Towards Real-world Event-guided Low-light Video Enhancement and Deblurring – Supplementary Materials –

Taewoo Kim, Jaeseok Jeong, Hoonhee Cho, Yuhwan Jeong, and Kuk-Jin Yoon

Korea Advanced Institute of Science and Technology
{intelpro, jason.jeong, gnsgnsgml, jeongyh98, kjyoon}@kaist.ac.kr

Abstract. Due to the lack of space in the main paper, we provide more details in the supplementary materials. In Sec. A, we cover additional experimental results not addressed in the main paper. In Sec. B, we provide more detailed information about the RELED dataset. In Sec. C, we provide more visual results and video demos.

A Additional experimental results

A.1 implementation details

We implement our framework using PyTorch [8]. For training, we utilized the Adam [7] optimizer with momentum term of (0.9, 0,999) to optimize networks with initial learning rate 1×10^{-4}. We utilized the charbonnier [3] loss function to supervise the multi-scale outputs. We use batch size of 8 with four RTX 3090 GPUs for training. We empirically set multi-scale weight $\lambda_s = \{1, 0.5, 0.25\}$ for each scale. We set the standard deviation σ to 7 in Eq. 3 of the main paper. We set the event voxel grid size to 16 for all experiments. For data augmentation, we perform random crop at the same positions for both low-light blurred videos and event voxel data, resulting in cropped blurred patch and voxel bin of size 256×256. To train our networks, we utilized our training split of the RELED dataset and conduct training for 200 epochs. Regarding other methods, we performed retraining on the RELED dataset for the same number of epochs to conduct quantitative and qualitative comparisons. For quantitative comparison, we used widely-used metrics such as PSNR and SSIM [14].

A.2 Qualitative results on the real-world low-light blurred videos.

To assess the generalization ability of the model trained on the RELED dataset in real-world scenarios, we captured actual low-light videos, including nighttime conditions, by removing ND filters. During this process, we varied the settings such as gain, exposure time, and utilized different cameras not present in the dataset to ensure a diverse range of testing conditions. Visual comparison between UEVD [6] and ours are presented in Fig. A1. It is evident that ours consistently delivers superior detail restoration, even when confronted with real-world low-light data.

Table A1: Comparison with other temporal alignment modules.

Methods	PSNRs	SSIMs	Params(MB)
Baseline	29.59	0.902	1.8
Baseline + PCD [13]	30.05	0.907	5.3
Baseline + (SpyNet [9]+ Resblock)	29.30	0.899	5.6
Baseline + RNN-MBP [2]	30.14	0.908	5.0
Baseline + ED-TFA(Ours)	**30.78**	**0.916**	5.0

Table A2: Computational complexity and performance analysis on the RELED dataset. Runtimes are measured using an RTX-3090 GPU on 1024×768 size inputs.

Methods	MPRNet [17]	LLFormer [12]	SNRNet [16]	Retinexformer [1]	MIMOUNet+ [5]	RNN-MBP [2]
PSNR	26.89	26.62	26.47	26.66	26.52	29.52
Params(MB)	20.13	13.15	40.08	1.61	16.11	14.16
Runtime(s)	1.179	0.716	0.080	0.240	0.246	1.288
Methods	REDNet [15]	GEM [18]	UEVD [6]	EFNet [10]	REFID [11]	Ours
PSNR	29.19	26.04	29.93	29.85	30.1	31.3
Params(MB)	9.7	2.36	27.88	8.47	15.9	12.8
Runtime(s)	0.256	0.072	0.636	0.234	1.183	0.718

A.3 Qualitative ablations

In the main paper, we confirmed the extent of performance improvement for each component through an ablation study. Moreover, to delve into a qualitative analysis of the impact of each component, we have visualized the results in Fig. A2. In the figure, (a) corresponds to a low-light blurry image, (b) represents low-light events, (c) corresponds to Ver.1 (baseline) of Tab. 3 in the main paper, (d) corresponds to the Ver.2 (baseline with the ED-TFA module), and (e) corresponds to Ver.4 of Tab. 3 in the main paper, which represents the Ours-full model, respectively. As observed in the results in the Fig. A2, our model demonstrates progressive improvement in qualitative results as each module is inserted. An important point to highlight is that, consistent with the motivation behind the SFCM-FE module mentioned in the main paper, its incorporation results in a more effective restoration of the main structural information within the scene.

