

Seeing Motion at Nighttime with an Event Camera

Haoyue Liu¹, Shihan Peng¹, Lin Zhu², Yi Chang^{1*}, Hanyu Zhou¹, Luxin Yan¹

¹ National Key Lab of Multispectral Information Intelligent Processing Technology
 School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
²Beijing Institute of Technology

{liuhy, pengshihan, yichang, hyzhou, yanluxin}@hust.edu.cn, linzhu@bit.edu.cn

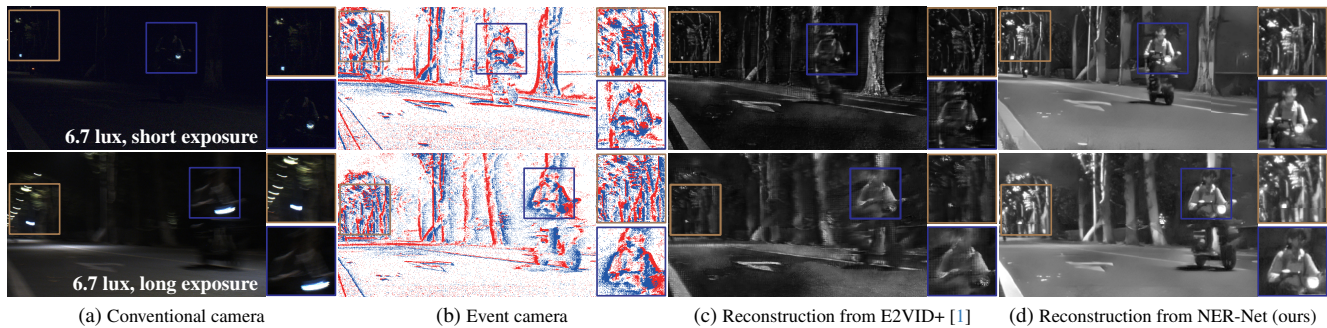


Figure 1. Comparison between frame-based imaging and event-based imaging. (a) The extended exposure time of conventional cameras leads to motion blur in low-light conditions (**6.7 lux**). (b) Event cameras can capture high-speed and high-dynamic-range scene information. (c) and (d) are reconstruction results of E2VID+ [1] and the proposed method.

Abstract

We focus on a very challenging task: imaging at nighttime dynamic scenes. Most previous methods rely on the low-light enhancement of a conventional RGB camera. However, they would inevitably face a dilemma between the long exposure time of nighttime and the motion blur of dynamic scenes. Event cameras react to dynamic changes with higher temporal resolution (microsecond) and higher dynamic range (120dB), offering an alternative solution. In this work, we present a novel nighttime dynamic imaging method with an event camera. Specifically, we discover that the event at nighttime exhibits temporal trailing characteristics and spatial non-stationary distribution. Consequently, we propose a nighttime event reconstruction network (NER-Net) which mainly includes a learnable event timestamps calibration module (LETC) to align the temporal trailing events and a non-uniform illumination aware module (NIAM) to stabilize the spatiotemporal distribution of events. Moreover, we construct a paired real low-light event dataset (RLED) through a co-axial imaging system, including 64,200 spatially and temporally aligned image GTs and low-light events. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art methods in terms of visual quality and generalization ability on

real-world nighttime datasets. The project are available at: <https://github.com/Liu-haoyue/NER-Net>.

1. Introduction

Imaging in nighttime dynamic scenes is crucial for various applications, including autonomous driving and video surveillance. Conventional cameras often require longer exposure times to capture information in low-light environments, but this inevitably leads to motion blur, which is a significant contradiction for conventional cameras, as shown in Fig. 1(a). While existing low-light image/video enhancement techniques can enhance contrast [2–14], the loss of essential scene structural information leads to suboptimal imaging results. Thus, achieving high-quality imaging in nighttime dynamic scenarios remains an intricate challenge.

Event cameras [15–20] overcome the challenge of the trade-off between seeing faster and seeing clearer, especially in nighttime scenarios. This sensor activates each pixel independently in response to changes in brightness and possesses several notable advantages, including high temporal resolution ($1\mu\text{s}$) and high dynamic range (120dB), as shown in Fig. 1(b). Therefore, reconstructing images from events offers an attractive approach for nighttime imaging.

Early event-based reconstruction methods rely on hand-

crafted priors [21–24]. Recently, learning-based event reconstruction methods [1, 25–34] have made remarkable progress. Sufficient data is the cornerstone of learning methods, but high-quality paired data is often hard to acquire. Unsupervised and self-supervised methods attempt to relax the dependency of paired data through knowledge transfer [32, 33] or the construction of self-supervised loss functions [34]. However, these methods ignore the non-stationary changes of events under complex lighting when constructing the unsupervised learning strategy, resulting in poor performance.

Otherwise, supervised methods [1, 25, 27, 30, 31] commonly utilize simulators [35–38] to generate paired data. While the above simulators provide powerful capabilities, their greatest enemy when applied in real-world scenarios is the diversity and non-uniformity of lighting, especially in the presence of artificial light sources at night. Simulation events tend to have a more uniform signal characteristic, whereas the signal characteristics of events are complex and varied in the real world. To address this issue, many works [28, 39, 40] have set up co-axial imaging systems to collect real-world paired data, but these data were taken in good light and do not provide high-quality GTs for low-light events. Although these state-of-the-art methods have achieved impressive results in well-illuminated conditions, they cannot well generalize to real nighttime environments due to the lack of high-quality paired data.

