

Joint Motion Deblurring and Low-Light Image Enhancement on Raw Images

Weijie Cai and Yu-Hua Chen

Abstract—Capturing high-quality images in low-light environments remains a fundamental challenge in computational photography. While motion blur is often mitigated in these settings by employing fast shutter speeds and high ISO gains, it remains an unavoidable artifact in specific scenarios, particularly in video acquisition where exposure time is intrinsically constrained by the frame rate. In this work, we investigate the adaptability of state-of-the-art (SOTA) models to the task of joint low-light enhancement and deblurring in the RAW domain. Specifically, we examine whether DarkIR, a recent SOTA model in the sRGB domain, can be effectively applied to RAW sensor data, and we attempt to extend ExposureDiffusion, the current SOTA in low-light enhancement, to handle motion blur. To facilitate this, we construct a new synthetic dataset containing realistic motion blur and extend the DarkIR architecture with a *Metadata-Aware Encoder* that leverages raw image metadata. Additionally, we attempt to extend ExposureDiffusion by incorporating blur kernels into its training process. Our experiments reveal that both models exhibit distinct limitations in this domain: pretrained DarkIR struggles significantly with the heavy noise inherent to low-light RAW images, though fine-tuning with metadata proves substantially beneficial. Conversely, the pretrained ExposureDiffusion model fails to handle blur effectively and proves difficult to fine-tune for geometric degradations without architectural modifications.

Index Terms—Computational Photography, Low-Light Image Enhancement, Motion Deblurring, Raw Image Processing



1 INTRODUCTION

Acquiring high-quality images in low-light environments remains a central problem in computational photography. When lighting conditions are poor, the resulting images usually have three issues: excessive noise, wrong colors, and blurriness from long exposures or camera shake. Addressing these issues requires a system capable of solving all three problems simultaneously.

Current learning-based approaches generally fall into two categories: those operating on processed sRGB images and those operating on raw sensor data. In the sRGB domain, methods like DarkIR [1] have shown promise but are limited by the non-linearities of the Image Signal Processing (ISP) pipeline. Furthermore, these models typically treat restoration as a purely visual task, ignoring critical capture metadata (such as ISO) that correlates with noise levels. Conversely, in the raw domain, ExposureDiffusion [2] has established itself as the state-of-the-art on the Seeing in the Dark (SID) dataset by leveraging the linearity and high bit-depth of raw signals. Nevertheless, these robust raw-domain models excel at denoising and low-light enhancement but fail to address motion blur.

To address these challenges, we develop a comprehensive approach that advances both sRGB and raw-domain methods for low-light deblurring. Recognizing that existing datasets lack images with both extreme noise and realistic motion blur, we create a new synthetic dataset that fills this gap and enables proper training of joint deblurring-denoising-LLIE models. For sRGB-domain processing, we augment the DarkIR architecture with ISO-aware mechanisms that provide physical guidance during training and inference. In the raw domain, we adapt ExposureDiffusion,

which was previously successful only for static scenes, to handle dynamic content through fine-tuning on our synthetic data.

Our main contributions are summarized as follows:

- **Synthetic Dataset Generation:** We construct a comprehensive dataset derived from SID, providing synthetically generated training and testing sets that simulate realistic motion blur for low-light scenarios.
- **Metadata-Guided sRGB Architecture:** We propose a novel enhancement to the sRGB-based DarkIR model by injecting ISO metadata, allowing the network to adaptively handle noise without altering the underlying domain.
- **ExposureDiffusion Adaptation:** We fine-tune the state-of-the-art raw-domain model, ExposureDiffusion, on our synthetic dataset, extending its capabilities to robustly handle joint deblurring and denoising.

2 RELATED WORK

2.1 Joint Low-Light Enhancement and Deblurring

Simultaneous low-light enhancement and deblurring is a challenging task due to the complex degradation relationship between noise and blur. Early attempts, such as LEDNet [3], noted that a direct combination or sequential application of enhancement and deblurring models lacks synergy; specifically, enhancement steps often magnify blur or remove cues necessary for effective deconvolution. To address this, they proposed an end-to-end model to perform the tasks jointly.

