Towards Robust Event-guided Low-Light Image Enhancement: A Large-Scale Real-World Event-Image Dataset and Novel Approach –Supplementary Material–

Guoqiang Liang[1] Kanghao Chen[1] Hangyu Li[1] Yunfan Lu[1] Lin Wang[1,2]
[1]AI Thrust, HKUST(GZ) [2]Dept. of Computer Science and Engineering, HKUST
{gliang041,kchen879,hli886,ylu066}@connect.hkust-gz.edu.cn, linwang@ust.hk

1. More details of our SDE Dataset

Details of Device To ensure precise spatial-temporal alignment, we design a robotic alignment system. This system is composed of several key elements: a DAVIS346 event camera [8], a robotic arm Universal Robot UR5e, an ND 8 filter from NiSi, and an illuminance meter, as shown in Fig. 1. The specific settings for both the robotic arm and event camera are contingent upon predefined paths and the lighting conditions within the scenes. In practical terms, the camera's exposure time varies between 10 ms and 80 ms, maintaining a fixed frame interval of 5 ms [1].

Details of Matching Alignment Strategy. Fig. 2 (a) illustrates the variance between intervals a and b (highlighted in blue regions), despite identical camera and robot settings, leading to unpredictable temporal alignment errors. Our approach to mitigating this issue involves the introduction of a novel matching alignment strategy. In Fig. 2 (b), we exemplify the alignment process by capturing six paired event-image sequences for each scene—three under low-light and three under normal-light conditions. These sequences are precisely trimmed to match the start and end timestamps of the predefined trajectory, ensuring consistent content across all videos. Subsequently, we perform matching between the 1st low-light sequence and the 2nd normal-light sequence, achieving minimal absolute errors of $1ms$.

Details of Distribution of Our SDE dataset. Fig. 3 visualizes the temporal alignment error distribution (measured in seconds), the video length distribution within our dataset, and the illumination variations across filming environments. Exceeding the measured Lux levels of the SDSD dataset [10], we capture sequences in well-illuminated environments due to the limitations imposed by the lens (with an aperture of f2.0) and sensor size.

Benefits of complex motion. Our SDE dataset, rich in motion variety, is critical for future research in event-guided low-light video enhancement, providing challenging and real-world scenarios to test temporal consistency. This positions it as a fundamental resource for the event-based vision community.

[1]DV software may adjust the interval between frame and exposure to suit the desired exposure duration.

Figure 1. The devices used for collecting our dataset. From the left to right is the Universal Robot UR5e, the DAVIS 346 event camera, an ND8 filter, and an illuminance meter.

Input	Method	FLOPs (G)	Params (M)
Event Only	E2VID+ (ECCV'20) [7]	27.99	10.71
Image Only	SNR-Net (CVPR'22) [13]	26.35	4.01
	Uformer (CVPR'22) [11]	12.00	5.29
	LLFlow-L-SKF (CVPR'23) [12]	409.50	39.91
	Retinexformer (ICCV'23) [1]	15.57	1.61
Image+Event	ELIE (TMM'23) [2]	440.32	33.36
	eSL-Net (ECCV'20) [9]	560.94	0.56
	Liu *et al.* (AAAI'23) [3]	44.71	47.06
	Ours	180.90	22.73

Table 1. Comparison of computational complexity.

Method	ELIE [2]	eSL-Net [9]	Liu *et al.* [3]	Ours
Inference time	108.42 ms	37.05 ms	27.39 ms	30.76 ms

Table 2. Inference time with the resolution of 256×256.

