

Event-boosted Deformable 3D Gaussians for Fast Dynamic Scene Reconstruction

Wenhai Xu

Wenming Weng

Yueyi Zhang

Ruikang Xu

Zhiwei Xiong

University of Science and Technology of China

{wh-xu, wmweng, xurk}@mail.ustc.edu.cn, {zhyuey, zwxiong}@ustc.edu.cn

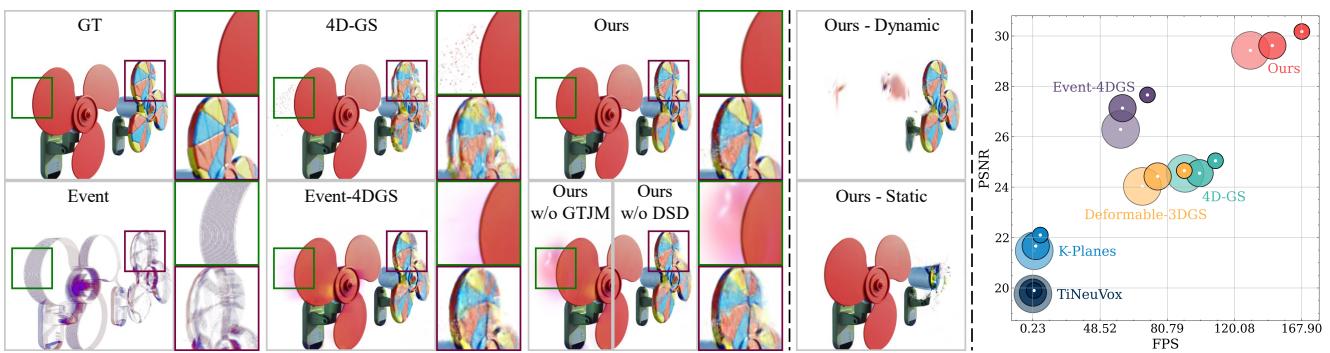


Figure 1. **Left:** Quality comparison with the baselines 4D-GS [40], Event-4DGS (the event-extended version of [44]), and our variants, highlighting the superior rendering quality of our method. Our GS-threshold joint modeling (GTJM) effectively eliminates event-induced artifacts (particularly the purple haze), while our dynamic-static decomposition (DSD) improves the quality of dynamic regions. **Middle:** Separate renderings of dynamic and static Gaussians from our DSD. **Right:** The scatter plot illustrates our method’s ability to achieve both high fidelity and fast rendering, where dot radii correspond to different resolutions (400×400 , 600×600 , and 800×800).

Abstract

3D Gaussian Splatting (3D-GS) enables real-time rendering but struggles with fast motion due to low temporal resolution of RGB cameras. To address this, we introduce the first approach combining event cameras, which capture high-temporal-resolution, continuous motion data, with deformable 3D-GS for fast dynamic scene reconstruction. We observe that threshold modeling for events plays a crucial role in achieving high-quality reconstruction. Therefore, we propose a GS-Threshold Joint Modeling (GTJM) strategy, creating a mutually reinforcing process that greatly improves both 3D reconstruction and threshold modeling. Moreover, we introduce a Dynamic-Static Decomposition (DSD) strategy that first identifies dynamic areas by exploiting the inability of static Gaussians to represent motions, then applies a buffer-based soft decomposition to separate dynamic and static areas. This strategy accelerates rendering by avoiding unnecessary deformation in static areas, and focuses on dynamic areas to enhance fidelity. Our approach achieves high-fidelity dynamic reconstruction at 156 FPS with a 400×400 resolution on an RTX 3090 GPU.

1. Introduction

Dynamic scene reconstruction and novel view synthesis are essential for immersive applications in virtual/augmented reality and entertainment [20, 23, 35, 40, 43, 44, 46]. While Neural Radiance Fields (NeRF) [2, 26, 37] offer unprecedented photorealism, they are constrained by slow training and rendering speeds. Despite recent advances in optimization techniques [4, 7, 9, 10], real-time rendering remains elusive. Recently, 3D Gaussian Splatting (3D-GS) [17] addresses this limitation through efficient differentiable rasterization, yet existing dynamic extensions [24, 40, 43, 44] struggle with fast motions due to RGB cameras’ inherent limitations of low frame rates and motion blur.

In this paper, we present event-boosted deformable 3D Gaussians for fast dynamic scene reconstruction. Event cameras [11, 19], with their microsecond-level temporal resolution, can capture rapid scene changes that traditional RGB cameras often miss. This capability makes them particularly valuable for fast dynamic scene reconstruction, as they provide continuous motion information and near-infinite viewpoints between sparse RGB frames.

However, integrating events into 3D scene reconstruction faces new challenges. Specifically, event supervision

for 3D-GS relies on an accurate event generation model [3, 18], where the threshold undergoes complex variations across polarity, space, and time [6, 18]. Previous methods [3, 5, 18, 25, 30, 32, 41, 42, 45] adopt a constant threshold, yet this simplification significantly degrades the quality of event supervision. While recent works [15, 21] attempt to model threshold variations using event data alone, they achieve limited success due to the inherent binary nature of events, which only indicate intensity change directions. To address this challenge, we propose a novel GS-threshold joint modeling strategy. First, we leverage the intensity change values from RGB frames to estimate the initial threshold. Second, since the sparsity of RGB frames weakens supervision, we use 3D-GS rendered results as pseudo-intermediate frames to enhance the supervision. This finally creates a mutually reinforcing process where RGB-optimized threshold enables better event supervision for 3D-GS, while improved 3D-GS in turn provides accurate geometric constraints for threshold refinement.

