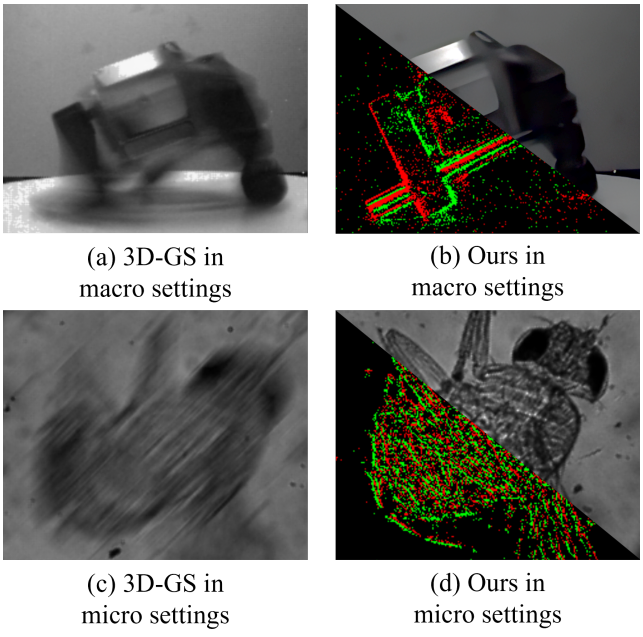
SweepEvGS: Event-Based 3D Gaussian Splatting for Macro and Micro Radiance Field Rendering from a Single Sweep

Jingqian Wu, Shuo Zhu, Chutian Wang, Boxin Shi, *Senior Member, IEEE* and Edmund Y. Lam, *Fellow, IEEE*

***Abstract*—Recent advancements in 3D Gaussian Splatting (3D-GS) have demonstrated the potential of using 3D Gaussian primitives for high-speed, high-fidelity, and cost-efficient novel view synthesis from continuously calibrated input views. However, conventional methods require high-frame-rate dense and high-quality sharp images, which are time-consuming and inefficient to capture, especially in dynamic environments. Event cameras, with their high temporal resolution and ability to capture asynchronous brightness changes, offer a promising alternative for more reliable scene reconstruction without motion blur. In this paper, we propose SweepEvGS, a novel hardware-integrated method that leverages event cameras for robust and accurate novel view synthesis across various imaging settings from a single sweep. SweepEvGS utilizes the initial static frame with dense event streams captured during a single camera sweep to effectively reconstruct detailed scene views. We also introduce different real-world hardware imaging systems for real-world data collection and evaluation for future research. We validate the robustness and efficiency of SweepEvGS through experiments in three different imaging settings: synthetic objects, real-world macro-level, and real-world micro-level view synthesis. Our results demonstrate that SweepEvGS surpasses existing methods in visual rendering quality, rendering speed, and computational efficiency, highlighting its potential for dynamic practical applications.**



(a) 3D-GS in macro settings (b) Ours in macro settings (c) 3D-GS in micro settings (d) Ours in micro settings

Fig. 1. Frame-based devices capture experience blur in real-world macro imaging settings when undergoing fast camera/object movement (a). Blur also occurs in real-world microimaging settings such as microscopy when undergoing drastic object displacement (b). Our SweepEvGS is able to render sharp and accurate detailed radiance field representation across various real-world hardware imaging settings (c) and (d).

I. Introduction

Novel View Synthesis with densely multi-view images of a given 3D scene generates new and realistic 2D viewpoints [1], [2]. Recently, 3D Gaussian Splatting (3D-GS) [3] has introduced an unstructured 3D Gaussian radiance field that uses 3D Gaussian primitives to achieve impressive success in fast, high-quality, and cost-effective novel view synthesis from densely input views [4], [5]. However, accurate reconstruction of a radiance field typically requires capturing dense and high-quality training frames. While using videos as input can be an efficient way, blur may occur, producing poor ground truth, shown in Figure 1(a) and (c), and inaccurate camera calibration [6], [7]. Some methods [3], [8]–[11] involve recording videos at a slow pace to minimize motion blur, ensuring that each frame is clear and usable for reconstruction. However, this approach has significant drawbacks. The process is not only slow and cumbersome but also prone to inefficiencies, particularly in dynamic environments where rapid movements are involved. Therefore, efficiently and accurately obtaining such training information has become a vital topic.

Event cameras, on the other hand, are a new imaging sensor that asynchronously captures per-pixel brightness changes [12], offering advanced properties and representing a major paradigm shift from conventional frame-based cameras across various vision tasks [13]–[20]. The nature of event cameras makes them highly efficient for fast and dynamic scene capturing. The captured event streams at the microsecond level without motion blur are highly effective for dense radiance field reconstruction.

These gifts from event cameras make, in theory, good imaging devices for fast imaging and capturing. However, using event cameras to enable robust and accurate real-world radiance field rendering in different imaging settings still faces several technical challenges. First, as there are no mature event-based COLMAP [6] to generate accurate camera pose, coordinate, and point cloud, designing hardware systems that fit different imaging settings while acquiring precise camera calibration for 3D-GS and corresponding clear ground truth frames for quantitative metric calculation is not trivial. Second,

Jingqian Wu, Shuo Zhu, Chutian Wang and Edmund Y. Lam are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China (e-mail: jingqianwu@connect.hku.hk, zhushuo@hku.hk, ctwang@eee.hku.hk, elam@eee.hku.hk).

Boxin Shi is with the State Key Laboratory for Multimedia Information Processing and National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100871, China (e-mail: shiboxin@pku.edu.cn)

Corresponding author: Edmund Y. Lam.

utilizing real-world data and formulating them into accurate supervisory signals across different imaging settings to reconstruct accurate radiance fields is also challenging. This is because object scopes and noise levels differed from each imaging setting, making it harder to unify across all of them. Third, designing an event-based 3D-GS algorithm for efficient and effective radiance field rendering in these different imaging settings is vital, because such a unified algorithm would benefit brother applications. To address these challenges, we propose SweepEvGS: a novel hardware-integrated and event-based 3D-GS approach that leverages the capabilities of event cameras for efficient and accurate novel view synthesis across three different imaging settings with just a single sweep of the camera. By using only the first static frame captured and combining it with the dense stream of events generated during the camera sweep, we employ 3D-GS to reconstruct frames of the swept view efficiently and accurately. We summarize our technical contributions as follows:

- We introduce SweepEvGS, marking the first hardware-integrated approach that enables efficient radiance field rendering from a monocular event camera with only a single camera sweep across different imaging settings.
- We design a unified and generalized multi-modal data utilization, and robust supervision for 3D-GS pipeline that enables accurate radiance field reconstitution in different imaging settings.
- We propose different hardware imaging systems designed for various settings of real-world radiance field rendering, including detailed steps for data collection and evaluation for future research.

To demonstrate the robustness of our proposed method, we conducted experiments in three different imaging settings and scales: (1) synthetic objects view synthesis, (2) real-world macro-level view synthesis, for example, Figure 1(b), and (3) real-world microscopic-level view synthesis, for example, Figure 1(d). Extensive experiments in both simulated and different real-world settings validate that our method surpasses the visual rendering quality of existing pipeline approaches both quantitatively and qualitatively. Additionally, it achieves around 150 times faster rendering speeds and requires about 3 times fewer computational resources than the previous NeRF-based approach. These results demonstrate the robustness and efficiency of our approach in both synthetic and real-world scenarios, highlighting its potential for practical applications in dynamic environments.

II. Related Work

A. Radiance Field Rendering

Radiance field rendering and novel-view synthesis are fundamental tasks in graphics and computer vision, with significant applications in robotics, virtual reality, and other areas. Neural Radiance Fields (NeRF) [21] and its variants [22], [23] have significantly advanced these tasks by implicitly modeling scenes with MLP-based neural networks and utilizing differentiable volume rendering. This approach achieves state-of-the-art rendering results with high fidelity and fine details. However, the necessity to sample a large number of points to accumulate the color of each pixel results in low rendering efficiency and prolonged training times.

To address these challenges, recent research has explored alternative 3D representations for better efficiency and visual fidelity. 3D-GS employs a set of optimized Gaussian splats to achieve state-of-the-art reconstruction quality and rendering speed. Initialized from sparse Structure-from-Motion (SfM) point clouds, 3D-GS is trained via differentiable rendering to adaptively control the density and refine the shape and appearance parameters. However, accurate radiance field reconstruction usually demands dense, high-quality training frames, which are time-consuming and inefficient to capture. Reducing view constraints can lead to incorrect geometry learning, resulting in poor novel view synthesis and rendering quality. Some methods [3] record videos slowly to minimize motion blur, ensuring clear frames for reconstruction. Yet, this approach is slow, cumbersome, and inefficient, especially in dynamic environments with rapid movements. Therefore, it is vital to seek solutions that overcome these challenges and synthesize novel views with simple imaging processes across different settings.