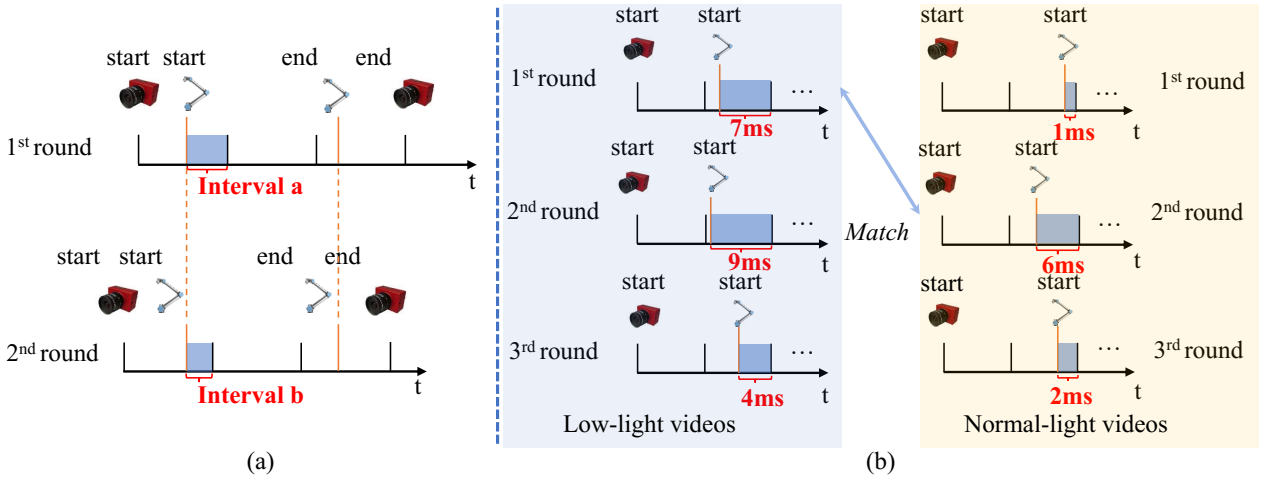


Figure 2. (a) An illustration of the variable time interval between the start timestamp of the trajectory and the first frame timestamp in each sequence. (b) An example of the matching alignment strategy.

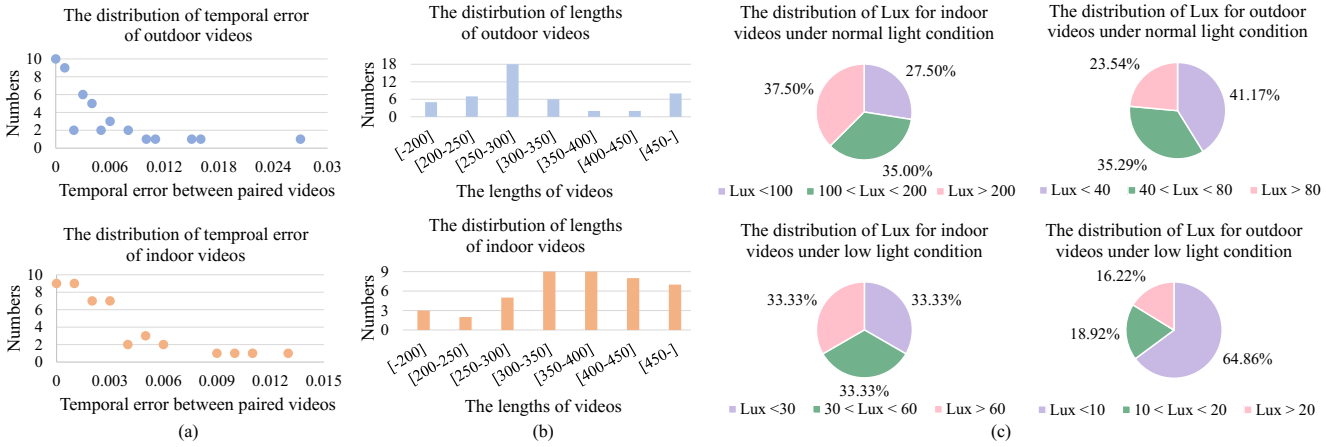


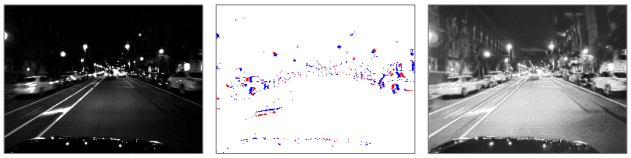
Figure 3. (a) Distribution of temporal alignment error (measured in seconds) of our dataset. (b) Distribution of video length of our dataset. (c) Illumination distribution in the filming environment.

