Mitigating Motion Blur in Neural Radiance Fields with Events and Frames

Marco Cannici and Davide Scaramuzza
Robotics and Perception Group, University of Zurich, Switzerland

Abstract

Neural Radiance Fields (NeRFs) have shown great potential in novel view synthesis. However, they struggle to render sharp images when the data used for training is affected by motion blur. On the other hand, event cameras excel in dynamic scenes as they measure brightness changes with microsecond resolution and are thus only marginally affected by blur. Recent methods attempt to enhance NeRF reconstructions under camera motion by fusing frames and events. However, they face challenges in recovering accurate color content or constrain the NeRF to a set of predefined camera poses, harming reconstruction quality in challenging conditions. This paper proposes a novel formulation addressing these issues by leveraging both model- and learning-based modules. We explicitly model the blur formation process, exploiting the event double integral as an additional model-based prior. Additionally, we model the event-pixel response using an end-to-end learnable response function, allowing our method to adapt to non-idealities in the real event-camera sensor. We show, on synthetic and real data, that the proposed approach outperforms existing deblur NeRFs that use only frames as well as those that combine frames and events by $+6.13dB$ *and* $+2.48dB$*, respectively.*

Multimedial Material: For videos, datasets and more visit https://github.com/uzh-rpg/evdeblurnerf.

1. Introduction

Neural Radiance Fields (NeRFs) [22] have completely revolutionized the field of 3D reconstruction and novel view synthesis, achieving unprecedented levels of details [2, 3, 35]. As a result, they have quickly found applications in many subfields of computer vision and robotics, such as pose estimation and navigation [29, 44, 49], image processing [11, 20, 23, 38], scene understanding [15, 19, 42], surface reconstruction [1, 39, 45], and many others.

On many occasions, however, images must be captured while in motion, which can impact picture quality, often resulting in motion blur. In such circumstances, NeRFs struggle to reconstruct sharp radiance fields, thereby hindering their practical application in real-world scenes. Although recent works [6, 16, 20, 40] have shown promising results in reconstructing radiance fields from motion-blurred images by learning to infer the camera motion during the exposure time, the task of recovering motion-deblurred NeRFs still remains significantly ill-posed. Existing image-based approaches typically fail when the camera undergoes similar motion trajectories during the exposure of training images [20], and they are inherently limited by the presence of motion ambiguities and loss of texture that cannot be recovered from blurry images alone.

Figure 1. Ev-DeblurNeRF combines blurry images and events to recover sharp radiance fields. A motion-aware NeRF recovers motion and a learnable event camera response function models real camera's non-idealities, enabling high-quality reconstructions.

In this regard, recent works have shown that event-based cameras can substantially aid the task of deblurring standard images [27, 30, 36, 46], thanks to their ability to measure brightness changes with microsecond precision and their robustness to motion blur [10]. Motivated by these advantages, the literature has recently looked into the possibility of recovering NeRFs from events [4, 12, 14, 25, 26]. While most of the works [4, 12, 26] focus on event data only, research on fusing motion-blurred images with events is still limited, with only two prior works [14, 25].

In this work, depicted in Fig. 1, we propose Ev-DeblurNeRF, a novel event-based deblurring NeRF formulation combining learning and model-based components. Inspired by E-NeRF [14], it exploits continuous event-by-event supervision to recover sharp radiance fields. But it departs from E-NeRF in that it models the blur formation process explicitly, exploiting the direct relationship between events triggered during the exposure time and the resulting blurred frames, i.e., the so-called Event Double Integral (EDI) [24]. Unlike E^2NeRF [25], our approach employs this relation as additional training supervision, adding an

end-to-end learnable camera response function that enables diverging from the model-based solution whenever inaccurate, thus resulting in higher-quality reconstructions.

To summarize, our contributions are:

- A novel approach for recovering a sharp NeRF in the presence of motion blur, incorporating both model-based priors and novel learning-based modules.
- A model +2.48dB more accurate and 6.9× faster to train than previous event-based deblurring NeRF methods.
- Two new datasets, one simulated and one collected using a Color-DAVIS346 [17] event camera, featuring precise ground truth poses for accurate quality assessment.