Subsequent works improved upon this foundation. FourierDiff [4] achieved better perceptual metrics by embedding Fourier priors into a pre-trained diffusion model. More

• W. Cai and Y.-H. Chen are with the Department of Computer Science, University of Toronto, Toronto, ON, M5S 1A1.
E-mail: {wjcai, chenjoachim}@cs.toronto.edu

recently, DarkIR [1] set a new state-of-the-art (SOTA) by designing a lightweight network utilizing both spatial and frequency information. However, a significant limitation of these methods is their operation in the sRGB domain. sRGB images have compressed dynamic range and lack critical metadata about capture settings, limiting the model’s ability to recover details in extreme low-light conditions.

2.2 Image Deblurring in the RAW Domain

While deep learning-based image restoration has flourished in the sRGB domain, research specifically targeting RAW-to-RAW deblurring remains comparatively nascent, despite the inherent advantages of operating in linear, device-independent space. Foundational work by Liang [5] demonstrated that deblurring models trained solely on RAW data significantly outperform their sRGB counterparts, while also offering superior generalization to unseen sensors through fine-tuning. Addressing the need for physically accurate training data, the *RawIR* framework [6] introduced a comprehensive degradation pipeline modeling sensor noise, motion blur, and camera shake. By integrating this pipeline with the NAFNet architecture, *RawIR* achieves robust restoration capabilities on real-world inputs. Concurrently, architectural innovations have focused on efficiency; *ELMformer* [7] proposes a Locally Multiplicative Transformer tailored to the specific characteristics of RAW data, achieving top-tier performance on denoising and deblurring benchmarks with minimal computational cost. Most recently, frequency-domain approaches have gained traction, with *FrENet* [8] utilizing adaptive frequency modulation to surpass existing state-of-the-art methods in both restoration quality and efficiency. Collectively, these studies validate the efficacy and promise of performing deblurring directly in the RAW domain.

2.3 Low-light Image Enhancement in the RAW Domain

The superiority of processing low-light images in the RAW domain over the processed sRGB domain is well-established, driven by the preservation of linear signal characteristics and high dynamic range. Early data-driven approaches sought to replace traditional Image Signal Processing (ISP) pipelines; notably, Chen et al. [9] pioneered the “Seeing in the Dark” (SID) framework, demonstrating that fully convolutional networks operating directly on raw sensor data could recover viable images where traditional pipelines fail. Building on this, Wei et al. [10] argued that the performance of such networks is heavily dependent on noise assumptions, proposing the “Extreme Low-light Denoising” (ELD) framework which utilizes a physics-based noise formation model to characterize complex sensor noise, thereby achieving higher restoration accuracy. More recently, Wang et al. [2] addressed the ill-posed nature of enhancement by introducing *ExposureDiffusion*, a method that integrates a physics-based exposure formulation into a diffusion-based restoration model to learn a continuous exposure process in the RAW domain, progressively refining low-light inputs into well-exposed images while maintaining photometric consistency. However, while these methods excel at photometric enhancement and denoising, they predominantly model the scene as static, and therefore

do not address geometric degradations—specifically motion blur—that can arise in practice from the slightly longer exposure times used in low-light photography.

3 PROPOSED METHOD

We formally define the task of joint motion deblurring and low-light image enhancement as follows. Let $x \in \mathbb{R}^{H \times W}$ denote the input RAW image with a Bayer pattern. Let $\mathbf{m} \in \mathbb{R}^{d_m}$ represent the associated metadata vector extracted from the RAW file (e.g., ISO, lens specifications), where d_m is the dimensionality of the parameter space. Correspondingly, let $y \in \mathbb{R}^{H \times W \times 3}$ denote the ground-truth normal-exposure, deblurred sRGB image.

Our objective is to learn a mapping function $\mathcal{F}$ that reconstructs a clean, high-quality image from the degraded RAW input and its metadata:

$$\hat{y} = \mathcal{F}(x, \mathbf{m}) \approx y \quad (1)$$

The remainder of this section is organized as follows: Section 3.1 details the methodology used to construct our synthetic dataset for training and benchmarking. Section 3.2 presents our architecture, which adapts DarkIR [1] for RAW image processing. Finally, Section 3.3 describes how we leverage this dataset within the *ExposureDiffusion* framework to improve performance on blurry inputs.

