eRetinexGS: Retinex Modeling for Low-Light Scene Enhancement via Event Streams and 3D Gaussian Splatting

Supplementary Materials

7. Overview

This supplementary material includes: Sec. 8: Background on 3D Gaussian Splatting. Sec. 9: Additional details on our synthetic and real-world datasets. Sec. 10: Extended experiments and analysis of our method.

8. Preliminary

3D Gaussian Splatting. 3D Gaussian Splatting (3DGS) [8] uses a collection of gaussians to represent the 3D scene, where each Gaussian is specified by the 3D covariance matrix $\Sigma \in \mathbb{R}^{3\times3}$ and the center point $\mu \in \mathbb{R}^3$:

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), \tag{15}$$

where Σ is decomposed into the matrices S and R as:

$$\Sigma = \mathbf{RSS}^T\mathbf{R}^T. \tag{16}$$

$\mathbf{S}$ and $\mathbf{R}$ are obtained from the scaling factor $\mathbf{s} \in \mathbb{R}^3$ and the rotation quaternion $\mathbf{r} \in \mathbb{R}^4$, respectively. The 3D Gaussians are projected to 2D image plane [21] to render images. The covariance matrix Σ' in camera coordinates can be expressed by the Jacobian matrix $\mathbf{J}$ and the viewing transform matrix $\mathbf{W}$ as:

$$\Sigma' = \mathbf{JW}\Sigma\mathbf{W}^T\mathbf{J}^T. \tag{17}$$

α-blending is applied to compute the color of each pixel:

$$C = \sum_{i=1}^{N} c_i\alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \tag{18}$$

where c_i is the color of each point and α_i is the opacity after multiplying with the transformed 2D Gaussian.

9. More Details

9.1. More Dataset Details

Fig. 8 illustrates our experimental setup, including the input-view and novel-view settings.

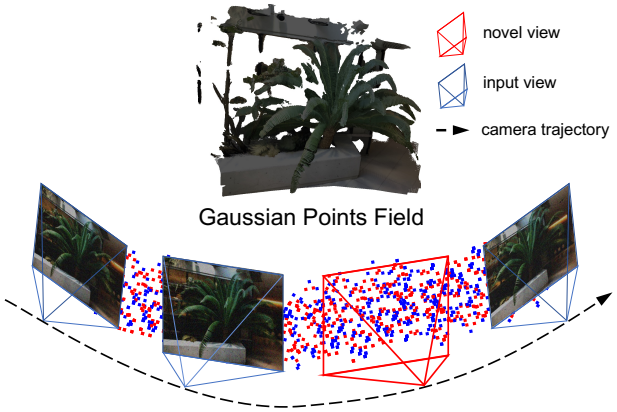


Figure 8. Experimental setup. Inputs are low-light frames from the input views and synchronized events .We report normal-light renderings for input-view synthesis (rendering the enhanced scene at the input views) and for novel-view synthesis (rendering at held-out novel views).

Table 3. Comparison of training efficiency and rendering speed on Synthetic data. In the Input column, "F" denotes frame and "E" denotes event.

Methods	Input	Train(GPU · h) ↓	FPS↑
Aleth-NeRF [2]	F	5.22	0.26
LLNeRF [19]	F	6.93	0.19
LuSh-NeRF [15]	F	4.66	0.39
ReDDiT [9]	F	-	6
Luminance-GS [3]	F	0.56	75
IncEventGS [6]	E	0.8	81
Ours	F+E	**0.75**	**83**

Synthetic Data. The spatial resolution of the LLFF dataset used in our synthetic experiments is 352×480. We synthesize events using the v2e simulator [5] with color event generation [18]. To ensure high temporal resolution, we upsample the original image sequences by 32× using Super SloMo [7].

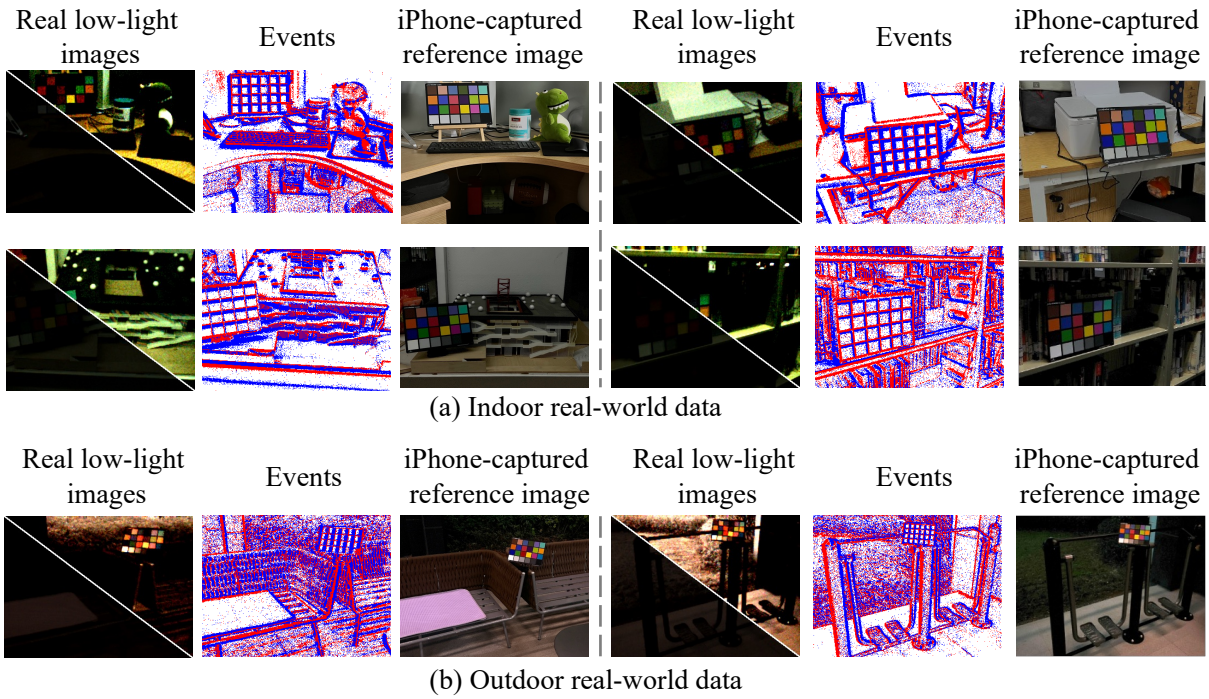


Figure 9. Real-world data capture details. (a) Examples of indoor scenes. (b) Examples of outdoor scenes. The reference image is captured using an iPhone 13 Pro Max under low-light conditions with long exposure, and enhanced by the smartphone's advanced computational photography algorithms.

