Low-Light Image Enhancement Using Event-Based Illumination Estimation Supplementary Material

Lei Sun[1] Yuhan Bao[2] Jiajun Zhai[2] Jingyun Liang[3] Yulun Zhang[4†] Kaiwei Wang[2†]
Danda Pani Paudel[1] Luc Van Gool[1]

[1]INSAIT, Sofia University "St. Kliment Ohridski" [2]Zhejiang University
[3]Alibaba Group [4]Shanghai Jiao Tong University

1. Why Using Event Camera for Low-Light Image Enhancement?

Event cameras, with their distinct advantages, have been widely utilized to enhance RGB images captured by conventional image sensors, contributing to tasks such as deblurring [7, 11, 13], video frame interpolation [14, 15], and high-dynamic-range (HDR) imaging. In this work, we aim to transfer the low-light responsiveness and HDR imaging capabilities of event cameras to RGB sensors. Looking ahead, as dynamic vision sensors (DVS) become integrated into mass-produced cameras, they are expected to play a significant role in advancing various aspects of image enhancement. With a mechanical shutter, our method can seamlessly integrate with the traditional image exposure process, enabling widespread application across various devices equipped with DVS.

2. More Details about T2I Module

In our Low-Light Degradation model, in the synthetic temporal-mapping events from ground-truth normal-light images, we introduce latency to the original timestamps in the temporal domain, which is inversely proportional to the intensity of the image. In the dark area of the images, some event timestamps may be clipped to the maximum time, resulting "dead pixel" in the images. *e.g.*, 40% of the pixels with grayscale values below 10 in an 8-bit image are randomly set to 0. The contrast threshold c is randomly sampled from $\mathcal{N}(0.2, 0.08)$.

To model blurriness, we define the kernel size as a discrete uniform variable, randomly chosen from $\{7 \times 7, 9 \times 9, \ldots, 21 \times 21\}$. For the isotropic Gaussian blur, the kernel width is sampled from a uniform distribution within [0.05, 2.8]. In the case of anisotropic Gaussian blur, the rotation angle follows a uniform distribution over [0, π], while the axis lengths are independently drawn from [0.5, 6].

For downsampling, we randomly apply either bicubic or bilinear interpolation to reduce the image to a randomly chosen scale within [1/2, 1] of its original size, followed by upsampling back to the original resolution.

For Gaussian-Poisson hybrid noise, we random sample the σ for Gaussian noise from [1/255, 25/255] and the σ for Poisson noise in [1, 10].

All the processes above are different from EvTemMap [1]. Besides, we also differ our T2I module with EvTemMap from estimating Illumination instead of the pixel values directly.

3. Qualitative Results for Ablation Study

Compared to the baseline method without event information, our final results exhibit enhanced texture quality and more accurate illumination estimation, demonstrating the effectiveness of incorporating temporal-mapping events for low-light image enhancement. Furthermore, the well-estimated illumination map, along with the Illumination-aided Reflectance Enhancement (IRE) module, significantly refines the reflectance component R, reducing underexposed regions. The improvement in both I (illumination) and R (reflectance) collectively contributes to a more visually compelling final result.

Additionally, compared to the variant without the proposed LLDM, our method exhibits superior noise suppression, producing a smoother illumination map. Moreover, the reduction of outliers in illumination estimation enhances the balance of I, as evident in the visualized illumination maps.

4. Qualitative Results for Different Low-Light Conditions

In our EvLowLight, images with varying low-light levels are captured. Fig 8 shows the estimated I, R, and final results. Our RETINEV demonstrates robustness across different low-light conditions, preserving fine-grained details.

† Yunlun Zhang: yulun100@gmail.com, Kaiwei Wang: wangkaiwei@zju.edu.cn.

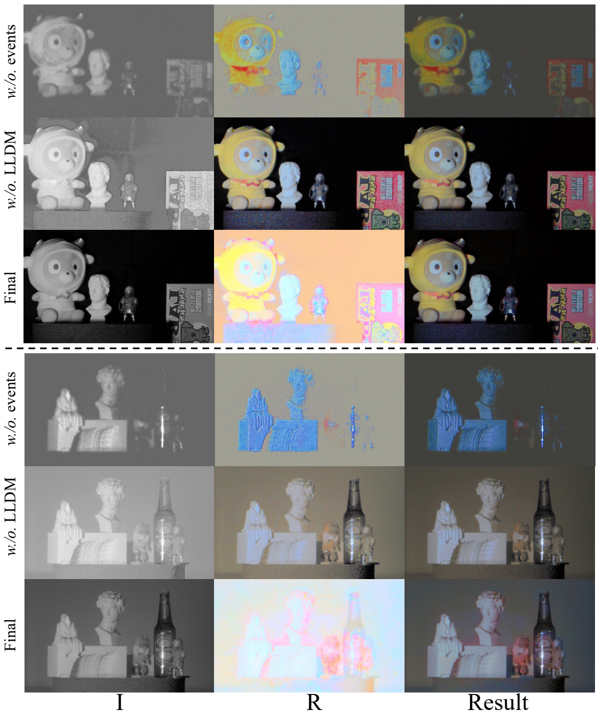


Figure 7. Qualitative results for ablation study. The estimation of I is better in our results, and leads to a better R with our proposed IRE module.