2. Related Works

In recent years, event-based cameras have become increasingly popular [8, 21, 32, 33, 41] due to their high dynamic range and temporal resolution. Several methods have been proposed to exploit the unique characteristics of event cameras for image deblurring, featuring both model-based approaches, such as the event-based double integral (EDI) [24, 24], as well as learning-based ones [9, 13, 30, 31, 37, 43, 47]. Recently, event-based cameras have also been used to recover sharp NeRFs [12, 26] from events only, or combining event-based cameras with motion-blurred images. E-NeRF [14] shows that incorporating an event supervision loss can enhance the recovery of sharp edges, but it struggles to restore sharp colors due to the lack of explicit blur modeling. Inspired by the success of recent image-only deblur NeRFs, such as [6, 16, 20, 40], E^{2}NeRF [25] follows Deblur-NeRF [20] by modeling the camera motion during the exposure time. Notably, in our approach, we exploit continuous event-by-event supervision and employ a novel learnable camera response function that better adapts to real data, resulting in improved performance under fast motion.

3. Method

The proposed Ev-DeblurNeRF aims to recover a latent sharp scene given a sequence of timestamped blurry images $\{(\mathbf{C}_i^{\text{blur}}, t_i)\}_{i=1}^{N_I}$ and events $\mathcal{E} = \{\mathbf{e}_j = (\mathbf{u}_j, t_j, p_j)\}_{j=1}^{N_E}$, specifying that either an increase or decrease in brightness (as indicated by the polarity $p_j \in \{-1, 1\}$) has been detected at a certain time instant t_j and pixel $\mathbf{u}_j = (u_j, v_j)$.

Event-Aided Deblur-NeRF. Our architecture takes inspiration from prior works [6, 16, 20], and is depicted in Figure 2. We aim to recover the scene as a radiance field, implemented by an MLP F_Ω, blindly, by directly modeling the blur formation process at each exposure. A blurry color observation generated by the ray $\mathbf{r}(\mathbf{u}, t_i)$ cast by pixel $\mathbf{u}$ during its exposure can be described as the integral of the sharp colors observed by the ray in a time interval $\Delta T_i = [t_i - \tau/2, t_i + \tau/2]$.

Similarly to [16], we learn to estimate the motion of each ray using a neural module $G_{\mathbf{\Phi}}$. We discretize the motion in a finite set of M observations and learn an $SE(3)$ field that rigidly warps pixel rays to discrete positions q:

$$(\mathbf{e^r}_q, \mathbf{t}_q, w_q) = G_{\mathbf{\Phi}}(\mathcal{R}(\boldsymbol{l}_i); \mathcal{T}(\boldsymbol{l}_i); \mathcal{W}(\boldsymbol{l}_i)), \tag{1}$$

where $\boldsymbol{l} \in \mathbb{R}^E$ is a shared learned image embedding, and $\mathcal{R}$, $\mathcal{T}$ and $\mathcal{W}$ are independent MLPs that predict, respectively, a set of rotation matrices $\mathbf{e}_q^r \in SO(3)$, translation vectors $\mathbf{t}_q \in \mathcal{R}^3$, and view weights $w_q \in \mathcal{R}$, one for each discrete position q. The warped rays can thus be finally obtained as $\hat{\mathbf{r}}_q = \mathbf{e^r}_q \mathbf{r}(\mathbf{u}, t_i) + \mathbf{t}_q$.