To cope with the above issue, we first construct a paired real low-light event dataset (RLED). Specifically, we design a co-axial imaging system comprising an event camera with a neutral density(ND) filter and a conventional camera, which allows for the simultaneous acquisition of low-light events and high-quality images. RLED provides 64,200 images with corresponding events, capturing events illuminance levels ranging from 0.5 lux to 1000 lux.

In addition, we propose a nighttime event reconstruction network (NER-Net) including a learnable event timestamps calibration module (LETC) and a non-uniform illumination aware module (NIAM). We first explore the spatiotemporal distribution patterns of events in nighttime lighting conditions based on the sensor circuit characteristics. The sensor exhibits a lower cutoff frequency in low-light conditions, leading to the trailing events effect. As events distribute similarly to the response of an RC low-pass filter [36], the time interval between adjacent events at the same pixel increases gradually over time. Based on this characterization, we design an event trail suppression (ETS) method that aligns the delayed-trigger events to their correct timestamps and utilizes a LETC to learn the alignment process of events in the time domain, and LETC pre-trained using data pre-processed by ETS. Besides, non-uniform artificial lighting results in significant variations in the distribution of events in different regions. Regions with higher illuminance often have clearer textures and higher event density, and vice

versa. In this work, we provide a NIAM to model the non-stationary state of events. NIAM can perform adaptive non-uniform correction based on event density cues and model the non-stationary state of signal distribution by aggregating long-term memory parameters and hierarchical spatial information. As shown in Fig. 1(d), NER-Net can effectively reconstruct the intensity of the nighttime dynamic scene. Our main contributions can be summarized as follows:

- We provide an event-based nighttime imaging solution under non-uniform illumination and construct a paired multi-illumination level real-world dataset. To our best knowledge, RLED is the first dataset that provides high-quality pixel-aligned GTs for low-light events.
- We discover the core reasons for the bad performance of existing methods are the temporal trailing effect and the spatial non-uniform response of events, and propose a NER-Net to model the non-stationary status of events under nighttime non-uniform illumination, which effectively improves the reconstruction quality.
- Extensive experimental results show that our NER-Net outperforms SOTA methods on three real-world nighttime datasets. Besides, the proposed method exhibits the ability to generalize to various lighting scenarios.

2. Related Work

Nighttime Imaging. Frame-based nighttime imaging approaches such as low-light image enhancement (LLIE) have been widely explored. Early LLIE methods were mostly based on histogram equalization [41–44] and Retinex theory [45–47]. In recent years deep learning approaches have achieved remarkable success [2–6, 13, 14, 48–50]. In addition, several works explored infrared nighttime imaging [51, 52] or hybrid RGB-VIR imaging approaches [53, 54], which enhanced nighttime information acquisition capabilities by broadening the spectral bands. To see moving in the dark, some works [8, 9, 12] collected paired low-light and day-light video datasets to enhance generalization in real dynamic scenes, and focused on the temporal consistency of videos. Unfortunately, setting a short exposure time for fast-moving scenes leads to significant loss of scene details. Frame-based methods cannot recover lost information.

Recently, nighttime imaging based on events became a hot topic. DVS-Dark [33] learns common features (scene structures) between the day and night and transfers daytime domain-specific knowledge (detailed textures) to the nighttime domain. Furthermore, event-guided image/video enhancement approaches have been explored for nighttime imaging [55–58]. These works skillfully leveraged the high dynamic range and high speed of event cameras to help images recover scene intensity. However, it is worth noting that the spatiotemporal distribution of events in real nighttime scenes is complex and variable. Neglecting these factors can

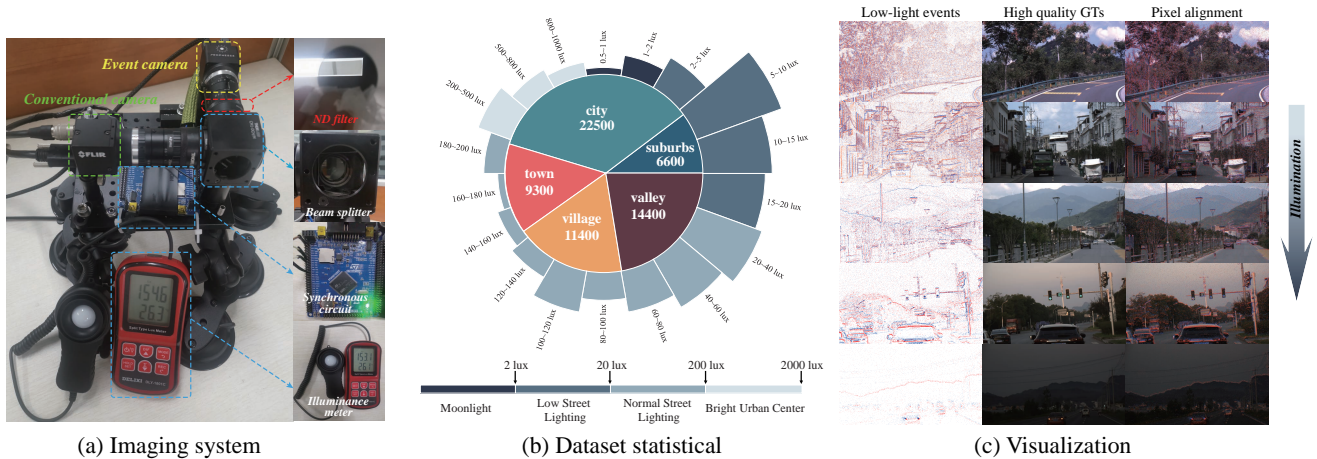


Figure 2. Features of the proposed RLED. (a) The implementation of our coaxial imaging system. (b) Distribution of illumination and scene of the proposed dataset. (c) Visualization of RLED, which collects low-light events aligned with high-quality images at the pixel level.

significantly impact the quality of imaging. In this work, we analyze the spatiotemporal distribution patterns of events in nighttime lighting based on the sensor circuit characteristic and design a NER-Net to address the temporal trailing effect and spatial non-uniform distribution of events.

