EDyGS: Event Enhanced Dynamic 3D Radiance Fields from Blurry Monocular Video

Mengxu Lu[1,2*], **Zehao Chen**[1,2*], **Yan Liu**[1,2], **De Ma**[1,2†], **Huajin Tang**[1,2], **Qian Zheng**[1,2†] and **Gang Pan**[1,2]

[1]The State Key Lab of Brain-Machine Intelligence, Zhejiang University
[2]College of Computer Science and Technology, Zhejiang University

Abstract

The task of generating novel views in dynamic scenes plays a critical role in the 3D vision domain. Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting (3DGS) have shown great promise in this domain but struggle with motion blur, which often arises in real-world scenarios due to camera or object motion. Existing methods address camera motion blur but fall short in dynamic scenes, where the coupling of camera and object motion complicates multi-view consistency and temporal coherence. In this work, we propose EDyGS, a model designed to reconstruct sharp novel views from event streams and monocular videos of dynamic scenes with motion blur. Our approach introduces a motion-mask 3D Gaussian model that assigns each Gaussian an additional attribute to distinguish between static and dynamic regions. By leveraging this motion mask field, we separate and optimize the static and dynamic regions independently. A progressive learning strategy is adopted, where static regions are reconstructed by jointly optimizing camera poses and learnable 3D Gaussians, while dynamic regions are modeled using an implicit deformation field alongside learnable 3D Gaussians. We conduct both quantitative and qualitative experiments on synthetic and real-world data. Experimental results demonstrate that EDyGS effectively handles blurry inputs in dynamic scenes.

1 Introduction

Novel view synthesis [Zhou *et al.*, 2016] plays a critical role in the 3D vision domain and enables more immersive experiences in applications such as augmented reality/virtual reality (AR/VR), autonomous driving, and 3D content creation. Recently, methods based on reconstructing radiance fields, such as Neural Radiance Field (NeRF) [Mildenhall *et al.*, 2021] and 3D Gaussians Splatting (3DGS) [Kerbl *et al.*, 2023], have demonstrated remarkable potential in novel view synthesis.

Such methods leverage multi-view consistency to reconstruct radiance fields, relying on high-quality calibrated images as inputs [Cannici and Scaramuzza, 2024]. However, such inputs are rarely available in real-world scenarios. In dynamic scenes, the presence of motion blur, caused by both camera motion and object motion, further complicates the reconstruction process, preventing NeRF and 3DGS from producing sharp radiance fields and significantly limiting their practical applicability.

Recent works [Ma *et al.*, 2022; Lee *et al.*, 2023; Wang *et al.*, 2023] address motion blur caused by camera motion by estimating camera movement during exposure time. However, these methods are not well-suited for dynamic scenes, where motion blur arises from a combination of camera and object motion. Simulating camera motion alone is insufficient to ensure multi-view consistency, making it difficult to accurately represent object motion and maintain temporal coherence. A straightforward approach is to separate the sources of blur: camera motion and object motion. However, decoupling these two types of blur from blurry images is a challenging task. Some recent works [Sun *et al.*, 2024] attempt to address this by using pre-trained networks to predict 2D-level object motion masks from blurry images. However, this approach is constrained by the severity of the blur and the absence of multi-view consistency (Figure 1 (a)).

Event cameras [Gallego *et al.*, 2020] are high-resolution neuromorphic sensors that capture motion cues during image exposure. We observe that in dynamic scenes, event streams maintain consistency across viewpoints in static regions but lose consistency in dynamic regions due to object motion (Figure 1 (b)). By analyzing the multi-view information from event streams, we can differentiate between static and dynamic regions, further decoupling the blur caused by the two types of motion. This facilitates the reconstruction of dynamic radiance fields.

In this work, we propose EDyGS[1], a model specifically designed to take event streams and monocular videos of dynamic scenes with motion blur as inputs and generate sharp novel views. To tackle the challenge of decoupling the two types of motion blur, we leverage the multi-view information from the event stream to construct a motion mask field that differentiates between static and dynamic regions. This allows us to handle deblurring and reconstruction separately for each region. Specifically, we introduce a motion-mask

[1]Source code and Supplementary Materials are available at: https://github.com/zju-bmi-lab/EDyGS