Following NeRF [7], we render the color at each ray with a pair of MLPs, one coarse- and the other fine-grained, F_Ω^c and F_Ω^f. Inspired by the hybrid design in [6], we enhance the capabilities of F_Ω^c and F_Ω^f by incorporating dedicated TensoRF [5] volumes, which we employ as additional input feature spaces for the MLPs. In particular, given a ray $\mathbf{r_u}$ and a set of coarse and fine points $\{\mathbf{x}_k^c\}_{k=1}^S$ and $\{\mathbf{x}_k^f\}_{k=1}^S$ along the ray, we first sample feature volumes:

$$\begin{aligned} f_{s_k}^{\,c} &= \mathcal{V}_s(\mathbf{x}_k^c), \;\; f_{s_k}^{\,f} = \mathcal{V}_s(\mathbf{x}_k^f), \\ f_{l_k}^{\,c} &= \mathcal{V}_l(\mathbf{x}_k^c), \;\; f_{l_k}^{\,f} = \mathcal{V}_l(\mathbf{x}_k^f), \end{aligned} \tag{2}$$

with $\mathcal{V}_s$ and $\mathcal{V}_l$, respectively, a small and a large TensoRF [5] volume. We use $f_{s_k}^{\,c}$ as additional features in F_Ω^c, and use all of the features with the fine-level MLP F_Θ^f.

Volumetric rendering is then finally used to estimate colors $\hat{\mathbf{C}}_q$ at the predicted camera positions, which are finally fused into a blurry observation

$$\hat{\mathbf{C}}^{\text{blur}}(\mathbf{r}(\mathbf{u}, t_i)) = g\left(\sum_{q=1}^{M-1} w_q \hat{\mathbf{C}}_q\right), \tag{3}$$

where $g(\cdot)$ is a gamma correction function. Inspired by [16], we further refine the composite weights using an adaptive weight proposal network $\lambda_q = \mathcal{AWP}(\zeta_q, \boldsymbol{l}_i, \mathbf{d}_q)$, which takes the ray's samples features ζ_q, directions $\mathbf{d}_q$ and image embedding $\mathbf{l}_i$ to produce refined weights. We use these refined weights in Equation (3) in place of w_q to obtain refined colors $\tilde{\mathbf{C}}^{\text{blur}}$.

The thus rendered synthetic blurry pixel is finally supervised with a ground truth observation $\mathbf{C}_{\text{gt}}$ through:

$$E^b(\mathbf{C}_{\mathbf{r}}^{\text{blur}}) = \left\| \mathbf{C}_{\mathbf{r}}^{\text{blur}} - \mathbf{C}_{\text{gt}}^{\text{blur}}(\mathbf{r}) \right\|_2^2 \tag{4}$$

$$\mathcal{L}_{\text{blur}} = \frac{1}{|\mathcal{R}_b|} \sum_{\mathbf{r} \in \mathcal{R}_b} E^b(\hat{\mathbf{C}}_{\mathbf{r}_c}^{\text{blur}}) + E^b(\hat{\mathbf{C}}_{\mathbf{r}_f}^{\text{blur}}) + E^b(\tilde{\mathbf{C}}_{\mathbf{r}_f}^{\text{blur}}), \tag{5}$$

where we consider a batch of pixels $\mathcal{R}_b$, and rewrite $\mathbf{C}_{\mathbf{r}}^{\text{blur}} = \mathbf{C}^{\text{blur}}(\mathbf{r})$. The subscripts c and f indicate values obtained through F_Ω^c or F_Ω^f, and $\tilde{}$ if adaptive weights are used.

Event-based supervision via learned event-CRF. We now exploit blur-free microsecond-level event measurements. Let's denote the brightness at a pixel $\mathbf{u}$ on a given

