Supplementary Material for "E^{2}NeRF: Event Enhanced Neural Radiance Fields from Blurry Images"

Yunshan Qi[1] Lin Zhu[2*] Yu Zhang[3] Jia Li[1,4*]

[1]State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University
[2]Beijing Institute of Technology [3]SenseTime and Tetras.AI [4]Peng Cheng Laboratory

{qi_yunshan, jiali}@buaa.edu.cn, linzhu@bit.edu.cn, zhangyulb@gmail.com

1. Synthetic data details

In our framework, we use synthetic data to evaluate our model. As shown in Fig. 10, camera shaking is applied to each view of a synthetic scene to generate a sequence of sharp frames, with a total number of 17 in our work. The pose information for the synthetic data is obtained from the five poses of frame numbers 1, 5, 9, 13, and 17.

To synthesize the blurred image, we first convert the 17 frames into raw images through the inverse ISP process. Since the exposure time and interval time of each image in the sequence are identical, the average of 17 raw images is taken directly as the blurred raw image. Finally, after ISP processing, we obtain the final synthetic blurred image.

Next, we utilize the v2e simulator to generate the corresponding event stream to the blurred image. The "dvs model" option in v2e is set to "noisy", which adds motion blur, latency and noise of event data during the simulation process, resulting in simulated event data that closer to the real data captured by the event camera. The input of the v2e simulator are the same 17 frames and eventually, it obtains the synthetic event data.

2. Additional quantitative analysis

The detailed quantitative results on six synthetic scenarios are shown in Tab. 7 and Tab. 8. We divide the experimental results into two groups: blur view and novel view. Our method shows better performance, especially on novel view, which indicates that our method can learn a more precise 3D representation of the scene with event data. Although our E^{2}NeRF does not achieve the best result in several scenes and metrics, the average result of our method is the best as shown in Tab. 1 and Tab. 2 of the main manuscript, which proves the effectiveness of our model.

*Correspondence should be addressed to Lin Zhu and Jia Li. Website: https://cvteam.buaa.edu.cn

3. Additional qualitative analysis

We show the result of five real scenes in Fig. 11. Our E^{2}NeRF effectively utilizes the internal relationship between events and blurry images to learn a sharp NeRF. As a result, our results are not affected by the noise of event data and achieve an impressive image deblurring effect.

In Fig. 12, Fig. 13, Fig. 14, Fig. 15 and Fig. 16, we provide a detailed comparison of "camera", "toys", "letter", "lego" and "plant" scenes. The Results clearly demonstrate that our method yields the best and most stable deblurring results. Both EDI and EDI-NeRF have limitations when exposed to event data noise, introducing more noise into their output. Furthermore, the EDI algorithm struggles with color images, resulting in color deviation along the edges of toys, tables, and other objects. Deblur-NeRF performs poorly in the case of very serious blurring. The effect of state-of-the-art image-based deblur method MPR and event-image-based method D2net are also limited. Affected by this, MPR-NeRF and D2net-NeRF also perform poorly.

Fig. 17 and Fig. 18 show the detailed comparison of blur view and novel view on synthetic data. Like the results of real-world data, our method shows the best deblurring results. EDI and EDI-NeRF are still limited by the noise on the edge of the object and the color deviation. Other methods cannot produce accurate deblurring results.

4. Supplementary video of additional results

We provide a video at https://icvteam.github.io/E2NeRF.html. For synthetic scenes, we only show the results of NeRF, MPR-NeRF, D2net-NeRF, EDI-NeRF and our E^{2}NeRF, because Deblur-NeRF cannot learn a 360° 3D representation on our synthetic dataset. For the real scenes, we show the comparison of the results of all mentioned methods. It is obvious that the results of our E^{2}NeRF have less cloudy material, noise and sharper texture details compared to other methods on synthetic and real scenes.