Furthermore, we observe that existing dynamic 3D-GS methods inefficiently model both static and dynamic regions with dynamic Gaussians [1, 13, 14, 22, 23, 40, 44]. This unified treatment leads to reduced rendering speed, wasted deformation field capacity, and degraded reconstruction quality. While some methods have explored dynamic-static decomposition, they are limited by either inaccurate dynamic Gaussians initialization [20] or constraints in multi-view scenarios [35]. To address these limitations, we propose a novel dynamic-static decomposition strategy that first identifies dynamic regions based on the inherent inability of static Gaussians to represent motions, and then employs a buffer-based soft decomposition to adaptively search for the optimal decomposition boundary. This decomposition not only accelerates rendering by eliminating unnecessary deformation computations in static regions but also enhances reconstruction quality by focusing the deformation field exclusively on dynamic regions.

Our main contributions can be summarized as follows:

- We present the first method integrating event cameras with deformable 3D-GS for fast dynamic scene reconstruction, enabling high-fidelity and fast rendering.
- We propose a novel GS-threshold joint modeling strategy that combines RGB-assisted initial estimation with GS-boosted refinement, creating a mutually reinforcing process that significantly improves both threshold modeling and 3D reconstruction.
- We introduce an effective dynamic-static decomposition strategy that not only accelerates rendering through selective deformation computation but also enhances reconstruction quality by focusing on dynamic regions.
- We contribute a comprehensive dynamic scene dataset along with its creation pipeline to facilitate future research in event-based vision.

2. Related Work

Neural Rendering for Dynamic Scenes. Neural rendering techniques have revolutionized dynamic scene reconstruction in recent years. Pioneering works like D-NeRF [29] and Nerfies [28] extend Neural Radiance Fields [26] through deformation field, mapping observations into a canonical space for modeling non-rigid motion. Despite their impressive reconstruction quality, these methods are constrained by extensive computational demands due to dense MLP evaluations during training and rendering. Various acceleration strategies have been proposed to address these limitations. K-Planes [10] introduces an efficient explicit representation using six feature planes, while Tensor4D [34] and DTensoRF [16] employ tensor decomposition techniques to achieve compact spatiotemporal encoding. A significant breakthrough came with 3D Gaussian Splatting (3D-GS) [17], which leverages efficient differentiable rasterization for real-time rendering. This advancement has spawned several dynamic scene extensions, including 4D-GS [40], Deformable-3DGs [44], and related works [23, 35], which successfully achieve real-time rendering for high-quality dynamic scene reconstruction.

Event-based Neural Rendering. The integration of neural representations with event-based 3D reconstruction has emerged as a promising research direction. Pioneering works like EventNeRF [32], E-NeRF [18], and Ev-NeRF [15] first demonstrated the potential of pure event-based static scene reconstruction, albeit with different assumptions about event camera characteristics. Building upon these, Robust e-NeRF [21] advanced the field by modeling spatially-varying threshold and refractory period, enhancing robustness against event noise. To further improve reconstruction quality, E2NeRF [30] and Ev-DeblurNeRF [3] incorporated blurry RGB images alongside events. A significant milestone was achieved by DE-NeRF [25], which pioneered the combination of events and RGB frames for dynamic scene reconstruction. The recent advent of 3D Gaussian Splatting [17] has catalyzed new developments in event-based methods. While Ev-GS [41] adapted the pure event-based paradigm to 3D-GS, subsequent works including E2GS [5], EaDeblur-GS [39], and Event3DGs [42] primarily addressed deblurring challenges. Our work distinguishes itself as the first to combine events with 3D-GS for monocular fast dynamic scene reconstruction, enabling high-fidelity and fast rendering.

3. Method

3.1. Event Cameras for 3D-GS

3D Gaussian Splatting Preliminary. 3D Gaussian Splatting (3D-GS) [17] represents a scene as anisotropic 3D Gaussians, each characterized by a covariance matrix Σ and center position μ : $GS(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}$. The co-

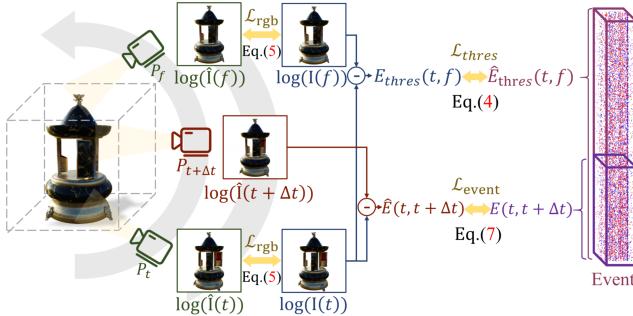


Figure 2. Overview of GS-threshold joint modeling strategy. \mathcal{L}_{rgb} optimizes 3D-GS, \mathcal{L}_{thres} optimizes the threshold, and \mathcal{L}_{event} jointly optimizes both 3D-GS and threshold.

variance matrix Σ is parameterized using scaling matrix S and rotation matrix R to ensure positive semi-definiteness: $\Sigma = RSS^T R^T$. Each Gaussian is further defined by spherical harmonic coefficients C and opacity σ . Final pixel colors c are computed using differentiable tile-based rasterization

$$c = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (1)$$

where c_i denotes spherical harmonic color, and α_i combines opacity σ with projected $GS(\mathbf{x})$.

Event/RGB Rendering Loss. Event cameras [11, 19] are novel sensors, that asynchronously capture pixel-wise intensity changes with microsecond-level temporal resolution. Their high temporal precision enables capturing crucial motions between RGB frames and provides near-infinite viewpoint supervision, making them ideal for monocular fast dynamic scene reconstruction.

