This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Low-Light Image Enhancement Using Event-Based Illumination Estimation

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

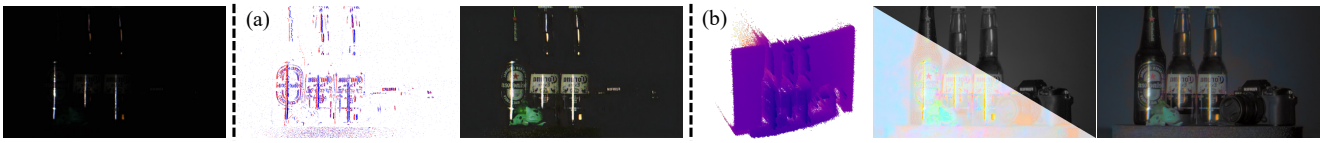


Figure 1. Comparison between (a) event-based LLIE methods using motion events [21] and (b) our RETINEV using temporal mapping events. Best viewed when zoomed in.

Abstract

Low-light image enhancement (LLIE) aims to improve the visibility of images captured in poorly lit environments. Prevalent event-based solutions primarily utilize events triggered by motion, i.e.*, "motion events" to strengthen only the edge texture, while leaving the high dynamic range and excellent low-light responsiveness of event cameras largely unexplored. This paper instead opens a new avenue from the perspective of estimating the illumination using "temporal-mapping" events,* i.e.*, by converting the timestamps of events triggered by a transmittance modulation into brightness values. The resulting fine-grained illumination cues facilitate a more effective decomposition and enhancement of the reflectance component in low-light images through the proposed Illumination-aided Reflectance Enhancement module. Furthermore, the degradation model of temporal-mapping events under low-light conditions is investigated for realistic training data synthesis. To address the lack of datasets under this regime, we construct a beam-splitter setup and collect EvLowLight dataset that includes images, temporal-mapping events, and motion events. Experiments across 5 synthetic datasets and our real-world EvLowLight dataset substantiate that the devised pipeline, dubbed* RETINEV*, excels in producing well-illuminated, high dynamic range images, outperforming previous state-of-the-art event-based methods by up to 6.62 dB, while maintaining an efficient inference speed of 35.6 frames-per-second on a* 640×480 *image. Codes and datasets:* https://github.com/AHupuJR/RetinEV.

1. Introduction

Images captured under insufficient lighting suffer from reduced visibility, resulting in an unsatisfactory visual experience and a decline in performance of computer vision systems optimized for well-lit images [20]. Addressing this issue, low-light image enhancement (LLIE) has become a pivotal yet challenging task [4, 8, 10, 45, 47]. The primary objective is to enhance visibility, Signal-to-Noise Ratio (SNR), and contrast for images taken under suboptimal illumination, while mitigating various distortions such as noise, artifacts, and color inaccuracies. Beyond image sensors, researchers have explored leveraging additional modalities to enhance outcomes further. In this paper, we aim to utilize event cameras for the same.

Event cameras, also known as Dynamic Vision Sensors (DVSs), are neuromorphic sensors that respond to light changes [3, 7, 27, 28, 33, 54, 59], capturing "*event*" information, at high temporal resolution (in the order of μs). Different from conventional frame-based cameras, the asynchronous architecture of event cameras endows them with high dynamic range (HDR) and exceptional low-light responsiveness. Events are triggered if a change occurs in the scene, even under insufficient illumination, providing clues for LLIE. Prevalent event-based methods [14, 21] focus on fusing low-light images with events, in which the information derived from *motion events* plays a crucial role in enhancing low-light images. We will refer to these approaches

†Corresponding authors.

as *motion-based* methods.

However, motion-based methods suffer from several limitations: **First**, in low-light environments, where longer exposure times are generally required, motion-based methods still rely on camera motion to generate events, *resulting in degradations combining low-light effects and motion blur*. **Second**, events in motion-based methods only deliver edge information, which is *insufficient for LLIE tasks that require fine-grained illumination estimation at both edges and within regions*. **Third**, these methods are also prone to generating *artifacts*, as shown in Fig. 1 (a). Hence, the full potential of event cameras in practical low-light image enhancement remains to be explored.

Recent advance in the event-to-image conversion task [2] produces impressive high-quality images from only events, which can be ascribed to triggering events with transmittance modulation instead of motion in the scene. This motivates us to rethink the current *de facto* paradigm in event-based LLIE: Can we utilize events triggered by transmittance modulation, *i.e*., the temporal-mapping events for LLIE? Intrinsically, illumination[1] Information encoded in temporal-mapping events is expected to offer higher density and illumination precision than motion events.

With this insight, we propose a *fast* technique, dubbed RETINEV, that does not need scene motion to enhance low-light RGB images using event cameras. We capture temporal-mapping events of the scene by changing the transmittance of the optical system. These 'temporal mapping events' refer to the brightness changes that are induced when the transmittance changes (opening a shutter in our case), and are harnessed to enhance the low-light RGB image by exploiting the Retinex theory [19]. This theory posits that an observed image results from the product of reflectance and illumination, where reflectance represents an intrinsic characteristic that remains constant under varying illumination conditions. Illumination component is estimated from temporal mapping events, while also aiding in the decomposition of the reflectance component, as shown in Fig. 1 (b). A coefficient is designed for arbitrary illumination manipulation in practical use, and the characteristics of events under low light are also considered in the degradation model for the model training on synthetic data. Unlike previous Retinex-based methods that focus solely on enhancing the illumination component [4, 47, 48, 55], our approach refines the reflectance component leveraging high-quality illumination components as prior information through a cross-modal attention mechanism.

To instantiate our RETINEV pipeline, we build a beam-splitter system consisting of a DVS and an RGB image sensor to capture the aligned events and low-light images. During data capture, the opening process of the mechanical shutter, which exposes the image sensor, also produces temporal-mapping events. This allows for the synchronization of image exposure and the capture of temporal-mapping events, completed as fast as 2 ms. With this setup, we construct a real-world dataset named EvLowLight, comprising data captured under 60 extremely low illumination and high contrast scenes. The proposed method shows consistently superiority over state-of-the-art (SOTA) LLIE methods on five synthetic datasets and our real-world EvLowLight dataset, across objective reference and no-reference metrics, as well as subjective evaluations from user studies. The simple yet effective architecture also endows RETINEV with a real-time inference speed of 35.6 frames-per-second on a 640×480 image. The impressive results shed light on the promises of event cameras in LLIE.