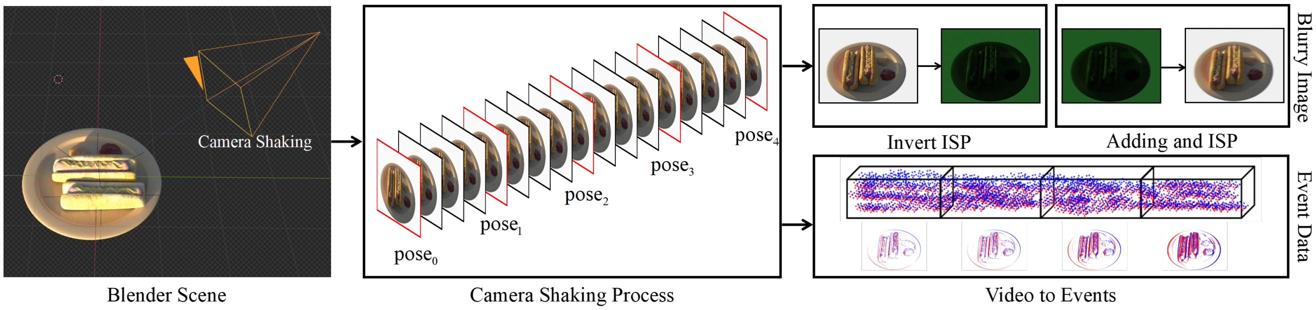

Figure 10: Process of generating synthetic data.

Blur View	Chair			Ficus			Hotdog			Lego			Mic			Materials		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	24.29	.9357	.1254	22.98	.9023	.1037	27.75	.9546	.1158	21.95	.8548	.2103	19.99	.9108	.1512	20.50	.8854	.1579
Deblur-NeRF	25.87	.9373	.2185	22.86	.8982	.1541	24.62	.9396	.2138	24.47	.8756	.2053	20.54	.9012	.2562	11.92	.7249	.3706
D2Net	29.14	.9632	.0811	27.20	.9441	.0591	32.57	.9753	.0797	26.70	.9281	.1170	25.12	.9497	.0897	26.15	.9495	.0937
D2Net-NeRF	28.92	.9606	.0900	26.77	.9377	.0740	32.42	.9733	.0904	26.51	.9165	.1364	24.75	.9437	.1159	25.37	.9381	.1104
EDI	29.31	.9585	.0760	27.55	.9455	.0888	33.83	.9729	.0700	26.68	.9217	**.0813**	24.48	.9391	.0928	25.42	.9327	.1068
EDI-NeRF	29.53	.9642	.0713	27.65	.9503	.0504	33.52	.9760	.0728	26.80	.9250	.0823	24.67	.9451	**.0829**	25.49	.9375	.0880
MPR	29.23	.9625	.0871	29.64	**.9665**	.0552	31.89	.9705	.0897	27.92	**.9444**	.0997	24.41	.9437	.0965	25.62	.9410	.0906
MPR-NeRF	29.24	.9644	.0818	28.97	.9599	.0516	31.70	.9715	.0940	27.88	.9373	.1123	24.34	.9433	.0990	25.42	.9383	.0905
E^{2}NeRF^{25}	**31.45**	.9735	.0667	29.14	.9596	.0492	32.98	.9748	.0845	27.16	.9211	.1357	26.90	.9485	.1100	26.77	.9435	.0859
E^{2}NeRF*	30.67	.9701	.0780	29.58	.9628	.0433	**34.76**	**.9804**	**.0645**	27.56	.9272	.1232	26.81	.9537	.0985	26.91	.9458	.0796
E^{2}NeRF	31.28	**.9749**	**.0608**	**30.00**	.9663	**.0362**	34.34	.9784	.0660	**28.11**	.9339	.1078	**27.27**	**.9570**	.0919	**27.60**	**.9496**	**.0724**

Table 7: Detailed quantitative results on blur view. The average results of the six synthetic scenes are shown in Tab. 1 in the main manuscript. We use **bold** and underline to mark the best and second best data. E^{2}NeRF^{25} represents training E^{2}NeRF with only 25 blurry images as in Deblur-NeRF. E^{2}NeRF* denotes training E^{2}NeRF without event loss.

Novel View	Chair			Ficus			Hotdog			Lego			Mic			Materials		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	23.75	.9319	.1291	22.32	.8959	.1088	26.84	.9504	.1201	21.31	.8484	.2149	19.39	.9054	.1541	20.00	.8788	.1628
Deblur-NeRF	22.80	.9162	.2435	20.84	.8751	.1761	24.45	.9372	.2213	21.71	.8311	.2275	17.80	.8649	.3028	12.01	.7259	.3724
D2Net-NeRF	29.04	.9619	.0900	26.62	.9382	.0773	26.84	.9504	.1201	26.46	.9196	.1372	25.15	.9455	.1162	25.79	.9406	.1116
EDI-NeRF	30.63	.9704	.0715	27.80	.9568	.0970	27.87	.9676	.0982	28.19	**.9444**	**.0808**	26.36	.9563	**.0807**	26.62	.9473	.0881
MPR-NeRF	29.06	.9644	.0825	28.19	.9560	.0553	31.50	.9725	.0955	27.30	.9353	.1136	24.79	.9462	.0989	25.40	.9386	.0919
E^{2}NeRF^{25}	**31.73**	.9765	.0679	27.91	.9560	.0534	33.25	.9772	.0848	27.74	.9385	.1340	26.80	.9492	.1113	27.40	.9513	.0857
E^{2}NeRF*	29.20	.9661	.0796	27.77	.9543	.0483	33.03	.9782	.0667	26.66	.9231	.1246	24.90	.9462	.1065	26.32	.9436	.0831
E^{2}NeRF	31.30	**.9769**	**.0613**	**29.02**	**.9649**	**.0389**	**33.67**	**.9794**	**.0662**	**28.20**	.9424	.1039	**27.06**	**.9569**	.0931	**28.13**	**.9556**	**.0721**