Each event is represented as $e_{xy}(\tau) = p\delta(\tau)$, where (x, y) is pixel position, τ is timestamp, $p \in \{+1, -1\}$ indicates intensity change direction relative to threshold C , and $\delta(t)$ is a unit integral impulse function. Omitting pixel subscripts, the intensity change over interval Δt can be formulated as

$$E(t, t + \Delta t) = \int_t^{t + \Delta t} C \cdot e(\tau) d\tau. \quad (2)$$

This change can also be estimated from rendered intensities:

$$\hat{E}(t, t + \Delta t) = \log(\hat{I}(t + \Delta t)) - \log(I(t)). \quad (3)$$

where \hat{I} and I respectively denote rendered and ground truth intensities. The event rendering loss is

$$\mathcal{L}_{event} = \|E(t, t + \Delta t) - \hat{E}(t, t + \Delta t)\|_2^2. \quad (4)$$

Similarly, we utilize the RGB rendering loss [17] combining L1 and D-SSIM losses as

$$\mathcal{L}_{rgb} = (1 - \lambda_s) \|\hat{I}(t) - I(t)\|_1 + \lambda_s \mathcal{L}_{D-SSIM}(\hat{I}(t), I(t)), \quad (5)$$

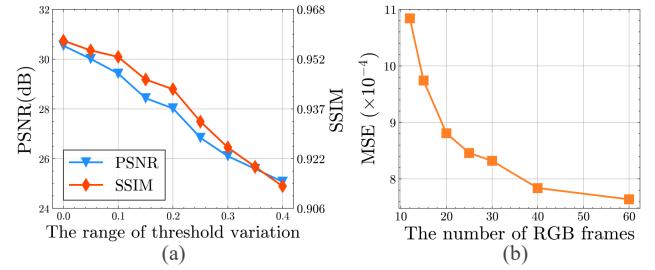


Figure 3. (a) The effect of different ranges of threshold variation for 3D reconstruction. (b) The effect of different number of RGB frames for threshold estimation.

where λ_s is a weighting factor that controls the balance. 3D-GS is optimized jointly with \mathcal{L}_{event} and \mathcal{L}_{rgb}

$$GS^* = \underset{GS}{\operatorname{argmin}} (\mathcal{L}_{event} + \mathcal{L}_{rgb}). \quad (6)$$

3.2. GS-threshold Joint Modeling

As shown in Eq. (2), the threshold C critically affects event integration and supervision quality. While existing methods [3, 18, 25, 30, 32] typically assume a constant threshold, real event cameras exhibit threshold variations across polarity, space, and time [6, 18]. Fig. 3 (a) demonstrates how increasing threshold variation significantly degrades reconstruction quality. To model threshold variations, we propose a *GS-threshold joint modeling (GTJM) strategy* (see Fig. 2), consisting of RGB-assisted threshold estimation and GS-boosted threshold refinement.

RGB-assisted Threshold Estimation. Recent works attempt to model threshold variations solely from event data but face significant limitations due to the difficulty of inferring intensity changes from events, which only indicate the direction of change. We propose to leverage intensity change values from RGB frames for robust threshold estimation. Given two RGB frames $I(t)$ and $I(f)$ at times t and f , we define the threshold modeling loss as

$$\mathcal{L}_{thres} = \left\| E_{thres}(t, f) - \hat{E}_{thres}(t, f) \right\|_2^2, \quad (7)$$

where $\hat{E}_{thres}(t, f) = \int_t^f C \cdot e(\tau) d\tau$, $E_{thres}(t, f) = \log(I(f)) - \log(I(t))$. In practice, we adopt a simple way to compute $\hat{E}_{thres}(t, f)$. We first accumulate events to obtain event count maps [12] $ECM_{t,f} \in \mathbb{R}^{B \times P \times H \times W}$, where B and P denote time bins and event polarity respectively. Using learnable threshold parameters $C_{t,f} \in \mathbb{R}^{B \times P \times H \times W}$, we compute

$$\hat{E}_{thres}(t, f) = \sum_{B,P} ECM_{t,f} \odot C_{t,f}, \quad (8)$$

where threshold $C_{t,f}$ is optimized by minimizing Eq. (7) in an end-to-end manner.

Table 1. Step-by-step validation of mutual boosting between threshold modeling (TM) and 3D reconstruction (3D Rec.). Abbreviations: “Fro.”: “Frozen”; “Ft.”: “Fine-tuning”.

TM for 3D Rec.		3D Rec. for TM			
Stage1	Stage2	3D Rec.	Stage1	Stage2	TM
TM	3D Rec. (Fro. C)	PSNR↑	3D Rec.	TM (Fro. GS)	MSE↓ ($\times 10^{-4}$)
×	✓	24.46	×	✓	8.317
✓	✓	26.63	✓	✓	7.077

Joint TM and 3D Rec. Optimization					
Stage1	Stage2	3D Rec.	TM		
TM	3D Rec. & TM (Ft. C & Ft. GS)	PSNR↑	MSE↓ ($\times 10^{-4}$)		
✓	✓	28.01		6.322	

We observe that *accurate threshold modeling improves 3D reconstruction quality*. As shown in “**TM for 3D Rec.**” in Tab. 1, our RGB-assisted threshold optimization approach benefits threshold estimation, thus significantly enhancing 3D-GS reconstruction quality and achieving a 2.17 dB PSNR improvement.

