# Dark-EvGS: Event Camera as an Eye for Radiance Field in the Dark

Jingqian Wu, Peiqi Duan, *Member, IEEE*, Zongqiang Wang, Changwei Wang, Boxin Shi, *Senior Member, IEEE*, and Edmund Y. Lam, *Fellow, IEEE*

***Abstract*—In low-light environments, conventional cameras often struggle to capture clear multi-view images of objects due to dynamic range limitations and motion blur caused by long exposure. Event cameras, with their high-dynamic range and high-speed properties, have the potential to mitigate these issues. Additionally, 3D Gaussian Splatting (GS) enables radiance field reconstruction, facilitating bright frame synthesis from multiple viewpoints in low-light conditions. However, naively using an event-assisted 3D GS approach still faced challenges because, in low lights, events are noisy, frames lack quality, and the color tone may be inconsistent. To address these issues, we propose Dark-EvGS, the first event-assisted 3D GS framework that enables the reconstruction of bright frames from arbitrary viewpoints along the camera trajectory. Triplet-level supervision is proposed to gain holistic knowledge, granular details, and sharp scene rendering. The color tone matching block is proposed to guarantee the color consistency of the rendered frames. Furthermore, we introduce the first real-captured dataset for the event-guided bright frame synthesis task via 3D GS-based radiance field reconstruction. Experiments demonstrate that our method achieves better results than existing methods, conquering radiance field reconstruction under challenging low-light conditions. The code and sample data are included in the supplementary material.**

***Index Terms*—Event camera, event-based vision, 3D Gaussian splatting, radiance field reconstruction.**

Received 24 September 2025; revised 4 February 2026; accepted 9 March 2026. Date of publication 20 March 2026; date of current version 25 March 2026. This work was supported in part by the Research Grants Council of Hong Kong under Grant GRF 17201822; in part by the Theme-Based Research Scheme under Grant T45-701/22-R; in part by the National Natural Science Foundation of China under Grant 62402014 and Grant 62136001; in part by Beijing Natural Science Foundation under Grant L233024; in part by Beijing Municipal Science and Technology Commission, Administrative Commission of Zhongguancun Science Park, under Grant Z241100003524012; in part by Taishan Scholars Program under Grant TSQN202507241; and in part by Shandong Provincial Natural Science Foundation for Young Scholars Program under Grant ZR2025QC1627. The work of Peiqi Duan was supported in part by China National Postdoctoral Program for Innovative Talents under Grant BX20230010 and in part by China Postdoctoral Science Foundation under Grant 2023M740076. The associate editor coordinating the review of this article and approving it for publication was Dr. Junhui Hou. *(Corresponding authors: Peiqi Duan; Edmund Y. Lam.)*

Jingqian Wu and Edmund Y. Lam are with the Department of Electrical and Computer Engineering, The University of Hong Kong, Hong Kong, SAR, China (e-mail: elam@eee.hku.hk).

Peiqi Duan and Boxin Shi are with the State Key Laboratory of Multimedia Information Processing and the National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100080, China (e-mail: duanqi0001@pku.edu.cn).

Zongqiang Wang is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

Changwei Wang is with the Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology, Jinan, Shandong 250013, China.

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2026.3674360, provided by the authors.

Digital Object Identifier 10.1109/TIP.2026.3674360

## I. Introduction

LOW-LIGHT radiance field reconstruction plays a crucial role in various real-world applications, including nighttime photography [1], [2], [3], surveillance [4], and autonomous driving [5], [6], [7]. However, traditional frame-based cameras struggle to capture high-quality images in dark environments due to their limited dynamic range and reliance on long exposure times. These constraints result in significant noise and motion blur, making it difficult to capture clear, bright frames of objects from multiple viewpoints in low-light conditions.

Event cameras are a novel type of sensor that captures brightness changes asynchronously per pixel, marking a shift from conventional frame-based imaging [8], [9], [10], [11], [12]. With high dynamic range (HDR) and high temporal resolution capabilities, they can simultaneously capture bright and dark regions with no motion blur, enabling the recovery of visual details often lost in frame-based imaging, and even supporting radiance field reconstruction, particularly under low-light conditions. Event-based video reconstruction methods are proposed [13], [14], [15] to output consecutive frames from events. Nevertheless, either they are limited to grayscale images [13], [14], or primarily focus on the short exposure problem [15]. EvLight [16] leverages event data for single-image enhancement rather than focusing on video, and may exhibit partial chromatic aberrations. Event data have also been applied to radiance field reconstruction, but existing approaches [17], [18], [19] are only designed for normal light conditions (around 300 lux).

Reconstructing a radiance field in the dark is particularly challenging due to several factors. First, the captured signals lack quality: dark frames contain extremely low intensity and brightness, while event streams become significantly noisier and sparser because of photon scarcity (Fig. 1 (a) and (c)). This makes both frame-only methods (e.g., 3D GS [20]) and event-based radiance field approaches [17], [18], [19] ineffective under such conditions. Second, ensuring color tone consistency across rendered views is non-trivial. Existing event-based video reconstruction works [13], [14], [15] either output grayscale sequences or focus on short-exposure enhancement, and thus cannot be directly adapted for radiance field reconstruction in dark environments. Third, accurate camera pose estimation and data availability present additional

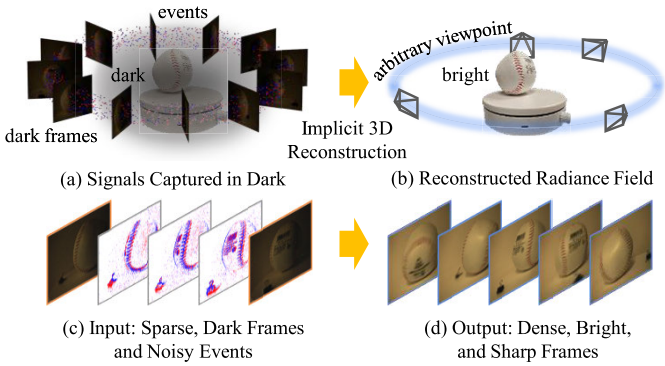


Fig. 1. In the dark, only noisy events and dark blurred frames can be captured (a), which is challenging for existing approaches for radiance field reconstruction. Our approach takes the raw events and sparse dark frames as input (c), and it reconstructs a radiance field, enabling arbitrary viewpoint synthesis (b). Our approach is capable of synthesizing dense, bright, and sharp views, enhancing visibility and detail (d).

obstacles. On the one hand, structure-from-motion pipelines struggle in dark environments because frames lack sufficient intensity for reliable feature extraction, and no mature event-based COLMAP alternative exists that can robustly handle noisy, sparse events. On the other hand, the absence of dedicated datasets further limits progress: most prior studies rely on synthetic short/long exposure pairs and simulated events [21], which fail to capture the noise characteristics, dynamic lighting, and inconsistencies of real-world low-light data. As a result, even when accurate ground-truth poses are provided in controlled experiments, such settings do not reflect real-world deployments, and models trained on synthetic data generalize poorly to real-world low light datasets and environments.

In this paper, we propose Dark-EvGS by integrating the event camera and 3D Gaussian Splatting (GS) [20], a radiance field reconstruction technique, enabling bright frame reconstruction from arbitrary viewpoints along the camera trajectory. The high-dynamic sensing capability of event cameras enables Dark-EvGS to have a precise perception of camera motion, which can provide valuable guidance for radiance field reconstruction performance (Fig. 1 (b)) and render bright frames from multiple viewpoints of objects in low-light conditions (Fig. 1 (d)).

Dark-EvGS conquers the mentioned challenges in three directions: 1) We propose a triplet-level supervision mechanism specifically tailored for handling noisy event and frame signals in low-light radiance field reconstruction; 2) the Color Tone Matching Block (CTMB) is proposed to guarantee color consistency of the rendered frames; 3) we collect a dataset for the event-guided bright frame synthesis task via 3D GS-based radiance field reconstruction, which includes paired low-light frames, bright ground truth frames, event streams, and corresponding camera poses. Extensive experiments demonstrate that our method outperforms previous state-of-the-art techniques across all samples. The outlined key technical contributions are:

- We introduce Dark-EvGS, the first event-guided bright frame synthesis method via 3D GS-based radiance field reconstruction from arbitrary viewpoints in the dark.
- We present a triplet-level supervision strategy to recover missed holistic structures, restore fine-grained details, and enhance sharpness for 3D GS training and for rendering color-consistent bright frames in low-light environments.
- We collect the first real-world event-based low-light radiance field reconstruction dataset with paired low-light, bright-light frames, event streams, and corresponding camera poses. We will release the full dataset and code to facilitate future research.

## II. Related Work

### A. General and Dark Radiance Field

Radiance field rendering and novel-view synthesis are fundamental tasks in graphics and computer vision with extensive applications in fields like robotics and virtual reality. Neural Radiance Fields (NeRF) [22] and 3D GS [20] have made substantial progress in these tasks. While NeRF produces high-fidelity renderings with intricate detail, the high number of samples required to accumulate the color information for each pixel results in low rendering efficiency and extended training times [20]. 3D GS, on the other hand, employs a set of optimized Gaussians to achieve state-of-the-art quality in reconstruction and rendering speed [17]. Starting from sparse point clouds generated by Structure-from-Motion (SfM), 3D GS uses differentiable rendering to control the density and refine parameters adaptively.

Most NeRF and 3D GS-based methods use frame data from traditional cameras [20], [23]. Few works have attempted to reconstruct radiance fields in dark and low-light environments due to the limited dynamic range of a frame-based camera. Zhang et al. [24] attempt to solve robot exploration in the dark using 3D Gaussians Relighting, but a constant illumination placed in front of the robot is necessary. Mildenhall et al. [23] use NeRF to reconstruct the radiance field in the dark by treating it as a denoising problem from HDR frames. However, it requires an HDR sensor in real applications and does not explore the effect of motion blur in low-light conditions. Some works [17], [25], [26], [27] have devoted to event-based radiance field reconstruction. But they all reconstruct under normal lighting conditions. Thus, limited work has been devoted to leveraging the dynamic range and temporal resolution properties of an event camera to solve such an issue.

### B. Event-Based Video Reconstruction

Many works have been devoted to the field of events to video reconstruction [13], [14], [28]. However, these methods do not work well for reconstructing frames from low-light events. EvLowLight [15] attempts to reconstruct bright-light video by incorporating event data. However, the focus is primarily on the short-to-long exposure problem rather than the dark-to-bright reconstruction. In their case, the darkness of the frames is caused by short exposure times, not actual low-light conditions, meaning the scene remains well-lit from the event camera's perspective. EvLight [16] utilizes event data to enhance single images rather than addressing video reconstruction, and its results may still exhibit partial chromatic aberrations. In contrast, our task focuses on using an event

camera to reconstruct bright radiance fields in real low-light environments. Others [29], [30] have tried to use an event to guide video reconstruction in the dark. However, estimating absolute intensity values in a video solely from brightness changes recorded in events is a highly ill-posed problem [15]. Thus, directly applying any of these models to our task is not feasible.