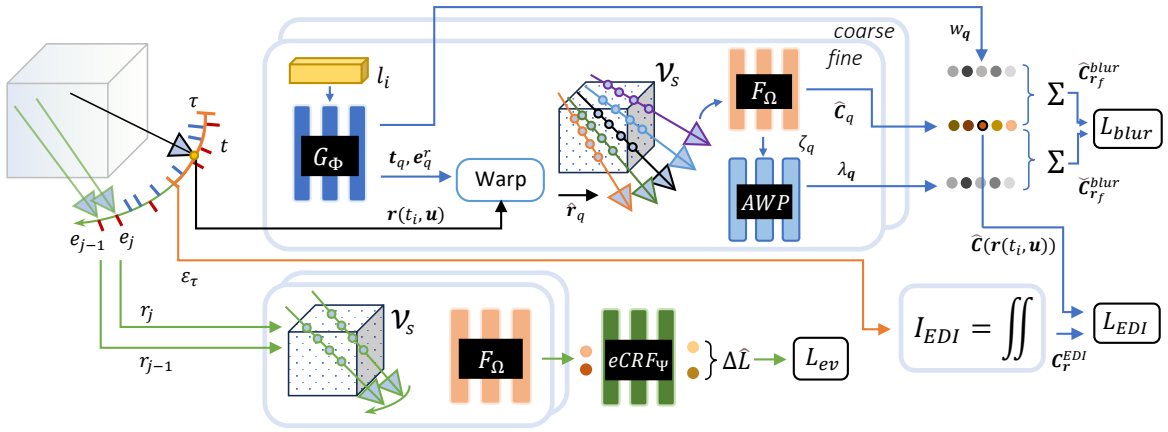

Figure 2. For each given ray $\mathbf{r}(\mathbf{u}, t)$, we estimate a set of warped rays $\mathbf{r}_q$ using G_Φ. We then render blurry colors through weighted averaging with L_{blur}, by evaluating F_Ω and additional explicit features $\mathcal{V}$. We supervise the color at mid-exposure through $\mathcal{L}_{EDI}$ by recovering a prior-based sharp color using the event double integral, considering all events in the exposure time. Finally, we sample a pair of two consecutive events, and supervise their brightness difference, modulated by eCRF, using the observed polarity value via $\mathcal{L}_{ev}$.

time t as $I(\mathbf{u}, t)$. An event $\mathbf{e}_j$ indicates that at time t_j, the log-brightness has changed by $p_j \cdot \Theta_{p_j}$ from the last time t_{j-1} an event has been generated from the pixel. The quantity $\Theta_{p_j} \in \mathbb{R}^+$ is a predefined threshold that controls the sensitivity to brightness. It follows that $L(\mathbf{u}, t_j) - L(\mathbf{u}, t_{j-1}) = p_j \cdot \Theta_{p_j}$, where we shorten $log(I(\cdot))$ as $L(\cdot)$.

We compute the left-hand side through volumetric rendering, while we take the right-hand side as a ground truth supervision, given recorded event pairs. In particular, we estimate the log-brightness at each event $\mathbf{e}_j$, observed by the pixel $\mathbf{u}$ at time t_j, as:

$$\hat{L}(\mathbf{u}, t_j) = log(h(eCRF_\Psi(\hat{\mathbf{C}}(\mathbf{r}_j), p_j))), \tag{6}$$

where we obtain $\hat{\mathbf{C}}(\mathbf{r}_j)$ via volumetric rendering [22] by rendering the ray $\mathbf{r}_j = \mathbf{r}(\mathbf{u}, t_j)$ cast from the camera pose $\mathbf{T}(t_j) \in SE(3)$, approximated via spherical linear interpolation [28] of the available known camera poses. Here, $eCRF_\Psi$ is an MLP that produces a modulated signal $\hat{\mathbf{C}}_e \in \mathbb{R}^3$ from the rendered color $\hat{\mathbf{C}}$ and the polarity p_j, while $h(\cdot)$ is a luma conversion function [34].

Given a pair of consecutive events at time t_{j-1} and t_j, we compute $\mathcal{L}_{\text{event}}$ as follows:

$$E^e(\Delta\hat{L}^t_{\mathbf{u}}) = \left\| \Delta\hat{L}^t_{\mathbf{u}} - \Delta L^t_{\mathbf{u}} \right\|_2^2 \tag{7}$$

$$\mathcal{L}_{\text{event}} = \frac{1}{|\mathcal{U}_e|} \sum_{(t,\mathbf{u}) \in \mathcal{U}_e} E^e(\Delta\hat{L}^t_{\mathbf{u}_c}) + E^e(\Delta\hat{L}^t_{\mathbf{u}_f}) + E^e(\Delta\tilde{L}^t_{\mathbf{u}_f}) \tag{8}$$

