EvHDR-NeRF: Building High Dynamic Range Radiance Fields with Single Exposure Images and Events

Zehao Chen[1,2], Zhanfeng Liao[1,2], De Ma[1,2], Huajin Tang[1,2], Qian Zheng[1,2]*, Gang Pan[1,2]*

[1]The State Key Lab of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China
[2]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
{zehao, zhanfengliao, made, htang, qianzheng, gpan}@zju.edu.cn

Abstract

We present EvHDR-NeRF to recover a High Dynamic Range (HDR) radiance field from event streams and a set of Low Dynamic Range (LDR) views with single exposures. Using the EvHDR-NeRF, we can generate both novel HDR views and novel LDR views under different exposures. The key to our method is to model the new relationship between events streams and LDR images, which considers both the Camera Response Function (CRF) and exposure time. Based on this relationship, we categorize events into inter-frame events and intra-exposure. The former is utilized for building HDR radiance field and the latter is used to deblur potentially blurred images. Compared to existing methods, this method can effectively reconstruct the HDR radiance field even when the input images are degraded. Experimental results demonstrate that our method achieves state-of-the-art HDR reconstruction, providing a more adaptable and accurate solution for complex imaging applications.

Introduction

HDR imaging aims to restore real-world scenes as accurately as possible. With the advent of Neural Radiance Fields (NeRF) (Mildenhall et al. 2021), researchers begin to explore HDR reconstruction for 3D scenes (Huang et al. 2022; Jun-Seong et al. 2022; Wu et al. 2024). Unlike HDR image/video reconstruction tasks, HDR reconstruction of 3D scenes enables the generation of HDR views beyond the initial input perspectives. This research improves the realism of virtual representations and holds promise applications in areas such as film production, 3D gaming, and virtual reality.

HDR reconstruction methods for 3D scenes currently rely on multi-exposure images as input (Huang et al. 2022; Jun-Seong et al. 2022). This method requires capturing multi-exposure images from different viewpoints, which increases the operational complexity for the user. Additionally, this type of method has specific requirements for the range of multi-exposure inputs (Fig. 1(b))[1], which limits the usage scenarios. For instance, in dimly lit indoor environments, short exposures might yield images with poor signal-to-noise ratios, while long exposures taken with handheld devices might lead to blurred images. As a result, these conditions can compromise image quality and lead to suboptimal reconstruction outcomes. Using single-exposure, multi-view images as input can reduce the operational complexity for users. However, performing HDR reconstruction with single-exposure images is an ill-posed problem.

Event cameras (Gallego et al. 2020) are neuromorphic sensors that have a high temporal resolution and a high dynamic range, and perform well in low light scenarios. Their superior performance in diverse lighting conditions offers broader applicability and increased convenience for HDR imaging tasks. However, current works (Han et al. 2020; Yang et al. 2023) on HDR reconstruction are based on a single image or video. Directly applying these event-guided methods to 3D scene HDR reconstruction is ineffective because they fail to ensure multi-view consistency (Fig. 1(d)). Moreover, these methods face additional challenges: They predefine the relationship between events and images, ignoring the impact of the Camera Response Function (CRF) and exposure time on the relationship between LDR images and events, leading to different HDR results for LDR images with different exposures (Fig. 1(c)).

Therefore, to apply single-exposure images and events to HDR reconstruction of 3D scenes, we need to establish a new relationship between events and images that 1) incorporates factors like exposure times and the CRF, and 2) is adaptable, adjusting to changes in data distribution.

Our key observations are that "Events indicate logarithmic changes in radiance from a scene, while images represent the integration of radiance over the exposure time, which is then processed by the CRF." Based on these observations, we employ radiance as a bridge to establish a relationship between events and images. In this relationship, we consider the CRF, exposure times, intra-exposure and inter-frame events. Taking into account these factors, our method can handle scenarios with degraded image quality.

Based on this relationship, we propose an optimization method based on NeRF. Specifically, we propose establishing relationships between viewpoint and radiance, and between radiance and LDR image, using two separate MLPs. The first MLP is designed to model the relationship between changes in the view of the scene and radiance. This relationship is primarily constrained by inter-frame events due to their ability to capture high dynamic range information

*Corresponding author.
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
[1]For specific details, please refer to Fig. 8

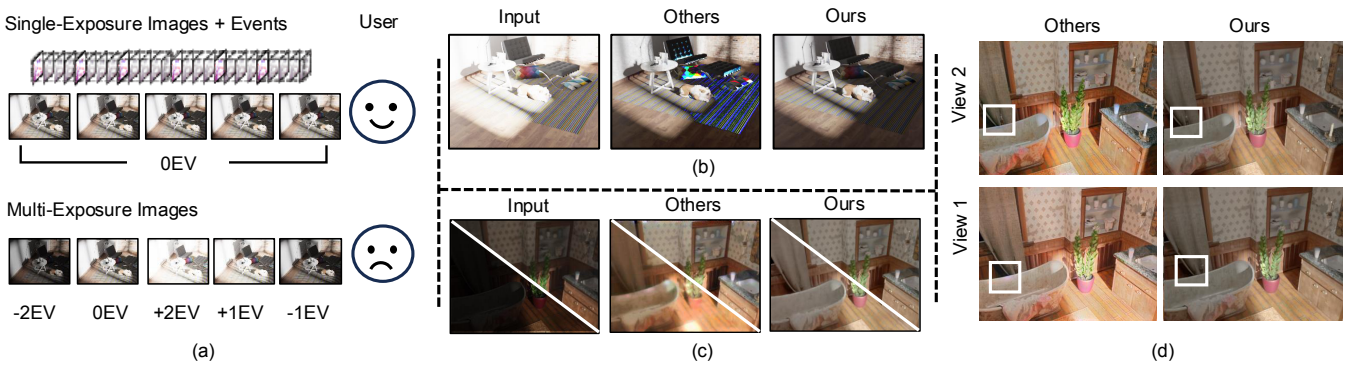

Figure 1: (a). We recover a high dynamic range neural radiance field from single-exposures LDR views and events. Our method is simpler to operate compared to multi-exposure techniques. (b). When the exposure range of input images in multi-exposure method (Huang et al. 2022) is narrow, the rendering of novel exposure time can fail. In contrast, our method does not require a specific exposure range. (c). The event-guided method (Yang et al. 2023) does not take exposure time into account, resulting in the reconstruction of HDR images with different brightness levels from LDR images with varying exposures. (d). The event-guided method (Yang et al. 2023) cannot guarantee multi-view consistency; at the white box, the colors from two views are inconsistent.