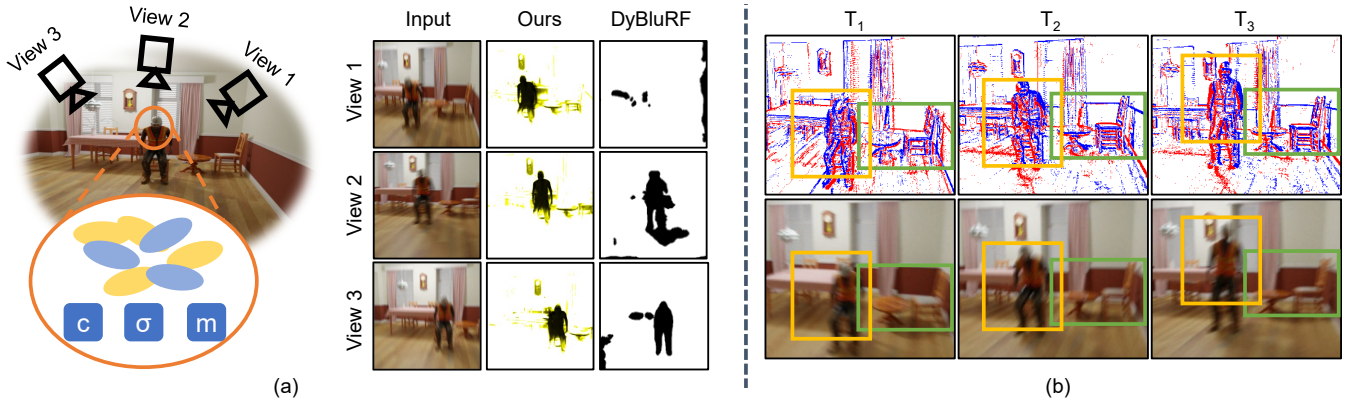

Figure 1: (a) Our method generates a motion mask field by adding attributes to 3D Gaussians to distinguish between static and dynamic regions. This mask enables the separation of two types of motion blur: camera motion and object motion. The right side shows a comparison between the 2D masks rendered from our method's motion mask field from different viewpoints (different time) and the 2D masks generated by other methods [Sun *et al.*, 2024]. It can be seen that the mask we generate ensures multi-view consistency and temporal continuity across different time frames. (b) Event streams in the static regions of dynamic scenes exhibit consistency across viewpoints at different time instances (green box). However, in the dynamic regions, due to object motion, event streams lack consistency across viewpoints at different time instances (yellow box). In contrast, blurry images lack multi-view consistency in both static and dynamic regions.

3D Gaussian Splatting model, where each Gaussian is assigned an additional attribute to indicate whether it belongs to a static or dynamic region. By leveraging event streams to supervise this model, we create a 3D-level mask that guides the reconstruction process. We then adopt a progressive learning strategy, sequentially reconstructing the static and dynamic regions. For static regions, we optimize both the camera pose and the learnable 3D Gaussians. For dynamic regions, we train an implicit deformation field alongside learnable 3D Gaussians. Experimental results demonstrate that EDyGS outperforms existing dynamic 3D radiance field methods, offering a significant advancement in reconstructing sharp, dynamic scenes.

In summary, the major contributions of our work are:

- We introduce EDyGS, the event-enhanced dynamic 3DGS model specifically designed to effectively tackle motion-blurred monocular video.
- We propose a motion-mask 3D Gaussian Splatting model that uses event streams to reconstruct the motion mask field, separating static and dynamic regions for targeted deblurring and reconstruction.
- Experimental results demonstrate that our method achieves the state-of-the-art performance.

2 Related Work

Novel View Synthesis for Dynamic Scenes. Synthesizing novel views from a series of 2D images is an important and challenging task in 3D static [Liao *et al.*, 2024; Zhang *et al.*, 2024] and dynamic reconstruction. For dynamic reconstruction, some NeRF-based methods [Park *et al.*, 2021a; Park *et al.*, 2021b; Pumarola *et al.*, 2021] incorporate a deformation field to map point coordinates to a canonical space that corresponds to a specific timestamp. They synthesize high-quality novel views, but the training and rendering speeds are slow. Besides, some Gaussian-based methods [Yang *et al.*, 2024; Wu *et al.*, 2024b; Duan *et al.*, 2024; Li *et al.*, 2024; Yang *et al.*, 2025] are proposed, achieving both fast training and high rendering quality. Both NeRF-based and Gaussian-based methods heavily rely on sharp inputs. When dealing with images affected by motion blur due to rapid objects or camera movement, they perform poorly because they lack accurate spatial information in the image to guide the supervision.

Our method accounts for camera and object motion blur and handles the motion blur effectively.

Novel View Synthesis from Blurry Images. In static 3D scene reconstruction, some methods [Ma *et al.*, 2022; Wang *et al.*, 2023; Cannici and Scaramuzza, 2024; Lee *et al.*, 2025; Chen *et al.*, 2025a; Chen *et al.*, 2025b] are proposed to address motion blur caused by camera movement. However, in dynamic 3D scene reconstruction, motion blur occurs due to the combined effect of camera motion and object motion. DyBluRF [Sun *et al.*, 2024] and Deblur4DGS [Wu *et al.*, 2024c] attempt to recover the sharp scene from blurry images. However, they rely on inaccurate blur information with a lack of multi-view consistency, thereby limiting their effectiveness. Additionally, since DyBluRF is based on NeRF, its training and rendering process are relatively slow (2 days for training on an RTX A6000 GPU).

We integrate event streams with more precise information into reconstructing 3D dynamic scenes from blurry images with 3DGS (1.5 hours for training on an RTX 3090 GPU).

Event-based Novel View Synthesis. The event camera [Brandli *et al.*, 2014], also called the Dynamic Vision Sensor (DVS), captures the intensity change of each pixel. The high temporal resolution of event cameras enables them to record events with precise information at a high frequency [Yan *et al.*, 2025; Wu *et al.*, 2024d]. As a result, event cameras are commonly used in static 3D scene reconstruction tasks [Chen *et al.*, 2021; Chen *et al.*, 2024]. Assuming the camera motion is predefined, some 3D scene re-

construction techniques [Qi *et al.*, 2023; Klenk *et al.*, 2023; Low and Lee, 2023; Cannici and Scaramuzza, 2024] leverage multi-view data from events to refine the 3D scene. Other methods [Qi *et al.*, 2024; Yu *et al.*, 2024; Ma *et al.*, 2024] jointly optimize both camera motion and the 3D scene. However, dynamic scenes introduce additional challenges due to object motion, which makes traditional methods unsuitable for such scenarios.

We leverage observations from event streams to separately consider camera motion blur and object motion blur, thereby optimizing 3D dynamic scenes from blurry images.