Table 8: Detailed quantitative results on novel view. The average results of the six synthetic scenes are shown in Tab. 2 in the main manuscript. We use **bold** and underline to mark the best and second best data. E^{2}NeRF^{25} represents training E^{2}NeRF with only 25 blurry images as in Deblur-NeRF. E^{2}NeRF* denotes training E^{2}NeRF without event loss.

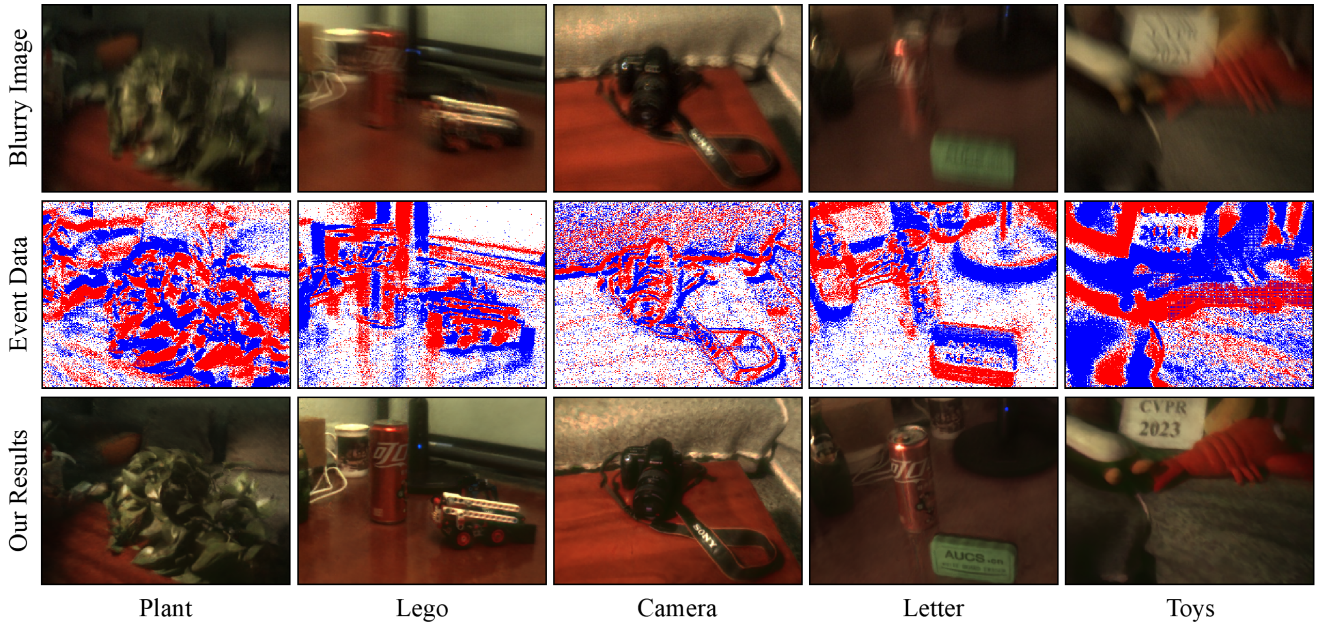

Figure 11: Five scenes of real-world data. Our E^{2}NeRF effectively utilizes the internal relationship between events and blurry images to learn a sharp NeRF. The results are not affected by the noise of event data and achieve an impressive image deblurring effect.

