Event-3DGS: Event-based 3D Reconstruction Using 3D Gaussian Splatting

Hanqian Han **Jianing Li** **Henglu Wei** **Xiangyang Ji***
Tsinghua University

Abstract

Event cameras, offering high temporal resolution and high dynamic range, have brought a new perspective to addressing 3D reconstruction challenges in fast-motion and low-light scenarios. Most methods use the Neural Radiance Field (NeRF) for event-based photorealistic 3D reconstruction. However, these NeRF methods suffer from time-consuming training and inference, as well as limited scene-editing capabilities of implicit representations. To address these problems, we propose Event-3DGS, the first event-based reconstruction using 3D Gaussian splatting (3DGS) for synthesizing novel views freely from event streams. Technically, we first propose an event-based 3DGS framework that directly processes event data and reconstructs 3D scenes by simultaneously optimizing scenario and sensor parameters. Then, we present a high-pass filter-based photovoltage estimation module, which effectively reduces noise in event data to improve the robustness of our method in real-world scenarios. Finally, we design an event-based 3D reconstruction loss to optimize the parameters of our method for better reconstruction quality. The results show that our method outperforms state-of-the-art methods in terms of reconstruction quality on both simulated and real-world datasets. We also verify that our method can perform robust 3D reconstruction even in real-world scenarios with extreme noise, fast motion, and low-light conditions. Our code is available in `https://github.com/lanpokn/Event-3DGS`.

1 Introduction

3D reconstruction [8] plays a crucial role in various cutting-edge fields, such as robot vision, virtual reality, and augmented reality systems. It usually enables the creation of accurate 3D models from ideal frame sequences. Nevertheless, with conventional cameras, 3D reconstruction performance has suffered from a significant drop in some challenging conditions [13, 36] (e.g., fast motion blur and low light). Thus, how to use a new visual sensing paradigm for 3D reconstruction to overcome the shortcomings of conventional cameras remains a partially unsolved issue.

Event cameras [9, 21, 30], namely bio-inspired dynamic vision sensors, fundamentally differ from conventional cameras that capture frames at fixed intervals. Event cameras operate asynchronously, recording light changes with dynamic events at the microsecond level. This unique property endows event cameras with high temporal resolution, high dynamic range, low power consumption, and low latency. These advantages have driven their application in various challenging vision tasks [12, 20, 23, 41, 43, 50], including recent efforts in 3D reconstruction [32].

Despite efforts [3, 17, 42, 51] to use event cameras for 3D reconstruction, real-world performance in terms of quality, robustness, and real-time capabilities still needs improvement. Traditional non-learning optimization-based methods [3, 17, 32, 51] serve as the foundation for event-based 3D reconstruction, but they often struggle with robustness and rendering quality. Recently, Neural

*Corresponding author: xyji@tsinghua.edu.cn

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

 https://doi.org/10.52202/079017-4069

Radiance Fields (NeRF) [10, 28, 40] have gained popularity for scene representation and novel view synthesis from event data, utilizing a Multi-Layer Perceptron (MLP) and differentiable rendering. Although these NeRF-based methods [1, 2, 6, 14, 18, 24, 25, 26, 27, 31, 35, 52] achieve impressive results in photorealistic 3D reconstruction from neuromorphic cameras, they suffer from time-consuming training and inference processes. Additionally, their implicit representations limit scene editing capabilities. Moreover, NeRF have primarily been investigated using simulated data and high-quality real-world images captured under ideal conditions (e.g., optimal lighting and minimal noise), posing limitations on real-world 3D reconstruction. In contrast, the emergence of 3D Gaussian Splatting (3DGS) [5, 11, 15, 46, 47] presents a compelling alternative, boasting high reconstruction accuracy and swift inference speed. However, 3DGS has predominantly been utilized with image or video data for 3D reconstruction, with limited exploration in event streams.

To address this gap, we propose Event-3DGS, the first event-based reconstruction framework utilizing 3DGS for synthesizing novel views from event streams. More specifically, we introduce an event-based 3DGS framework, enabling direct processing of event data and reconstruction of 3D scenes while simultaneously optimizing scenario and sensor parameters. Then, we present a high-pass filter-based photovoltage estimation module, effectively reducing noise in event data to enhance the robustness of our method in real-world scenarios. Finally, we propose an event-based 3D reconstruction loss to optimize the parameters of our method for better reconstruction quality. Extensive experiments show that our method outperforms state-of-the-art methods in reconstruction quality on simulated and real-world datasets. This pioneering work in event-based 3D reconstruction with 3DGS sets a new benchmark, opening new avenues for high-quality, efficient, and robust 3D reconstruction in challenging real-world scenarios, such as extreme noise, fast motion, and low light. Our contributions can be summarized as follows:

- We introduce Event-3DGS, the first framework that combines event cameras with 3DGS technology, enabling 3D reconstruction in challenging real-world scenarios.
- We present a high-pass filter-based photovoltage contrast estimation module, which effectively estimates photovoltage contrast by reducing noise in event streams for robust 3D reconstruction.
- We design a novel event-based 3D reconstruction loss to optimize the parameters of our method for better reconstruction quality.

2 Related Works

Event-based 3D Reconstruction. Early attempts [3, 16, 17, 32, 51] at using event cameras for 3D reconstruction typically relied on geometric models and handcrafted features. However, these non-learning, optimization-based methods often struggle to achieve robust and high reconstruction quality. A growing trend is the use of neural radiance fields (NeRF) for scene representation and novel view synthesis. For instance, Ev-NeRF [14] and E-NeRF [18] are some of the earliest works to apply NeRF for event-based 3D reconstruction. These methods render images at different times, generate events through differencing, and compare them with actual events. Further advancements include DeNeRF [25] and EvDNeRF[1], which introduce Deformable NeRF for dynamic scene reconstruction. EventNeRF [35] extends this by enabling colored rendering through the incorporation of three-channel events into NeRF. Some methods like E2NeRF [31] and Ev-DeblurNeRF [2] perform hybrid reconstruction to mitigate motion blur by combining blurred images with events. However, these NeRF-based approaches face significant challenges, including the time-consuming generation of novel views and limited scene editing capabilities due to implicit representations. Additionally, NeRF have primarily been explored using simulated data and high-quality images captured under ideal conditions, leaving a considerable gap between the models and real-world scenarios. Therefore, our goal is to design a novel event-based 3D reconstruction framework that ensures high-quality, efficient, and robust performance in real-world scenarios.