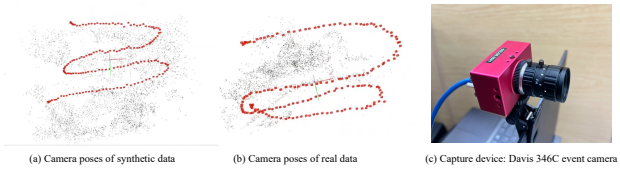


Figure 10. Visualization of camera trajectories. (a) Estimated camera poses on synthetic data using COLMAP, visualizing motion patterns; (b) Estimated camera poses on real-world data using COLMAP; (c) Capture device used for real-world data: DAVIS346C camera.

Real-World Data Collection. Fig. 9 presents examples of our captured real-world low-light scenes. We record six low-light scenes using the DAVIS346C [18] camera(Fig. 10(c)) with a resolution of 260×346, comprising four indoor scenes (Fig. 9(a)) and two outdoor scenes (Fig. 9(b)). Exposure times are set to approximately 5 ms for indoor scenes and 10 ms for outdoor scenes. As shown in Fig. 11, the captured low-light images have pixel intensities mostly below 50, consistent with the settings used in LLNeRF [19]. Camera trajectories follow the LLFF [13] dataset design, and Fig. 10(c) visualizes the continuous camera poses estimated from the captured low-light images using COLMAP.

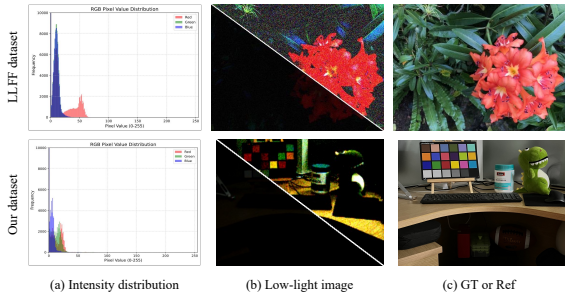
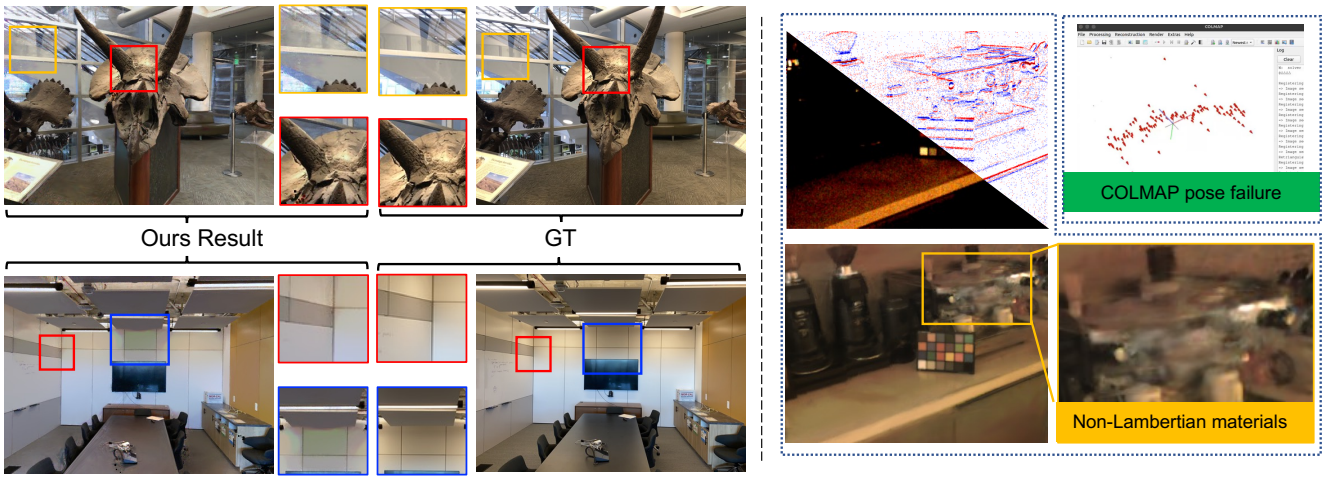


Figure 11. Intensity distribution and representative samples from our dataset.

10. Additional Experimental and Analysis

10.1. Comparison with Baselines and Efficiency Analysis

We evaluate the training and inference efficiency of our method and the baselines on the synthetic dataset using an NVIDIA RTX 3090 GPU. Tab. 3 reports the comparison in terms of training time (GPU·h) and rendering speed (FPS). Our method achieves significantly lower training time and higher inference FPS than NeRF-based methods, primarily due to the real-time rendering capabilities of 3DGS.



(a) Limitation of eRetinexGS (b) Failure cases of eRetinexGS

Figure 12. Limitations and failure cases of eRetinexGS. (a) Typical limitations. **Red box:** the gray-world prior may introduce a mild color bias compared with the ground truth in some scenes; stronger color priors (e.g., diffusion-based guidance) may help alleviate this issue. **Blue box:** in regions with slowly varying illumination (e.g., diffuse shadows), event observations may be sparse, which can cause soft-boundary artifacts. **Yellow box:** specular or non-Lambertian surfaces violate the Retinex assumption and may therefore lead to reconstruction artifacts. (b) Failure cases. Inaccurate COLMAP poses may prevent reliable reconstruction, and large specular or non-Lambertian regions may further cause severe artifacts.