3 Proposed Method

In this section, we demonstrate how to reconstruct dynamic 3D radiance fields using monocular videos with motion blur $\mathcal{B} = \{B_0, B_1, ..., B_{n-1}\}$ and their corresponding event streams $\mathcal{E}_{T_{0,0} \to T_{n-1,k-1}}$. For each frame B_p and the corresponding event stream $\mathcal{E}_{T_{p,0} \to T_{p,k-1}}$, the exposure start time and end time are $T_{p,0}$ and $T_{p,k-1}$ respectively.

We first describe the method for decoupling two types of motion blur (camera motion blur and object motion blur) in Section 3.1. We then introduce a 4D field $\mathcal{M}$ to represent dynamic and static regions and propose motion-mask 3D Gaussians to implement this field in Section 3.2. Finally, we demonstrate how to utilize this field to reconstruct dynamic 3D radiance fields from blurry monocular videos $\mathcal{B}$ and their corresponding event streams $\mathcal{E}$ in Section 3.3.

3.1 Decoupling Two Types of Motion Blur

In dynamic scenes, motion blur can be categorized into two types according to its source: camera motion and object motion. Camera motion blur occurs when the camera moves during the exposure time, while object motion blur arises from objects in the scene moving during the same period [Sun *et al.*, 2024]. Distinguishing between these two types of blur in an image with motion blur is a challenging task.

To tackle this problem, we propose dividing a dynamic scene into dynamic regions and static regions. Dynamic regions correspond to parts of the scene that undergo changes over time t, such as moving objects. On the other hand, static regions remain invariant over time t, representing stationary background elements. This division allows us to associate motion blur more specifically with its source.

We find that motion blur caused by the camera movement is confined to static regions, while blur caused by the object movement is limited to dynamic regions. By leveraging this distinction between dynamic and static regions, we can decouple the two types of motion blur in an image. This decoupling is not precise but serves as a robust starting point for further analysis.

Distinction Between Dynamic and Static Regions. As a camera moves through a dynamic scene, it naturally generates a monocular video. Each frame in the video represents a specific moment in time and a distinct viewpoint. For static regions, these areas remain unchanged over time, so they appear the same when observed from different viewpoints (or moments in time). In contrast, dynamic regions change over time, resulting in differences when viewed from different viewpoints (or moments). This fundamental distinction between dynamic and static regions allows us to utilize multi-view relationships inherent in images to identify them. Static regions will maintain consistent appearance and structure across views, while dynamic regions will exhibit changes, making them distinguishable.

However, blurry images do not provide accurate multi-view relationship information [Cannici and Scaramuzza, 2024]. In contrast, event streams, with their high temporal resolution and lack of blur [Gallego *et al.*, 2020], can provide precise information. Therefore, we can utilize the multi-view relationships of event streams to identify static and dynamic regions in a dynamic scene (Shown in Figure 1 (b)).

3.2 Motion Mask Field

We propose a motion mask field $\mathcal{M}$ to represent the dynamic and static regions in the dynamic scene. This motion mask field represents the probability that a point $\mathbf{x}$ in the canonical space belongs to a dynamic region at time t,

$$m = \mathcal{M}(\mathbf{x}, t), \tag{1}$$

where m ranges between 0 and 1. A value of $m = 0$ indicates that the point belongs entirely to a static region, while $m = 1$ means that the point is fully within a dynamic region.

This motion mask field plays a crucial role in decoupling motion blur caused by camera motion and object motion. By leveraging this field, we can optimize the dynamic and static components of the scene separately, ensuring accurate reconstruction and rendering for both regions.

Motion-mask 3D Gaussian Splatting. We approximate this motion mask field representation using motion-mask 3D Gaussians, an extension of the vanilla 3DGS framework [Kerbl *et al.*, 2023]. Unlike vanilla 3DGS, which focuses solely on rendering images C, our proposed 3D Gaussians model is capable of rendering both images C and 2D mask images for each viewpoint. These masks enable the distinction between dynamic and static regions within the scene, providing an essential tool for decoupling motion blur.

To achieve this, we introduce an additional attribute m for each Gaussian. This attribute represents the probability that the corresponding Gaussian belongs to a dynamic region. The 2D mask image M is rendered using the alpha blending of the splatted 3D Gaussians onto the 2D image plane. The process is defined as:

$$M = \sum_{t \in \mathcal{N}} m_t \alpha_t \prod_{u=1}^{t-1} (1 - \alpha_u), \tag{2}$$

where m_t represents the dynamic probability for each Gaussian, α_t is the alpha value of the Gaussian, and the product term $\prod_{u=1}^{t-1}(1 - \alpha_u)$ accounts for the accumulated transparency of the Gaussians layered before t.

This field enables the generation of view-specific motion masks, which are crucial for capturing the temporal and spatial variation of dynamic regions in a scene. Unlike other methods [Sun *et al.*, 2024], which obtain masks from blurry images using a pre-trained model, our approach generates masks at the 3D level, ensuring multi-view consistency.