about a scene. The second MLP is designed to model the relationship between radiance and the resulting LDR image. This relationship is constrained not only by the parameters learned by the MLP, such as the CRF but also by intra-exposure events. These events, which occur within the exposure time of a single frame, provide valuable information that can help constrain and enhance the quality of the reconstructed LDR image. The results of our experiments indicate that we have achieved state-of-the-art performance in HDR reconstruction of 3D scenes. The main contributions of this paper can be summarized as follows,

1. We have established a new relationship between events and frames that considers the CRF and exposure time.
2. Based on this relationship, we propose an optimization framework that utilizes events and single-exposure images to build high dynamic range radiance fields. This method can effectively reconstruct 3D scenes in HDR even when the input images are degraded.
3. Our method can estimate the CRF from events and single-exposure images and can generate images with novel views and varying exposure times.

Related Work

Neural Radiance Fields

The domain of Neural Rendering has experienced a significant increase in interest following the recent introduction of NeRF (Mildenhall et al. 2021; Liao et al. 2024).

NeRF for HDR. (Mildenhall et al. 2022) introduced a method for reconstructing the high dynamic range neural radiance field from raw data. (Jun-Seong et al. 2022) and (Huang et al. 2022) devised techniques to recover the high dynamic range neural radiance field from a collection of LDR images with different exposures. (Wu et al. 2024) propose a dynamic HDR NeRF framework which can learn 3D scenes from dynamic 2D images captured with various exposures. (Lu et al. 2024) proposes a NeRF method for HDR panoramic imaging. However, these methods increase operational complexity for users.

In contrast, our approach utilizes single-exposure images and events as inputs. Due to the high temporal resolution of events and their advantageous signal-to-noise ratio in low-light conditions, our method can tolerate degraded images.

NeRF for Events. Recent research (Rudnev et al. 2023; Hwang, Kim, and Kim 2023; Low and Lee 2023; Qi et al. 2023; Klenk et al. 2023; Ma et al. 2023) has sought to integrate events with NeRF to improve neural radiance fields reconstruction. Given that event cameras only capture changes in intensity rather than absolute values, methods (Rudnev et al. 2023; Hwang, Kim, and Kim 2023; Low and Lee 2023) that exclusively depend on events as inputs require the manual assignment of a background intensity value. This approach could potentially lead to inaccuracies in the intensity of the reconstructed scene. In response to this issue, other techniques (Qi et al. 2023; Klenk et al. 2023; Ma et al. 2023) employ images and events as inputs. However, these methods presuppose that events are generated by differencing between two consecutive LDR frames. (Cannici and Scaramuzza 2024) considers the gap between events and images, but does not account for exposure time. Furthermore, (Chen et al. 2024) propose an intrinsic decomposition framework that takes advantage of events for stable decomposition under extreme scenarios.

In contrast to these previous methods, our model understands the physical imaging relationship between events and frames, factoring in exposure time and the CRF.

HDR Reconstruction

Frame-based HDR Reconstruction. Frames-based HDR reconstruction methods can be divided into two categories: multi-exposure methods and single-exposure meth-

ods. Multi-exposure methods (Nayar and Mitsunaga 2000; Tocci et al. 2011; Kang et al. 2003; Kalantari and Ramamoorthi 2019; Chen et al. 2021a; Chung and Cho 2023) merge a stack of LDR images with multi-exposure into an HDR image. Compared to multi-exposure methods, single-exposure methods (Endo, Kanamori, and Mitani 2017; Rempel et al. 2007; Eilertsen et al. 2017) are simpler to operate. However, these methods lack physical guidance in over/under-exposed regions, making it an ill-posed problem. Previous HDR reconstruction methods were based on images or videos (Yan et al. 2019; Liu et al. 2022; Song et al. 2022; Chung and Cho 2023; Shu et al. 2024; Xu et al. 2024). With the emergence of NeRF, researchers have begun to explore HDR reconstruction for 3D scenes (Huang et al. 2022; Jun-Seong et al. 2022). Currently, these methods rely on multi-exposure as input.

In contrast, our approach utilizes single-exposure images and events as inputs.

Event-guided HDR Reconstruction. Event cameras, unlike conventional RGB cameras, possess the unique ability to asynchronously record logarithmic changes in radiance, thereby providing a significantly higher dynamic range and are applied in many fields (Chen et al. 2021b; Lele et al. 2021; Li et al. 2018). Recent studies (Wang et al. 2019, 2020; Pan et al. 2020; Shaw et al. 2022; Wang et al. 2021) have investigated the potential of using events to aid in the transformation of LDR images or videos into HDR format. (Han et al. 2020) proposed a method that initially reconstructs intensity maps from events, followed by their fusion with the LDR image in the intensity domain. Subsequently, (Yang et al. 2023) introduced a multimodal learning strategy aimed at converting LDR videos and events into HDR videos. (Cui et al. 2024) utilizes color events in single-exposure HDR imaging. (Zou et al. 2024) utilizes a recurrent convolutional neural network that reconstruct high-speed HDR videos from event steams.

The difference between our method and others lies in the fact that our method is an optimization method and can also learn the CRF. Because of this, we can achieve consistent HDR results using inputs from LDR images captured under various exposure settings and are not affected by data bias.

