Supplementary Materials: Exploring in Extremely Dark: Low-Light Video Enhancement with Real Events

Anonymous Authors

1 INTRODUCTION

In this supplementary material, we first provide detailed configuration of V2E. Then we provide computational cost and further discussions of the REN. In addition, we present the complete experimental results of unsupervised methods on SDSD testset. We show more visual comparisons of different low-light enhancement approaches on real and synthetic datasets, and give the full enhanced videos of real data in the zip file. Finally, we provide more visual results of ablation studies.

2 THE DETAILED CONFIGURATION OF V2E

We employ V2E [4] to simulate events from SDSD [8] dataset. Considering the diverse luminance levels across different scenes, it is essential to tailor the V2E by employing distinct parameter configurations. As shown in Table 1, we employ two distinct sets of parameters to simulate events in the indoor and outdoor subsets of the SDSD dataset. The pos_thres, neg_thres and sigma_thres represent distinct thresholds within the V2E, setting dvs_params as clean means turning off noise, while cutoff_hz means pixel cutoff frequency in Hz.

3 ADDITIONAL ANALYSIS

Computational cost To compare the computational complexity of our approach with other low-light video enhancement methods, we report floating point operations (denoted by FLOPs) in Table 2. FLOPs are computed based on video frames with a resolution of 640×480. We also provide PSNR values and SSIM values for reference. It can be observed that REN achieves the best results under moderate computational complexity.

More results of ablation study The Event-Image Attention consists of two components: the cross-modal attention and the establishment of spatial mapping between image and event features. The former is employed to emphasize event features in extremely dark regions, while the latter is utilized for spatial alignment of these two modalities.

We conduct more detailed ablation studies on REN_{sd} and provide the ablation studies on REN for reference. The REN_{sd} is trained exclusively on simulated events that are aligned with images, while the REN's training dataset comprises misaligned images and real events. We employ two modified models based on them and the results are shown in Table 3. Full Model represents the complete network architecture, while W/o EIA and W/o W_p mean remove Event-Image Attention module and remove W_p from the module respectively, which are same to the main paper.

The PSNR value of W/o W_p exhibits minimal variation compared to Full Model based on REN_{sd}, while the PSNR value of W/o W_p shows a noticeable decrease based on REN. The results indicate that W_p enhances performance on misaligned real data, suggesting its effectiveness in handling misalignments between events and images. We observe that in the experiments on REN_{sd}, the PSNR of the

Full model is lower than the result of the W/o W_p. The performance decreases due to the inaccurate aligning, the W_p aligns images and events based on similarity matrix, it is less accurate than direct summation on aligned data. Meanwhile, the EIA module demonstrates effective performance improvement on both simulated and real data.

The complete results of unsupervised methods on the SDSD testset. As unsupervised methods can be trained on SDSD dataset or SDSD+DSEC datasets, which is similar to ours, we only present the best results of these methods in Table 1 of the main paper. We present the complete experimental results in Table 4.

Table 1: Detailed configuration of V2E.

Setting	Indoor	Outdoor
pos_thres	0.15	0.001
neg_thres	0.15	0.001
sigma_thres	0.003	0.0002
dvs_params	clean	clean
cutoff_hz	15	15

Table 2: Computational cost of different low-light video enhancement methods.

Method	PSNR↑	SSIM↑	FLOPs ↓
MBLLEN [6]	21.79	0.65	210.34 G
SGLLIE [11]	23.89	0.70	3.30 G
SMID [1]	24.09	0.69	0.16 G
SDSDNet [8]	24.92	0.73	214.44 G
LAN [2]	27.25	0.85	110.48 G
REN	29.03	0.88	178.68 G

Table 3: More detailed ablation studies on REN_{sd} and REN.

Model	Method	PSNR↑	SSIM↑	LPIPS ↓
REN_{sd}	W/o EIA	26.81	0.86	0.10
	W/o W_p	28.52	0.87	0.10
	Full Model	28.45	0.88	0.09
REN	W/o EIA	26.29	0.87	0.09
	W/o W_p	27.63	0.87	0.09
	Full Model	29.03	0.88	0.09

4 MORE VISUAL COMPARISON RESULTS

We present more visual results of different low-light enhancement methods on synthetic data and real data. Fig. 1 shows the visual results of different methods on SDSD synthetic dataset. Fig. 2 shows the visual results of different unsupervised methods on DSEC real

Table 4: The complete results of unsupervised methods on the SDSD testset.