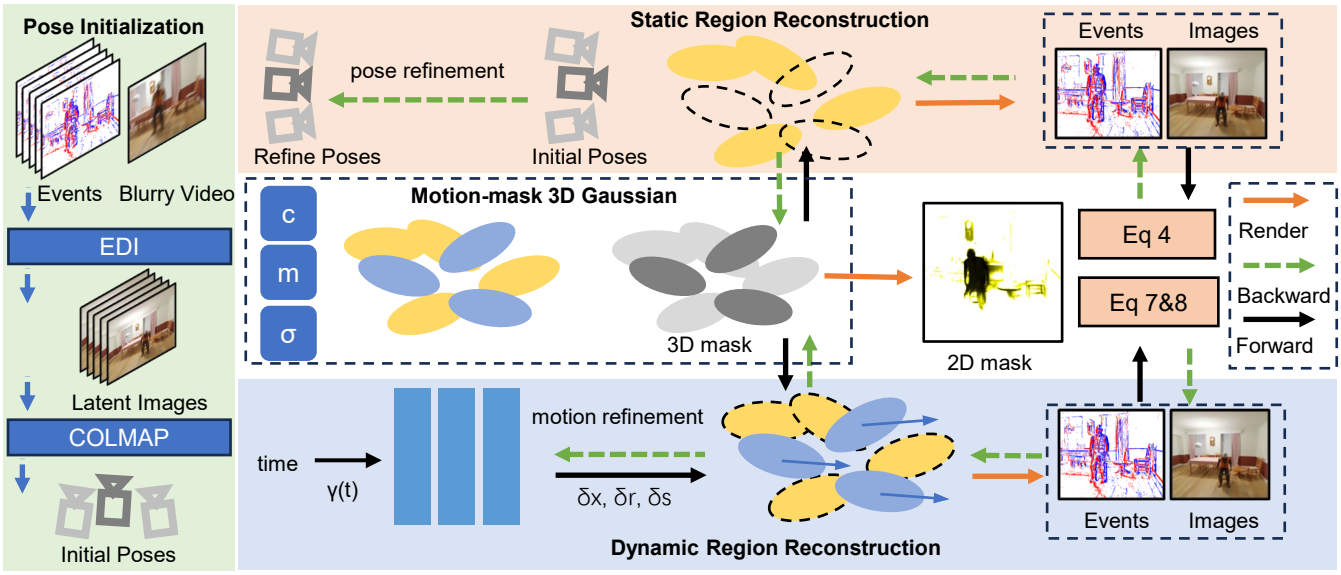

Figure 2: **Overall pipeline.** Our framework is divided into three stages: **Pose Initialization**, **Static Region Reconstruction**, and **Dynamic Region Reconstruction**. The event stream $\mathcal{E}$ and the blurry monocular video $\mathcal{B}$ first pass through the Pose Initialization stage, generating relatively latent sharp images and corresponding initial poses. In the Static Region Reconstruction stage, our method jointly optimizes the static region, motion mask field, and initial poses. In the Dynamic Region Reconstruction stage, we focus on reconstructing the dynamic region while simultaneously refining the motion mask field and poses.

3.3 EDyGS

Overview. Figure 2 illustrates the framework of our method. Our framework is divided into three stages: Pose Initialization, Static Region Reconstruction, and Dynamic Region Reconstruction. In the Pose Initialization stage, the event stream and the blurry monocular video are processed to generate relatively latent sharp images and corresponding initial poses. In the Static Region Reconstruction stage, our method jointly optimizes the static region, motion mask field, and initial poses. In the Dynamic Region Reconstruction stage, we focus on reconstructing the dynamic region while simultaneously refining the motion mask field and poses.

Pose Initialization. To obtain the initial trajectory of the event camera and the point cloud of the dynamic scene for 3DGS optimization, we follow the approach outlined in E2NeRF [Qi *et al.*, 2023]. First, we preprocess the n motion-blurred images $\{B_p\}_{p=0}^{n-1}$ using the EDI model [Pan *et al.*, 2019]. Specifically, for each motion-blurred image B_p, we uniformly sample $k-2$ timestamps to segment the corresponding event streams $\mathcal{E}_{T_{p,0}\to T_{p,k-1}}$. Using the EDI model, we then generate nk latent sharp images $\{\hat{B}_i\}_{i=0}^{nk-1}$, which capture more texture details compared to the original blurry image. Finally, we employ COLMAP [Schonberger and Frahm, 2016] to obtain the initial trajectory $\{P_i\}_{i=0}^{nk-1}$ of the event camera and initial point cloud of the dynamic scene.

Static Region Reconstruction. Given that the static regions within dynamic scenes exhibit consistency across different viewpoints and time instances, we focus solely on optimizing the static region of the Gaussian field and the trajectory of the event camera in the first stage. We first construct a differentiable path from the static region of the Gaussian field to the trajectory of the event camera. According to Equation 2, the rendered result related to the static region of the Gaussian field is given by $\hat{C}^i_{\text{st}} = \hat{C}_i(1-M_i)$, where $\hat{C}_i$ is the rendered image. Then we build the partial derivatives between $\hat{C}^i_{\text{st}}$ and the trajectory $\{P_i\}_{i=0}^{nk-1}$ of event camera:

$$\frac{\partial \hat{C}^i_{\text{st}}}{\partial P_i} = \frac{\partial \hat{C}_i}{\partial P_i} - M_i \frac{\partial \hat{C}_i}{\partial P_i} - \hat{C}_i \frac{\partial M_i}{\partial P_i}. \tag{3}$$

Using the above differentiable path (as detailed in the Supplementary Materials), we jointly optimize the camera trajectory and the static part of the Gaussian field. Meanwhile, to adaptively distinguish between the static and dynamic regions of the Gaussian field, we bias M towards classifying as static, so that the 3D mask can better identify the parts that satisfy multi-view consistency.