In a nutshell, we make the following main contributions:

- Based on Retinex theory, we propose a new and more effective paradigm named RETINEV for practical event-based low-light image enhancement, where low-light event characteristics are considered and resulted brightness can be controlled.
- An Illumination-aid Reflectance Enhancement module is designed, where the illumination component is introduced to enhance reflectance with cross-modal attention.
- We construct a beam-splitter system and present EvLowLight, a real-world dataset consisting of aligned events and images under challenging illumination conditions, which provides an evaluation setting for LLIE methods.
- Quantitative and qualitative experiments show that RETINEV outperforms image-only methods as well as earlier SOTA event-based methods with real-time inference speed.

2. Related Work

Frame-Only Low-Light Image Enhancement. Conventional LLIE methods fall into two main categories: histogram equalization [1, 31] and Retinex-based approaches [8, 15, 19]. Histogram equalization enhances contrast by adjusting intensity distributions [34], while Retinex-based methods decompose images into illumination and reflectance, enabling illumination correction at different scales [15, 19]. However, these methods struggle with real-world variations and noise. Deep learning methods have significantly advanced LLIE. RetinexNet [47] and its variants [48, 53] use CNNs to decompose images and enhance illumination maps. Other approaches leverage illumination estimation [45], semantic priors [49], and signal-to-noise ratio modeling [50], *etc*. Despite their effectiveness, CNN-based methods [45, 49–51, 61] often suffer from limited generalization under varying lighting conditions. Recent transformer-based models [22, 26, 58] have demonstrated improved performance on benchmark

[1]The CMOS sensor of a camera actually measures irradiance rather than illuminance, and in this paper, any mention of "illuminance" specifically refers to irradiance.

datasets. Meanwhile, generative models such as EnlightenGAN [13], Diff-Retinex [55], GSAD [10], and DiffLL [12] leverage GANs and diffusion models to enhance perceptual quality. However, these methods often introduce increased computational complexity.

Event-Based Image Enhancement employs a hybrid sensor model, where event data aids image quality improvement, as in our method. This approach has been explored in various tasks, including image deblurring [5, 17, 24, 32, 35, 36, 38, 39, 44, 60, 62], HDR imaging [30], video frame interpolation [37, 40, 41], and real-time photometric stereo [56], *etc*. These systems share a common focus on motion-related image enhancement. The HDR and low-light responsiveness of the event camera have inspired its use in low-light image and video enhancement. Jiang *et al*. [14] and Liang *et al*. [21] utilize motion-free low-light images and events near exposure times for enhancement. However, prolonged exposure in low-light conditions can blur motion-triggered events, and events primarily capture high-contrast edges without grayscale information, limiting their utility. Recent works [23, 25] model video coherence using events, facilitating frame alignment and fusion. However, these methods primarily build upon existing image enhancement techniques by incorporating multi-frame alignment. As a result, the potential of event cameras in low-light image enhancement remains largely unexplored.

3. RETINEV

We first establish the mathematical relationship between illumination and the timestamps of temporal-mapping events in §3.1. Leveraging the estimated illumination from events, we incorporate Retinex theory [19] to construct the architecture of RETINEV in §3.2. We then detail the T2I and IRE modules in §3.3 and §3.4, respectively. Finally, loss functions are designed in §3.5.

3.1. Problem Formulation

In event cameras, an event, typically represented as a tuple $e = (x, y, t, p)$, is triggered once the change in light intensity $\mathcal{I}$ exceeds a predefined threshold in the log domain:

$$p = \begin{cases} +1, \text{if} \log\left(\frac{\mathcal{I}_t(x,y)}{\mathcal{I}_{t-\Delta t}(x,y)}\right) > c, \\ -1, \text{if} \log\left(\frac{\mathcal{I}_t(x,y)}{\mathcal{I}_{t-\Delta t}(x,y)}\right) < -c, \end{cases} \tag{1}$$

where x, y, p, t, c represent the coordinates, polarity, timestamp, and the contrast threshold of the event, resp.

Previous work [2] actively changes the transmittance of the optical system to modify the illuminance received by the sensor. It then uses the timestamp of the first positive event (FPE) generated at each pixel location to estimate the local illumination. In our case, we set the transmittance function to be a step function. The FPE of a pixel is triggered when the energy of the photoelectric conversion at the pixel reaches the threshold voltage, at which point the energy stored in the capacitor is reached, that is,

$$\eta \cdot E \cdot A \cdot t_{fpe} = \frac{C \cdot U_{thd}^2}{2}, \tag{2}$$

where η, E, A, t_{fpe}, C, and U_{thd} represent photoelectric conversion efficiency, illuminance, the photosensitive area of the pixel, timestamp of FPE, pixel capacitance, and pixel threshold voltage, resp.

From Eq. (2), the relationship between the illuminance E, the physical quantity to be measured, and the time t_{fpe} can be expressed as a conversion function $f(\cdot)$:

$$E = \frac{C \cdot U_{thd}^2}{2\eta \cdot A \cdot t_{fpe}} = \frac{k}{t_{fpe}} = f(t_{fpe}). \tag{3}$$

Due to various complex factors, the illuminance E in real-world scenarios *cannot* be precisely determined using Eq. (3). However, we can leverage the general relationship between E and t_{fpe} to construct our model, which will be introduced in the following section.

3.2. General Architecture

According to the classic Retinex theory [19], an observed image S can be decomposed into two components: reflectance component R and illumination component I.

$$S = R \cdot I. \tag{4}$$

I is determined by the lighting conditions. Prevalent image-only Retinex-based methods [4, 47, 48] estimate the I_{low} and R_{low} from low-light images and enhance I. However, sensor response degradation in low light leads to suboptimal I_{low} estimation, with *defects further amplified* during illumination enhancement for $\hat{I}_{low}$.