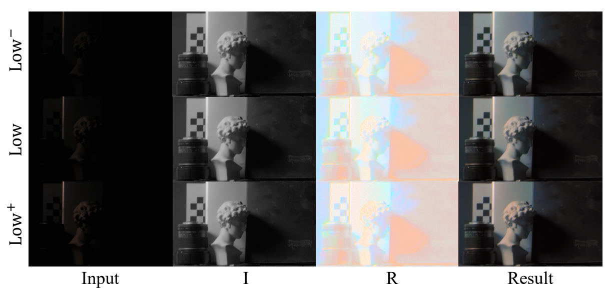


Figure 8. Qualitative results for different low-light conditions. Our RETINEV shows robustness,

5. Experiments on Synthetic Datasets

5.1. Synthetic Events

To generate synthetic events for the LOL v1 [17], LOL v2 [19], and SDSD [16] datasets, we utilize the open-source V2E tool [5] to map grayscale values from images to temporal mapping events. Normal-light images are first converted to grayscale and then transformed into videos with increasing brightness to simulate the effect of a mechanical shutter. The brightness increment rate is set in alignment with the intensity values of the normal-light images. The spatial dimensions of the generated images are preserved to match those in the original datasets. We synthesize events with contrast threshold c set randomly following Gaussian distribution $\mathcal{N}(\mu = 0.2, \sigma = 0.03)$ and $\mathcal{N}(\mu = 0.2, \sigma = 0.05)$ during training and testing stages to simulate scenarios encountered during testing with different camera settings. For all other parameters, the default settings in V2E are used.

5.2. More Qualitative Results

More qualitative results are presented in this Section. Figure 9 and Figure 10 shows results from LIME [3], URetinex-Net [18], GSAD [4], Retinexformer [2], and our RETINEV. Please zoom in for a better view. Previous methods show color distortions (LIME), unnatural artifacts (URetinex-Net, Retinexformer), and less visibility in dark areas (GSAD), while Our methods enhance visibility and contrast in low-light areas, eliminate noise cleanly without producing artifacts or spots, and ensure robust color preservation.

5.3. More Comparisons

5.3.1. Comparison with Retinexformer and MambaLLIE

(1) Retinexformer and MambaLLIE are designed for image only; they revise the original Retinex model by adding perturbation terms or incorporating state space module. They did not adopt decomposition because when the inputs are *single-modality*, and directly improving the illumination of the image is more effective. (2) But our method processes *dual modalities*. Since temporal-mapping events from the event branch provide high-quality illumination information, the most direct and effective way is to decompose Reflectance from the image branch. The inherent characteristics of each modality require extracting different types of information, with the image branch specifically involving a *decomposition process*. (3) Experimental proof: Following your suggestion, we integrated our T2I module into Retinexformer and MambaLLIE and retrained them. As shown in Tab. 6, the "+T2I" variants outperform their baselines due to the high-quality illumination provided by our T2I module. However, they still underperform compared to RETINEV, as they are not designed to handle two modalities. Image-based illumination may interfere with T2I features without proper decomposition.

5.3.2. Equipping Other Methods with Our Temporal-Mapping Events

We clarify that using temporal-mapping events instead of motion events for LLIE is a key contribution of our work. We conducted experiments with synthesized motion events (*from GT images*), shown in Tab.6, rows 1 and 2. We also fed temporal-mapping events into ELIE and EvLight, replacing their original motion events. These variants, denoted as ELIE (TME) and EvLight (TME) in Tab.6, outperform their motion-event counterparts, but still lag behind