3D Gaussian Splatting for 3D Reconstruction. 3D Gaussian Splatting (3DGS) [15] has significantly advanced in 3D scene representation, offering notable advantages over NeRF by capturing complex geometries and lighting effects more accurately and efficiently. These advantages make 3DGS highly suitable for real-time and real-world applications. Extended works like 4DGS [46, 47] and D-3DGS [48] further enhance dynamic scene rendering. Besides, some works integrate with Simultaneous Localization and Mapping (SLAM) [39] and text-to-3D models [4] to expand 3DGS capabilities. However, these methods have primarily been applied to image and video data, leaving

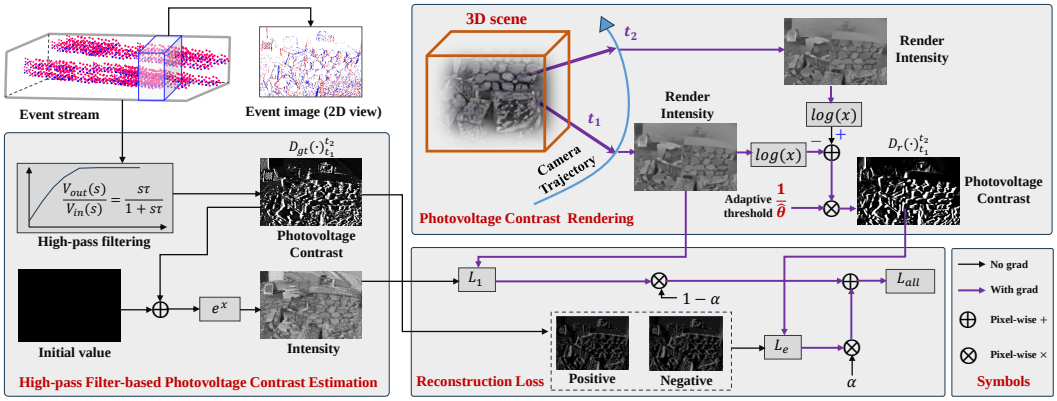

Figure 1: The pipeline of **Event-based 3D Reconstruction using 3D Gaussian Splatting (Event-3DGS)**. The proposed event-based 3DGS framework enables direct processing of event data and reconstructs 3D scenes while simultaneously optimizing scenario and sensor parameters. A high-pass filter-based photovoltage contrast estimation module is presented to reduce noise in event data, enhancing the robustness of our method in real-world scenes. An event-based 3D reconstruction loss is designed to optimize the parameters of our method for better reconstruction quality.

their potential with event cameras largely unexplored. Thus, designing a novel 3DGS model to directly process asynchronous events for 3D reconstruction remains an open challenge.

3 Method

3.1 Event-3DGS Architecture

To achieve high-quality, efficient, and robust 3D reconstruction in challenging real-world scenarios, we propose a novel event-based 3D reconstruction framework using 3D Gaussian Splatting (Event-3DGS). As shown in Fig. 1, our framework mainly consists of three modules: **high-pass filter-based photovoltage contrast estimation**, **photovoltage contrast rendering**, and **event-based 3D reconstruction loss**. More precisely, we first present a high-pass filter-based photovoltage contrast estimation module that reduces noise in event data to enhance the robustness of our method in real-world scenes (see Sec. 3.2). Then, we design a photovoltage contrast rendering module that obtains the photovoltage contrast image by calculating the difference in light intensity in 3DGS. After obtaining two contrast estimations, we propose a novel event-based 3D reconstruction loss to measure the differences (see Sec.3.3). Finally, our method optimizes the 3D scene and camera parameters by propagating gradients through backpropagation.

3DGS [15] demonstrates superior 3D reconstruction capabilities by rapidly converting input images into highly detailed 3D point clouds, accurately representing the scene. For a specific 3D scene represented by 3D Gaussian functions, the forward process of 3DGS can be regarded as a mapping function $G(\mathbf{T})$, which gets the rendered image by alpha blending in the corresponding camera pose $\mathbf{T}$ in time t. For a single pixel on a single channel, alpha blending can be described as follows:

$$L = \sum_{i=1}^{N} l_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \tag{1}$$

where L denotes the pixel value result, which can be intensity or one of the three channels. l_i and α_i are the color and opacity of each point mapped to this pixel, respectively.

Event cameras operate on a fundamentally distinct imaging principle, generating event data in the form of sparse points (see Sec. 3.2). This disparity prevents the direct integration of asynchronous events into the original 3DGS formulation. To bridge this gap, we integrate event data seamlessly with the output of 3DGS by leveraging photovoltage contrast (i.e., intensity changes). Considering two close-in-time instances t_1 and t_2, the camera poses corresponding to these two moments are $\mathbf{T_1}$ and $\mathbf{T_2}$. We can obtain its photovoltage contrast between two moments using the proposed high-pass

https://doi.org/10.52202/079017-4069

filter-based photovoltage contrast estimation module, and it can be formulated as:

$$D_{gt}(\cdot)_{t_1}^{t_2} = \frac{1}{\theta}((V(\cdot, t_2) - V(\cdot, t_1)), \tag{2}$$

where $V(p, t)$ refers to the photovoltage in pixel p and time t.

Correspondingly, we need to use 3DGS to render the photovoltage contrast and light intensity, and then compare these with the ground truth obtained from the event data for subsequent reconstruction. The photovoltage contrast image can be obtained by the proposed photovoltage contrast rendering module. The rendering process can be mathematically described as follows:

$$D_r(\cdot)_{t_1}^{t_2} = \frac{1}{\hat{\theta}}(\log(G(\mathbf{T_2}) + \epsilon) - \log(G(\mathbf{T_1}) + \epsilon)), \tag{3}$$

where $G(\mathbf{T})$ denotes the intensity result in 3DGS. $D_r(p)$ is the normalized contrast value in the pixel p, and ϵ is a small number that avoids log(0). $\hat{\theta}$ is a learnable parameter, which is an estimate of the threshold in the event sensor. For the sake of convenience in writing, we do not strictly distinguish between time t and the corresponding pose T.

Given the intrinsic characteristics of event data, it's essential to highlight that intensity estimation methods [37, 45] are unlikely to outperform photovoltage contrast estimation techniques. Even data-driven learning-based models [7, 34] may only yield visually appealing results without ensuring physical accuracy and generality. To address this challenge, we proposed a dynamic adjustment strategy for intensity to use high-quality photovoltage while ensuring robustness and stability.

3.2 High-pass Filter-based Photovoltage Contrast Estimation

To gain a deeper understanding of the high-pass filter-based photovoltage contrast estimation module, it's essential to begin with a fundamental understanding of the Dynamic Vision Sensor (DVS) [9]. The DVS is designed to capture changes in light intensity pixel-wise and asynchronously. It accomplishes this by converting the light spectrum into photocurrent as follows:

$$I(p, t) = \int \lambda \cdot QE(\lambda) \cdot L(p, t, \lambda) \cdot d\lambda, \tag{4}$$

where $L(p, t, \lambda)$ denotes the spectrum, t is time, p refers to the pixel coordinate $\{x, y\}$, and λ is the wavelength of light. $I(p, t)$ is the photocurrent or intensity in pixel p and time t. QE is quantum efficiency, which represents the weighting of different wavelength light converted into photocurrent.

Subsequently, the photocurrent will be converted into photovoltage through a logarithmic function. When the voltage change surpasses a predefined threshold θ, an event will be triggered [30] as:

$$V(p, t_2) - V(p, t_1) = \sigma_i^p \theta, \tag{5}$$

where $V(p, t)$ refers to the logarithmic operation of the photocurrent $I(p, t)$. σ_i^p is either +1 or -1 by comparing $V(p, t_2)$ and $V(p, t_1)$.

