Mitigating Motion Blur in Neural Radiance Fields with Events and Frames
Supplementary Material

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

1. Implementation Details

Training. We implement Ev-DeblurNeRF building upon the DP-NeRF [5] official codebase implemented in PyTorch [9], and incorporating additional features from PDRF [2] and TensoRF [1]. We train both our Ev-DeblurNeRF network and the baselines on full-resolution images using either an NVIDIA V100, an NVIDIA RTX A6000, or an NVIDIA A100 GPU. In particular, we use 600×400 images for Ev-DeblurBlender and 346×260 for Ev-DeblurCDAVIS. Similar to [2, 5, 7], we warm up the training for the first $1,200$ iterations, by using at first only the $\mathcal{L}_{EDI}$ and $\mathcal{L}_{\text{ev}}$ losses and without utilizing the $eCRF$ module. Subsequently, we introduce $\mathcal{L}_{blur}$, along with the proposed $eCRF$, which we initialize as the identity function, and the blur estimation module G_Φ, keeping the λ parameters ($\lambda_b = \lambda_{EDI} = 1$, and λ_e = 0.1) fixed for the entire duration of the training. To implement L_{EDI}, we precompute $C^{\mathbf{r}}_{EDI}$ images using Eq. (4) and directly sample them during training. When using the $\mathcal{L}_{\text{ev-color}}$ loss, we weigh the events' contributions by 0.4, 0.2, or 0.4 depending on whether the event corresponds to a red, green, or blue channel, as green pixels appear twice as often in an RGBG Bayer pattern. We use symmetric constant thresholds for the events, setting $\Theta = 0.2$ for synthetic events, and $\Theta = 0.25$ when using a real camera.

Architecture. The motion estimation module G_Φ is implemented following DP-NeRF [5] hyperparameters' choice, and using $M = 9$ exposure poses. Differently from [5], we implement the image embedding $\boldsymbol{l}_i$ using a simple set of learnable 32-dimensional parameters, instead of predicting them through an additional 4-layers MLP. We found this design to be easier to optimize and yield overall better results. We follow [5] to implement the refinement AWP module and employ the coarse-to-fine scheduling strategy to weight $\hat{\mathbf{C}}^{\text{blur}}_{\mathbf{r}_f}$ and $\tilde{\mathbf{C}}^{\text{blur}}_{\mathbf{r}_f}$ in $\mathcal{L}_{\text{blur}}$. However, we weigh their contribution equally in $\mathcal{L}_{\text{ev}}$ through the whole training, as we found the coarse-to-fine scheduling strategy not to improve the results. We implement F^c_Ω as a 2-layers MLP with ReLU activation, hidden dimension 64, and output dimension 16, followed by a 3-layers MLP with the same activation and hidden dimension, but output dimension 3. We use one of the output channels of the first MLP as the predicted density, while the rest is used by the second MLP to predict colors. The structure of F^f_Ω is analogous, but we use an output dimension of 128 for the first MLP and a 256 hidden dimension for both MLPs. We implement $\mathcal{V}_s$ and $\mathcal{V}_l$ with vector-matrix decomposition [1], using 16.7 million voxels in $\mathcal{V}_s$ and 134.2 million voxels in $\mathcal{V}_l$, and setting to $\{64, 16, 16\}$ the channel dimensions of the decomposed $\{X, Y, Z\}$ axes in both $\mathcal{V}_s$ and $\mathcal{V}_l$. The proposed Ev-DeblurNeRF architecture trains in around 3 hours and 30 minutes on an NVIDIA A100 GPU.

2. Extended Analysis on Ev-DeblurCDAVIS

State-of-the-art comparison. Section 4.2 of the paper provides an analysis on the Ev-DeblurCDAVIS dataset focused on the top-performing architectures selected from the synthetic evaluation. For completeness, we report in Table 4 of this supplementary material a comprehensive evaluation against all other baselines used in the paper. The trend follows that of the synthetic analysis, with image-only baselines performing worse than networks making use of either images deblurred through events or event-enhanced NeRFs. Notice that, we could not finetune EFNet [13] on Ev-DeblurCDAVIS as, differently from simulation, we do not have corresponding sharp images for each blurry training view. We designed the Ev-DeblurCDAVIS dataset in such a way as to ensure reliable ground truth collection, but also to showcase the ability of our network to tackle a known limitation of image-only DeblurNeRF-like architectures. While these networks work particularly well on random motion patterns, they fail in the presence of consistent blur, i.e., when the motion pattern is similar in each exposure. This is the case of Ev-DeblurCDAVIS, where image-only baselines such as DP-NeRF [5] and PDRF [2] struggle to remove blur (see Figures 5 and 9 of this supplementary material and Figure 3 of the paper). For similar reasons, BAD-NeRF diverges after a few training iterations on this dataset. We address this by fixing the rotation matrix to ground truth and optimizing the translation vector

