Supplementary Material for "BeNeRF: Neural Radiance Fields from a Single Blurry Image and Event Stream"

Wenpu Li[1,5*], Pian Wan[1,2*], Peng Wang[1,3*], Jinghang Li[4], Yi Zhou[4], and Peidong Liu[1†]

[1]Westlake University, [2]EPFL, [3]Zhejiang University, [4]Hunan University
[5]Guangdong University of Technology

A Introduction

In this supplementary materials, we provide additional details and analysis on the ablation experiments. Furthermore, we present a more detailed quantitative comparison on synthetic datasets and an extensive qualitative comparison on both synthetic and real-world datasets by showcasing the deblurred images produced by various methods, facilitating a comprehensive evaluation. Finally, we provide the latent sharp videos recovered from a single blurred image using our method.

B More ablation studies

Effect of event stream. We evaluate the effect of event stream on Livingroom dataset. Quantitative results are shown in Table 1 and qualitative results are presented in Fig. 1. It demonstrate that event streams can constrain the ill-posed problem caused by motion blur and only single view image as training data. Our method effectively incorporates event stream to guide NeRF to learn the correct underlying scene representation., significantly improving performance. Thus, the introduction of event streams is highly motivated.

Table 1: Ablation studies on event stream. The results demonstrate that leveraging event stream can dramatically boost the performance of BeNeRF. We validate that introducing event streams is an effective method to constrain the ill-posed problem caused by motion blur and limited to a single-view image as training data.

	Livingroom		
	PSNR↑	SSIM↑	LPIPS↓
w/o event stream	24.40	.6612	.4712
w/ event stream	37.11	.9370	.0632

⋆ Equal contribution: {liwenpu, wangpeng}@westlake.edu.cn, pianwan@outlook.com
† Corresponding author: Peidong Liu(liupeidong@westlake.edu.cn).

Fig. 1: Qualitative results of ablation studies on event stream. Here we visualize how our method benefits from event stream. The qualitative results indicate that using event stream can lead to improved model performance.

Effect of the coarse-to-fine strategy for training. We experiment with the effect of introducing the coarse-to-fine strategy used in BARF [4] to train our model. The results in Table 2 show that incorporating the coarse-to-fine strategy slightly decreases model performance. It might be due that the coarse-to-fine strategy is more suit-able for multi-view case.

Table 2: Ablation studies on coarse-to-fine strategy for training. The results demonstrate that using coarse-to-fine strategy to train model does not further improve the performance

	Livingroom		
	PSNR↑	SSIM↑	LPIPS↓
w/o coarse-to-fine	37.11	.9370	.0632
w/ coarse-to-fine	34.20	.8931	.1450

C More quantitative results

Due to space constraints, we did not report the comparative results of SSIM metrics for each method on the synthetic dataset in the main text. In the supplementary material, we provide comprehensive quantitative comparison results, including metrics such as PSNR, SSIM, and LPIPS. The detailed quantitative results on our synthetic datasets are shown in Table 3, Table 4, Table 5. The experimental results presented in Table 3 demonstrate a significant superiority of our method over existing state-of-the-art single image deblurring methods. Similarly, the results in Table 4 exhibit superior performance of our method compared to event-enhanced single image deblurring methods. Additionally, the experimental results in Table 5 showcase the remarkable efficacy of our approach. Despite utilizing only a single blurred image and an event stream of a limited time interval, our method achieves performance comparable to E^{2}NeRF [8], which employs multi-view images and a longer event stream, particularly in terms of the PSNR metric. Furthermore, our method even surpasses E^{2}NeRF [8] in terms of the LPIPS metric.

Table 3: Detailed quantitative comparisons on single image deblurring with synthetic datasets. The results demonstrate that our method significantly performs better than prior learning-based methods in terms of image quality. For HINet and NAFNet, we tests pre-trained weights from both GoPro and REDS datasets(*).