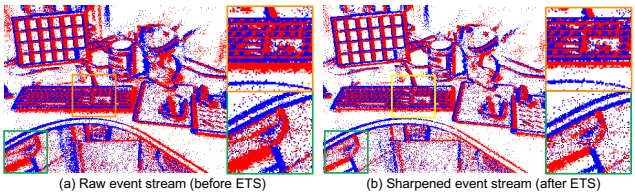


Figure 13. Visualization of Event Trailing Suppression. (a) Raw event stream before ETS preprocessing. (b) Enhanced event representation after ETS, showing sharper edges and reduced trailing artifacts.

Importantly, incorporating our decomposition module and lightweight MLP on top of vanilla 3DGS does not compromise its real-time performance.

Among event-only methods, we identify Event-3DGS [4] and IncEventGS [6] as the most relevant baselines. However, we found that the public implementation of Event-3DGS lacks essential components, including intensity image generation, intensity differencing, and the high-pass filtering step. Therefore, we adopt IncEventGS as the representative event-only method in our comparative experiments.

10.2. Limitations, Practicality, and Robustness

Although eRetinexGS performs robustly in most multi-view low-light scenes, several limitations remain. First, the gray-world prior may introduce a mild color bias in some cases (Fig. 12(a), red box). Second, the event-guided reflectance smoothness prior becomes less reliable when event observations are sparse, especially in regions with slowly varying illumination such as diffuse shadows, which may lead to soft-boundary artifacts (Fig. 12(a), blue box). This issue is more pronounced in a single view, while multi-view supervision generally helps alleviate the ambiguity across views. Third, specular or non-Lambertian surfaces violate the Retinex assumption and may therefore cause reconstruction artifacts (Fig. 12(a), yellow box).

We also observe failure cases caused by inaccurate camera poses. In particular, when COLMAP pose estimation fails, reliable scene reconstruction may not be obtained (Fig. 12(b), top). In addition, large specular or non-Lambertian regions can further amplify artifacts (Fig. 12(b), bottom). In practice, our method remains reasonably robust under multi-view supervision, but these cases highlight the current limitations of the formulation. When pose estimation becomes unreliable, event-based pose estimation methods such as IncEventGS [6] may provide a practical fallback.

10.3. Visualization of Event Trailing Suppression

In low-light scenes, event signals are degraded by leak noise, shot noise, and delayed responses arising from the limited bandwidth of photoreceptors. Among these factors, the trailing artifact [12, 20], which is a direct consequence of limited photoreceptor bandwidth, is a significant form of degradation. Inspired by NeR-Net [12], we adopt the Event Trailing Suppression (ETS) module to preprocess event data

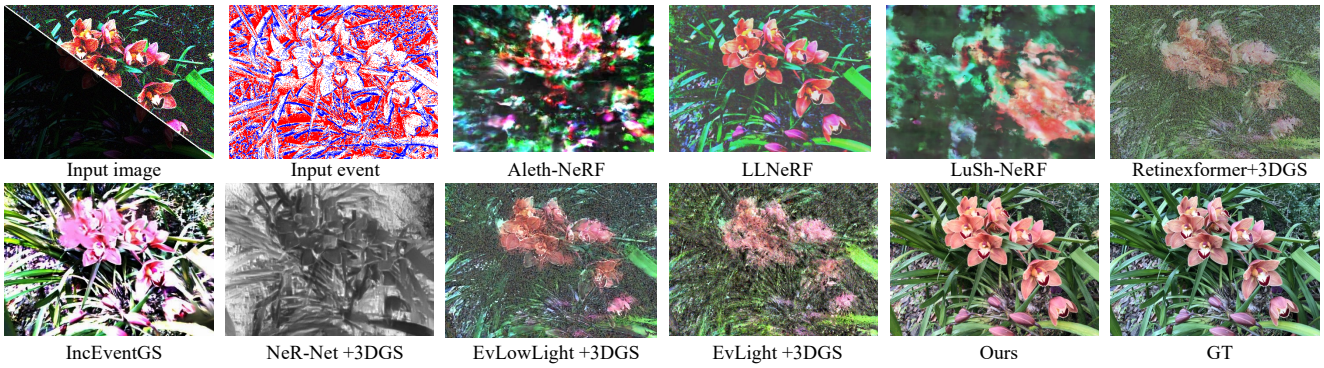


Figure 14. More qualitative results about novel view synthesis in different views.

for both synthetic and real-world inputs.

Fig. 13 visualizes the effect of ETS before and after processing. As shown, the event edges become noticeably sharper after ETS, demonstrating its effectiveness in alleviating trailing degradation under low-light conditions.

10.4. Novel View Synthesis Results

Fig. 14 presents additional qualitative results for novel view synthesis, evaluated by holding out 1/8 of the input views following the standard protocol in [14]. Existing methods generally struggle to preserve structural features and details in extremely dark regions due to poor reconstruction quality. Specifically, ***NeRF-based Method***, Aleth-NeRF [2] and LuSh-NeRF [15] fail in these scenes, primarily due to their lack of robustness to texture loss under low-light conditions [17]. LLNeRF [19] also underperforms, as it lacks sufficient information in severely underexposed regions. ***Event-only methods*** such as IncEventGS [6] and NeR-Net+3DGS [12] suffer from brightness and color inaccuracies, leading to unsatisfactory novel view synthesis results. In addition, ***hybrid pipelines*** that combine enhancement methods with 3DGS, such as RetinexFormer+3DGS [1], EvLowLight+3DGS [11], and EvLight+3DGS [10], tend to suffer from data bias, and their outputs often degrade significantly under novel low-light viewpoints.

In contrast, our method introduces event signals into the scene reconstruction pipeline, providing complementary texture information in low-light regions. Moreover, the incorporation of multi-view consistency and event guided Retinex decomposition enables better generalization to novel viewpoints and avoids the bias inherent in purely learning-based methods.