In general, the threshold θ is an inherent attribute of the DVS. It can be defined as follows:

$$\theta = \frac{C_{diff} V_{th}}{U_T A_v}, \tag{6}$$

where C_{diff} is determined by the capacitance in the ,V_{th} is a fixed constant, U_T is thermal voltage, and A_v is voltage gain factor. This formula illustrates that θ often varies with sensor settings and environmental changes, making it generally difficult to obtain its true value. In this work, we will optimize the threshold θ together with the 3D scene.

Intuitively, asynchronous events appear as sparse points [19] in the spatiotemporal domain:

$$E(p, t) = \sum_{i=1}^{N} \sigma_i^p \theta \delta(t - t_i^p), \tag{7}$$

where $\delta(\cdot)$ refers to the Dirac delta function, with $\int \delta(t)\, dt = 1$ and $\delta(t) = 0, \forall t \neq 0$.

In this work, we implement the reverse process of dynamic event generation by extracting photovoltage contrast and intensity from the event stream $E(p,t)$. The photovoltage is given as:

$$V(p,t) - V(p,t_i) = E(p,t_j) + w(p,t), \tag{8}$$

where t_i and t_j are the times corresponding to two adjacent events around t. $w(p,t)$ denotes an uncertainty term, which is determined by the mathematical principles of the event camera and is unrelated to noise. This formula indicates that the event camera cannot accurately reconstruct the intensity of light outside the triggering event's timing. Moreover, since $V(p,t_i)$ is unknown, it is also impossible to accurately obtain the intensity of light at each event's timing. In short, only the intensity difference between each event's triggering times can be accurately obtained.

For the photovoltage between two events, we can simply assume:

$$w(p,t) = -E(p,t_j), t < t_j. \tag{9}$$

The above assumption will not interfere with the voltage at the time of event triggering. If $E(p,t)$ is ideal, V can be directly obtained from E through pure integration, that is:

$$\hat{V_d}(p,t) = V(p,t) - V(p,t_0) = \int_0^t E(p,t)dt, \tag{10}$$

where $\hat{V_d}(p,t)$ is the photocurrent contrast to be estimate. Ideally, this method can accurately provide the photovoltage contrast between the triggering moments of each event.

However, when event cameras are applied in real-world 3D reconstruction, they often encounter various types of noise, making it challenging to accurately estimate the photovoltage contrast through pure integration [37]. To address this issue, we use high-pass filtering to process the event data. The high-pass filtering can be depicted as follows:

$$\frac{\mathcal{V}_{out}(s)}{\mathcal{V}_{in}(s)} = \frac{s\tau}{1+s\tau}, \tag{11}$$

where $V_{out}(s)$ and $V_{in}(s)$ are the input-output signals after the Laplace transformation. τ is the time constant related to the cutoff frequency. We treat $\int_0^t E(p,t)dt$ as a noisy input V_{in} and $\hat{V_d}(p,t)$ as the output V_{out}. When substituted into Eq. 11 and transformed back into the time domain, we obtain that:

$$\hat{V_d}(p,t) = E(p,t) - \frac{1}{\tau}\dot{\hat{V}}_d(p,t), \tag{12}$$

where $\dot{\hat{V}}_d(p,t)$ is the differential with respect to the timestamp.

By simultaneously solving with Eq. 7, we can estimate the photovoltage contrast between any two corresponding moments. Once the photovoltage is obtained, the intensity can be computed as:

$$I(p,t) = e^{\hat{V_d}(p,t)+\hat{V_d}(p,0)}. \tag{13}$$

$V(p,t_0)$ does not affect photovoltage contrast estimation. Therefore, our high-pass filter-based method typically provides more accurate results than restoring pure photovoltage or intensity.

3.3 Event-based 3D Reconstruction Loss

For better reconstruction quality, we use a loss function with two key components: intensity estimation and photovoltage estimation. The light intensity is evaluated as follows:

$$l_i^{t_1} = l_1(I(\cdot,t_1), G(\mathbf{T_1})), \tag{14}$$

where l_1 is a loss function that computes the average of the absolute values between each pixel.

For the evaluation of photovoltage contrast, it's important to consider that 3DGS uses a rasterization method to generate the entire image at once. This means the rendering sampling time often does not match the event triggering time for most pixels. According to Eq. 8, the ground truth of the photovoltage contrast inherently contains some errors. Additionally, despite applying filtering methods, event data still has significant noise that cannot be entirely eliminated. Consequently, if we directly use the L_1 loss to compare the rendered photovoltage contrast with the photovoltage contrast calculated from event data, these errors will be strictly considered during the reconstruction process,

 https://doi.org/10.52202/079017-4069

Table 1: Performance comparison on the DeepVoxels synthetic dataset [38]. our Event-3DGS outperforms two state-of-the-art methods and our baseline using pure integration without filtering.

Sequence	E2VID [34]			E2VID [34]+3DGS [15]			PI-3DGS			Event-3DGS		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
mic	0.938	19.965	0.048	0.946	19.955	0.068	0.955	21.979	0.060	0.952	21.127	0.063
ship	0.808	16.556	0.108	0.825	16.681	0.122	0.792	16.750	0.177	0.818	17.815	0.147
materials	0.872	18.302	0.084	0.885	18.325	0.094	0.925	20.053	0.062	0.933	20.506	0.060
lego	0.883	19.744	0.075	0.899	20.002	0.084	0.928	23.853	0.056	0.925	23.046	0.058
ficus	0.932	19.795	0.043	0.935	19.626	0.056	0.939	19.880	0.050	0.940	19.939	0.049
drums	0.908	18.312	0.071	0.915	18.288	0.085	0.953	22.643	0.041	0.951	22.568	0.042
chair	0.939	23.842	0.040	0.949	23.866	0.050	0.954	27.024	0.042	0.953	27.336	0.050
Average	0.897	19.502	0.067	0.908	19.535	0.080	0.921	21.740	0.070	**0.925**	**21.762**	**0.067**

which can actually degrade the reconstruction quality. Thus, if a new loss can be designed that allows for a certain tolerance in the estimation of photovoltage contrast, it would enable better completion of the reconstruction task. For simplification, we denote $L_g^p = D_{gt}(p)_{t_1}^{t_2}$, $L_r^p = D_r(p)_{t_1}^{t_2}$. The loss function of photovoltage contrast can be formulated as:

$$l_e^{p,t_1,t_2}(L_g^p, L_r^p) = \begin{cases} R(|L_r^p - L_g^p - \beta| - \beta) & \text{if } L_g^p > 0 \\ R(|L_g^p - L_r^p - \beta| - \beta) & \text{if } L_g^p < 0 \end{cases}, \tag{15}$$

where $R(x) = \max(0, x)$, and β is a measure of the tolerance in the estimated photovoltage contrast obtained from rendering. When β is infinitely large, the loss becomes a trivial constant zero, imposing no constraints. Conversely, if β is zero, the loss degenerates into a very strict L_1 loss. Theoretical analysis shows that setting β to 0.5 yields excellent results in the real-world dataset. When β is set to 0.5, it best aligns with the characteristics described in Eq. 8.