Event-based image and video reconstruction. Recovering high speed and high dynamic range intensity images from events is an efficient imaging approach. Early methods attempted to construct mathematical models of event generation or introduced artificial priors to recover image intensity [21–24], but suffered from severe artifacts or over-smoothed results. With the development of deep learning technology, the quality of reconstructed images has seen an impressive improvement [1, 25–34]. Rebecq *et al.* [25] proposed an effective recurrent network architecture and leveraged a large amount of simulated data to learn the mapping from events to images. However, the sensor noise can lead to serious reconstruction artifacts due to the limited diversity of simulated data. Thus, Stoffregen *et al.* [1] and Liu *et al.* [31] attempt to improve generalization in real-world scenarios by enhancing the simulation process. In contrast, Zou *et al.* [28] collected a paired events and images dataset through a co-axial imaging system, and the HDR ground truth was synthesized by combining two low dynamic range (LDR) images with different brightness levels. Unfortunately, this method cannot generate high-quality ground truth at night.

Although these methods have achieved impressive results in well-illuminated scenes, the lack of low-light paired data makes them ineffective in real nighttime scenes. In this work, we contribute a multi-illumination paired dataset, which provides high-quality pixel-aligned GTs for low-light events.

3. Nighttime Event Signal Characteristics

As mentioned above, paired low-light event and high-quality image datasets are still lacking, and the spatiotempo-

ral non-stationary characteristics of nighttime events remain under-explored. In this section, we first present the real low-light event dataset (RLED). Then, we analyze the response properties of events in the temporal domain and point out that the core reason for the event trailing effect is the reduced sensor cutoff frequency. Furthermore, we discuss the impact of non-uniform artificial lighting on the spatial distribution of events, higher illumination near the light source typically leads to higher event density, and vice versa.

3.1. Real Low-light Event Dataset

Obtaining high-quality paired data in nighttime dynamic scenes is virtually an impossible task because conventional cameras have poorer imaging capabilities at night. To address this issue, we collect data under optimal lighting conditions during the daytime, while also equipping the event camera with an ND filter to obtain low-light events.

First, we built a co-optical axis imaging system comprising an event camera (Prophesee EVK4, 1280*720), a conventional camera (FLIR BFS-U3-32S4C, 2048*1536), a beam splitter (Thorlabs BSW26R) and a ND filter (Thorlabs ND20A), as illustrated in Fig. 2(a). The beam splitter divides the incoming light into two equal parts and sends them to the event camera and the conventional camera respectively. The ND filter is placed between the event camera and the beam splitter to filter out the majority of incoming light (99%). In addition, we provided external trigger signals to the cameras through a programmable synchronous circuit, enabling precise synchronization of the timestamps of both cameras. Finally, we achieved pixel alignment between the two cameras through the stereo rectification process.

RLED contains 64,200 images and corresponding events, we utilized a photometer to continuously measure scene illumination, and calculate the illuminance value after attenuation at the event camera. As shown in Fig. 2(b), the

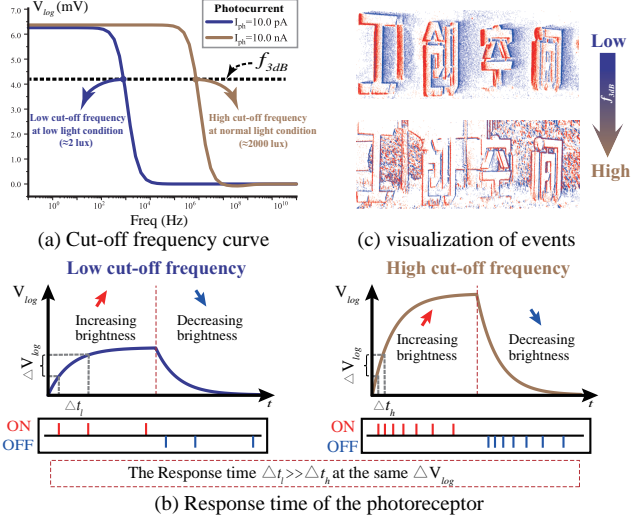


Figure 3. This figure illustrates the mechanism of trailing events. (a) Photoreceptor cut-off frequency decreases with decreasing illumination. (b) The response time of the photoreceptor increases as the cut-off frequency decreases, requiring more time ($\Delta t_l > \Delta t_h$) to reach the same voltage change V_{log} at low illumination. (c) The increase in response time leads to the effect of trailing events.

illumination range for the event data was between 0.5 lux to 1000 lux. The capture scenes included city (35.0%), suburbs (10.3%), town (14.5%), village (17.8%), and valley (22.4%).

3.2. Trailing Events in Temporal Domain

The signal characteristics of event cameras can be influenced by external factors such as temperature, notably illumination. The photocurrent of the photodiode I_{ph} is directly related to the intensity of illumination,

$$I_{ph} = R \times A \times E, \quad (1)$$

where R is the responsivity of the photodiode, A is the photodiode active area, E is the illuminance of the incident light. The magnitude of photocurrent directly influences cutoff frequency f_{3dB} of the sensor, which is a key indicator used to evaluate signal response speed. The magnitude of the cutoff frequency is directly correlated with the photocurrent [16],

$$f_{3dB} = \frac{1}{2\pi C_{DMfb}} \frac{I_{ph}}{U_t}, \quad (2)$$

where C_{DMfb} is the gate-drain capacitance of transistor M_{fb} and U_t is the thermal voltage. When the temperature is constant, f_{3db} is proportional to I_{ph} . Therefore, when the environmental illumination changes, it directly affects the cutoff frequency of the sensor, thereby altering the signal characteristics of events.