3.1 Synthetic Dataset

We adopt the physics-based noise model for digital RAW images introduced in the ELD dataset [10], formulated as:

$$D = KI + N \quad (2)$$

where K denotes the overall system gain, I represents the number of incident photoelectrons, and N encapsulates the total noise introduced during image capture.

The SID dataset [9] consists of short-exposure and long-exposure image pairs captured under identical settings, differing only in exposure time. Let $r \in [100, 1000]$ denote the ratio between the long and short exposure times. We express the relationship for both RAW images as:

$$D_{short} = KI + N_{short} \quad (3)$$

$$D_{long} = rKI + N_{long} \quad (4)$$

By scaling the long-exposure image by $1/r$, we can approximate the clean signal:

$$\frac{1}{r}D_{long} = KI + \frac{1}{r}N_{long} \approx KI \quad (5)$$

Here, we approximate the noise term $\frac{1}{r}N_{long}$ to zero, as r is a sufficiently large value, rendering the scaled noise negligible. Consequently, we can isolate the noise component of the short-exposure image, N_{short} , by subtracting the approximated clean signal from the raw input:

$$N_{short} = D_{short} - KI \approx D_{short} - \frac{1}{r}D_{long} \quad (6)$$

Finally, to generate pairs of blurred, low-light images and their clear counterparts, we apply a random motion

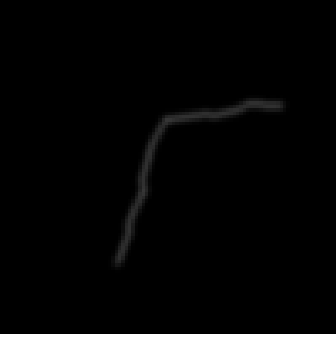


Fig. 1: An example of a synthetically generated motion kernel ($d_x = d_y$).

kernel M based on the implementation by [11]. The kernel generation begins by sampling a maximum trajectory length:

$$L_{max} = [\mathcal{U}(0, 1) + \mathcal{U}(0, \lambda^2)] \times \sqrt{d_x^2 + d_y^2} \times 0.75 \quad (7)$$

where λ represents the intensity and (d_x, d_y) denotes the kernel dimensions. The trajectory is constructed by iteratively sampling the step size r_t and angle θ_t until the maximum length is reached:

$$r_t = \text{Beta}(1, 30) \times (1 - \lambda + \epsilon) \times \sqrt{d_x^2 + d_y^2} \quad (8)$$

where $\epsilon = 0.1$ is a constant added for numerical stability. The trajectory angle is sampled as:

$$\theta_{max} = \mathcal{U}(0, \lambda\pi) \quad (9)$$

$$\theta_t = \text{Triangular}(0, \lambda\theta_{max}, \theta_{max} + \epsilon) \quad (10)$$

To simulate jitter, the trajectory direction is inverted relative to the previous step with a probability $J \sim \text{Beta}(2, 20)$. We generate a unique kernel for each image instance, sampling square dimensions $d_x = d_y \sim \mathcal{U}(30, 80)$ and fixing the intensity at $\lambda = 0.5$. A representative motion kernel is illustrated in Figure 1.

Given the generated kernel M , we synthesize the blurry, low-light RAW image D'_{short} by convolving the clean signal approximation with the kernel and injecting the extracted noise:

$$D'_{short} = (KI) * M + N_{short} \approx \left(\frac{1}{r} D_{long}\right) * M + \left(D_{short} - \frac{1}{r} D_{long}\right) \quad (11)$$

To generate the corresponding ground truth, we process the long-exposure RAW image D_{long} through a standardized ISP pipeline (implemented via `rawpy`) to obtain the clean sRGB image y . Consequently, our final dataset consists of pairs mapping the synthetic blurry RAW input D'_{short} to the high-quality sRGB target y .