Finally, the total loss function for event-based 3D reconstruction can be described as follows:

$$l_{all}(t_1, t_2) = \alpha \sum_{p \in R^2} \frac{l_e^{p,t_1,t_2}(L_g^p, L_r^p)}{W * H} + (1 - \alpha) l_i^{t_1}, \tag{16}$$

where W and H are the width and height of the image, respectively. α is a parameter that controls the weight of the intensity. For the first 8000 epoch, α is set to 0 to give an initialization of the reconstruction. Then, α is generally set near 1 to improve the quality of 3D reconstruction.

4 Experiments

4.1 Experimental Setting

Datasets. To evaluate the effectiveness of our Event-3DGS, we conduct experiments on the DeepVoxels synthetic dataset [38] and the real-world Event-Camera dataset[29]. For the synthetic dataset, we use seven sequences with continuous 180-degree image rotations on a gray background as the ground truth for reconstruction. These sequences are processed by the VOLT simulator [22] to generate event data, offering a more realistic simulation than ESIM [33] with higher noise levels. For the real-world dataset, we select five typical sequences that provide aligned image and event data under fast motion and low-light conditions. For longer sequences, we typically utilize the initial 100 images for training and evaluate performance on separate data not employed during reconstruction.

Implementation Details. We set τ to 0.05 for the high-pass filter-based photovoltage contrast estimation module. In the loss function, we set α to 0.9. For synthetic experiments with low noise, β is set to 0, while for real data with higher noise, β is set to 0.5. We utilize E2VID [34] for initial intensity estimation. We use E2VID+3DGS as a baseline for event-based 3D reconstruction, comparing it with our full method to validate its efficacy. All experiments are conducted on an AMD Ryzen Threadripper 3970X 32-Core CPU and an NVIDIA GeForce RTX 3080 Ti GPU. The evaluation metrics use the Peak Signal Noise Ratio (PSNR), the Structural Similarity (SSIM) [44], and the Learned Perceptual Image Patch Similarity (LPIPS) [49].

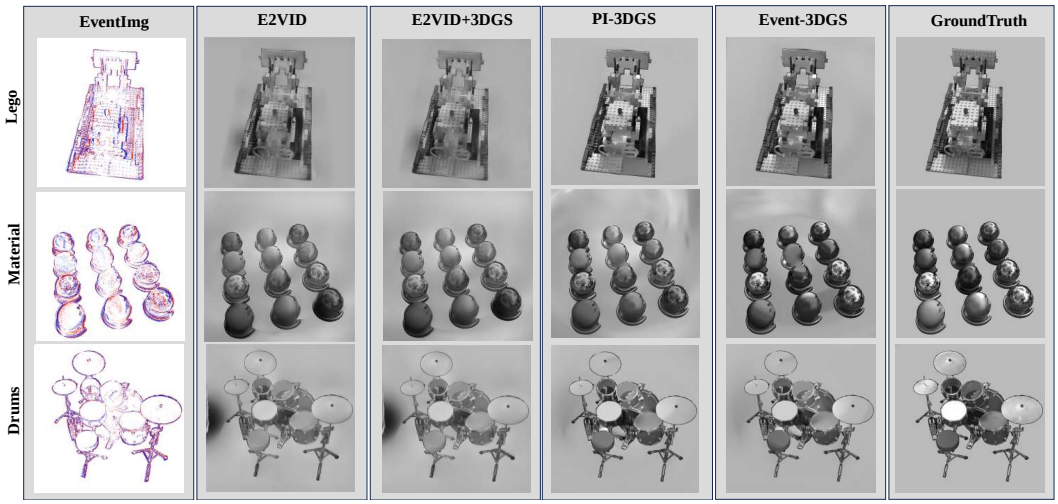

Figure 2: Representative visualization results on the DeepVoxels synthetic dataset [38]. Obviously, our Event-3DGS produces visually pleasing images with fine details and fewer artifacts.

Table 2: Performance comparison on the real-world Event-Camera dataset[29]. Note that, our Event-3DGS surpasses three comparative methods on three metrics.

Sequence	E2VID [34]			E2VID [34]+3DGS [15]			PI-3DGS			Event-3DGS		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
boxes	0.356	8.705	0.320	0.408	8.841	0.228	0.224	8.364	0.749	0.575	17.696	0.260
office_zigzag	0.326	7.550	0.287	0.346	7.649	0.233	0.310	10.009	0.424	0.430	14.043	0.183
slider_depth	0.353	7.634	0.317	0.356	7.620	0.309	0.477	13.509	0.276	0.497	12.448	0.261
outdoors_walking	0.156	3.657	0.508	0.179	3.499	0.361	0.138	3.429	0.738	0.271	10.583	0.300
calibration	0.239	5.669	0.293	0.276	5.742	0.316	0.270	11.698	0.714	0.312	11.065	0.222
Average	0.286	6.643	0.345	0.313	6.670	0.290	0.284	9.402	0.580	**0.417**	**13.167**	**0.245**

4.2 Effective Test

Evaluation on Synthetic Data. To verify the effectiveness of our method, we select two state-of-the-art methods (i.e., E2VID [34] and reconstructed images for 3DGS [15]) and our baseline using Pure Integration without filtering for 3DGS (PI-3DGS) as three comparison methods. As shown in Table 1, our method and our baseline outperform the representative reconstruction method (i.e., E2VID [34]) on the DeepVoxels synthetic dataset [38]. More precisely, our method has improved by 0.017 and 2.227 respectively compared to the competitor (i.e., E2VID [34]+3DGS [15]) in SSIM and PSNR, while decreasing by 0.013 in LPIPS. Furthermore, we present some representative visualization results on the DeepVoxels synthetic dataset [38] in Fig. 2. Note that, our method produces visually pleasing images with fine details and fewer artifacts. For clarification, the visualization is for comparative analysis, our method is capable of synthesizing novel views.

Evaluation on Real-world Data. To evaluate the performance of our method in real-world scenarios, we present the comparative results of the real-world Event-Camera dataset[29] in Table 2. We select three representative comparison methods to highlight the performance of Event-3DGS. Our baseline (i.e., PI-3DGS) performs worse because the pure integration method for estimating light intensity is sensitive to noise. Our method, with its high-pass filter-based photovoltage contrast estimation module, effectively filters out noise in real scenes and improves reconstruction quality (see Fig. 3). To demonstrate its robustness in extreme scenarios such as high-speed motion or low lighting, we selected two typical scenarios in Fig. 4. We can find that conventional cameras struggle in low-light conditions and produce significant motion blur in high-speed scenarios. These examples show that our method can reconstruct 3D scenes from the event stream to overcome the limitations of conventional cameras in challenging conditions.