### C. Event-Based Radiance Field Reconstruction

Event cameras, with their unique characteristics, such as motion-blur resistance, high dynamic range, low latency, and energy efficiency, are increasingly utilized in computer vision and computational imaging applications [31], [32], [33], [34]. Approaches integrating NeRF with event data [18], [35] apply volumetric rendering using event supervision, which takes advantage of NeRF's inherent multi-view consistency to extract coherent scene structures. However, the computational demands for training and optimizing event-based NeRF pipelines remain substantial [17]. Recent works have aimed to integrate 3D GS with event data [17], [19], [26], [36], [37], [38]. Regardless of the base approach, existing event-involved approaches were designed for general scenes. For those who take event streams as input only, the noisier and more random nature of events under dark and low-light environments makes these approaches unrobust and ineffective for reconstruction. For others who also utilize blurred frames as additional input, it further suffers from the low intensity in low-light frames.

## III. Methodology

### A. Preliminary on 3D Gaussian Splatting

The 3D Gaussian Splatting (3D GS) approach [20] models detailed 3D scenes as point clouds, where each point is represented as a Gaussian, defining the structure of the scene. Each Gaussian is characterized by a 3D covariance matrix $\Sigma$ and a central location $\mathbf{x}$:

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

where the central location $\mathbf{x}$ serves as the Gaussian's mean. To enable differentiable optimization, $\Sigma$ is decomposed into a rotation matrix $R$ and a scaling matrix $S$:

$$\Sigma = RSS^TR^T. \tag{2}$$

Rendering from various viewpoints utilizes the splatting technique from [39], which projects Gaussians onto camera planes. This technique, based on the method in [40], involves the viewing transformation matrix $W$ and the Jacobian $J$ of the affine projective transformation. In-camera coordinates, the covariance matrix $\Sigma'$ is then represented as:

$$\Sigma' = JW\ \Sigma\ W^TJ^T. \tag{3}$$

To summarize, each Gaussian point includes several attributes: position $\mathbf{x} \in \mathbb{R}^3$, color encoded via spherical harmonics coefficients $\mathbf{c} \in \mathbb{R}^k$ (with $k$ indicating the degrees of freedom), opacity $\alpha \in \mathbb{R}$, rotation quaternion $\mathbf{q} \in \mathbb{R}^4$, and scaling factor $\mathbf{s} \in \mathbb{R}^3$. For each pixel, the combined color and opacity values of all Gaussians are computed using the Gaussian representation from Equation 1. The blending of colors $C$ for $N$ ordered points projected onto a pixel is:

$$C = \sum_{i\in N}\mathbf{c}_i\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j), \tag{4}$$

where $\mathbf{c}_i$ and $\alpha_i$ represent the color and density of each point, respectively. These values are influenced by each Gaussian's covariance matrix $\Sigma$ and modified through per-point opacity and spherical harmonics color coefficients.

### B. Noisy Events in Dark as Supervisory Signals

The event sensor output can be formulated as:

$$E_{gt} = \Gamma\left\{\log\left(\frac{I_t + c}{I_{t-w} + c}\right), \epsilon\right\}, \tag{5}$$

where $E_{gt}$ is the captured events, $I_t$ and $I_{t-w}$ are intensity images at the two timestamps, $\Gamma\{\theta, \epsilon\}$ represents the conversion function from log-intensity to events, $c$ is an offset value to prevent log(0), $\epsilon$ is the event triggering threshold. $\Gamma\{\theta, \epsilon\} = 1$ when $\theta \geq \epsilon$, indicating positive event triggered, and $\Gamma\{\theta, \epsilon\} = -1$ when $\theta \leq -\epsilon$, indicating negative event triggered [41]. Else, no events are triggered.

Events triggered in low-light environments are much noisier [21], [42]. To address noise, which is often more prevalent in microsecond-level event data [43], we preprocess the event stream using the $y$-noise filter from [44]. As shown in Fig. 3 (a), raw events captured in the dark contain extreme noise and randomness. After applying the noise filter, the cleaned event stream becomes more helpful as a supervision signal, as shown in Fig. 3 (b). Noise filtering is essential for accurate view rendering, as shown by the ablation study section.

Our objective, depicted in Fig. 2, is to reconstruct a radiance field using differentiable 3D Gaussian functions $G$, guided by event and corresponding blurred dark frames. We accumulate the ground truth event signal between timestamps $t_1$ and $t_2$ to form the supervisory signal using Eq. 5 by aggregating the polarities of all events occurring between times $t$ and $t - w$, indexed by their pixel location.

At each timestamp, $t_k$, a rendering result is produced with the camera at pose $p_k$. Thus, we calculate two rendered frames from the 3D GS model: $I_1 = G(p_1)$ and $I_2 = G(p_2)$, where $I_1$ and $I_2$ are RGB renderings at times $t_1$ and $t_2$, respectively. Here, $G$ denotes the 3D GS model, and $p_1$ and $p_2$ are the camera poses at timestamps $t_1$ and $t_2$. The logarithmic image representation is defined as $L(I_t) = \log\left((I_t)^g + \kappa\right)$, where $\kappa = 1 \times 10^{-5}$ (to prevent NaN values) and $g$ is a gamma correction factor set to 2.2 across all experiments, following [18], [45]. From this, we obtain the predicted cumulative difference $E_{pred} = L(I_2) - L(I_1)$.

### C. Color Tone Consistency Module

Pure event utilization and supervision are insufficient for radiance field reconstruction on real data [17], [46], let alone reconstruction under low-light environments. Therefore, prior knowledge is required in the 3D GS training pipeline. Following EvLowLight [15], we deploy feature encoders for

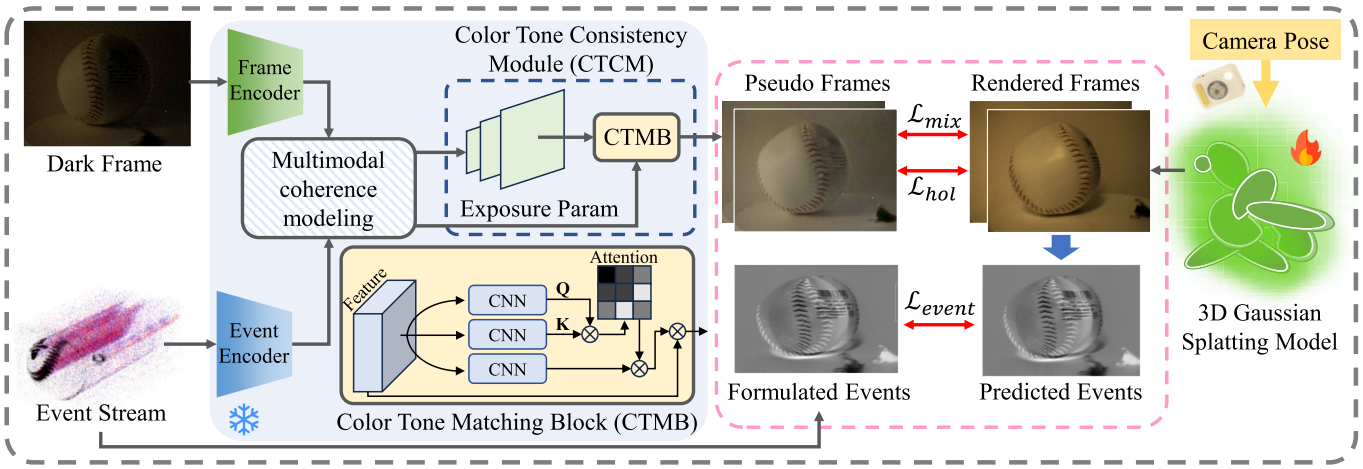


Fig. 2. Overview of the Dark-EvGS pipeline for radiance field reconstruction in the dark. We obtain dark frames, events, and camera parameters using an event camera under low light. The frame encoder and event encoder will extract features, which will then be forwarded to multimodal coherence modeling. The proposed Color Tone Consistency Module takes the features and decodes them into pseudo bright frames with consistent color tone (described in Sec. III-C). We used pre-trained weights from [15] for the initialization of encoders and the multimodal coherence modeling block. The ice icon denotes frozen weights, and the fire icon denotes optimizing weights. Via the 3D GS model, we render bright and shape frames when given the corresponding camera position and pose, and formulate (blue arrow) the rendered results to predicted event signals. The same process applies when formulating captured events to ground truth supervision signals. To supervise accurate training, the proposed triplet-level loss (red lines) provides a holistic view while keeping the rendered results sharp and accurate in detail. As a result, our method enables high-quality radiance field reconstruction in the dark.

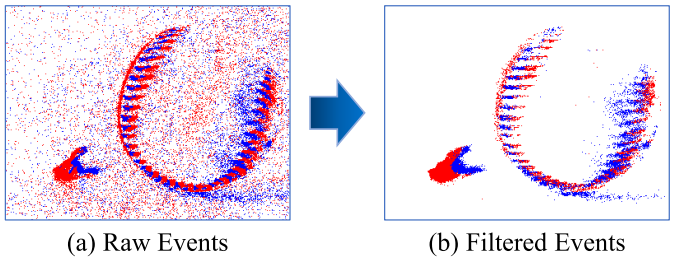


Fig. 3. Frame-based cameras are unable to capture enough signals for residence reconstruction in the dark. Event cameras, on the other hand, can capture ample signals, including ignored details by frame-based cameras (a). However, these raw events captured are rather noisy and random (a), which is one of the biggest challenges for existing methods [13], [15], [17], [18]. To solve this problem, we proposed a utilization for noisy events for accurate and clean supervisory signals (b).

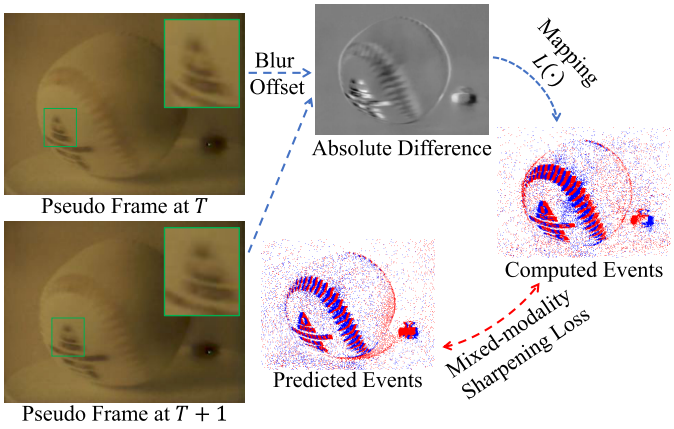


Fig. 4. Illustration of our proposed mix-modality sharpening loss. The blurred regions (green boxes) can be offset, thus producing sharper rendering results.

both frame and event modality along with the Multimodal Coherence Modeling block (MCM), with corresponding pre-trained weights for initialization. We then propose the Color Tone Consistency Module (CTCM), which consists of a proposed Color Tone Matching Block (CTMB) along with the decoders from MCM. CTMB takes the decoded frames during the decoding process and corrects the color tone. Finally, the CTCM outputs the predicted pseudo frames using the corrected frames and the exposure parameter predicted from the MCM. Specifically, the CTMB, inspired by recent advancements in transposed self-attention [47], takes the last feature map $F \in \mathbb{R}^{H\times W\times C}$, and derives query $Q$, key $K$, and value $V$ representations using a $1 \times 1$ convolution followed by a depth-wise convolution. Subsequently, $Q$ and $V$ are reshaped, and an attention map $M \in \mathbb{R}^{C\times C}$ is computed through matrix multiplication. The output of the CTMB $F_{\text{out}}$ represents the color-corrected feature, and the bright light frame with corrected color tone can be derived from $F_{\text{out}}$ and the exposure parameter from MCM. This allows CTCM to generate global color correction via channel-wise self-attention while enhancing local color adjustments through CNNs.