3.2 Adaptation of DarkIR

We construct an end-to-end pipeline based on the DarkIR framework [1], which has achieved state-of-the-art results in joint low-light image enhancement (LLIE) and deblurring benchmarks. To bridge the domain gap between RAW

inputs and the sRGB-trained architecture, we introduce two key modifications, as illustrated in Figure 2.

First, we address the bit-depth limitation. The original DarkIR model operates on 8-bit sRGB inputs, confining the dynamic range to 255 discrete levels. To mitigate this, we configure the ISP (implemented via `rawpy`) to output a 16-bit array. Since inputs are normalized to $[0, 1]$ prior to processing, this adjustment preserves high-fidelity spectral details while maintaining compatibility with the model’s expected input domain.

Second, we design a **Metadata-Aware Encoder**. Camera metadata provides critical physical context for restoration, yet is often discarded in traditional pipelines. We extract and normalize the following five attributes to condition the model:

- **Exposure Time t :** This parameter governs signal accumulation and is the most significant factor in low-light imaging. We apply logarithmic normalization:

$$\hat{t} = \frac{\log_{10}(t) + 4}{6} \quad (12)$$

- **ISO Value S_{ISO} :** The ISO setting directly correlates with signal gain and noise magnitude. It is normalized logarithmically as:

$$\hat{S}_{ISO} = \frac{\log_{10}(S_{ISO})}{6} \quad (13)$$

- **F-Number N_f :** The aperture size determines both incoming light flux and depth of field. Conditioning on N_f assists the model in distinguishing optical defocus from motion blur. We apply linear normalization:

$$\hat{N}_f = \frac{N_f}{32} \quad (14)$$

- **Focal Length f :** Jointly with the F-number, the focal length defines the field of view. We normalize it linearly:

$$\hat{f} = \frac{f}{500} \quad (15)$$

- **White Balance $(r_{mul}, g_{mul}, b_{mul})$:** Since the pre-training domain (DarkIR) and our target domain (SID) may differ in color space, explicit white balance information aids in learning the correct color mapping. We normalize the red and blue gains relative to the green gain:

$$\hat{r}_{ratio} = \frac{r_{mul}}{5g_{mul}}, \quad \hat{b}_{ratio} = \frac{b_{mul}}{5g_{mul}} \quad (16)$$

These parameters are integrated into the architecture via a linear projection layer added to each frequency-domain feed-forward layer (FreMLP). Building on the observation from [1], [4] that illumination information is primarily concentrated in the spectral amplitude, we project the metadata vector to match the channel dimension of the FreMLP module. This projected vector is then spatially broadcast and added directly to the amplitude component of the feature map, as illustrated in Figure 3.

In our experiments, we evaluate three configurations: (1) direct inference using the pre-trained DarkIR model, (2) fine-tuning DarkIR on our dataset without modification, and (3) fine-tuning the proposed metadata-aware DarkIR. We adopt

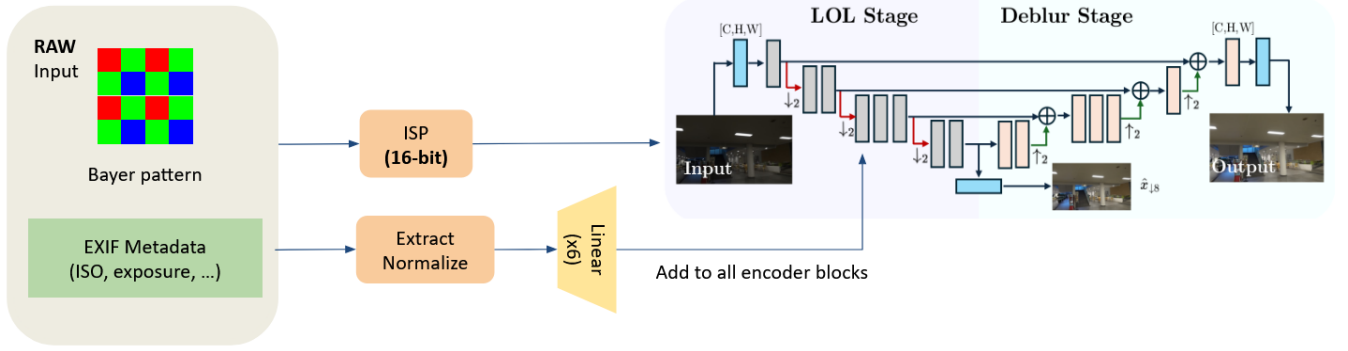


Fig. 2: Overview of our adapted DarkIR architecture. Note that the linear layers are distinct for different encoder blocks in LOL stage.