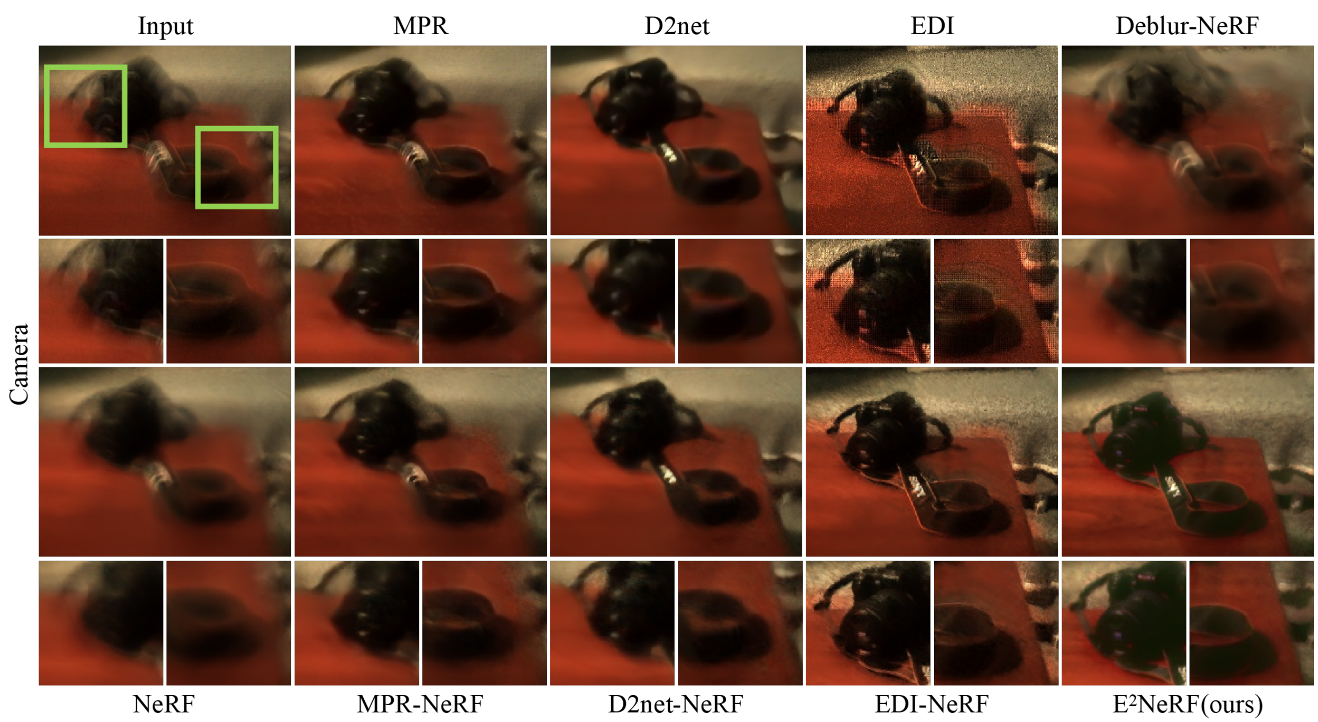

Figure 12: Detailed qualitative comparison for “camera” scene of real-world data.