Table 4. Extended 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↑
BAD-NeRF* [14]	27.32	0.26	0.82	26.42	0.32	0.79	27.84	0.31	0.81	26.96	0.31	0.81	28.21	0.35	0.77	27,.5	0.31	0.80
DP-NeRF [5] + TensoRF [1]	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
PDRF [2]	26.82	0.25	0.81	25.79	0.31	0.77	27.70	0.31	0.81	26.72	0.29	0.81	27.80	0.33	0.77	26.96	0.30	0.79
MPRNet [15] + NeRF	27.99	0.21	0.83	26.89	0.23	0.78	27.20	0.28	0.80	26.98	0.23	0.80	28.51	0.29	0.79	27.52	0.25	0.80
PVDNet [12] + NeRF	24.65	0.30	0.72	23.50	0.30	0.66	25.04	0.32	0.72	24.21	0.31	0.69	25.92	0.33	0.72	24.66	0.31	0.70
EFNet [13] + NeRF	29.85	0.13	0.88	29.10	0.13	0.87	30.28	0.18	0.88	29.72	0.14	0.88	30.62	0.17	0.85	29.91	0.15	0.87
EDI [8] + 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
ENeRF [4]	27.85	0.26	0.73	27.91	0.21	0.76	27.79	0.25	0.73	28.28	0.25	0.77	29.05	0.18	0.77	28.17	0.23	0.75
E^{2}NeRF [10]	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 5. Qualitative ablation study of the main components of the proposed Ev-DeblurNeRF network. Tables below each picture are drawn from Table 3 of the paper, and report the configuration used and the PSNR metric achieved in each case.

only (reported as BAD-NeRF* in the table). Despite this, our method still significantly outperforms BAD-NeRF. Our architecture, indeed, eliminates ambiguities in motion estimation as it leverages additional event-based supervision to further constrain the NeRF recovery, resulting in significantly higher performance.

Effect of using eCRF. In Figure 5 of this document, we complement the ablation study reported in Table 3 of the paper with a qualitative assessment of our network's key components. As discussed in the previous paragraph, the image-only architecture struggles in consistent blur conditions. Notably, incorporating event supervision significantly aids in the recovery of sharp details, as evident when comparing the first two settings in Figure 5. The performance further increases when adding the proposed eCRF module, as can be noticed in the checkerboard patterns on the background, the globe in the foreground, and the facial details of the figures. However, as discussed in the main paper, this improvement comes at the cost of over-augmented details and increased contrast, which are not present in the ground truth reference images. We attribute this phenomenon to the under-constrained optimization setting, which allows the eCRF module to freely augment these details as long as they appear correct once blurred though $\mathcal{L}_{\text{blur}}$. We solve this issue by adding an additional prior, in the form of $\mathcal{L}_{\text{EDI}}$, which further constrains the network in reconstructing accurate details. The improved quality is clearly demonstrated in Figure 5, where over-augmented details are removed, but without compromising essential details.

Event-by-event vs. Event-window loss. In this section, we extend the analysis of the robustness to training views

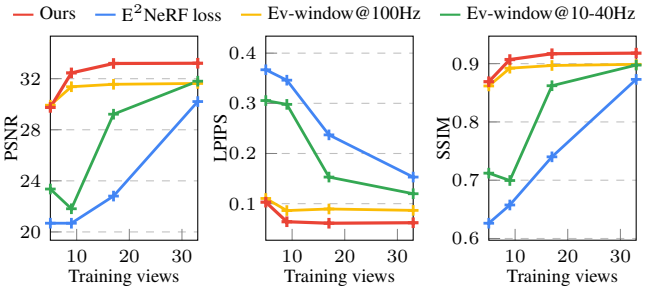

Figure 6. Analysis on event-by-event vs. event-batch losses.

reported in Figure 4-left of the paper. Specifically, utilizing our Ev-DeblurNeRF network, we examine the impact of implementing event supervision on an event-by-event basis, as we suggest in the paper and proposed in [4, 6], in contrast to accumulating events occurring over temporal windows [3, 11], as well as applying supervision only at specific times during the exposure time, as in E^{2}NeRF [10]. Results are reported in Figure 6. As the supervision frequency decreases, especially in sparse training views regimes, the performance also decreases. This observation aligns with the findings in [6], which suggest that noise effects and threshold variations in the event stream amplify with event accumulation, ultimately leading to a decrease in overall performance. Moreover, when only a few images are available for training, leveraging the continuous event stream to propagate absolute brightness measurements across unseen image views proves crucial for achieving top performance. Leveraging event-by-event supervision and incorporating a learnable camera response function to mitigate noise effects, our

Table 5. Extended study on motor encoder's vs. COLMAP's poses on Ev-DeblurCDAVIS. Best results in bold, second-best underlined.