In RETINEV (shown in Fig. 2), temporal-mapping events with *better low-light responsiveness* are harnessed for the estimation and enhancement of I_{low}. Note that different from well-defined physics, the Retinex theory is based on human perception. As a consequence, there is no precise conversion between illumination I (subjective measure) and illuminance E (objective physical quantity, derived from t_{fpe} in Eq. (3)). Hence, we design a Time-to-Illumination (T2I) module for the conversion and enhancement of illumination:

$$\hat{I} = \mathcal{F}_{T2I}(t_{fpe}(x, y); \Theta_1), \quad \forall (x, y) \in \Omega. \tag{5}$$

Here, Θ_1, as well as Θ_2 and Θ_3 mentioned later in this section, represent the learnable parameters of the corresponding mappings. Ω represents the spatial domain of the image sensor. The T2I module consists of the conversion function $f(\cdot)$, a DenoiseNet, Multilayer Perceptron (MLP) layers for non-linear mapping, and Gamma encoding [6] to ensure alignment with RGB images, as shown in Fig 2.

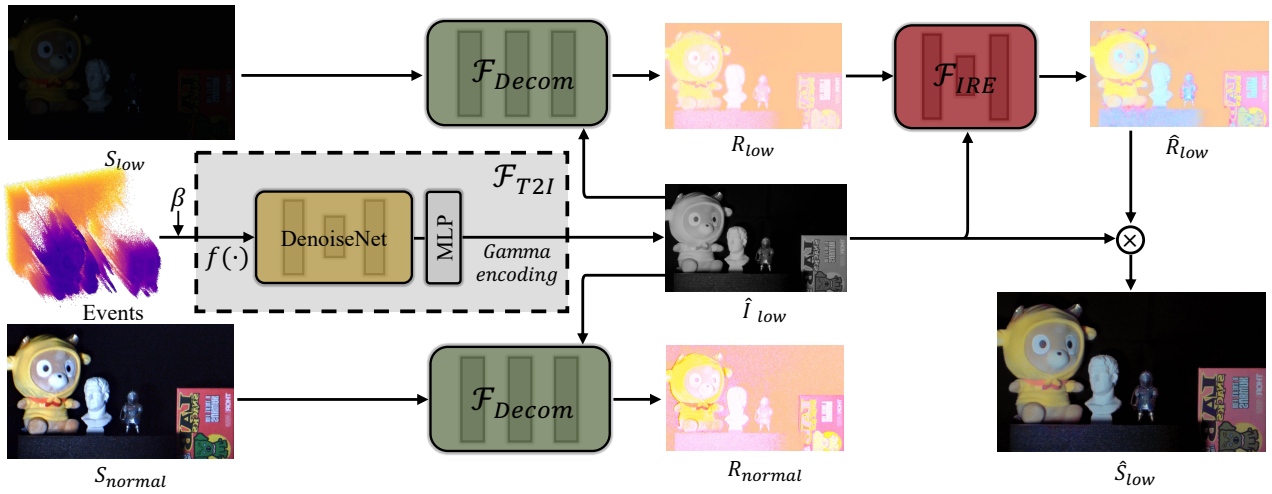


Figure 2. **The architecture of our RETINEV.** $\mathcal{F}_{Decom}$ for low-light images and normal-light images share the same weights. β is added for brightness manipulation. "IRE": Illumination-aid Reflectance Enhancement, "S_{low}": low-light image, "S_{normal}": normal-light image, "I": illumination component, "R": Reflectance component, "$\hat{S}_{low}$": the predicted enlightened image.

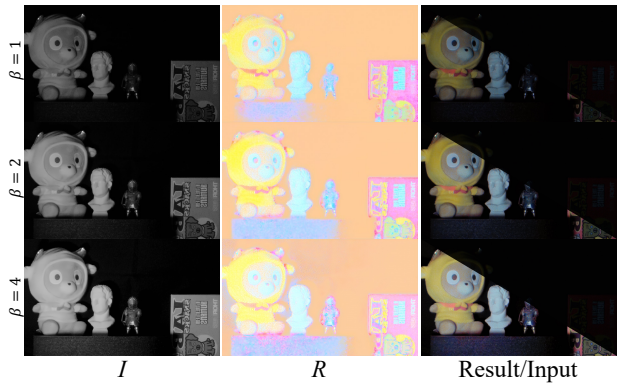


Figure 3. **Visualization of the effect of β on the brightness of I and the result**.

Regarding the reflectance component R, classical Retinex theory asserts that R is an intrinsic property of the captured object and remains invariant under varying illumination conditions [19, 47]. However, in practice, similar to the illumination component I extracted from RGB camera under low-light conditions, the reflectance component R also *suffers severe degradation*. Prior studies have demonstrated that accurately estimating I significantly facilitates the determination of R [4, 48]. To this end, we introduce the well-estimated illumination $\hat{I}$ from *events* into the decomposition process:

$$R_{low} = \mathcal{F}_{Decom}(S_{low}, \hat{I}; \Theta_2), \tag{6}$$

and subsequently enhance R_{low} as follows:

$$\hat{R}_{low} = \mathcal{F}_{IRE}(R_{low}, \hat{I}; \Theta_3). \tag{7}$$

Here, the Illumination-aided Reflectance Enhancement (IRE) module $\mathcal{F}_{IRE}$ captures long-range dependencies between the enhanced reflectance $\hat{R}_{low}$ and the illumination estimate $\hat{I}$, as detailed further in Section 3.4.

Finally, the predicted $\hat{S}_{low}$ that approximates normal-light image S_{normal} is derived by:

$$\hat{S}_{low} = \hat{I} \cdot \hat{R}_{low}. \tag{8}$$

The general architecture of RETINEV is shown in Fig. 2.

3.3. Time-to-Illumination Module

To leverage the illumination prior, the Time-to-Illumination (T2I) module is designed to convert temporal events to illumination for normal-light image synthesis, as shown in Fig. 2. Though event cameras exhibit superior low-light responsiveness, their performance still degrades under low-light conditions.