4.3 Ablation Test

https://doi.org/10.52202/079017-4069

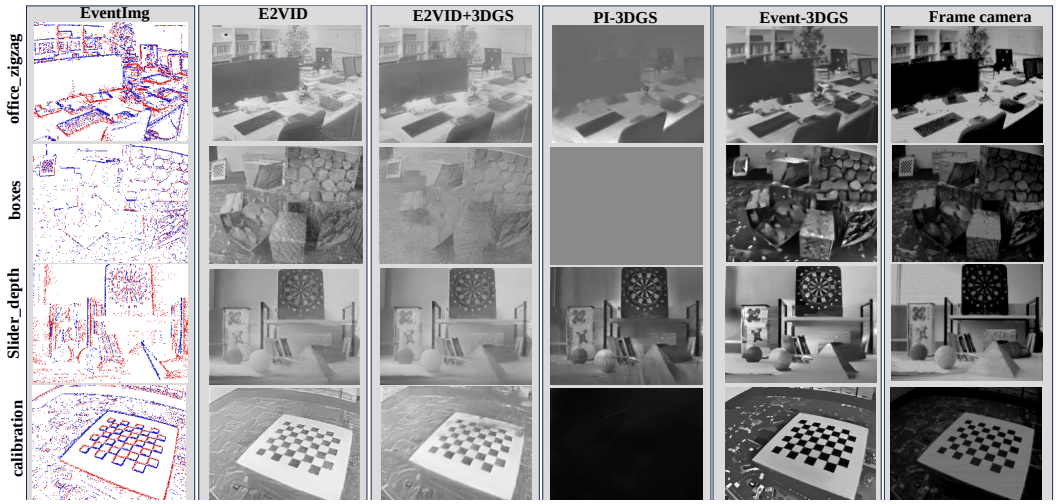

Figure 3: Representative visualization results on the real-world Event-Camera dataset[29]. Note that, our Event-3DGS achieves better reconstruction quality than the three comparative methods.

Contribution of Each Component. To explore the impact of each component on the final reconstruction performance, we chose the pure integration image without the adaptive threshold and event loss as the baseline. As illustrated in Table 3, We use three different strategies (e.g., adaptive thresholding, high-pass filtering, and loss function) to enhance our baseline. As a result, our method achieves the best performance among all competitors. In other words, our method employs these effective components to process event streams for 3D reconstruction.

Influence of the Parameter α. To analyze the hyperparameter α of the loss function, we set the hyperparameter α with various values (e.g., 0.05, 0.2, 0.4, 0.6, 0.8, 0.9, and 1). As shown in Table 4, large values of α pose a risk of the training deviating to suboptimal points. Small values render the main phase ineffective, resulting in degraded outcomes similar to the initial phase.

Influence of the Parameter β. We report the impact of varying α in Table 5. The results indicate that the reconstruction performs optimally when β is set to 0.5. Deviations from this value led to a decline in performance, consistent with theoretical predictions.

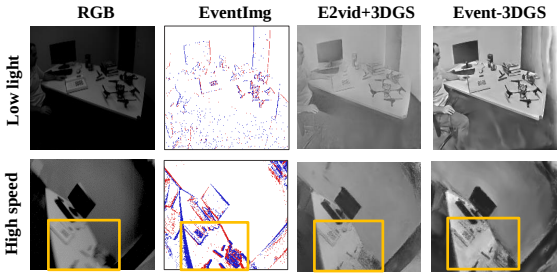

Figure 4: Representative visualization examples on low-light and high-speed motion blur scenarios.

Table 3: The contribution of each component.

Threshold				✓		✓	✓	✓
Filtering			✓		✓		✓	✓
Loss		✓			✓	✓		✓
SSIM	0.219	0.317	0.376	0.221	0.410	0.310	0.393	**0.430**
PSNR	6.713	11.197	9.594	6.878	11.715	10.009	9.840	**14.043**
LPIPS	0.767	0.463	0.249	0.765	0.217	0.424	0.191	**0.183**

Table 4: The influence of the Parameter α

α	1	0.99	0.9	0.8	0.6	0.4	0.2	0.05
SSIM	0.945	0.953	**0.956**	0.955	0.954	0.953	0.952	0.952
PSNR	25.688	**27.302**	25.974	25.066	24.540	24.300	24.146	24.148
LPIPS	0.063	0.051	0.039	0.040	**0.038**	0.038	0.039	0.039

4.4 Scalability Test

Event-3DGS for Motion Deblurring. Our Event-3DGS can be further expanded to motion deblurring. By integrating event data with RGB frames, our method can achieve deblurring effects using the hybrid reconstruction manner. As shown in Fig. 6, We test the hybrid framework on some simulated sequences [15] using VOLT [22]. Note that, a blurred image is generated through integration to serve as the RGB input. Our Event-3DGS leverages event data to achieve deblurred color reconstruction.

https://doi.org/10.52202/079017-4069

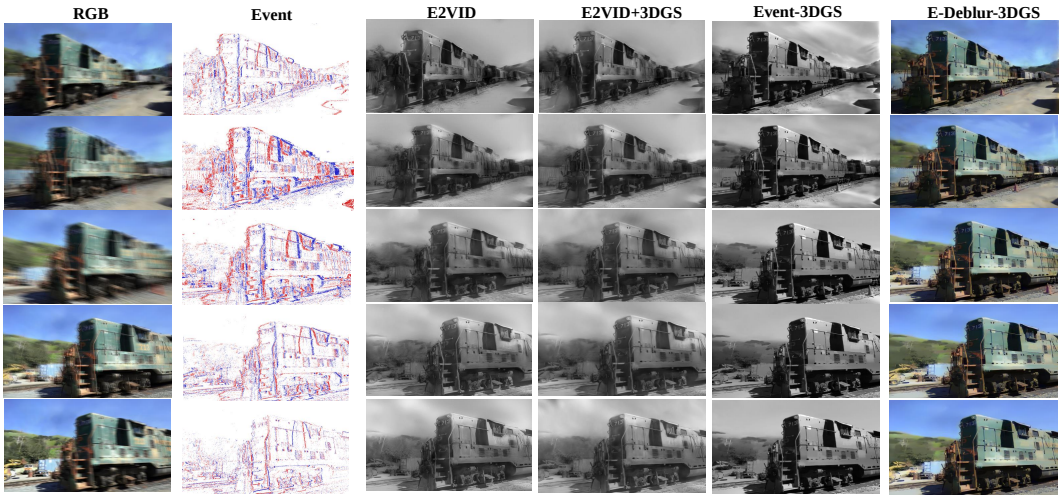

Figure 6: Representative visualization examples of motion deblurring. Note that, our Event-3DGS can be extended for high-quality hybrid reconstruction using events and frames with motion blur.

Event-3DGS for Color Reconstruction. In general, adding color information to 3D reconstructed images is crucial for visual appeal and downstream applications. To achieve this, we extend Event-3DGS from a single channel to three channels to enable color reconstruction. As illustrated in Fig 5, we selected a video sequence [15] and utilized the VOLT simulator [22] to convert the RGB channels into events. Using our method framework, we jointly reconstructed these three channels. The results in Fig. 5 demonstrate that our method can achieve high-quality color reconstruction.

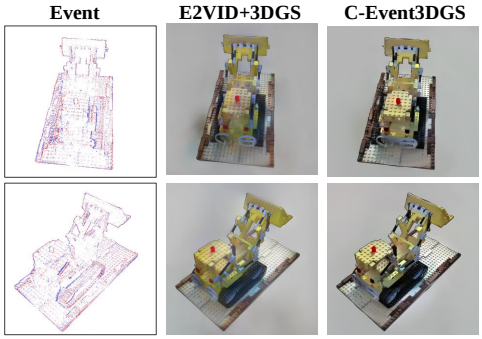

Figure 5: Representative examples of colorful event-based 3D reconstruction.