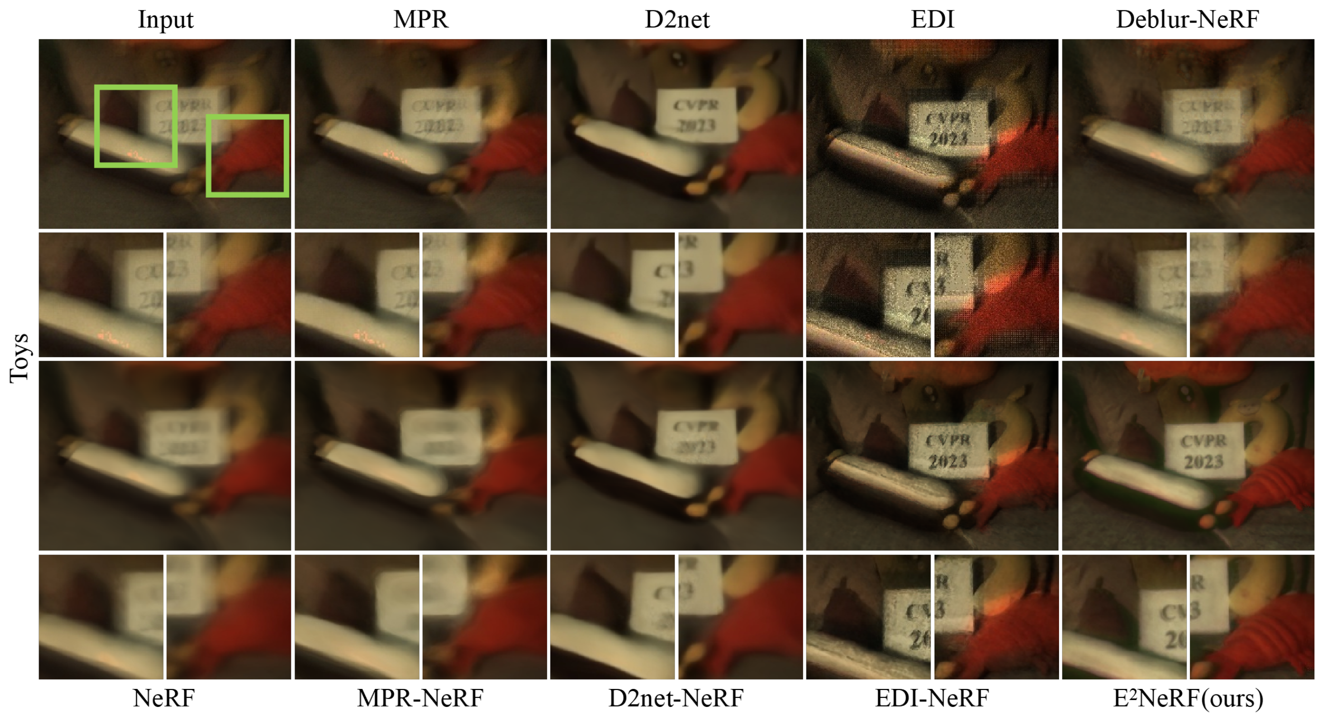

Figure 13: Detailed qualitative comparison for “toys” scene of real-world data.

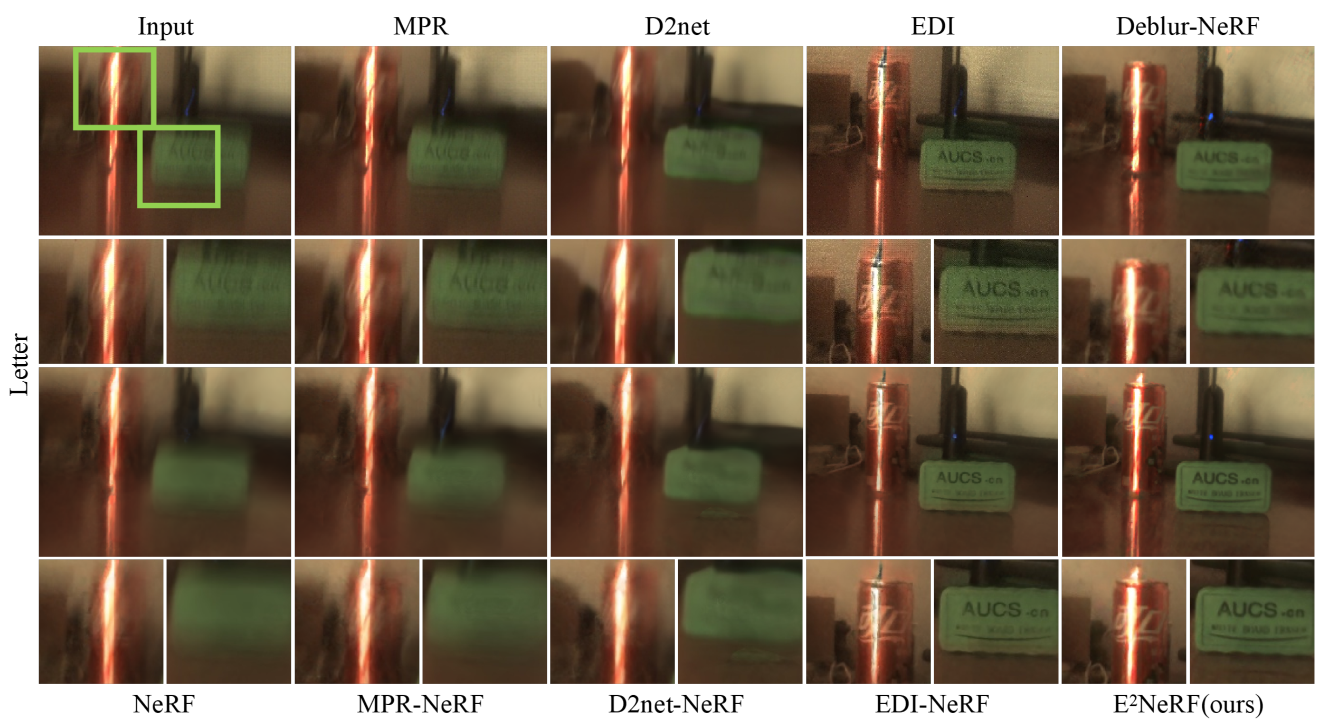

Figure 14: Detailed qualitative comparison for “letter” scene of real-world data.