Low-Light Degradation Model (LLDM). Different from the temporal-mapping events in EvTemMap [2], which are captured under normal-light conditions, events in low-light environments exhibit distinct characteristics [11], mainly including: (*i*) longer latency and (*ii*) increased dark current noise. We design and formulate LLDM to make synthetic data in the training process more closely approximate real-world data. In the training phase, (1) the GT images are first applied with degradation on the intensity of pixels, *i.e.*, *spatial domain*, including: *blurriness* to synthesize the diffraction effects, *downsampling* for the sensor limitation, and *poisson-Gaussian* hybrid noise for the low-light dark current noise. Then, the images are converted to the *temporal domain*, introduced with *timestamp latency*, *dead pixel*, and *random contrast threshold* degradation. Latency and

the probability of dead pixels are directly proportional to the timestamp, *i.e.*, a larger timestamp corresponds to higher latency and an increased probability of dead pixels.

All the degradation is randomly sampled within a range, and randomly shuffled in each domain to cover real-world conditions. For more details kindly *cf*. §5.2 and §2 in supp.

Illumination Manipulation. The preferred illumination intensities can vary significantly across individuals and applications. To make RETINEV more practical, an illumination manipulation coefficient β is introduced in the event-to-intensity conversion to adjust the illumination strengths:

$$t_{norm} = \frac{t_{fpe} + \beta}{max(t_{fpe}) + \beta}. \tag{9}$$

With t_{norm}, the value of E is obtained using Eq. (3). By choosing different β, we can adjust the intensity of I as well as the resulting images, as shown in Fig. 3. As the change of I, R should be invariant, which is also indicated in the visualized R.

3.4. Illumination-aid Reflectance Enhancement

Image-only Retinex-based LLIE methods [4, 47, 48] focus mainly on the enhancement of illumination. For our method, with the high-quality illumination from temporal-mapping events, we improve the reflectance from the limited image sensor in IRE module. After layer normalization and 1×1 convolution, queries ($\mathbf{Q}$), keys ($\mathbf{K}$) and values ($\mathbf{V}$) are derived. In our case, the $\mathbf{Q}$ is from R component while $\mathbf{K}$ and $\mathbf{V}$ are from I (denoted as $\mathbf{Q}_R$, $\mathbf{K}_I$, and $\mathbf{V}_I$, resp.). Because R and I are the same size as the input image, this results in a substantial computational burden. Hence, different from vanilla attention mechanism [42], the $\mathbf{Q}_R$ is transposed before multi-head attention:

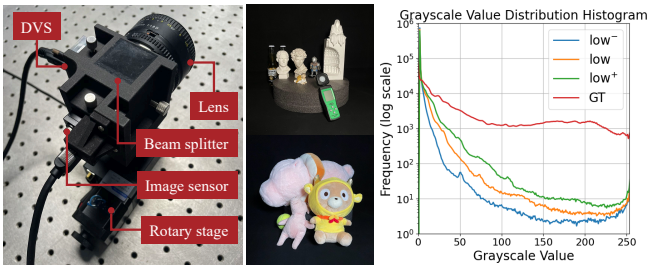
$$\text{Attention}(\mathbf{Q}_R, \mathbf{K}_I, \mathbf{V}_I) = \mathbf{V}_I \text{softmax}\left(\frac{\mathbf{Q}_R^{\intercal}\mathbf{K}_I}{\sqrt{d_k}}\right), \tag{10}$$

where d_k denotes the dimensionality of the key (and query) vectors in each attention head. The attention map in Eq. (10) is $c \times c$, reducing the spatial complexity from $\mathcal{O}(h^2w^2)$ to $\mathcal{O}(c^2)$, where h, w, and c denotes height, width, and channel of the input feature maps, resp. Long-range dependency is better captured between two modalities while maintaining a low computational cost.

3.5. Loss Functions

In our training stage, both low-light images and normal-light images are forwarded (shown in Fig. 2). The normal-light image should be reconstructed by combining $\hat{I}$ from events and R from both normal-light image S_{normal} and low-light image S_{low}, resulting in a reconstruction loss:

$$\mathcal{L}_{recon} = ||\hat{I} \cdot \hat{R}_{low} - S_{normal}||_1 + ||\hat{I} \cdot R_{normal} - S_{normal}||_1. \tag{11}$$



(a) Capture setup (b) Example scenes (c) Gray value histogram

Figure 4. **Details** about EvLowLight dataset. The beam splitter is positioned between the lens and the sensors.

Table 1. **Comparison** with existing datasets.

	Motion events	Temporal-mapping events	Resolution
LOL v1 [BMVC2018] [47]	✘	✘	600 × 400
SDSD [ICCV2021] [46]	✘	✘	512 × 960
ELIE [TMM2023] [57]	✔	✘	256 × 256
SDE [CVPR2024] [21]	✔	✘	260 × 346
EvLowLight (Ours)	✔	✔	1216 × 720

Following Retinex theory, the reflectance remains constant under varying illuminations, leading to the concept of an invariant reflectance loss:

$$\mathcal{L}_R = ||R_{low} - R_{normal}||_1 + ||\hat{R}_{low} - R_{normal}||_1. \tag{12}$$

Finally, perceptual loss is added to guide the model toward generating outputs that are perceptually closer to the target, capturing finer textures and structural details:

$$\mathcal{L}_{final} = \lambda_1\mathcal{L}_{recon} + \lambda_2\mathcal{L}_R + \lambda_3\mathcal{L}_{percep}, \tag{13}$$

where λ_1, λ_2, and λ_1 are hyper-parameters.

4. EvLowLight Dataset

Our RETINEV establishes a new paradigm for event-based LLIE. Existing datasets [14, 21], relying on motion events, are incompatible due to the absence of temporal-mapping events as shown in Table 1.

Hence, in this work, we construct a share-lens beamsplitter system, as shown in Fig. 4 (a). Different from previous dual-camera setup [40] or beam-splitting-first style beamsplitter configurations [41], the 50R/50T beamsplitter in our setup is placed between the lens and sensors. This positioning provides both sensors with an imaging beam that preserves identical optical characteristics, enabling spatial alignment at any depth using the same homography matrix [9]. For more details, please *cf*. supp.