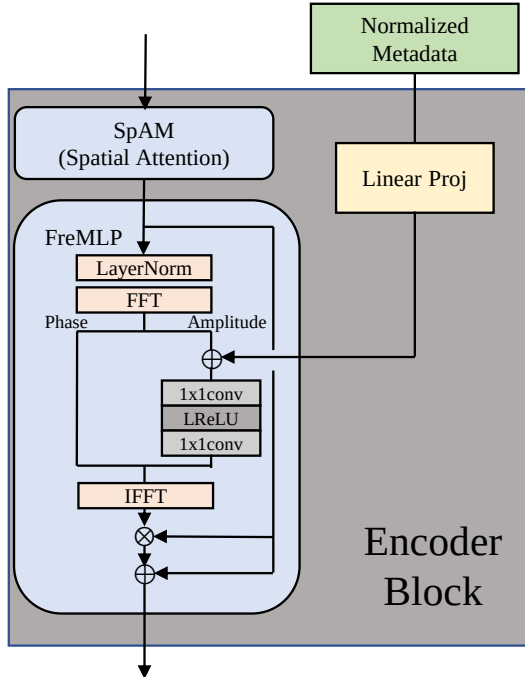


Fig. 3: Metadata-guided encoder block architecture. We build upon DarkIR’s [1] SpAM and FreMLP modules, injecting normalized ISO information through a linear projection layer that modulates the spectral amplitude within the frequency-domain feed-forward network.

the hybrid loss function from the original implementation, comprising a reconstruction loss ($\mathcal{L}_{main}$, Charbonnier), perceptual loss ($\mathcal{L}_{percep}$, VGG-based), edge loss ($\mathcal{L}_{edge}$, Laplacian), and a multi-scale reconstruction loss ($\mathcal{L}_{lol}$). The total objective is formulated as:

$$\mathcal{L} = \lambda_p \cdot \mathcal{L}_{main} + \lambda_{pe} \cdot \mathcal{L}_{percep} + \lambda_{ed} \cdot \mathcal{L}_{edge} + \mathcal{L}_{lol} \quad (17)$$

Following the settings in [1], we adopt the weighting coefficients $\lambda_p = 1$, $\lambda_{pe} = 0.01$, and $\lambda_{ed} = 50$. For all settings, we fine-tuned on our synthetic dataset (training split) for 10 epochs.

3.3 Fine-tuning ExposureDiffusion

To address the limitations of deterministic feed-forward networks and the computational inefficiency of vanilla diffusion models in low-light image enhancement, Wang et al. [2] propose *ExposureDiffusion*, a framework that integrates a physics-based exposure model into the diffusion process. Unlike standard diffusion approaches that initiate restoration from pure Gaussian noise, ExposureDiffusion treats the input low-light image as an intermediate noisy state (X_T) generated by a simulated exposure degradation process. This allows the reverse diffusion process to start directly from the low-light input, significantly reducing the number of inference steps and accelerating convergence. The method employs a diffusion model to approximate the exposure process and introduces an adaptive residual layer to effectively manage the trade-off between noise removal and detail preservation in well-exposed regions. By leveraging this physical prior, the model achieves robust enhancement and denoising on raw data without the need for extensive Gaussian denoising steps common in traditional diffusion-based restoration.

To address both low-light conditions and motion blur, we modify the degradation model by injecting blur kernels into the training set. Consequently, the initial state X_T in the diffusion process encapsulates both photometric (low exposure, noise) and geometric (blur) degradations. This formulation requires the reverse diffusion process to simultaneously handle exposure adjustment and deconvolution.