Preliminary

Event generation model. An event $e_t = (\mathbf{x}, t, p)$ at pixel position $\mathbf{x} = (u, v)$ and time t with polarity $p \in \{-1, +1\}$ is generated when the logarithmic change of brightness L since the last event at the pixel $\mathbf{x}$ and time $t - \Delta\mathrm{t}$ exceeds a threshold Θ ($\Theta > 0$) and the event e can be represented:

$$e_t = \lfloor (\ln(\mathrm{L_t}) - \ln(\mathrm{L_{t-\Delta t}}))/\Theta \rfloor . \tag{1}$$

Consistent with the perspective of (Yang et al. 2023), we regard brightness L as the radiance within the scene.

Neural Radiance Field. NeRF (Mildenhall et al. 2021) is formulated as: $(\mathbf{h}(\mathbf{r}), \sigma(\mathbf{r})) = F(\mathbf{r})$. Radiance $\mathbf{h}$ and density σ represent the outputs of the radiance field F, with ray $\mathbf{r}$ defined by $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$, where $\mathbf{o}$ denotes the ray's origin, $\mathbf{d}$ its direction, and s denotes a position on the ray. The expected color $\hat{C}$ is defined as:

$$\hat{C}(\mathbf{r}) = \int_{s_{\mathrm{near}}}^{s_{\mathrm{far}}} \exp\left(-\int_{\mathrm{s_{near}}}^{\mathrm{s}} \sigma(\mathbf{r}(\mathrm{p}))\mathrm{dp}\right) \sigma(\mathbf{r}(\mathrm{s}))\mathbf{h}(\mathbf{r}(\mathrm{s}))\mathrm{ds}. \tag{2}$$

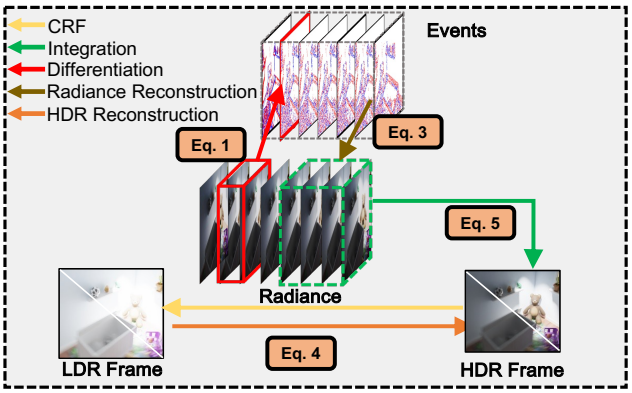

Figure 2: Schematic of the events-radiance-frames relationship. Radiance can serve as a bridge connecting events and frames. Within this relationship between events and frames, we consider factors such as exposure time and the Camera Response Function (CRF).

Proposed Method

In this section, we establish the relationship between events streams and single-exposure frames. Based on this relationship, we propose an optimization framework, called EvHDR-NeRF, to build high dynamic range radiance fields.

Modeling

In response to the requirements outlined in the introduction regarding the relationship between events and images, we model this relationship. Our observations are that "Events indicate logarithmic changes in radiance from a scene, while frames represent the integration of radiance over the exposure time, which is then processed by the CRF". Based on this observation, we employ radiance as a bridge to establish a relationship between events and frames (shown in Fig. 2).

Following Eq. (1), the radiance R can be expressed as,

$$R_{t_j} = R_{t_i} \exp(\Theta \hat{\mathrm{E}}_{\mathrm{t_i} \to \mathrm{t_j}}), \tag{3}$$

where $R_{t_{\{i,j\}}}$ stands for radiance at time $t_{\{i,j\}}$, Θ is a threshold and $\hat{E}_{t_i \to t_j}$ represents the **observed** accumulation of events from time t_i to t_j, which is in the form of a matrix[2].

The relationship between events streams and frames can be mathematically expressed as,

$$\mathbf{I}_{\mathrm{LDR}} = f_{\mathrm{CRF}}(\mathbf{I}_{\mathrm{HDR}}), \tag{4}$$

$$\mathbf{I}_{\mathrm{HDR}} = \int_{t_i}^{t_j} R_{t_i} \exp(\Theta \hat{\mathrm{E}}^{\mathrm{exp}}_{\mathrm{t_i} \to \mathrm{t}})\mathrm{dt}, \tag{5}$$

[2]In this paper, we use $\hat{\bullet}$ to denote the observed data.

where $\mathbf{I}_{\mathrm{LDR}}$ represents LDR frame, f_{CRF} is CRF, $\mathbf{I}_{\mathrm{HDR}}$ represents HDR frame (the integration of radiance over the exposure time) and the term $\hat{E}^{\mathrm{exp}}_{t_i \to t}$ represents **observed** accumulation of events during the exposure time (from t_i to t_j).

To better decouple the variables, following (Huang et al. 2022), we reformulate Eq. (4) in the logarithmic space,

$$\mathbf{I}_{\mathrm{LDR}} = \mathcal{G}(\ln(\mathrm{R}_{\mathrm{t_i}}) + \ln(\Delta \mathrm{t}) + \ln(\mathcal{E})), \tag{6}$$

$$\mathcal{E} = \left(\frac{1}{\Delta t} \int_{t_i}^{t_j} \exp(\Theta \hat{\mathrm{E}}^{\mathrm{exp}}_{\mathrm{t_i} \to \mathrm{t}}) dt \right), \tag{7}$$

where $\mathcal{G} = (\ln f^{-1}_{\mathrm{CRF}})^{-1}$ and $\Delta \mathrm{t} = \mathrm{t_j} - \mathrm{t_i}$[3]. Unlike (Huang et al. 2022), Eq. (6) takes into account the events streams that occur within the exposure time.