Using the shared-lens beamsplitter setup, we collected a novel real-world event-based low-light image enhancement dataset, EvLowLight, detailed in Table 2. The dataset includes 60 scenes with illuminance levels ranging from 2.5

Table 2. **Details of EvLowLight dataset.** Image: image sensor, T.M. events: temporal-mapping events, M. events: motion events.

Device	Size	Groups	Illuminance	Image exposure time in each group (ms)						T.M. events	M. events
				low$^-$	low	low$^+$	ldr$^-$	ldr	ldr$^+$		
DVS + Image	1216×720	60	2.5–6 lux	5–8	10–15	15–20	120–180	200–400	400–700	✔	✔

Table 3. **Quantitative comparisons** on LOL [47, 52] (v1, v2-real, and v2-syn), and SDSD [46] (in and out) datasets. Best results in **bold**.

Method	Events	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	#Params (M)
		LOL-v1		LOL-v2-real		LOL-v2-syn		SDSD-in		SDSD-out		
LIME [TIP2016] [8]	✘	16.76	0.440	15.24	0.470	16.88	0.776	-	-	-	-	**0.77**
RetinexNet [BMVC2018] [47]	✘	16.77	0.560	20.06	0.816	22.05	0.905	23.25	0.863	25.28	0.804	0.84
DeepUPE [CVPR2019] [45]	✘	14.38	0.446	13.27	0.452	15.08	0.632	21.70	0.662	21.94	0.698	1.02
KinD [MM2019] [61]	✘	20.86	0.790	14.74	0.641	13.29	0.578	21.95	0.672	21.97	0.654	8.02
EnGAN [TIP2021] [13]	✘	17.48	0.650	18.32	0.617	16.57	0.734	20.02	0.604	20.10	0.616	114.35
Restormer [CVPR2022] [58]	✘	22.43	0.823	19.94	0.827	21.41	0.830	25.67	0.827	24.79	0.802	26.13
SNR-Net [CVPR2022] [50]	✘	24.61	0.842	21.48	0.849	24.14	0.928	29.44	0.894	28.66	0.866	4.01
Retinexformer [ICCV2023] [4]	✘	25.16	0.845	22.80	0.840	25.67	0.930	29.77	0.896	29.84	0.877	1.61
ELIE [TMM2023] [14]	✔	-	-	-	-	-	-	27.46	0.879	23.29	0.742	204.95
Liu *et al.* [AAAI2023] [25]	✔	-	-	-	-	-	-	27.58	0.888	23.51	0.726	-
EvLight [CVPR2024] [21]	✔	-	-	-	-	-	-	28.52	0.913	26.67	0.836	22.73
RETINEV (Ours)	✔	**28.60**	**0.877**	**30.32**	**0.929**	**32.06**	**0.951**	**33.65**	**0.960**	**33.29**	**0.958**	3.44

to 6 lux, with example scenes shown in Fig. 4 (b). For each scene, we captured two sets of images: low-light images and normal-light images. Each set consists of 3 images with short, regular, and long exposure time, representing extreme-underexposure, normal, and slightly brighter low-light/normal-light conditions. Normal-light images with different exposure time are utilized for exposure fusion [29] for producing reference HDR images. Grayscale value distribution is shown in Fig 4 (c). Theoretically, temporal-mapping event acquisition could *seamlessly* integrate with image exposure when the shutter opens. Finally, we generated moving events by rotating the rotary stage, which served as input events for the event-based competitors [21].

5. Experiments

5.1. Datasets

LOL v1 & v2. The images in the LOL v1 [47] and LOL v2 Real [52] datasets have a resolution of 600×400, whereas the images in the LOL v2 Synthetic [52] dataset have a resolution of 384×284. The training and testing sets for LOL v1, LOL v2 Real, and LOL v2 Synthetic datasets are divided in the ratios of 485:15, 689:100, and 900:100, resp.

SDSD. The SDSD dataset [46] comprises both indoor and outdoor subsets. We utilize 70:12 video pairs for training and testing in the indoor subset and 80:13 video pairs for the outdoor subset. All images have a resolution of 512×960.

EvLowLight. 60 groups of data are present in EvLowLight. Each set consists of 3 low-light images, 1 HDR image, temporal-mapping events, and motion events.

5.2. Training Details

The network is trained with a patch size of 128×128. The denoising network in T2I module is pre-trained for 100k iterations with the proposed degradation model and geometric augmentation (random flips and rotations) for data augmentation. Learning rate is set to 2×10^{-4} with Adam [18] optimizer and minimum learning rate of 10^{-7}. The learning rate is the same as pre-training except for the denoising network, which is set to one-tenth of that for the other components. We train the model with a batch size of 32 for 150k iterations on a single NVIDIA L4 GPU. FPS are evaluated with a single NVIDIA A100 GPU with 640×480 images. For more details kindly *cf*. supp.

5.3. Comparisons on Synthetic Datasets

We compare our method with SOTA techniques on five synthetic datasets. PSNR results are computed without using the mean value from ground truth [49, 61]. Motion-based methods, such as ELIE [14], Liu *et al*. [25], and EvLight [21], cannot be applied to the LOL v1 and v2 datasets as they require video sequences to generate motion events.