GS-boosted Threshold Refinement. While RGB frames facilitate threshold estimation, their effectiveness is constrained by low frame rate. As illustrated in Fig. 3 (b), sparse RGB frames lead to longer integration intervals, reducing supervision quality and threshold estimation accuracy. To overcome this limitation, we found that once a 3D-GS is trained first by Eq. (4) and Eq. (5), it can be used to render intermediate frames as additional pseudo-supervision. Specifically, we freeze the trained 3D-GS and reuse Eq. (4) to enhance Eq. (7) for optimizing threshold C

$$C^* = \underset{C}{\operatorname{argmin}}(\mathcal{L}_{thres} + \mathcal{L}_{event}). \quad (9)$$

We observe that *the incorporation of 3D-GS significantly enhances threshold estimation accuracy*. The underlying reason is that events may provide unreliable supervision in some regions with inaccurate thresholds or noise, whereas 3D-GS can correct these errors via geometric consistency. As demonstrated in “**3D Rec. for TM**” part in Tab. 1, using trained and frozen 3D-GS for threshold modeling substantially reduces MSE, leading to more precise threshold estimation.

Joint Threshold and GS Optimization. Having demonstrated the mutual benefits between threshold modeling and 3D reconstruction, we propose jointly optimizing both threshold C and 3D Gaussians GS through

$$C^*, GS^* = \underset{C, GS}{\operatorname{argmin}}(\mathcal{L}_{thres} + \mathcal{L}_{event} + \mathcal{L}_{rgb}). \quad (10)$$

We observe that *this joint optimization enables a beneficial cycle where optimized thresholds enhance event su-*

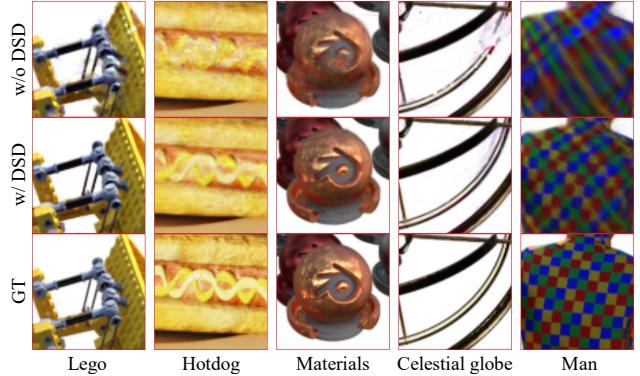


Figure 4. The effect of dynamic-static decomposition strategy, which improves the rendering quality of dynamic regions.

pervision for 3D-GS, while improved 3D-GS refines threshold estimates through geometric consistency. As shown in “**Joint TM and 3D Rec. Optimization**” part in Tab. 1, this approach achieves superior threshold modeling and reconstruction quality.

In summary, our optimization proceeds in two stages: first optimizing the threshold using \mathcal{L}_{thres} , then jointly optimizing both threshold and 3D-GS using all three losses \mathcal{L}_{thres} , \mathcal{L}_{event} , and \mathcal{L}_{rgb} .

3.3. Dynamic-static Decomposition

Dynamic scenes typically contain substantial static regions (e.g., tables, walls) that require no deformation. Unlike existing methods [1, 13, 14, 22, 23, 40, 44] that use dynamic Gaussians throughout, we separately model dynamic and static regions with corresponding Gaussian types. This decomposition offers dual benefits: accelerated rendering by bypassing deformation field computation for static Gaussians, and enhanced deformation fidelity through focused MLP capacity optimization for dynamic regions, as demonstrated in Fig. 4.

The key challenge lies in accurately initializing Gaussians in their respective regions. We address this through a proposed *dynamic-static decomposition (DSD) strategy*, as illustrated in Fig. 5.

Dynamic-static Decomposition on 2D. We leverage the inherent limitation of static Gaussians in representing motion to decompose dynamic and static regions in 2D images. During the first 3k iterations, we perform scene reconstruction using only static Gaussians for initialization. This naturally results in poor reconstruction in dynamic regions while achieving high fidelity in static areas (illustrated in Fig. 5 (a)). This distinct performance difference enables effective decomposition of rendered images into dynamic and static regions.

Specifically, using a pretrained VGG19 [36] network \mathcal{F}_ϕ , we extract multi-scale features from both rendered image

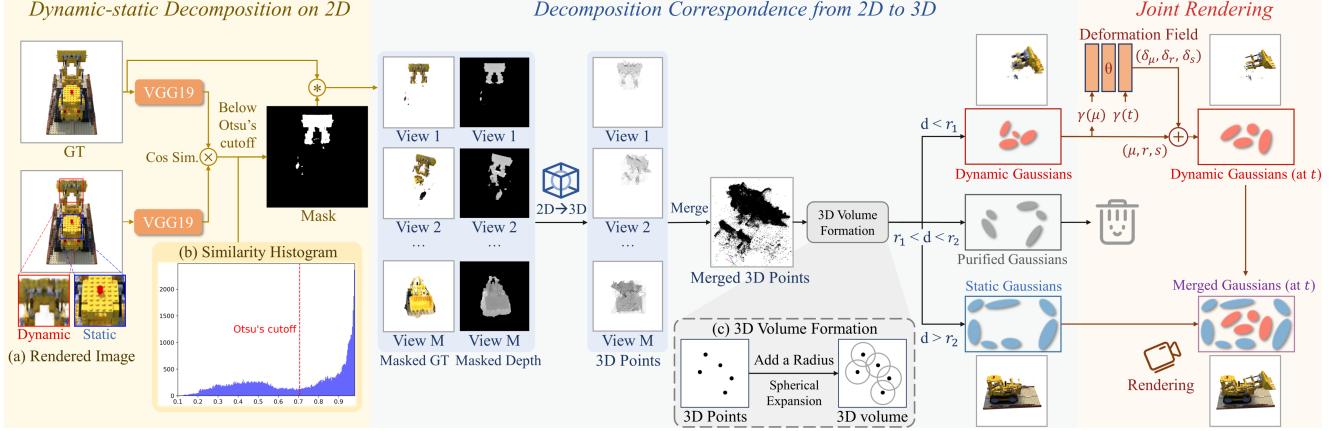


Figure 5. Overview of dynamic-static decomposition strategy. First, we decompose dynamic and static regions in 2D images based on the inherent inability of static Gaussians to represent motions. Next, we establish a correspondence to extend 2D decomposition to 3D Gaussians. Finally, the decomposed dynamic and static Gaussians are jointly rendered to reconstruct the complete dynamic scene.