where we use the compact form $\hat{L}^t_{\mathbf{u}}$ for $\hat{L}(t, \mathbf{u})$, and apply the supervision on both fine and coarse levels, as well as on adaptively refined colors. $\mathcal{U}_e$ selects pairs of pixels $\mathbf{u}$ and timestamps t corresponding to received events. Our proposed event CRF function $eCRF_\Psi$ learns to compensate for potential mismatches between the ideal model and that of the camera at hand, filling the gap between the RGB color space and that of the event sensor.

Double integral supervision. The eCRF just introduced provides an effective way of handling unmodeled event pixel behaviors. However, blindly recovering the event camera response to colors is not trivial since the only direct source of color supervision comes from Equation (4). Inspired by recent works [18], which exploit additional priors, we propose here to exploit a model-based deblurring solution to further constrain the NeRF training. In particular, we first deblur training images utilizing the event-based double integral (EDI) [24, 24], which exploits the relationship between a blurry image and the events triggered during its exposure. We thus obtain $\mathbf{C}^{EDI}_{\mathbf{r}}$ deblurred colors corresponding to every original ray $\mathbf{r} = \mathbf{r}(\mathbf{u}, t_i)$ in $\mathcal{R}_b$, sampled when optimizing Eq. (4). We use this color as a prior:

$$E^{\text{EDI}}(\hat{\mathbf{C}}_r) = \left\| \hat{\mathbf{C}}_{\mathbf{r}} - \mathbf{C}^{\text{EDI}}_{\mathbf{r}} \right\|_2^2 \tag{9}$$

$$\mathcal{L}_{\text{EDI}} = \frac{1}{|\mathcal{R}_b|} \sum_{\mathbf{r} \in \mathcal{R}_b} E^{\text{EDI}}(\hat{\mathbf{C}}_{\mathbf{r}_c}) + E^{\text{EDI}}(\hat{\mathbf{C}}_{\mathbf{r}_f}) \tag{10}$$

4. Experiments

We evaluate our method on four synthetic scenes derived from the original DeblurNeRF [20] work, as well as a novel dataset composed of 5 scenes captured with a Color-DAVIS346 [17], mounted on a motor-controlled linear slider that moves the camera and collects ground truth poses. Additional details on the training and the datasets, named Ev-DeblurBlender and Ev-DeblurCDAVIS, are provided in the supplementary material. We evaluate our

Table 1. Quantitative comparison on the synthetic Ev-DeblurBlender dataset. Best results are reported in bold.

	Factory			Pool			Tanabata			Trolley			Average		
	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑
DeblurNeRF [20]	24.52	0.25	0.79	26.02	0.34	0.69	21.38	0.28	0.71	23.58	0.22	0.79	23.87	0.27	0.74
BAD-NeRF [40]	21.20	0.22	0.64	27.13	0.23	0.70	20.89	0.25	0.65	22.76	0.18	0.73	22.99	0.22	0.68
PDRF [6]	27.34	0.17	0.87	27.46	0.32	0.72	24.27	0.20	0.81	26.09	0.15	0.86	26.29	0.21	0.81
DP-NeRF [16]	26.77	0.20	0.85	29.58	0.24	0.79	27.32	0.11	0.85	27.04	0.14	0.87	27.68	0.17	0.84
ENeRF [14]	22.46	0.19	0.79	25.51	0.28	0.72	22.97	0.16	0.83	21.07	0.20	0.80	23.00	0.21	0.79
E^{2}NeRF [25]	24.90	0.17	0.78	29.57	0.18	0.78	23.06	0.19	0.74	26.49	0.10	0.85	26.00	0.16	0.78
(Ours) Ev-DeblurNeRF	**31.79**	**0.06**	**0.93**	**31.51**	**0.14**	**0.84**	**28.67**	**0.08**	**0.90**	**29.72**	**0.07**	**0.92**	**30.42**	**0.08**	**0.90**

Table 2. Quantitative comparison on the real-world Ev-DeblurCDAVIS dataset. Best results are reported in bold.

