiffBIR: Toward Blind Image Restoration with Generative Diffusion Prior. In *Computer Vision – ECCV 2024*, 430–448. Cham: Springer Nature Switzerland.
- Lipman, Y.; Chen, R. T. Q.; Ben-Hamu, H.; Nickel, M.; and Le, M. 2023. Flow Matching for Generative Modeling. arXiv:2210.02747.
- Liu, C.; Wang, X.; Xu, X.; Tian, R.; Li, S.; Qian, X.; and Yang, M.-H. 2024. Motion-Adaptive Separable Collaborative Filters for Blind Motion Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25595–25605.
- Liu, M.; Cui, Y.; Ren, W.; Zhou, J.; and Knoll, A. C. 2025. LIEDNet: A Lightweight Network for Low-Light Enhancement and Deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(7): 6602–6615.
- Lv, X.; Zhang, S.; Wang, C.; Zheng, Y.; Zhong, B.; Li, C.; and Nie, L. 2024. 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)*, 25378–25388.
- Perez, E.; Strub, F.; de Vries, H.; Dumoulin, V.; and Courville, A. 2018. FiLM: Visual Reasoning with a General Conditioning Layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695.
- Su, M.; Hu, J. Y.-C.; Lee, Y.-C.; Zhu, N.; Chung, J.-H.; Wu, S.; Song, Z.; Chen, M.; and Liu, H. 2025. High-Order Flow Matching: Unified Framework and Sharp Statistical Rates. In Belgrave, D.; Zhang, C.; Lin, H.; Pascanu, R.; Koniusz, P.; Ghassemi, M.; and Chen, N., eds., *Advances in Neural Information Processing Systems*, volume 38, 45820–45932. Curran Associates, Inc.
- Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023. Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(3): 2654–2662.

- Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep Retinex Decomposition for Low-Light Enhancement. arXiv:1808.04560.
- Wei, K.; Fu, Y.; Yang, J.; and Huang, H. 2020. A Physics-Based Noise Formation Model for Extreme Low-Light RAW Denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. 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)*, 5901–5910.
- Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-Aware Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17714–17724.
- Yi, X.; Xu, H.; Zhang, H.; Tang, L.; and Ma, J. 2023. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12302–12311.
- Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5728–5739.
- Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the Darkness: A Practical Low-light Image Enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, 1632–1640. Association for Computing Machinery.
- Zhao, Y.; Xu, Y.; Yan, Q.; Yang, D.; Wang, X.; and Po, L.-M. 2022. D2HNet: Joint Denoising and Deblurring with Hierarchical Network for Robust Night Image Restoration. In *Computer Vision – ECCV 2022*, 91–110. Cham: Springer Nature Switzerland.
- Zhou, S.; Li, C.; and Loy, C. C. 2022. LEDNet: Joint Low-Light Enhancement and Deblurring in the Dark. In *Computer Vision – ECCV 2022*, 573–589. Cham: Springer Nature Switzerland.
- Zhu, N.; Wang, J.; Zhang, Y.; Wang, H.; and Han, T. 2024. An Adversarial Training Framework Based on Unsupervised Feature Reconstruction Constraints for Crystalline Silicon Solar Cells Anomaly Detection. *IEEE Transactions on Instrumentation and Measurement*, 73: 1–13.
- Zhu, Y.; Zhang, K.; Liang, J.; Cao, J.; Wen, B.; Timofte, R.; and Van Gool, L. 2023. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1219–1229.