$\hat{I}(t)$ and ground truth $I(t)$. The cosine similarities computed at each scale are upsampled to a uniform resolution and averaged to generate a fused similarity map

$$Sim = \sum_l Up \left(\frac{\mathcal{F}_\phi^l(\hat{I}(t)) \cdot \mathcal{F}_\phi^l(I(t))}{\|\mathcal{F}_\phi^l(\hat{I}(t))\| \|\mathcal{F}_\phi^l(I(t))\|} \right), \quad (11)$$

where $\mathcal{F}_\phi^l(\cdot)$ represents the l -th layer output of VGG19, and $Up(\cdot)$ indicates bilinear upsampling. The histogram of the resulting similarity map exhibits a bimodal distribution as shown in Fig. 5 (b), enabling dynamic region mask generation through Otsu’s [27] method

$$Mask = \mathbf{1}_{Sim < Otsu(Sim)}, \quad (12)$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, which returns 1 if the condition is true. The mask is then multiplied by the ground truth image to extract the dynamic region.

Decomposition Correspondence from 2D to 3D. To extend the 2D dynamic-static decomposition to 3D Gaussians, we establish view-independent correspondences by leveraging the depth information obtained during 3D-GS rendering. By unprojecting pixels from masked dynamic regions across multiple views and merging the resulting 3D points, we obtain a comprehensive representation of dynamic regions in 3D space.

Next, we map merged points to dynamic Gaussians based on spatial proximity. Each point expands spherically with radius r to form a 3D volume (Fig. 5 (c)), initially classifying enclosed Gaussians as dynamic and others as static. To overcome potential decomposition inaccuracies and radius sensitivity, we implement a *buffer-based soft decomposition strategy* using two radii, r_1 and r_2 . Gaussians within r_1 are marked as dynamic, beyond r_2 as static, while those

between are pruned to create a buffer zone. This strategy enables 3D-GS to optimize decomposition boundaries through adaptive density control (ADC) [17], enhancing both rendering quality and speed. As demonstrated in Fig. 8, the strategy also exhibits improved robustness to radius parameter selection.

It should be noted that, our DSD method is performed only once during the entire training process and requires only about one minute, introducing minimal overhead to the training pipeline.

Joint Rendering of Dynamic and Static Gaussians. With the decomposed dynamic and static Gaussians, we jointly render the entire dynamic scene. Particularly, a deformation field [44] learns to map dynamic Gaussians from canonical space to arbitrary time. Taking time t and the center position μ of dynamic Gaussians as inputs, the deformation field outputs the displacement of their position δ_μ , rotation δ_r , and scaling δ_s

$$(\delta_\mu, \delta_r, \delta_s) = \mathcal{F}_\theta(\gamma(sg(\mu)), \gamma(t)), \quad (13)$$

Where $sg(\cdot)$ indicates a stop-gradient operation and $\gamma(\cdot)$ denotes the positional encoding [44]. Then, the deformed dynamic gaussians can be addressed as

$$(\mu', r', s') = (\mu + \delta_\mu, r + \delta_r, s + \delta_s). \quad (14)$$

Finally, static Gaussians bypass the deformation field and are merged with the deformed dynamic Gaussians as inputs to the rasterizer, enabling high-frame-rate dynamic rendering.

4. Experiment

4.1. Experimental Settings

Implementation Details. Our method builds upon the differentiable rasterizer of 3D-GS [17] and is implemented in



Figure 6. Qualitative comparisons on our dataset. Please see the supplementary video for details.

PyTorch, running on an NVIDIA RTX 3090 GPU. Given sparse RGB frames, we initialize from random point clouds rather than structure-from-motion (SfM) [33] points. The training process consists of three phases: 7k iterations of RGB-assisted initial estimation, 3k iterations of static Gaussian warm-up, and 47k iterations of joint optimization of all components (threshold, dynamic/static Gaussians, and deformation field) after decomposition. Detailed hyperparameters are provided in the supplementary materials.

Datasets. Current datasets for event-based dynamic scene reconstruction are limited, with only three unpublished scenes [25]. Existing RGB-based datasets [28, 29] are unsuitable for event simulation due to discontinuous [29] or insufficient [28] frames. To address this gap, we introduce a comprehensive benchmark featuring eight dynamic scenes

with varying complexity, intricate structures, and fast motions, suitable for dynamic reconstruction evaluation.

Using Blender [8], we generate one-second, 360° monocular camera rotations at 400×400 resolution, producing thousands of continuous frames per scene. These high-temporal-resolution sequences are processed through ESIM [31] to generate event streams. For training, we sample 30 RGB frames (30 FPS) from each sequence.

Our code, benchmark, and dataset creation pipeline will be publicly released, with detailed scene descriptions available in the supplementary materials.

Baselines. For RGB-only settings, we benchmarked our method against the representative NeRF baselines K-Planes [10] and TiNeuVox [7], along with Gaussians baselines 3D-GS [17], 4D-GS [40], and Deformable-3DGS [44]. For

Table 2. Quantitative results on our dataset with the 400×400 rendering resolution. Event-4DGS is an extension of Deformable-3DGS [44] by incorporating events.