The variation of the cutoff frequency with changes in illumination is depicted in Fig. 3(a). The cutoff frequency represents the ability of the photodiode voltage V_{log} to track changes in photocurrent I_{ph} . When the cutoff frequency is high, the voltage can rapidly respond to current changes, leading to the quick generation of events, and vice versa. As

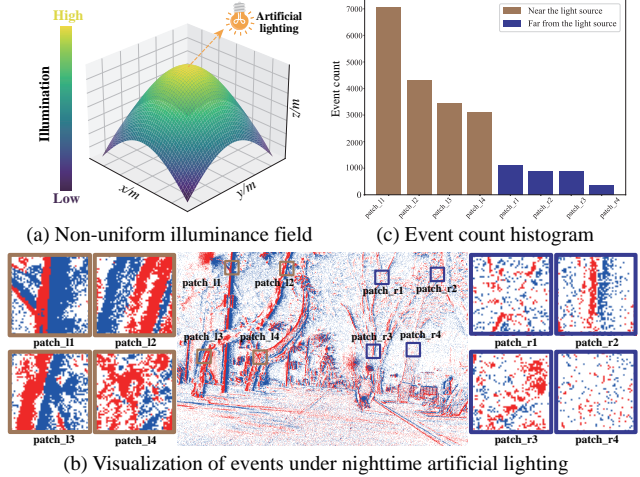


Figure 4. The spatial characteristics of events at night. (a) The illuminance field under artificial nighttime lighting is non-uniform and exhibits significant variations with different distances from the light source. (b) Events are more densely distributed in the vicinity of the artificial light source. (c) Greater luminance difference between nearer the light source and the dark background, resulting in more events being triggered, and vice versa.

shown in Fig. 3(b), when the cutoff frequency is lower, it requires more response time for the same voltage change ΔV_{log} . This behavior is analogous to the response process of a first-order RC low-pass filter [36], where the intervals between consecutive events progressively increase. Fig. 3(c) illustrates the trailing events resulting from the lower cutoff frequency of the sensor under low light conditions, which significantly blurs the edges of the scene.

3.3. Non-uniform Illumination in Spatial Domain

In nighttime autonomous driving or surveillance, artificial lighting sources such as street lamps are often present. One significant characteristic of artificial light sources is their non-uniformity. The illuminance formula is

$$I = \frac{\Phi}{4\pi d^2}, \quad (3)$$

where I is the illuminance in the scene, Φ is the luminous flux, which depends on the power of the light source, and d is the distance from the light source to the scene. The power of the lighting can be regarded as constant, then the only factor affecting the illuminance in the scene is the distance from the light source. The illuminance field of the artificial light source is shown in Fig. 4(b), the closer the distance to the light source, the higher the illuminance.

The event generation process [34] can be formulated as

$$\log I(x, y, t) - \log I(x, y, t - \Delta t) = pC, \quad (4)$$

where $\log I(x, y, t)$ is the logarithmic illuminance at pixel (x, y) and time t , Δt is the time interval between consecutive events, $p \in [-1, 1]$ is the polarity of events, and C is the contrast threshold of the event camera. This process indicates

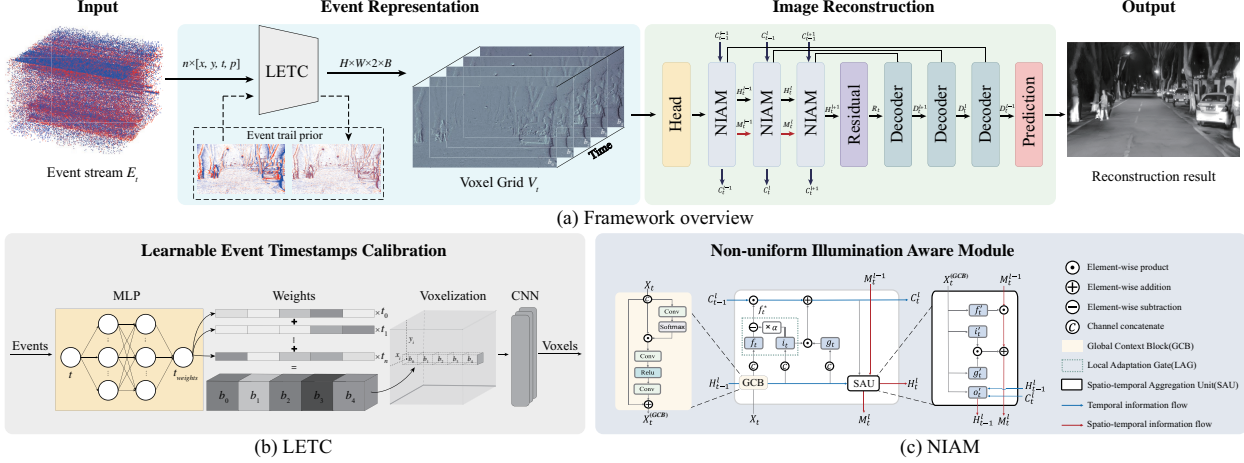


Figure 5. The overall architecture of the proposed **Nighttime Event Reconstruction (NER)** network, which can model the non-stationary spatiotemporal distribution of nighttime events. (a) NER contains a Learnable Event Timestamps Calibration (LETC) and a U-shaped image reconstruction network with Non-uniform Illumination Aware Module (NIAM) encoders. (b) LETC generates voxels with sharp edges by redistributing the weights of event timestamps within different voxel units. (c) NIAM performs local and global illumination sensing and regulation via the Local Adaptation Gate (LAG) and Global Context Block (GCB), respectively. By using a Spatiotemporal Aggregation Unit (SAU), NIAM can adaptively exploit and fuse multi-scale spatial information and long-term temporal information.