### D. Triplet Supervision for Radiance Field in Dark

There are three major challenges in supervision under low light conditions: First, events are too noisy to provide accurate supervision alone. Second, pseudo-bright frames generated by CTCM cannot provide robust details due to the data gap between pre-trained knowledge and actual training scenarios. Third, in low light environments, captured dark frames are easier to motion blur as a longer exposure time is required [48], [49]. Thus, a sharpening module is needed because the pseudo-bright frames may also contain motion blur when given blurred dark frames.

To tackle these challenges, we propose triplet-level supervision to train and reconstruct radiance field representation in the dark accurately. Specifically, frame-level holistic supervision,

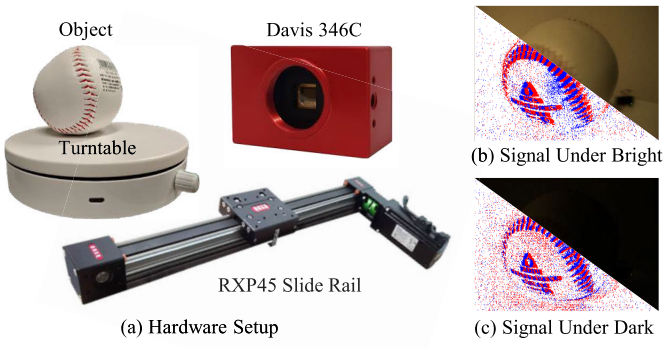


Fig. 5. Demonstration of hardware setup and dataset collection. Objects are captured either on a turntable or using a linear translation stage, with an event camera observing the scene (a). Signals are collected under both bright (b) and dark (c) illumination conditions. In low-light environments, frame-based images suffer from severe information loss and motion blur, while event signals become noisy.

event-level granular supervision, and mixed-modality sharpening supervision. At the first level, Dark-EvGS leverages pseudo-bright frames, though it may be inaccurate in detail, to obtain a holistic view of knowledge through frame-level supervision. At the second level, filtered events serve as a supervisory signal to refine granular details that are ignored in previously used frames. Finally, at the third level, to mitigate motion blur effects in dark scenes, a combined event and frame supervision approach is applied using a mixed-modality sharpening strategy.

*1) Frame-Based Holistic Supervision:* Pure events supervision brings additional noise under low lights and does not lead to good reconstruction results (Fig. 7 (f)). Though the pseudo bright light frame, generated by CTCM, may contain blur and minor defects in details, as ground truth, it provides a holistic overview of how the object in the current frame looks and supervises the rendered frame so it follows that high-level holistic framework.

At a time window with range $t_1$ and $t_2$, given two low-light dark frames $I^{t_1}_{low}$ and $I^{t_2}_{low}$ captured at these two timestamps, and the ground truth accumulated events $E_{gt}$ between the time window, the Color Tone Consistency Module, CTCM, generates two estimated bright frames $B^{t_1}$, and $B^{t_2}$, where $(B^{t_1}, B^{t_2}) = \text{CTCM}(I^{t_1}_{low}, I^{t_2}_{low}, E_{gt})$. We then use $N$ rendered frames from the GS model to compute $\mathcal{L}2$ loss with an estimated bright frame $B$, where $N$ is the number of viewpoints. Therefore, the frame-based holistic loss is defined as:

$$\mathcal{L}_{hol}(I, B) = \frac{1}{N}\sum_{i=1}^{N}(B_i - I_i)^2. \tag{6}$$

*2) Event-Level Granular Supervision:* Granular details are ignored by the frame camera when captured in the dark due to limited dynamic range. Furthermore, the estimated pseudo-bright frame from CTCM can only give a holistic view, ignoring the same region of sharp and accurate details. Event streams captured by the event camera play a critical role in supervising details and minors because they are high in temporal resolution, so details between frames can be captured, and they are wide in dynamic range, so unrevealed information can be seen. We use a y-noise filter [44] $Y$ to form ground truth event $E_{gt} = Y(E_{raw})$ from raw captured events $E_{raw}$. To learn $E_{pred}$ with the formulated clean event signals $E_{gt}$, we adopt the classic event supervision loss from [18], where $N$ is the number of rendering viewpoints:

$$\mathcal{L}_{event}(E_{pred}, E_{gt}) = \frac{1}{N}\sum_{i=1}^{N}(E^i_{pred} - E^i_{gt})^2. \tag{7}$$

This helps to refine details that the frame-based holistic supervision failed to reconstruct.

*3) Mixed-Modality Sharpening Supervision:* Because the generated pseudo-bright frames from CTCM may contain motion blur, applying a strict $\mathcal{L}2$ loss leads to the rendered results being blurry as well. To correct motion blur in the rendering process, we propose a mixed-modality sharpening loss. The key idea is: though individual generated bright frames from CTCM contain motion blur, consecutive frames retain temporal consistency - motion blur and sharp areas of nearby frames belong to similar area and thus can be compensated by subtraction (Fig. 4 green boxes). Mathematically speaking, the subtraction (Fig. 4 absolute difference) between two generated bright frames is sharp and stands for the pixel shift in spatial. Thus, it can be a good guide for the spatial shift to rendered GS results. Therefore, we map the sharp subtraction results using the logarithmic function $L$, to formulate the computed event signal and then supervise the predicted events from the 3D GS network. This effectively eliminates the blurring effect and produces sharper results during rendering. Here, $I^{t_1}$ and $I^{t_2}$ denote two rendered intensity images from the 3D Gaussian Splatting model at timestamps $t_1$ and $t_2$, respectively, while $B^{t_1}$ and $B^{t_2}$ are the corresponding pseudo-bright frames generated by the CTCM. The operator $L(\cdot)$ represents the logarithmic intensity mapping used to compute log-intensity differences. The index $i$ enumerates frames, and $N$ denotes the total number of frames. The mixed-modality sharpening loss $\mathcal{L}_{mix}$ is defined as:

$$\begin{aligned}&\mathcal{L}_{mix}(I^{t_1}, I^{t_2}, B^{t_1}, B^{t_2})\\ &= \frac{1}{N}\sum_{i=1}^{N}((L(I^{t_1}) - L(I^{t_2})) - (L(B^{t_1}) - L(B^{t_2})))^2.\end{aligned} \tag{8}$$

*4) Final Loss:* The final loss function is expressed as:

$$\mathcal{L} = \mathcal{L}_{hol} + \lambda_1\mathcal{L}_{event} + \lambda_2\mathcal{L}_{mix}, \tag{9}$$

where $\lambda_1$ and $\lambda_2$ is set to 0.25 across all experiments.

Noise filtering using the $y$-noise filter effectively removes interfering signals, especially in dark and low-light, while the triplet supervision keeps the training stable across varying conditions. This stability ensures consistent rendering performance, regardless of data source.

## IV. Experiments

This section provides details of our data collection, experiment setup, evaluation results, and analysis.

### A. Data Collection

In this subsection, we introduce the real-world data collection process in detail. Fig. 5 (a) demonstrates the hardware

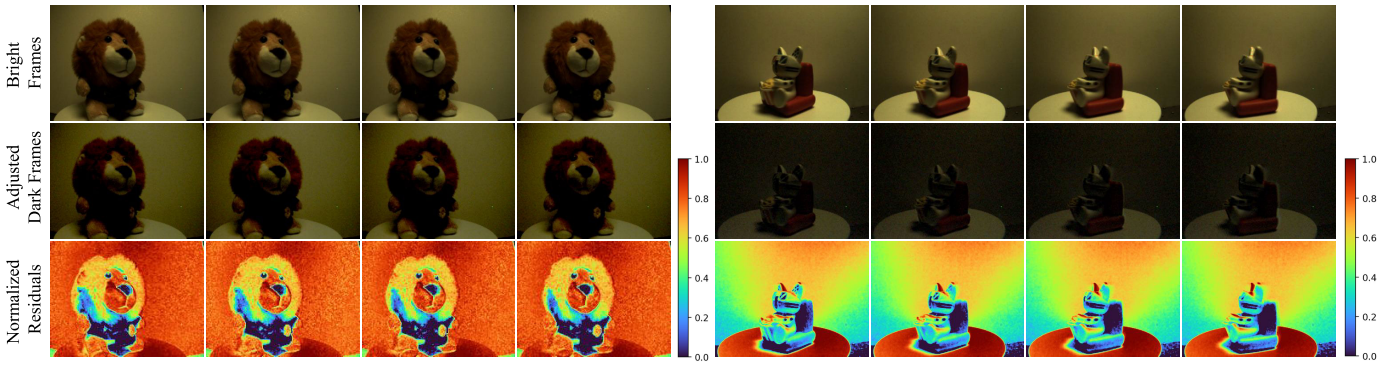


Fig. 6. Paired bright and dark frame samples from the proposed real-world dataset. The first row shows bright frames captured under normal illumination. The second row shows the corresponding dark frames after brightness and contrast adjustment (+ 40% brightness, −40% contrast), which is applied only for visualization and residual analysis to better reveal structural correspondence. The third row visualizes the residual heatmaps computed between each bright frame and its adjusted dark counterpart. As shown, the residuals are dominated by global intensity differences, while object contours and structural edges remain well aligned, indicating that the paired bright–dark frames are temporally synchronized and pixel-wise aligned, differing primarily in illumination conditions rather than spatial misalignment.

setup. Our real-world dataset consists of two complementary parts: (i) single-object turntable scenes and (ii) forward-facing scenes with complex objects. The former enables controlled analysis under low-light conditions, while the latter is introduced to evaluate the generalization ability of the proposed method beyond turntable-based settings.

*1) Single-Object Turntable Scenes:* To collect the single-object dataset, we place a static object on a constant, unknown-speed turntable. A Davis 346 Color event camera is held steady and points toward the object throughout the capture process. We use a TES 1339R light meter to measure the illumination level of the environment in lux. Six single-object scenes are collected, and for each scene, we set two lighting conditions to form paired data: bright (around 300 lux) and dark (less than 40 lux). All scenes are fully covered by curtains to eliminate natural light interference, and a controllable artificial light source is placed directly above the object.