2. Additional Analysis

More Implementation Details. As outlined in the main paper, our approach utilizes the dynamic version of the SDSD dataset [10] to synthesize event streams. Notably, this dataset's training and testing sets differ from the static version employed in SNR-Net [13] and Retinexformer [1]. Consequently, we undertake the task of retraining all these methods instead of relying on quoted results from Retinexformer [1]. In addition, we use the mixed precision [4] training tool provided by PyTorch [5] , which can speed up our training and reduce memory usage.

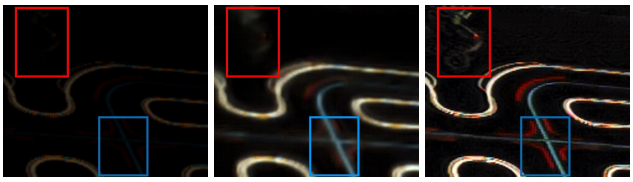
Computational Cost. In order to evaluate the computational complexity of our proposed method in contrast to several state-of-the-art (SOTA) methodologies, we present both the floating point operations (FLOPs) and the total number of parameters (Params) in Tab. 1. The computation of FLOPs has been carried out at a resolution of 256 $\times$ 256. It is important to note that the channel numbers in Liu *et al*. [3] have not been explicitly stated; therefore, for the purpose of maintaining a fair comparison, we have assumed the channel number in each layer to be identical to that in our method. In Tab. 2, we compare the inference time of our method with other event-guided LIE methods, all tested on an NVIDIA A30 GPU with mixed precision [4]. Our inference time is comparable to that of Liu *et al*. [3] and shorter than ELIE's [2] and eSL-Net's [9].

Generalization Ability. In order to assess the generalization capability of our proposed approach, we carry out an experiment in the CED dataset [6] with the model trained in our SDE dataset. Given that the CED dataset lacks paired normal-light images to serve as ground truth, we offer qual-



(a) Input image (b) Input event (c) Ours

Figure 4. Qualitative results on MVSEC [14] dataset.



(a) Input low-light image (b) Retinexformer [4] (c) Ours

Figure 5. Qualitative results of generalization evaluation.

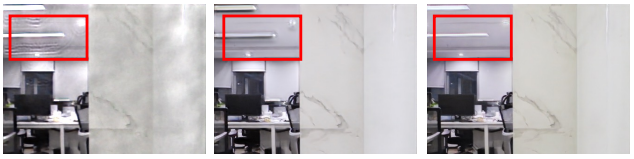
itative results derived from challenging examples captured from a moving vehicle within a tunnel, under varying lighting conditions such as with or without sunlight. As depicted in Fig. 7 and Fig. 8, our method capably recovers details obscured in the low-light image, while simultaneously preventing overexposure as seen in (d) and color distortion in (f) and (i). Besides, even the structured details of the top of the tunnel are reconstructed by our method as shown in Fig. 8 (g). This effectively exhibits the robustness of our methodology. Ultimately, we have conducted two experiments: one on the MVSEC, as detailed by [14], and another wherein the model, having been trained on the synthetic events from the SDSD dataset [10], underwent evaluation against the real events captured within our SDE dataset. The visual results of these experiments are shown in Fig. 4 and 5, respectively.

Performance with higher resolution. To explore our performance with the resolution of 512×512, we follow the same process in the main paper to synthesize data with SDSD dataset [10]. As shown in Tab. 3 and Fig. 6, our method achieves the **SOTA** performance and the **best** visual result (see red box in Fig. 6).

Method	eSL-Net [9]	Liu *et al*. [3]	Ours
PSNR / SSIM	25.83 / 0.8441	28.12 / 0.8791	**28.71 / 0.8923**

Table 3. Numerical results on SDSD [10] with a resolution of 512×512.

More Visual Comparison Results. We provide more visual comparisons on both our SDE dataset and SDSD [10] dataset. We compare our method with recent methods with three different settings: **(I)** the experiment with events as input, including E2VID+ [7]. **(II)** the experiment with a RGB image as input, including SNR-Net [13], Uformer [11], LLFlow-L-SKF [12], and Retinexformer [1]. **(III)** the experiment with a RGB image and paired events as inputs,



(a) eSL-Net [35] (b) Liu et al. [23] (c) Ours

Figure 6. Visual results on SDSD [10] with the resolution of 512×512.

including ELIE [2], eSL-Net [9], and Liu *et al*. [3]. Visual comparison results on our SDE dataset are shown in Fig. 9 and Fig.10, while results on SDSD [10] dataset are shown in Fig. 11 and Fig. 12.

