DiET-GS: Diffusion Prior and Event Stream-Assisted Motion Deblurring 3D Gaussian Splatting

Supplementary Material

In this supplementary material, we provide more implementation details, introduce additional evaluations on various tasks, and conduct further ablation studies with more qualitative and quantitative analysis.

- More implementation details are provided in Sec. A.
- We conduct the user study in Sec. B to further evaluate the visual quality of our method.
- Additional evaluations on single image deblurring task is introduced in Sec. C.
- Further ablation studies for our DiET-GS++ are presented in Sec. D to validate our design choice.
- More qualitative and quantitative results on novel-view synthesis are provided in Sec. E.

A. Implementation Details

Training. During training, we follow the configuration of the original 3DGS. The learnable camera response function $\mathrm{CRF}(\cdot)$ is introduced after a 1,500-iteration warm-up. Similarly, the regularization term $\mathcal{L}_{\text{edi\_simul}}$ is employed after a 7,000-iteration warm-up, since DiET-GS should be able to simulate the blurry images properly. Following [1], we leverage the color events in $\mathcal{L}_{\text{ev}}$ during training on real-world scenes, where the color events record color intensity changes following a Bayer pattern [6]. In this case, the luma conversion $h(\cdot)$ is dropped from Eq. 6 and $\mathcal{L}_{\text{ev}}$ is directly applied to the color channel responsible for triggering events. Furthermore, since green pixels appear twice as often in an RGBG Bayer pattern, we weigh the events' contributions by 0.4, 0.2, and 0.4 for each of the RGB channels.

Leveraging Diffusion Prior. We use Stable Diffusion $\times 4$ Upscaler (SD$\times$4) [11] as a pretrained diffusion model to provide diffusion prior. SD$\times$4 is originally designed to upscale the image while recovering high-resolution details, with the low-resolution image as a conditional input to the diffusion UNet. However, we find that SD$\times$4 is also effective at enhancing edge details at the same resolution. During the RSD optimization, we sample uniform random crops of 128$\times$128 resolution in latent space for fast optimization speed, following [5]. A constant learning rate of $1e{-}2$ is employed for all learnable parameters $\mathbf{f_g}$ in Stage 2.

Color Correction. As also noted in [3, 14], we empirically find that leveraging diffusion prior alone in Stage 2 can exhibit color shifts. To address this issue, we adopt wavelet-based color correction proposed in [14] as a post-processing step. Specifically, let us denote the two images $\hat{C}$ and $\tilde{C}$ rendered from DiET-GS and DiET-GS++ respectively as follows:

$$\tilde{C} = \mathcal{D}(\mathbf{f}_{2\text{D}} + \mathcal{E}(\hat{C})), \quad \hat{C} = g(p), \tag{8}$$

where $g(\cdot)$ is the pretrained 3D Gaussians from DiET-GS with a rendering function and p is the given camera pose. We assume that $\hat{C}$ is capable of preserving the accurate color due to the color guidance from ground-truth blurry images and EDI prior in Stage 1. In contrast, $\tilde{C}$ from DiET-GS++ tends to show a color shift since it solely relies on diffusion prior while the edge details are effectively enhanced. To combine the accurate color information from $\hat{C}$ and sharp edge details of $\tilde{C}$, we first decompose both images into high-frequency and low-frequency components via the wavelet decomposition. Considering that color information belongs to the low-frequency components while fine-grained details are mostly high-frequency components, we simply incorporate the low-frequency parts of $\hat{C}$ and high-frequency parts of $\tilde{C}$ to obtain the final output. More details about wavelet-based color correction can be found in [14].

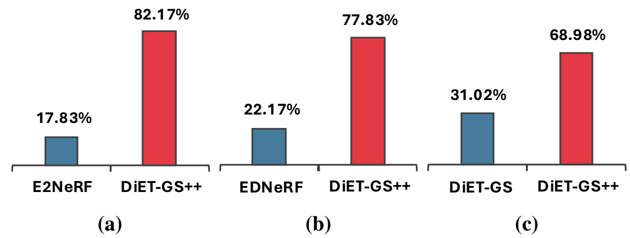


Figure 7. User study. DiET-GS++ is compared to E2NeRF, Ev-DeblurNeRF (denoted as EDNeRF) and DiET-GS by 60 evaluators for each pair. DiET-GS++ gains significantly higher votes against the baselines, showing at least 37.96% difference.