Table 6. Extended experimental results on synthetic datasets, read as Tab.3

Method	Motion Events	Temporal Mapping Events	LOL-v1 PSNR	LOL-v1 SSIM	LOL-v2-real PSNR	LOL-v2-real SSIM	LOL-v2-syn PSNR	LOL-v2-syn SSIM	SDSD-in PSNR	SDSD-in SSIM	SDSD-out PSNR	SDSD-out SSIM	#Params (M)
ELIE [TMM2023]	✔	✘	23.52	0.852	21.39	0.861	23.97	0.933	27.46	0.879	23.29	0.742	204.95
EvLight [CVPR2024]	✔	✘	24.61	0.867	23.51	0.875	25.12	0.934	28.52	0.913	26.67	0.836	22.73
ELIE (TME) [TMM2023]	✘	✔	26.63	0.863	27.31	0.907	28.38	0.941	31.48	0.952	31.19	0.949	204.95
EvLight (TME) [CVPR2024]	✘	✔	26.93	0.870	28.19	0.915	29.98	0.943	31.53	0.953	31.56	0.942	22.73
RetinexFormer (+T2I) [ICCV2023]	✘	✔	27.47	0.864	29.69	0.921	31.45	0.942	33.06	0.957	32.68	0.955	27.33
MambaLLIE (+T2I) [NeurIPS2024]	✘	✔	27.54	0.866	29.75	0.923	31.61	0.946	32.78	0.948	32.36	0.954	3.38
RETINEV (Ours)	✘	✔	**28.60**	**0.877**	**30.32**	**0.929**	**32.06**	**0.951**	**33.65**	**0.960**	**33.29**	**0.958**	3.44

our method. This is because their models are designed to *align* event edges with images, while our RETINEV estimates illumination from temporal-mapping events and uses them to further enhance the Reflectance component.

6. More Details on EvLowLight Dataset

Beam-splitter Setup. Figure 11 shows our shared-lens beam-splitter setup from different angles. We use the Nikon F-mount, with a flange focal distance of 46.5 mm. Then the light beam is divided into two with equal energy. In this way, both DVS and image sensor share the same view robustly. The Nikkor 50mm f/1.8D lens was specially modified with a leaf shutter to adjust the transmission of the optical system, enabling the generation of temporal-mapping events. Prophesee EVK4 (1280×720) and MindVision MV-SUA134GC (1280×1024) serve as DVS and image sensor, respectively. The entire system is mounted within a rigid plastic casing, connected at the base to a rotary stage and secured to an optical platform.

Dataset Collection. We utilize the mechanical shutter to generate temporal-mapping events for our RETINEV. In contrast, to produce motion events required by EvLight [9], we employ a rotary stage to adjust the camera's yaw. It is important to emphasize that this setup is *exclusively* for EvLight, as RETINEV *does not* rely on such motion events.

HDR Image Synthesize. In each scene, we capture three low-light images and three single-exposure low dynamic range (ldr) images. The ldr images are combined using the exposure fusion strategy from [10] to generate high dynamic range (HDR) images for reference. Examples from our dataset are illustrated in Figure 12. It is important to note that these HDR images serve solely as visual references and are not considered the "true" ground truth.

Why Normal-Light Image Is Not the "True" Ground-Truth? The intrinsic differences between event cameras and RGB cameras result in their distinct dynamic ranges. Although the HDR images in our dataset are generated using three ldr images, the dynamic range of the event camera remains higher than that of RGB images. As illustrated in Figure 12, backlit areas are still excessively dark, while well-lit regions suffer from overexposure (*e.g.*, the doll's face in the fourth example). Consequently, the PSNR and SSIM values reported in Tab. 3 of the main manuscript are less reliable compared to those on synthetic datasets and should be considered for reference only.

7. Experiments on Real-World Datasets

7.1. Downstream Computer Vision Tasks

The goal of LLIE is to improve the visibility for both human perception and downstream computer vision tasks. Hence, we qualitatively evaluate two representative tasks—semantic segmentation and object detection—on the enhanced results from our method and other approaches. Figure 13 presents the semantic segmentation results using the Segment Anything Model [8], while Figure 14 illustrates object detection outcomes using YOLO v3 [12].

For semantic segmentation, as illustrated in Figure 13, various LLIE methods markedly improve the performance on initial low-light images. Among them, segmentation maps derived from our enhanced results exhibit the highest accuracy and the most detailed edges.

Regarding object detection, as depicted in Figure 14, RetinexNet and EvLight generate incorrect predictions (*e.g.*, misidentifying a bench), while all other methods fail to detect any objects. In contrast, our method accurately identifies both the sculpture and the clock.

These qualitative examples underscore how RETINEV significantly improves the performance of downstream computer vision tasks.

7.2. More Qualitative Results

Figure 15 shows more results on our dataset. Results from our method show better visibility in the dark area. Moreover, the dynamic ranges of our result images are also higher than the dynamic range of other results, *e.g.*, the textures of the statue are clearly visible in both the illuminated and shadowed areas (fourth example image), and the letters on the bag is also clear (fifth example image).