Under bright lighting, we use the highest illumination level and capture the corresponding frames, event streams, and camera information during a full 360-degree rotation. Subsequently, we repeat the same mechanically controlled rotation trajectory under low-light conditions, leaving only limited illumination on the object, and capture the same set of data. Among the six scenes, three are categorized as moderate lighting (between 20 and 40 lux, including "Baseball", "House", and "Lion"), and three are categorized as challenging lighting (less than 20 lux). For each scene, paired dark–bright frames, event streams, and camera poses are prepared. Bright-light frames are used as ground-truth supervision for evaluation, while only dark frames and/or dark events are available during training and supervision.

Fig. 5 (b) and (c) show example data captured under bright and dark lighting conditions, respectively. As illustrated, frames captured under normal lighting are sharp and informative, while frames captured under low-light conditions suffer from severe information loss and motion blur, making them difficult to interpret for both humans and machines. At the same time, event streams captured in dark environments exhibit significantly higher noise levels.

A visual illustration of paired bright and dark frames is provided in Fig. 6, where each bright–dark pair corresponds to the same camera viewpoint along the rotation trajectory. Since both sequences are captured using the same camera with fixed intrinsics and an identical mechanically controlled motion, the paired frames are temporally synchronized and pixel-wise aligned, differing only in illumination conditions.

*2) Forward-Facing Scenes With Complex Objects:* While turntable-based capture is commonly adopted in radiance field reconstruction for controlled benchmarking, it does not fully reflect forward-facing camera motion encountered in practical scenarios. To address this limitation, we additionally collect forward-facing scenes with complex objects using the RXP45 Slide Rail linear translation stage (shown in Fig. 5 (a)). In this setting, the event camera undergoes a translational motion along a linear trajectory while observing cluttered scenes composed of multiple objects and background structures. This setup simulates forward-facing viewpoints and unbounded scene configurations. The translation stage is controlled by an Arduino-based pulse generator. The stepper driver is configured with 6400 pulses per revolution, and the stage is driven at a constant speed of 20 RPM or 2.67cm/s. During acquisition, the camera is placed on a fixed location of the translation stage, moving along with it, and illumination remains the same, and we record event streams and frames at a constant sensor configuration. These forward-facing scenes are captured under low-light conditions following the same acquisition protocol, and the corresponding bright-light frames are used solely for evaluation. By incorporating this additional data, we aim to demonstrate that the proposed framework is not limited to turntable-based scenarios but can also generalize to more realistic camera trajectories and complex scene layouts.

## B. Implementation Details

Our implementation builds upon the 3D GS framework [20], leveraging its primary structure and functionalities. As 3D GS requires a point cloud input, we randomly initialize $10^3$ points to create the starting point cloud, as the structure-from-motion initialization [50] in the original setup is not directly applicable to event data. We use the pre-trained weights

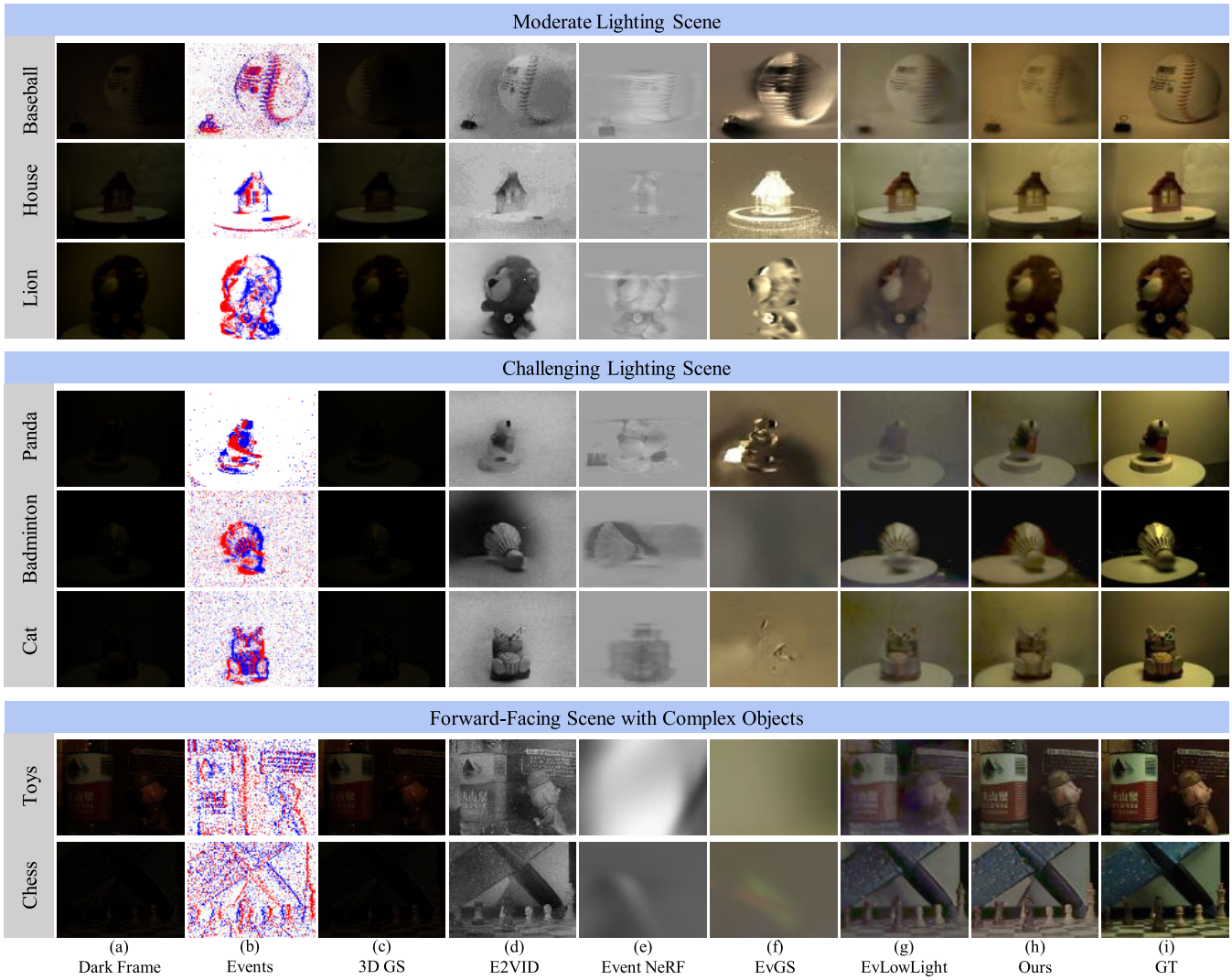


Fig. 7. A visual comparison of different approaches on different scenes captured: (a) dark frames captured by a frame-based camera under low lights; (b) filtered events captured under low lights by an event camera; (c) results from 3D GS [20] using dark frames; (d) rendering results from 3D GS using reconstructed frames from E2VID [13]; (e) results from EventNeRF [18]; (f) results from Ev-GS [17]; (g) results from Ev-LowLight; (h) results from ours; (i) ground truth frames captured under bright lights. The top two rows correspond to object-centric scenes under moderate and challenging low-light conditions, respectively, while the bottom row shows forward-facing scenes with complex objects captured using a translational camera trajectory. Our approach renders brighter and sharper results purely using signals captured under dark environments compared to existing approaches across all scenes.

from EvLowLight on its existing modules with randomly initialized CTMB weights as a start. They are then tuned with five hundred unrelated real-world dark-bright data pairs and corresponding events[1] for 5 epochs to provide robust prior knowledge on pseudo frames. All experiments were conducted on an NVIDIA RTX 3090 GPU. We used specific hyperparameters to ensure optimal performance. Training was conducted for a total of 30,000 iterations for all scenes. For position optimization, the learning rate was scheduled to decay from $1.6 \times 10^{-4}$ to $1.6 \times 10^{-6}$. The learning rates for optimizing features, opacity, scaling, and rotation were set to $2.5 \times 10^{-3}$, $5 \times 10^{-2}$, $5 \times 10^{-3}$, and $1 \times 10^{-3}$, respectively.

### C. Quantitative Evaluation

We compare the radiance field reconstruction quality and efficiency of various methods, including frame-based method (F): 3D GS [20]; pure-event based methods (E): E2VID [13], EventNeRF [18], and Ev-GS [17]; and hybrid methods (E+F): EvLowLight [15], Sweep-EvGS [26], and E2GS [38], on our real-world dataset. For approaches that do not directly perform radiance field reconstruction (e.g., E2VID and EvLowLight), their enhanced frame outputs are further processed by a 3D Gaussian Splatting pipeline to obtain the final 3D reconstruction results used for evaluation. To further ensure a fair comparison, we additionally report results of fine-tuned baseline methods (denoted by *), where E2VID and EvLowLight are fine-tuned on the same type of auxiliary real-world dark–bright data pairs as CTCM training, with no overlap with the evaluation scenes. Following the standard protocol, we adopt an 8:2 train-test split to ensure reliable evaluation across both moderate and challenging lighting scenes. The evaluation considers the following two key aspects:

(i) *Reconstruction Quality:* We use PSNR, SSIM, and LPIPS as metrics to quantify the fidelity of reconstructed

[1]These are extra-captured natural scene data using Davis 346C.

TABLE I

QUANTITATIVE COMPARISON ON THE REAL DATASET. METRICS INCLUDE PSNR, SSIM, AND LPIPS ACROSS VARIOUS SCENES. ↑ MEANS HIGHER IS BETTER AND ↓ MEANS LOWER IS BETTER. THE FIRST COLUMN (**INPUT**) INDICATES THE INPUT MODALITY: **F** = FRAME ONLY, **E** = EVENT ONLY, **E+F** = EVENT + FRAME. FOR METHODS THAT ARE NOT DESIGNED FOR RADIANCE FIELD RECONSTRUCTION (E.G., E2VID AND EVLOWLIGHT), THEIR ENHANCED OUTPUTS ARE USED AS INPUT TO A SUBSEQUENT 3D GAUSSIAN SPLATTING STAGE FOR 3D RECONSTRUCTION AND EVALUATION. RESULTS MARKED WITH * INDICATE MODELS FINE-TUNED ON THE SAME REAL-WORLD DARK–BRIGHT DATA PAIRS USED FOR CTCM TRAINING. BEST RESULTS ARE IN BOLD AND SECOND BEST ARE UNDERLINED