4 EXPERIMENTAL RESULTS

In this section, we present the experimental results to validate the performance of our adapted DarkIR model and the fine-tuned ExposureDiffusion model. For each architecture, we first report the quantitative comparisons across datasets, as summarized in Table 1, to evaluate its robustness against synthetic blur. Then, we present a quantitative and qualitative analysis of the fine-tuning process to demonstrate the effectiveness of our adaptation methods.

4.1 Adaptation of DarkIR

4.1.1 Analysis of Blur Robustness

We first examined the pretrained DarkIR model’s sensitivity to motion blur. We observed that the original model tends

TABLE 1: Quantitative comparison of DarkIR and ExposureDiffusion on the SID and SID-Blurred datasets. “FT” denotes Fine-Tuned models.

Method	PSNR $\uparrow$	SSIM $\uparrow$
<i>Dataset: SID (Original)</i>		
DarkIR (Pretrained)	14.16	0.3937
DarkIR (Denoise + Pretrained)	13.35	0.5025
ExposureDiffusion (Pretrained)	36.54	0.8823
<i>Dataset: SID-Blurred</i>		
DarkIR (Pretrained)	11.36	0.2340
DarkIR (Denoise + Pretrained)	13.49	0.5622
DarkIR (FT w/o EXIF)	17.83	0.6117
DarkIR (FT w/ EXIF)	18.65	0.6240
ExposureDiffusion (Pretrained)	33.15	0.8377
ExposureDiffusion (FT)	28.57	0.7345

to amplify noise, leading to severe visual degradation. To address this, we also evaluated a pipeline where the image is denoised using NAFNet [12] before being processed by DarkIR.

As shown in Table 1, the original model suffers a performance drop, with PSNR falling from 14.16 dB on the clean SID dataset to 11.36 dB on the SID-Blurred dataset. However, when the input is denoised first, the model maintains comparable PSNR and achieves even higher SSIM on the blurred dataset. This suggests that the DarkIR model itself is robust to blur, but its performance is primarily hurt by the high-ISO noise and color space discrepancies in the dataset.

4.1.2 Fine-tuning Results and Analysis

To address the low PSNR observed in both baseline models, we fine-tuned the original DarkIR model on our synthetic dataset with RAW image metadata (EXIF) information injected into the architecture. As shown in Table 1, the fine-tuned model significantly improves PSNR from 11.36 dB (pretrained model) to 18.65 dB, while SSIM increases from 0.2340 to 0.6240. Notably, it also surpasses the two-stage pipeline that first preprocesses images via denoising, which achieves a PSNR of 13.49 dB and SSIM of 0.5622. These quantitative results demonstrate that our architectural adaptation and fine-tuning strategy successfully enhance DarkIR’s performance on the joint task of motion deblurring and low-light image enhancement.

To further investigate whether the injected EXIF information contributes to this improved performance, we conduct an ablation study by fine-tuning a DarkIR model with the original architecture (without EXIF injection). As shown in Table 1, the model with EXIF information outperforms the vanilla architecture on both PSNR and SSIM metrics. This confirms that encoding camera settings metadata meaningfully contributes to improving DarkIR’s performance on the SID-Blurred dataset.

4.1.3 Qualitative Analysis

In Figure 4e, 4f, 4g, we present a visual comparison between the pretrained DarkIR model, the two-stage pipeline, and our fine-tuned model with EXIF injection. We observe that the pretrained DarkIR model fails to handle the domain shift, amplifying the severe noise to produce a heavily distorted, unusable image. In contrast, the two-stage pipeline

(denoising prior to DarkIR) successfully suppresses the noise, revealing the underlying structural content, though it fails to fully correct the color shift.

Finally, our fine-tuned model achieves the best visual quality, successfully recovering the scene structure and restoring a color space consistent with the ground truth. However, some blur remains in the predicted image. We hypothesize that this is due to the competing objectives of the joint task: given that DarkIR is a lightweight model, simultaneously handling severe noise removal and large-motion deblurring (up to 80 pixels) creates a capacity bottleneck. The optimization process likely prioritizes denoising and global color recovery, which significantly reduce the loss. This results in slightly over-smoothed textures and a compromise in high-frequency detail restoration.