8. Limitations and Future Works

Our RETINEV leverages a transmittance-modulating device to generate temporal-mapping events. However, for cameras equipped with digital shutters, such as those in

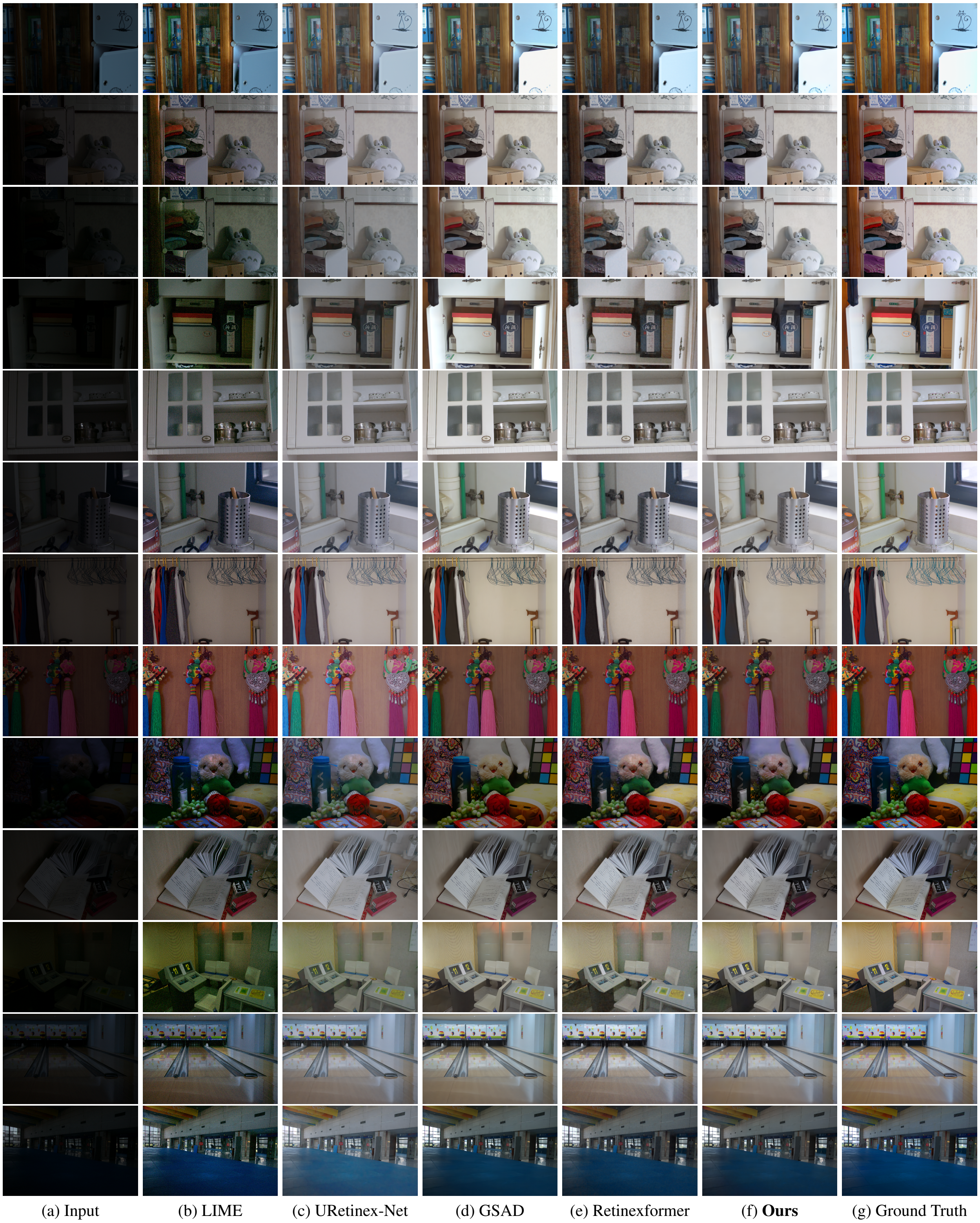
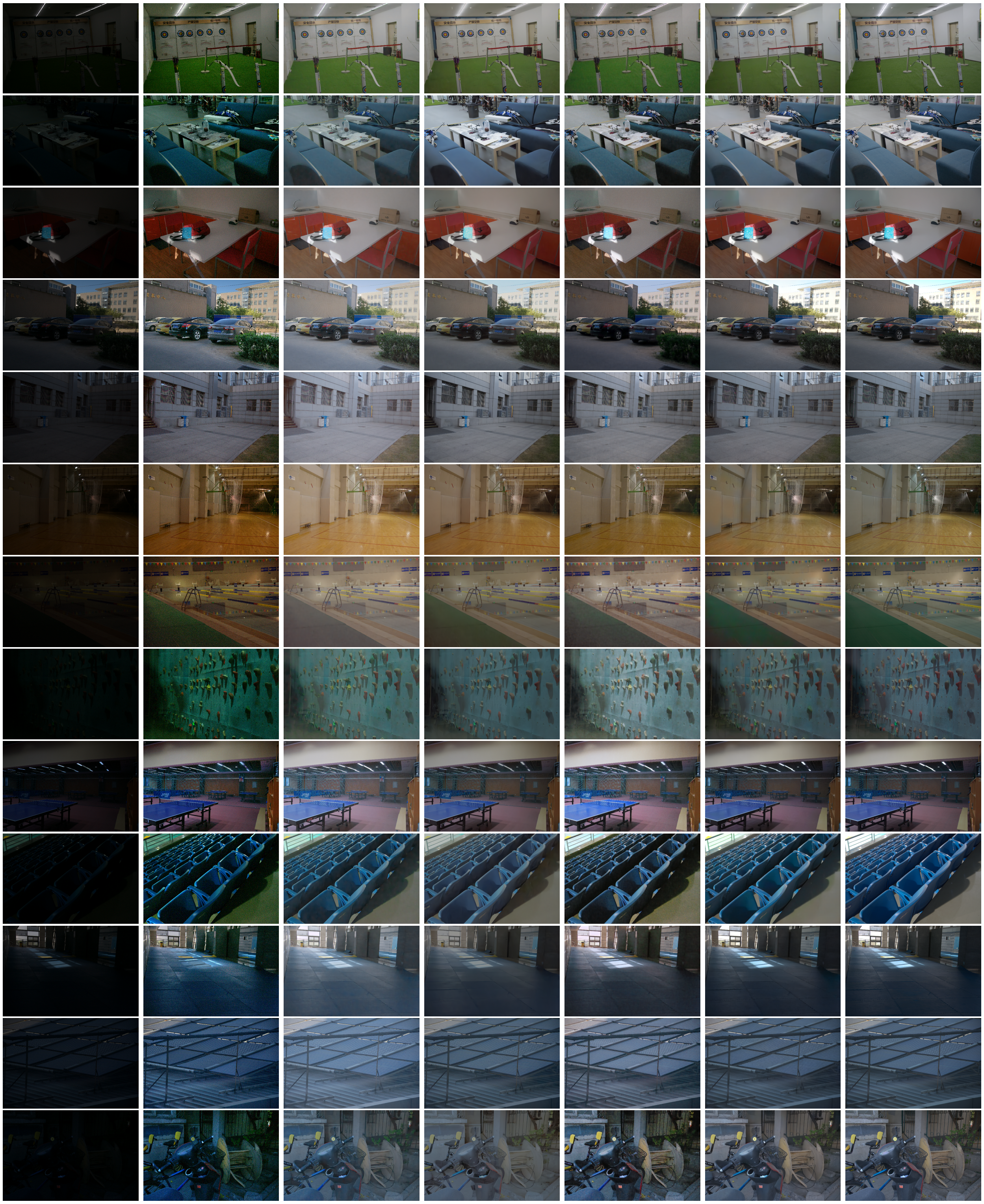
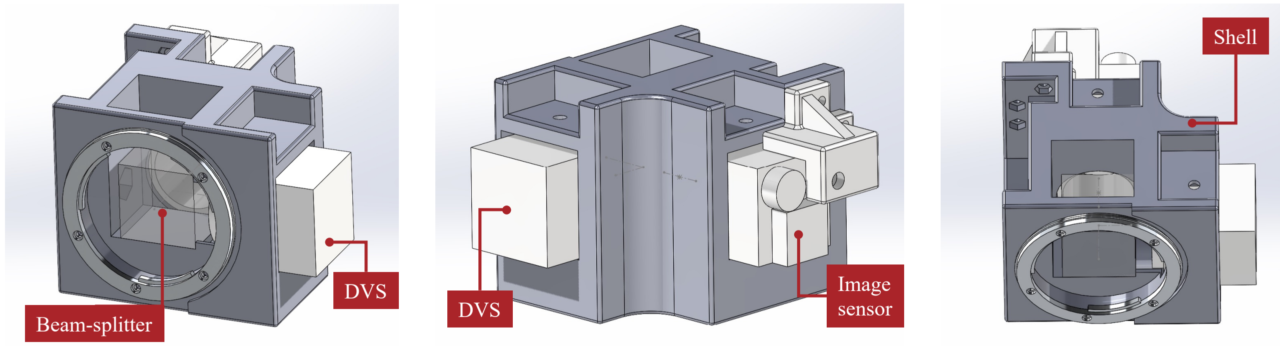