| **Moderate Lighting Scene (20 - 40 lux)** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Input** | **Method** | Baseball (28 lux) | | | House (38 lux) | | | Lion (32 lux) | | |
| | | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| F | 3D GS [20] | 9.91 | 0.15 | 0.52 | 8.79 | 0.21 | 0.55 | 9.32 | 0.23 | 0.58 |
| E | E2VID [13] | 9.78 | 0.37 | 0.65 | 9.32 | 0.49 | 0.69 | 9.54 | 0.46 | 0.72 |
| | E2VID* | 11.24 | 0.41 | 0.64 | 10.62 | 0.50 | 0.67 | 9.93 | 0.47 | 0.71 |
| | EventNeRF [18] | 9.05 | 0.59 | 0.58 | 10.06 | 0.62 | 0.67 | 10.43 | 0.60 | 0.68 |
| | EvGS [17] | 15.82 | 0.72 | 0.51 | 13.27 | 0.67 | 0.57 | 14.95 | 0.68 | 0.60 |
| E+F | E2GS [38] | 14.90 | 0.70 | 0.55 | 14.31 | 0.68 | 0.55 | 14.19 | 0.70 | 0.62 |
| | SweepEvGS [26] | 20.25 | 0.79 | 0.40 | 19.29 | 0.72 | 0.41 | 20.37 | 0.75 | 0.45 |
| | EvLowLight [15] | 20.34 | 0.83 | 0.38 | 19.31 | 0.75 | 0.45 | 18.86 | 0.74 | 0.49 |
| | EvLowLight* | 21.02 | 0.81 | 0.37 | 19.17 | 0.70 | 0.48 | 18.94 | 0.74 | 0.47 |
| | Ours | **26.58** | **0.87** | **0.31** | **23.26** | **0.81** | **0.36** | **31.22** | **0.85** | **0.30** |
| **Challenging Lighting Scene (<20 lux)** | | | | | | | | | | |
| **Input** | **Method** | Panda (17 lux) | | | Badminton (14 lux) | | | Cat (16 lux) | | |
| | | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| F | 3D GS [20] | 8.73 | 0.08 | 0.69 | 14.69 | 0.58 | 0.35 | 10.61 | 0.09 | 0.46 |
| E | E2VID [13] | 9.38 | 0.38 | 0.70 | 8.26 | 0.15 | 0.65 | 8.92 | 0.35 | 0.64 |
| | E2VID* | 10.14 | 0.40 | 0.67 | 9.12 | 0.31 | 0.65 | 9.15 | 0.39 | 0.60 |
| | EventNeRF [18] | 9.23 | 0.56 | 0.69 | 5.91 | 0.14 | 0.48 | 5.95 | 0.33 | 0.46 |
| | EvGS [17] | 12.37 | 0.63 | 0.60 | 10.89 | 0.17 | 0.35 | 10.58 | 0.21 | 0.37 |
| E+F | E2GS [38] | 14.20 | 0.69 | 0.50 | 13.72 | 0.68 | 0.60 | 11.73 | 0.27 | 0.39 |
| | SweepEvGS [26] | 17.22 | 0.73 | 0.47 | 18.84 | 0.33 | 0.28 | 18.56 | 0.72 | 0.45 |
| | EvLowLight [15] | 16.28 | 0.76 | 0.46 | 18.87 | 0.31 | 0.20 | 19.67 | 0.78 | 0.38 |
| | EvLowLight* | 16.51 | 0.75 | 0.48 | 19.20 | 0.32 | 0.22 | 19.23 | 0.74 | 0.40 |
| | Ours | **20.84** | **0.80** | **0.36** | **23.70** | **0.42** | **0.19** | **21.64** | **0.83** | **0.33** |

images compared to the bright-light ground truth. Higher PSNR and SSIM values correspond to better reconstruction in terms of brightness and structural similarity, while lower LPIPS values indicate better perceptual similarity. As reported in Table I and Table II, our method consistently achieves the best performance across all scenes. For example, in the *Lion* scene under moderate conditions, our approach improves PSNR to 31.22 and SSIM to 0.85, outperforming the second-best EvLowLight [15] by a clear margin, while reducing LPIPS to 0.30. Similar advantages are observed in challenging lighting scenes, such as the *Cat* sequence, where our method achieves the highest perceptual similarity and structural fidelity.

(ii) *Efficiency:* To evaluate practical usability, we measure training time, rendering FPS, and GPU memory consumption. As shown in Table III, our method strikes a favorable balance between efficiency and accuracy. Specifically, Dark-EvGS requires only 4 min of training, significantly faster than hybrid methods such as E2GS [38] (30 min) and Sweep-EvGS [26] (8 min), while achieving the highest rendering FPS (41 and 35 FPS in moderate and challenging lighting scenes, respectively). In terms of GPU memory usage, our framework also remains lightweight (1.2 to 1.7 GB), close to the most efficient 3D GS [20] baseline, and much lower than NeRF-based approaches such as EventNeRF [18] (12 GB).

Overall, these results highlight that Dark-EvGS is not only more accurate in reconstructing radiance fields under dark environments but also practical for real-time deployment, owing to its fast training, high rendering speed, and modest memory requirements.

### D. Ablation Studies

We conduct ablation studies on the effect of rendering quality from the following three aspects: the supervision module and the event utilization module; we then provide the ablation study on the effect of efficiency from each component.

*1) Ablation Study on Supervision Module:* To understand the individual contributions of each component in our proposed framework, Dark-EvGS, we conducted a detailed ablation study. Table IV presents the results of the PSNR and SSIM metrics on four test samples, with various combinations of loss components applied: $\mathcal{L}_{hol}$, $\mathcal{L}_{event}$, and $\mathcal{L}_{mix}$. The study highlights the unique impact of each component, demonstrating their complementary roles in improving rendering quality.

The holistic supervision loss, $\mathcal{L}_{hol}$, significantly improves global coherence, as shown by the notable increase in both PSNR and SSIM when applied alone. The event-based supervision loss, $\mathcal{L}_{event}$, contributes to capturing dynamic details, while the mix supervision loss, $\mathcal{L}_{mix}$, refines local textures.

TABLE II

QUANTITATIVE COMPARISON ON FORWARD-FACING SCENES WITH COMPLEX OBJECTS. METRICS INCLUDE PSNR, SSIM, AND LPIPS. ↑ INDICATES HIGHER IS BETTER AND ↓ INDICATES LOWER IS BETTER. THE FIRST COLUMN (**INPUT**) INDICATES THE INPUT MODALITY: **F** = FRAME ONLY, **E** = EVENT ONLY, AND **E+F** = EVENT + FRAME. FOR METHODS NOT ORIGINALLY DESIGNED FOR RADIANCE FIELD RECONSTRUCTION (E.G., E2VID AND EVLOWLIGHT), THEIR ENHANCED OUTPUTS ARE FED INTO A SUBSEQUENT 3D GAUSSIAN SPLATTING STAGE FOR RECONSTRUCTION. RESULTS MARKED WITH * INDICATE MODELS FINE-TUNED ON THE SAME REAL-WORLD DATA PAIRS USED FOR CTCM TRAINING. BEST RESULTS ARE SHOWN IN BOLD AND SECOND-BEST RESULTS ARE UNDERLINED

| Input | Method | Toys (30 lux) | | | Chess (15 lux) | | |
|---|---|---|---|---|---|---|---|
| | | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| F | 3D GS | 10.12 | 0.22 | 0.58 | 9.84 | 0.19 | 0.60 |
| E | E2VID | 10.45 | 0.39 | 0.66 | 10.02 | 0.36 | 0.68 |
| | E2VID* | 14.36 | 0.63 | 0.53 | 13.88 | 0.55 | 0.55 |
| | EventNeRF | 10.21 | 0.55 | 0.60 | 9.76 | 0.52 | 0.62 |
| | EvGS | 11.92 | 0.48 | 0.59 | 11.37 | 0.45 | 0.57 |
| E+F | E2GS | 15.31 | 0.71 | 0.46 | 14.85 | 0.69 | 0.48 |
| | SweepEvGS | 18.74 | 0.77 | 0.41 | 18.29 | 0.75 | 0.43 |
| | EvLowLight | 18.96 | 0.79 | 0.37 | 18.12 | 0.77 | 0.41 |
| | EvLowLight* | 19.42 | 0.78 | 0.40 | 18.65 | 0.74 | 0.43 |
| | Ours | **22.85** | **0.84** | **0.30** | **21.97** | **0.82** | **0.32** |

TABLE III

COMPARISON OF TRAINING TIME (TRAIN), RENDERING PERFORMANCE (FPS), AND GPU MEMORY USAGE (MEM) ACROSS DIFFERENT METHODS ON MODERATE LIGHTING SCENE AND CHALLENGING LIGHTING SCENE. THE METRICS INCLUDE TRAINING TIME (MINUTES AND HOURS), FRAMES PER SECOND (FPS) FOR REAL-TIME RENDERING, AND GPU MEMORY USAGE (GB) DURING TRAINING. ↑ INDICATES THAT HIGHER VALUES ARE BETTER, WHILE ↓ INDICATES THAT LOWER VALUES ARE BETTER. THE METRICS IN BOLD ARE RANKED FIRST, AND THE UNDERLINED METRICS ARE RANKED SECOND

| Method | Moderate Lighting | | | Challenging Lighting | | |
|---|---|---|---|---|---|---|
| | Train↓ | FPS↑ | Mem↓ | Train↓ | FPS↑ | Mem↓ |
| 3D GS | **3min** | 40 | **1GB** | **3min** | 32 | **1.1GB** |
| E2VID | 5min | 35 | 1.8GB | 5min | 30 | 1.8GB |
| EventNeRF | 19h | 0.5 | 12GB | 26h | 0.3 | 12GB |
| EvGS | 4.5min | 33 | 1.5GB | 4.5min | 31 | 2.1GB |
| E2GS | 30min | 30 | 5GB | 32min | 28 | 5GB |
| SweepEvGS | 8min | 38 | 2GB | 8min | 40 | 2GB |
| EvLowLight | 11h | 3.5 | 2GB | 10h | 3.5 | 3GB |
| Ours | 4min | **41** | 1.2GB | 4min | **35** | 1.7GB |

When combined, these components work synergistically, with $\mathcal{L}_{hol} + \mathcal{L}_{event}$ balancing global and dynamic details and the addition of $\mathcal{L}_{mix}$ enhancing fine-grained reconstruction.

The full framework, combining all three components, achieves the best results across all samples, with the highest PSNR and SSIM values, such as 31.22 and 0.85 for the "Lion" scene. These results confirm that each component plays a vital role and that their combination maximizes rendering quality for low-light radiance field reconstruction.

*2) Ablation Study on Utilization Module:* To evaluate the effectiveness of the proposed noise filter in the event utilization module, we conducted an ablation study comparing the performance of our framework with and without the noise filter. In this work, we adopt the y-noise filter from [44], which is specifically designed to suppress background activity noise while preserving meaningful event structures. The core idea is that noise events tend to appear sparsely and randomly in both space and time, whereas signal events form locally dense clusters corresponding to real intensity changes. Concretely, raw events are first grouped within a short temporal window $\Delta T$. For each event, a local spatial neighborhood of size $L \times L$ is considered to construct an event density map. Events with a local density below a predefined threshold are considered noise and are removed. This process effectively filters out isolated events while retaining temporally and spatially coherent event patterns generated by actual scene motion or intensity variation. In our ablation study, we further investigate the impact of the temporal window size $\Delta T$ on reconstruction performance, as it directly controls the trade-off between noise suppression and detail preservation in event density estimation.