	Train poses	Test-time refine	Batteries PSNR↑	Batteries LPIPS↓	Batteries SSIM↑	Power Supplies PSNR↑	Power Supplies LPIPS↓	Power Supplies SSIM↑	Lab Equipment PSNR↑	Lab Equipment LPIPS↓	Lab Equipment SSIM↑	Drones PSNR↑	Drones LPIPS↓	Drones SSIM↑	Figures PSNR↑	Figures LPIPS↓	Figures SSIM↑
Ours	Motor	–	33.17	**0.05**	0.92	**32.35**	**0.06**	**0.91**	33.01	**0.08**	**0.91**	**32.89**	0.52	**0.92**	33.39	0.07	0.90
Ours	Motor	✓	33.10	**0.05**	0.92	32.31	**0.06**	**0.91**	**33.05**	**0.08**	**0.91**	32.77	**0.05**	**0.92**	33.58	0.08	0.90
Ours	COLMAP	✓	**33.43**	**0.05**	**0.93**	32.18	**0.06**	**0.91**	33.01	**0.08**	**0.91**	32.69	**0.05**	0.91	**33.88**	**0.06**	**0.91**

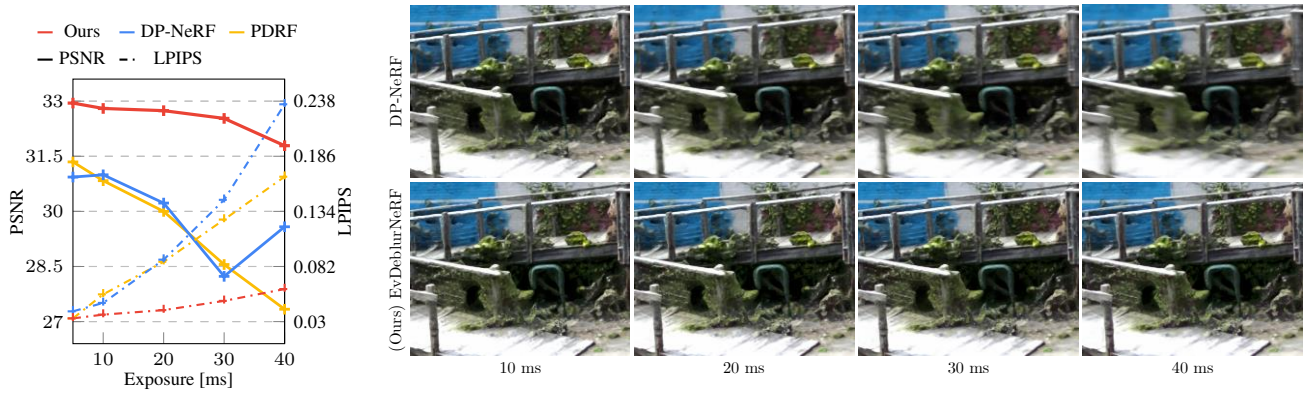

Figure 7. Robustness to motion blur analysis on the *factory* sample of Ev-DeblurBlender (left). Figures on the right show a qualitative comparison between DP-NeRF and Ev-DeblurNeRF among different exposures.

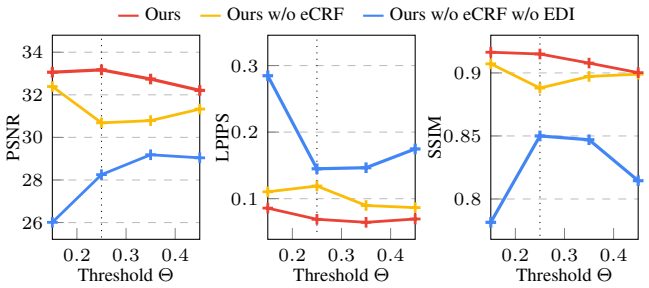

Figure 8. Analysis on the robustness to model mismatches.

approach achieves the best performance compared to other solutions.