	PSNR ↑					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
DeblurGANv2 [3]	29.26	27.64	23.16	20.09	26.89	25.41
SRN-deblur [11]	30.86	27.59	23.12	19.89	27.79	25.85
MPRNet [14]	28.57	26.49	21.60	18.20	27.02	24.38
HINet [2]	28.56	26.27	21.91	18.59	26.70	24.41
HINet* [2]	27.55	22.89	20.25	18.15	27.14	23.20
NAFNet [1]	29.92	28.16	22.41	18.96	26.75	25.24
NAFNet* [1]	28.18	23.67	20.85	18.38	27.52	23.72
Restormer [13]	29.48	27.39	22.22	18.82	27.35	25.05
BeNeRF	37.11	32.95	29.68	32.14	36.38	33.65
	SSIM ↑					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
DeblurGANv2 [3]	.8121	.7235	.7043	.4964	.6123	.6697
SRN-deblur [11]	.8437	.7396	.7043	.5111	.6572	.6912
MPRNet [14]	.7937	.7301	.6547	.4258	.6253	.6459
HINet [2]	.7920	.6950	.6625	.4411	.6235	.6428
HINet* [2]	.7822	.6122	.6019	.4155	.6211	.6066
NAFNet [1]	.8306	.7874	.6896	.4665	.6255	.6799
NAFNet* [1]	.7991	.6422	.6175	.4230	.6407	.6245
Restormer [13]	.8262	.7314	.6803	.4596	.6352	.6665
BeNeRF	.9370	.8651	.8593	.9015	.9039	.8934
	LPIPS ↓					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
DeblurGANv2 [3]	.2087	.1989	.2608	.3934	.3100	.2744
SRN-deblur [11]	.2529	.2503	.3245	.4260	.3594	.3226
MPRNet [14]	.2621	.2564	.3586	.4173	.3679	.3325
HINet [2]	.2468	.2620	.3500	.4024	.3355	.3193
HINet* [2]	.3327	.3602	.3789	.5265	.4397	.4076
NAFNet [1]	.2268	.1991	.3058	.3908	.3280	.2901
NAFNet* [1]	.3182	.3566	.3943	.5271	.4257	.4044
Restormer [13]	.2391	.2493	.3373	.4248	.3664	.3234
BeNeRF	.0632	.0788	.0761	.0515	.0677	.0675

Table 4: Detailed quantitative comparisons on event-enhanced single image deblurring with synthetic datasets. The results demonstrate that our method performs better than both EDI and eSLNet.

	PSNR ↑					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
eSLNet [12]	14.22	10.81	10.49	8.86	11.80	11.24
EDI [7]	32.61	30.33	27.24	24.87	31.64	29.34
BeNeRF	37.11	32.95	29.68	32.14	36.38	33.65
	SSIM ↑					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
eSLNet [12]	.3527	.2156	.2903	.1658	.2181	.2485
EDI [7]	.8871	.8152	.8356	.7564	.8044	.8197
BeNeRF	.9370	.8651	.8593	.9015	.9039	.8934
	LPIPS ↓					
	Livingroom	Whiteroom	Pinkcastle	Tanabata	Outdoorpool	*Average*
eSLNet [12]	.3981	.4236	.4902	.5067	.4676	.4572
EDI [7]	.0904	.1020	.0779	.1039	.1409	.1030
BeNeRF	.0632	.0788	.0761	.0515	.0677	.0675

Table 5: Detailed quantitative comparisons on NeRF-based image deblurring with synthetic datasets from E^2NeRF. The results indicate that our method outperforms both NeRF and Deblur-NeRF, and exhibits performance comparable to E^2NeRF in terms of the PSNR metric. Moreover, our method even surpasses E^2NeRF with the LPIPS metric.