Figure 9. **More Visual comparison with state-of-the-art methods on LOL v1 dataset**. Results from RETINEV show the best visibility. Best viewed on a screen and zoomed in.



(a) Input (b) LIME (c) URetinex-Net (d) GSAD (e) Retinexformer (f) **Ours** (g) Ground Truth

Figure 10. **More Visual comparison with state-of-the-art methods on LOL v2 Real dataset**. Results from RETINEV show the best visibility. Best viewed on a screen and zoomed in.



(a) 30-degree front view (b) 45-degree rear view (c) Front view

Figure 11. The 3D model of our shared-lens beam-splitter setup.


(a) low^{-} (b) low (c) low^{+} (d) ldr^{-} (e) ldr (f) ldr^{+} (g) HDR

Figure 12. Example images from EvLowLight datasets.

smartphones, producing temporal-mapping events is currently challenging. We believe this limitation could be addressed by integrating a charge-clearing module into the CMOS circuit, which would require adjustments to the CMOS peripheral circuitry. Additionally, as with all LLIE methods, color drifting is inevitable to some extent due to the inherent characteristics of the sensor. Compared to other approaches, our RETINEV effectively mitigates color drifting, achieving more stable color reproduction.

Additionally, while DVS exhibits exceptional low-light performance, it has a lower operational limit of approximately 0.01 lux, as observed in our experiments. Below this threshold, no events are generated. We are confident that future advancements in DVS technology will further enhance its capabilities.

References

[1] Yuhan Bao, Lei Sun, Yuqin Ma, and Kaiwei Wang. Temporal-mapping photography for event cameras. *arXiv preprint arXiv:2403.06443*, 2024. 1

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 2, 7, 8, 9

[3] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 2, 7, 8, 9

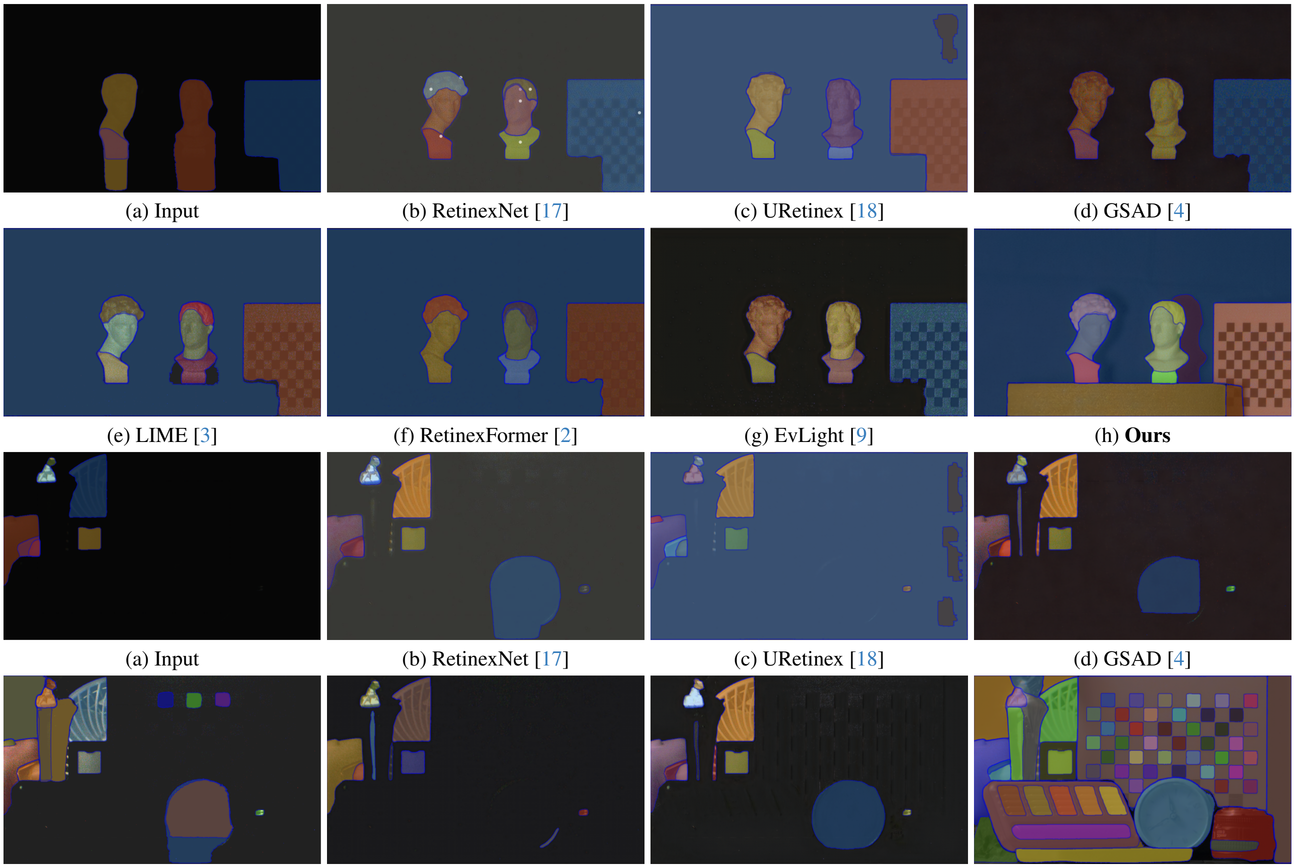
[4] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 7, 8, 9

[5] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proc. CVPR*, pages 1312–1321, 2021. 2

[6] Yu Jiang, Yuehang Wang, Siqi Li, Yongji Zhang, Minghao Zhao, and Yue Gao. Event-based low-illumination image enhancement. *IEEE Transactions on Multimedia*, 2023. 9