10.5. More analysis of Event and Image Complementary

Low-light images and events degrade differently but offer complementary cues. Events capture temporal radiance changes in the logarithmic domain [16], making them more

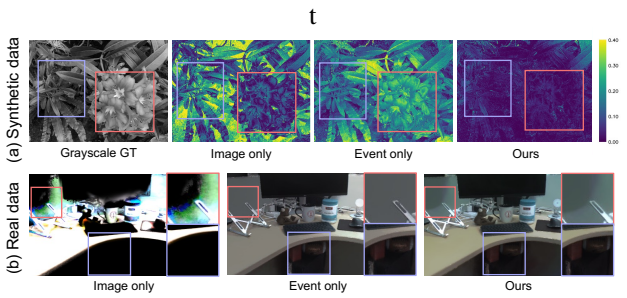


Figure 15. Complementary photometric information from image and event modalities. (a) On synthetic data, **photometric error** visualizations show that event-only inputs yield lower errors in dark regions (purple box), while image-only inputs perform better in bright regions (red box). (b) A similar trend is observed on real-world data: event-only constraints (middle) better recover dark-region details, while image-only constraints (left) preserve structure in moderately lit areas (e.g., the monitor shadow behind frosted glass in red box). By leveraging both modalities, our method enables superior reconstruction in both synthetic and real scenes.

sensitive in dark regions, while images quantize radiance in the linear domain(before gamma correction) and preserve more reliable information in brighter areas where events may saturate. As illustrated in Fig. 15, on synthetic data, Fig. 15(a) shows that event-only inputs yield lower photometric error in dark regions, whereas image-only inputs perform better in bright areas. A similar trend is observed on real-world data in Fig. 15(b), which uses the input data of the "cues2" sequence from the bottom row of Fig. 1 in the main paper: event-only constraints better recover dark-region details, while image-only constraints preserve structure in moderately lit regions. Exploiting this complementarity within a multi-view reconstruction setting helps suppress random noise and yields superior reconstructions in both synthetic and real scenes.

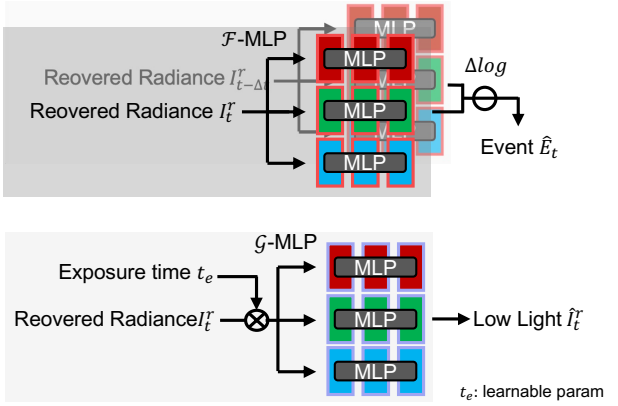


Figure 16. Architectures of the learned degradation mappings. $\mathcal{F}$-MLP applies a lightweight per-channel mapping to recovered radiance to align it with the event sensor response and match event brightness changes. $\mathcal{G}$-MLP applies a lightweight per-channel tone mapping (with a learnable exposure) to map recovered radiance to low-light frames, compensating exposure/ISP-induced color shifts. The exposure is regularized by the brightness loss, and the per-channel mappings are constrained by the color prior

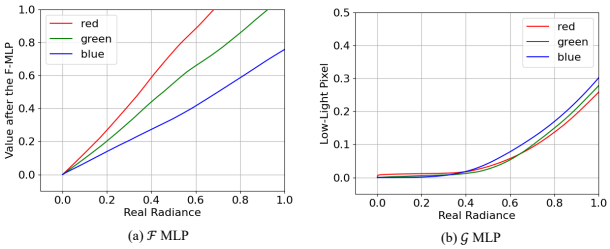


Figure 17. An example visualization of $\mathcal{F}$ and $\mathcal{G}$ degradation mapping curves on real-world data. (a) $\mathcal{F}$ mapping bridging the gap between the rendered real radiance and the brightness changes perceived by the event camera; (b) $\mathcal{G}$ mapping model the degradation process from real radiance to low-light images and handle color shifts by learning separate RGB channels.

10.6. More analysis of Degradation Curves

Fig. 16 shows the network architectures of the $\mathcal{F}$-MLP and $\mathcal{G}$-MLP. Fig. 17 shows the learned degradation curves of $\mathcal{F}$ and $\mathcal{G}$ in our framework. For low-light images, signal-dependent Gaussian–Poisson noise is further distorted by the non-linear camera ISP pipeline. Under multi-view constraints, this leads to convergence toward a biased color distribution [19]. To address this, the degradation function $\mathcal{G}$ learns a channel-wise RGB mapping to model the transformation from real radiance I^r to degraded low-light images I^l.

Similarly, in low-light conditions, event cameras are primarily affected by shot noise, which occurs at approximately 1 Hz per pixel [5], whereas leak noise is much lower, typically around 0.1 Hz. Given this disparity, we neglect the influence of leak noise in our modeling. Additionally, the nonlinear log-intensity encoding in event cameras can further amplify noise, introducing reconstruction bias under multi-view constraints. To address these degradations, the function $\mathcal{F}$ is trained to model the transformation from scene radiance I^r to the observed event signal E under low-light conditions.

Table 4. Ablation study under different settings. Quantitative results demonstrate the effectiveness of each component.

Methods	PSNR ↑	SSIM ↑	LPIPS ↓
Ours w/o E, $\mathcal{L}_{tv}$	17.21	0.5030	0.5027
Ours w/o I^l	17.88	0.5404	0.5403
Ours w/o $\mathcal{L}_{tv}$	21.83	0.7282	0.2047
Ours w/o $\mathcal{L}_{brightness}$	19.52	0.7684	0.1843
Ours w/o $\mathcal{L}_{gray}$	22.73	0.8185	0.1916
Ours w/o $\mathcal{F}, \mathcal{G}$	20.10	0.7801	0.1201
Ours	**23.45**	**0.8312**	**0.0888**

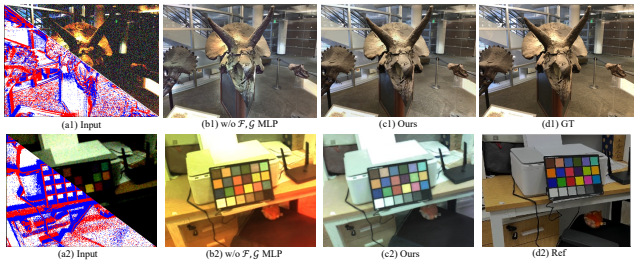


Figure 18. Visualization of the ablation study on degradation mappings. Removing the mappings $\mathcal{F}$ and $\mathcal{G}$ leads to noticeable color shifts, biased radiance distributions, and additional artifacts. Results are shown for synthetic data (top) and real-world data (bottom), demonstrating the necessity of modality alignment through $\mathcal{F}$ and $\mathcal{G}$ for stable reconstruction quality.