In exploring the relationship between radiance and frames, we consider two key factors: the CRF and the integration of radiance over the exposure time. Compared to other event-guided HDR reconstruction methods (Yang et al. 2023), our approach takes into account the integration of radiance over the exposure time. Additionally, we have access to the events $\hat{E}^{\mathrm{exp}}_{t_i \to t}$, enabling us to effectively handle situations with motion blur (see Fig. 4 and Fig. 5). Due to the higher signal-to-noise ratio of events compared to images under low-light conditions. The relationship we have established can tolerate image degradation.

EvHDR-NeRF

Overview. In this section, we propose our HDR reconstruction framework, called EvHDR-NeRF. The function $\mathcal{G}$ is difficult to estimate from observations that have a constant exposure time, so we treat it as a learnable variable and use an MLP is used to interpolate its value for unobserved exposure times. Additionally, directly computing the high dynamic range radiance R_t is challenging due to its complexity. Therefore, we represent R_t as a high dynamic range radiance field and use another MLP, named $\mathcal{F}$, to learn and model its characteristics. The architecture of our model, including these components, is detailed in Fig. 3.

Events for Radiance Estimation. Our objective is to capture the radiance that has a high dynamic range. Events inherently exhibit this high dynamic range. Consequently, events serve as an ideal constraint for guiding MLP $\mathcal{F}$ during the learning process. We can revise Eq. (3) to express the computation of events in the context of radiance values,

$$E_{t_i \to t_j} = \frac{\ln(\mathrm{R}_{\mathrm{t_j}}) - \ln(\mathrm{R}_{\mathrm{t_i}})}{\Theta}. \tag{8}$$

We employ accumulated events $\hat{E}_{t_i \to t_j}$ as a supervisory signal to guide the MLP $\mathcal{F}$. This method serves to significantly reduce the effect of noise intrinsic to the events. As a result, it substantially improves the network's resilience. Like (Rudnev et al. 2023), we optimize the implicit functions $\mathcal{F}$ using events by constraining them by minimizing the mean squared error (MSE). The events reconstruction loss is formulated as,

$$\mathcal{L}_e = \sum_{\mathbf{r} \in \mathcal{R}(\mathbf{P})} \| \hat{E}_{t_i \to t_j} - E_{t_i \to t_j} \|^2. \tag{9}$$

[3] $\Delta \mathrm{t}$ represents the exposure time and can be obtained from event timestamps.

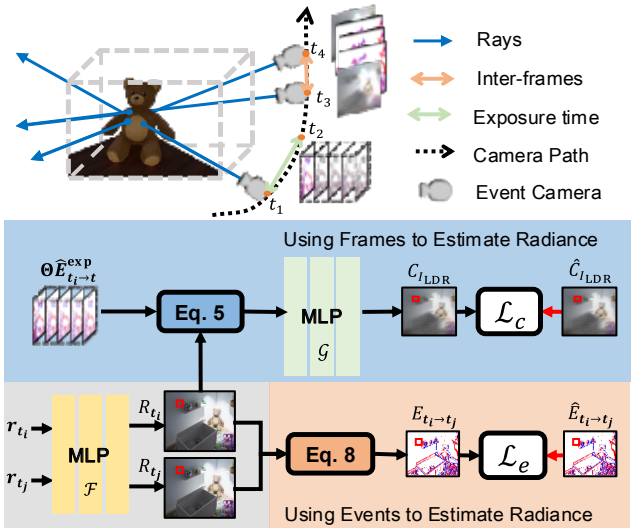

Figure 3: (Top) Setup. Consider a scenario where a hybrid camera system is continually recording data within a scene, yielding a series of LDR frames and their corresponding events; (Bottom) Framework Overview. The two rays $r_{\{t_i,t_j\}}$ sequentially enter the HDR radiance field (MLP $\mathcal{F}$) to generate radiance $R_{\{t_i,t_j\}}$ from two viewpoints. The MLP $\mathcal{G}$ receives radiance R and the accumulation of events during an exposure time $\hat{E}^{\mathrm{exp}}_{t_i \to t}$ as its inputs and produces the LDR color $C_{I_{\mathrm{LDR}}}$ as its output. In the diagram, we have omitted the process of sampling the 3D points on each ray and integrating the output colors of each 3D point.

Frames for Radiance Estimation. However, relying solely on events to estimate radiance proves to be insufficient, as events provide only differential information about the radiance as indicated by Eq.(8), without the ability to establish absolute radiance values. Fortunately, LDR frames offer these 'absolute radiance values', even though they are influenced by the CRF. Therefore, it is essential to also utilize LDR frames to impose constraints on the network. Incorporating this constraint allows us to mitigate the differential nature of Eq. (8) and achieve a complete representation of the radiance.

Specifically, we optimize the implicit functions $\mathcal{G}$ using frames by constraining them through the minimization of MSE. The image reconstruction loss is formulated as,

$$\mathcal{L}_c = \sum_{\mathbf{r} \in \mathcal{R}(\mathbf{P})} \| \hat{C}_{I_{\mathrm{LDR}}}(\mathbf{r}) - C_{I_{\mathrm{LDR}}}(\mathbf{r}) \|^2, \tag{10}$$

where $C_{I_{\mathrm{LDR}}}(\mathbf{r})$ represents the color value of a pixel in LDR image I_{LDR}.

Additionally, we use the unit exposure loss $\mathcal{L}_u$ from (Huang et al. 2022) to better constrain $\mathcal{G}$. And the unit exposure loss $\mathcal{L}_u$ can be defined as,

$$\mathcal{L}_u = \| \mathcal{G}(0) - C_0 \|_2^2, \tag{11}$$

where C_0 is the midpoint of the pixel value.