Table V reports the PSNR and SSIM results on six test scenes under different noise filtering settings. The results clearly demonstrate the necessity of incorporating event noise filtering. Without the noise filter, the framework struggles to suppress background noise, leading to degraded reconstruction quality across all scenes. For example, on the "Baseball" scene, the model achieves a PSNR of 24.23 and an SSIM of 0.85 without filtering, which improves significantly to 26.58 and 0.87, respectively, when an appropriate noise filter is applied. Moreover, the temporal window size $\Delta T$ plays a critical role in filtering effectiveness. When $\Delta T$ is too small (e.g., $1000\,\mu$s), noise is not sufficiently suppressed, resulting in limited performance gains. Conversely, an overly large temporal window (e.g., $4000\,\mu$s) leads to over-smoothing and loss of informative event structures, even degrading performance below the no-filter baseline. Among all settings, $\Delta T = 2000\,\mu$s consistently achieves the best results across all scenes, indicating an optimal balance between noise suppression and detail preservation.

Overall, these results highlight the critical role of event noise filtering in stabilizing event supervision and refining fine-grained details, thereby improving global coherence and texture fidelity in the reconstructed radiance fields under extreme low-light conditions.

*3) Ablation Study on Efficiency:* To evaluate the efficiency of Dark-EvGS, we conduct an ablation study focusing on the computational cost introduced by each component. Table VI presents the training time (in minutes) and frames per second (FPS) rendered during inference for various configurations of the loss terms.

Our analysis highlights that each component is computationally lightweight and introduces minimal overhead. For example, applying $\mathcal{L}_{hol}$ alone achieves training times of approximately 3.2 minutes and maintains a steady FPS across

TABLE IV

DETAILED ABLATION STUDY ON DARK-EvGS. ✓ DENOTES APPLIED COMPONENTS. THE RESULTS OF PSNR AND SSIM ON FOUR SAMPLES, ADDING EACH COMPONENT INDIVIDUALLY AND APPLYING THE COMPONENTS SEQUENTIALLY, ARE REPORTED. BOLD DENOTES THE HIGHEST SCORE

| $\mathcal{L}_{hol}$ | $\mathcal{L}_{event}$ | $\mathcal{L}_{mix}$ | Baseball | | House | | Lion | | Panda | | Badminton | | Cat | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ✓ | | | 20.37 | 0.83 | 18.57 | 0.74 | 25.36 | 0.80 | 16.31 | 0.76 | 18.21 | 0.28 | 15.44 | 0.71 |
| | ✓ | | 16.67 | 0.73 | 14.96 | 0.69 | 20.73 | 0.77 | 13.19 | 0.64 | 15.12 | 0.19 | 13.11 | 0.62 |
| | | ✓ | 14.61 | 0.70 | 13.21 | 0.67 | 19.11 | 0.75 | 11.89 | 0.51 | 13.12 | 0.15 | 11.61 | 0.49 |
| ✓ | ✓ | | 24.93 | 0.85 | 22.18 | 0.79 | 29.92 | 0.83 | 19.23 | 0.78 | 22.03 | 0.40 | 19.51 | 0.79 |
| ✓ | | ✓ | 22.61 | 0.84 | 20.27 | 0.76 | 27.84 | 0.81 | 18.37 | 0.77 | 20.17 | 0.36 | 17.71 | 0.76 |
| | ✓ | ✓ | 18.23 | 0.78 | 15.72 | 0.71 | 22.45 | 0.78 | 14.41 | 0.67 | 15.73 | 0.21 | 14.23 | 0.66 |
| ✓ | ✓ | ✓ | **26.58** | **0.87** | **23.26** | **0.81** | **31.22** | **0.85** | **20.84** | **0.80** | **23.70** | **0.42** | **21.64** | **0.83** |

TABLE V

ABLATION STUDY ON EVENT NOISE FILTERING WITH DIFFERENT TEMPORAL WINDOW $\Delta T$. THE RESULTS OF PSNR AND SSIM ON SIX SCENES ARE REPORTED. BOLD DENOTES THE HIGHEST SCORE

| Method | Baseball | | House | | Lion | | Panda | | Badminton | | Cat | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| w/o Noise Filter | 24.23 | 0.85 | 21.01 | 0.78 | 29.78 | 0.83 | 19.45 | 0.78 | 20.22 | 0.37 | 20.27 | 0.80 |
| with Noise Filter ($\Delta T = 1000\,\mu s$) | 25.61 | 0.86 | 22.35 | 0.80 | 30.54 | 0.84 | 20.12 | 0.79 | 22.01 | 0.40 | 21.08 | 0.82 |
| with Noise Filter ($\Delta T = 2000\,\mu s$) | **26.58** | **0.87** | **23.26** | **0.81** | **31.22** | **0.85** | **20.84** | **0.80** | **23.70** | **0.42** | **21.64** | **0.83** |
| with Noise Filter ($\Delta T = 3000\,\mu s$) | 25.94 | 0.86 | 22.71 | 0.80 | 30.88 | 0.84 | 20.31 | 0.79 | 22.64 | 0.41 | 21.21 | 0.82 |
| with Noise Filter ($\Delta T = 4000\,\mu s$) | 23.81 | 0.84 | 20.36 | 0.77 | 28.92 | 0.82 | 18.97 | 0.77 | 19.45 | 0.35 | 19.88 | 0.79 |

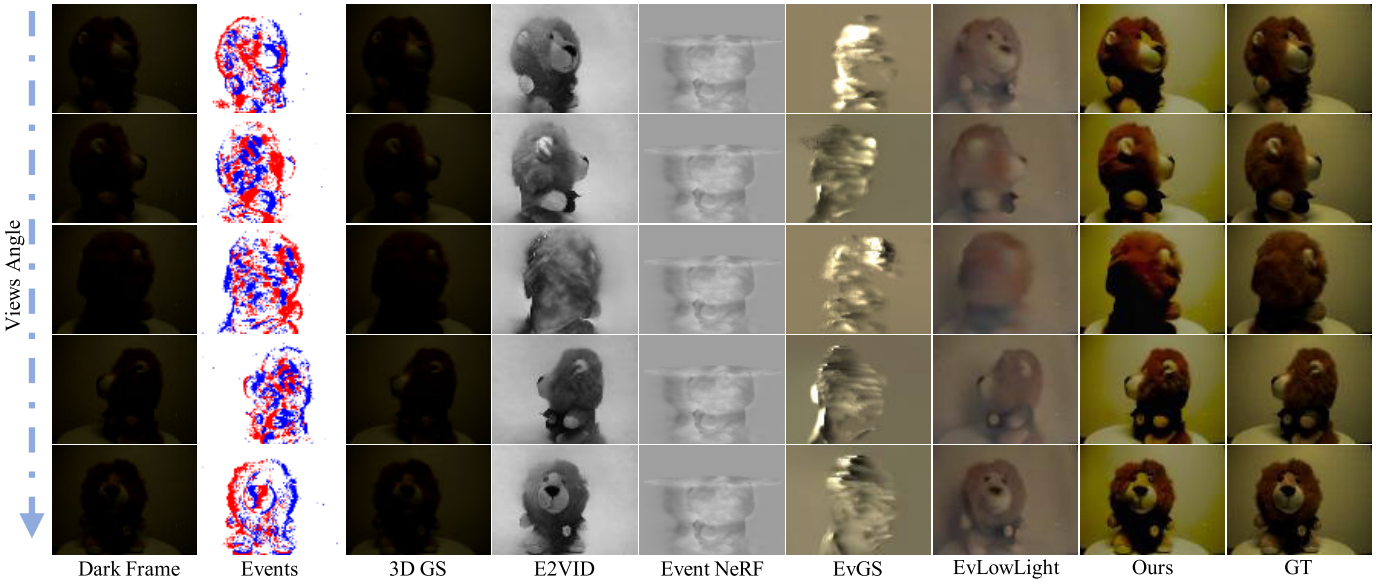


Fig. 8. Multi-view comparison of rendered results from different methods on a low-light scene. For each viewpoint, Dark-EvGS consistently produces sharper, brighter, and more structurally accurate frames compared to baseline methods, demonstrating its effectiveness in reconstructing radiance fields under low-light conditions across diverse camera perspectives.

all samples. Adding $\mathcal{L}_{event}$ or $\mathcal{L}_{mix}$ results in a slight increase in training time but ensures the FPS remains consistent or only marginally reduced. When all components are applied jointly, the total training time remains well within practical limits (approximately 4 minutes), while the FPS remains competitive, reaching up to 41 FPS for moderate lighting scenes. These results demonstrate that the design of Dark-EvGS achieves a favorable trade-off between computational cost and efficiency. Each loss component contributes unique benefits without significantly increasing the computational burden, making the model both effective and resource-efficient.

### E. Qualitative Evaluation

*1) Qualitative Visualization:* We provide a comprehensive qualitative evaluation here. As shown in Fig. 7, frames captured in low light lose detail, while event data preserves fine structures regardless of illumination. Leveraging this

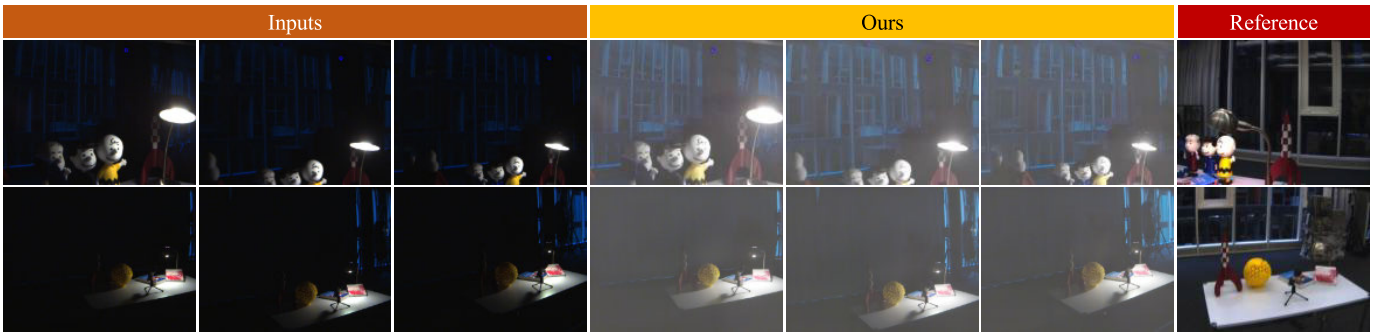


Fig. 9. Qualitative evaluation of our approach on the open benchmark [51]. The orange rows represent the dark input frames, while the yellow rows show our rendered results. The red column contains bright reference frames from the same scene, which are unpaired and used only for qualitative comparison. Our method successfully reconstructs bright and sharp intensity frames even in complex backgrounds and scenes, demonstrating its robustness on an unrelated open benchmark.