Limitation. our method rendering module's adaptive threshold learns a threshold for each scene but doesn't account for variations within the same scene over time. Additionally, our current 3DGS lacks support for dynamic scenarios, where 4DGS may be a solution. Future research will address these limitations to enhance Event-3DGS practicality.

5 Conclusion

This paper introduces Event-3DGS, a pioneering event-based 3D reconstruction framework that utilizes 3D Gaussian Splatting (3DGS) to directly process event streams for synthesizing novel views. We present a high-pass filter-based photovoltage estimation module to effectively reduce noise in event data, enhancing the robustness of our method in real-world scenarios. Additionally, we design an event-based 3D reconstruction loss to optimize the parameters of our method. Our results demonstrate that our method surpasses state-of-the-art methods in both reconstruction quality and computational speed on simulated and real-world datasets. We also verify that our method can perform robust 3D reconstruction even in real-world cases with extreme noise, fast motion, and low-light conditions. We believe that our method establishes a new benchmark for using 3DGS with event data, paving the way for high-quality, efficient, and robust 3D reconstruction in challenging real-world scenarios.

Table 5: The influence of the Parameter β.

β	0	0.1	0.15	0.2	0.25	0.5	0.75	1
SSIM	0.405	0.418	0.416	0.459	0.459	**0.497**	0.430	0.477
PSNR	8.523	8.874	8.904	10.754	10.688	**12.448**	9.508	11.441
LPIPS	0.324	0.310	0.310	0.285	0.280	**0.261**	0.296	0.271

https://doi.org/10.52202/079017-4069

Acknowledgements

This work was supported by National Natural Science Foundation of China under Grant 61827804, 62131011.

References

[1] Anish Bhattacharya, Ratnesh Madaan, Fernando Cladera, Sai Vemprala, Rogerio Bonatti, Kostas Daniilidis, Ashish Kapoor, Vijay Kumar, Nikolai Matni, and Jayesh K Gupta. Evdnerf: Reconstructing event data with dynamic neural radiance fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5846–5855, 2024.

[2] Marco Cannici and Davide Scaramuzza. Mitigating motion blur in neural radiance fields with events and frames. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[3] João Carneiro, Sio-Hoi Ieng, Christoph Posch, and Ryad Benosman. Event-based 3d reconstruction from neuromorphic retinas. *Neural Networks*, 45:27–38, 2013.

[4] Zilong Chen, Feng Wang, and Huaping Liu. Text-to-3d using gaussian splatting. In *arXiv*, 2023.

[5] Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3d gaussian splatting in few-shot images. In *arXiv*, 2023.

[6] Gaole Dai, Zhenyu Wang, Qinwen Xu, Wen Cheng, Ming Lu, Boxing Shi, Shanghang Zhang, and Tiejun Huang. Spikenvs: Enhancing novel view synthesis from blurry images via spike camera. In *arXiv*, 2024.

[7] Burak Ercan, Onur Eker, Canberk Saglam, Aykut Erdem, and Erkut Erdem. Hypere2vid: Improving event-based video reconstruction via hypernetworks. *IEEE Transactions on Image Processing*, 2024.

[8] Anis Farshian, Markus Götz, Gabriele Cavallaro, Charlotte Debus, Matthias Nießner, Jón Atli Benediktsson, and Achim Streit. Deep-learning-based 3-d surface reconstruction—a survey. *Proceedings of the IEEE*, 2023.

[9] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2020.

[10] Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. In *arXiv*, 2022.

[11] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *arXiv*, 2023.

[12] Jesse Hagenaars, Federico Paredes-Vallés, and Guido De Croon. Self-supervised learning of event-based optical flow with spiking neural networks. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 7167–7179, 2021.

[13] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014.

[14] Inwoo Hwang, Junho Kim, and Young Min Kim. Ev-nerf: Event based neural radiance field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 837–847, 2023.

[15] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.

[16] Hanme Kim, Ankur Handa, Ryad Benosman, Sio-Hoi Ieng, and Andrew J Davison. Simultaneous mosaicing and tracking with an event camera. In *Proceedings of the British Machine Vision Conference*, pages 1–14, 2014.

[17] Hanme Kim, Stefan Leutenegger, and Andrew J Davison. Real-time 3d reconstruction and 6-dof tracking with an event camera. In *Proceedings of the European Conference on Computer Vision*, pages 349–364, 2016.

[18] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-nerf: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 8(3):1587–1594, 2023.

[19] Jianing Li, Yihua Fu, Siwei Dong, Zhaofei Yu, Tiejun Huang, and Yonghong Tian. Asynchronous spatiotemporal spike metric for event cameras. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4):1742–1753, 2021.

[20] Jianing Li, Jia Li, Lin Zhu, Xijie Xiang, Tiejun Huang, and Yonghong Tian. Asynchronous spatio-temporal memory network for continuous event-based object detection. *IEEE Transactions on Image Processing*, 31:2975–2987, 2022.

[21] Jianing Li and Yonghong Tian. Recent advances in neuromorphic vision sensors: A survey. *Chinese Journal of Computers*, 44(6):1258–1286, 2021.

[22] Songnan Lin, Ye Ma, Zhenhua Guo, and Bihan Wen. Dvs-voltmeter: Stochastic process-based event simulator for dynamic vision sensors. In *Proceedings of the European Conference on Computer Vision*, pages 578–593, 2022.

[23] Xiuhong Lin, Changjie Qiu, Siqi Shen, Yu Zang, Weiquan Liu, Xuesheng Bian, Matthias Müller, Cheng Wang, et al. E2pnet: Event to point cloud registration with spatio-temporal representation learning. In *Proceedings of the Advances in Neural Information Processing Systems*, 2024.

[24] Weng Fei Low and Gim Hee Lee. Robust e-nerf: Nerf from sparse & noisy events under non-uniform motion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18335–18346, 2023.

[25] Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. Deformable neural radiance fields using rgb and event cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3590–3600, 2023.

[26] Sazan Mahbub, Brandon Feng, and Christopher Metzler. Multimodal neural surface reconstruction: Recovering the geometry and appearance of 3d scenes from events and grayscale images. In *Proceedings of the Advances in Neural Information Processing Systems Workshops*, 2023.

[27] Mana Masuda, Yusuke Sekikawa, and Hideo Saito. Event-based camera tracker by nerf. *IEEE Access*, 2023.

[28] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[29] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 36(2):142–149, 2017.

[30] Christoph Posch, Teresa Serrano-Gotarredona, Bernabe Linares-Barranco, and Tobi Delbruck. Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proceedings of the IEEE*, 102(10):1470–1484, 2014.

[31] Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. E2nerf: Event enhanced neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13254–13264, 2023.

[32] Henri Rebecq, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza. Emvs: Event-based multi-view stereo—3d reconstruction with an event camera in real-time. *International Journal of Computer Vision*, 126(12):1394–1414, 2018.