A.4 Comparisons on the frame-based alignment methods.

We have recorded the comparison results of various other temporal feature alignment methods in Tab. A1. For comparative analysis, we incorporated other temporal feature alignment modules (PCD [13] alignment, spynet [9], RNN-MBP [2] module) into our network baseline (Ver.1 of the ablation study Tab. 3 of the main paper) to assess the extent of performance improvement. We noticed a decline in performance with spynet [9], which can be attributed to inaccurate visual correspondence (optical flows) between adjacent video frames caused by factors such as low visibility and severe motion blur. In the case of RNN-MBP [2] and PCD [13] alignment modules, performance improvement was observed. However, relying solely on frame information without the aid of events leads to sub-optimal results. Due to the high dynamic range and high temporal resolution properties of events, our ED-TFA module can accurately estimate temporal correspondence

even in situations of low visibility and motion blur. This results in superior temporal alignment performance compared to other alignment modules.

A.5 Runtime/Params and performance comparisons.

In Tab. A2, we present a comparative analysis of parameters, runtime, and performance results for recent methods on the RELED dataset. As shown in the results, we achieve superior performance at reasonable computational costs.

A.6 Loss function

As in last line of Sec. 4.1 in the main paper(L.256-258), our decoder outputs normal-light sharp images at multiple visual scales. We utilized the charbonnier [3] loss function to supervise these multi-scale outputs.

$$\mathcal{L}_{total} = \sum_{s=0}^{2} \lambda_s \sqrt{\|S_s^t - G_s^t\|^2 + \epsilon^2} \tag{1}$$

where S_s^t represents the estimated normal-light sharp image at scale s, and G_s^t represents the ground-truth normal-light sharp image at scale s, respectively. We empirically set ϵ to 10^{-3} for all experiments.

B RELED dataset

B.1 More information about the dataset.

The RELED dataset contains 6,258 pairs of images, encompassing both low-light blurry images and normal-light sharp images, alongside event stream data that correlates with the exposure time of blurred images. To select a beam-splitter, we opt for Edmund Optics 50mm VIS, 50R/50T, Non-Polarizing Cube Beamsplitter over the plate-based alternative to mitigate beam-shifting problems. For the RGB cameras, we've chosen two FLIR BFS-U3-16S2C-CS RGB cameras, capable of recording videos at a resolution of 1440×1080, while also offering support for an external trigger interface. Alongside RGB cameras, we've also selected the EVK4 HD Prohesee Gen4.1 HD event camera, which can capture events at a resolution of 1280×720. To achieve temporal synchronization across multiple devices, we employ an ATmega328 microcontroller as an external trigger. With this external trigger, the cameras are synchronized at the hardware level, ensuring that the RGB camera starts exposure at a rising edge signal and ends exposure at a falling edge signal. Moreover, the event camera captures data from the rising edge to the falling edge signal, corresponding to the exposure time of blurry images. Finally, considering the field of view for both the event camera and each RGB camera, we adjusted the resolution of the images from the two RGB cameras and the events to a size of 1024×768. We have depicted the samples of the dataset in the Fig. A3.

C Visual results

C.1 More visual results

In Fig. A4, A5, A6, we show more qualitative results on the RELED dataset. In the figure, we compare our method with state-of-the-art frame-based LLE methods LLFormer [12], SNRNet [16], frame-based deblurring methods NAFNet [4], and event-guided deblurring methods GEM [18] and UEVD [6]. As evident from the figure, our method consistently achieves superior qualitative restoration results in terms of scene details, such as letters, compared to other methods.

C.2 Video demos

We produced demo videos on the RELED datasets. These videos, labeled as Video_demo.mp4, demonstrate a comparison of qualitative video demos between our methods and other methods.

Fig. A1: Qualitative results on the real-world low-light blurred videos. From left to right: input, UEVD [6], Ours. Zoom in for better view.