[7] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *Proc. CVPR*, pages 3320–3329, 2020. 1

[8] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer White-



(a) Input (b) RetinexNet [17] (c) URetinex [18] (d) GSAD [4]

(e) LIME [3] (f) RetinexFormer [2] (g) EvLight [9] (h) **Ours**

(a) Input (b) RetinexNet [17] (c) URetinex [18] (d) GSAD [4]

(e) LIME [3] (f) RetinexFormer [2] (g) EvLight [9] (h) **Ours**

Figure 13. **Segmentation maps generated using the Segment Anything Model [8] with results from different methods as inputs.** Our method produces the most fine-grained segmentation results.

head, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 3, 7

[9] Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, and Lin Wang. Towards robust event-guided low-light image enhancement: A large-scale real-world event-image dataset and novel approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23–33, 2024. 3, 7, 8, 9

[10] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390. IEEE, 2007. 3

[11] 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 *Proc. CVPR*, pages 6820–6829, 2019. 1

[12] Joseph Redmon. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 3, 8

[13] 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 *Proc. ECCV*, pages 412–428. Springer, 2022. 1

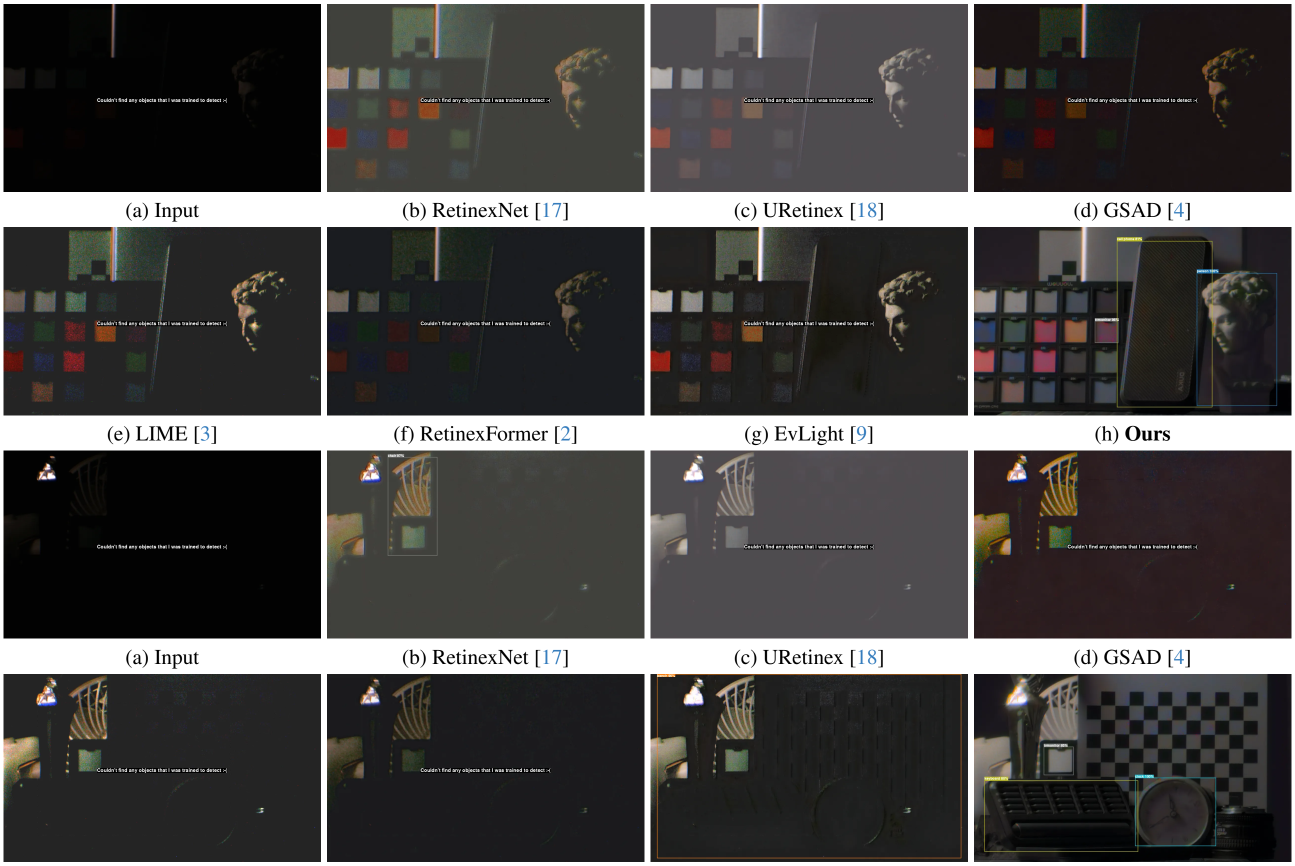
[14] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *Proc. CVPR*, pages 16155–16164, 2021. 1