B. Neuromorphic Radiance Field Rendering

The distinctive features of event cameras, including their ability to prevent motion blur, provide a high dynamic range, ensure low latency, and consume minimal power, have spurred their adoption in computer vision and computational imaging [24]–[28]. Several approaches have been proposed to address the view synthesis challenge using NeRF [21] with event data [29], [30], leveraging volumetric rendering with either pure event or semi-event (blurred RGB involved) supervision. These methods utilize the inherent multi-view consistency of NeRFs, providing a robust self-supervision signal for extracting coherent scene structures from raw event data. However, a significant drawback is the time-consuming optimization of NeRF. The computational demands of training and optimizing an event NeRF pipeline, in terms of both training time and GPU memory, are approximately 80 times greater than those of the 3D-GS pipeline. Additionally, the use of high-dimensional multilayer perceptron networks in the NeRF architecture results in a view-rendering speed that is around 190 times slower, which limits its applicability for real-time rendering.

A few recent works have attempted to overcome these efficiency issues by combining 3D-GS with event data. [31] proposed a collaborative learning framework for event depth estimation, event-to-frame reconstruction, and event-based 3D-GS. However, this approach requires prior features and outputs from previous depth estimation and frame reconstruction tasks. The entire training framework also necessitates ground truth depth maps and frame references, limiting its usage, especially in scenarios where only event streams are available. [32], [33] proposed a pure event-based 3D-GS end-to-end method for novel view synthesis. Nonetheless, the rendering quality was not satisfactory as the work was mainly designed for synthetic event data, and the method did not explore different real-world hardware-involved imaging setups.

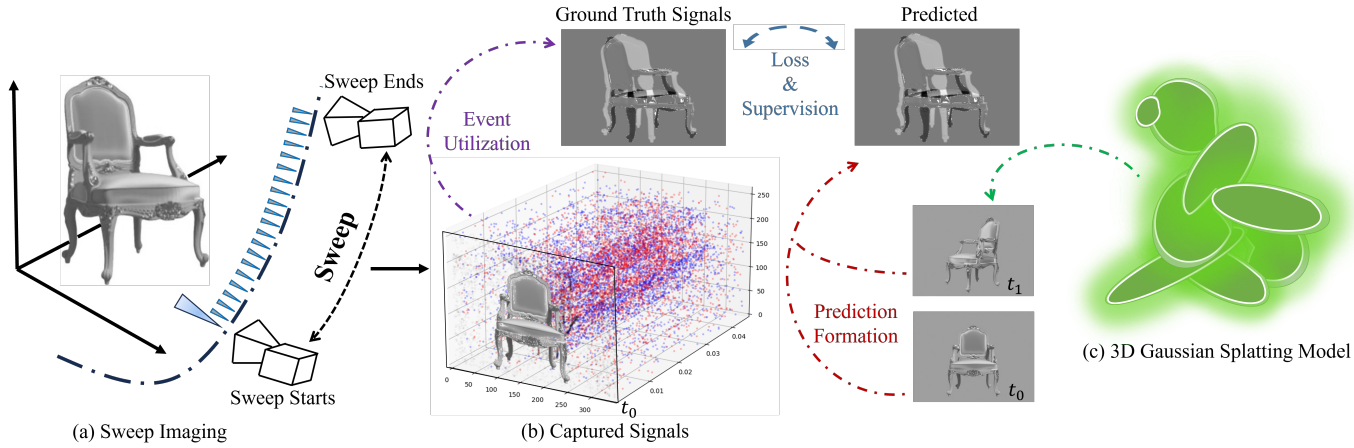


Fig. 2. Overview of the SweepEvGS pipeline for novel view synthesis. The process begins with capturing a static object or scene using an event camera (a). The camera captures signals (b) including the initial frame before the sweep starts when static, and the dense event streams representing intensity changes during the camera sweep. On the one hand, these signals are utilized to form (purple arrow) the ground truth signals for training and supervision, as described in Sec. III-B. On the other hand, the 3D-GS model, described in Sec. III-A, renders (green arrow) when given the corresponding camera position and pose, and formalizes (red arrow) the rendered results to predictions. The designed loss supervises the training pipeline, as described in Sec. III-C. Our method efficiently bridges the gap between sparse event data and dense scene reconstruction, enabling high-quality novel view synthesis.

C. Radiance Field Rendering with Real-World Data

Motivated by the potential of radiance field rendering and novel view synthesis, extensive work has been conducted on real-world datasets. NeRF has been widely applied to real-world object-level and scene-level rendering [21], [34], [35]. Additionally, NeRF has been utilized in microscopy imaging settings [36]. Although 3D-GS has not yet been applied to microscopy and other optical settings, it has been successfully used for real-world object and scene rendering [3], [37], [38]. Currently, there has been no research exploring the application of real-world event data with 3D-GS across various real-world imaging systems. To address this gap, our work examines the use of 3D-GS with event data in three different levels of imaging setups. We conduct both quantitative and qualitative analyses, and we provide detailed information about our different hardware setups for real-world data collection. we hope to provide valuable insights for future research in this area.

III. Methodology

A. Preliminary on 3D Gaussian Splatting

3D-GS [3] represents detailed 3D scenes using point clouds, where Gaussians are used to define the structure of the scene. Each Gaussian is characterized by a 3D covariance matrix Σ, and a central point $\mathbf{x}$:

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

where the central point $\mathbf{x}$ is referred to as the mean value of the Gaussian. To facilitate differentiable optimization, the covariance matrix Σ can be decomposed into a rotation matrix R and a scaling matrix S:

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

To generate renderings from different viewpoints, the splatting technique described in [39] is used to project the Gaussians onto the camera planes. This method, initially introduced in [40], involves a viewing transformation matrix W and the Jacobian J of the affine approximation of the projective transformation. Using these, the covariance matrix Σ' in camera coordinates is given by:

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

To summarize, each Gaussian point in the model is defined by several attributes, i.e., its position $\mathbf{x} \in \mathbb{R}^3$, color represented by spherical harmonics coefficients $\mathbf{c} \in \mathbb{R}^k$ (where k denotes the degrees of freedom), opacity $\alpha \in \mathbb{R}$, a rotation quaternion $\mathbf{q} \in \mathbb{R}^4$, and a scaling factor $\mathbf{s} \in \mathbb{R}^3$. For each pixel, the color and opacity of all Gaussians are computed based on the Gaussian representation outlined in Equation 1. The blending process for the color C of N-ordered points overlapping a pixel follows this formula:

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

where $\mathbf{c}_i$ and α_i denote the color and density of a specific point, respectively. These values are influenced by a Gaussian with a covariance matrix Σ, which is subsequently adjusted by per-point opacity and spherical harmonics color coefficients.

B. Noisy Events as 3D-GS Supervisory Signals

Each event e_k is described as a tuple $(\mathbf{x}_k, t_k, p_k)$, which occurs asynchronously at pixel $\mathbf{x}_k = (x_k, y_k)$ at a microsecond timestamp t_k. The polarity $p_k \in \{-1, +1\}$ denotes either an increase or decrease in the logarithmic brightness $L(\mathbf{x}_k, t_k)$ by the contrast threshold A. Specifically, an event at time t_k is triggered if the following condition is met:

$$\Delta L \triangleq L(\mathbf{x}_k, t_k) - L(\mathbf{x}_k, t_{k-1}) = p_k A, \tag{5}$$

where t_{k-1} is the timestamp of the previous event at pixel $\mathbf{x}_k$. As we utilize event stream in a unifying way across all imaging settings and microscopic settings tend to introduce more noise [41], we apply the y-noise filter from [42] for preprocessing noisy event. Noise filtering is vital for accurate view reconstruction [43], as illustrated in the supplementary material. As illustrated in Figure 2, our goal is to render a radiance field representation using differentiable 3D Gaussian functions under pure event signal supervision with only the starting frame for camera calibration, without involving other frame-based data. To achieve this, we need to convert the ground truth event data into a differentiable supervisory signal and train a 3D-GS model to render this representation. Our algorithm generates a rendering result at t_k with camera pose p_k the timestamp where the camera is currently at. With the known initial frame at timestamp t_0 with camera pose p_0, meaning before the camera begins to sweep, the supervisory signal is formed by accumulating ground truth event signal between those two timestamps.