We separately utilize events and latent sharp images to supervise the joint training process, which includes the static region of the Gaussian field, the motion mask field, and the trajectory of the event camera:

$$\begin{aligned}
\mathcal{L} &= (1-\lambda)\mathcal{L}_1 + \lambda \mathcal{L}_{\text{D-SSIM}}, \\
\mathcal{L}_{\text{event}} &= \mathcal{L}((\ln(\hat{C}_{i+1}) - \ln(\hat{C}_i)), \sum_{\mathbf{e}\in\mathcal{E}_{i\to i+1}} \mathbf{e}\Theta), \\
\mathcal{L}_{\text{edi}} &= \mathcal{L}(\hat{C}_i, \hat{B}_i) + \mathcal{L}(\hat{C}^i_{\text{st}}, \hat{B}_i(1-M_i)), \\
\mathcal{L}_{\text{reg}} &= ||M_i||_1,
\end{aligned} \tag{4}$$

where Θ is the threshold of the event camera, and λ is 0.2.

Dynamic Region Reconstruction. Following the approach in [Yang *et al.*, 2024], we control the Gaussian using the deformation field $\mathcal{F}_\theta$. First, we predict the motion of each

Gaussian, which transforms the canonical 3D Gaussian to the deformed space:

$$(\Delta\mu_i, \Delta r_i, \Delta s_i) = \mathcal{F}_\theta(\gamma(sg(\mu_i)), \gamma(t)), \tag{5}$$

where $sg(\cdot)$ is the stop-gradient operation, γ is the positional encoding, $\Delta\mu_i$, Δr_i, and Δs_i represent the offset of per-Gaussian center point, rotation, and scale. Then we use the motion mask field obtained from the first stage to suppress the motion of the static part of the Gaussian field:

$$(\Delta\mu_i^{\text{st}}, \Delta r_i^{\text{st}}, \Delta s_i^{\text{st}}) = (\Delta\mu_i, \Delta r_i, \Delta s_i)(1 - m_i). \tag{6}$$

Finally, we refine the motion of the Gaussian field by regularizing the motion of the static part:

$$\mathcal{L}_{\text{motion}} = ||\Delta\mu_i^{\text{st}}||_2^2 + ||\Delta r_i^{\text{st}}||_2^2 + ||\Delta s_i^{\text{st}}||_2^2. \tag{7}$$

This enables the dynamic part of the Gaussian field to more effectively learn the object motion in the dynamic scene from the event streams. Meanwhile, we continue to fine-tune the mask field and the camera trajectory using both events and latent sharp images, which helps to better decouple the camera motion from the object motion:

$$\begin{aligned} \mathcal{L}_{\text{event}} &= \mathcal{L}((\ln(\hat{C}_{i+1}) - \ln(\hat{C}_i)), \sum_{\mathcal{E}\in\mathbf{e}_{i\to i+1}} \mathcal{E}\Theta), \\ \mathcal{L}_{\text{edi}} &= \mathcal{L}(\hat{C}_i, \hat{B}_i), \\ \mathcal{L}_{\text{reg}} &= ||M_i - \hat{M}_i||_1, \end{aligned} \tag{8}$$

where $\hat{M}_i$ is the rendered 2D mask image in the static region reconstruction stage.

4 Experiment

In this section, we introduce the datasets, experimental setup, comparisons with other methods, and ablation study to demonstrate the effectiveness of our approach. More experimental results are in the Supplementary Materials.

4.1 Datasets

Synthetic Data. Since no publicly available dynamic scene datasets contain both blurry images and event streams, we collect four synthetic scenarios. The data for these scenes is sourced from [Wu *et al.*, 2024a]. We first generate the corresponding events by applying Vid2E [Gehrig *et al.*, 2020] to sharp images with a resolution of 400 × 400, rendered in Blender. Next, we process nine sharp images using the simulation method in E2NeRF [Qi *et al.*, 2023] to produce one blurry image. As a result, we obtain 60 blurry images for each scenario, along with their corresponding event streams.

Real-world Data. We use the DAVIS346C [Taverni *et al.*, 2018] to capture six real-world dynamic scenes with both RGB frames and spatial-temporal aligned event streams. Each video is 5 seconds long, with the RGB frames having an exposure time of 100 milliseconds and a resolution of 346 × 260. The real-world dynamic scenes contain blurry images along with their corresponding events.

4.2 Experiment Settings

Metrics. We utilize standard metrics to evaluate the effectiveness of our method, including PSNR [Huynh-Thu and Ghanbari, 2008], SSIM [Wang *et al.*, 2004], and LPIPS [Zhang *et al.*, 2018].

	Input View			Novel View		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
D3DGS	25.78	.7796	.3430	25.13	.7657	.3269
DyBluRF	25.18	.7670	.3436	20.28	.6067	.4171
E2V+D3DGS	20.07	.7267	.3819	20.45	.7341	.3707
GEM+D3DGS	27.72	.8513	.2638	26.86	.8370	.2697
Ours	**31.84**	**.9050**	**.1923**	**30.90**	**.8993**	**.1950**

Table 1: Comparison of PSNR, SSIM, and LPIPS across different methods on deblurring input view and synthesizing novel view.

Baselines. We compare our method with four approaches: (1) Deformable 3DGS (D3DGS) [Yang *et al.*, 2024], a method used for dynamic scene reconstruction; (2) DyBluRF [Sun *et al.*, 2024], a method for dynamic scene reconstruction from blurry video; (3) E2V [Rebecq *et al.*, 2019]+D3DGS, a combined method that uses E2V to generate latent sharp images from events and then employs D3DGS for dynamic scene reconstruction; (4) GEM [Zhang *et al.*, 2023]+D3DGS, a combined method that uses an event-guided deblur method (GEM) to deblur images first, then employs D3DGS for dynamic scene reconstruction.