COLMAP poses on real scenes. Experiments on Ev-DeblurCDAVIS presented in the paper make use of the poses obtained from the motor encoder, which tracks the camera's movement along the slider. However, in a typical real-world setting, access to such precise camera poses may not be available, although they are required by our method to work. In this section, we investigate a more general scenario where training poses are estimated using COLMAP instead of relying on the motor encoder.

Inspired by [10], we deblur training images using the EDI in Eq. (4) of the paper and then use COLMAP to estimate their poses. Analogous to the experiments conducted in the paper, we use spherical linear interpolation of the COLMAP poses to obtain poses at events' timestamps during training. At test time, we obtain the test poses by aligning the ground truth trajectory with that estimated with COLMAP. Since the two trajectories might not perfectly align, we further refine the alignment via gradient descent before computing metrics, as done in BAD-NeRF [14], to ensure pixel-perfect aligned test poses. We also include results of our method trained on encoder poses but evaluated using refined test poses. Results are presented in Table 5. Our method using COLMAP poses yields results comparable to those obtained using motor encoder poses, thus proving its potential in scenarios where accurate poses are not available. While performance degradation may occur in scenarios with more complex motion than that found in the Ev-DeblurCDAVIS dataset, further investigation into this aspect is left for future research endeavors.

3. Additional Results

Robustness to blur. In Figure 7 of this document, we supplement the analysis in Figure 4 of the main paper by comparing our Ev-DeblurNeRF network against the top-performing image-based baselines under different blur. We utilize the *factory* sample of the EV-DeblurBlender dataset for this analysis since it allows us to easily control the blur intensity, and it does not constitute a corner case for the image-only baselines. We change the exposure time τ of the simulated camera in the range $\tau \in \{5, 10, 20, 30, 40\}$, which results in an average pixel displacement of $\{3, 5, 11, 16, 20\}$, and a maximum displacement of $\{15, 24, 50, 75, 96\}$ in each configuration, respectively. The quantitative and qualitative comparison in Figure 7 shows that using events not only helps in cases of extreme motion but also helps when the motion is not extreme.