Method	Lego				Hotdog				Materials				Music box			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
3D-GS [17]	23.60	0.918	0.088	223	30.01	0.951	0.064	260	28.07	0.967	0.061	262	19.20	0.905	0.122	239
TiNeuVox [7]	22.39	0.891	0.071	0.53	30.81	0.953	0.035	0.49	26.63	0.938	0.054	0.52	20.45	0.831	0.152	0.62
K-Planes [10]	24.55	0.931	<u>0.035</u>	2.34	31.36	0.958	<u>0.016</u>	2.35	30.62	0.976	0.009	2.29	20.77	0.858	0.071	2.39
4D-GS [40]	26.30	0.937	0.072	104	33.48	0.965	0.052	132	30.40	0.979	0.054	111	24.06	0.937	0.071	64
Deformable-3DGS [44]	23.79	0.923	0.053	73	32.91	0.962	0.017	132	34.00	0.986	<u>0.004</u>	91	22.08	0.924	0.052	51
Event-4DGS	<u>28.00</u>	<u>0.943</u>	0.040	54	<u>34.61</u>	<u>0.969</u>	0.019	96	<u>35.60</u>	<u>0.989</u>	0.006	74	<u>28.58</u>	<u>0.950</u>	<u>0.043</u>	42
Ours	31.85	0.967	0.018	<u>189</u>	36.15	0.974	0.013	<u>241</u>	38.02	0.993	0.003	<u>240</u>	30.78	0.963	0.029	<u>92</u>
Method	Celestial globe				Fan				Water wheel				Man			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
3D-GS [17]	19.05	0.915	0.110	182	21.26	0.891	0.118	270	19.43	0.887	0.109	215	20.79	0.870	0.114	210
TiNeuVox [7]	13.62	0.736	0.290	0.62	19.90	0.889	0.107	0.53	17.03	0.850	0.147	0.56	22.81	0.887	0.071	0.50
K-Planes [10]	15.49	0.857	0.088	2.46	22.10	0.909	0.062	2.38	20.96	0.920	<u>0.046</u>	2.41	21.02	0.857	0.073	2.33
4D-GS [40]	20.97	0.942	0.072	52	25.05	0.936	0.080	108	20.96	0.917	0.077	77	23.80	0.914	0.076	64
Deformable-3DGS [44]	23.07	<u>0.962</u>	0.036	41	24.66	0.929	0.051	90	20.79	0.912	0.051	43	23.06	0.906	<u>0.051</u>	37
Event-4DGS	<u>24.30</u>	0.948	0.045	36	<u>27.66</u>	<u>0.949</u>	<u>0.041</u>	71	<u>26.34</u>	<u>0.932</u>	0.052	30	<u>25.55</u>	<u>0.921</u>	0.063	33
Ours	28.83	0.976	0.020	<u>73</u>	30.18	0.964	0.025	<u>168</u>	28.47	0.950	0.033	<u>112</u>	28.21	0.943	0.037	<u>129</u>

event-assisted settings, DE-NeRF [25] is the only baseline; however, it could not be directly compared, as its code was still unavailable at the time of writing. DE-NeRF relies on conventional volume rendering techniques [26] and a deformation field without explicit structures [29], leading to predictably slow rendering speeds. Moreover, its reconstruction quality is also predictably limited due to the absence of threshold modeling for events. To provide a comparable baseline, we introduce Event-4DGS, an extension of Deformable-3DGS [44] that incorporates the event rendering loss in Eq. (4) for additional supervision.

Metrics. We evaluate the rendered image quality using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [38], and Learned Perceptual Image Patch Similarity (LPIPS) [47] with an AlexNet backbone, and assess rendering speed in terms of Frames Per Second (FPS).

4.2. Comparisons

The quantitative results of the comparison are summarized in Tab. 2. As shown, 3D-GS [17] fails to reconstruct dynamic scenes due to the absence of the deformation field, while grid-based deformable NeRF methods [7, 10] suffer from slow rendering speeds due to the cumbersome volume rendering. Although 4D-GS [40] and Deformable-3DGS [44] achieve relatively high FPS, their reconstruction quality is limited by the sparsity of the RGB training frames. In contrast, Event-4DGS leverages the rich intermediate motion and viewpoint information provided by events, significantly outperforming other baselines in reconstruction quality, with an average PSNR improvement of 3.28 dB over Deformable-3DGS across all scenes. This notable improvement underscores the importance of high-temporal-resolution event cameras for fast dynamic scene reconstruction. However, Event-4DGS still suffers from threshold

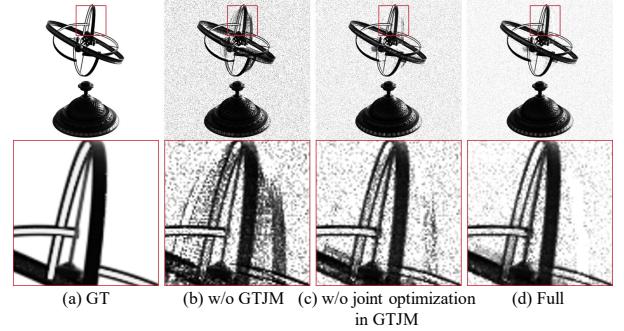


Figure 7. The effect of GS-threshold joint modeling strategy, which eliminates event artifacts caused by threshold variations.

variation, whereas our method with GTJM, enables accurate threshold modeling and better event supervision. Furthermore, our approach substantially improves rendering speed by avoiding unnecessary deformation computations for static Gaussians. In summary, our method achieves the highest rendering quality across all scenes and metrics, with an average PSNR improvement of 2.73 dB over Event-4DGS. Meanwhile, our method maintains exceptionally fast rendering speeds, averaging 1.71 times faster than 4D-GS in terms of FPS as demonstrated in Tab. 2.