4.3 Results

Quantitative Analysis on Synthetic Data. The results of the synthesis of novel views and input views are shown in Table 1. Compared to other methods, our quantitative results outperform them. D3DGS [Yang *et al.*, 2024], lacking a module specifically designed for blurry image inputs, fails to reconstruct sharp dynamic scenes under blurry conditions. DyBluRF [Sun *et al.*, 2024] leverages optical flow and depth information derived from blurry images but lacks precise spatial data, resulting in poor performance. E2V [Rebecq *et al.*, 2019]+D3DGS and GEM [Zhang *et al.*, 2023]+D3DGS suffer from multi-view inconsistency, preventing D3DGS from producing effective and high-quality results.

Qualitative Analysis on Synthetic Data. Figure 3 shows the visual comparison results of the input view and the novel view in synthetic scenes. Both D3DGS and DyBluRF lack precise spatial information as inputs, resulting in overall blurry images. E2V+D3DGS, due to significant image inconsistencies, deviates from the training images required by D3DGS, leading to more distorted renderings. Additionally, images generated by GEM exhibit certain inconsistencies, causing GEM+D3DGS to produce blurry details. Since these methods do not distinguish between coupled camera and object motion blur, they perform poorly in dynamic regions. Our method separates the two types of motion blur, resulting in good performance in both static and dynamic regions.

Qualitative Analysis on Real-world Data. We further validate the conclusions obtained from the synthetic datasets on the real-world datasets. Figure 4 shows the rendered images of several methods on the real-world datasets. The D3DGS rendered image is highly blurry. DyBluRF removes only a small amount of blur. The rendered images from E2V+D3DGS and GEM+D3DGS still contain some blurred details. Our method achieves better deblurring results, and the rendered images in the novel view synthesis task are also

Figure 3: Results on synthetic scenes. The red box highlights the detailed magnification of the dynamic region, while the green box shows the detailed magnification of the static region. The leftmost column shows inputs, including blurry images (top) and event streams (bottom).

notably better since we separate static and dynamic regions for targeted deblurring and reconstruction.

Depth Visualization. Figure 5 shows the visualization of depth maps for both real-world and synthetic scenes. It can be observed that our method reconstructs dynamic scenes with well-defined geometric structures, making it beneficial for the novel view synthesis task.

5 Ablation Study

To validate the effectiveness of each component of our framework, we conduct ablation experiments as presented in Table 2 and Table 3.

Effect of EDI+event. In Table 2, when using only the event loss, the sparse spatial information of events is insufficient to effectively supervise D3DGS in modeling dynamic scenes. Utilizing latent sharp images from EDI with dense spatial information as supervision results in an average PSNR improvement of 19.68. However, these images exhibit some color distortion and noise. By incorporating the event loss alongside the latent sharp images supervision to jointly supervise D3DGS, a further increase in PSNR is observed.

Effect of optimizing poses. To achieve more accurate camera motion, we optimize the estimated poses, resulting in an average PSNR improvement of 1.33. We further validate that our method can optimize the camera motion trajectory more accurately by comparing it with the ground truth pose in synthetic scenes. The quantitative result is shown in Table 3.

	PSNR↑	SSIM↑	LPIPS↓
event	7.94	.1122	.6407
EDI	27.62	.8440	.2687
EDI+event	28.95	.8714	.2195
pose, w/o mask	30.28	.8889	.2141
pose, mask	**30.90**	**.8993**	**.1950**

Table 2: Ablation study on EDI, event, optimizing poses and motion mask. By leveraging these components, our method achieves the highest PSNR of 30.90.

w/o optimizing poses			Ours		
RPE_t↓	RPE_r↓	ATE↓	RPE_t↓	RPE_r↓	ATE↓
0.3543	0.0248	0.0065	**0.1765**	**0.0148**	**0.0024**

Table 3: Ablation study on optimizing poses.

Effect of motion mask. To separate object motion blur from multi-view information and camera motion blur, we use a mask to learn the dynamic components. This approach results in an average PSNR increase of 0.62 in Table 2.