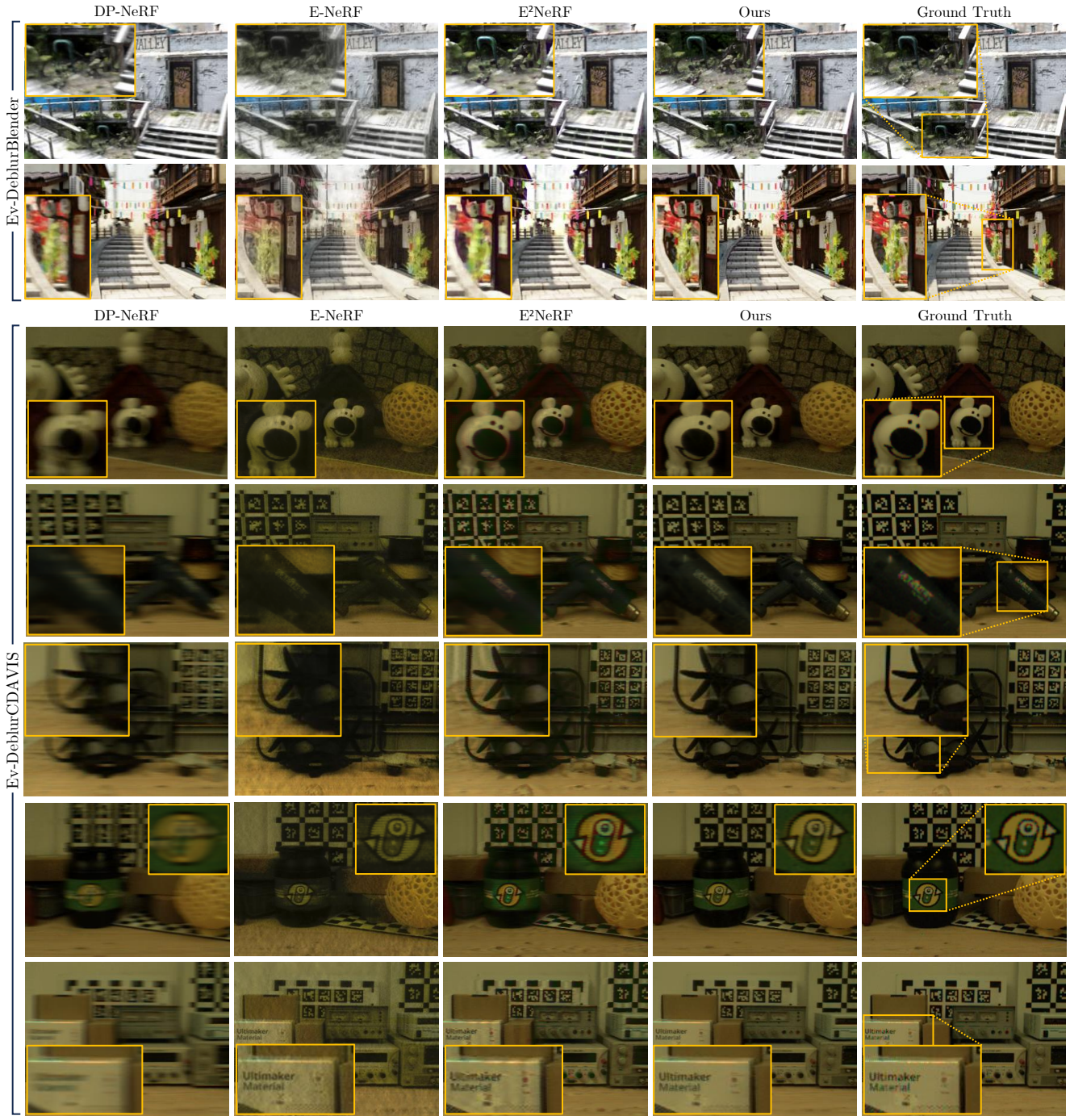

Figure 9. Qualitative comparison of synthetic (top) and real (bottom) motion blur. The remaining EV-DeblurBlender samples are provided in Figure 3 of the paper. Notice that, while the reference images can exhibit demosaicing artifacts around sharp edges, our reconstructs are less affected by these as we exploit multi-view supervision through NeRF and directly use color events without interpolation.

While image-only baselines recover details blindly, by trying to estimate the blur formation through a limited set of camera poses, our network can achieve higher-quality results as it exploits blur-free information carried by events at microseconds resolution. Notably, our solution shows great robustness to motion blur, while image-only performance decreases significantly as the blur increases. While these results consider synthetic data, where the effect of noise

and non-idealities is limited, they underscore the promise of event cameras as complementary sensors for attaining high-quality image synthesis even in non-ideal conditions.

Robustness to model mismatches. In this section, we analyze the proposed $eCRF$ module in terms of increased robustness to model mismatches. We do so in a real setup, i.e., on the *figures* sample of the Ev-DeblurCDAVIS dataset, by analyzing the sensitivity of our network to the Θ event-camera threshold. While in the paper we select Θ via manual inspection, i.e., by utilizing the event double integral [8] as visual feedback following [8], we evaluate here the performance of our model when the Θ used in $\mathcal{L}_{\text{ev}}$ deviates from this value. We compare our network against a configuration that does not use the proposed $eCRF$, as well as a network where we also remove $\mathcal{L}_{\text{EDI}}$. Results are reported in Figure 8. By acting in between the rendered color space and the $\mathcal{L}_{\text{ev}}$, $eCRF$ can modulate the brightness, or color, $\hat{L}^t$ used for computing the loss, acting as a residual between the model-based supervision (Equation (1) of the paper) and the brightness actually perceived. As a result, our solution achieves increased consistency across different choices of Θ, showcasing its ability to deviate from the model-based solution in case needed.

4. Qualitative Results and Video

We conclude this supplementary material by including extended qualitative results. In Figure 9, we complement Figure 3 of the paper by comparing the proposed method against top-performing networks across all remaining samples on the Ev-DeblurBlender and Ev-DeblurCDAVIS. Additionally, we provide a supplementary video showing qualitative results on all the samples of our proposed datasets. We strongly advise readers to watch our additional video where our method outperforms all baselines in rendering a novel-view continuous path, showing increased image quality and fewer artifacts than the other methods.

References

[1] 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. 1, 2

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

[3] 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. 2

[4] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-NeRF: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 8(3):1587–1594, 2023. 2

[5] 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

[6] Weng Fei Low and Gim Hee Lee. Robust e-nerf: Nerf from sparse & noisy events under non-uniform motion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18335–18346, 2023. 2

[7] 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

[8] 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. 2, 5

[9] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 1

[10] Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. E2nerf: Event enhanced neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13254–13264, 2023. 2, 3

[11] 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. 2

[12] Hyeongseok Son, Junyong Lee, Jonghyeop Lee, Sunghyun Cho, and Seungyong Lee. Recurrent video deblurring with blur-invariant motion estimation and pixel volumes. *ACM Transactions on Graphics*, 40(5):1–18, 2021. 2

[13] 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

[14] 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. 2, 3

[15] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14821–14831. IEEE, 2021. 2