4.2 Adaptation of ExposureDiffusion Model

4.2.1 Analysis of Blur Robustness

We conducted a comprehensive evaluation to assess the robustness of the original *ExposureDiffusion* model against motion blur. Quantitatively, the model exhibits a significant performance degradation, dropping from 36.54 dB on the clean SID dataset to 33.15 dB on the SID-Blurred dataset (see Table 1). To corroborate these metrics with qualitative evidence, we present a failure cases in Figure 4d. This visual example demonstrates that while the model successfully corrects exposure, it fails to deconvolve the geometric degradation, resulting in outputs that retain significant blur artifacts. This confirms that the pre-trained priors are specific to photometric noise and do not generalize to unseen geometric distortions.

4.2.2 Fine-tuning Results and Analysis

To address the observed limitations, we attempted to fine-tune the *ExposureDiffusion* model by injecting blur kernels directly into the training data. However, contrary to the expectation that the model would adapt to the new degradation, this strategy led to a significant performance regression: the PSNR dropped from 33.15 dB (baseline) to 28.57 dB after fine-tuning. We posit that this degradation stems from the complexity of the joint optimization task; requiring the model to simultaneously learn deblurring and exposure correction creates a challenging learning landscape.

Moreover, the original *ExposureDiffusion* formulation implicitly assumes a static-scene imaging model, in which the number of photons arriving at each pixel grows approximately proportionally with exposure time:

$$X_t = \lambda_t K I + N, \quad (18)$$

where λ_t represents the exposure time, K and N follow the same definition as Equation (2) while I represents the rate of the photoelectrons which is proportional to the scene irradiation. This assumption is violated once motion blur is introduced, as object and camera motion redistribute photon counts across neighboring pixels. In that case, the raw measurement is better modeled as

$$X_t = K \int_0^{\lambda_t} I(t) dt + N, \quad (19)$$

where $I(t)$ denotes the time-varying irradiance along the exposure trajectory. Because $I(t)$ now changes over the time

instead of remaining constant, the architecture—primarily tailored for pixel-wise photometric mapping under static conditions—struggles to disentangle and jointly resolve deblurring and exposure correction without explicit structural guidance or specialized loss constraints.

4.2.3 Cascaded Restoration with Pre-trained Models

As an alternative strategy, we investigated whether the geometric and photometric degradations could be effectively decoupled via a two-stage cascaded pipeline. Given the scarcity of off-the-shelf models pre-trained for RAW-domain deblurring, we employed *Restormer* [13], a strong and widely-adopted baseline model for transformer-based image restoration in the sRGB domain. Our experimental pipeline followed the sequence: *RAW input* $\rightarrow$ *ExposureDiffusion* $\rightarrow$ *ISP* $\rightarrow$ *Restormer* $\rightarrow$ *output*

We observed that the initial stage, *ExposureDiffusion*, successfully recovered the exposure and suppressed sensor noise, yielding well-lit but blurry sRGB outputs. Consequently, the input to the secondary stage was effectively reduced to a pure deblurring task. However, despite this simplified objective, quantitative evaluation revealed no distinct improvement in restoration quality. We hypothesize that this inefficacy is driven by a significant domain distribution shift arising from hardware disparities. The off-the-shelf *Restormer* checkpoint is optimized on the GoPro dataset, which is captured using action cameras with small sensors and specific optical characteristics. In contrast, our SID-Blurred benchmark is derived from the Sony A7S II, a full-frame mirrorless camera with a vastly different sensor size and signal profile. These differences in sensor characteristics likely prevent the pre-trained model from generalizing effectively to our specific data distribution.

5 CONCLUSION

In this work, we proposed the novel task of joint motion deblurring and low-light image enhancement on RAW images. To facilitate this, we introduced a synthetic dataset with both training and testing sets to serve as a benchmark. We evaluated two approaches adapted from DarkIR and ExposureDiffusion (ED), respectively.

Our analysis reveals distinct challenges for each method. While the original DarkIR model suffered from noise amplification, we successfully mitigated this by fine-tuning on our dataset and incorporating a RAW metadata encoding module. Conversely, while ED demonstrated superior performance in low-light enhancement, it failed to account for image blur effectively, even after fine-tuning. This suggests that this joint task requires a specialized mathematical formulation capable of simultaneously modeling illumination and motion blur, rather than simply adapting single-task models.