For each timestamp t_k, we calculate two associated rendering results from the 3D Gaussian model: $I_k = G(p_k)$ and $I_0 = G(c_k)$, where I_k and I_0 are the rendered grayscale frames at times t_0 and t_k, respectively, G is the 3D-GS model, and p_k and p_0 are the camera poses at times t_0 and t_k. We represent the logarithmic image as $L(I_t) = \log\big((I_t)^g + \epsilon\big)$, where $\epsilon = 1 \times 10^{-5}$ (to avoid NaN values), and g denotes a fixed gamma correction value set to 2.2 across all experiments. This conforms to the grayscale gamma curve, with Gamma 2.2 serving as the recommended smooth approximation [30], [44]. Consequently, we derive the predicted accumulative difference $E_{pred} = L(I_0) - L(I_k)$.

To utilize event data as a supervisory signal for E_{pred}, following [45], we aggregate the polarities of all events occurring between the selected times t and $t - w$ based on their positional information $\mathbf{x}_k$, such that

$$E_{gt} = \int_t^{t+w} e_k(\mathbf{x}_k, p_k, t_k)\, dt, \tag{6}$$

where E_{gt} is the aggregated result.

C. Supervising Events to Radiance Fields

Pure events as input give poor real-world radiance field rendering results [32]. Since events represent the intensity changes from previous timestamps, the clear and shaped initial frame captured by the static camera before the sweep is valuable knowledge of where the events changed from. Given a camera pose at a specific time t, the 3D-GS model learns and renders the corresponding frame I_t with supervision. Because the first frame I_0 is known for camera calibration, our task is to formulate the accumulated difference E_{pred} between I_t, I_0, and supervise it with the ground truth event E_{gt}. To effectively supervise E_{pred} from E_{gt}, we follow the methods described in [29], [32], [46], applying the linlog mapping to derive the predicted logarithmic brightness difference for both E_{pred} and E_{gt}. Using a normalized $\mathcal{L}_2$ loss, the loss $\mathcal{L}_{event}$ is calculated as:

$$\mathcal{L}_{event}(x, y) = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\big(\|L(x)\|_2^2 - \|L(y)\|_2^2\big)^2}{H \times W}, \tag{7}$$

where for an arbitrary u:

$$L(u) = \text{linlog}\big(I(u)\big) \triangleq \begin{cases} I(u) \times \ln(B)/B & \text{if } I(u) < B, \\ \ln\big(I(u)\big) & \text{otherwise,} \end{cases} \tag{8}$$

where the threshold B delineates the linear region where no logarithmic mapping is applied. The value of 20 is used for B in all our experiments. Here, x and y correspond to E_{pred} and E_{gt}, respectively.

Additionally, we retained the D-SSIM loss used in the original 3D-GS approach, as we found that a small coefficient improves the rendering result. Following previous works, we calculate the D-SSIM loss as:

$$\mathcal{L}_{D-SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{9}$$

where μ is the pixel sample mean, σ_x is the variance of x, σ_{xy} is the covariance of x and y, $c_1 = (k_1 q)^2$, and $c_2 = (k_2 q)^2$ are constants to stabilize the division with a weak denominator. q is the window range of the pixel values, and k_1 and k_2 are constants. Therefore, the loss function is derived as follows:

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{event}(E_{pred}, E_{gt}) + \\ \lambda(1 - \mathcal{L}_{D-SSIM}(E_{pred}, E_{gt})), \end{aligned} \tag{10}$$

where λ is set to 0.1 for all our experiments.

Our supervision method is designed to be robust across various types of data, including synthetic, real-world macro, and real-world microimaging data. Introducing the y-noise filter eliminates interference signals, especially for microimaging data collected under the microscope. Incorporating normalization techniques ensures the loss remains stable across different imaging conditions. This stability is crucial for maintaining consistent performance and achieving high-quality rendering results, regardless of the data source. More ablation studies in the supplementary materials support our claim.

IV. Experiments

A. Overview

This subsection provides an overview and structural layout of our experiments. Here, we briefly introduce the experiment setups. Experiments are conducted in three different imaging settings to demonstrate the robustness of our approach: 1) real-world macro-level sequences, 2) real-world microscopic-level sequences, and 3) synthetic sequences.

All real-world data were captured using the DAVIS 346C event camera. For real-world macro sequences, we introduce two types of imaging systems for different scenarios. For objects that involve depth information (such as a Lego toy

and an optical component), we fixed the event camera, and "swept" the objects using a fast turntable with a known constant speed, shown in Figure 3(a), to generate accurate camera poses and coordinates. For objects that are plain (such as a keyboard and a cloth), we mounted the event camera on a horizontal translation stage, shown in Figure 3(b), allowing linear movement at a constant speed to generate accurate camera poses and coordinates. For quantitative comparison, we captured two sets of data for each sequence: (1) a fast sweep (high-speed mode) to capture the event stream and RGB frames generated by the DAVIS 346C camera, and (2) a slow sweep (low-speed mode) to capture clear and sharp RGB frames used as ground truth references, since the fast sweep results in blurry RGB frames unsuitable for quantitative comparison. We aligned the time correspondence between these datasets to accurately calculate quantitative metrics.

For the real-world microscopic-level imaging setup, we deployed the DAVIS 346C camera on a microscope station, using the station's mobility control to "sweep" the object horizontally or vertically. Due to manual control of the object's movement, we collect the event stream and RGB frames under a fast sweep and clear and sharp RGB frames under a slow sweep with a static frame capturing process to collect the RGB ground truth references. We then manually paired the fast-slow data as in the other setups. Thus, we could also compare the visual rendering quality between the rendered result and a reference image of the object taken statically under quantitative metrics. Figure 3 shows a photograph of the setup, and detailed hardware information will be presented in Sec. IV-B.

For the synthetic sequences, we used 3D models from [21] to simulate the sweep process of a camera in a real-world setting. Each scene was captured by sweeping the camera 90 degrees around the object, resulting in 250 RGB images, which were then converted to grayscale. An event stream was generated from these images using the model from [47], and we applied the corresponding camera intrinsic and extrinsic in our approach.

The rendering results of different approaches are evaluated using the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM). PSNR assesses signal fidelity by comparing the maximum signal power to noise power, while SSIM evaluates image similarity based on luminance, contrast, and structure. These metrics provide quantitative insights into the quality and fidelity of the rendered images.

While the above provides an overview of the experiment setup, detailed hardware information and implementation details of the algorithm will be presented in Sec. IV-B and Sec. IV-C. The results of the three types of experiments mentioned will be presented in Sec. IV-D, Sec. IV-E, and Sec. IV-F. We also conduct an ablation study in Subsection IV-G to demonstrate the necessity and effectiveness of our method design. Last, qualitative analysis of the methods and results will be presented in Sec. IV-H.

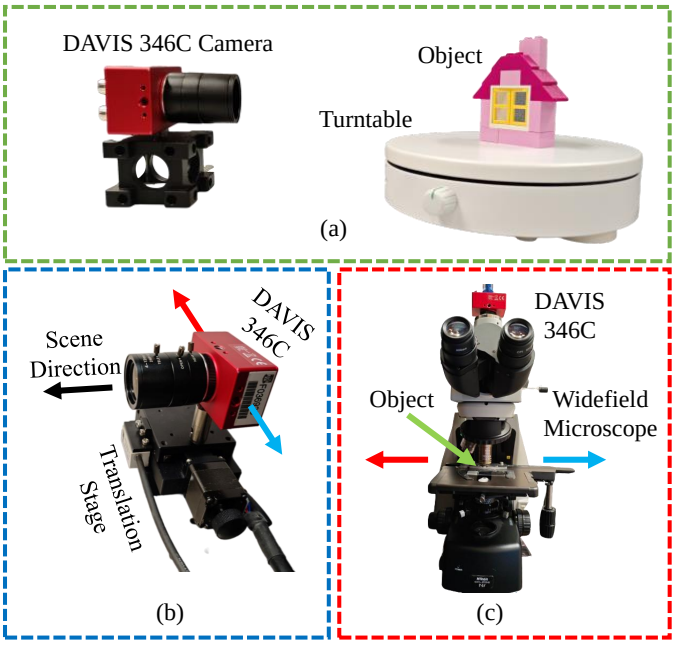


Fig. 3. Demonstration of Hardware Setup. (a) Real-world macro camera system setup for objects with depth: the DAVIS 346C event camera is fixed, and the object is placed on a fast turntable with known constant speed. The camera will capture events when there are changes in the view angle from the object. (b) Real-world macro camera system setup for plain objects: the DAVIS 346C event camera is placed on a translation station that enables horizontal linear movements in both red and blue directions. (c) Real-world microscopic camera system setup: the DAVIS 346C event camera is placed on a widefield microscope that enables horizontal linear movements in both red and blue directions. The camera will capture events and frames from the object placed on the observation plate.