that an event is triggered once the logarithmic illuminance change at a particular pixel exceeds the threshold C . Due to the extremely low background brightness, the closer the distance to the light source, the greater the difference in brightness between the object and the background, leading to the triggering of more events. As shown in Fig. 4(a), the distance to the light source is closer on the left side, and there is a significantly higher event density compared to the right side. The lower illumination on the right side results in a lower event density and weak texture as in Fig. 4(c).

4. Nighttime Event Reconstruction Network

Fig. 5 illustrates the overall framework of nighttime event reconstruction network (NER-Net).

4.1. Trail Prior Guided Event Representation

Event Trail Suppression. In Section 3, we discuss the response of events is similar to an RC low-pass filter, the rate of change in photocurrent voltage gradually decreases after an excitation. This phenomenon becomes more pronounced with lower cutoff frequencies, resulting in progressively longer time intervals between consecutive events. We design the ETS method based on this characteristic, and the conditions for identifying trailing events include: (1) the polarity of consecutive events remains unchanged, (2) the time intervals between consecutive events are gradually increasing, (3) the largest time interval among consecutive events is less than a specified threshold, as shown in Fig. 6(a).

Once a trailing event has been identified, ETS will correct the timestamp of the event. The corrected timestamps are rearranged at a specified interval, which is specified to be

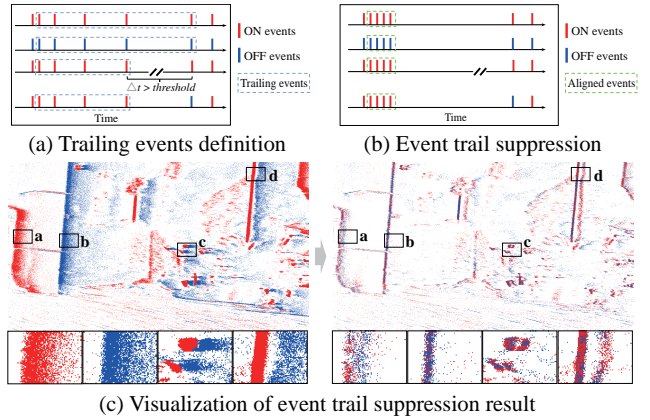


Figure 6. Method and visualization of Event Trail Suppression (ETS). (a) ETS is designed based on the event response characteristics in Section 3.2 with three conditions. (b) ETS aligns event timestamps to the correct position. (c) Events processed by ETS result in sharper edges.

$1\mu\text{s}$ based on statistics in daylight, as shown in Fig. 6(b). Fig. 6(c) illustrates the visual results of ETS.

Learnable Event Timestamps Calibration. To enhance the reconstruction quality with the prior of tail suppression, we design an end-to-end trainable LETC module. The LETC can transform the trailing events to sharp voxel grids by assigning weights to timestamps. The classic voxelization method transform events $\varepsilon_k = \{e_i\}_{i=0}^{N-1}$ into a tensor $V \in \mathbb{R}^{B \times H \times W}$ with B bins [59], which can be formulated as

$$V(k) = \sum_i p_i \max\left(0, 1 - \left|k - \frac{t_i - t_0}{t_N - t_0} (B - 1)\right|\right), \quad (5)$$

where N is the number of events, p_i and t_i represent the polarity and timestamp of the i -th event respectively, and

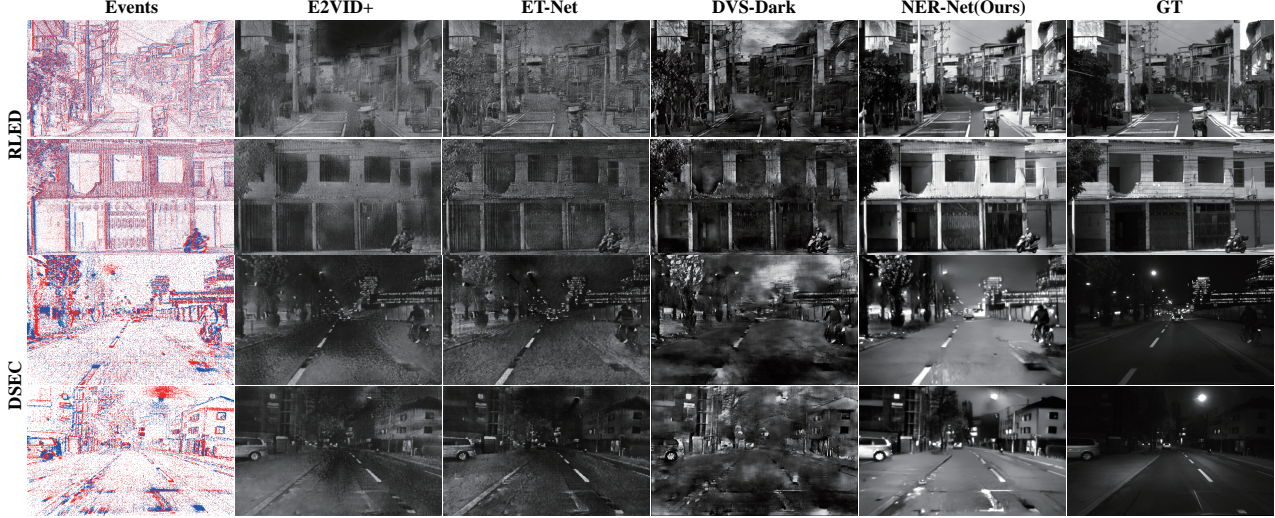


Figure 7. Visual comparison between other SOTA methods and proposed NER-Net on real-world datasets.