Methods	SDSD			SDSD + DSEC		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SCI [7]	19.67	0.69	0.29	19.18	0.65	0.16
NeRCo [10]	20.02	0.65	0.19	20.07	0.66	0.21
CLIP-LIT [5]	22.29	0.72	0.16	14.56	0.47	0.23
SGLLIE [11]	23.89	0.70	0.31	9.26	0.30	0.42
REN_{sd}	28.45	0.88	0.09	/	/	/
REN	/	/	/	29.03	0.88	0.09

dataset. Fig. 3 shows the visual performance of cross-dataset evaluations on ViViD++ dataset. Fig. 4 shows the visual results of ablation studies. As shown in the results, our methods can achieve excellent visual performance in both illumination and details, indicating that our proposed approach not only has better performance but also has the excellent generalization ability. We present the full enhanced videos in the zip file, including the results of our proposed method and the previous methods on DSEC real dataset and ViViD++ real dataset.

REFERENCES

[1] Chen Chen, Qifeng Chen, Minh N Do, and Vladlen Koltun. 2019. Seeing motion in the dark. In *Proceedings of the IEEE/CVF International conference on computer vision*. 3185–3194.

[2] Huiyuan Fu, Wenkai Zheng, Xicong Wang, Jiaxuan Wang, Heng Zhang, and Huadong Ma. 2023. Dancing in the Dark: A Benchmark towards General Low-light Video Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12877–12886.

[3] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. 2023. Learning a Simple Low-Light Image Enhancer From Paired Low-Light Instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22252–22261.

[4] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. 2021. v2e: From video frames to realistic DVS events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1312–1321.

[5] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. 2023. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8094–8103.

[6] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. 2018. MBLLEN: Low-Light Image/Video Enhancement Using CNNs.. In *BMVC*, Vol. 220. 4.

[7] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5637–5646.

[8] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. 2021. 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*. 9700–9709.

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

[10] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. 2023. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12918–12927.

[11] Shen Zheng and Gaurav Gupta. 2022. Semantic-guided zero-shot learning for low-light image/video enhancement. In *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*. 581–590.

(a) Low-light Frame   (b) Event   (c) SCI [7]   (d) NeRCo [10]

(e) CLIP-LIT [5]   (f) SDSDNet [8]   (g) SDSDNet + Event   (h) LAN [2]

(i) LAN + Event   (j) REN_{sd} (k) REN   (l) GroundTruth

(m) Low-light Frame   (n) Event   (o) SCI [7]   (p) NeRCo [10]

(q) CLIP-LIT [5]   (r) SDSDNet [8]   (s) SDSDNet + Event   (t) LAN [2]

(u) LAN + Event   (v) REN_{sd} (w) REN   (x) GroundTruth

Figure 1: Visual quality comparison of enhancement results on SDSD synthetic dataset.

Anonymous Authors

(a) Low-light Frame   (b) Event   (c) SCI [7]   (d) NeRCo [10]

(e) CLIP-LIT [5]   (f) SGLLIE [11]   (g) SMID [1]   (h) SNR-Aware [9]

(i) SDSDNet [8]   (j) LAN [2]   (k) REN_{sd} (l) REN

(m) Low-light Frame   (n) Event   (o) SCI [7]   (p) NeRCo [10]

(q) CLIP-LIT [5]   (r) SGLLIE [11]   (s) SMID [1]   (t) SNR-Aware [9]