Quantitative Comparison. The quantitative results are reported in Table 3. Compared to the best existing image-only and event-based methods, our RETINEV achieves 3.44/7.52/6.39/3.88/3.45 dB improvement in PSNR and 0.032/0.089/0.021/0.047/0.081 improvement in SSIM, respectively, with only 3.44 Mb parameters. With events as prior for illumination estimation, our method significantly outperforms image-only methods. As a representative method for event-based methods leveraging motion events for LLIE, EvLight [21] only achieves limited im-

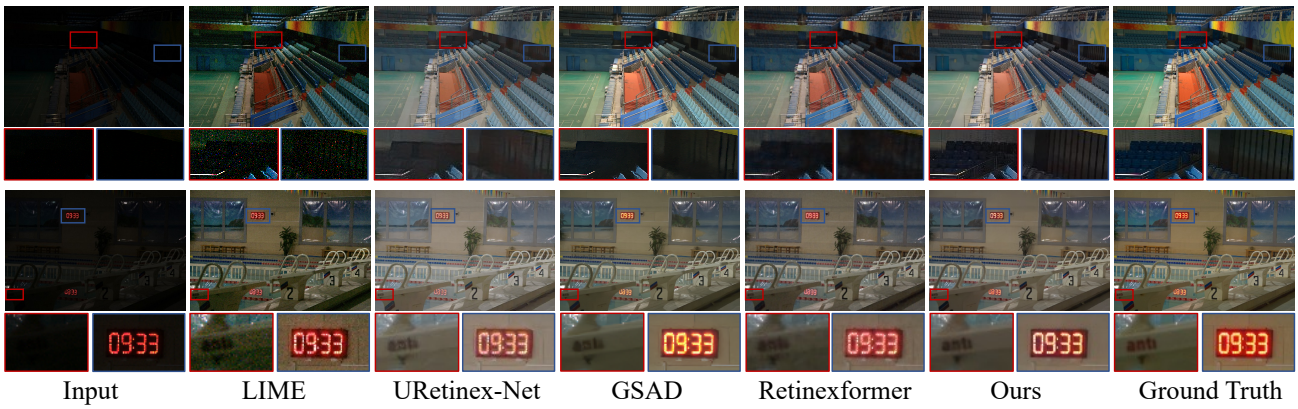


Figure 5. **Visual comparison on LOL-v1 [47] dataset**. Our method effectively enhances visibility and preserves fine-grained textures.

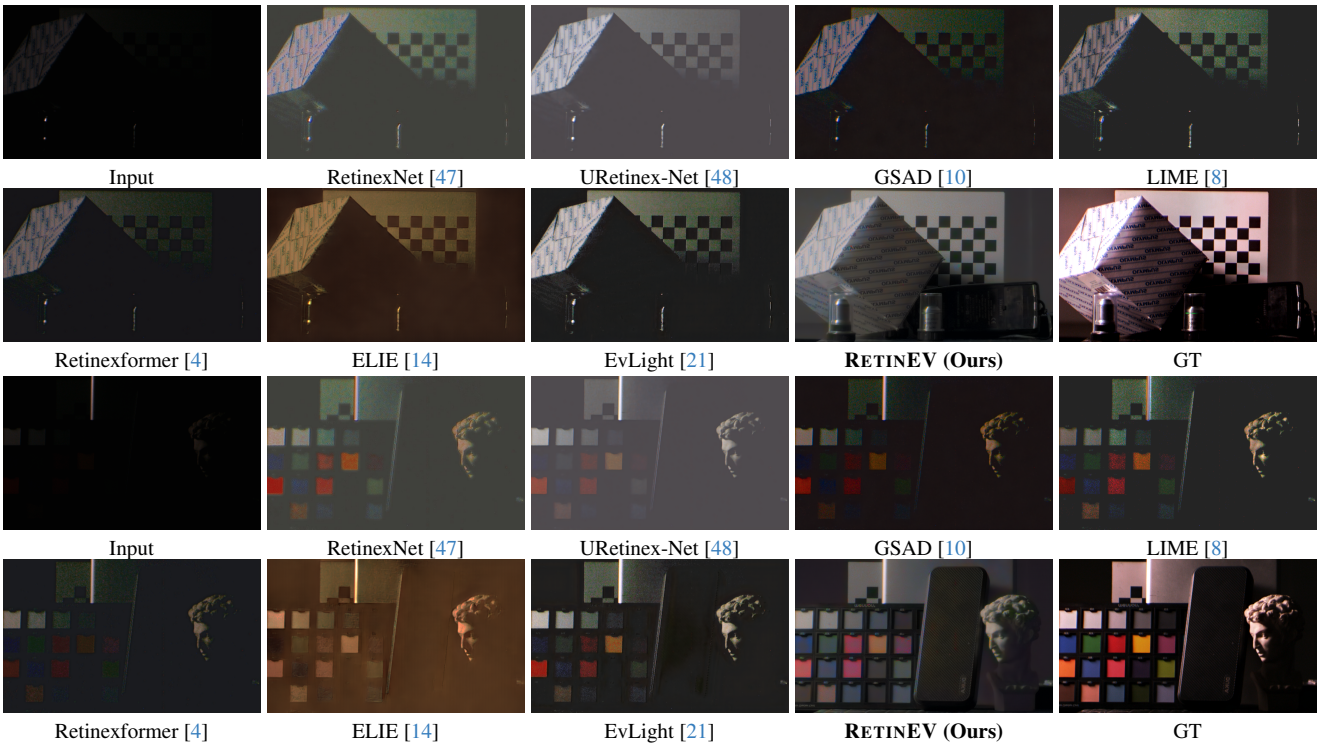


Figure 6. **Visual comparison with SOTA methods on EvLowLight**. RETINEV improves visibility and HDR performance.

provements in SSIM (0.017) due to the edge information from motion events. However, EvLight fails to outperform other image-only methods in other dimensions even with the event as an additional modality. By utilizing temporal-mapping events, our method fully harnesses the potential of event-based information with a SOTA performance among event-based methods.

Visual Comparison. Fig. 5 shows example results from LIME [8], URetinex-Net[48], GSAD [10], Retinexformer [4] and our method in LOL v1 dataset. Our method restores the best details in the poor-illumination area, *e.g.*, seats in the dark in the red box.

5.4. Comparisons on EvLowLight

To assess methods in real-world settings, we evaluate SOTA image-only and event-based approaches, including ELIE [14], EvLight [21], and our method on EvLowLight. Except for EvLight, trained on SDE-in [21], all methods are trained on LOL v1 and directly tested on EvLowLight, demonstrating their generalization to real-world scenes.