10.7. Ablation Study of Degradation Curves

We further conduct an ablation study by removing the degradation mappings $\mathcal{F}$ and $\mathcal{G}$. Quantitative results on synthetic data are reported in Tab. 4 (row "Ours w/o $\mathcal{F}, \mathcal{G}$").

Fig. 18 presents visual comparisons on synthetic data (top) and real data (bottom). We observe that removing the degradation mappings for the two modalities leads to severe color shifts. The recovered radiance exhibits biased brightness distributions and noticeable artifacts. These results demonstrate that using $\mathcal{F}$ and $\mathcal{G}$ to align the two degraded modalities with the recovered radiance is crucial for stable reconstruction quality.

10.8. Temporal Coherence of Our Results.

Fig. 19 compares the temporal consistency of our method with other event-based low-light video enhancement approach. In Fig. 19(a), which shows results on synthetic data, we observe that EvLowLight[11] fails to effectively

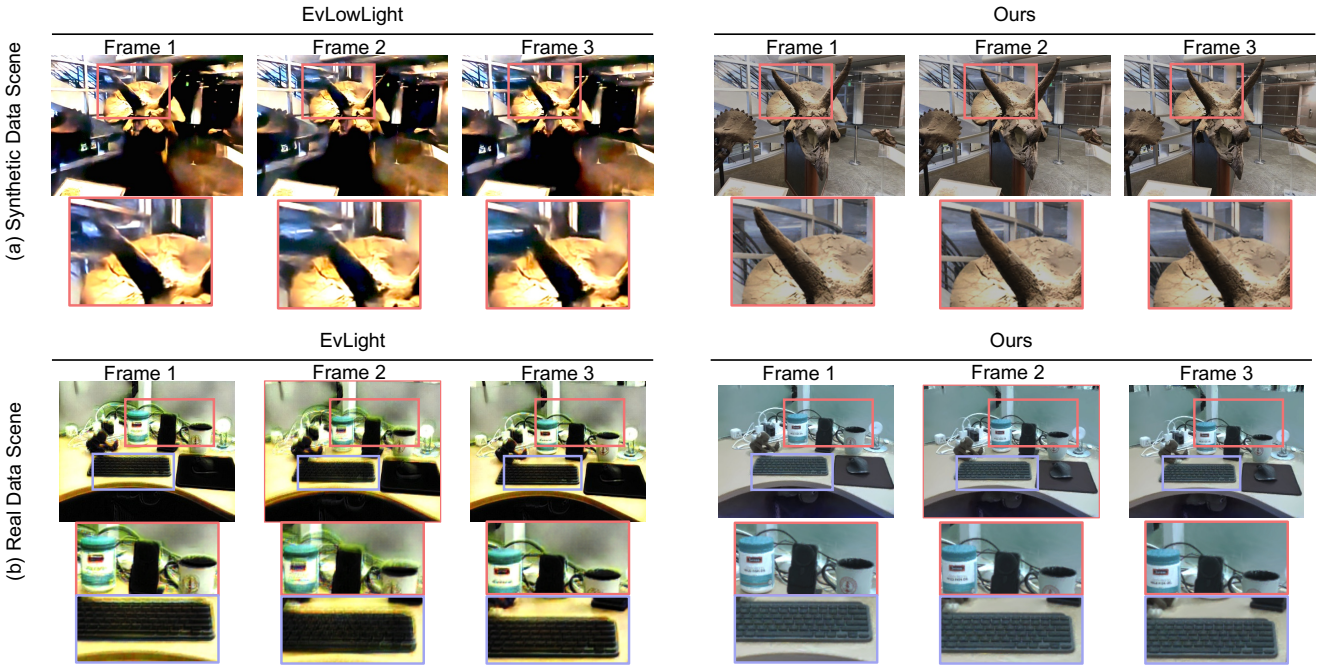


Figure 19. Qualitative comparisons of temporal consistency on synthetic (a) and real-world (b) data: (a) vs. EvLowLight [11], (b) vs. EvLight [10].

leverage the complementary information from images and events in extremely dark regions, such as the horn, resulting in noticeable flickering across frames. In contrast, our method benefits from multi-view geometric constraints and reconstructs a 3D radiance field with improved temporal coherence.

Fig. 19(b) presents the results on real-world data. For example, around the keyboard area, EvLight [10] exhibits inconsistent brightness across consecutive frames. Additionally, in regions like the cup, where significant inter-frame motion occurs due to camera shake, EvLight suffers from motion-induced artifacts. Our method, on the other hand, demonstrates strong temporal coherence even on real-world sequences, and remains robust to large inter-frame motion.

10.9. Additional Quantitative Results

Tab. 5 presents per-scene quantitative results for low-light scene enhancement.

10.10. Additional Qualitative Results

Fig. 20 and Fig. 21 show additional qualitative results on our real-world dataset. LLNeRF and LuSh-NeRF often fail under extreme low-light and are thus omitted from the main paper (see Fig. 5).

Image-only methods struggle to restore structural details due to weak signals, while event-only methods (e.g., IncEventGS) suffer from artifacts and color degradation caused by the lack of absolute radiometric cues. Learning-based fusion methods are sensitive to data bias and often misalign cross-modal information, leading to distortions.

In contrast, our method reliably recovers radiance, texture, and color in both indoor and outdoor low-light scenes. Reference images were captured by an iPhone 13 Pro Max using long exposure.

Fig. 22 further validates our effectiveness on synthetic data.

Table 5. Per-scene quantitative results for low-light enhancement(Input View Synthesis).