For a more visual assessment, we present qualitative results in Fig. 6. As shown, events enable our method and Event-4DGS to capture intricate motion details, while other baselines exhibit structural deficiencies and distortions. However, due to threshold variations, Event-4DGS suffers from event-based artifacts, leading to erroneous purple haze in the rendered outputs (discussed in Sec. 4.3). In contrast, our method significantly mitigates these artifacts through GTJM strategy.

Table 3. Ablation studies on our dataset with the 400×400 rendering resolution.

Method	Lego				Hotdog				Materials				Music box			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
w/o GTJM	29.49	0.955	0.028	177	34.72	0.970	0.017	261	35.72	0.990	0.006	243	28.73	0.952	0.042	79
w/o joint optimization in GTJM	30.92	0.962	0.021	185	35.67	0.973	0.014	238	37.27	0.992	0.004	239	30.15	0.959	0.033	87
w/o DSD	29.62	0.954	0.028	58	35.69	0.972	0.014	104	37.70	0.992	0.004	76	30.29	0.959	0.032	43
w/o buffer-based soft decomposition	30.27	0.959	0.025	168	35.95	0.973	0.015	216	37.31	0.991	0.004	207	30.77	0.963	0.030	82
Full	31.85	0.967	0.018	189	36.15	0.974	0.013	241	38.02	0.993	0.003	240	30.78	0.963	0.029	92
Method	Celestial globe				Fan				Water wheel				Man			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
w/o GTJM	25.06	0.955	0.038	67	28.16	0.955	0.035	169	26.57	0.932	0.051	113	26.67	0.937	0.053	118
w/o joint optimization in GTJM	27.49	0.970	0.025	74	29.54	0.961	0.028	165	28.14	0.945	0.038	110	27.76	0.940	0.041	120
w/o DSD	27.83	0.971	0.024	37	29.73	0.959	0.029	77	27.86	0.948	0.034	31	27.54	0.935	0.040	31
w/o buffer-based soft decomposition	28.16	0.973	0.022	78	29.65	0.959	0.029	153	28.36	0.949	0.033	90	27.67	0.938	0.038	106
Full	28.83	0.976	0.020	73	30.18	0.964	0.025	168	28.47	0.950	0.033	112	28.21	0.943	0.037	129

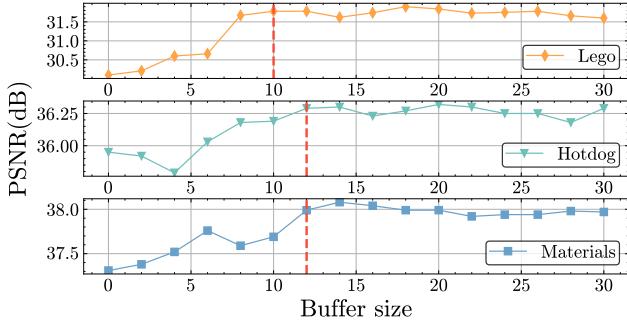


Figure 8. Sensitivity analysis on buffer size ($r_2 - r_1$).

4.3. Ablation Study

GS-threshold Joint Modeling. Using a constant threshold fails to properly neutralize opposing polarity events during accumulation, resulting in motion trajectory artifacts as shown in Fig. 7 (b). These artifacts, when used for Gaussian supervision, produce undesirable purple haze in rendered outputs, such as the Event-4DGS results in Fig. 6. Our RGB-assisted threshold estimation significantly reduces these artifacts (Fig. 7 (c)), while subsequent joint threshold and GS optimization effectively eliminates remaining distortions (Fig. 7 (d)). As demonstrated in Tab. 3, this improved event supervision yields a 2.17 dB average PSNR improvement across all scenes, validating our GTJM strategy’s effectiveness in handling threshold variations.

Dynamic-static Decomposition. Our DSD method successfully identifies dynamic regions of varying sizes and geometries, as demonstrated in Fig. 9. Modeling the entire scene with dynamic Gaussians without DSD misallocates deformation field capacity to static regions, compromising dynamic region reconstruction quality as shown in Fig. 4. Quantitative results in Tab. 3 show that removing DSD leads to a 0.78 dB decrease in average PSNR and reduces rendering speed to 37% of the original FPS. This underscores DSD’s crucial role in achieving both high-fidelity dynamic scene reconstruction and efficient rendering.



Figure 9. Rendering results of dynamic and static Gaussians separated by our dynamic-static decomposition strategy.

Buffer-based Soft Decomposition. Our buffer-based soft decomposition enables adaptive optimization of decomposition boundaries, yielding a 0.54 dB improvement in average PSNR (Tab. 3). Sensitivity analysis reveals that reconstruction quality stabilizes when buffer size ($r_2 - r_1$) exceeds approximately 12 basic units (normalized by average inter-Gaussian distance to account for scene variations), as shown in Fig. 8. This stability demonstrates the robustness of our DSD method through adaptive boundary search, highlighting the effectiveness of the buffer-based strategy.

5. Conclusion and Limitation

In this paper, we present an event-assisted deformable 3D Gaussian framework for high-quality monocular fast dynamic scene reconstruction. Our GS-threshold joint modeling effectively addresses threshold variation challenges, enabling reliable event supervision. The proposed dynamic-static decomposition method enhances both rendering efficiency and reconstruction quality through optimized resource allocation between static and dynamic regions.