Visual Comparison. As shown in Fig. 6. Other than improving the average brightness of the image, image-only methods fail to enhance the contrast in the poor illumination area. Benefiting from motion events, ELIE and EvLight

Table 4. **Quantitative comparison and user study on EvLowLight,** with the best results highlighted in **bold**. PSNR* and SSIM* are calculated with exposure fusion images. Avg ranking: the average ranking from the user study. FPS: Frame-per-second.

Method	Events	PSNR* ↑	SSIM* ↑	PIQE ↓	Avg Ranking ↓	Param (M) ↓	Flops (G)↓	FPS ↑
LIME [TIP2016] [8]	✘	13.31	0.400	41.68	2.46	-	**7.0**	0.77
RetinexNet [BMVC2018] [47]	✘	12.66	0.363	15.31	4.56	0.84	246.4	14.4
URetinex-Net [CVPR2022] [48]	✘	12.29	0.382	45.97	4.11	**0.34**	106.5	**48.1**
Retinexformer [ICCV2023] [4]	✘	11.79	0.403	19.55	5.93	1.61	235.1	3.7
GSAD [NeurIPS2024] [10]	✘	12.20	0.416	31.23	3.69	17.36	942.6	1.6
ELIE [TMM2023] [57]	✔	13.67	0.424	42.73	3.73	204.95	2116.0	0.9
EvLight [CVPR2024] [21]	✔	14.51	0.435	20.45	2.31	22.73	438.5	4.1
RETINEV (Ours)	✔	**15.39**	**0.470**	**9.41**	**1.21**	3.44	184.6	35.6

Table 5. **Ablation study of various components of our method** on LOL v1 dataset [47]. "IRE": Illumination-aid Reflectance Enhancement module, "LLDM": Low-light degradation model. "CrossAttn": Cross-modal attention.

	Events	T2I module	IRE	PSNR	SSIM
1	✘	n/a	n/a	16.98	0.565
2	✔	None	n/a	26.96	0.849
3	✔	*w./o.* LLDM	No fusion	27.25	0.862
4	✔	*w./o.* LLDM	CrossAttn. fusion	28.45	0.872
5	✔	*w.* LLDM	No fusion	27.83	0.865
6	✔	*w.* LLDM	Add fusion	28.13	0.870
7	✔	*w.* LLDM	Concat. fusion	28.01	0.867
8	✔	*w.* LLDM	Multiply fusion	27.98	0.867
9	✔	*w.* LLDM	CrossAttn. fusion	28.60	0.877

improve contrast better than image-only methods. However, it also suffers from edge artifacts caused by the events, *e.g.*, ghost artifacts around the camera in the second example. Our method demonstrates balanced illumination, making details in dark areas, such as the black box and dark side of the sculpture, clearly visible. Even the gray background with uniformly distributed illumination is well-defined. With the estimated illumination from the T2I module, the enhanced reflectance yields *smoother results*. In addition to low-light performance, our results *also demonstrate HDR capability*, benefiting from the inherent advantages of event cameras. Notably, the 'Olympus' logo remains clearly visible in both illuminated and backlit areas.

Quantitative Comparison. Table 4 shows quantitative results. Note that for PSNR and SSIM calculation, the ground truth image is generated from three single-exposure images using exposure fusion based on Mertens *et al*. [29]. Because of the inherent difference in dynamic range and response function between the frame-based camera and the event camera in our setup, the pixel-wise comparison may not be meaningful [9, 16], yet we report PSNR and SSIM (marked with *) as a reference. Our RETINEV outperforms all methods in PSNR, SSIM, and the non-reference metric Perception-based Image Quality Evaluator (PIQE) [43].

User Study. To quantify the human subjective visual perception quality of the enhanced images, we conduct a user study among 43 human subjects. 8 sets of results are randomly selected for the test. Testers are invited to rank the results from 1 (best) to 8 (worst). For each low-light image, we present the image alongside results enhanced by various algorithms, without displaying the algorithms' names, to the human testers. Results from RETINEV are most favored by the testers, followed by EvLight, and LIME.

5.5. Ablation Study

We conduct an ablation study on the LOL v1 dataset to verify the contributions of the proposed components and data augmentation methods, as shown in Table 5. The introduction of temporal-mapping events brings an improvement of 10.51 dB improvement in PSNR, which proves the huge potential of temporal-mapping events in LLIE. Adding the T2I module (row 3) improves the performance by 0.29 dB. The design of the low-light degradation model in the data augmentation of denoising network training brings an improvement of 0.58 dB in PSNR, evidencing the benefit of the modeling of event camera response under low-illumination (rows 3 and 5). Introducing illumination in the reflectance enhancement module with simple "add fusion" yields 0.64 dB improvements (rows 5 and 6). When changing to cross-modal attention, the improvement comes to 0.47/0.59/0.62 dB compared to "add", "concatenation", and "multiply" fusion. All the components yield a substantial improvement of 11.62 dB, setting a new SOTA in event-based LLIE.

6. Conclusion

In this work, we approach low-light image enhancement through event-based illumination estimation. Unlike previous LLIE methods that extract edge information from motion events, we exploit the event camera's superior low-light responsiveness by capturing and processing temporal-mapping events for illumination estimation. This well-estimated illumination further improves reflectance decomposition and refinement. Based on Retinex theory, the enhanced image is reconstructed by multiplying reflectance and illumination. We create a real-world dataset using a beam-splitter dual-sensor system. On five synthetic datasets and our EvLowLight, RETINEV surpasses SOTA methods, achieving real-time inference at 35.6 FPS.

Acknowledgments. This work was partially supported by the Ministry of Education and Science of Bulgaria, as part of the Bulgarian National Roadmap for Research Infrastructure (support for INSAIT) and the European Union's Horizon Europe – the Framework Programme for Research and Innovation, under grant agreement 101168521. Additional funding was provided by the Henan Province Key R&D Special Project (No. 231111112700), the Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX0102), and the Fundamental Research Funds for the Central Universities.