The final loss function is as follows,

$$\mathcal{L} = \mathcal{L}_{\mathrm{e}} + \lambda_{\mathrm{c}}\mathcal{L}_{\mathrm{c}} + \lambda_{\mathrm{u}}\mathcal{L}_{\mathrm{u}}, \tag{12}$$

where $\lambda_{\{\mathrm{c,u}\}} = \{1, 0.5\}$ denotes the weight of those loss.

Implementation Details. We base our implementation on (Huang et al. 2022), utilizing the Adam optimizer (Kingma and Ba 2014) with an initial learning rate of 5×10^{-4}, decaying exponentially to 5×10^{-5} over 200K iterations. We set the threshold Θ to 0.3. The optimization runs on a single NVIDIA RTX3090 GPU for around 16 hours.

Experiments

Dataset and Metrics

Metrics. Our experiments utilize standard metrics such as PSNR (Hore and Ziou 2010), SSIM (Hore and Ziou 2010), and LPIPS (Zhang et al. 2018) to evaluate HDR reconstruction. Consistent with common practice and following (Huang et al. 2022; Yang et al. 2023), we analyze HDR outputs tone-mapping using the μ-law. And the tone-mapping is expressed as:

$$M(E) = \frac{\log(1+\mu \mathrm{E})}{\log(1+\mu)}, \tag{13}$$

where μ is the compression level set at 5000, and E is the scaled HDR pixel value within [0, 1].

Synthetic Data. To compensate for the absence of a dataset comprising HDR images, blurred images, low-light images, and events, we augmented the synthetic dataset from (Huang et al. 2022) with blurred images, low-light images and events using methods from (Qi et al. 2023), (Liang et al. 2023) and (Yang et al. 2023). We crafted three data variants—over-exposed, blurred and under-exposed images—to facilitate cross-task comparisons, resulting in three composite datasets (Over-Exposed, Blur, Under-Exposed). The dataset features eight scenes, each with 35 viewpoints, using 17 images and their corresponding events from each scene for training, with the rest allocated for testing.

Real-World Data. To test the robustness of our approach, we collected real data using the DAVIS 346C sensor from three scenes, capturing 6-second monocular videos (each video is about 6 seconds and 200 frames) with events and frames. Frame poses were estimated using COLMAP (Schonberger and Frahm 2016). We sampled 35 evenly distributed frames to form our dataset, segmenting the events based on the frames' timestamps.

Comparison with State-of-the-Art Methods

We selected two types of SOTA HDR reconstruction methods in recent years for comparison: two NeRF-based methods (Huang et al. 2022; Wu et al. 2024) and one event-guided method (Yang et al. 2023). Furthermore, we used the event-guided approach (Zhang et al. 2023; Yang et al. 2023) to enhance the degraded images before entering them into the NeRF-based method (Huang et al. + X)[4]. Finally, we compared three event-based NeRF methods (Rudnev et al. 2023; Low and Lee 2023; Klenk et al. 2023) that exclusively use events as input.

[4]In the Blur and Under-exposed datasets, X represents the method (Zhang et al. 2023) and (Yang et al. 2023), respectively

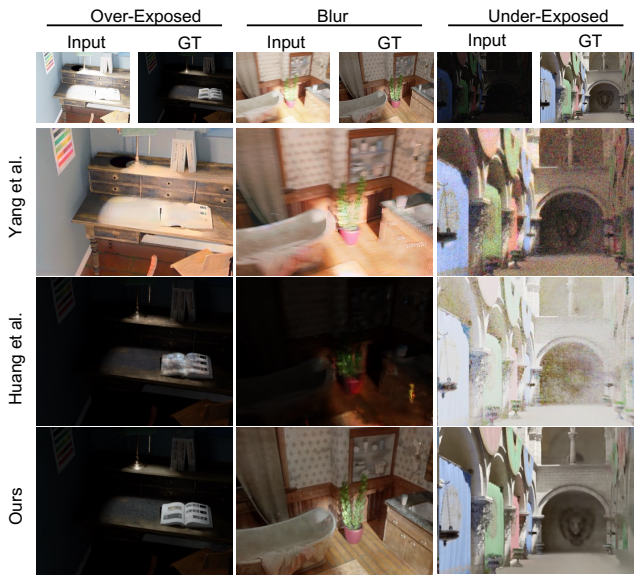

Figure 4: Qualitative comparison of HDR reconstruction on our synthetic datasets.(From left to right: Over-Exposed, Blur, Under-Exposed)

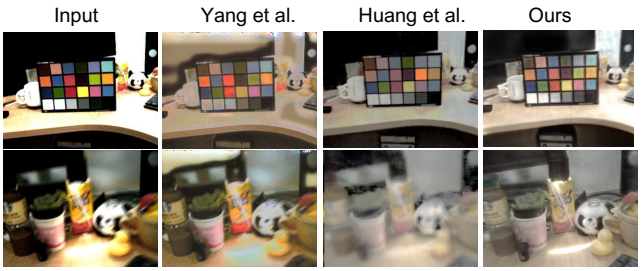

Figure 5: Qualitative comparison of HDR reconstruction in real-world data (top: sharp images as input; bottom: blurry images as input).

Comparison with HDR Methods. Table 1 showcases the quantitative results, indicating our method's superiority. Shown in Fig. 4, Huang et al. encounters difficulties in discerning details within areas of over-exposure. In contrast, Yang et al. capitalizes on the advantages of events to successfully reconstruct details in the over-exposed parts. However, as a data-driven approach, it tends to produce results with a yellow color shift. When dealing with blurry images, Huang et al. struggles as the misalignment between images taken at different exposures can lead to unsuccessful reconstruction. In underexposed scenarios, Huang's method struggles to reconstruct details in dark areas due to noise interference. In contrast, our approach leverages the high signal-to-noise ratio of events in low-light conditions, successfully reconstructing these details. Fig. 5 displays comparative re-

	Over-Exposed			Blur			Under-Exposed		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Yang et al.	16.91	0.7443	0.1972	15.86	0.6261	0.2950	16.48	0.3545	0.8270
Huang et al.[1]	24.15	0.7800	0.1599	15.73	0.4472	0.4816	17.9680	0.5506	0.4095
Wu et al.	24.70	0.6648	0.1509	14.31	0.4221	0.5285	17.62	0.4590	0.4911
Huang et al. + X	—	—	—	13.62	0.59018	0.3136	12.517	0.3074	0.7546
Ours	**28.95**	**0.9162**	**0.0586**	**28.3245**	**0.9002**	**0.0850**	**28.01**	**0.8831**	**0.1615**

[1] Huang et al. needs inputs of images with two-exposure. We doubled the original exposure to meet input requirements, using images from both exposures for input.