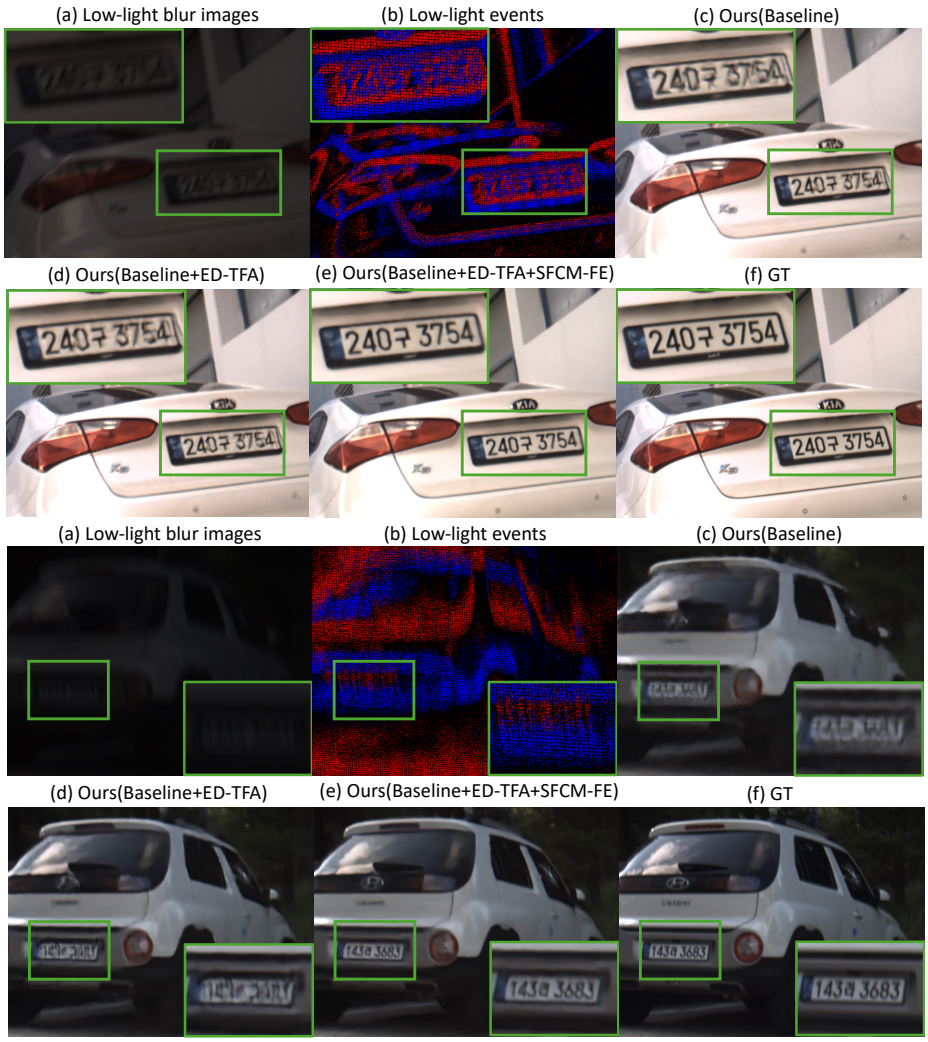

Fig. A2: Qualitative ablations results. Each subfigure represents: (a) low-light blurry image input, (b) low-light event input, (c) the baseline network, (d) the model with the addition of the ED-TFA module to the baseline, and (e) the model with the addition of both the ED-TFA module and the SFCM-FE module to the baseline. We can observe a gradual improvement in qualitative results from (c) to (e).