B. User Study

To evaluate the visual quality in terms of human perception, we conduct a user study with 60 evaluators. Specifically, we collect 30 pairs of samples from the test views of both synthetic and real-world datasets, where each pair consists of two images rendered from identical poses using different methods. During the user study, evaluators were asked to select the image with better quality between the two presented options for every pair.

Baselines. We compare our DiET-GS++ to event-based deblurring rendering methods, including E2NeRF [10] and Ev-DeblurNeRF [1]. Furthermore, DiET-GS++ is also

Methods	*Batteries*		*Powersupplies*		*Labequipment*		*Figures*		*Drones*		***Average***	
	MUSIQ↑	CLIP-IQA↑	MUSIQ↑	CLIP-IQA↑	MUSIQ↑	CLIP-IQA↑	MUSIQ↑	CLIP-IQA↑	MUSIQ↑	CLIP-IQA↑	MUSIQ↑	CLIP-IQA↑
MPRNet [16]	23.50	0.1074	34.01	0.1284	23.34	0.1082	24.10	0.1326	30.97	0.1034	27.18	0.1160
NAFNet [2]	34.43	0.2363	49.25	0.2045	35.47	0.2023	28.87	0.2264	41.64	0.1543	37.93	0.2047
Restormer [17]	28.71	0.1210	42.32	0.1154	32.13	0.1252	30.99	0.1631	36.37	0.0790	34.10	0.1207
EDI [9]	38.80	0.2013	49.63	0.2260	33.53	0.1338	41.77	0.2675	41.75	0.1991	41.10	0.2055
EFNet [12]	35.32	0.1503	45.23	0.1934	30.13	0.1435	38.45	0.2234	39.12	0.1762	37.65	0.1773
BeNeRF [7]	47.31	0.1704	56.06	0.2445	44.25	0.1789	46.97	0.2653	48.71	0.2054	48.66	0.2129
DiET-GS++	**51.23**	**0.2654**	**56.32**	**0.2598**	**45.14**	**0.2034**	**52.43**	**0.3012**	**50.34**	**0.2078**	**51.09**	**0.2475**

Table 3. Quantitative comparisons on single image deblurring with real-world datasets.

Methods	MUSIQ↑	Training time (hr)			Rendering time (s)
		Stage 1	Stage 2	Total	
E2NeRF [10]	39.47	24.3	-	24.3	2.4139
Ev-DeblurNeRF [1]	39.70	3.4	-	3.4	0.8861
DiET-GS	45.31	9.8	-	9.8	**0.0014**
DiET-GS++	**51.71**	9.8	0.17	10.0	1.8703
DiET-GS++-*light*	50.23	1.1	0.17	**1.3**	1.8703

Table 4. Comparison on training time and rendering time.

compared with DiET-GS trained from Stage 1 to demonstrate the efficacy of leveraging diffusion prior in Stage 2.

Results. As shown in Fig. 7, our DiET-GS++ gains at least 68.98% of the votes in each comparison, further validating the effectiveness of our framework. It also shows the clear gap of 37.96% over DiET-GS (*cf*. Fig. 7c), highlighting the efficacy of enhancing the edge details with diffusion prior in Stage 2.

C. Single Image Deblurring

We also conduct experiments on the single image deblurring task using the real-world Ev-DeblurCDAVIS dataset [1]. For evaluation, we randomly select 5 blurry images per scene and compare our DiET-GS++ against various single image deblurring baselines on these sampled images.