References

[1] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. *arXiv preprint arXiv:2303.06705*, 2023. 1, 2, 3

[2] Yu Jiang, Yuehang Wang, Siqi Li, Yongji Zhang, Minghao Zhao, and Yue Gao. Event-based low-illumination image enhancement. *IEEE Transactions on Multimedia*, 2023. 1, 2, 3

[3] Lin Liu, Junfeng An, Jianzhuang Liu, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, Yan Feng Wang, and Qi Tian. Low-light video enhancement with synthetic event guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1692–1700, 2023. 1, 2, 3

[4] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, et al. Mixed precision training. *arXiv preprint arXiv:1710.03740*, 2017. 2

[5] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 2

[6] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. Ced: Color event camera dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2, 5, 6

[7] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 534–549. Springer, 2020. 1, 3

[8] Gemma Taverni, Diederik Paul Moeys, Chenghan Li, Celso Cavaco, Vasyl Motsnyi, David San Segundo Bello, and Tobi Delbruck. Front and back illuminated dynamic and active pixel vision sensors comparison. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 65(5):677–681, 2018. 1

[9] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *Com-*

puter Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, pages 155–171. Springer, 2020. 1, 2, 3

[10] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9700–9709, 2021. 1, 2, 3, 9, 10

[11] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 1, 3

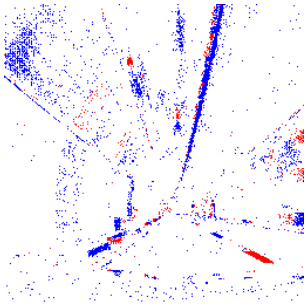
[12] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1662–1671, 2023. 1, 3

[13] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 1, 2, 3

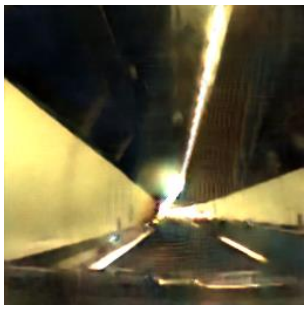
[14] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3 (3):2032–2039, 2018. 3

(a) Input image

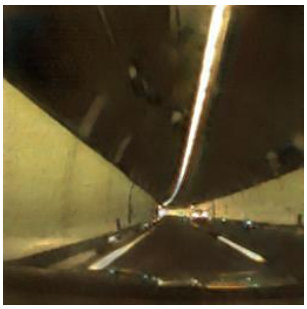
(b) Image after gamma correction



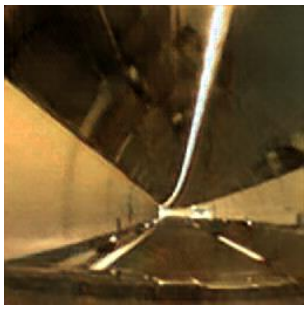
(c) Input event



(d) eSL-Net

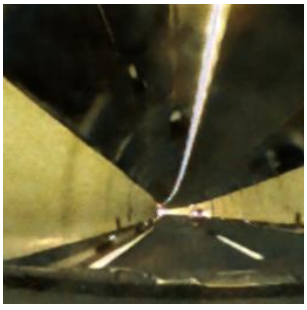


(e) Liu et al.

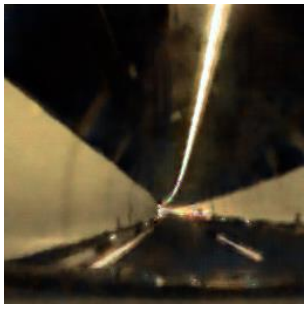


(f) ELIE

(g) ours



(h) Retinexformer

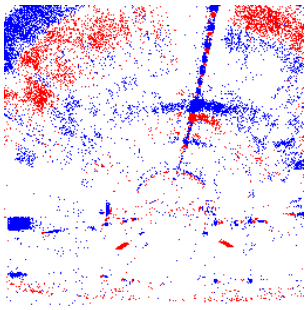


(i) SNR-Net

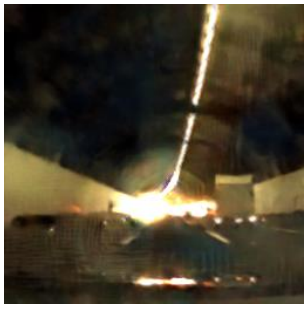
Figure 7. Generalization results in the driving tunnel sequence captured in CED [6] dataset.