[33] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Proceedings of the Conference on Robot Learning*, pages 969–982, 2018.

[34] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):1964–1980, 2019.

[35] Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. Eventnerf: Neural radiance fields from a single colour event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4992–5002, 2023.

[36] Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki. Dense 3d reconstruction from high frame-rate video using a static grid pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(9):1733–1747, 2014.

 https://doi.org/10.52202/079017-4069

[37] Cedric Scheerlinck, Nick Barnes, and Robert Mahony. Continuous-time intensity estimation using event cameras. In *Proceedings of the Asian Conference on Computer Vision*, pages 308–324, 2018.

[38] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019.

[39] Lisong C Sun, Neel P Bhatt, Jonathan C Liu, Zhiwen Fan, Zhangyang Wang, Todd E Humphreys, and Ufuk Topcu. Mm3dgs slam: Multi-modal 3d gaussian splatting for slam using vision, depth, and inertial measurements. In *arXiv*, 2024.

[40] Fabio Tosi, Youmin Zhang, Ziren Gong, Erik Sandström, Stefano Mattoccia, Martin R Oswald, and Matteo Poggi. How nerfs and 3d gaussian splatting are reshaping slam: A survey. In *arXiv*, 2024.

[41] Sai Vemprala, Sami Mian, and Ashish Kapoor. Representation learning for event-based visuomotor policies. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 4712–4724, 2021.

[42] Jiaxu Wang, Junhao He, Ziyi Zhang, and Renjing Xu. Physical priors augmented event-based 3d reconstruction. In *arXiv*, 2024.

[43] Xiao Wang, Jianing Li, Lin Zhu, Zhipeng Zhang, Zhe Chen, Xin Li, Yaowei Wang, Yonghong Tian, and Feng Wu. Visevent: Reliable object tracking via collaboration of frame and event flows. *IEEE Transactions on Cybernetics*, 2023.

[44] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[45] Ziwei Wang, Yonhon Ng, Cedric Scheerlinck, and Robert Mahony. An asynchronous kalman filter for hybrid event cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 448–457, 2021.

[46] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *arXiv*, 2023.

[47] Zeyu Yang, Hongye Yang, Zijie Pan, Xiatian Zhu, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. In *arXiv*, 2023.

[48] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. *arXiv preprint arXiv:2309.13101*, 2023.

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[50] Xu Zheng, Yexin Liu, Yunfan Lu, Tongyan Hua, Tianbo Pan, Weiming Zhang, Dacheng Tao, and Lin Wang. Deep learning for event-based vision: A comprehensive survey and benchmarks. In *arXiv*, 2023.

[51] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. Semi-dense 3d reconstruction with a stereo event camera. In *Proceedings of the European Conference on Computer Vision*, pages 235–251, 2018.

[52] Lin Zhu, Kangmin Jia, Yifan Zhao, Yunshan Qi, Lizhi Wang, and Hua Huang. Spikenerf: Learning neural radiance fields from continuous spike stream. In *arXiv*, 2024.

A Appendix / Efficiency Advantages over NeRF-based Method

To demonstrate the efficiency of our approach over existing NeRF-based event-based 3D reconstruction methods, we focus on three key aspects: the speed of rasterization compared to ray tracing, the superior parallel support for rasterization in current GPU hardware, and the efficiency of manually derived gradients over neural network backpropagation.

Firstly, the core algorithm of Event-3DGS relies on rendering the 3D Gaussian point cloud using rasterization. Rasterization is inherently faster than ray tracing, which is the method used in NeRF. This speed advantage is crucial for achieving efficient 3D reconstruction.

Secondly, modern GPU hardware is optimized for parallel processing in rasterization. This means that rasterization can be executed with greater efficiency and speed compared to ray tracing, which requires more computational resources and time to process each ray individually. The ability to leverage the parallel processing capabilities of GPUs allows Event-3DGS to perform more efficiently.

Lastly, Event-3DGS utilizes manually derived gradients for optimization, bypassing the complex and time-consuming process of neural network backpropagation used in NeRF. This not only reduces computational overhead but also speeds up the overall process, enabling faster and more efficient 3D reconstruction.

In summary, due to the faster nature of rasterization, better parallel support on GPUs, and the efficiency of manual gradient derivation, Event-3DGS is significantly more efficient than existing NeRF-based methods. This allows for high-quality reconstruction results to be achieved in less time.

B Appendix / The Principle of Event-Based Deblurring based on Event-3DGS

To achieve Event-Based Deblurring using Event-3DGS (E-Deblur-3DGS), the key is to associate blurred frames with event data. For a blurry frame camera, its imaging principle can be described as:

$$Y(p,t,C) = \int_{t-t_e}^{t+t_e} \int \lambda \cdot QE_C(\lambda) \cdot L(p,t_i,\lambda) \cdot d\lambda dt_i, \tag{17}$$

where C represents an additional color channel, typically red, green, and blue. QE_C denotes the response magnitude of this channel to different wavelengths of light, while t_e stands for the camera's exposure time, which is the primary cause of motion blur.

Comparing Equation 17 with Equation 4, we see that RGB cameras, while prone to blurring due to their exposure time, offer rich color information. Conversely, event cameras, with their extremely fast response times, capture almost no motion blur but only single-channel spectral information. By combining these strengths through a loss function, motion deblurring can be achieved.

For the event data, we directly apply the loss described in the main text. For the blurry camera, neglecting unit influence, we can sample the output of 3DGS, perform numerical integration, and compare it with the blurry camera's result, expressed as:

$$l_{blur}^{t_1} = \sum_p \frac{|Y(p,t_1,\cdot) - (\int_{t_1-t_e}^{t_1+t_e} G(T(t))dt)|}{W*H}, \tag{18}$$

where W and H are the image's width and height, $T(t)$ is the camera pose at time t, and $G(T(t))$ is the 3DGS render result. If the exposure time is not too long and the motion speed is not too fast, the integral can be approximated using numerical integration with a single sample.

Finally, combining the above formulas, the resulting loss is:

$$l_{E-Deblur}(t_1,t_2) = (1-\alpha_{blur}) \cdot l_{all}(t_1,t_2) + \alpha_{blur} \cdot l_{blur}, \tag{19}$$

where α_{blur} is a hyperparameter controlling the weight of l_{blur}. By calculating the gradient with this loss and jointly optimizing the scene and sensor parameters, a clear color image can be successfully reconstructed.