TABLE VI
EFFICIENCY ABLATION ON DARK-EVGS. ✓ DENOTES APPLIED COMPONENTS

| Loss config | | | Moderate Lighting | | Challenging Lighting | |
|---|---|---|---|---|---|---|
| $\mathcal{L}_{\text{hol}}$ | $\mathcal{L}_{\text{event}}$ | $\mathcal{L}_{\text{mix}}$ | Time (min)↓ | FPS↑ | Time (min)↓ | FPS↑ |
| ✓ | | | 3.2 | 35 | 3.2 | 35 |
| | ✓ | | 3.5 | 35 | 3.5 | 35 |
| | | ✓ | 3.0 | 40 | 3.0 | 43 |
| ✓ | ✓ | | 3.8 | 35 | 3.7 | 30 |
| ✓ | | ✓ | 3.5 | 38 | 3.4 | 38 |
| | ✓ | ✓ | 3.7 | 35 | 3.6 | 40 |
| ✓ | ✓ | ✓ | 4.0 | 41 | 4.0 | 35 |

advantage, our method produces clearer, more detailed renderings that better match the bright reference views. We also provide more visualization of different views for all samples in Fig. 10, and conduct a multi-view visualizations comparison against other methods in Fig. 8. For a given scene, we render outputs from multiple viewpoints along the camera trajectory using each competing method. The results consistently show that Dark-EvGS produces sharper, brighter, and more geometrically consistent frames across all angles. Compared to baseline methods, our reconstructions maintain clearer structural details and better visual fidelity under varying perspectives, highlighting the robustness and superiority of our approach in low-light radiance field reconstruction.

Although there is currently no benchmark for event-based radiance field reconstruction in the dark, we further demonstrated the robustness of our approach through tasks that are not directly related, such as event-based frame reconstruction in the dark. The EDS dataset [51] provides events and frames under a low-light environment, which is compatible with the input-output format of our approach. However, the dark frames have no paired bright-light ground truth data. The EDS [51] provides unpaired bright frames captured from a different camera perspective and trajectory. In this case, no quantitative evaluation can be performed since there are no paired data. However, using the bright light frames as a reference, we could qualitatively evaluate our approach to this open benchmark for fairness. Fig. 9 demonstrates our rendered results (yellow rows) using the dark frame as input (orange rows) under the peanut and rocket scenes in the EDS [51]. Sub figures in the red column are bright frames in the same scene, but could only be considered as a qualitative reference since they are not paired with the dark frames. Our approach renders bright and sharp intensity results even in complex backgrounds and scenes. The robust performance of Dark-EvGS on both our proposed dataset and another open benchmark again reflects the effectiveness of our methods.

*2) Analysis on Challenges and Why Existing Methods Fail:* Reconstructing real-world radiance fields in low-light and dark environments presents several core challenges that make the task uniquely difficult compared to standard lighting conditions. Below, we analyze the major challenges and discuss why existing methods fail to address them.

*i) Limited Dynamic Range:* Traditional frame-based cameras have a limited dynamic range, making it challenging to capture sufficient details in low-light settings [52], [53]. The captured frames often lack critical scene information, resulting in incomplete inputs for radiance field reconstruction. Although event cameras are capable of capturing unseen details in the dark due to their high dynamic range, directly applying frame-based Gaussian Splatting pipelines such as 3D GS [20] under dark conditions fails, as shown in Fig. 7 (c). Without sufficient intensity and brightness cues, these methods fail to capture fine details and produce degraded radiance fields.

*ii) Increased Noise in Event Data:* While event cameras are advantageous in low-light scenarios, they inevitably suffer from heightened noise levels under extreme dark conditions [54]. The increased randomness and sparsity of the event streams make it harder to extract meaningful information (Fig. 5 (c)), reducing the reliability of existing event-based radiance field approaches [17], [18], [26], [38]. The reason lies in physics: like all vision sensors, event cameras are subject to photon-counting noise [21]. Under low-light conditions, the scarcity of photons causes higher noise and slower event rates, which undermines methods relying purely on events [17], [18] or event-assisted radiance field reconstruction [26], [38]. These approaches cannot effectively handle the noise induced by dark environments, leading to blurry or unstable reconstructions.

*iii) Uncertain Camera Pose:* Accurate pose estimation is another major obstacle in dark environments. On the one hand, structure-from-motion pipelines struggle because dark

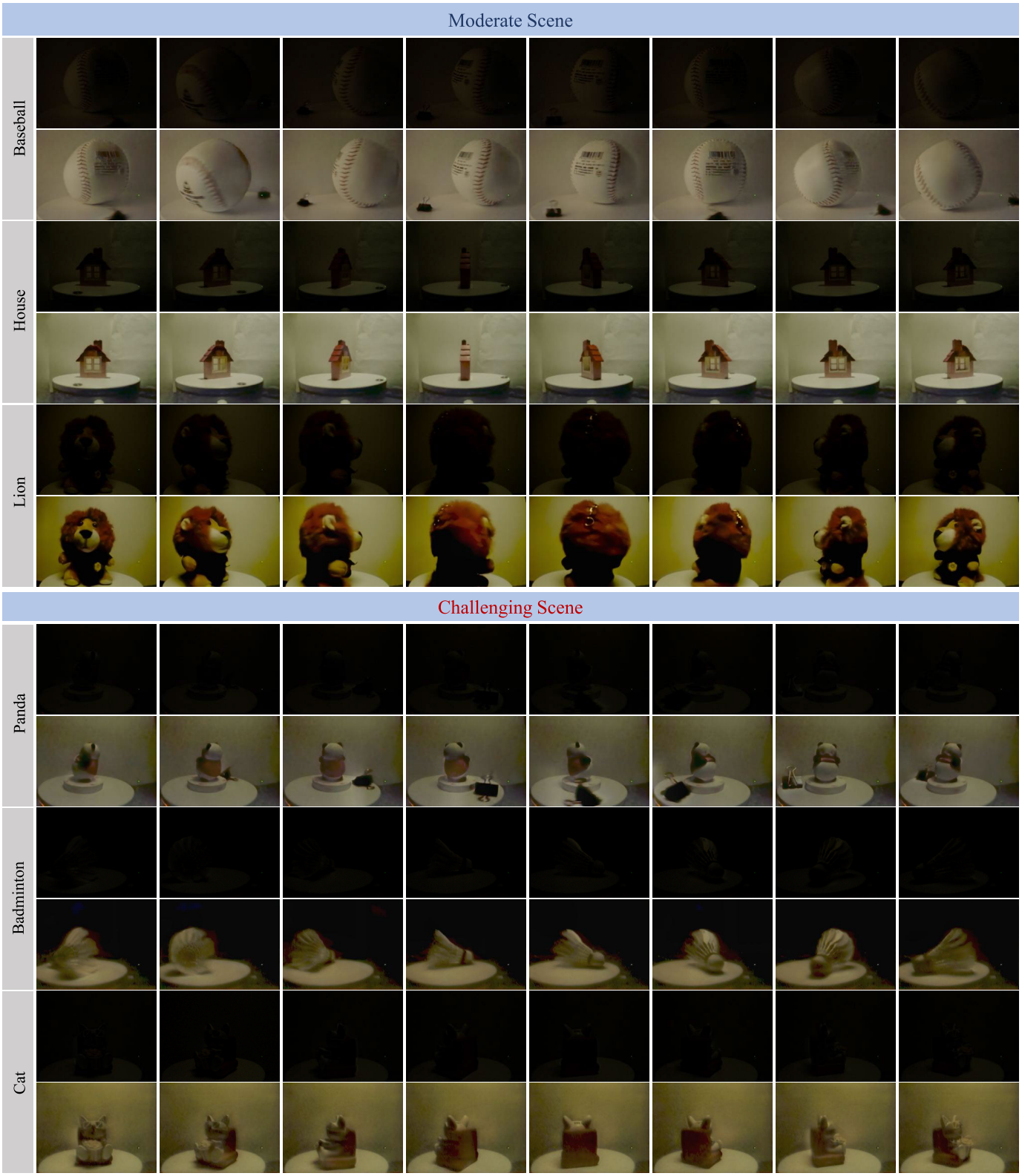


Fig. 10. Visual comparison showcasing the performance of our method across all six real-world scenes ("Baseball", "House", "Lion" as moderate lighting scenes, and "Panda", "Badminton", "Cat" as challenging lighting scenes). Each scene is viewed from multiple perspectives to comprehensively evaluate our method's ability to restore details, structure, and brightness under low-light conditions. In each example, the first row presents the input images captured in dark environments. In contrast, the second row shows the high-quality reconstructions generated by our approach, highlighting significant enhancements in visual clarity and realism.

frames lack sufficient intensity and brightness cues, making it difficult to extract reliable features for geometric matching and tracking. On the other hand, while event cameras provide high temporal resolution, there is currently no mature and robust event-based COLMAP alternative capable of handling noisy and sparse events in such conditions.

In our comparative experiments, for fairness, we supplied all competing methods with ground-truth accurate poses obtained from calibration and known trajectories. However, this setting does not reflect real-world deployment, where such high-quality poses are usually unavailable under dark conditions. In contrast, our pipeline circumvents this difficulty by generating pseudo-bright frames with sufficient structural and photometric cues, enabling COLMAP to recover accurate poses directly without additional steps or specialized tools. Moreover, our dataset contributes beyond benchmarking: it provides precise ground-truth poses alongside real low-light event data, which not only ensures reliable evaluation but also lays the foundation for advancing research toward future *pose-free* dark Gaussian Splatting methods.

*iv) Data Gap Between Synthetic and Real-World Data:* Most existing methods, whether event-based radiance field reconstruction [17], [18], [26], [38] or event-based video enhancement [13], [15], rely heavily on either synthetic datasets or real-world data with normal light. However, synthetic paired data fail to capture the complexities of real-world low-light conditions, such as diverse noise patterns, dynamic lighting changes, and motion blur. For instance, EvLowLight [15] is trained on synthetic paired short- and long-exposure frames, along with simulated events [21], but these data are too clean and lack realistic noise distributions [55]. Consequently, when applied to real-world dark frames, EvLowLight [15] fails to enhance inputs effectively and performs poorly when combined with 3D GS for radiance field reconstruction (Fig. 7, column g). This highlights the necessity of creating real-world datasets that contrast dark and bright scenes. Our dataset is the first to provide such real-world low-light event data with accurate poses, which not only supports the evaluation of Dark-EvGS but also contributes a valuable benchmark for advancing this research field.

*v) Summary:* In summary, low-light radiance field reconstruction suffers from limited dynamic range, increased noise in event data, uncertain camera poses, and a significant gap between synthetic and real-world datasets. These challenges collectively explain why existing methods fail in dark environments. Our Dark-EvGS framework is designed specifically to address these issues, leading to robust reconstruction quality under extreme low-light conditions.

## V. Future Work

For future work, we envision expanding Dark-EvGS for broader real-world applications. One promising direction is reconstructing large-scale scenes at night using handheld devices or vehicles. This extension would require optimizing the framework for dynamic scenarios, integrating robust motion compensation techniques, and enhancing the system's computational efficiency. Another important direction is to address unknown or inaccurate camera poses under dark environments. In such scenarios, with severe noise, motion blur, and a lack of visual features, accurate pose information can be unavailable. Future work could explore jointly optimizing camera pose estimation and radiance field reconstruction, potentially leveraging the high temporal resolution and robustness of event data to mitigate pose drift and misalignment in low-light conditions. This work could pave the way for impactful applications in areas such as autonomous navigation, AR/VR, and nighttime photography by enabling nighttime scene capture and reconstruction in diverse real-world contexts.