	PSNR ↑						
	Chair	Ficus	Hotdog	Lego	Materials	Mic	*Average*
NeRF [6]	24.29	22.98	27.75	21.95	19.99	20.50	22.91
Deblur-NeRF [5]	25.87	22.86	24.62	24.47	20.54	11.92	21.71
E^2NeRF [8]	31.28	30.00	34.34	28.11	27.27	27.60	29.77
BeNeRF	31.17	30.81	34.31	28.09	27.44	26.13	29.66
	SSIM ↑						
	Chair	Ficus	Hotdog	Lego	Materials	Mic	*Average*
NeRF [6]	.9357	.9023	.9546	.8548	.9108	.8854	.9072
Deblur-NeRF [5]	.9373	.8982	.9396	.8756	.9012	.7249	.8795
E^2NeRF [8]	.9749	.9663	.9784	.9339	.9570	.9496	.9600
BeNeRF	.9488	.9465	.9497	.8930	.9144	.9115	.9273
	LPIPS ↓						
	Chair	Ficus	Hotdog	Lego	Materials	Mic	*Average*
NeRF [6]	.1254	.1037	.1158	.2103	.1512	.1579	.1441
Deblur-NeRF [5]	.2185	.1541	.2138	.2053	.2562	.3706	.2364
E^2NeRF [8]	.0608	.0362	.0660	.1078	.0919	.0724	.0725
BeNeRF	.0500	.0299	.0539	.0745	.0708	.0738	.0588

D More qualitative results

We conducted a detailed comparison of our proposed method with single-image deblurring methods and event-enhanced single-image deblurring methods on our proposed synthetic dataset, consisting of six scenes(i.e. Livingroom, Whiteroom, Pinkcastle, Tanabata and Outdoorpool). The motion blur images are synthesized by importing real camera motion trajectories from ETH3D [10]. Thorough qualitative comparison in Fig. 2 and Fig. 3 illustrate that even under severe motion blur conditions, our method can effectively recover sharp images, demonstrating significant superiority over existing state-of-the-art single-image deblurring methods. Since the event stream synthesized using ESIM [9] is monochannel, the images recovered using EDI [7] are grayscale.

To facilitate a comparative analysis with a NeRF-based image deblurring method leveraging multi-view information, we conducted a detailed evaluation on the synthetic dataset proposed by E^{2}NeRF [8], which comprises six distinct scenes(i.e. Chair, Ficus, Hotdog, Lego, Materials, and Mic). Extensive qualitative comparisons in Fig. 4 and Fig. 5 demonstrate the superiority of our method, despite utilizing only a single blurred image and a short event stream, over methods that utilize multi-view images and long event streams.

Fig. 6 and Fig. 7 illustrate comprehensive comparisons on real-world datasets. The results demonstrate the superior performance of our method on real datasets, attributed to its enhanced capability in modeling the physical process of motion-blurred imaging.

E Supplementary videos

To showcase the effectiveness of our approach, we provide a supplementary video illustrating its capability to recover high-quality latent sharp video from a single blurry image and corresponding event stream, which encapsulates rich temporal information. These videos are available on the project page. Furthermore, our results highlight the superior performance of our method in comparison to previous state-of-the-art approaches.