Table 1: Quantitative comparisons of HDR reconstruction (input view) in our synthetic datasets. In the Blur and Under-exposed datasets, X represents the method Zhang et al. and Yang et al., respectively. ↑ (↓) means higher (lower) is better.

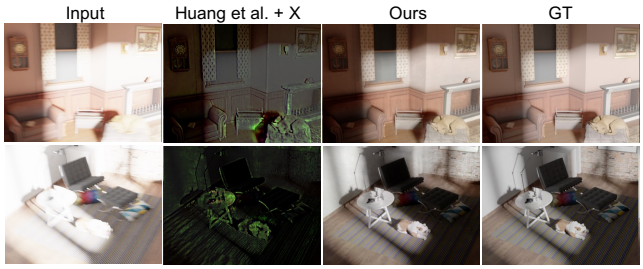

Figure 6: Qualitative comparisons with combination of SOTA methods.

sults from real-world scenarios. It is observable that Yang et al. retains a yellowish tint, in line with its performance on synthetic data. Huang et al. shows color inaccuracies, especially evident on the colorimetric card. Conversely, our method effectively recovers the texture of the overexposed desk area. When a blurry image is used as the input, our method successfully reconstructs HDR images.

Novel View & Novel Exposure Time. Similar to Huang et al., our method is also capable of rendering novel views and novel exposure times. Table. 2 outlines the quantitative evaluations for synthesizing novel views, with our approach maintaining strong performance across multiple datasets. Fig. 7 shows visualizations in new exposure settings, highlighting our model's effective scene interpretation and its ability to enhance LDR images to HDR quality by using events and single-exposure LDR frames. Nonetheless, we observed a limitation in Huang et al., particularly when the exposure times during training and testing diverge markedly. The network's performance may degrade, presumably because it learns mappings close to the training exposures and cannot easily extend to vastly different ones. Fig. 7 (right) compares our estimated CRF curves with those from Huang et al.. The ability to learn the CRF from multi-exposure images hinges on having a broad range of input exposures; otherwise, CRF estimation can be inaccurate.

Impact of Input Exposure Range on CRF Estimation. We investigated how varying the range of input exposures affects the CRF estimation in Huang et al. We experimented with three different exposures, expanding beyond the previously used double exposures. As evidenced in Fig. 8, when the exposure time range is limited to [2,16], Huang et al.

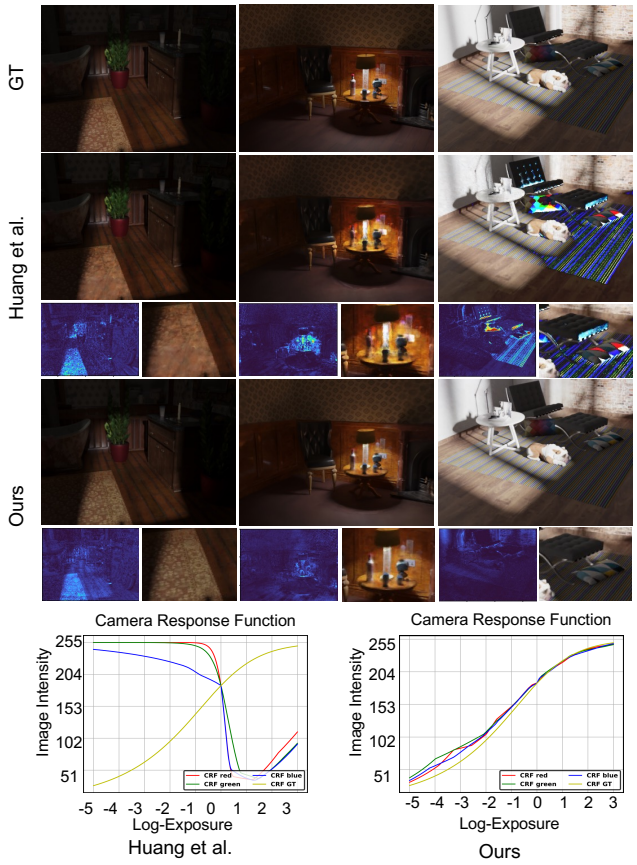

Figure 7: Top: Qualitative comparison of rendered novel LDR view with a novel exposure in Over-Exposed dataset. The error maps and zoom-in insets are provided at the bottom. Bottom: Discrete CRFs estimated by Huang et al. (Left) and by our method on synthetic dog scene (Right).

does not reconstruct the CRF accurately. Remarkably, our method can approximate the performance of Huang et al. with a wide exposure range of [0.5,16] using just a single type of exposure. This underscores our method's efficiency in CRF reconstruction with limited exposure diversity.