	Batteries			Power supplies			Lab Equipment			Drones			Figures			Average		
	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑
DP-NeRF [16] + TensoRF [5]	26.64	0.27	0.81	25.74	0.32	0.77	27.49	0.31	0.80	26.52	0.30	0.81	27.76	0.34	0.77	26.83	0.31	0.79
EDI [24] + NeRF	28.66	0.12	0.87	28.16	0.09	0.88	31.45	0.13	0.89	29.37	0.10	0.88	31.44	0.12	0.88	29.82	0.11	0.88
E^{2}NeRF	30.57	0.12	0.88	29.98	0.11	0.87	30.41	0.16	0.86	30.41	0.14	0.87	31.03	0.14	0.85	30.48	0.13	0.87
(Ours) Ev-DeblurNeRF	**33.17**	**0.05**	**0.92**	**32.35**	**0.06**	**0.91**	**33.01**	**0.08**	**0.91**	**32.89**	**0.05**	**0.92**	**33.39**	**0.07**	**0.90**	**32.96**	**0.06**	**0.91**

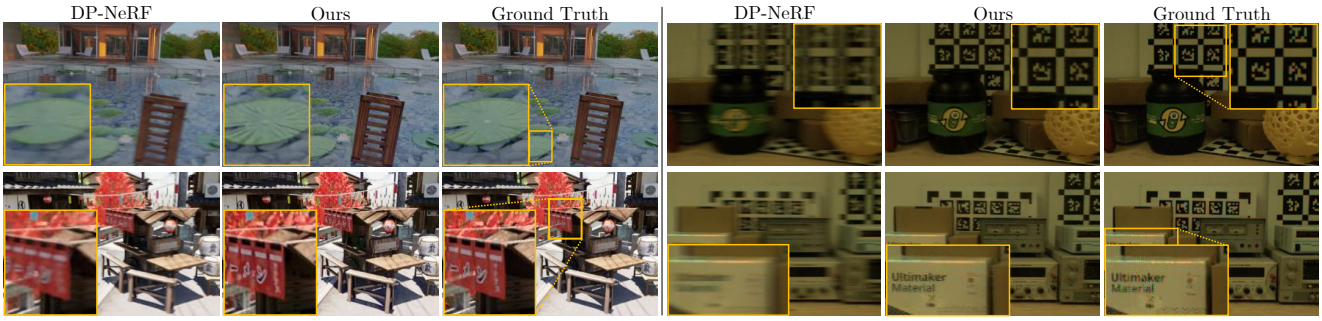

Figure 3. Qualitative comparison on the Ev-DeblurBlender (left) and Ev-DeblurCDAVIS (right) datasets.

method against Deblur-NeRF [20], BAD-NeRF [40], DP-NeRF [16] and PDRF [6]. Finally, we also consider event-based NeRFs, such as E-NeRF [14] and E^{2}NeRF [25].

Results. We report results on Ev-DeblurBlender in Table 1. Our proposed approach largely outperforms all other baselines, both event-based and frame-based. Compared to DP-NeRF [16], which uses a similar backbone architecture, our method achieves on average a $+3.34$dB higher PSNR, a 52.9% lower LPIPS [48] and 7.14% higher SSIM, highlighting the improvement gained by effectively integrating event-based supervision. Notably, ENeRF [14], which does not explicitly model the blur formation process, struggles to recover sharp color information, while E^{2}NeRF [25], exclusively employing event supervision during the exposure time, fails at fully exploiting event-based data.