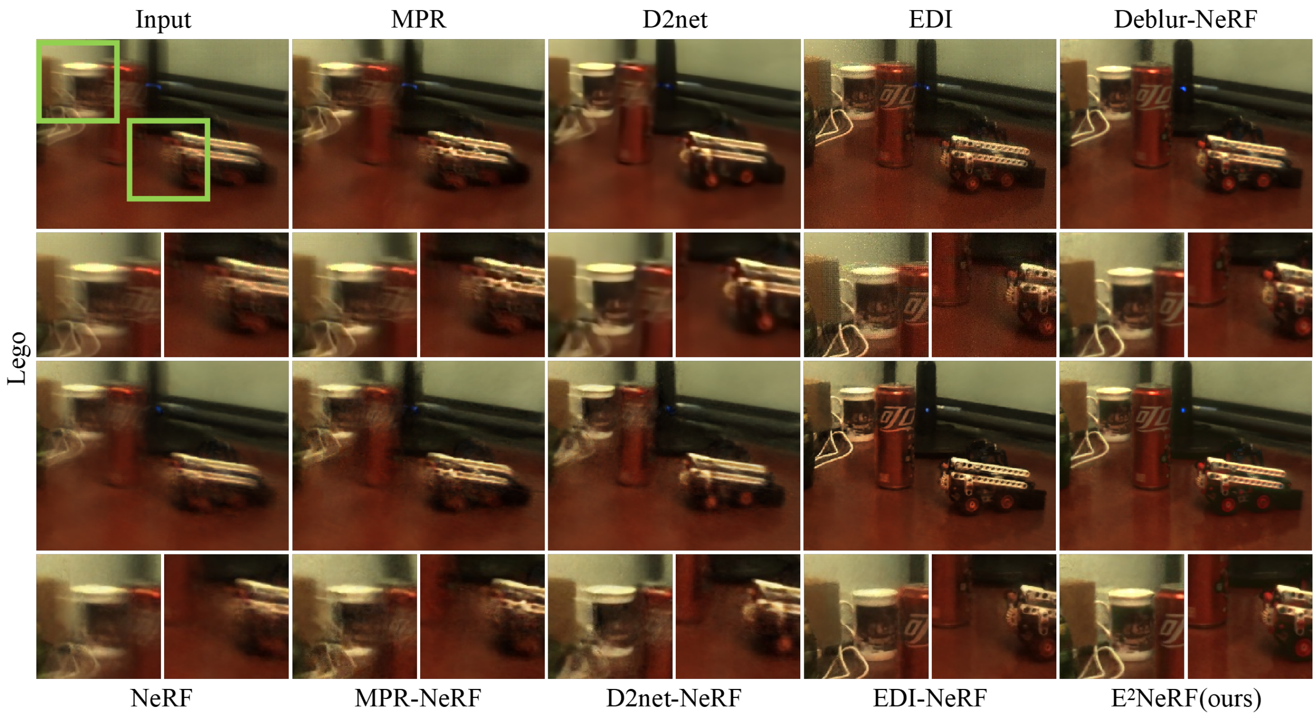

Figure 15: Detailed qualitative comparison for “lego” scene of real-world data.