B. Hardware Detailed Information

This subsection presents the detailed hardware information used in the real-world data collection phase to contribute to future research. A video is included as supplementary material to demonstrate the hardware setups used in this paper.

- **Event Camera**: Shown in Figure 3 (a), in all real-world experiments, we use DAVIS 346 Colorful event camera, a state-of-the-art imaging device designed to capture event stream and frame data. With a frame resolution of 346 pixels by 260 pixels, this camera offers a detailed view of the environment. One of the key features of the DAVIS 346 C is that it can react to events in the scene with microsecond precision. This makes it particularly effective for capturing fast-moving objects and dynamic events that would typically be blurred or missed by conventional cameras.
- **Translation Stage**: Shown in Figure 3 (b), in real-world macro experiments, we use WNMC400 motion controller to provide precise camera trajectory control. The WNMC400 motion controller is a compact and high-performance device designed for precise control of industrial automation systems. It offers robust motion control capabilities with features such as trajectory planning, electronic gearing, and synchronization, making it ideal for applications requiring accurate and coordinated motion. With its unique and precise control, we set the

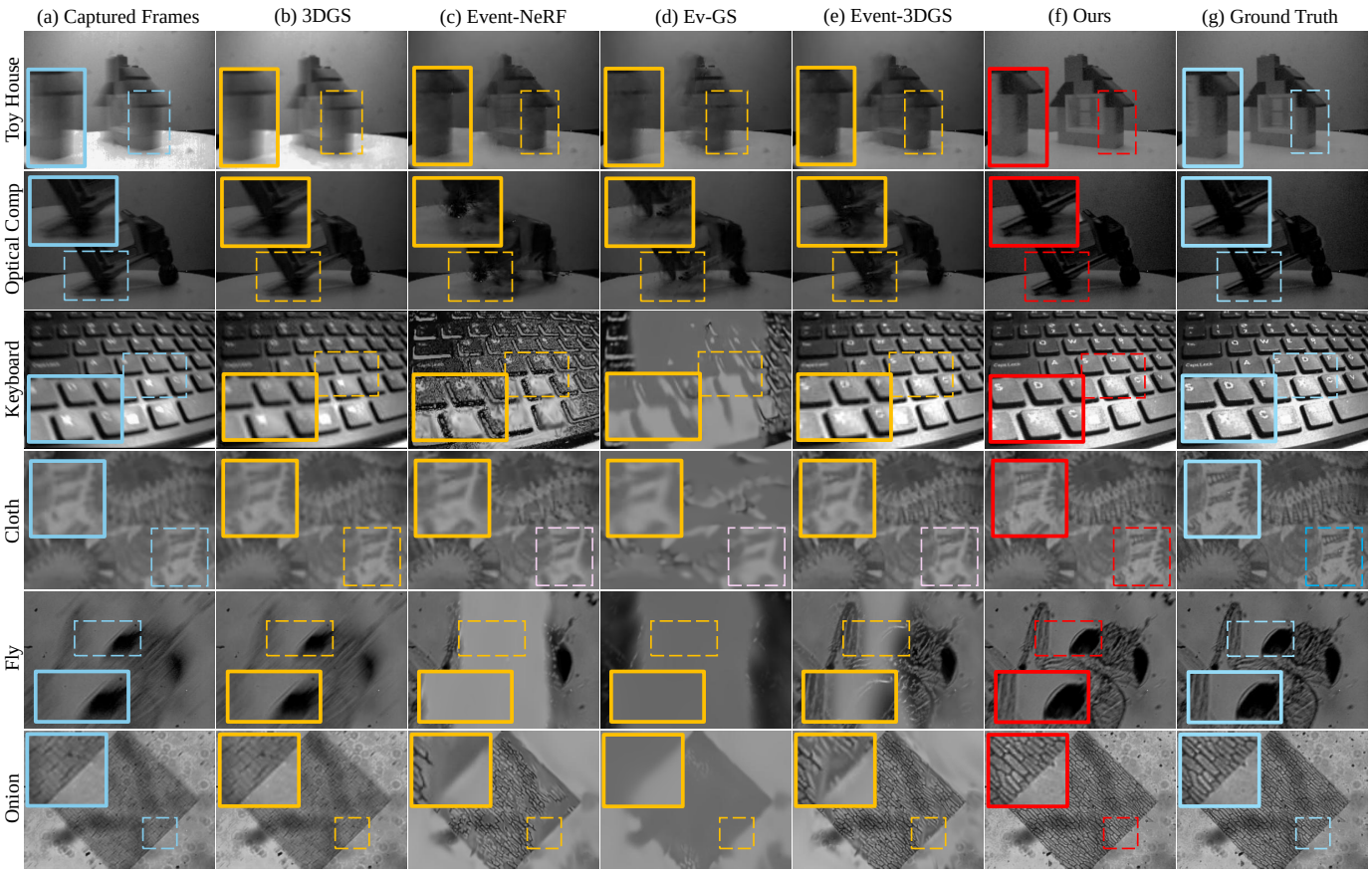


Fig. 4. A visual comparison of our approach against others on real-world datasets: the captured frames (a) from traditional frame-based camera results in motion blur; using traditional frame-based radiance field approach, such as 3D-GS (b) produce blurred radiance field reconstruction and view rendering; event-based NeRF approach (c), time and resource consuming while still produce accurate rendering results; further, existing event-based GS approaches (d and e) failed on real captured data as more noise and randomness are introduced in event streams; our approach (f), leveraging the proposed multi-modal data utilization and supervision, renders sharp and accurate results compared to the ground truth references (g). We visually compare the results across all methods on real-world macro sequences (toy, optical component, keyboard, and cloth) captured by hardware in Figure 3(a) and Figure 3(b); and also on real-world microscopy sequences captured by hardware in Figure 3(c). Dotted boxes indicate areas of interest, and the solid line boxes reveal zoomed results.

speed of the fast motion (input only) to 8 times faster than the speed of the slow motion (ground truth only). The known trajectory and speed relationship makes it easy for us to generate precise global coordinates and frame correspondence needed for metric computation.

- **Microscope**: Shown in Figure 3 (c), in real-world microscopic experiments, we use Nikon Model ECLIPSE Ni-U Microscopy equipment to observe microscopic data. The Nikon Model ECLIPSE Ni-U is an advanced research-grade microscope known for its exceptional optical performance and ergonomic design. In the context of microscopy, minor shifts in the specimen or the camera can lead to considerable displacement. Such movements can render frame-based cameras ineffective, producing blurred images. Consequently, this constraint impedes conventional radiance field rendering techniques from accurately reconstructing clear and detailed textural information within microscopic environments. We manually control the movement of objects using the microscope with linear movement, making it easy to generate global coordinates. However, we did not find a way to provide precise control of the speed like the macro setup, therefore, we manually aligned the fast sweep frame and the slow sweep ground truth frame to compute quantitative metrics.

C. Implementation Details

Our implementation is based on the 3D-GS codebase [3], utilizing its core structure and functionalities. The project was built using Python and CUDA toolkit [48]. Since 3D-GS requires a point cloud as input, we initialize 10^6 points randomly to form the initial point cloud, as the structure-from-motion initialization [6] used in the original model is not suitable for event data. All experiments were performed on an NVIDIA RTX 3090 GPU.

We adhered to specific hyperparameter settings for performance and optimization. The total number of training iterations was set to 50,000. For position optimization, the initial and final learning rates were 1.6×10^{-4} and 1.6×10^{-6}, respectively. The learning rates for feature, opacity, scaling, and rotation optimization were 2.5×10^{-3}, 5×10^{-2}, 5×10^{-3}, and 1×10^{-3}, respectively. Densification parameters included a percent density of 0.01, a densification interval of 100 iterations, and densification occurring between iterations 500

and $50,000$, triggered by a gradient threshold of 2×10^{-4}. The opacity was reset every $3,000$ iterations. These settings were chosen from the default 3D-GS settings to balance training stability and performance.

D. Real-World Macro Sequences

TABLE I
QUANTITATIVE COMPARISON ON REAL-WORLD SEQUENCES. THE TOP TWO RESULTS ARE MARKED WITH **BOLD** (1ST) AND UNDERLINE (2ND).