 https://doi.org/10.52202/079017-4069

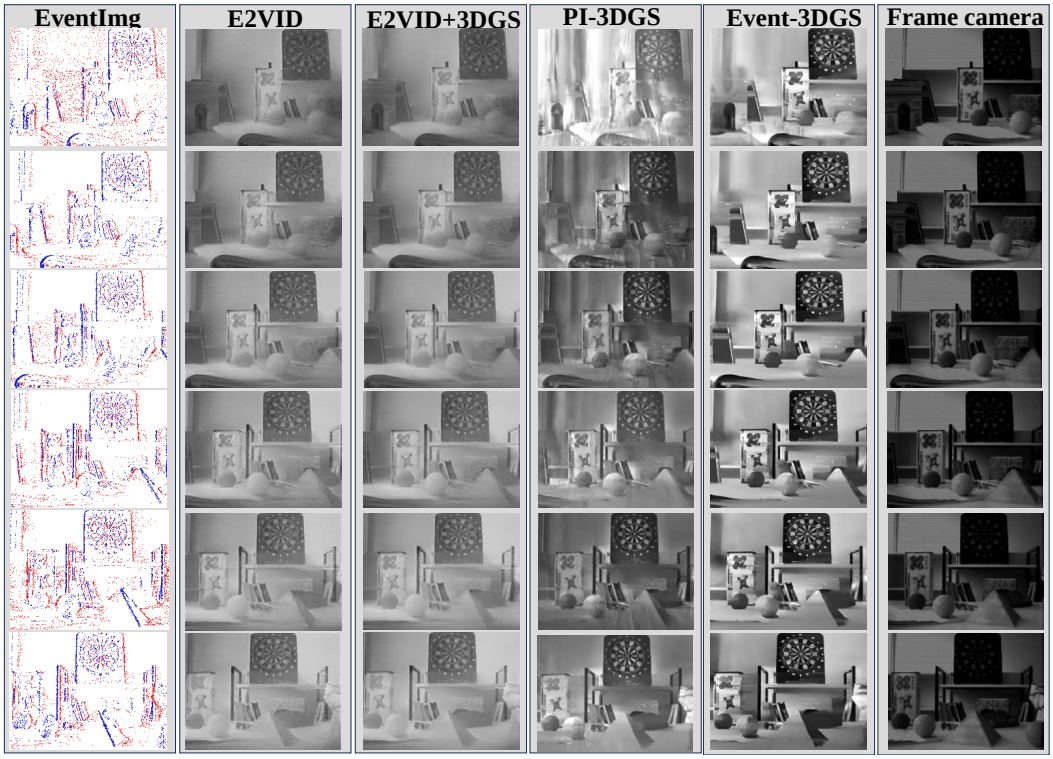

Figure 7: Representative visualization results series the real-world Event-Camera dataset[29]. Our Event-3DGS method produces clearer results compared to other methods.

C Appendix / The Principle of Event-3DGS for Color Reconstruction

Since conventional DVS cameras lack color information capture capability, a new sensor type capable of acquiring three-channel events is imperative. Its operational principle is defined as:

$$I_c(p,t,C) = \int \lambda \cdot QE_C(\lambda) \cdot L(p,t,\lambda) \cdot d\lambda, \tag{20}$$

where C signifies an additional color channel, typically denoted as red, green, and blue. QE_C denotes the magnitude of this channel's response to various light frequencies.

Expanding our method algorithm to C-Event-3DGS merely requires extending the loss function. The approach involves computing the loss for each channel's event using the Event-3DGS method and then aggregating these losses to derive the final loss:

$$l_{C-all}(t_1,t_2) = \alpha_R \cdot l_{all}(t_1,t_2)^R + \alpha_G \cdot l_{all}(t_1,t_2)^G + \alpha_B \cdot l_{all}(t_1,t_2)^B, \tag{21}$$

where $l_{all}(t_1,t_2)^R$, $l_{all}(t_1,t_2)^G$, and $l_{all}(t_1,t_2)^B$ represent the loss calculated for each channel's event data using the original Event-3DGS algorithm. The parameters α_R, α_G, and α_B are hyper-parameters adjusting the weights of each channel. With this loss formulation, C-Event-3DGS can reconstruct a color scene solely based on three-channel event data input. Note that in actual color event cameras, the Bayer pattern is often used to achieve the color effect. Therefore, additional post-processing is required during use, specifically interpolation to obtain the three-channel intensity differences

D Appendix / Additional Experiment Result

Visualization Results on Real Data. To further showcase the reconstruction prowess of Event-3DGS, we present the reconstruction outcomes from various camera poses within the same scene, denoted as "slider_depth". In Fig. 7, the results illustrate the real Event data reconstruction capabilities of Event-3DGS. Each row in the figure corresponds to the camera positioned in the same pose. As we progress from top to bottom, the camera gradually moves from left to right. Remarkably, Event-3DGS consistently upholds high-quality image reconstruction across a spectrum of poses, underscoring its robust 3D reconstruction capabilities.

Comparison with Ev-NeRF[14]. To highlight the accuracy superiority of our method over existing approaches, we conducted a comparative analysis with the Ev-NeRF method using a real dataset. Table 6 presents the performance evaluation of our method and Ev-NeRF in three typical real-world scenarios. Observing the results, it's evident that our method consistently outperforms Ev-NeRF across all metrics in the real dataset. This substantial improvement across various scenarios strongly attests to the superiority of our approach.

Table 6: Experiment results on real dataset between Ev-NeRF and our Event-3DGS. Our Event-3DGS outperforms Ev-NeRF in all metrics.

	Ev-NeRF			Ours Event-3DGS		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
office_zigzag	0.415	14.559	0.275	0.430	14.043	0.183
boxes	0.470	13.979	0.320	0.575	17.696	0.259
Dynamic_6dof	0.260	7.100	0.420	0.212	11.404	0.307
Average	0.382	11.879	0.338	**0.406**	**14.381**	**0.250**

Impact of Training Iterations on Performance and Time. To empirically demonstrate the time complexity of our proposed Event-3DGS, we investigate the impact of training iterations on its performance and time requirements. Table 7 presents the performance and time metrics of Event-3DGS using synthetic data. The "initial" stage in the table indicates the phase when the parameter α is set to zero. Following this stage, α is adjusted to 0.99, and training proceeds consistently until 7999 iterations. The table illustrates that Event-3DGS achieves high-quality reconstruction results in approximately five minutes. This indicates its potential for real-time applications.

Table 7: Experiment results on performance and time of different training iterations

iterations	initial	999	1999	2999	3999	4999	5999	6999	7999
SSIM	0.9493	0.9528	0.9544	**0.9549**	0.9546	0.9545	0.9539	0.9530	0.9534
PSNR	23.8661	26.5129	27.0038	27.1565	27.2864	27.2990	**27.3594**	27.1334	27.0929
LPIPS	0.0500	0.0448	0.0422	**0.0414**	0.0428	0.0450	0.0486	0.0509	0.0500
Running time(s)	123	141	220	267	314	361	407	453	500

https://doi.org/10.52202/079017-4069

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We claim the contributions of this work at the end of the Introduction, and elaborate on how we achieve them in the Methodology.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss about the limitations in the 9th page.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We make the assumption and proof in the method section.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We present the detailed experimental settings and the code in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

https://doi.org/10.52202/079017-4069

Answer: [Yes]

Justification: We have provided the code on the supplementary material.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We have present the detailed experimental settings.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We use the statistical significance in the evaluation metrics.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have described the computer resource in the experimental settings.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: There is no concern about ethics involved in this work.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no concern about broader impacts involved in this work.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

https://doi.org/10.52202/079017-4069

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: There is no concern about safeguards involved in this work.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We credit all works cited including dataset, code and model in the references.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We provide new asserts in the appendix.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/079017-4069