Our current evaluation is limited to synthetic data due to the absence of real-world event-based dynamic scene datasets. Future work will focus on establishing a real-world benchmark by using an event camera capturing system to further advance event-based vision research.

References

- [1] Jeongmin Bae, Seoha Kim, Youngsik Yun, Hahyun Lee, Gun Bang, and Youngjung Uh. Per-gaussian embedding-based deformation for deformable 3d gaussian splatting. *arXiv preprint arXiv:2404.03613*, 2024. 2, 4
- [2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021. 1
- [3] 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*, pages 9286–9296, 2024. 2, 3
- [4] Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2023. 1
- [5] Hiroyuki Deguchi, Mana Masuda, Takuya Nakabayashi, and Hideo Saito. E2gs: Event enhanced gaussian splatting. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 1676–1682. IEEE, 2024. 2
- [6] Tobi Delbruck, Yuhuang Hu, and Zhe He. V2e: From video frames to realistic dvs event camera streams. *arXiv e-prints*, pages arXiv–2006, 2020. 2, 3
- [7] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 1, 6, 7
- [8] The Blender Foundation. Blender 4.0. <https://www.blender.org/>. 6
- [9] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5501–5510, 2022. 1
- [10] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 1, 2, 6, 7
- [11] 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. 1, 3
- [12] Daniel Gehrig, Antonio Loquercio, Konstantinos G Derpanis, and Davide Scaramuzza. End-to-end learning of representations for asynchronous event-based data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5633–5643, 2019. 3
- [13] Zhiyang Guo, Wengang Zhou, Li Li, Min Wang, and Houqiang Li. Motion-aware 3d gaussian splatting for efficient dynamic scene reconstruction. *arXiv preprint arXiv:2403.11447*, 2024. 2, 4
- [14] Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4220–4230, 2024. 2, 4
- [15] 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. 2
- [16] Hankyu Jang and Daeyoung Kim. D-tensorf: Tensorial radiance fields for dynamic scenes. *arXiv preprint arXiv:2212.02375*, 2022. 2
- [17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1, 2, 3, 5, 6, 7
- [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. 2, 3
- [19] Chenghan Li, Christian Brandli, Raphael Berner, Hongjie Liu, Minhao Yang, Shih-Chii Liu, and Tobi Delbrück. Design of an rgbw color vga rolling and global shutter dynamic and active-pixel vision sensor. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 718–721. IEEE, 2015. 1, 3
- [20] Yiqing Liang, Numair Khan, Zhengqin Li, Thu Nguyen-Phuoc, Douglas Lanman, James Tompkin, and Lei Xiao. Gaufre: Gaussian deformation fields for real-time dynamic novel view synthesis. *arXiv preprint arXiv:2312.11458*, 2023. 1, 2
- [21] 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. 2
- [22] Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, and Yuchao Dai. 3d geometry-aware deformable gaussian splatting for dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8900–8910, 2024. 2, 4
- [23] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. *arXiv preprint arXiv:2308.09713*, 2023. 1, 2, 4
- [24] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *2024 International Conference on 3D Vision (3DV)*, pages 800–809. IEEE, 2024. 1
- [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. 2, 3, 6, 7
- [26] 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. 1, 2, 7
- [27] Nobuyuki Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285–296):23–27, 1975. 5
- [28] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 2, 6
- [29] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. 2, 6, 7
- [30] 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. 2, 3
- [31] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Conference on robot learning*, pages 969–982. PMLR, 2018. 6
- [32] 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. 2, 3
- [33] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 6
- [34] Ruizhi Shao, Zerong Zheng, Hanzhang Tu, Boning Liu, Hongwen Zhang, and Yebin Liu. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16632–16642, 2023. 2
- [35] Richard Shaw, Michal Nazarcuk, Jifei Song, Arthur Moreau, Sibi Catley-Chandar, Helisa Dhamo, and Eduardo Pérez-Pellitero. Swings: sliding windows for dynamic 3d gaussian splatting. ECCV, 2024. 1, 2
- [36] Karen Simonyan. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4
- [37] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 1
- [38] 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. 7
- [39] Yuchen Weng, Zhengwen Shen, Ruofan Chen, Qi Wang, and Jun Wang. Eadeblur-gs: Event assisted 3d deblur reconstruction with gaussian splatting. *arXiv preprint arXiv:2407.13520*, 2024. 2
- [40] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xianggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 1, 2, 4, 6, 7
- [41] Jingqian Wu, Shuo Zhu, Chutian Wang, and Edmund Y Lam. Ev-gs: Event-based gaussian splatting for efficient and accurate radiance field rendering. *arXiv preprint arXiv:2407.11343*, 2024. 2
- [42] Tianyi Xiong, Jiayi Wu, Botao He, Cornelia Fermuller, Yiannis Aloimonos, Heng Huang, and Christopher Metzler. Event3dgs: Event-based 3d gaussian splatting for high-speed robot egomotion. In *8th Annual Conference on Robot Learning*, 2024. 2
- [43] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*, 2023. 1
- [44] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20331–20341, 2024. 1, 2, 4, 5, 6, 7
- [45] Wangbo Yu, Chaoran Feng, Jiye Tang, Xu Jia, Li Yuan, and Yonghong Tian. Evagaussians: Event stream assisted gaussian splatting from blurry images. *arXiv preprint arXiv:2405.20224*, 2024. 2
- [46] Bowen Zhang, Yiji Cheng, Jiaolong Yang, Chunyu Wang, Feng Zhao, Yansong Tang, Dong Chen, and Baining Guo. Gaussiancube: Structuring gaussian splatting using optimal transport for 3d generative modeling. *arXiv preprint arXiv:2403.19655*, 2024. 1
- [47] 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. 7