Metric	PSNR↑	SSIM↑	Training Time↓	FPS↑	Memory↓
Scene			Toy House		
3D-GS	23.8	0.80	**7min**	**80.2**	**3GB**
EventNeRF	24.8	0.88	14h	0.30	15GB
Ev-GS	19.2	0.68	10min	59.7	5GB
Event-3DGS	26.7	0.85	15min	55.2	7GB
Ours	**28.7**	**0.90**	9min	62.7	5GB
Scene			Optical Component		
3D-GS	20.1	0.83	**7min**	**80.9**	**3GB**
EventNeRF	24.1	0.81	14h	0.33	15GB
Ev-GS	20.2	0.70	10min	55.3	5GB
Event-3DGS	25.8	0.82	15min	53.0	7GB
Ours	**26.3**	**0.87**	9min	63.0	5GB
Scene			Keyboard		
3D-GS	25.3	0.86	**7min**	**81.3**	**3GB**
EventNeRF	25.0	0.84	14h	0.32	15GB
Ev-GS	13.6	0.60	10min	61.2	5GB
Event-3DGS	25.2	0.79	15min	57.0	7GB
Ours	**30.5**	**0.92**	9min	62.3	5GB
Scene			Cloth		
3D-GS	24.1	0.81	**7min**	**81.8**	**3GB**
EventNeRF	25.5	0.86	14h	0.32	15GB
Ev-GS	11.6	0.58	10min	55.9	5GB
Event-3DGS	26.4	0.77	15min	53.2	7GB
Ours	**31.1**	**0.94**	9min	61.3	5GB

In this experiment, we demonstrate the efficacy of our rendering method compared to others in real-world macro sequences. Due to the fast sweep during the imaging and data collection process, the captured frames are blurred. Training with blurred and low-quality frames, Traditional 3D-GS struggles, leading to blurred and indistinct renderings, as shown in part (b) in Figure 4. Existing event-based approaches such as EventNeRF [30], Ev-GS [32], and Event-3DGS [33], on the other hand, are mainly designed for synthetic event data. Therefore, when dealing with real-world event streams that are noisy and with more randomness, they failed to produce high-quality radiance fields and rendering results, shown in parts (c, d, and e) in Figure 4. Our method, however, utilizes the captured event streams and the first static frame at the initial position of the camera. Event streams are, in nature, less sensitive to lighting changes. Therefore, with our designed approach, we can produce sharp and clear rendering results, as shown in sub-figure (f) in Figure 4. Our method shows a sharp rendering result with clear details and semantics. The quantitative comparison of rendering results is shown in Table I. Our method addresses the limitations of traditional 3D-GS and challenges faced by existing event-based radiance field approaches. The results demonstrate that our approach not only preserves but enhances the clarity and detail of the rendered images, providing a robust solution for real-world applications.

E. Real-World Microscopic Sequences

In microscopy settings, as demonstrated in Figure 3 part (b), even small movements of the scene or camera can cause significant displacements, resulting in blurry imaging when using frame-based cameras. This limitation often hampers traditional radiance field rendering methods from reconstructing sharp and meaningful textural information in microscopy settings. In contrast, our approach leverages event streams for radiance field rendering, making it particularly suitable for such challenging microscopy settings. The quantitative comparison results are shown in Table II, and our approach renders high-quality reconstruction results with decent PSNR, SSIM scores, while keeping a balanced efficiency in training and rendering. As shown in the right part of Figure 4, we present visual comparisons between our approach, Frame-Based 3D-GS, and existing event-based radiance field approaches.

We provide a visualization of different view angles from different samples across different imaging setups in Figure 5. To demonstrate the capability of novel view synthesis, we randomly add a small displacement on the input coordinate (horizontal and vertical directions) to simulate unseen camera directions and render the unseen view based on these unseen coordinates.

TABLE II
QUANTITATIVE COMPARISON ON REAL-WORLD MICROSCOPY SEQUENCES. THE TOP TWO RESULTS ARE MARKED WITH **BOLD** (1ST) AND UNDERLINE (2ND).

Metric	PSNR↑	SSIM↑	Training Time↓	FPS↑	Memory↓
Scene			Onion		
3D-GS	23.8	0.80	**7min**	**80.2**	**3GB**
EventNeRF	24.8	0.88	14h	0.30	15GB
Ev-GS	8.6	0.55	10min	62.0	5GB
Event-3DGS	18.9	0.75	15min	57.0	7GB
Ours	**28.7**	**0.90**	9min	62.7	5GB
Scene			Fly		
3D-GS	20.1	0.83	**7min**	**80.9**	**3GB**
EventNeRF	24.1	0.81	14h	0.33	15GB
Ev-GS	9.8	0.55	10min	60.0	5GB
Event-3DGS	24.1	0.82	15min	55.0	7GB
Ours	**26.3**	**0.87**	9min	63.0	5GB

F. Synthetic Sequences

As previously mentioned, we evaluated synthetic sequences sourced from [21], which include various scenes such as chairs, hot dogs, ficus, and mic. Table III presents a quantitative comparison with EventNeRF [30] and Ev-GS [32] [1]. Our method demonstrates significant advantages over EventNeRF across these scenes. Specifically, in terms of PSNR and SSIM, as shown in Table III, our approach consistently achieves higher values, indicating superior reconstruction fidelity and better structural similarity between the reconstructed and ground truth images. Notably, our method also boasts significantly reduced training times, requiring only approximately 10 minutes compared to EventNeRF's 14 hours for all scenes. This accelerated training time enables more efficient model development and experimentation. Moreover, our

[1]Comparison methods are evaluated with 360-degree data and the score was reported from the original paper, whereas our approach is evaluated on 90-degree data to replicate the sweep.

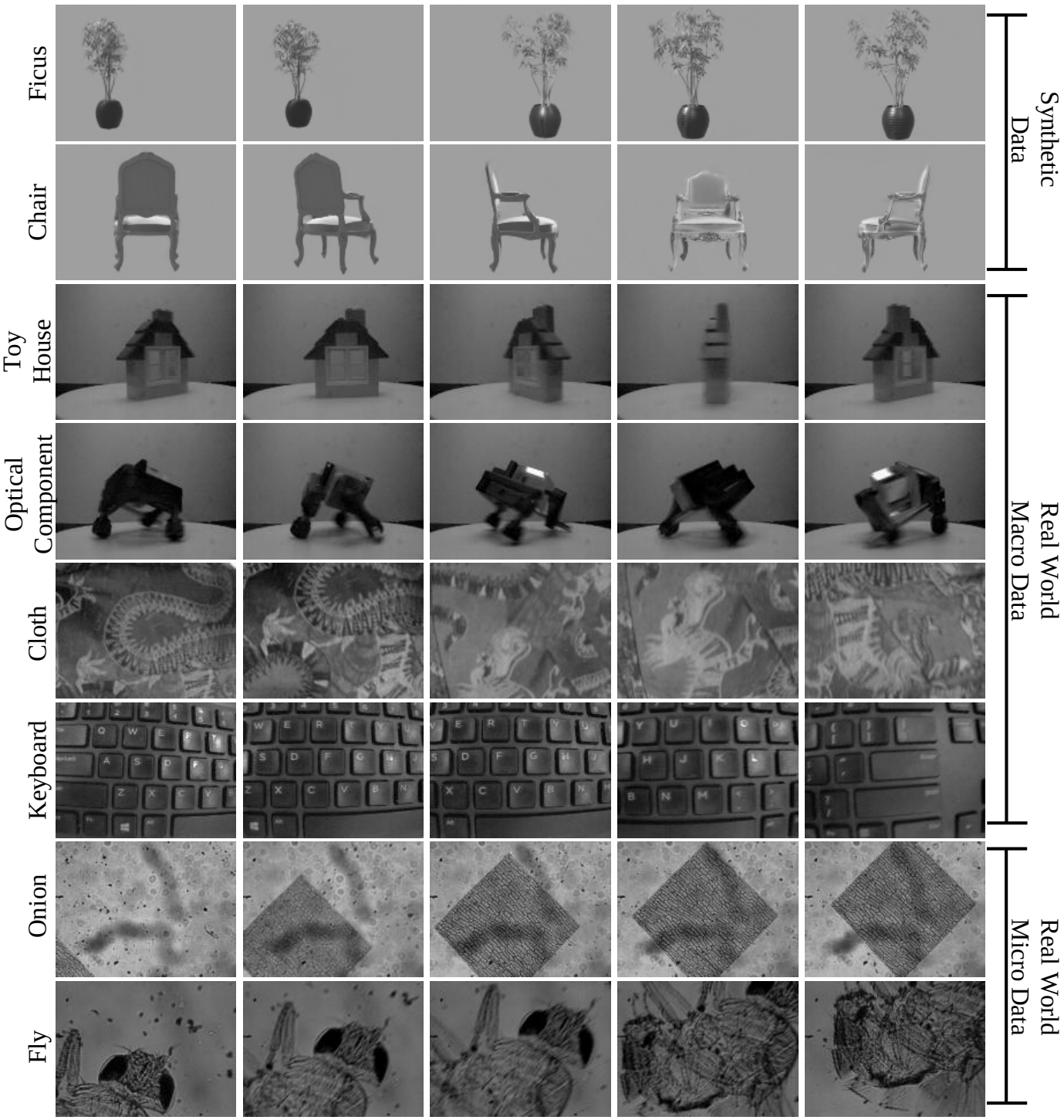


Fig. 5. More visualization of the rendered results on objects from different imaging settings. The top two rows are synthetic data. The middle four rows are real-world macro objects. The bottom two rows are real-world objects captured under the microscope. We showcase the rendering results from different view perspectives for each object. We also demonstrate the rendered results on unseen views of the keyboard and cloth data. This also showcases that our approach is able to reconstruct sharp unseen views with clear semantic information.