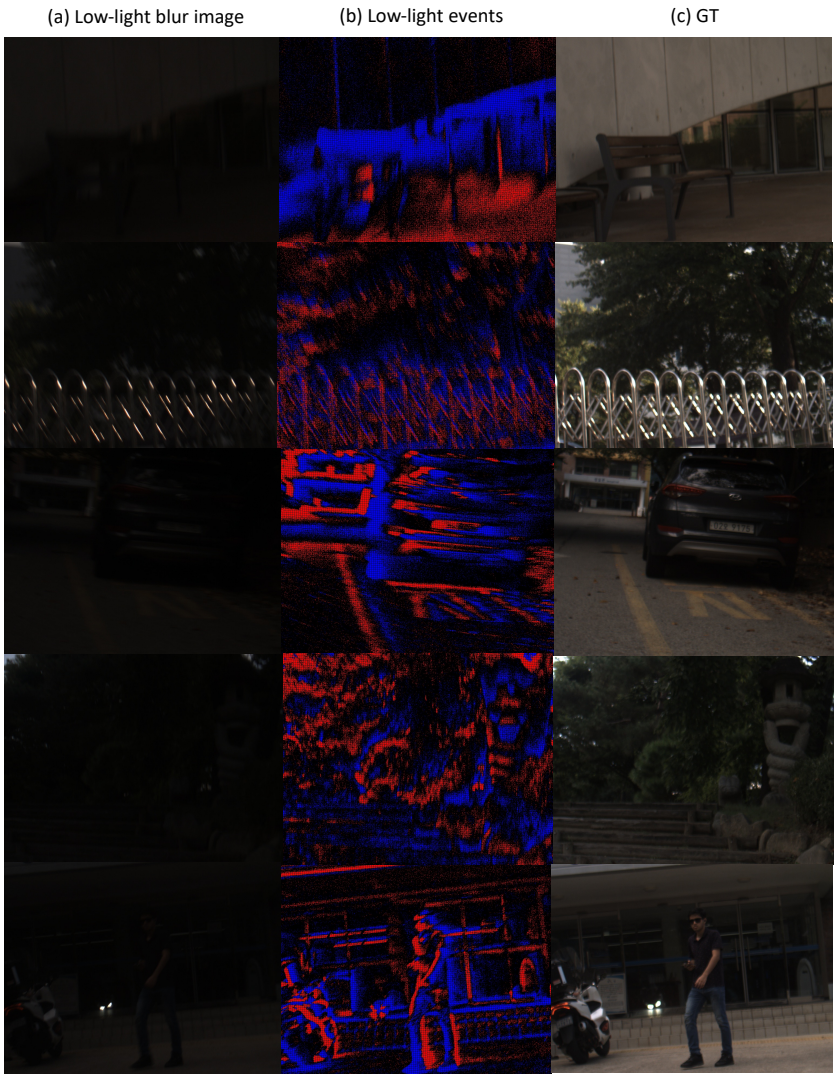

Fig. A3: The examples of our RELED dataset. Our dataset contains diverse scenes and motion, which consists of low-light blurry frames and normal-light sharp images and synchronized event streams.