Results on Ev-DeblurCDAVIS are reported in Table 2, where we select the top-performing NeRF models from the previous evaluation, namely E^{2}NeRF [25] and DP-NeRF [16], which we modify here by integrating the TensoRF modules discussed in Section 3 for a better comparison. Additionally, we include results obtained by initially deblurring images using the model-based EDI deblurring method, followed by NeRF. An extended analysis including all other baselines is provided in the supplementary materials. Once again, our proposed approach significantly outperforms all baselines, exhibiting an improvement of $+2.5$dB in PSNR and a 4.6% increase in SSIM. A qualitative comparison is provided in Figure 3 and in the supplementary materials.

5. Conclusions

We present Ev-DeblurNeRF, a novel deblur NeRF architecture that integrates a learnable event-based camera response function and ad-hoc event-based supervision that facilitates fine-grained details recovery. Ev-DeblurNeRF, despite being supervised by model-based priors, can adapt to non-idealities in the camera response, potentially departing from the model-based solution. We validate our method on both synthetic and real data, achieving an increase of $+4.42$dB and $+2.48\%$dB in PSNR, respectively, when compared to the previous best-performing event-based baseline, and an increase of $+2.74$dB and $+6.13$dB when compared to the top-performing image-only baseline.

Acknowledgements. This work was supported by the National Centre of Competence in Research (NCCR) Robotics (grant agreement No. 51NF40-185543) through the Swiss National Science Foundation (SNSF), and the European Research Council (ERC) under grant agreement No. 864042 (AGILEFLIGHT).

References

[1] Dejan Azinovic, Ricardo Martin-Brualla, Dan B Goldman, Matthias Niebner, and Justus Thies. Neural RGB-d surface reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6290–6301. IEEE, 2022. 1

[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 *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864. IEEE, 2021. 1

[3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5479. IEEE, 2022. 1

[4] Anish Bhattacharya, Ratnesh Madaan, Fernando Cladera, Sai Vemprala, Rogerio Bonatti, Kostas Daniilidis, Ashish Kapoor, Vijay Kumar, Nikolai Matni, and Jayesh K Gupta. Evdnerf: Reconstructing event data with dynamic neural radiance fields. *arXiv preprint arXiv:2310.02437*, 2023. 1

[5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial radiance fields. In *Lecture Notes in Computer Science*, pages 333–350. Springer Nature Switzerland, 2022. 2, 4

[6] Peng Cheng and Chellappa Rama. Pdrf: Progressively deblurring radiance field for fast and robust scene reconstruction from blurry images, 2023. 1, 2, 4

[7] Kyle Gao, Yina Gao, Hongjie He, Denning Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*, 2022. 2

[8] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. Neuromorphic camera guided high dynamic range imaging. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1730–1739. IEEE, 2020. 2

[9] Chen Haoyu, Teng Minggui, Shi Boxin, Wang YIzhou, and Huang Tiejun. Learning to deblur and generate high frame rate video with an event camera. *arXiv preprint arXiv:2003.00847*, 2020. 2

[10] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic DVS events. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1312–1321. IEEE, 2021. 1

[11] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. HDR-NeRF: High dynamic range neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18398–18408. IEEE, 2022. 1

[12] Inwoo Hwang, Junho Kim, and Young Min Kim. Ev-NeRF: Event based neural radiance field. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 837–847. IEEE, 2023. 1, 2

[13] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3320–3329. IEEE, 2020. 2

[14] 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. 1, 2, 4

[15] Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. Panoptic neural fields: A semantic object-aware neural scene representation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12871–12881. IEEE, 2022. 1

[16] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. DP-NeRF: Deblurred neural radiance field with physical scene priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12386–12396. IEEE, 2023. 1, 2, 4

[17] Chenghan Li, Christian Brandli, Raphael Berner, Hongjie Liu, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. 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, IEEE, 2015. 2, 3

[18] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6498–6508. IEEE, 2021. 3

[19] Zhizheng Liu, Francesco Milano, Jonas Frey, Roland Siegwart, Hermann Blum, and Cesar Cadena. Unsupervised continual semantic adaptation through neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3031–3040, 2023. 1

[20] 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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12861–12870. IEEE, 2022. 1, 2, 3, 4