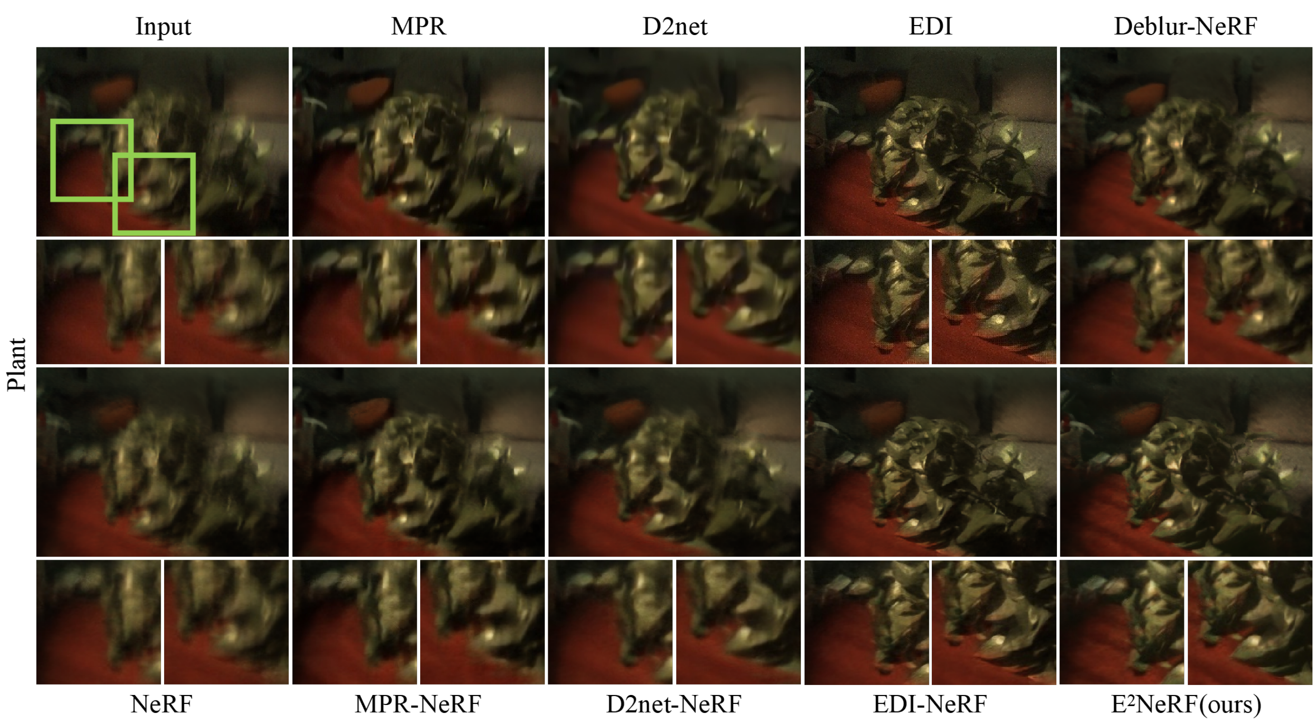

Figure 16: Detailed qualitative comparison for “plant” scene of real-world data.

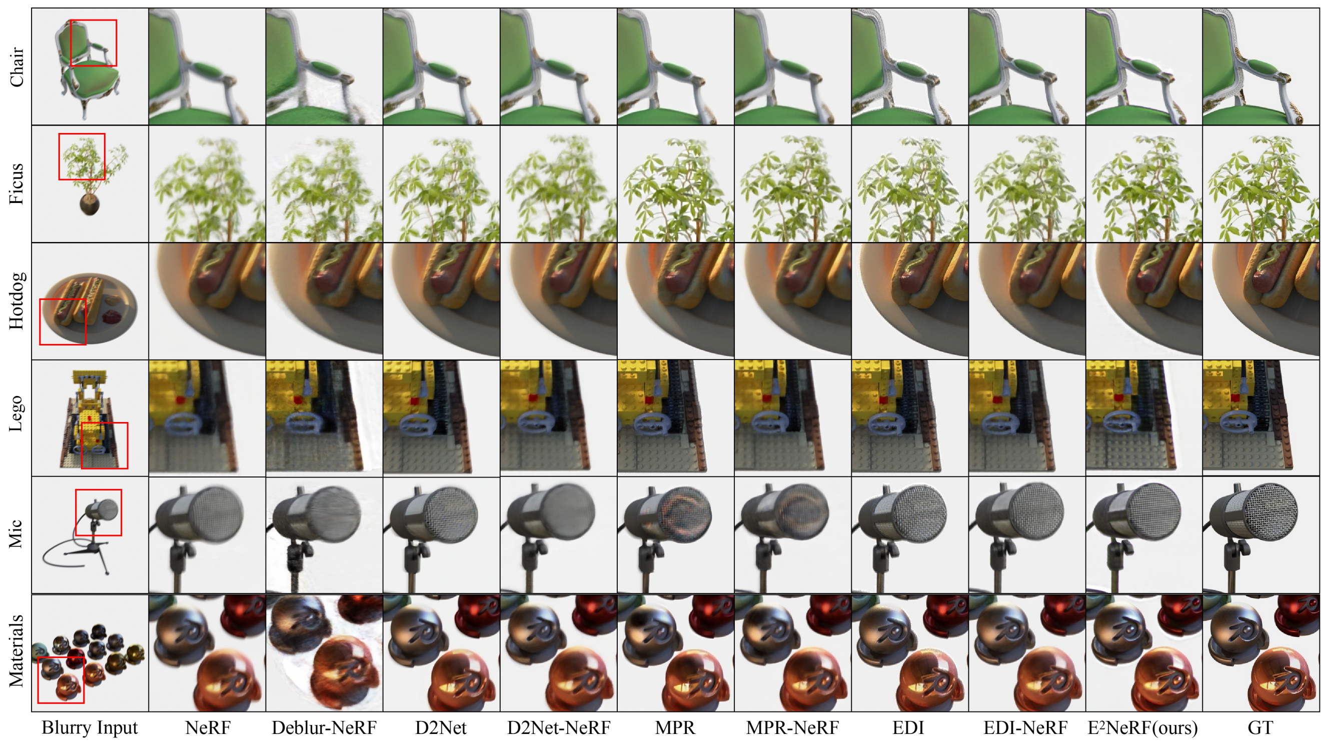

Figure 17: Detailed qualitative comparison for blur view of synthetic data.