Baselines. We classify the baselines into three categories. The first category is frame-based single image deblurring methods that rely solely on RGB frames to recover a clean image. MPRNet [16], NAFNet [2], and Restormer [17] are selected for this category. The second category is event-enhanced deblurring methods that utilize additional event data to improve visual quality, consisting of EDI [9] and EFNet [12]. The third category combines NeRF and events to tackle single image deblurring, where BeNeRF [7] is chosen for this category. BeNeRF reconstructs the 3D scenes by learning the camera trajectory from a single blurry image and corresponding event stream to deblur the given single view. Once we have trained BeNeRF, the deblurred image is produced by rendering the mid-exposure pose of the image along the estimated camera trajectory.

Evaluation metrics. Since real-world dataset lacks the ground-truth images for the mid-exposure poses of blurry views, we employ the No Reference Image Quality Assessment (NR-IQA) metrics: MUSIQ [4] and CLIP-IQA [13] for the evaluation.

Results. We present the quantitative comparisons in Tab. 3. Our DiET-GS++ consistently outperforms all baselines in every 5 real-world scenes. Specifically, compared to BeNeRF, performance is improved by an average of 2.43 and 0.0346 in MUSIQ and CLIP-IQA scores, respectively. Furthermore, we also present qualitative comparisons in Fig. 10. As shown in 2nd column, frame-based image deblurring method NAFNet often produces inaccurate details since it solely relies on blurry images to recover fine-grained details. EDI and BeNeRF recover more precise details, benefiting from the event-based cameras while severe artifacts are still exhibited. Our DiET-GS++ shows the best visual quality with cleaner and well-defined details by leveraging EDI and pretrained diffusion model as prior.

D. Ablation Study

We present additional ablation studies to thoroughly investigate each component of DiET-GS++. All the experiments are conducted on a real-world scene, namely, *Figures* sample.

D.1. Training and Rendering Efficiency.

We compare the optimization and rendering efficiency of our method to event-enhanced rendering methods, including E2NeRF [10] and Ev-DeblurNeRF [1] in Tab. 4. We present the training time of Stage 1 and Stage 2 separately, while the training time of Stage 2 remains blank if the corresponding method employs single-stage training. We observe from Tab. 4: 1) DiET-GS and DiET-GS++ require longer training time compared to Ev-DeblurNeRF. We find that RSD optimization in Stage 1 is the main factor of prolonged training time, since the gradient from the RSD loss flows to the 3D Gaussians through the pretrained VAE encoder, which introduces significant computational overhead. We thus propose the light variant of our DiET-GS++ in the 5th row by simply excluding the RSD loss in Stage 1, which we refer to as DiET-GS++-*light*. Despite a slight performance drop in MUSIQ scores, our variant DiET-GS++-*light* shows the fastest optimization speed with a $\times 2.6$ speedup in convergence compared to Ev-DeblurNeRF. 2) Training time for Stage 2 in DiET-GS++ only requires 0.17 hours, while showing a significant improvement in MUSIQ scores compared to DiET-GS. In contrast to RSD optimization in Stage 1, the learnable latent residual is directly ren-

$\mathrm{CRF}(\cdot)$	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	CLIP-IQA↑
✘	32.93	0.8703	0.1123	38.93	0.2000
✓	**34.89**	**0.9049**	**0.0600**	**45.31**	**0.2471**

Table 5. Ablation on camera response function $\mathrm{CRF}(\cdot)$.

(a) wo $\mathrm{CRF}(\cdot)$ **(b)** w $\mathrm{CRF}(\cdot)$

Figure 8. Qualitative analysis on camera response function.

(a) wo color correction **(b)** w color correction **(c)** GT

Figure 9. Ablation on wavelet-based color correction.

dered from DiET-GS without exploiting the VAE encoder, which thus leads to faster gradient computation. 3) DiET-GS enables real-time rendering, benefiting from the explicit representations of 3DGS. However, our DiET-GS++ exhibits longer rendering time compared to Ev-DeblurNeRF, since the rendered image is further refined through the VAE encoder and decoder. Nonetheless, leveraging diffusion prior is reasonable given the performance improvement of 12.0 MUSIQ scores compared to Ev-DeblurNeRF.