(a) Input image

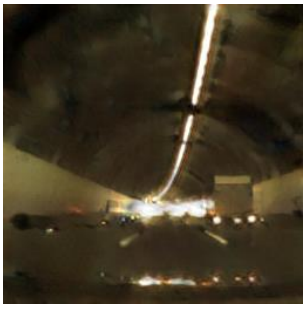
(b) Image after gamma correction



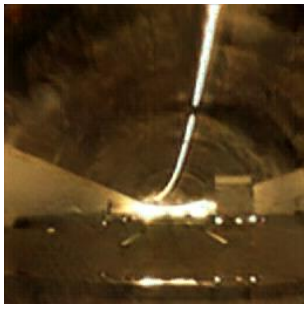
(c) Input event



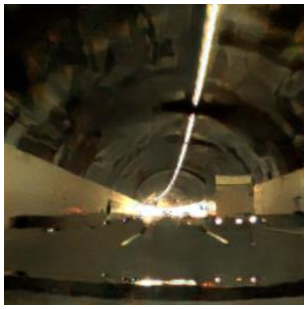
(d) eSL-Net



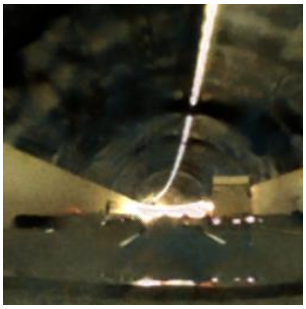
(e) Liu et al.