(u) SDSDNet [8]   (v) LAN [2]   (w) REN_{sd} (x) REN
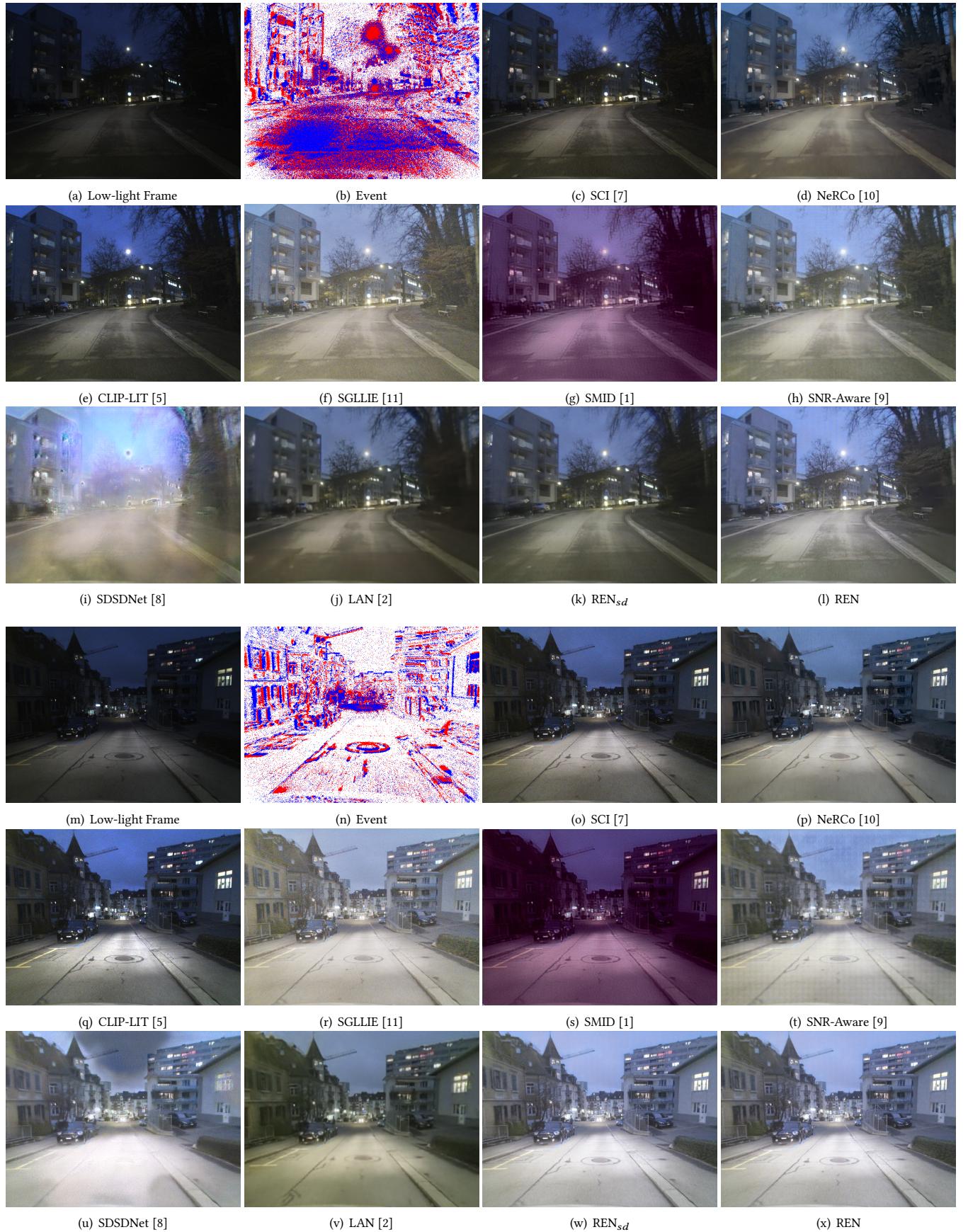
Figure 2: Visual quality comparison of enhancement results on DSEC real dataset.

(a) Low-light Frame

(b) Event

(c) SCI [7]

(d) NeRCo [10]

(e) CLIP-LIT [5]

(f) Pair-LIE [3]

(g) SMID [1]

(h) MBLLEN [6]

(i) SDSDNet [8]

(j) LAN [2]

(k) REN_{sd}

(l) REN

(m) Low-light Frame

(n) Event

(o) SCI [7]

(p) NeRCo [10]

(q) CLIP-LIT [5]

(r) Pair-LIE [3]

(s) SMID [1]

(t) MBLLEN [6]

(u) SDSDNet [8]

(v) LAN [2]

(w) REN_{sd}

(x) REN

Figure 3: Visual quality comparison of enhancement results on ViViD++ real dataset