D.2. Camera Response Function

Tab. 5 shows the effectiveness of leveraging the learnable camera response function $\mathrm{CRF}(\cdot)$, showing the performance improvement in all 5 metrics. Furthermore, Fig. 8 demonstrates that the $\mathrm{CRF}(\cdot)$ function is capable of restoring more well-defined details. As noted in [1], employing the learnable camera response function naturally fills the gap between the RGB space and the brightness change perceived by the event camera, thus effectively restoring the intricate details.

Discussion. Although we adopt the similar strategy with [1] by leveraging learnable camera response function, we differ from [1] as follows: We further combine the $\mathrm{CRF}(\cdot)$ function into the EDI formulation, proposing the novel EDI constraint for enhancing fine-grained details. As shown in the Fig. 5b, EDI color guidance $\mathcal{L}_{\mathrm{edi\_color}}$ proposed by [1] often yields over-smoothed details since it treats each RGB channel as brightness which deviates from the real-world setting. To compensate the well-defined details, we propose $\mathcal{L}_{\mathrm{edi\_gray}}$ by modeling the EDI in the brightness domain with exploiting learnable $\mathrm{CRF}(\cdot)$ function. Using the $\mathcal{L}_{\mathrm{edi\_color}}$ and $\mathcal{L}_{\mathrm{edi\_gray}}$ together enables the mutual compensation between accurate color and well-defined details, achieving the best visual quality as shown in the Fig. 5c.

Conditional input	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	CLIP-IQA↑
EDI-processed image	34.25	0.8964	0.0754	41.47	0.2233
ground-truth blurry image	**34.89**	**0.9049**	**0.0600**	**45.31**	**0.2471**

Table 6. Ablation on conditional input of RSD optiimzation in Stage 1.

D.3. Conditional Input of Diffusion UNet

In Tab. 6, we explore the various options for conditional input of the diffusion UNet during the RSD optimization in Stage 1. The 1st row of Tab. 6 exploits the EDI-processed image as conditional input, while the sharp image rendered from 3DGS is given as the input to the diffusion process. However, this choice leads to inferior performance compared to leveraging the ground-truth blurry image as conditional input (2nd row). We postulate that this is because the unnatural artifacts introduced by EDI are often detrimental to the noise inference of the diffusion UNet. Despite the motion blur in the image, the ground-truth blurry image provides more natural guidance to noise prediction, such as accurate color prior, since it is real-captured from the frame-based camera.

D.4. Wavelet-based Color Correction

Fig. 9 presents the effectiveness of wavelet-based color correction. It effectively mitigates the color shift introduced from diffusion prior, achieving better color.

E. More Results on Novel-View Synthesis

Quantitative Results. In Tab. 7 and Tab. 8, we present the additional quantitative results on novel-view synthesis for each scene in both real-world and synthetic datasets. In most cases, our DiET-GS and DiET-GS++ significantly outperform existing baselines across all five evaluation metrics, showing the effectiveness of our framework.

Qualitative Results. In Fig. 11 and Fig. 12, we present more qualitative comparisons on novel-view synthesis in both real-world and synthetic datasets. Our DiET-GS++ is capable of restoring: 1) *accurate color*, 2) *fine-grained details* and 3) *clean texture*, thus achieving the best visual quality compared to the existing baselines.

F. Limitation

Following previous works [1, 10], we structure DiET-GS assuming uniform-speed camera motion and dense, low-noise events. While real-world scenarios may not always meet these ideal conditions, advanced techniques like [8] could extend our method's applicability.