(f) ELIE



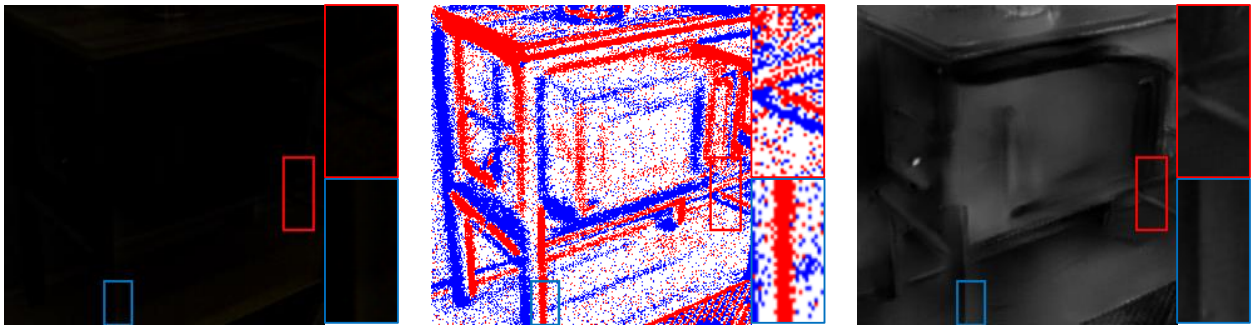
(g) ours



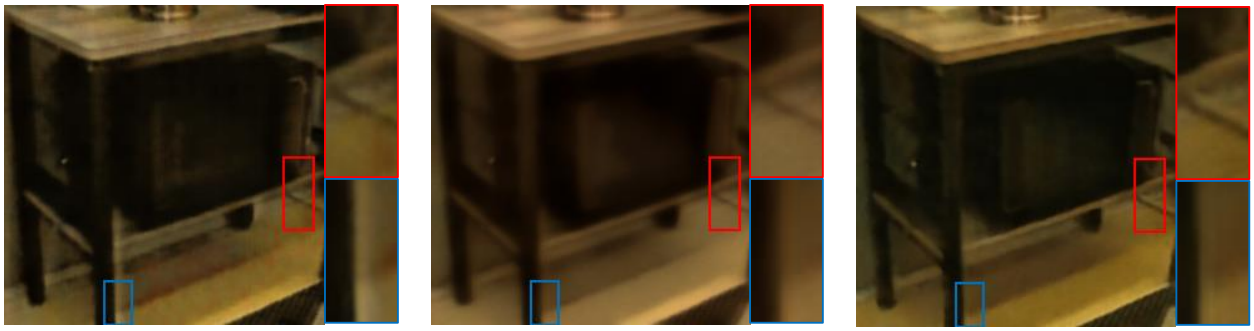
(h) Retinexformer

(i) SNR-Net

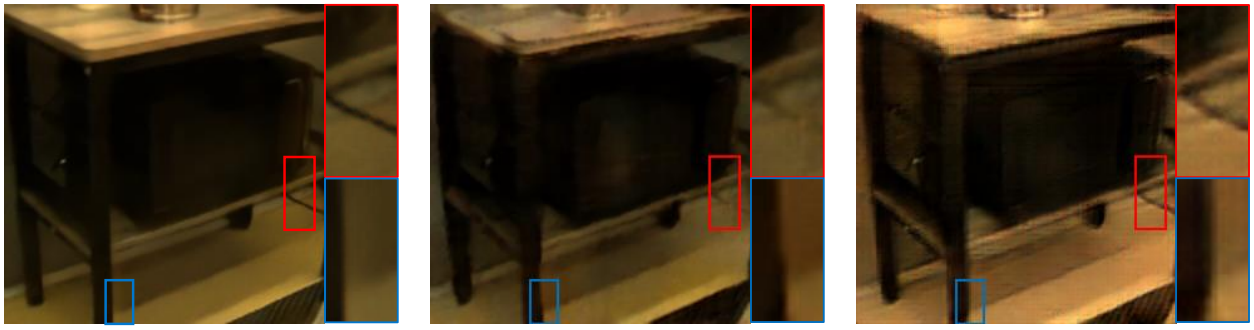
Figure 8. Generalization results in the driving tunnel sun sequence captured in CED [6] dataset.



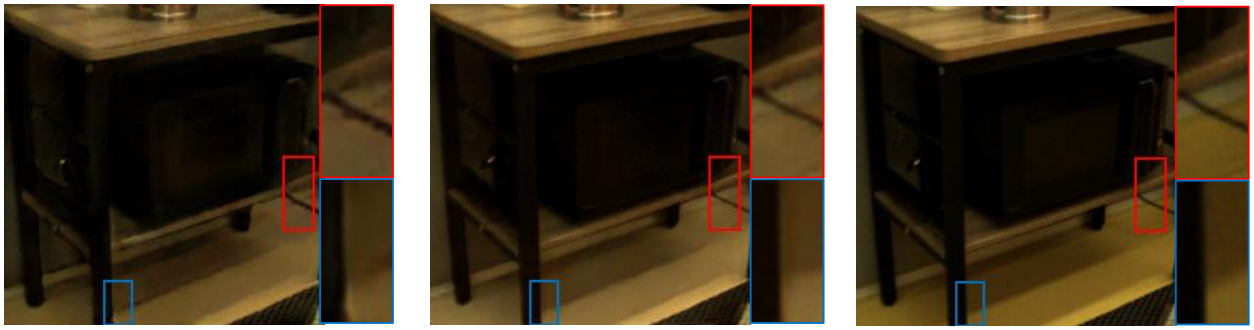
(a) Input image (b) Input event (c) E2VID+