## VI. Limitation

While Dark-EvGS demonstrates significant advancements in radiance field reconstruction under low-light conditions, it is not without limitations. A key challenge is that the current framework only supports static radiance field reconstruction as it was built on the 3D GS pipeline, which does not account for temporal variations effectively. For future work, we will explore dynamic radiance field reconstruction in the dark.

## VII. Conclusion

In this work, we introduce Dark-EvGS, the first event-guided 3D Gaussian Splatting framework for bright frame synthesis from arbitrary views in low-light environments. Leveraging the high dynamic range and speed of event cameras, Dark-EvGS overcomes noise, motion blur, and poor illumination challenges faced by conventional methods. We propose a triplet-level supervision strategy to enhance scene understanding, detail refinement, and sharpness restoration, along with a Color Tone Matching Block to ensure color consistency. Additionally, we introduce the first real-world dataset for event-guided bright frame synthesis in low-light settings. Extensive experiments show that Dark-EvGS outperforms prior approaches, establishing it as a robust solution for low-light radiance field reconstruction.

## References

[1] X. Liu et al., "NTIRE 2024 challenge on low light image enhancement: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 6571–6594.

[2] Z. Zhu, J. Hou, J. Li, J. Wu, and J. Hou, "Modeling state shifting via local–global distillation for event-frame gaze tracking," *IEEE Trans. Mobile Comput.*, vol. 24, no. 11, pp. 11614–11627, Nov. 2025.

[3] X. Lyu and J. Hou, "Enhancing low-light light field images with a deep compensation unfolding network," *IEEE Trans. Image Process.*, vol. 33, pp. 4131–4144, 2024.

[4] H. Liu, S. Peng, L. Zhu, Y. Chang, H. Zhou, and L. Yan, "Seeing motion at nighttime with an event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25648–25658.

[5] D. Gehrig and D. Scaramuzza, "Low-latency automotive vision with event cameras," *Nature*, vol. 629, no. 8014, pp. 1034–1040, May 2024.

[6] J. Wu, R. Xu, Z. Wood-Doughty, C. Wang, S. Xu, and E. Y. Lam, "Segment anything model is a good teacher for local feature learning," *IEEE Trans. Image Process.*, vol. 34, pp. 2097–2111, 2025.

[7] Y. Lu et al., "SEE: See everything every time-adaptive brightness adjustment for broad light range images via events," 2025, *arXiv:2502.21120*.

[8] Y. Yang, J. Liang, B. Yu, Y. Chen, J. S. Ren, and B. Shi, "Latency correction for event-guided deblurring and frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 24977–24986.

[9] Z. Ding et al., "Spatio-temporal recurrent networks for event-based optical flow estimation," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 1, pp. 525–533.

[10] J. Chen, Y. Wang, Y. Cao, F. Wu, and Z.-J. Zha, "Progressivemotionseg: Mutually reinforced framework for event-based motion segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 303–311.

[11] C. Ge, X. Fu, K. Wang, and Z.-J. Zha, "Event-based video reconstruction with deep spatial-frequency unfolding network," *IEEE Trans. Image Process.*, vol. 34, pp. 1779–1794, 2025.

[12] Z. Liu, B. Guan, Y. Shang, Q. Yu, and L. Kneip, "Line-based 6-DoF object pose estimation and tracking with an event camera," *IEEE Trans. Image Process.*, vol. 33, pp. 4765–4780, 2024.

[13] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3857–3866.

[14] B. Ercan, O. Eker, C. Saglam, A. Erdem, and E. Erdem, "HyperE2 VID: Improving event-based video reconstruction via hypernetworks," *IEEE Trans. Image Process.*, vol. 33, pp. 1826–1837, 2024.

[15] J. Liang, Y. Yang, B. Li, P. Duan, Y. Xu, and B. Shi, "Coherent event guided low-light video enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jul. 2023, pp. 10615–10625.

[16] G. Liang, K. Chen, H. Li, Y. Lu, and L. Wang, "Towards robust event-guided low-light image enhancement: A large-scale real-world event-image dataset and novel approach," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 23–33.

[17] J. Wu, S. Zhu, C. Wang, and E. Y. Lam, "EV-GS: Event-based Gaussian splatting for efficient and accurate radiance field rendering," in *Proc. IEEE 34th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2024, pp. 1–6.

[18] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik, "EventNeRF: Neural radiance fields from a single colour event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 4992–5002.

[19] T. Xiong et al., "Event3DGS: Event-based 3D Gaussian splatting for high-speed robot egomotion," in *Proc. Conf. Robot Learn.*, 2024, pp. 4100–4118.

[20] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1–139, Aug. 2023.

[21] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic DVS events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1312–1321.

[22] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, Jan. 2022.

[23] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, "NeRF in the dark: High dynamic range view synthesis from noisy raw images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, May 2022, pp. 16190–16199.

[24] T. Zhang, K. Huang, W. Zhi, and M. Johnson-Roberson, "DarkGS: Learning neural illumination and 3D Gaussians relighting for robotic exploration in the dark," 2024, *arXiv:2403.10814*.

[25] J. Huang, C. Dong, X. Chen, and P. Liu, "IncEventGS: Pose-free Gaussian splatting from a single event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 26933–26942.

[26] J. Wu, S. Zhu, C. Wang, B. Shi, and E. Y. Lam, "SweepEvGS: Event-based 3D Gaussian splatting for macro and micro radiance field rendering from a single sweep," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 12, pp. 12734–12746, Dec. 2025.

[27] C. Feng et al., "E-4DGS: High-fidelity dynamic reconstruction from the multi-view event cameras," in *Proc. 33rd ACM Int. Conf. Multimedia*, Oct. 2025, pp. 7356–7365.

[28] Q. Qu, Y. Shen, X. Chen, Y. Y. Chung, and T. Liu, "E2HQV: High-quality video generation from event camera via theory-inspired model-aided deep learning," in *Proc. AAAI Conf. Artif. Intell.*, 2024, pp. 4632–4640.

[29] S. Zhang, Y. Zhang, Z. Jiang, D. Zou, J. Ren, and B. Zhou, "Learning to see in the dark with events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*Switzerland: Springer, Jun. 2020, pp. 1–16.

[30] F. Paredes-Vallés and G. C. H. E. de Croon, "Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3445–3454.

[31] J. Han et al., "Hybrid high dynamic range imaging fusing neuromorphic and conventional images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8553–8565, Jul. 2023.

[32] P. Duan et al., "NeuroZoom: Denoising and super resolving neuromorphic events and spikes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 15219–15232, Dec. 2023.

[33] P. Zhang, H. Liu, Z. Ge, C. Wang, and E. Y. Lam, "Neuromorphic imaging with joint image deblurring and event denoising," *IEEE Trans. Image Process.*, vol. 33, pp. 2318–2333, 2024.

[34] S. Zhu, Z. Ge, C. Wang, J. Han, and E. Y. Lam, "Efficient non-line-of-sight tracking with computational neuromorphic imaging," *Opt. Lett.*, vol. 49, no. 13, pp. 3584–3587, 2024.

[35] S. Klenk, L. Koestler, D. Scaramuzza, and D. Cremers, "E-NeRF: Neural radiance fields from a moving event camera," *IEEE Robot. Autom. Lett.*, vol. 8, no. 3, pp. 1587–1594, Mar. 2023.

[36] W. Yu et al., "EvaGaussians: Event stream assisted Gaussian splatting from blurry images," 2024, *arXiv:2405.20224*.

[37] S. Zahid, V. Rudnev, E. Ilg, and V. Golyanik, "E-3DGS: Event-based novel view rendering of large-scale scenes using 3D Gaussian splatting," 2025, *arXiv:2502.10827*.

[38] H. Deguchi, M. Masuda, T. Nakabayashi, and H. Saito, "E2GS: Event enhanced Gaussian splatting," in *Proc. IEEE Int. Conf. Image Process.*, Aug. 2024, pp. 1676–1682.

[39] W. Yifan, F. Serena, S. Wu, C. Öztireli, and O. Sorkine-Hornung, "Differentiable surface splatting for point-based geometry processing," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–14, Dec. 2019.

[40] M. Zwicker, H. Pfister, J. van Baar, and M. Gross, "Surface splatting," in *Proc. 28th Annu. Conf. Comput. Graph. Interact. Techn.*, Aug. 2001, pp. 371–378.

[41] P. Duan, Z. W. Wang, X. Zhou, Y. Ma, and B. Shi, "EventZoom: Learning to denoise and super resolve neuromorphic events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 12824–12833.

[42] H. Fang, J. Wu, L. Li, J. Hou, W. Dong, and G. Shi, "AEDNet: Asynchronous event denoising with spatial–temporal correlation among irregular data," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1427–1435.

[43] R. Yang et al., "SCI: A spectrum concentrated implicit neural compression for biomedical data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, 2023, pp. 4774–4782.

[44] Y. Feng, H. Lv, H. Liu, Y. Zhang, Y. Xiao, and C. Han, "Event density based denoising method for dynamic vision sensor," *Appl. Sci.*, vol. 10, no. 6, p. 2024, Mar. 2020.

[45] *Multimedia Systems and Equipment-color Measurement and Management–Part 2–1*, I. E. Commission, Brussels, Belgium, 1999.

[46] H. Han, J. Li, H. Wei, and X. Ji, "Event-3DGS: Event-based 3D reconstruction using 3D Gaussian splatting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 37, 2025, pp. 128139–128159.

[47] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[48] Z. Zhu, J. Hou, H. Liu, H. Zeng, and J. Hou, "Learning efficient and effective trajectories for differential equation-based image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 10, pp. 9150–9168, Oct. 2025.

[49] K. Chen, G. Liang, Y. Lu, H. Li, and L. Wang, "EvLight++: Low-light video enhancement with an event camera: A large-scale real-world dataset, novel method, and more," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 48, no. 2, pp. 1608–1625, Feb. 2026.

[50] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4104–4113.

[51] J. Hidalgo-Carrió, G. Gallego, and D. Scaramuzza, "Event-aided direct sparse odometry," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5771–5780.

[52] Z. Wan, Y. Dai, and Y. Mao, "Learning dense and continuous optical flow from an event camera," *IEEE Trans. Image Process.*, vol. 31, pp. 7237–7251, 2022.

[53] Q. Zhan, G. Liu, X. Xie, R. Tao, M. Zhang, and H. Tang, "Spiking transfer learning from RGB image to neuromorphic event stream," *IEEE Trans. Image Process.*, vol. 33, pp. 4274–4287, 2024.

[54] P. Duan et al., "EventAid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 8, pp. 6959–6973, Aug. 2025.

[55] T. Stoffregen et al., "Reducing the sim-to-real gap for event cameras," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 534–549.