Scene	Metric	MRPNet+GS [16]	EDI+GS [9]	EFNet+GS [12]	BAD-NeRF [15]	BAD-GS [18]	E2NeRF [10]	Ev-DeblurNeRF [1]	DiET-GS (Ours)	DiET-GS++ (Ours)
batteries	PSNR↑	28.42	33.11	31.30	28.29	28.73	31.49	32.63	**34.52**	33.51
	SSIM↑	0.7518	0.8994	0.8556	0.8086	0.8217	0.8715	0.8938	**0.9304**	0.9118
	LPIPS↓	0.1948	0.0613	0.0804	0.2245	0.1651	0.0932	0.0443	**0.0435**	0.0444
	MUSIQ↑	22.13	37.90	35.51	17.71	20.20	37.48	42.99	45.66	**49.89**
	CLIP-IQA↑	0.2338	0.2182	0.2293	0.1887	0.1918	0.2445	0.2292	0.2327	**0.2603**
figures	PSNR↑	28.18	33.51	31.28	29.31	30.12	32.59	32.82	**34.89**	33.86
	SSIM↑	0.7311	0.8723	0.8317	0.7703	0.7767	0.8543	0.8577	**0.9049**	0.8846
	LPIPS↓	0.2146	0.0977	0.1324	0.2935	0.2438	0.1108	0.0687	**0.0600**	0.0634
	MUSIQ↑	23.45	38.48	37.13	19.50	22.13	39.47	39.70	45.37	**51.71**
	CLIP-IQA↑	0.2418	0.2384	0.2218	0.1836	0.1898	0.2624	0.2441	0.2584	**0.2955**
drones	PSNR↑	27.13	33.02	31.18	28.51	29.19	31.03	31.62	**34.08**	32.92
	SSIM↑	0.7634	0.9025	0.8617	0.8123	0.8317	0.8780	0.8866	**0.9339**	0.9152
	LPIPS↓	0.2012	0.0832	0.1293	0.2122	0.1687	0.1075	0.0538	**0.0387**	0.0396
	MUSIQ↑	28.38	42.35	41.18	19.05	22.20	39.00	41.81	47.58	**50.17**
	CLIP-IQA↑	0.1718	0.1633	0.1526	0.1723	0.1743	0.1877	0.1773	0.1778	**0.2028**
powersupplies	PSNR↑	26.37	32.10	30.92	27.35	28.38	31.06	32.05	**33.54**	32.37
	SSIM↑	0.7513	0.8955	0.8516	0.7953	0.8071	0.8820	0.8980	**0.9271**	0.9108
	LPIPS↓	0.1824	0.0657	0.1029	0.2756	0.2247	0.0826	0.0492	0.0460	**0.0459**
	MUSIQ↑	31.48	46.04	44.15	24.68	24.90	45.17	47.97	50.25	**55.83**
	CLIP-IQA↑	0.2477	0.2307	0.2219	0.1762	0.1701	0.2373	0.2501	0.2078	**0.2531**
labequipment	PSNR↑	27.47	33.00	30.18	28.89	29.19	31.51	32.36	**34.06**	33.13
	SSIM↑	0.7598	0.8911	0.8512	0.8042	0.8276	0.8578	0.8772	**0.9150**	0.8971
	LPIPS↓	0.2138	0.0871	0.1262	0.2563	0.2037	0.1355	0.0696	0.0599	**0.0575**
	MUSIQ↑	20.18	35.54	33.18	18.84	21.19	32.95	34.14	40.21	**44.60**
	CLIP-IQA↑	0.1722	0.1534	0.1418	0.1749	0.1804	0.1854	**0.2048**	0.1708	0.1958