(d) SNR-Net (e) LLFlow-L-SKF (f) Retinexformer

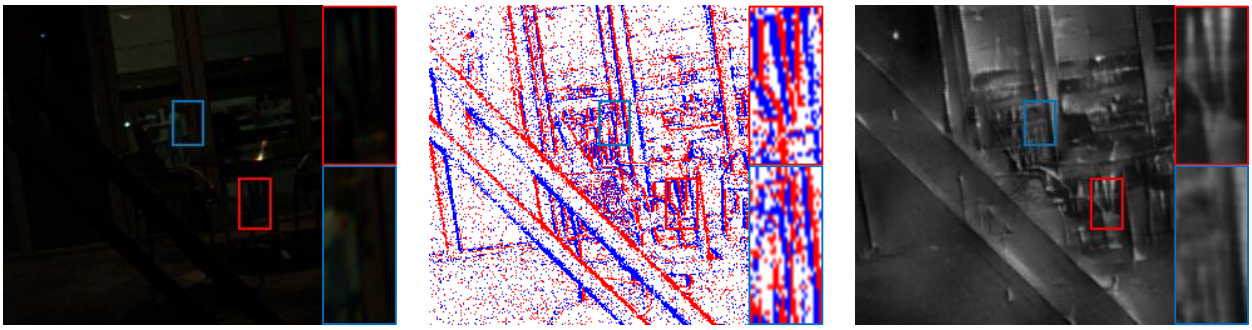


(g) Uformer (h) eSL-Net (i) ELIE



(j) Liu et al. **(k) Ours** (l) GT

Figure 9. Qualitative results on our SDE dataset.



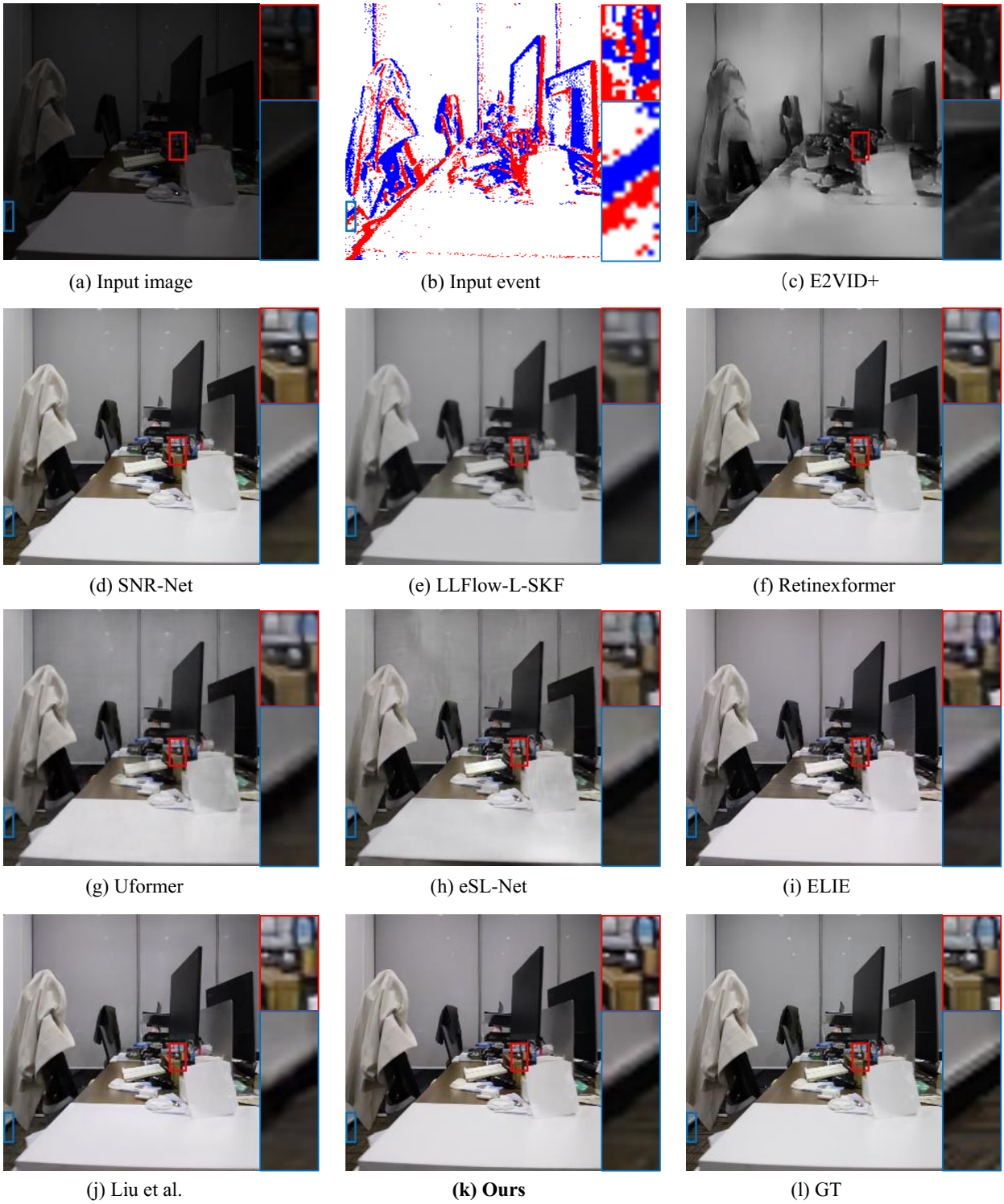
(a) Input image (b) Input event (c) E2VID+