[15] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In *Proc. CVPR*, pages 17755–17764, 2022. 1

[16] 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. 2

[17] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018. 2, 7, 8, 9

[18] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2022. 2, 7, 8, 9



(a) Input (b) RetinexNet [17] (c) URetinex [18] (d) GSAD [4]

(e) LIME [3] (f) RetinexFormer [2] (g) EvLight [9] (h) **Ours**

(a) Input (b) RetinexNet [17] (c) URetinex [18] (d) GSAD [4]

(e) LIME [3] (f) RetinexFormer [2] (g) EvLight [9] (h) **Ours**

Figure 14. **Object detection results generated using YOLO v3 [12] with inputs from different methods.** Our method achieves the most accurate detection results.

[19] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. 2

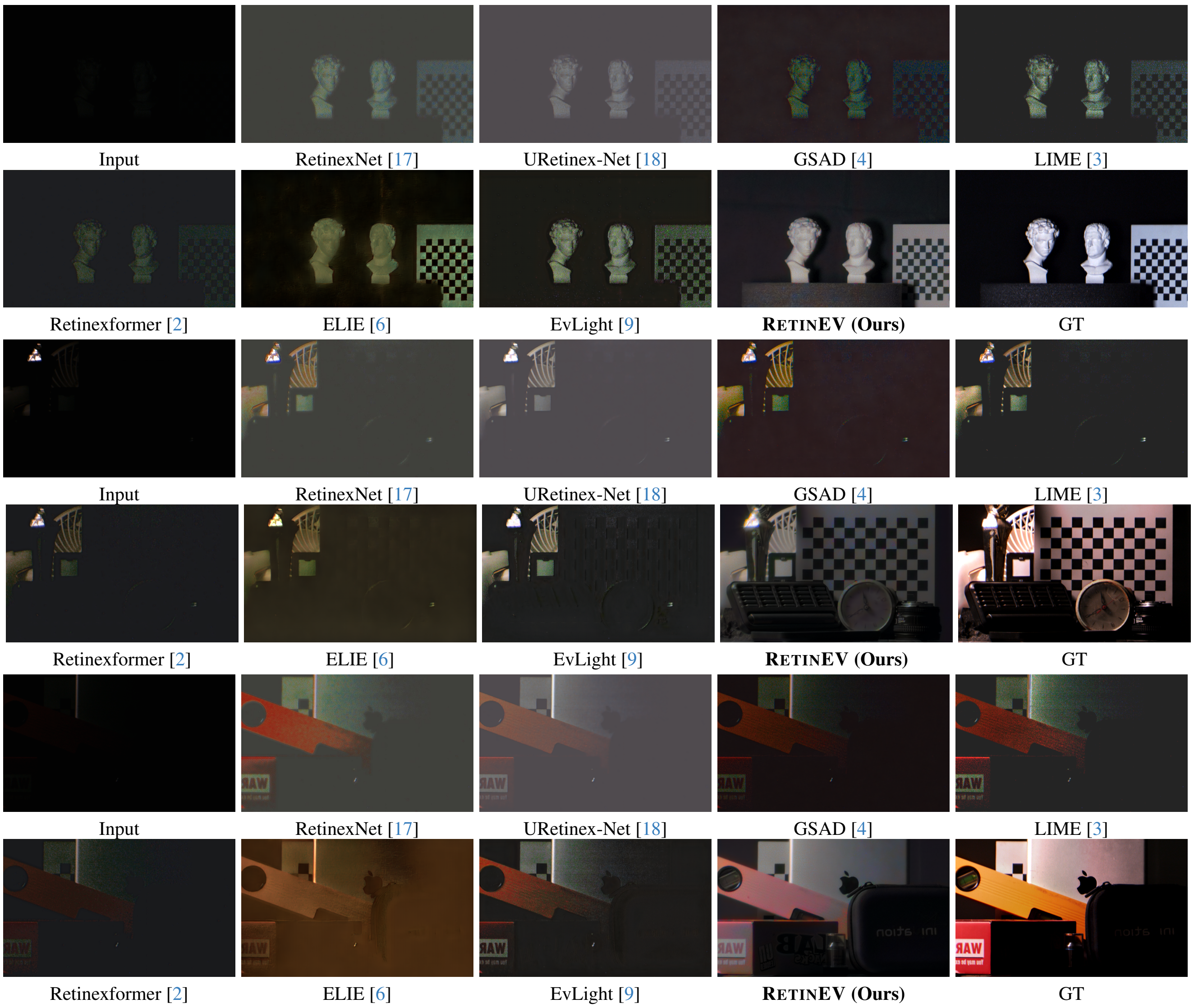


Figure 15. **Visual comparison with state-of-the-art methods on EvLowLight**. Best viewed on a screen and zoomed in.