Table 7. Quantitative comparisons on novel-view synthesis in 5 real-world scenes

Scene	Metric	MRPNet+GS [16]	EDI+GS [9]	EFNet+GS [12]	BAD-NeRF [15]	BAD-GS [18]	E2NeRF [10]	Ev-DeblurNeRF [1]	DiET-GS (Ours)	DiET-GS++ (Ours)
factory	PSNR↑	17.44	22.46	19.74	18.81	21.35	22.28	23.33	**26.54**	26.00
	SSIM↑	0.5918	0.7629	0.6415	0.6038	0.6709	0.7822	0.8189	**0.8856**	0.8707
	LPIPS↓	0.3817	0.1448	0.3319	0.2822	0.2391	0.1838	0.1858	**0.0898**	0.0962
	MUSIQ↑	26.18	56.74	37.12	27.43	36.19	45.88	41.58	54.24	**57.62**
	CLIP-IQA↑	0.2211	0.2177	0.2118	0.1668	0.1718	0.2014	0.1947	0.2215	**0.2270**
pool	PSNR↑	19.49	24.83	21.79	25.58	26.18	**27.63**	27.26	27.40	26.5880
	SSIM↑	0.4718	0.6496	0.5238	0.6888	0.7418	0.7488	0.7440	**0.7512**	0.7283
	LPIPS↓	0.3219	0.1897	0.3718	0.2601	0.2118	0.1995	0.2230	0.1895	**0.1827**
	MUSIQ↑	15.19	47.12	26.14	30.81	39.14	47.68	44.63	51.01	**53.03**
	CLIP-IQA↑	0.1729	0.2106	0.2037	0.1860	0.1911	0.2473	**0.2635**	0.2126	0.2384
tanabata	PSNR↑	18.54	23.02	20.80	16.91	20.18	23.43	23.74	**26.18**	25.90
	SSIM↑	0.6203	0.8088	0.6817	0.6483	0.7661	0.8156	0.8059	**0.8965**	0.8896
	LPIPS↓	0.3645	0.1232	0.3128	0.2175	0.1608	0.1505	0.1727	0.0754	**0.0712**
	MUSIQ↑	28.71	59.13	39.28	17.56	27.19	47.81	41.53	63.94	**65.54**
	CLIP-IQA↑	0.2818	0.3288	0.2518	0.2018	0.2167	0.1914	0.2572	0.3115	**0.3504**
trolley	PSNR↑	19.58	24.43	21.79	17.81	21.20	24.83	24.70	**26.67**	26.43
	SSIM↑	0.6811	0.8563	0.7182	0.6114	0.7064	0.8505	0.8465	**0.9094**	0.9026
	LPIPS↓	0.3499	0.0923	0.2691	0.2362	0.1931	0.1157	0.1335	**0.0708**	**0.0708**
	MUSIQ↑	26.39	57.56	37.98	18.71	27.19	47.87	41.80	61.48	**63.43**
	CLIP-IQA↑	0.2894	0.3434	0.2583	0.2007	0.2176	0.2113	0.2047	0.3618	**0.3683**

Table 8. Quantitative comparisons on novel-view synthesis in 4 synthetic scenes

References

[1] Marco Cannici and Davide Scaramuzza. Mitigating motion blur in neural radiance fields with events and frames. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9286–9296, 2024. 1, 2, 3, 4, 6, 7