[21] Nico Messikommer, Stamatios Georgoulis, Daniel Gehrig, Stepan Tulyakov, Julius Erbach, Alfredo Bochicchio, Yuanyou Li, and Davide Scaramuzza. Multi-bracket high dynamic range imaging with event cameras. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 547–557. IEEE, 2022. 2

[22] 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. In *ECCV*, 2020. 1, 3

[23] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16190–16199. IEEE, 2022. 1

[24] 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 *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6820–6829. IEEE, 2019. 1, 2, 3, 4

[25] 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. 1, 2, 4

[26] Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. Eventnerf: Neural radiance fields from a single colour event camera. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 1, 2

[27] Wei Shang, Dongwei Ren, Dongqing Zou, Jimmy S Ren, Ping Luo, and Wangmeng Zuo. Bringing events into video deblurring with non-consecutively blurry frames. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4531–4540. IEEE, 2021. 1

[28] Ken Shoemake. Animating rotation with quaternion curves. *ACM SIGGRAPH Computer Graphics*, 19(3):245–254, 1985. 3

[29] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J. Davison. *iMAP: Implicit Mapping and Positioning in Real-Time*, pages 6229–6238. IEEE, 2021. 1

[30] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *Lecture Notes in Computer Science*, pages 412–428. Springer, Springer Nature Switzerland, 2022. 1, 2

[31] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18043–18052, 2023. 2

[32] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16155–16164. IEEE, 2021. 2

[33] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric nonlinear flow and multi-scale fusion. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17755–17764. IEEE, 2022. 2

[34] International Telecomunication Union. Studio encoding parameters of digital television for standard 4: 3 and widescreen 16: 9 aspect ratios. bt.709. *Technical Report, International Telecomunication Union*, 2011. 3

[35] 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, IEEE, 2022. 1

[36] Patricia Vitoria, Stamatios Georgoulis, Stepan Tulyakov, Alfredo Bochicchio, Julius Erbach, and Yuanyou Li. Event-based image deblurring with dynamic motion awareness. In *Eur. Conf. Comput. Vis.* Springer Nature Switzerland, 2022. 1

[37] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *Computer Vision – ECCV 2020*, pages 155–171. Springer International Publishing, 2020. 2

[38] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. NeRF-SR: High quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454. ACM, 2022. 1

[39] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *NeurIPS*, 2021. 1

[40] 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*, pages 4170–4179, 2023. 1, 2, 4

[41] Song Wu, Kaichao You, Weihua He, Chen Yang, Yang Tian, Yaoyuan Wang, Ziyang Zhang, and Jianxing Liao. Video interpolation by event-driven anisotropic adjustment of optical flow. In *Lecture Notes in Computer Science*, pages 267–283. Springer Nature Switzerland, 2022. 2

[42] Christopher Xie, Keunhong Park, Ricardo Martin-Brualla, and Matthew Brown. FiG-NeRF: Figure-ground neural radiance fields for 3d object category modelling. In *2021 International Conference on 3D Vision (3DV)*, pages 962–971. IEEE, IEEE, 2021. 1

[43] Fang Xu, Lei Yu, Bishan Wang, Wen Yang, Gui-Song Xia, Xu Jia, Zhendong Qiao, and Jianzhuang Liu. Motion deblurring with real events. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2583–2592. IEEE, 2021. 2

[44] Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. iNeRF: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330. IEEE, IEEE, 2021. 1

[45] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1

[46] Hongguang Zhang, Limeng Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Event-guided multi-patch network with self-supervision for non-uniform motion deblurring. *International Journal of Computer Vision*, 131(2):453–470, 2022. 1

[47] Limeng Zhang, Hongguang Zhang, Jihua Chen, and Lei Wang. Hybrid deblur net: Deep non-uniform deblurring with event camera. *IEEE Access*, 8:148075–148083, 2020. 2

[48] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep

features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595. IEEE, 2018. 4

[49] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. NICE-SLAM: Neural implicit scalable encoding for SLAM. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12–786. IEEE, 2022. 1