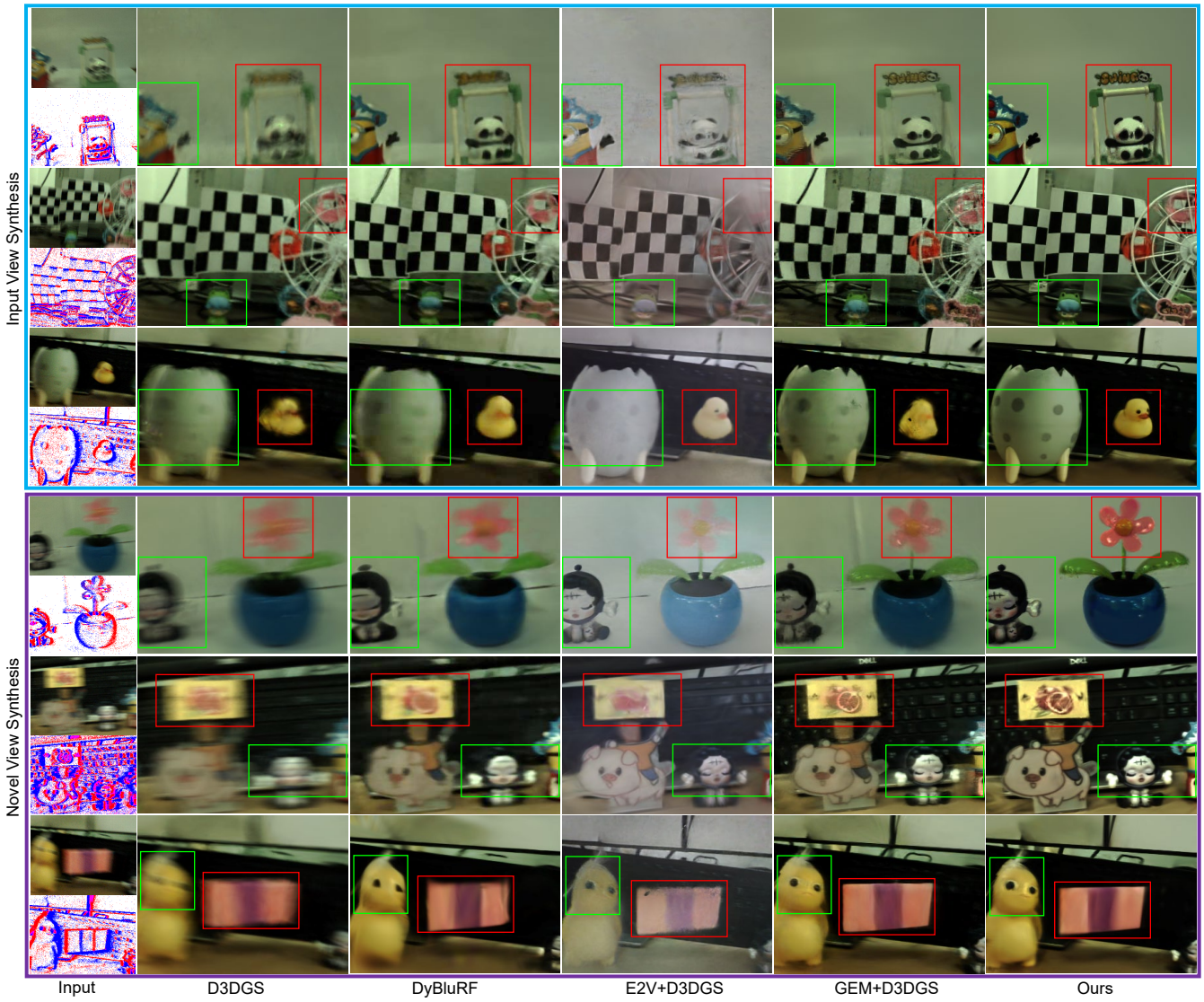

Figure 4: Results on real-world scenes. The red box highlights the detailed magnification of the dynamic region, while the green box shows the detailed magnification of the static region. The leftmost column shows the inputs (blurry images and the event streams).

6 Conclusion

In conclusion, we present EDyGS, a method for recovering a sharp dynamic scene representation from event streams and blurry monocular video. With EDyGS, we can deblur input views and generate sharp novel views of dynamic scenes. The core of our approach is the decoupling of two types of motion blur by leveraging multi-view information from the event stream, accounting for both camera and object motion blur. Experimental results demonstrate that EDyGS achieves the state-of-the-art performance in dynamic scene reconstruction with blurry images as inputs.

Limitations. Although EDI can effectively leverage event data to perform a coarse deblurring of blurry images, the latent sharp images after EDI still exhibit some gap compared to the sharp ground truth images.

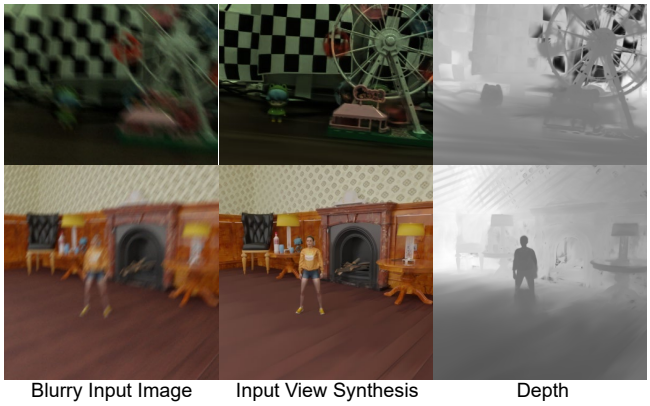

Figure 5: Depth visualization. The first row shows the real-world scene data, and the second row shows the synthetic data.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (No. 2024YDLN0005), in part by the National Natural Science Foundation of China (62376247, U20A20220, and 62334014), and in part by the grant from Key R&D Program of Zhejiang (2022C01048).

Contribution Statement

* These authors contributed equally. † Corresponding author.

References

[Brandli *et al.*, 2014] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 2014.

[Cannici and Scaramuzza, 2024] 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.

[Chen *et al.*, 2021] Zehao Chen, Qian Zheng, Peisong Niu, Huajin Tang, and Gang Pan. Indoor lighting estimation using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[Chen *et al.*, 2024] Zehao Chen, Zhan Lu, De Ma, Huajin Tang, Xudong Jiang, Qian Zheng, and Gang Pan. Eventid: Intrinsic decomposition using an event camera. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024.