$k \in [0, B - 1]$. This method uniformly fills events into the two nearest bins using the fixed interpolation approach.

In order to make the event representation process more adaptable to specific tasks, Gehrig *et al.* [60] introduced a representation method called event spike tensor (EST) that allows each event to be filled into all B bins, and the filling values are learned end-to-end by an MLP. Motivated by this, the proposed LETC module can dynamically assign weights based on the degree of trailing events to generate sharp voxels dynamically allocates weights based on the degree of event trailing, thereby generating sharp voxels. Differing from their approach of directly embedding the representation module into the task network for joint training [60], the LETC module includes a tail CNN layer to integrate information for voxel generation, and then it uses ETS-processed events as labels to train LETC, enabling LETC to have a priori knowledge for trailing suppression. Finally, LETC is integrated into the reconstruction network for fine-tuning.

4.2. Non-uniform Illumination Aware Module

Events are temporally continuous and spatially sparse, the restoration of scene intensity relies on modeling the continuous spatiotemporal context. Non-uniform illumination disrupts the spatiotemporal distribution stability of events, leading to significant reconstruction artifacts. Hence, stabilizing the spatiotemporal distribution of events is crucial for restoring scene intensity. Inspired by the human vision system (HVS), which can globally regulate light intake through the expansion and contraction of the pupil, retinal cells can locally adjust based on light intensity through the response compression mechanism[61, 62]. We design NIAM encoders comprising GCB, LAG, and SAU modules.

GCB [63] captures long-range dependencies in non-

uniform illumination fields, which can be formulated as

$$\mathcal{X} = x + \text{Conv}_{1 \times 1}(\text{ReLU}(\text{LN}(\text{Conv}_{1 \times 1}(\sum_{j=1}^N \gamma_j x_j)))), \quad (6)$$

where N is the number of positions in the feature map, $\gamma_j = \frac{e^{\text{Conv}_{1 \times 1}(x_j)}}{\sum_m e^{\text{Conv}_{1 \times 1}(x_m)}}$ is the weight for global attention pooling.

Local Adaptation Gate. The standard convLSTM [64] can effectively handle temporal information. Still, we aim to preserve weak texture area while eliminating redundant information in the temporal memory parameter. The event density encoded information is additionally introduced in the LAG and acts as negative feedback to the forgot gate. This allows for adaptive control of the memory parameters based on the density of input events. The architecture of LAG is illustrated in the green box of Fig. 5(c), formulated as

$$f_t^{\text{LAG}} = \sigma(f_t - \alpha i_t), \quad (7)$$

where $\alpha = e^{\sigma(\text{Conv}(x_t))}$ is the event density encoded information, $\sigma(\cdot)$ denotes sigmoid function, f_t and i_t represent the forgot gate and input gate of convLSTM respectively. Note that, event density depends on both illumination and object motion. LAG regulates information flow based on event density without requiring their decoupling. The effectiveness of the proposed method is shown in Fig. 7.

Spatiotemporal Aggregation Unit. ConvLSTM-based encoder are widely used in event-based image reconstruction networks[1, 25, 27–29, 34], but they focus more on modeling temporal variations and memory states being updated repeatedly over time inside each LSTM unit, resulting in information blockage between different layers. Inspired by PredRNN [65], we design the SAU to integrates long-term temporal information along with hierarchical spatial information, as shown within the black dashed box Fig. 5(c). \mathcal{C}_t^l is the temporal cell including GAB and LAG, which being updated repeatedly over time. \mathcal{M}_t^l is the spatiotemporal memory that can aggregate hierarchical features across lay-

| Datasets | Methods | E2VID+ | | | ET-Net | | | DVS-Dark | NER-Net(ours) | | |
|-------------|---------|--------|--------|---------------|--------|--------|--------------|--------------|---------------|--------|---------------|
| | | *ESIM | V2E | RLED | *ESIM | V2E | RLED | *DVS-Dark | ESIM | V2E | RLED |
| RLED | MSE ↓ | 0.099 | 0.100 | 0.019 | 0.051 | 0.096 | 0.075 | 0.095 | 0.077 | 0.087 | 0.011 |
| | SSIM ↑ | 0.298 | 0.404 | 0.688 | 0.307 | 0.421 | 0.534 | 0.355 | 0.179 | 0.081 | 0.717 |
| | LPIPS ↓ | 0.560 | 0.582 | 0.393 | 0.559 | 0.544 | 0.411 | 0.591 | 0.628 | 0.749 | 0.309 |
| DSEC-night | LOE ↓ | 1383.4 | 1704.2 | 1165.1 | 1271.1 | 1782.9 | 1227.6 | 1757.7 | 1421.2 | 1694.9 | 1031.2 |
| | SSIM ↑ | 0.311 | 0.304 | 0.323 | 0.331 | 0.272 | 0.411 | 0.337 | 0.153 | 0.305 | 0.340 |
| | LPIPS ↓ | 0.586 | 0.554 | 0.534 | 0.581 | 0.552 | 0.534 | 0.608 | 0.622 | 0.536 | 0.502 |
| MVSEC-night | LOE ↓ | 1292.9 | 1708.5 | 1283.8 | 1588.9 | 1739.7 | 1344.3 | 1604.2 | 1534.1 | 1743.7 | 1198.4 |
| | SSIM ↑ | 0.078 | 0.104 | 0.141 | 0.113 | 0.085 | 0.197 | 0.120 | 0.062 | 0.092 | 0.205 |
| | LPIPS ↓ | 0.579 | 0.644 | 0.564 | 0.573 | 0.658 | 0.507 | 0.598 | 0.606 | 0.645 | 0.482 |
| VECTor-hdr | LOE ↓ | 1495.0 | 1824.8 | 1384.5 | 1617.9 | 1932.4 | 1526.5 | 1759.2 | 1439.8 | 1839.8 | 1139.8 |
| | SSIM ↑ | 0.224 | 0.184 | 0.269 | 0.208 | 0.153 | 0.232 | 0.317 | 0.197 | 0.219 | 0.344 |
| | LPIPS ↓ | 0.695 | 0.707 | 0.652 | 0.701 | 0.711 | 0.683 | 0.723 | 0.658 | 0.682 | 0.634 |