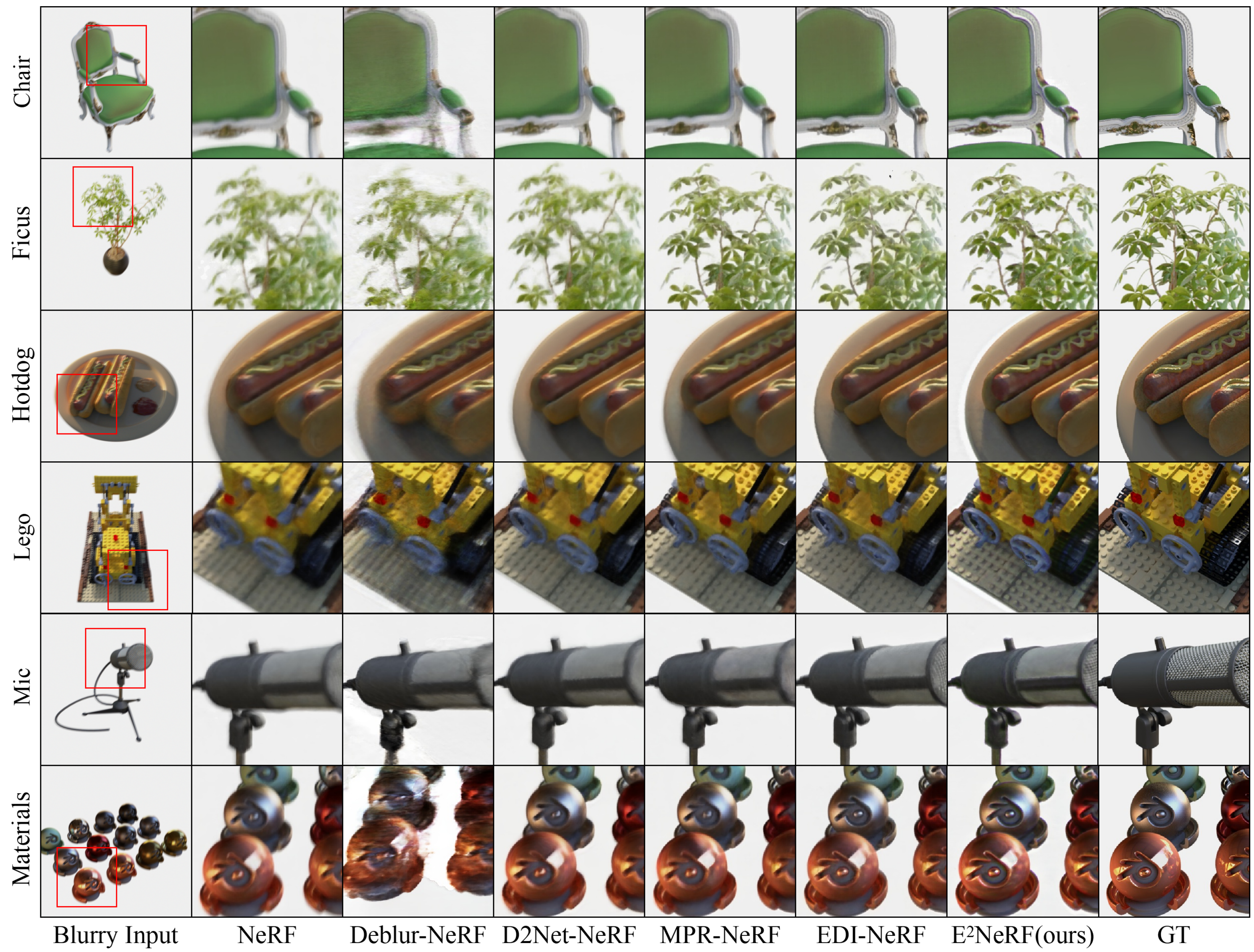

Figure 18: Detailed qualitative comparison for novel view of synthetic data.