Scene	PSNR ↑	SSIM ↑	LPIPS ↓
fern	24.70	0.8017	0.0873
flower	24.39	0.8635	0.0637
fortress	27.20	0.7760	0.1184
horns	24.40	0.8812	0.0636
leaves	18.71	0.7187	0.1635
orchids	22.32	0.8347	0.0776
room	21.97	0.8637	0.0892
trex	23.89	0.9105	0.0470
Avg.	**23.45**	**0.8312**	**0.0888**

10.11. SDE Dataset Results

Although the SDE dataset [10] provides real-world low-light video data, its unreliable camera pose estimation

Table 6. Quantitative comparison between our method and EvLight on the SDE dataset.

Scenes	EvLight			Ours		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
o_69	**21.55**	**0.7057**	**0.1005**	21.07	0.6852	0.1217
o_126	**21.25**	**0.7446**	**0.1249**	20.80	0.7370	0.1496

makes it less suitable for scene reconstruction. Therefore, we conduct experiments only on scenes from the SDE test set where camera poses can be reliably estimated from low-light images.

Fig. 23 presents qualitative comparisons between our method and EvLight on two such scenes. Each example includes the predicted results from both methods, their corresponding error maps relative to GT, and visual comparisons. Unlike EvLight, which relies on supervised learning and exhibits limited generalization(performing poorly on both our synthetic and real-world datasets) our method leverages the same unsupervised optimization framework across all scenes, requiring no ground-truth supervision and optimizing per scene.

As shown in the top row of Fig. 23, our method produces more accurate results than EvLight in texture-rich regions such as tables and chairs. Similarly, in the bottom scene, our method achieves lower error around the handrail structure. However, in texture-less regions (e.g., the mirror in the top scene or the ceiling in the bottom scene), our method exhibits slightly higher errors due to the lack of structural cues.

Quantitative results are provided in Tab. 6. Although our unsupervised method yields slightly lower PSNR compared to the learning-based EvLight, it demonstrates superior generalization without requiring any ground truth. Moreover, the visual results highlight our advantage in preserving structural detail in textured regions.

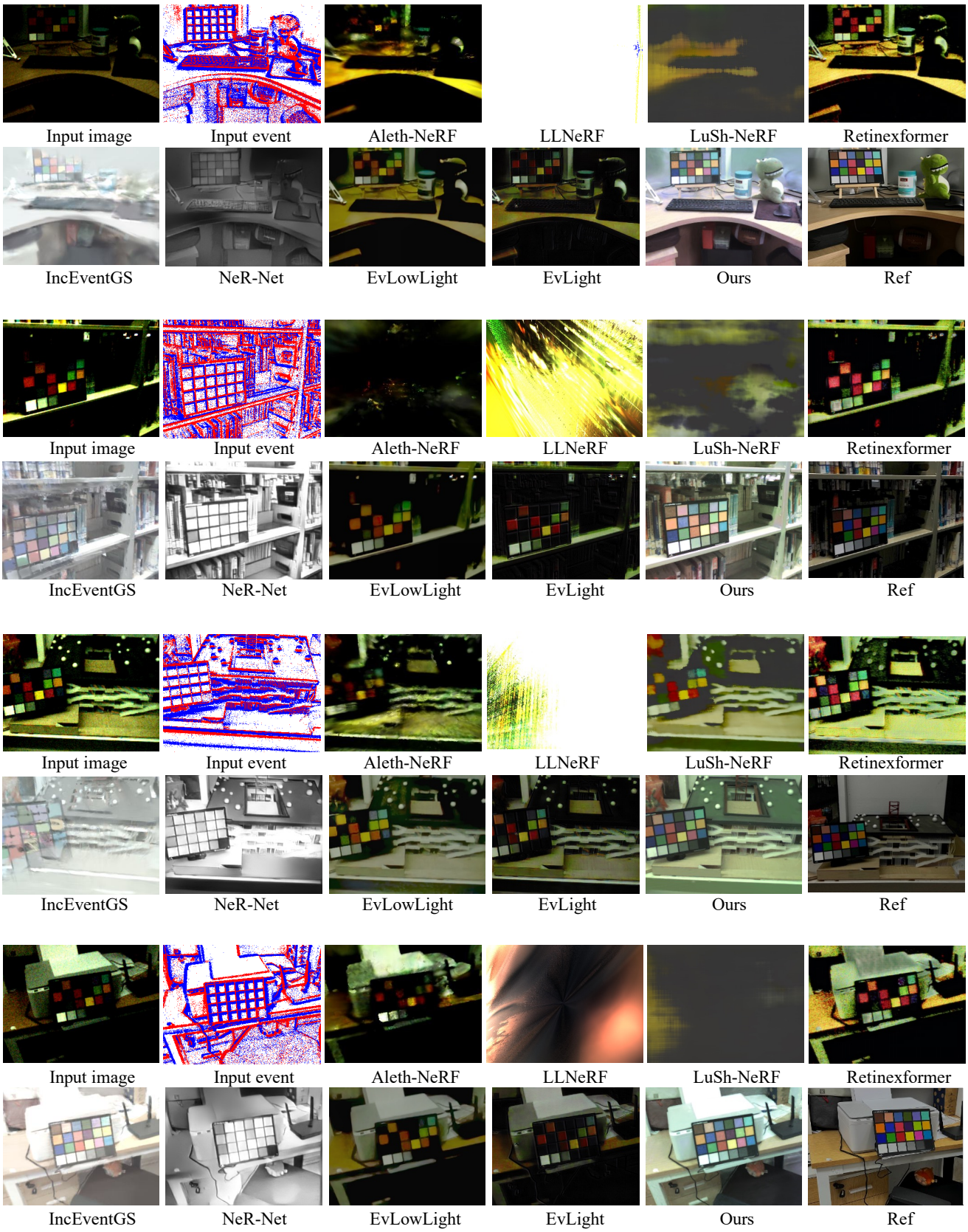


Figure 20. More qualitative results for low-light scene enhancement on Real Data-1. The reference image, captured using an iPhone 13 Pro Max with a more advanced sensor and powerful computational photography pipeline under long exposure, is not strictly aligned with the low-light inputs.