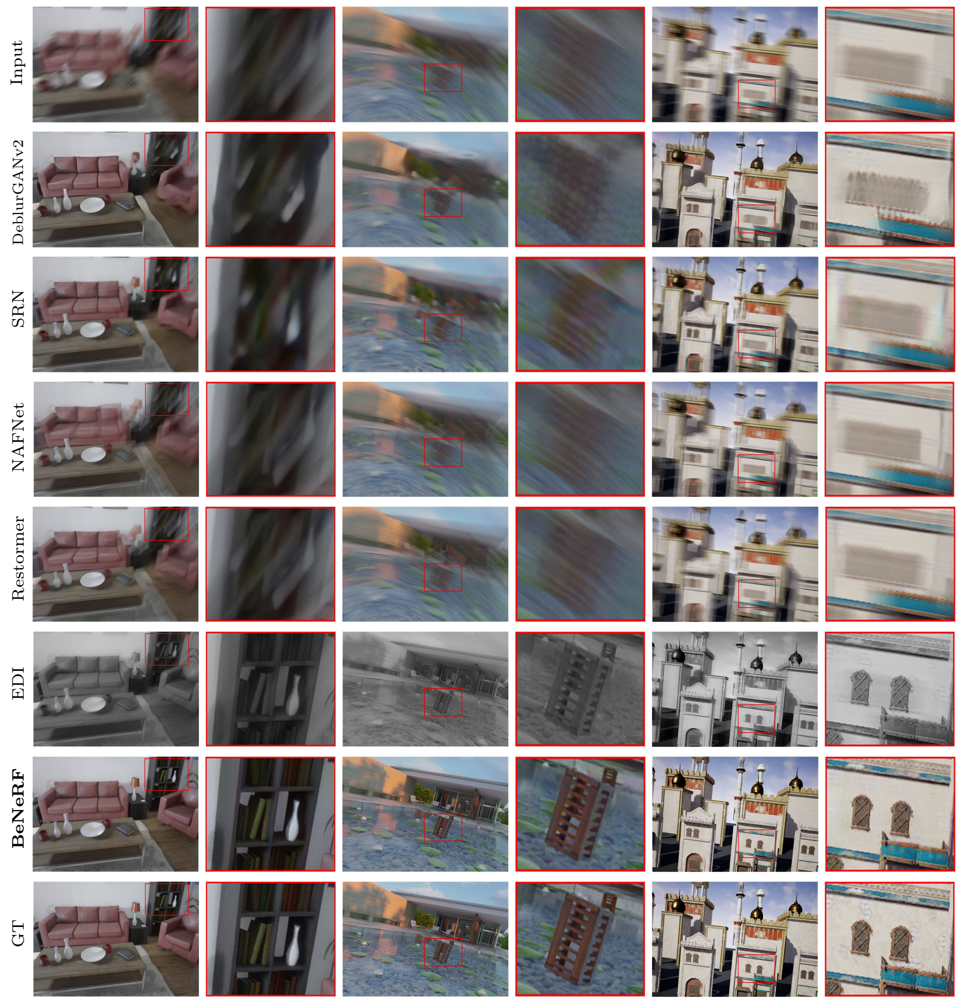

Fig. 2: Qualitative results of different methods with synthetic datasets. Detailed qualitative comparison for “Livingroom”, "Outdoorpool" and "Pinkcastle" scene of synthetic dataset.

Fig. 3: Qualitative results of different methods with synthetic datasets. Detailed qualitative comparison for "Tanabata" and "Whiteroom" scene of synthetic dataset.

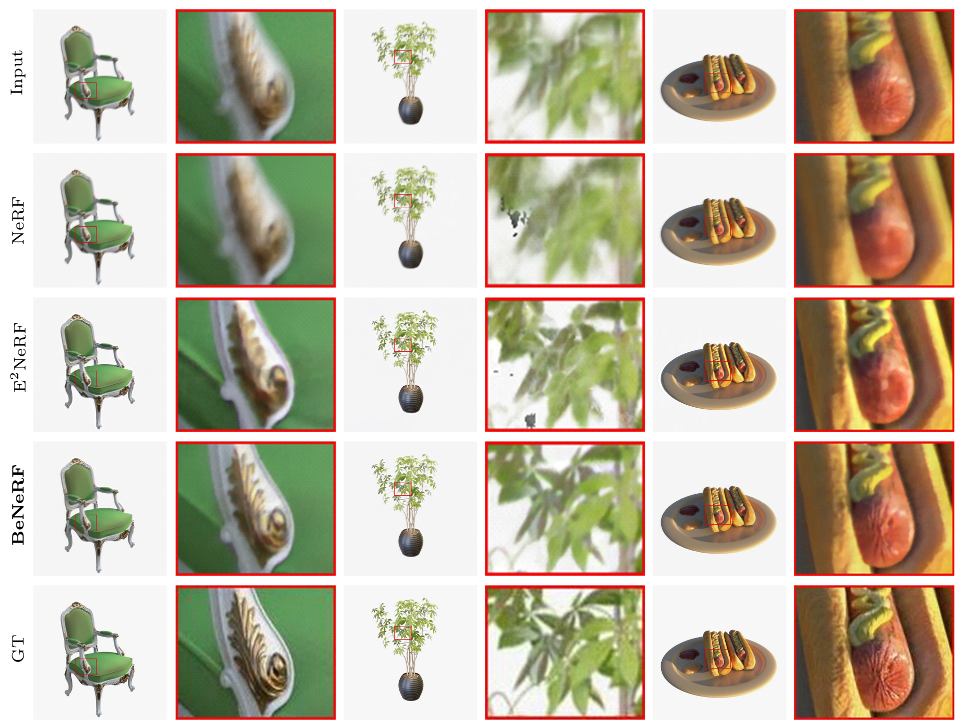

Fig. 4: Qualitative results of different methods with synthetic datasets proposed by E^{2}NeRF. Detailed qualitative comparison for “Chair”, "Ficus" and "Hotdog" scene of synthetic dataset from E^{2}NeRF.