[Chen *et al.*, 2025a] Zehao Chen, Zhanfeng Liao, De Ma, Huajin Tang, Qian Zheng, and Gang Pan. Evhdr-nerf: Building high dynamic range radiance fields with single exposure images and events. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Chen *et al.*, 2025b] Zehao Chen, Zhan Lu, De Ma, Huajin Tang, Xudong Jiang, Qian Zheng, and Gang Pan. Evhdrgs: Event-guided hdr video reconstruction with 3d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Duan *et al.*, 2024] Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes. In *ACM SIGGRAPH 2024 Conference Papers*, 2024.

[Gallego *et al.*, 2020] 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*, 2020.

[Gehrig *et al.*, 2020] Daniel Gehrig, Mathias Gehrig, Javier Hidalgo-Carrió, and Davide Scaramuzza. Video to events: Recycling video datasets for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[Huynh-Thu and Ghanbari, 2008] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 2008.

[Kerbl *et al.*, 2023] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 2023.

[Klenk *et al.*, 2023] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-nerf: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 2023.

[Lee *et al.*, 2023] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. Dp-nerf: Deblurred neural radiance field with physical scene priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[Lee *et al.*, 2025] Byeonghyeon Lee, Howoong Lee, Xiangyu Sun, Usman Ali, and Eunbyung Park. Deblurring 3d gaussian splatting. In *European Conference on Computer Vision*. Springer, 2025.

[Li *et al.*, 2024] Deqi Li, Shi-Sheng Huang, Zhiyuan Lu, Xinran Duan, and Hua Huang. St-4dgs: Spatial-temporally consistent 4d gaussian splatting for efficient dynamic scene rendering. In *ACM SIGGRAPH 2024 Conference Papers*, 2024.

[Liao *et al.*, 2024] Zhanfeng Liao, Yan Liu, Qian Zheng, and Gang Pan. Spiking nerf: Representing the real-world geometry by a discontinuous representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[Low and Lee, 2023] 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*, 2023.

[Ma *et al.*, 2022] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[Ma *et al.*, 2024] Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. Continuous pose for monocular cameras in neural implicit representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[Mildenhall *et al.*, 2021] 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*, 2021.

[Pan *et al.*, 2019] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event

camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[Park *et al.*, 2021a] 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*, 2021.

[Park *et al.*, 2021b] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: a higher-dimensional representation for topologically varying neural radiance fields. *ACM Transactions on Graphics (TOG)*, 2021.

[Pumarola *et al.*, 2021] 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*, 2021.

[Qi *et al.*, 2023] 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*, 2023.

[Qi *et al.*, 2024] Yunshan Qi, Lin Zhu, Yifan Zhao, Nan Bao, and Jia Li. Deblurring neural radiance fields with event-driven bundle adjustment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024.

[Rebecq *et al.*, 2019] 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*, 2019.

[Schonberger and Frahm, 2016] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[Sun *et al.*, 2024] Huiqiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, and Zhiguo Cao. Dyblurf: Dynamic neural radiance fields from blurry monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[Taverni *et al.*, 2018] Gemma Taverni, Diederik Paul Moeys, Chenghan Li, Celso Cavaco, Vasyl Motsnyi, David San Segundo Bello, and Tobi Delbruck. Front and back illuminated dynamic and active pixel vision sensors comparison. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018.

[Wang *et al.*, 2004] 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*, 2004.

[Wang *et al.*, 2023] Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[Wu *et al.*, 2024a] Guanjun Wu, Taoran Yi, Jiemin Fang, Wenyu Liu, and Xinggang Wang. Fast high dynamic range radiance fields for dynamic scenes. In *2024 International Conference on 3D Vision (3DV)*. IEEE, 2024.

[Wu *et al.*, 2024b] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[Wu *et al.*, 2024c] Renlong Wu, Zhilu Zhang, Mingyang Chen, Xiaopeng Fan, Zifei Yan, and Wangmeng Zuo. Deblur4dgs: 4d gaussian splatting from blurry monocular video. *arXiv preprint arXiv:2412.06424*, 2024.

[Wu *et al.*, 2024d] Xiaoshan Wu, Weihua He, Man Yao, Ziyang Zhang, Yaoyuan Wang, Bo Xu, and Guoqi Li. Event-based depth prediction with deep spiking neural network. *IEEE Transactions on Cognitive and Developmental Systems*, 2024.

[Yan *et al.*, 2025] Haojie Yan, Zhan Lu, Zehao Chen, De Ma, Huajin Tang, Qian Zheng, and Gang Pan. Evstvsr: Event guided space-time video super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Yang *et al.*, 2024] 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*, 2024.

[Yang *et al.*, 2025] Xiaodong Yang, Weixing Xie, Sen Peng, Yihang Fu, Wentao Fan, Baorong Yang, and Xiao Dong. 4d gaussian splatting for high-fidelity dynamic reconstruction of single-view scenes. *Neurocomputing*, 2025.

[Yu *et al.*, 2024] 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.

[Zhang *et al.*, 2018] 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*, 2018.

[Zhang *et al.*, 2023] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing event-based motion deblurring in real-world scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[Zhang *et al.*, 2024] Weixing Zhang, Zongrui Li, De Ma, Huajin Tang, Xudong Jiang, Qian Zheng, and Gang Pan. Spiking gs: Towards high-accuracy and low-cost surface reconstruction via spiking neuron-based gaussian splatting. *arXiv preprint arXiv:2410.07266*, 2024.

[Zhou *et al.*, 2016] Tinghui Zhou, Shubham Tulsiani, Weilun Sun, Jitendra Malik, and Alexei A Efros. View synthesis by appearance flow. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016*. Springer, 2016.