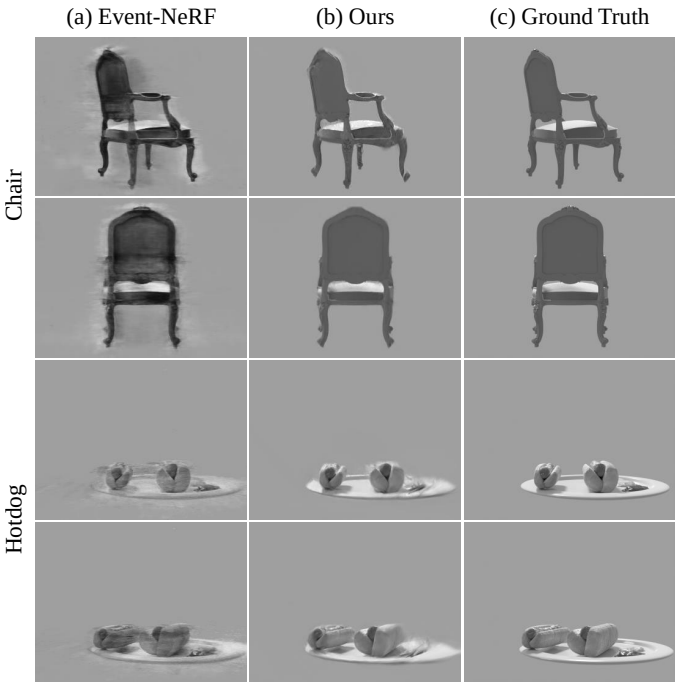


Fig. 6. Visual Comparison between our approach and [30]

method achieves a higher temporal resolution of around 60 fps, compared to EventNeRF's 0.32 fps. This high frame rate is crucial for smooth visual experiences, reducing motion sickness and enhancing productivity and competitiveness. Lastly, our method consumes less memory during the rendering stage, utilizing only 5 GB compared to EventNeRF's 15 GB, which is advantageous for memory-constrained environments and facilitates scalability.

Qualitatively, we further compare visual results for two sequences. As shown in Figure 6, our Ev-GS effectively captures view-dependent effects, structures, and intensity in the frames compared to previous methods.

TABLE III
QUANTITATIVE COMPARISON AGAINST EVENT-NERF [30] AND EV-GS [32] ON SYNTHETIC SEQUENCES.

Metric	PSNR↑	SSIM↑	Training Time↓	FPS↑	Memory↓
Scene			Chair		
EventNeRF	25.6	0.91	14h	0.32	15GB
Ev-GS	28.1	0.93	9min	53.1	**5GB**
Ours	**31.7**	**0.94**	**8min**	**65.7**	**5GB**
Scene			Ficus		
EventNeRF	27.1	0.91	14h	0.33	15GB
Ev-GS	28.1	0.92	9min	63.0	**5GB**
Ours	**29.1**	**0.93**	**8min**	**65.9**	**5GB**
Scene			Hotdog		
EventNeRF	26.0	0.92	14h	0.32	15GB
Ev-GS	25.7	0.93	12min	55.3	**5GB**
Ours	**29.3**	**0.94**	**8min**	**65.3**	**5GB**
Scene			Mic		
EventNeRF	25.0	0.91	14h	0.32	15GB
Ev-GS	24.5	0.92	11min	**66.0**	**5GB**
Ours	**27.5**	**0.93**	**8min**	64.0	**5GB**

G. Ablation Study

1) End-To-End Against Two Stages: To assess the necessity and effectiveness of our proposed end-to-end design, we conducted an ablation study. Specifically, we compared our rendering results to a two-stage baseline approach: first converting event data to frame data using E2VID [49] and then applying 3D-GS for rendering. Table IV summarizes the quantitative evaluation results for two different scenes, Chair and Ficus, based on PSNR and SSIM metrics. As shown in Table IV, our end-to-end design consistently outperforms the two-stage E2VID+3D-GS approach across both scenes and evaluation criteria. The limitations of the E2VID model become apparent when confronted with frames unseen during training. Since E2VID relies on the frames it was trained on, its intermediate results may become inaccurate when presented with novel frames. This inaccuracy propagates through the subsequent 3D-GS stage, resulting in poorer rendering quality, as illustrated in Figure 7. These results underscore the necessity and effectiveness of our end-to-end design, seamlessly integrating event-based video input processing with 3D-GS reconstruction. This integration leads to significantly improved reconstruction quality and fidelity compared to the two-stage approach.

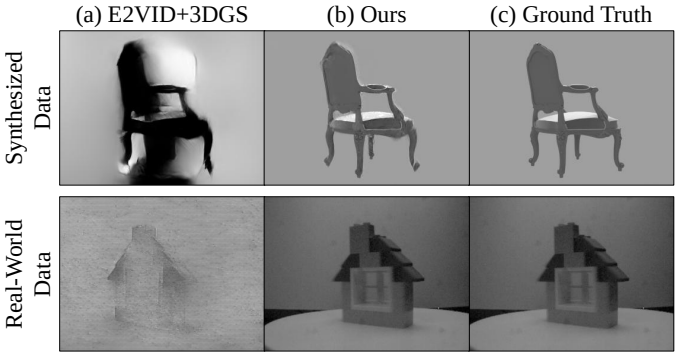


Fig. 7. Visual Comparison between our end-to-end approach and Two-Stage baseline. (a) Use E2VID [49] to generate training frames for 3D-GS; (b) the rendered results of our end-to-end approach; (c) the ground truth frame. We showcase this on both synthetic data and real-world data.

TABLE IV
ABLATION STUDY ON OUR END-TO-END APPROACH AGAINST THE TWO-STAGE BASELINE OF USING [49] PLUS 3D-GS. BOTH EVALUATION METRICS INDICATES THAT OUR END-TO-END APPROACH IS MORE EFFECTIVE AND NECESSARY FOR TRAINING AND RENDERING ACROSS BOTH SYNTHESIZED AND REAL-WORLD DATA.

Scene	Synthesized		Real-World	
Criteria | Metric	PSNR↑	SSIM↑	PSNR↑	SSIM↑
E2VID+3D-GS	15.2	0.72	13.2	0.68
Ours	**29.4**	**0.94**	**29.2**	**0.91**

2) How Event Noise Affects Radiance Field Reconstruction: Designing an approach that is robust across different real-world imaging settings is not trivial. Different hardware imaging systems introduce various challenges due to their unique characteristics and limitations. For instance, real-world macro imaging settings might deal with varying lighting conditions, while microscopy imaging settings often face higher levels

of noise [41]. These variations make it difficult to maintain consistent performance across different scenarios. To address these challenges, we introduced the y-noise filter [42] to eliminate event noise. This simple yet effective modification enhanced the overall performance of our method.

In our experiments, we compared the performance of our method with and without the noise filter. Qualitatively, the input event is more clean on noncritical regions, indicated by Table 8. An input event stream with less noise leads to better reconstruction of radiance field representations. The quantitative results are summarized in Table V. The metrics used for evaluation are the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), both of which are standard measures for assessing image quality. The results indicate that incorporating a noise filter significantly improves the rendering performance. For synthetic data, the PSNR increased from 27.8 to 29.4, and the SSIM improved from 0.90 to 0.94. Similarly, the PSNR increased from 26.9 to 29.2 for real-world data, and the SSIM improved from 0.88 to 0.91. These improvements highlight the effectiveness of our approach in enhancing the quality of the rendered radiance fields, thereby making our method more robust across different types of data.

TABLE V
IMPACT OF NOISE FILTERING ON THE RENDERING PERFORMANCE FOR SYNTHETIC AND REAL-WORLD DATA. INCORPORATING A NOISE FILTER SIGNIFICANTLY IMPROVES BOTH PSNR AND SSIM VALUES.