We acknowledge certain limitations in our current benchmark. Due to resource constraints, our synthesis method was applied only to the Sony subset of the SID dataset; however, the proposed pipeline is theoretically agnostic and can be generalized to any RAW sensor data. Future work will aim to expand the benchmark to other datasets, such as ELD [10] and incorporate more complex degradation types, including out-of-focus blur and localized object motion.

REFERENCES

- [1] D. Feijoo, J. C. Benito, A. Garcia, and M. V. Conde, “Darkir: Robust low-light image restoration,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 10 879–10 889.
- [2] Y. Wang, Y. Yu, W. Yang, L. Guo, L.-P. Chau, A. C. Kot, and B. Wen, “ExposureDiffusion: Learning to Expose for Low-light Image Enhancement,” Aug. 2023, arXiv:2307.07710 [cs]. [Online]. Available: <http://arxiv.org/abs/2307.07710>
- [3] S. Zhou, C. Li, and C. Change Loy, “Lednet: Joint low-light enhancement and deblurring in the dark,” in *European conference on computer vision*. Springer, 2022, pp. 573–589.
- [4] X. Lv, S. Zhang, C. Wang, Y. Zheng, B. Zhong, C. Li, and L. 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*, 2024, pp. 25 378–25 388.
- [5] C.-H. Liang, Y.-A. Chen, Y.-C. Liu, and W. H. Hsu, “Raw Image Deblurring,” Dec. 2020, arXiv:2012.04264 [eess]. [Online]. Available: <http://arxiv.org/abs/2012.04264>
- [6] M. V. Conde, F. Vasluianu, and R. Timofte, “Toward Efficient Deep Blind RAW Image Restoration,” Sep. 2024, arXiv:2409.18204 [eess]. [Online]. Available: <http://arxiv.org/abs/2409.18204>
- [7] J. Ma, S. Yan, L. Zhang, G. Wang, and Q. Zhang, “ELMformer: Efficient Raw Image Restoration with a Locally Multiplicative Transformer,” in *Proceedings of the 30th ACM International Conference on Multimedia*, Oct. 2022, pp. 5842–5852, arXiv:2208.14704 [cs]. [Online]. Available: <http://arxiv.org/abs/2208.14704>
- [8] W. Jiao, B. Li, W. Shang, P. Wang, and D. Ren, “Efficient RAW Image Deblurring with Adaptive Frequency Modulation,” Sep. 2025, arXiv:2505.24407 [eess]. [Online]. Available: <http://arxiv.org/abs/2505.24407>
- [9] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to see in the dark,” *arXiv*, 2018.
- [10] K. Wei, Y. Fu, J. Yang, and H. Huang, “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*, 2020, pp. 2758–2767.
- [11] L. Borodenko, “motionblur: Generate authentic motion blur kernels and apply them to images,” <https://github.com/LeviBorodenko/motionblur>, 2020, gitHub repository.
- [12] L. Chen, X. Chu, X. Zhang, and J. Sun, “Simple baselines for image restoration,” in *European conference on computer vision*. Springer, 2022, pp. 17–33.
- [13] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient transformer for high-resolution image restoration,” in *CVPR*, 2022.



(a) Ground Truth



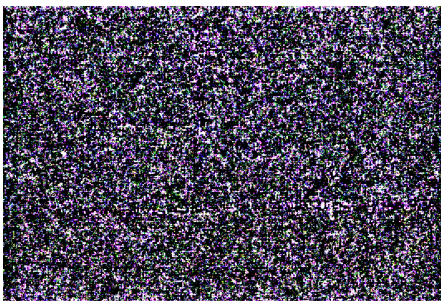
(b) Blurry Synthetic Input



(c) Pretrained ED on Sharp Input



(d) Pretrained ED on Blur Input



(e) Pretrained DarkIR on sRGB after ISP



(f) Pretrained DarkIR on denoised sRGB after ISP



(g) DarkIR (FT w/ EXIF) on sRGB after ISP

Fig. 4: Comparison of ExposureDiffusion and DarkIR on the same scene.