References

[1] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on image processing*, 18(9):1921–1935, 2009. 2

[2] Yuhan Bao, Lei Sun, Yuqin Ma, and Kaiwei Wang. Temporal-mapping photography for event cameras. *arXiv preprint arXiv:2403.06443*, 2024. 2, 3, 4

[3] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014. 1

[4] 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. 1, 2, 3, 4, 5, 6, 7, 8

[5] Haoyu Chen, Minggui Teng, Boxin Shi, Yizhou Wang, and Tiejun Huang. Learning to deblur and generate high frame rate video with an event camera. *arXiv preprint arXiv:2003.00847*, 2020. 3

[6] Paul Debevec, Erik Reinhard, Greg Ward, and Sumanta Pattanaik. High dynamic range imaging. In *ACM SIGGRAPH 2004 Course Notes*, pages 14–es. 2004. 3

[7] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2020. 1

[8] 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. 1, 2, 6, 7, 8

[9] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 5, 8

[10] 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. 1, 3, 7, 8

[11] iniVation A G. Understanding the performance of neuromorphic event-based vision sensors. *Tech. Rep.*, 2020. 4

[12] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 3

[13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021. 3, 6

[14] 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, 3, 5, 6, 7

[15] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 2

[16] Seon Joo Kim, Hai Ting Lin, Zheng Lu, Sabine Süsstrunk, Stephen Lin, and Michael S Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2289–2302, 2012. 8

[17] Taewoo Kim, Jungmin Lee, Lin Wang, and Kuk-Jin Yoon. Event-guided deblurring of unknown exposure time videos. *arXiv preprint arXiv:2112.06988*, 2021. 3

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[19] Edwin H Land and John J McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971. 2, 3, 4

[20] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):9396–9416, 2021. 1

[21] 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. 1, 3, 5, 6, 7, 8

[22] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 2

[23] Jinxiu Liang, Yixin Yang, Boyu Li, Peiqi Duan, Yong Xu, and Boxin Shi. Coherent event guided low-light video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10615–10625, 2023. 3

[24] Songnan Lin, Jiawei Zhang, Jinshan Pan, Zhe Jiang, Dongqing Zou, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning event-driven video deblurring and interpolation. In *Proc. ECCV*, pages 695–710. Springer, 2020. 3

[25] 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. 3, 6

[26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[27] Xingyu Lu, Lei Sun, Diyang Gu, and Kaiwei Wang. Sge: structured light system based on gray code with an event camera. *Optics Express*, 32(26):46044–46061, 2024. 1

[28] Carver Mead. Neuromorphic engineering: In memory of misha mahowald. *Neural Computation*, 35(3):343–383, 2023. 1

[29] 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. 6, 8

[30] Nico Messikommer, Stamatios Georgoulis, Daniel Gehrig, Stepan Tulyakov, Julius Erbach, Alfredo Bochicchio, Yuanyou Li, and Davide Scaramuzza. Multi-bracket high dynamic range imaging with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 547–557, 2022. 3

[31] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 445–449. IEEE, 2013. 2

[32] 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. 3

[33] Lichtsteiner Patrick, Christoph Posch, and Tobi Delbruck. A 128×128 120 dB 15μ s latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 2008. 1

[34] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 2

[35] Wei Shang, Dongwei Ren, Dongqing Zou, Jimmy S Ren, Ping Luo, and Wangmeng Zuo. Bringing events into video deblurring with non-consecutively blurry frames. In *Proc. ICCV*, pages 4531–4540, 2021. 3

[36] 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. 3

[37] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 22871. IEEE, 2023. 3

[38] Lei Sun, Daniel Gehrig, Christos Sakaridis, Mathias Gehrig, Jingyun Liang, Peng Sun, Zhijie Xu, Kaiwei Wang, Luc Van Gool, and Davide Scaramuzza. A unified framework for event-based frame interpolation with ad-hoc deblurring in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3

[39] Lei Sun, Andrea Alfarano, Peiqi Duan, Shaolin Su, Kaiwei Wang, Boxin Shi, Radu Timofte, Danda Pani Paudel, Luc Van Gool, Qinglin Liu, et al. Ntire 2025 challenge on event-based image deblurring: Methods and results. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1324–1341, 2025. 3

[40] 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. 3, 5

[41] 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. 3, 5

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NIPS*, 30, 2017. 5

[43] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*, pages 1–6. IEEE, 2015. 8

[44] Patricia Vitoria, Stamatios Georgoulis, Stepan Tulyakov, Alfredo Bochicchio, Julius Erbach, and Yuanyou Li. Event-based image deblurring with dynamic motion awareness. *arXiv preprint arXiv:2208.11398*, 2022. 3

[45] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 1, 2, 6

[46] 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. 5, 6

[47] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018. 1, 2, 3, 4, 5, 6, 7, 8

[48] 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, 3, 4, 5, 7, 8

[49] 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 Vi-*

sion and Pattern Recognition, pages 1662–1671, 2023. 2, 6

[50] 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. 2, 6

[51] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9893–9903, 2023. 2

[52] 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. 6

[53] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 2

[54] Yaozu Ye, Hao Shi, Kailun Yang, Ze Wang, Xiaoting Yin, Lei Sun, Yaonan Wang, and Kaiwei Wang. Towards anytime optical flow estimation with event cameras. *Sensors*, 25(10): 3158, 2025. 1

[55] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023. 2, 3

[56] Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, and Boxin Shi. Eventps: Real-time photometric stereo using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9602–9611, 2024. 3

[57] Lei Yu, Bishan Wang, Xiang Zhang, Haijian Zhang, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Learning to super-resolve blurry images with events. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 5, 8

[58] 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, 6

[59] Shaobo Zhang, Lei Sun, and Kaiwei Wang. A multi-scale recurrent framework for motion segmentation with event camera. *IEEE Access*, 11:80105–80114, 2023. 1

[60] Xiang Zhang and Lei Yu. Unifying motion deblurring and frame interpolation with events. In *Proc. CVPR*, pages 17765–17774, 2022. 3

[61] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 2, 6

[62] Chu Zhou, Minggui Teng, Jin Han, Jinxiu Liang, Chao Xu, Gang Cao, and Boxin Shi. Deblurring low-light images with events. *International Journal of Computer Vision*, 131(5): 1284–1298, 2023. 3