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 2, 5

[3] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11472–11481, 2022. 1

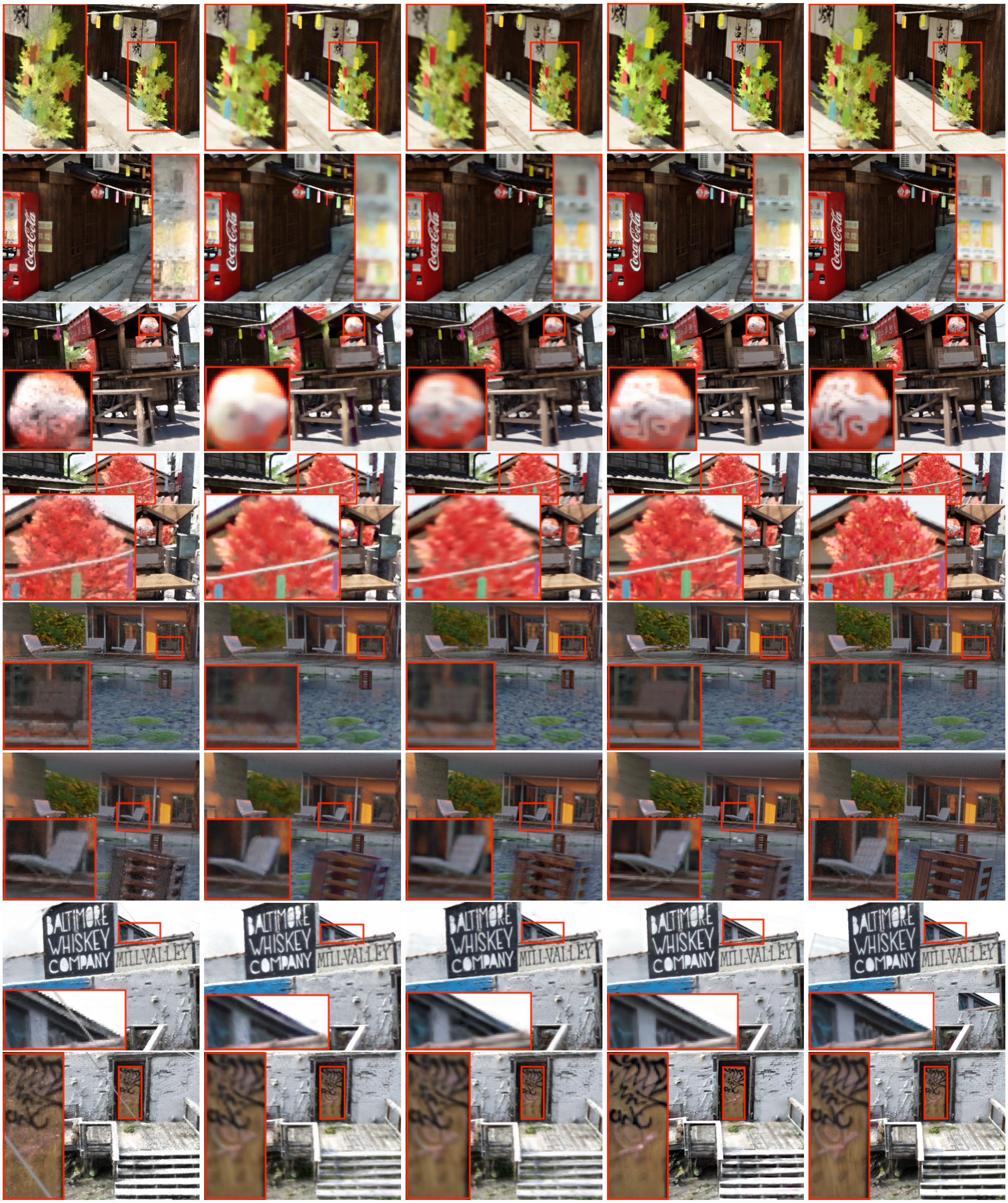
[4] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 2

(a) Blur Image **(b)** NAFNet [2] **(c)** EDI [9] **(d)** BeNeRF [7] **(e)** DiET-GS++ (Ours)

Figure 10. Qualitative comparisons on single image deblurring with real-world datasets.

(a) EDI+GS [9] **(b)** E2NeRF [10] **(c)** Ev-DeblurNeRF [1] **(d)** DiET-GS++ (Ours) **(e)** GT

Figure 11. More qualitative comparisons on novel-view synthesis in real-world datasets.



(a) EDI+GS [9] **(b)** E2NeRF [10] **(c)** Ev-DeblurNeRF [1] **(d)** DiET-GS++ (Ours) **(e)** GT

Figure 12. More qualitative comparisons on novel-view synthesis in synthetic datasets.

[5] Jie Long Lee, Chen Li, and Gim Hee Lee. Disr-nerf: Diffusion-guided view-consistent super-resolution nerf. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20561–20570, 2024. 1

[6] 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, 2015. 1

[7] Wenpu Li, Pian Wan, Peng Wang, Jinghang Li, Yi Zhou, and Peidong Liu. Benerf: neural radiance fields from a single blurry image and event stream. In *European Conference on Computer Vision*, pages 416–434. Springer, 2025. 2, 5

[8] 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. 3

[9] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6820–6829, 2019. 2, 4, 5, 6, 7

[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. 1, 2, 3, 4, 6, 7

[11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1

[12] 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 *European conference on computer vision*, pages 412–428. Springer, 2022. 2, 4

[13] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 2

[14] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. 2024. 1

[15] 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. 4

[16] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 2, 4

[17] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2

[18] Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, pages 233–250. Springer, 2025. 4