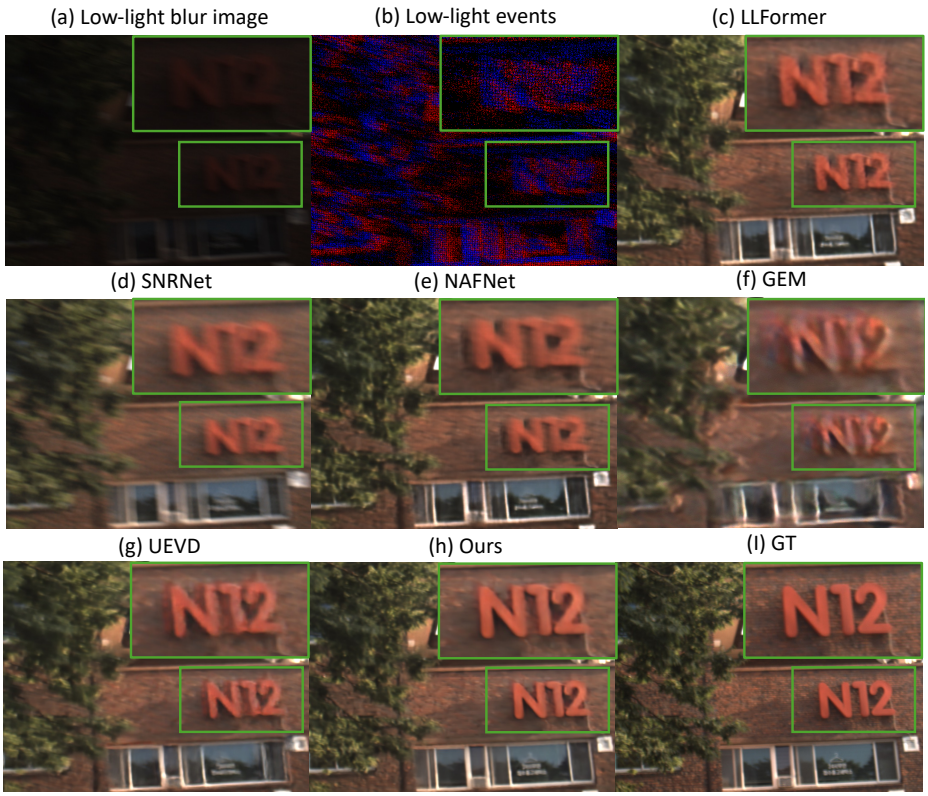

Fig. A4: Visual comparisons of different methods on the RELED datasets. In the Fig, restoration results from (c) to (I) are as follows: LLformer [12], SNRNet [16], NAFNet [4], GEM [18], UEVD [6], Ours, GT.

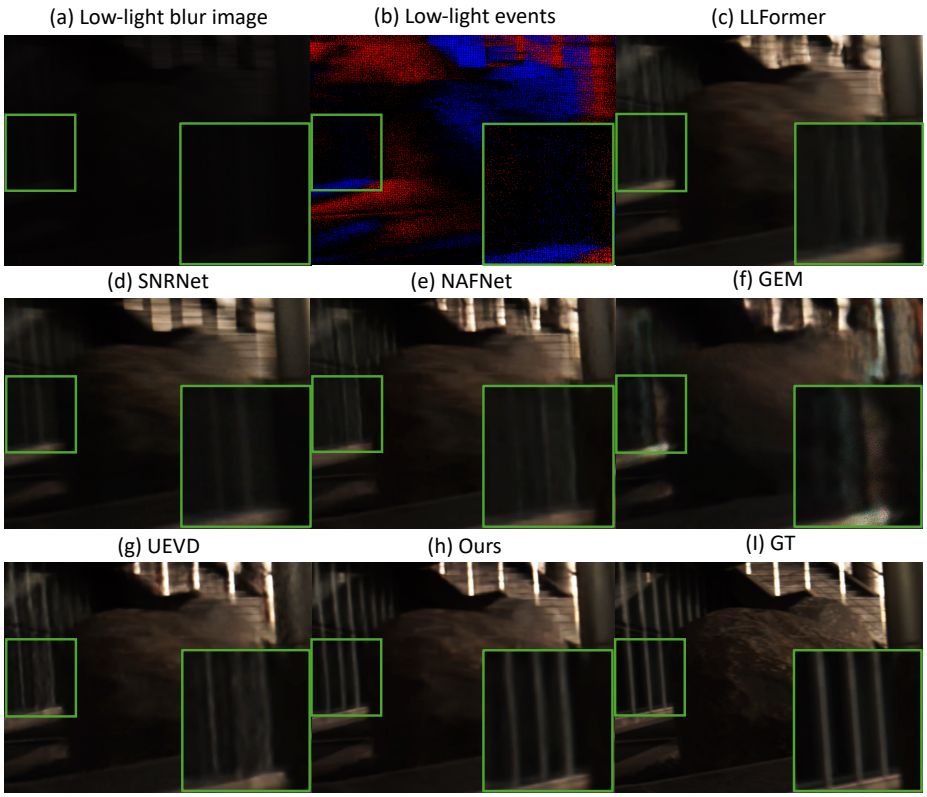

Fig. A5: Visual comparisons of different methods on the RELED datasets. In the Fig, restoration results from (c) to (I) are as follows: LLformer [12], SNRNet [16], NAFNet [4], GEM [18], UEVD [6], Ours, GT.