(d) SNR-Net (e) LLFlow-L-SKF (f) Retinexformer

(g) Uformer (h) eSL-Net (i) ELIE

(j) Liu et al. **(k) Ours** (l) GT

Figure 10. Qualitative results on our SDE dataset.



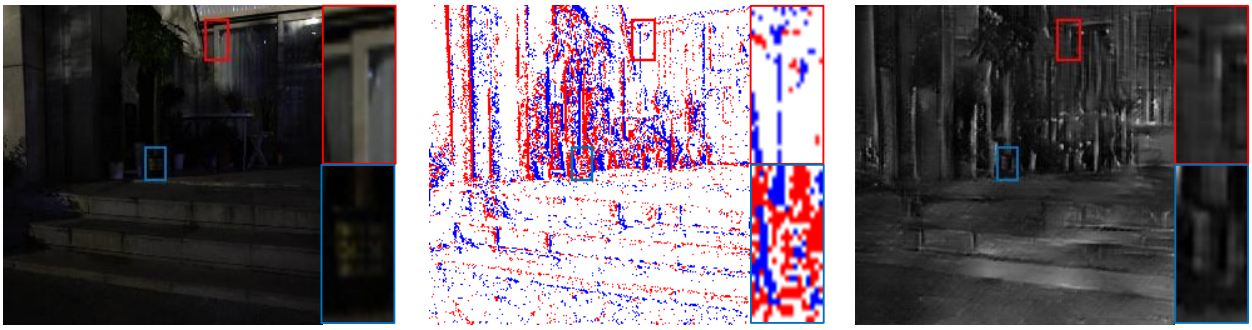
(a) Input image (b) Input event (c) E2VID+

(d) SNR-Net (e) LLFlow-L-SKF (f) Retinexformer

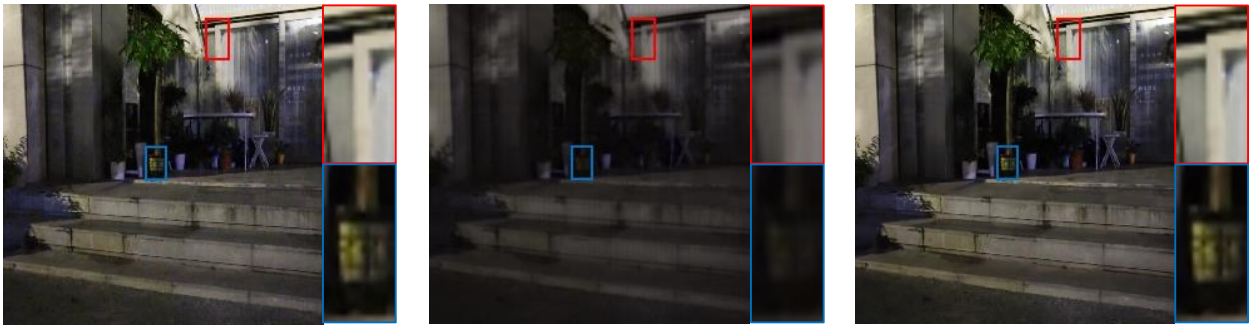
(g) Uformer (h) eSL-Net (i) ELIE

(j) Liu et al. **(k) Ours** (l) GT

Figure 11. Qualitative results on SDSD dataset [10].



(a) Input image (b) Input event (c) E2VID+



(d) SNR-Net (e) LLFlow-L-SKF (f) Retinexformer

(g) Uformer (h) eSL-Net (i) ELIE

(j) Liu et al. **(k) Ours** (l) GT

Figure 12. Qualitative results on SDSD dataset [10].