(a) Low-light Frame
(b) Event

(c) Only Video
(d) W/o EIA
(e) W/o DRM
(f) W/o W_p

(g) W/o L_{con}
(h) W/o L_t
(i) REN_{sd}
(j) REN

(k) Low-light Sequence

(l) W/o L_t Sequence

(m) REN Sequence
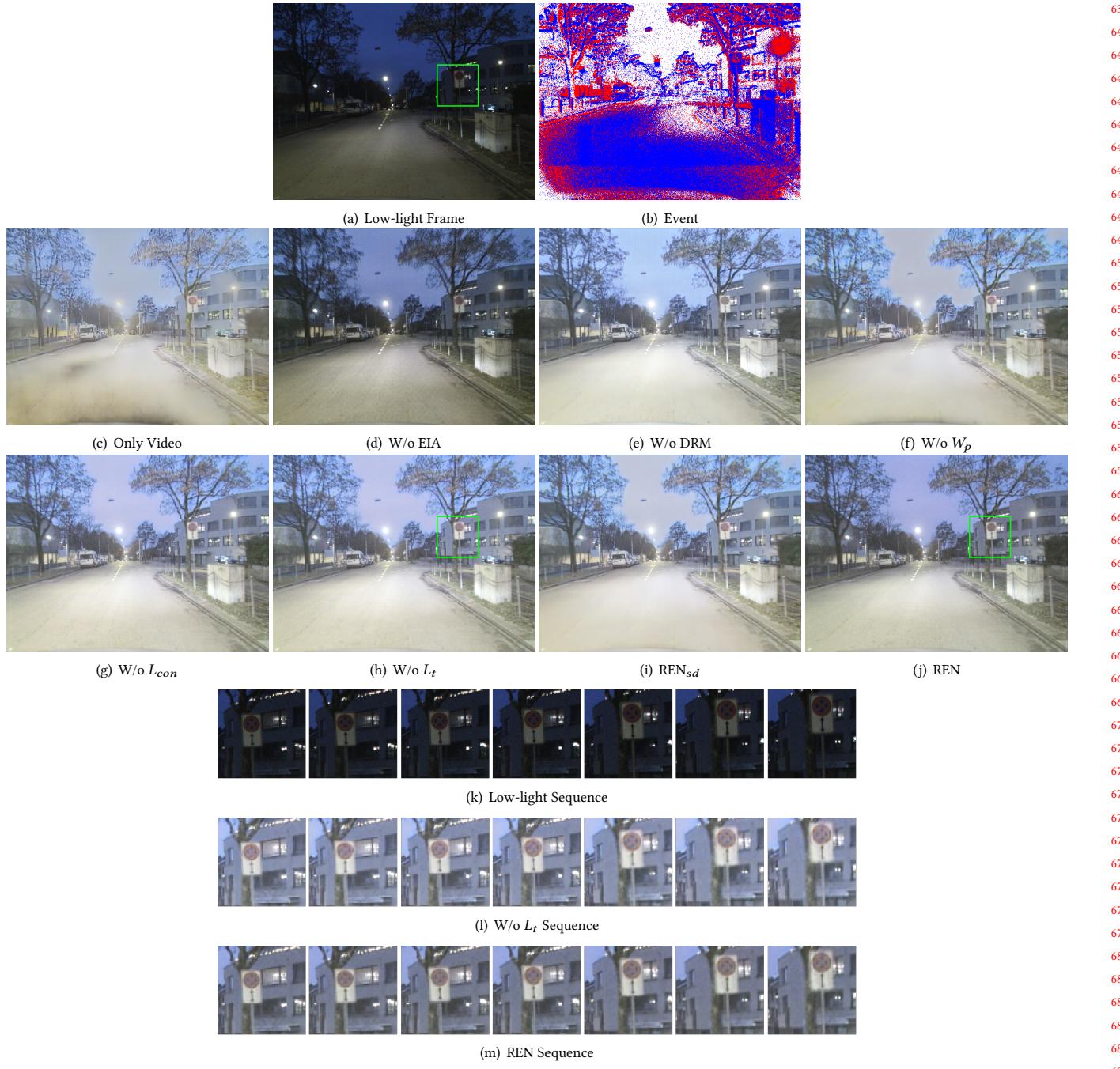
Figure 4: Visual quality comparison of ablation studies. (k), (l) and (m) present comparison between W/o L_t and REN in video sequence. The results of W/o L_t exhibit noticeable artifacts and discontinuities, it indicates the contribution of L_t in temporal consistency.