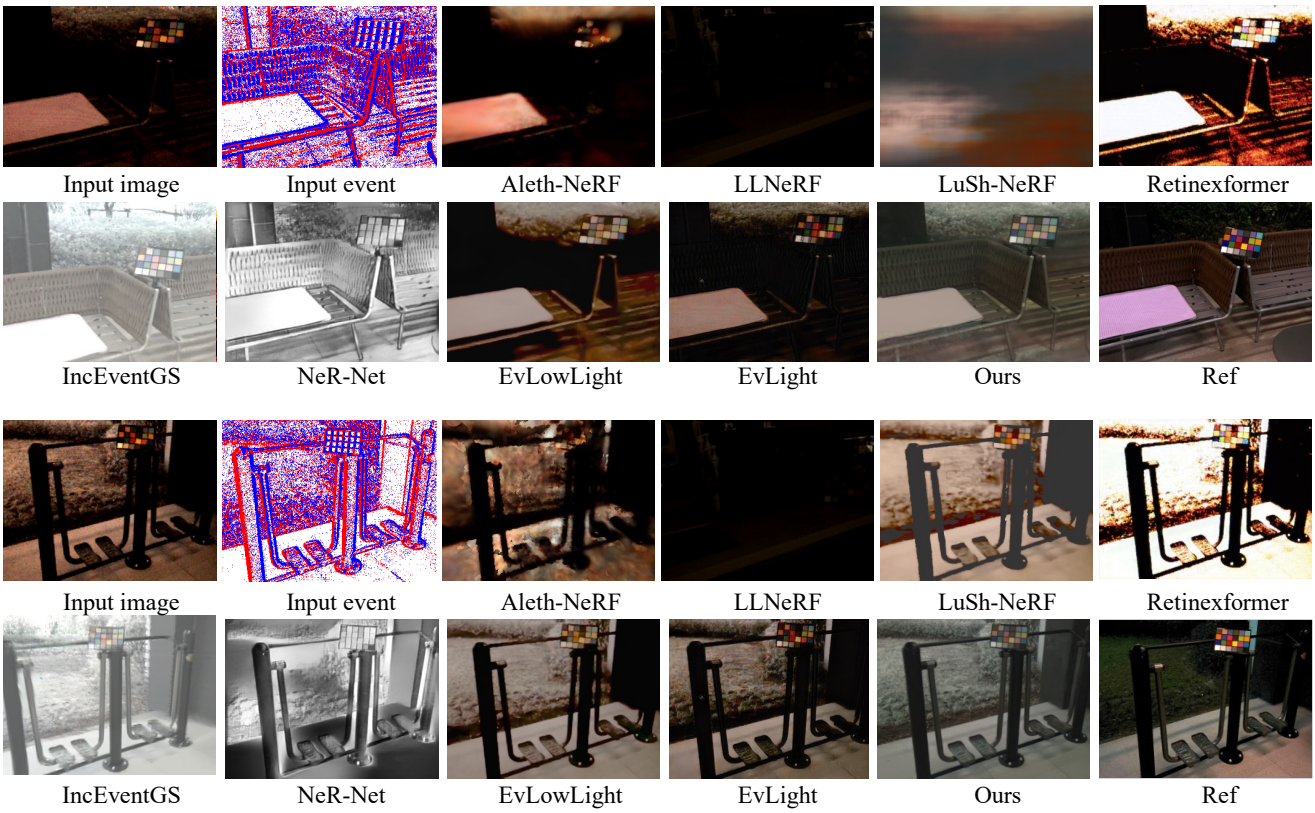


Figure 21. More qualitative results for low-light scene enhancement on Real Data-2. The reference image, captured using an iPhone 13 Pro Max with a more advanced sensor and powerful computational photography pipeline under long exposure, is not strictly aligned with the low-light inputs.