* Using the pretrained model provided in the original paper.

Table 1. Quantitative comparisons with SOTA methods on three real-world datasets: RLED, DSEC-night, and MVSEC-night. The top two results are colored in **blue** and **green**.

ers, and it also transmit the spatiotemporal memory from the top layer at time $t - 1$ with rich semantic information to the bottom layer at time t through a progressive upsampling. Thus, NIAM can effectively perceive the non-stationary state of signal distribution, and model the dynamic information in non-uniform illumination environments. The detailed network architecture can be found in supplementary material.

4.3. Training Details

Loss Function. We use LPIPS [66] to evaluate image quality, and temporal consistency(TC) loss [67] is employed to mitigate the temporal artifacts between adjacent reconstructed images. The total loss is

$$\sum_{i=0}^L \mathcal{L}_i^{LPIPS} + \lambda_{TC} \sum_{i=L_0}^L \mathcal{L}_i^{TC}, \quad (8)$$

where L and L_0 are the iteration time step and the starting index for TC loss calculations. We set $L = 40$, $L_0 = 2$, and $\lambda_{TC} = 2$ respectively.

Implementation. Our network is implemented using the Pytorch framework [68] and use ADAM [69] with a learning rate of 0.0001. Patches at the size of 160×160 are randomly cropped from training samples. We train our model for 200 epochs using an NVIDIA A100 GPU.

5. Experiments

5.1. Datasets and Experimental Settings

Datasets. Since there is a gap between the simulated data and the real world data, our target is to assess the applicability of methods in real scenarios. Thus, we choose four real low-light datasets to validate the performance of our method: RLED, DSEC [70] MVSEC [71] and VECTor [72].

Comparison Methods. We compare our models with three SOTA methods E2VID+ [1], ET-Net [27], and DVS-Dark [33]. To ensure the fairness of the experiments, in addition to using the pre-trained model provided in the original paper, we re-train the above model using paired data from RLED and V2E [36] simulated data based on RLED images using default noisy settings.

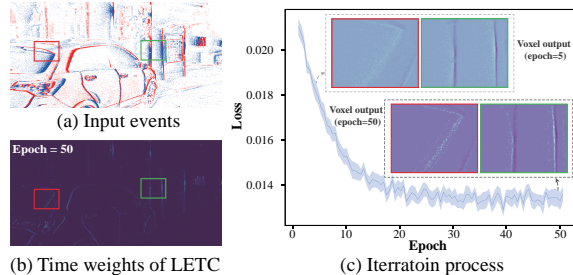


Figure 8. Effectiveness of LETC. (a) Input events. (b) Visualization of $t_{weights}$ at the 50th epoch. (c) LETC can generate sharp voxels from trailing events by training with ETS data.

Evaluation Metrics. For RLED, we use mean squared error (MSE), structural similarity (SSIM) [73], and perceptual similarity (LPIPS) [66] as evaluation metrics. For the other three datasets, considering the low quality of nighttime images, we opt for a no-reference image quality assessment metric known as lightness-order-error (LOE) [74] instead of MSE. LOE effectively measures the naturalness of images in non-uniform illumination conditions.

5.2. Quantitative and Qualitative Evaluation

Qualitative Evaluation. Fig. 7 depicts a visual comparison of results between other SOTA methods and proposed NER-Net in low-light dynamic scenes. E2VID+ [1] and ET-Net [27] exhibit temporal memory errors in the presence of noise and artificial light sources, leading to severe black artifacts. DVS-Dark [33] is a GAN-based method that experiences model breakdowns when encountering diverse data distributions, resulting in unnatural reconstruction results. NER-Net achieves more visually pleasing high dynamic range results in non-uniform illumination and high-noise environments.

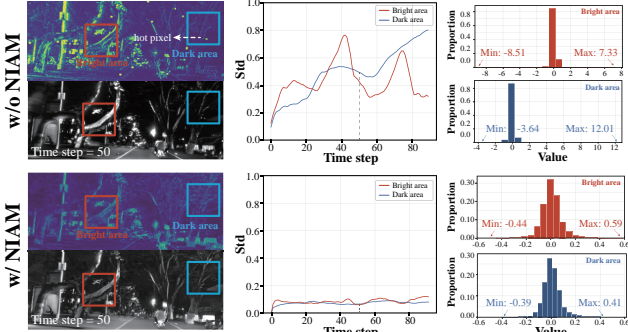
Quantitative Evaluation. The quantitative results are reported in Table 1. It is observed that all the top two results were trained on RLED. Note that, models trained on a single noise setting using V2E [36] often exhibit poor performance. The reason is that the spatiotemporal distribution of nighttime events is non-uniform and dynamically changing, and there is a large gap with simulated events. NER-Net outperforms the SOTA methods on all three benchmarks, which confirms the effectiveness of our method with event timestamps calibration and non-uniform illumination awareness. More importantly, NER-Net achieved better generalization across four datasets captured by different sensors. We also validate the generalizability of the proposed method during the day. Please refer to supplementary material for details.