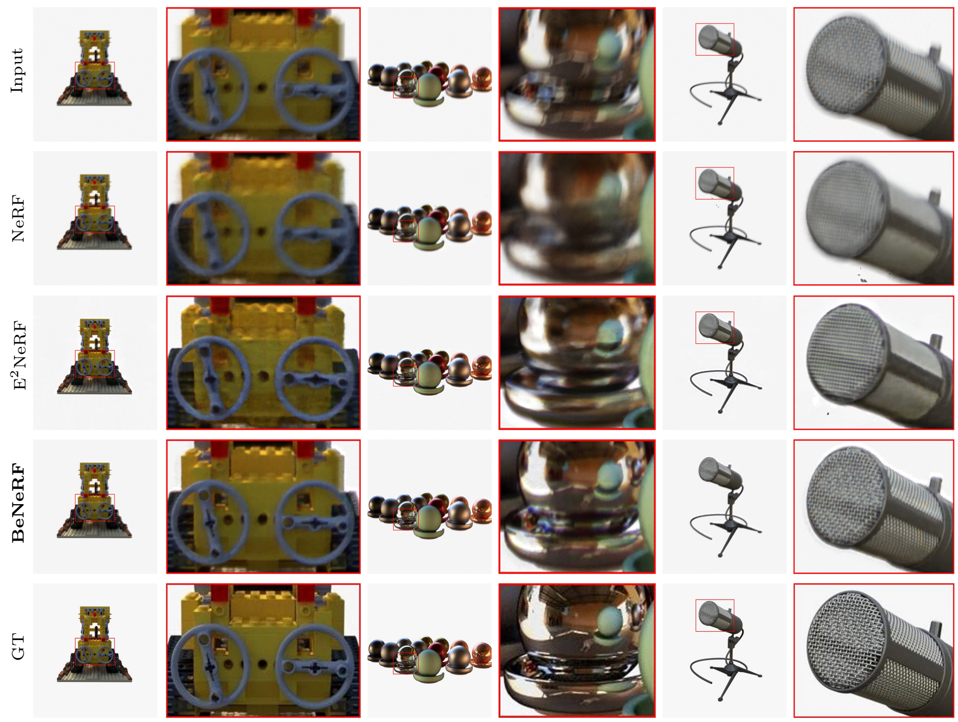

Fig. 5: Qualitative results of different methods with synthetic datasets proposed by E^{2}NeRF. Detailed qualitative comparison for "Lego", "Materials" and "Mic" scene of synthetic dataset from E^{2}NeRF.