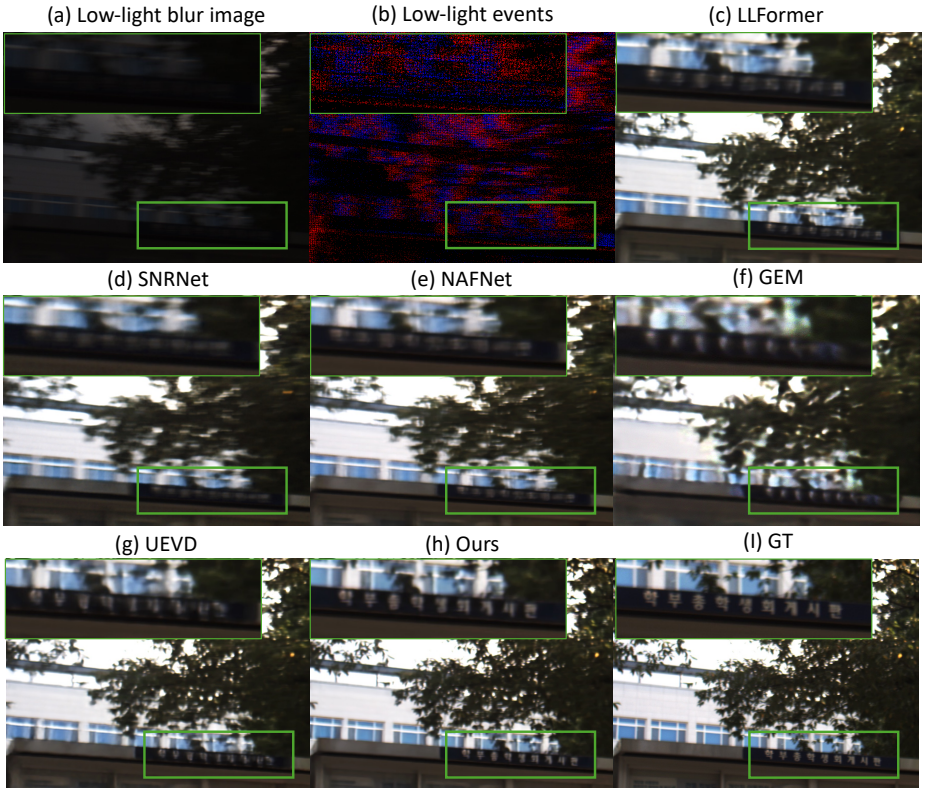

Fig. A6: Visual comparisons of different methods on the RELED datasets. In the Fig, restoration results from (c) to (I) are as follows: LLformer [12], SNRNet [16], NAFNet [4], GEM [18], UEVD [6], Ours, GT.

References

1. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV (2023) 2
2. Chao, Z., Hang, D., Jinshan, P., Boyang, L., Yuhao, H., Lean, F., Fei, W.: Deep recurrent neural network with multi-scale bi-directional propagation for video deblurring. In: AAAI (2022) 2
3. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of 1st international conference on image processing. vol. 2, pp. 168–172. IEEE (1994) 1, 3
4. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European Conference on Computer Vision. pp. 17–33. Springer (2022) 4, 8, 9, 10
5. Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4641–4650 (2021) 2
6. Kim, T., Lee, J., Wang, L., Yoon, K.J.: Event-guided deblurring of unknown exposure time videos. In: European Conference on Computer Vision. pp. 519–538. Springer (2022) 1, 2, 4, 5, 8, 9, 10
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 1
8. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017) 1
9. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4161–4170 (2017) 2
10. Sun, L., Sakaridis, C., Liang, J., Jiang, Q., Yang, K., Sun, P., Ye, Y., Wang, K., Gool, L.V.: Event-based fusion for motion deblurring with cross-modal attention. In: European Conference on Computer Vision. pp. 412–428. Springer (2022) 2
11. Sun, L., Sakaridis, C., Liang, J., Sun, P., Cao, J., Zhang, K., Jiang, Q., Wang, K., Van Gool, L.: Event-based frame interpolation with ad-hoc deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18043–18052 (2023) 2
12. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2654–2662 (2023) 2, 4, 8, 9, 10
13. Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: Edvr: Video restoration with enhanced deformable convolutional networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019) 2
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004) 1
15. Xu, F., Yu, L., Wang, B., Yang, W., Xia, G.S., Jia, X., Qiao, Z., Liu, J.: Motion deblurring with real events. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2583–2592 (2021) 2

16. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022) 2, 4, 8, 9, 10
17. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021) 2
18. Zhang, X., Yu, L., Yang, W., Liu, J., Xia, G.S.: Generalizing event-based motion deblurring in real-world scenarios. In: ICCV (2023) 2, 4, 8, 9, 10