We do note a visible difference between our CRF curve estimation and the ground truth at the extremity of short exposure times (left side). This is due to two main reasons:

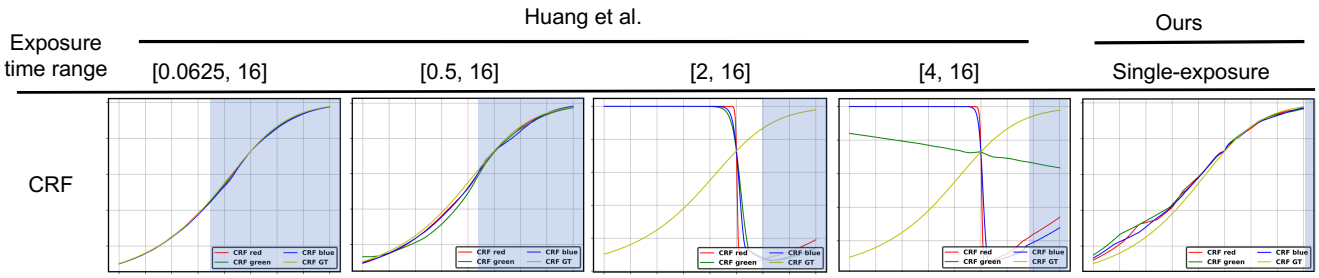

Figure 8: Discrete CRFs estimated by Huang et al. with varying exposure range and qualitative results of HDR reconstruction on our Over-exposed dataset. (The blue region indicates the exposure range covered in the logarithmic domain). When the exposure time range is limited to [2,16], Huang et al. does not reconstruct the CRF accurately. Remarkably, our method can approximate the performance of Huang et al. with a wide exposure range of [0.5,16] using single-exposure images.

	Over-Exposed			Blur			Under-Exposed		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Huang et al.	24.0275	0.7718	0.1639	16.0738	0.4555	0.4914	17.8692	0.7417	0.1883
Ours	**29.4243**	**0.9210**	**0.0562**	**28.8037**	**0.9086**	**0.0784**	**27.7480**	**0.8763**	**0.1642**

Table 2: Quantitative performance comparison of novel view in our synthetic datasets. ↑ (↓) means higher (lower) is better.

Figure 9: Qualitative comparison with the event-based NeRF methods in our synthetic datasets.

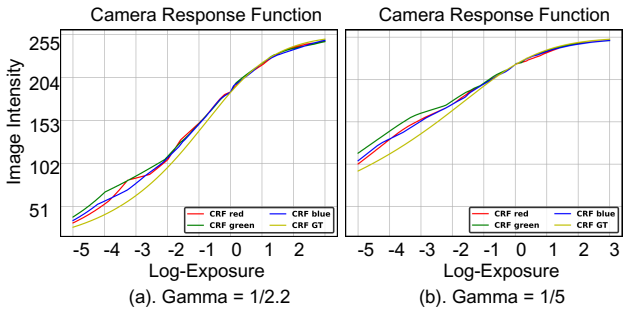

Figure 10: Ablation study: Discrete CRFs estimated by our method on synthetic dog scene: (a) the gamma value is set to 1/2.2; (b) the gamma value is set to 1/5.

first, events only provide incremental information, which is insufficient for full HDR capture. Second, the overexposure bias in our inputs skews the exposure times towards the right segment of the CRF curve. This leads to a more precise fit on the right than on the left of the estimated CRF curve.

Comparison with Event-based NeRF. We also specifically compare our approach with event-based NeRF methods. As seen in Fig 9, methods relying solely on events as input result in significant color deviation from the ground truth. This arises from events encapsulating only relative brightness relationships, requiring these methods to impose constraints with a fixed background value.

Effectiveness of Estimating Discrete CRFs. To authenticate if our methodology has indeed grasped the camera response function, we reconstructed LDR images utilizing HDR images. This process distinguishes itself from previous approaches as we refined the transformation function from HDR to LDR (the gamma value was adjusted from $1/2.2$ to $1/5$). As depicted in Fig. 10 (right), our methodology continues to successfully learn the fresh mapping relationship.

Conclusion

In conclusion, we present EvHDR-NeRF to recover a high dynamic range radiance field from event streams and a set of LDR views with single exposures. Using the EvHDR-NeRF, we can generate both novel HDR views and novel LDR views under different exposures. The key to our method is to model the new relationship between events streams and LDR images, which considers both the CRF and exposure time. Experimental results demonstrate that our method achieves state-of-the-art HDR reconstruction, providing a more adaptable and accurate solution for complex imaging applications.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61925603, 62376247, U20A20220, and 62334014), in part by the grant from Key R&D Program of Zhejiang (2022C01048), and in part by the Fundamental Research Funds for the Central Universities.

References

Cannici, M.; and Scaramuzza, D. 2024. Mitigating motion blur in neural radiance fields with events and frames. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9286–9296.

Chen, G.; Chen, C.; Guo, S.; Liang, Z.; Wong, K.-Y. K.; and Zhang, L. 2021a. HDR Video Reconstruction: A Coarse-to-fine Network and A Real-world Benchmark Dataset. *ICCV*.

Chen, Z.; Lu, Z.; Ma, D.; Tang, H.; Jiang, X.; Zheng, Q.; and Pan, G. 2024. Event-ID: Intrinsic Decomposition Using an Event Camera. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 10095–10104.

Chen, Z.; Zheng, Q.; Niu, P.; Tang, H.; and Pan, G. 2021b. Indoor lighting estimation using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14760–14770.

Chung, H.; and Cho, N. I. 2023. Lan-hdr: Luminance-based alignment network for high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12760–12769.

Cui, M.; Wang, Z.; Wang, D.; Zhao, B.; and Li, X. 2024. Color Event Enhanced Single-Exposure HDR Imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1399–1407.

Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R. K.; and Unger, J. 2017. HDR image reconstruction from a single exposure using deep CNNs. *ACM transactions on graphics (TOG)*, 36(6): 1–15.

Endo, Y.; Kanamori, Y.; and Mitani, J. 2017. Deep reverse tone mapping. *ACM Trans. Graph*, 36(6): 1–10.

Gallego, G.; Delbrück, T.; Orchard, G.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A. J.; Conradt, J.; Daniilidis, K.; et al. 2020. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1): 154–180.

Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; and Shi, B. 2020. Neuromorphic camera guided high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1730–1739.

Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*, 2366–2369. IEEE.

Huang, X.; Zhang, Q.; Feng, Y.; Li, H.; Wang, X.; and Wang, Q. 2022. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18398–18408.

Hwang, I.; Kim, J.; and Kim, Y. M. 2023. Ev-NeRF: Event based neural radiance field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 837–847.

Jun-Seong, K.; Yu-Ji, K.; Ye-Bin, M.; and Oh, T.-H. 2022. Hdr-plenoxels: Self-calibrating high dynamic range radiance fields. In *European Conference on Computer Vision*, 384–401. Springer.

Kalantari, N. K.; and Ramamoorthi, R. 2019. Deep HDR video from sequences with alternating exposures. In *Computer graphics forum*, volume 38, 193–205. Wiley Online Library.

Kang, S. B.; Uyttendaele, M.; Winder, S.; and Szeliski, R. 2003. High dynamic range video. *ACM Transactions On Graphics (TOG)*, 22(3): 319–325.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klenk, S.; Koestler, L.; Scaramuzza, D.; and Cremers, D. 2023. E-nerf: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 8(3): 1587–1594.

Lele, A.; Fang, Y.; Ting, J.; and Raychowdhury, A. 2021. An end-to-end spiking neural network platform for edge robotics: From event-cameras to central pattern generation. *IEEE Transactions on Cognitive and Developmental Systems*, 14(3): 1092–1103.

Li, H.; Li, G.; Ji, X.; and Shi, L. 2018. Deep representation via convolutional neural network for classification of spatiotemporal event streams. *Neurocomputing*, 299: 1–9.

Liang, J.; Yang, Y.; Li, B.; Duan, P.; Xu, Y.; and Shi, B. 2023. Coherent Event Guided Low-Light Video Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10615–10625.

Liao, Z.; Liu, Y.; Zheng, Q.; and Pan, G. 2024. Spiking NeRF: Representing the Real-World Geometry by a Discontinuous Representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13790–13798.

Liu, Z.; Wang, Y.; Zeng, B.; and Liu, S. 2022. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on computer vision*, 344–360. Springer.

Low, W. F.; and Lee, G. H. 2023. Robust e-NeRF: NeRF from Sparse & Noisy Events under Non-Uniform Motion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 18335–18346.

Lu, Z.; Zheng, Q.; Shi, B.; and Jiang, X. 2024. Pano-NeRF: Synthesizing High Dynamic Range Novel Views with Geometry from Sparse Low Dynamic Range Panoramic Images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3927–3935.

Ma, Q.; Paudel, D. P.; Chhatkuli, A.; and Van Gool, L. 2023. Deformable Neural Radiance Fields using RGB and Event Cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3590–3600.

Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P. P.; and Barron, J. T. 2022. Nerf in the dark: High dynamic

range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16190–16199.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Nayar, S. K.; and Mitsunaga, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, 472–479. IEEE.

Pan, L.; Hartley, R.; Scheerlinck, C.; Liu, M.; Yu, X.; and Dai, Y. 2020. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2519–2533.

Qi, Y.; Zhu, L.; Zhang, Y.; and Li, J. 2023. E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13254–13264.

Rempel, A. G.; Trentacoste, M.; Seetzen, H.; Young, H. D.; Heidrich, W.; Whitehead, L.; and Ward, G. 2007. Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM transactions on graphics (TOG)*, 26(3): 39–es.

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

Schonberger, J. L.; and Frahm, J.-M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Shaw, R.; Catley-Chandar, S.; Leonardis, A.; and Pérez-Pellitero, E. 2022. HDR reconstruction from bracketed exposures and events. *arXiv preprint arXiv:2203.14825*.

Shu, Y.; Shen, L.; Hu, X.; Li, M.; and Zhou, Z. 2024. Towards Real-World HDR Video Reconstruction: A Large-Scale Benchmark Dataset and A Two-Stage Alignment Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2879–2888.

Song, J. W.; Park, Y.-I.; Kong, K.; Kwak, J.; and Kang, S.-J. 2022. Selective transhdr: Transformer-based selective hdr imaging using ghost region mask. In *European Conference on Computer Vision*, 288–304. Springer.

Tocci, M. D.; Kiser, C.; Tocci, N.; and Sen, P. 2011. A versatile HDR video production system. *ACM Transactions on Graphics (TOG)*, 30(4): 1–10.

Wang, B.; He, J.; Yu, L.; Xia, G.-S.; and Yang, W. 2020. Event enhanced high-quality image recovery. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, 155–171. Springer.

Wang, L.; Ho, Y.-S.; Yoon, K.-J.; et al. 2019. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10081–10090.

Wang, Z.; Ng, Y.; Scheerlinck, C.; and Mahony, R. 2021. An asynchronous kalman filter for hybrid event cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 448–457.

Wu, G.; Yi, T.; Fang, J.; Liu, W.; and Wang, X. 2024. Fast High Dynamic Range Radiance Fields for Dynamic Scenes. In *2024 International Conference on 3D Vision (3DV)*, 862–872. IEEE.

Xu, G.; Wang, Y.; Gu, J.; Xue, T.; and Yang, X. 2024. HDR-Flow: Real-Time HDR Video Reconstruction with Large Motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24851–24860.

Yan, Q.; Gong, D.; Shi, Q.; Hengel, A. v. d.; Shen, C.; Reid, I.; and Zhang, Y. 2019. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1751–1760.

Yang, Y.; Han, J.; Liang, J.; Sato, I.; and Shi, B. 2023. Learning event guided high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13924–13934.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhang, X.; Yu, L.; Yang, W.; Liu, J.; and Xia, G.-S. 2023. Generalizing Event-Based Motion Deblurring in Real-World Scenarios. In *ICCV*.

Zou, Y.; Fu, Y.; Takatani, T.; and Zheng, Y. 2024. EventHDR: From Event to High-Speed HDR Videos and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.