Metric	PSNR↑	SSIM↑
Data Type	Synthetic	
w/o filter	27.8	0.90
w noise filter	**29.4**	**0.94**
Data Type	Real-World	
w/o filter	26.9	0.88
w noise filter	**29.2**	**0.91**

3) Robust Loss Function across Various Real-World Imaging Settings: In this ablation study, we compare the performance of our proposed loss function against the traditional Mean Squared Error (MSE) loss. Our loss function incorporates a more sophisticated approach to supervising the rendering process. Incorporating normalization techniques into our loss function is essential for maintaining stability across various imaging conditions. Normalization helps to mitigate the impact of differing data distributions and scaling effects, ensuring that the loss function behaves consistently regardless of the specific imaging environment. This stability is particularly important for preserving the quality of rendering results, as it allows the model to perform reliably and effectively across diverse data sources and imaging settings. By stabilizing the loss, our approach ensures that high-quality results are achieved consistently, whether the data comes from synthetic or real-world imaging scenarios. In contrast, the traditional MSE loss focuses solely on minimizing the pixel-wise squared differences between predicted and ground truth images. This simple metric often fails to capture structural and perceptual nuances, which can result in lower visual quality.

Table VI summarizes the quantitative comparison between our loss function and the traditional MSE loss for both synthetic and real-world data. As illustrated, our loss function outperforms the MSE loss significantly. For synthetic data, the PSNR improves from 25.2 to 29.4 and the SSIM from 0.78 to 0.94. Similarly, for real-world data, the PSNR increases from 25.7 to 29.2 and the SSIM from 0.81 to 0.91. These results highlight the effectiveness of our loss function in capturing and preserving important visual and structural details, leading to higher rendering quality compared to the traditional MSE loss. The improved performance underscores the advantages of our approach in producing more accurate and visually appealing results.

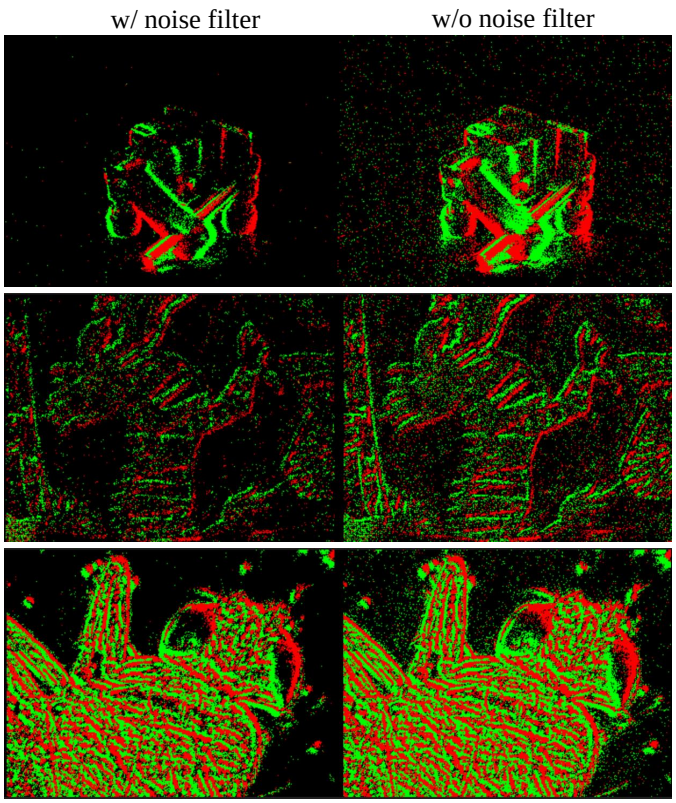


Fig. 8. Visual comparison between noise-filtered events (left) and raw events captured (right). Raw events, especially triggered under different imaging settings, contain different noise and randomness patterns. To adapt Sweep-EvGS for data from different settings, noise is needed to be filtered and balanced.

TABLE VI
QUANTITATIVE COMPARISON OF RENDERING PERFORMANCE USING TRADITIONAL MSE LOSS VERSUS OUR PROPOSED LOSS FUNCTION. OUR LOSS FUNCTION SIGNIFICANTLY IMPROVES PSNR AND SSIM METRICS FOR BOTH SYNTHETIC AND REAL-WORLD DATA.

Metric	PSNR↑	SSIM↑
Data Type	Synthetic	
MSE Loss	25.2	0.78
Our Loss	**29.4**	**0.94**
Data Type	Real-World	
MSE Loss	25.7	0.81
Our Loss	**29.2**	**0.91**

H. Qualitative Analysis

As shown above, we provide rendering results of different kinds of sequences, demonstrating the different capabilities of our approach. Here, we provide more detailed analyses of the unique capability of our approach.

- **Reconstruction of Sharpe Frames**: The first sequence involves a toy house. Traditional frame-based cameras capture blurred frames under fast camera movements, leading the traditional 3D-GS method to learn and render significant blurring, making it difficult to discern the object's details, as shown in the Toy House sequence in Figure 4. In contrast, our method effectively uses the event stream data to reconstruct a sharp and well-defined image of the toy house. This highlights our method's capability to maintain clarity and detail, even in challenging imaging environments.
- **Reconstruction of Challenging Surface Materials**: The second sequence, shown in the introduction of the main text, showcases an optical component with vitric material that is reflective of lights. Traditional methods often struggle to capture the true geometry and appearance of reflective objects accurately. However, our approach excels in rendering these difficult materials, producing clear and accurate reconstructions.
- **Reconstruction of Semantic Information**: The third sequence, shown in the Keyboard sequence in Figure 4, is the classical keyboard scene. We demonstrate this challenging sequence to showcase that our method is not only capable of reconstructing texture information but, compared to existing methods, is also more capable of reconstructing semantic information with clear and detailed visual quality.
- **Reconstruction of Complex Texture:** The fourth sequence, shown in Figure 5, presents a cloth with complex textures. Our approach provides a high-fidelity reconstruction that is particularly challenging for event-based radius field rendering. Unlike conventional methods that may smooth over fine details, our method retains the rich detail and variability inherent in such textures. This capability is crucial for rendering applications in any field requiring accurate representation of complex patterns and materials.

V. Conclusion and Discussion

This paper introduces SweepEvGS, a novel method for robust and accurate novel view synthesis using an event camera. By harnessing the unique advantages of event cameras, SweepEvGS surpasses the limitations of traditional methods that rely on dense and high-quality training frames. It enables efficient novel view synthesis across various imaging settings with minimal data collection requirements. Experimental results demonstrate that SweepEvGS outperforms existing methods in terms of visual rendering quality, rendering speed, and computational efficiency across synthetic data, real-world scenarios, and microscopy settings. By leveraging event streams, SweepEvGS achieves sharper and clearer reconstructions, highlighting its effectiveness in challenging imaging environments. Overall, SweepEvGS represents a significant advancement in novel view synthesis techniques, offering practical solutions for dynamic and varied imaging applications.

VI. Limitations and Future Work

Although SweepEvGS has shown effectiveness in novel view synthesis across different imaging setups with high efficiency, the data collection process is less adaptable. This limitation arises from the reliance on precise camera pose and extrinsic coordinate information for 3D-GS methods. Currently, our real data collection can be performed on machines that can transform the event camera with a known trajectory and speed, making it possible and easy to simulate coordinates. However, an ideal industrial-level algorithm should enable sweeps to be performed by human hands, with flexibility in trajectory and speed, while still producing high-quality view synthesis and reconstruction. A straightforward approach might be to utilize event-based stereo vision to generate coordinates, thereby making the process independent of specific hardware configurations. Future work could focus on further refining this technology to achieve the goal of a flexible, high-quality view synthesis and reconstruction system that is not tied to a particular hardware setup.

VII. Acknowledgments

This work is supported by the Theme-based Research Scheme (Grant No. T45-701/22-R), National Natural Science Foundation of China (Grant No. 62136001, 62088102), Beijing Natural Science Foundation (Grant No. L233024), and Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012).

References

[1] K. Gao, Y. Gao, H. He, D. Lu, L. Xu, and J. Li, "Nerf: Neural radiance field in 3d vision, a comprehensive review," *arXiv preprint arXiv:2210.00379*, 2022.

[2] H. Zhang, Y. Zhang, L. Zhu, and W. Lin, "Deep learning-based perceptual video quality enhancement for 3D synthesized view," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 5080–5094, 2022.

[3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1–14, 2023.

[4] Y. Bao, T. Ding, J. Huo, Y. Liu, Y. Li, W. Li, Y. Gao, and J. Luo, "3d gaussian splatting: Survey, technologies, challenges, and opportunities," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[5] S. Guo, Q. Wang, Y. Gao, R. Xie, L. Li, F. Zhu, and L. Song, "Depth-guided robust point cloud fusion nerf for sparse input views," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[6] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104–4113.