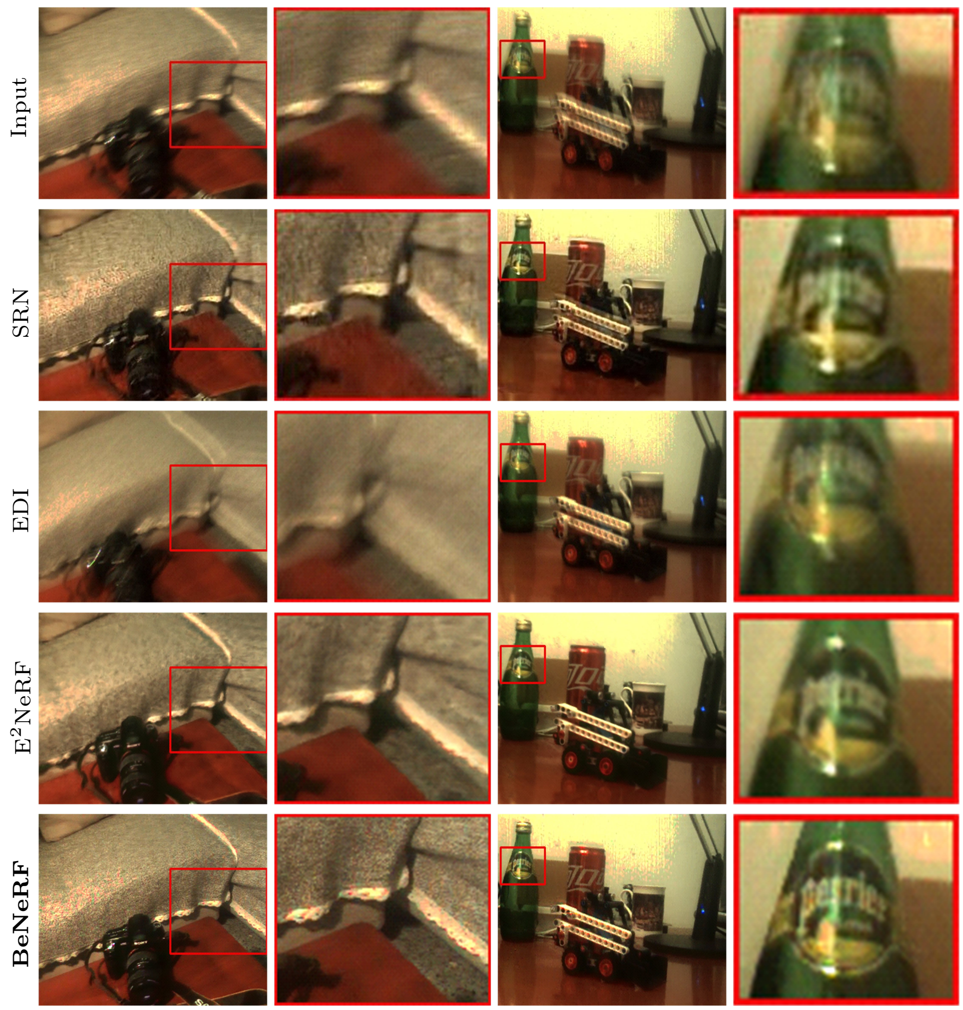

Fig. 6: Qualitative results of different methods with real-world datasets. Detailed qualitative comparison for “Camera” and "Lego" scene of real-world dataset.

Fig. 7: Qualitative results of different methods with real-world datasets. Detailed qualitative comparison for “Letter”, "Plant" and "Toys" scene of real-world dataset.

References

1. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple Baselines for Image Restoration. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 17–33. Lecture Notes in Computer Science, Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-20071-7_2
2. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: HINet: Half Instance Normalization Network for Image Restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 182–192 (2021)
3. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8878–8887 (2019)
4. Lin, C.H., Ma, W.C., Torralba, A., Lucey, S.: BARF: Bundle-Adjusting Neural Radiance Fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5741–5751 (2021)
5. Ma, L., Li, X., Liao, J., Zhang, Q., Wang, X., Wang, J., Sander, P.V.: Deblur-NeRF: Neural Radiance Fields From Blurry Images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12861–12870 (2022)
6. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis (Aug 2020). https://doi.org/10.48550/arXiv.2003.08934
7. Pan, L., Scheerlinck, C., Yu, X., Hartley, R., Liu, M., Dai, Y.: Bringing a Blurry Frame Alive at High Frame-Rate With an Event Camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6820–6829 (2019)
8. Qi, Y., Zhu, L., Zhang, Y., Li, J.: E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13254–13264 (2023)
9. Rebecq, H., Gehrig, D., Scaramuzza, D.: ESIM: An Open Event Camera Simulator. In: Proceedings of The 2nd Conference on Robot Learning. pp. 969–982. PMLR (Oct 2018)
10. Schops, T., Sattler, T., Pollefeys, M.: BAD SLAM: Bundle Adjusted Direct RGB-D SLAM. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 134–144 (2019)
11. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-Recurrent Network for Deep Image Deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8174–8182 (2018)
12. Wang, B., He, J., Yu, L., Xia, G.S., Yang, W.: Event Enhanced High-Quality Image Recovery. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 155–171. Lecture Notes in Computer Science, Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58601-0_10
13. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient Transformer for High-Resolution Image Restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022)
14. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-Stage Progressive Image Restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021)