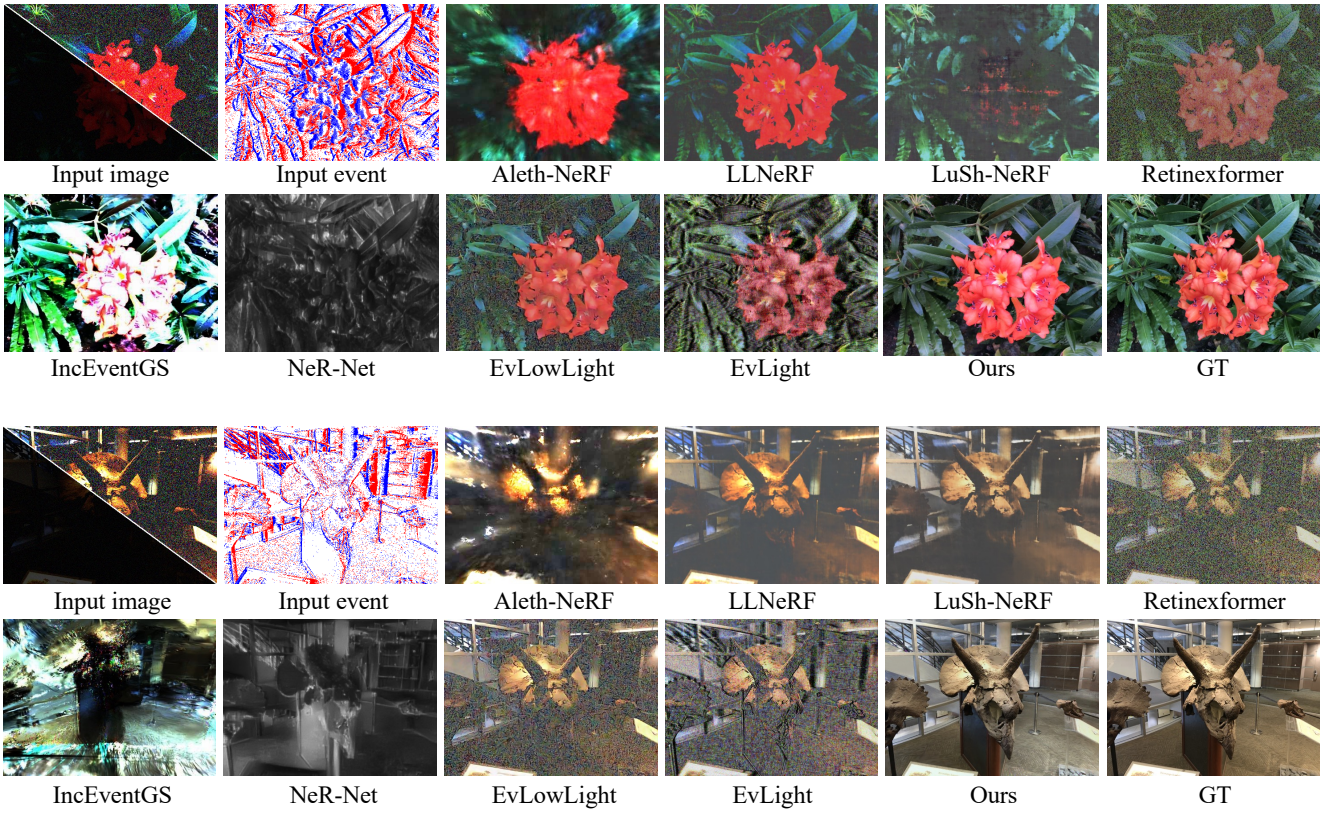


Figure 22. More qualitative results for low-light scene enhancement on synthetic data.

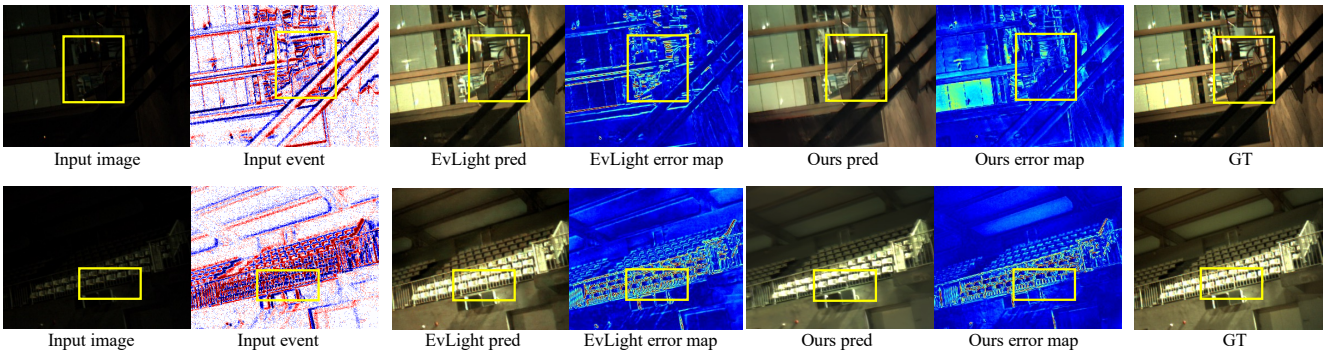


Figure 23. Visualization results compared on the SDE dataset [8]. Since the SDE dataset is designed for low-light video enhancement tasks, only a small subset of scenes in the test set allow for camera pose reconstruction. Therefore, we conducted experiments on these scenes. Additionally, the ground truth (GT) in the SDE dataset was captured using a DAVIS346 camera, which inherently introduces color bias. As shown in the scenes below, both the GT and the EvLight predictions exhibit yellowish hues in the bleachers. In contrast, our method recovers more visually accurate white seats.

References

[1] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, pages 12504–12513, 2023. 4

[2] Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-NeRF: illumination adaptive NeRF with concealing field assumption. In *AAAI*, pages 1435–1444, 2024. 1, 4

[3] Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment. In *CVPR*, pages 26472–26482, 2025. 1

[4] Haiqian Han, Jianing Li, Henglu Wei, and Xiangyang Ji. Event-3dgs: Event-based 3d reconstruction using 3d gaussian splatting. *NeurIPS*, 37:128139–128159, 2024. 3

[5] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: from video frames to realistic DVS events. In *CVPR*, pages 1312–1321, 2021. 1, 5

[6] Jian Huang, Chengrui Dong, Xuanhua Chen, and Peidong Liu. IncEventGS: Pose-free Gaussian Splatting from a single event camera. In *CVPR*, pages 26933–26942, 2025. 1, 3, 4

[7] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *CVPR*, pages 9000–9008, 2018. 1

[8] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1

[9] Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, and Bin Zhao. Efficient diffusion as low light enhancer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21277–21286, 2025. 1

[10] Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, and Lin Wang. Towards robust event-guided low-light image enhancement: a large-scale real-world event-image dataset and novel approach. In *CVPR*, pages 23–33, 2024. 4, 6

[11] Jinxiu Liang, Yixin Yang, Boyu Li, Peiqi Duan, Yong Xu, and Boxin Shi. Coherent event guided low-light video enhancement. In *ICCV*, pages 10615–10625, 2023. 4, 5, 6

[12] Haoyue Liu, Shihan Peng, Lin Zhu, Yi Chang, Hanyu Zhou, and Luxin Yan. Seeing motion at nighttime with an event camera. In *CVPR*, pages 25648–25658, 2024. 3, 4

[13] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. 2

[14] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, pages 405–421, 2020. 4

[15] Zefan Qu, Ke Xu, Gerhard Petrus Hancke, and Rynson WH Lau. Lush-NeRF: Lighting up and sharpening NeRFs for low-light scenes. *NeurIPS*, 2024. 1, 4

[16] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *CVPR*, pages 3857–3866, 2019. 4

[17] Fabio Remondino, Ali Karami, Ziyang Yan, Gabriele Mazzacca, Simone Rigon, and Rongjun Qin. A critical analysis of nerf-based 3d reconstruction. *Remote Sensing*, 15(14): 3585, 2023. 4

[18] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. Ced: Color event camera dataset. In *CVPRW*, pages 0–0, 2019. 1, 2

[19] Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *ICCV*, pages 12632–12641, 2023. 1, 2, 4, 5

[20] Ziran Zhang, Yongrui Ma, Yueting Chen, Feng Zhang, Jinwei Gu, Tianfan Xue, and Shi Guo. From sim-to-real: Toward general event-based low-light frame interpolation with per-scene optimization. In *Proc. of ACM SIGGRAPH Asia*, pages 1–10, 2024. 3

[21] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In *Proceedings Visualization, 2001. VIS'01.*, pages 29–538. IEEE, 2001. 1