[7] Z. Guo, W. Zhou, L. Li, M. Wang, and H. Li, "Motion-aware 3D gaussian splatting for efficient dynamic scene reconstruction," *arXiv preprint arXiv:2403.11447*, 2024.

[8] Y. Yan, H. Lin, C. Zhou, W. Wang, H. Sun, K. Zhan, X. Lang, X. Zhou, and S. Peng, "Street gaussians for modeling dynamic urban scenes," 2024.

[9] X. Liu, C. Wu, J. Liu, X. Liu, C. Zhao, H. Feng, E. Ding, and J. Wang, "GVA: Reconstructing vivid 3D Gaussian avatars from monocular videos," *CoRR*, 2024.

[10] J. Wen, X. Zhao, Z. Ren, A. G. Schwing, and S. Wang, "Gomavatar: Efficient animatable human modeling from monocular video using gaussians-on-mesh," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2059–2069.

[11] Y. Deng, H. Chen, and Y. Li, "MVF-Net: A multi-view fusion network for event-based object classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8275–8284, 2021.

[12] B. Xie, Y. Deng, Z. Shao, Q. Xu, and Y. Li, "Event voxel set transformer for spatiotemporal representation learning on event streams," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[13] Y.-F. Zuo, J. Yang, J. Chen, X. Wang, Y. Wang, and L. Kneip, "DEVO: Depth-event camera visual odometry in challenging conditions," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2179–2185.

[14] Y. Yang, J. Liang, B. Yu, Y. Chen, J. S. Ren, and B. Shi, "Latency correction for event-guided deblurring and frame interpolation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 24 977–24 986.

[15] S. Zhu, C. Wang, H. Liu, P. Zhang, and E. Y. Lam, "Computational neuromorphic imaging: Principles and applications," in *Computational Optical Imaging and Artificial Intelligence in Biomedical Sciences*, vol. 12857, SPIE, 2024, pp. 4–10.

[16] Z. Chen, J. Wu, J. Hou, L. Li, W. Dong, and G. Shi, "ECSnet: Spatio-temporal feature learning for event camera," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 2, pp. 701–712, 2022.

[17] M. Liu and T. Delbruck, "EDFLOW: Event driven optical flow camera with keypoint detection and adaptive block matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5776–5789, 2022.

[18] X. Long, X. Zhu, F. Guo, C. Chen, X. Zhu, F. Gu, S. Yuan, and C. Zhang, "Spike-BRGNet: Efficient and accurate event-based semantic segmentation with boundary region-guided spiking neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[19] Z. Liu, J. Wu, G. Shi, W. Yang, W. Dong, and Q. Zhao, "Motion-oriented hybrid spiking neural networks for event-based motion deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[20] P. Zhang, S. Zhu, C. Wang, Y. Zhao, and E. Y. Lam, "Neuromorphic imaging with super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[21] 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," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[22] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14 124–14 133.

[23] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. Sajjadi, A. Geiger, and N. Radwan, "RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5480–5490.

[24] J. Han, Y. Yang, P. Duan, C. Zhou, L. Ma, C. Xu, T. Huang, I. Sato, and B. Shi, "Hybrid high dynamic range imaging fusing neuromorphic and conventional images," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 45, no. 7, pp. 8553–8565, 2023.

[25] P. Duan, Y. Ma, X. Zhou, X. Shi, Z. W. Wang, T. Huang, and B. Shi, "NeuroZoom: Denoising and super resolving neuromorphic events and spikes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

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

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

[28] C. Wang, S. Zhu, P. Zhang, J. Huang, K. Wang, and E. Y. Lam, "Neuromorphic shack-hartmann wave normal sensing," *arXiv preprint arXiv:2404.15619*, 2024.

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

[30] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik, "EventNeRF: Neural radiance fields from a single colour event camera," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4992–5002.

[31] J. Wang, J. He, Z. Zhang, M. Sun, J. Sun, and R. Xu, "EvGGS: A collaborative learning framework for event-based generalizable gaussian splatting," *arXiv preprint arXiv:2405.14959*, 2024.

[32] J. Wu, S. Zhu, C. Wang, and E. Y. Lam, "Ev-GS: Event-based gaussian splatting for efficient and accurate radiance field rendering," in *2024 IEEE 34rd International Workshop on Machine Learning for Signal Processing (MLSP)*, 2024, pp. 1–6.

[33] H. Han, J. Li, H. Wei, and X. Ji, "Event-3dgs: Event-based 3d reconstruction using 3d gaussian splatting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 128 139–128 159, 2024.

[34] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–14, 2019.

[35] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-NeRF 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5470–5479.

[36] H. Zhou, B. Y. Feng, H. Guo, S. S. Lin, M. Liang, C. A. Metzler, and C. Yang, "Fourier ptychographic microscopy image stack reconstruction using implicit neural representations," *Optica*, vol. 10, no. 12, pp. 1679–1687, 2023.

[37] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3D Gaussian splatting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 19 447–19 456.

[38] J. Gao, C. Gu, Y. Lin, H. Zhu, X. Cao, L. Zhang, and Y. Yao, "Relightable 3D Gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing," *arXiv preprint arXiv:2311.16043*, 2023.

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

[40] M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "Surface splatting," in *Conference on Computer Graphics and Interactive Techniques*, 2001, pp. 371–378.

[41] R. Yang, T. Xiao, Y. Cheng, Q. Cao, J. Qu, J. Suo, and Q. Dai, "Sci: A spectrum concentrated implicit neural compression for biomedical data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 4774–4782.

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

[43] P. Duan, Z. W. Wang, B. Shi, O. Cossairt, T. Huang, and A. K. Katsaggelos, "Guided event filtering: Synergy between intensity images and neuromorphic events for high performance imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8261–8275, 2021.

[44] I. E. Commission *et al.*, "Multimedia systems and equipment-color measurement and management-part 2-1," *Color management-Default RGB color space-sRGB*, 1999.

[45] A. Zihao Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based optical flow using motion compensation," in *European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

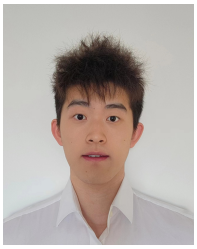
[46] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic dvs events," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1312–1321.

[47] D. Gehrig, M. Gehrig, J. Hidalgo-Carrió, and D. Scaramuzza, "Video to events: Recycling video datasets for event cameras," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3586–3595.

[48] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with cuda: Is cuda the parallel programming model that application developers have been waiting for?" *Queue*, vol. 6, no. 2, pp. 40–53, 2008.

[49] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1964–1980, 2019.

VIII. Biography Section

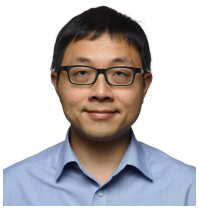


Jingqian Wu received the B.S. degree from Wake Forest University, USA, in 2022, and the M.S. degree from Northwestern University, USA, in 2024. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include computational imaging, computer vision and event-based vision.

SHUO ZHU received his B.S. degree from the Changchun University of Science and Technology in 2016, M.S. degree from the University of Shanghai for Science and Technology in 2019, and Ph.D. degree in optical engineering from the Nanjing University of Science and Technology in 2023. He is now a postdoctoral fellow at the University of Hong Kong. His research interest is computational neuromorphic imaging and its optical applications.

CHUTIAN WANG received the B.S. degree in Huang Kun Elite Class from the University of Science & Technology Beijing in 2020, and the M.S. degree in the major of Optics and Photonics in Imperial College London in 2021. He was a research assistant in Zhejiang University until 2022. He is currently working towards his PhD degree with the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include computational imaging and neuromorphic imaging.

Boxin Shi (Senior Member, IEEE) received the B.E. degree from the Beijing University of Posts and Telecommunications, the M.E. degree from Peking University, and the PhD degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently a Boya Young Fellow Associate Professor (with tenure) and Research Professor at Peking University, where he leads the Camera Intelligence Lab. Before joining PKU, he did research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University, National Institute of Advanced Industrial Science and Technology, from 2013 to 2017. His papers were awarded as Best Paper, Runners-Up at CVPR 2024, ICCP 2015 and selected as Best Paper candidate at ICCV 2015. He is an associate editor of TPAMI/IJCV and an area chair of CVPR/ICCV/ECCV. He is a senior member of IEEE.



EDMUND Y. LAM (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University. He was a Visiting Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. He is currently a Professor of electrical and electronic engineering at The University of Hong Kong. He also serves as the Computer Engineering Program Director and a Research Program Coordinator with the AI Chip Center for Emerging Smart Systems. His research interest includes computational imaging algorithms, systems, and applications. He is a fellow of Optica, SPIE, IS&T, and HKIE, and a Founding Member of the Hong Kong Young Academy of Sciences.