5.3. Ablation Study and Discussion

How does LETC work? As shown in Fig. 8, the attention of LETC focuses on the correct timestamps as the model iterates. Consequently, the generated voxels exhibit sharper edges. In Table 2, one can see that directly reconstructing ETS data will yield poorer results, as there is a larger gap

| LETC | | NIAM | | | RLED | | |
|------|------|------|-----|-----|--------------|--------------|--------------|
| ETS | LETC | LAG | GCB | SAU | MSE ↓ | SSIM ↑ | LPIPS ↓ |
| | | | | | 0.019 | 0.688 | 0.393 |
| ✓ | | | | | 0.045 | 0.586 | 0.423 |
| ✓ | ✓ | | | | 0.017 | 0.692 | 0.368 |
| ✓ | ✓ | | | | 0.016 | 0.704 | 0.354 |
| ✓ | | ✓ | | | 0.014 | 0.710 | 0.331 |
| ✓ | ✓ | ✓ | ✓ | | 0.015 | 0.712 | 0.324 |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.011 | 0.717 | 0.309 |

Table 2. Ablation studies of NER-Net.



(a) Visualization of C_t^l and result (b) Standard deviation of C_t^l (c) Distribution of C_t^l

Figure 9. Effectiveness of NIAM. (a) Visualization of memory cell C_t^l and reconstruction results. (b) Standard deviation of C_t^l within 90 time steps. (c) The distribution of C_t^l within 90 time steps.

between the ETS-processed data and real-world data. LETC with trail prior achieves better results through end-to-end training, because the network can learn the mapping from tail-suppressed events to images during the training process. **Effectiveness of non-uniform illumination awareness.** We study how NIAM models non-stationary spatiotemporal information flow as shown in Fig. 9. The length of the reconstructed image sequence is 90. For the model w/o NIAM, the standard deviation of C_t^l exhibits significant fluctuations under non-uniform illumination and noise disturbance (Fig. 9 (b)), leading to outliers in temporal propagation as shown in Fig. 9 (c), and resulting in contrast distortion in the reconstructed images. NIAM can suppress such abnormal disturbances, and ensure the stability of temporal parameters in both bright and dark areas, thereby reconstructing high-quality images. As shown in Table 2, LAG and SAU play more significant roles in improving the reconstruction quality because they can adaptively perceive changes in event density and stabilize the distribution of events.

Nighttime dynamic scene imaging: event vs image. We capture 6 real-world challenging sequences at nighttime with fast-moving objects and random oscillations of the imaging system. The illumination in the scene ranges between 3 and 20 lux. We carefully set the exposure time of the conventional camera to ensure there was no noticeable motion blur. We compared the results with three SOTA low-light enhancement methods: KinD++ [6], SCI [13], and URetinex-Net [50]. No-reference metrics such as LOE [74], NIQE [75], and SPAQ [76] are adopted for evaluation. Besides, we perform a user study (US) to quantify the subjective visual. We

| Method | LOE ↓ | NIQE ↓ | SPAQ ↑ | US ↑ |
|-------------------|--------------|--------------|----------------|-------------|
| KinD++ [6] | 113.02 | 26.24 | 3708.02 | 5.03 |
| SCI [13] | 99.78 | 12.49 | 1658.48 | 3.92 |
| URetinex-Net [50] | 123.45 | 13.19 | 9228.63 | 4.34 |
| NER-Net(Ours) | 92.81 | 11.78 | 8205.78 | 6.02 |

Table 3. Quantitative results on nighttime dynamic scene.

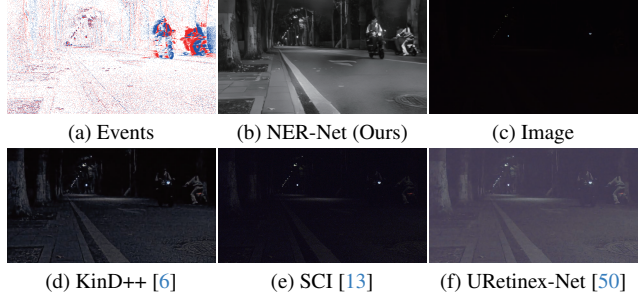


Figure 10. Visual comparisons on nighttime dynamic scenes.

invite 20 human subjects to score the visual quality on a scale from 1 to 10, higher score represents better quality.

Table 3 reports that NER-Net achieves the best performance. From Fig. 10, it is observed that the excessively short exposure time of conventional cameras used to capture fast-moving objects, leads to the loss of information that is challenging to restore. In contrast, NER-Net can reconstruct scene intensity well with events.

Limitation. NER-Net performs well in the presence of non-uniform artificial lighting and severe sensor noise but fails in extreme low-light scenarios (*e.g.* <0.5 lux). The reason is that the drastic reduction of events triggered by object motion under extremely low illumination, it is insufficient for reconstructing reasonable scene details. In addition, the proposed method does not take into account the color information. In the future, we attempt to employ frame-based cameras to provide additional scene information.

6. Conclusion

In this work, we propose an event-based solution for nighttime dynamic scene imaging and propose the first paired real-world dataset comprising high-quality pixel-aligned GTs for low-light events. Moreover, we demonstrate that the non-stationary status of events under nighttime non-uniform illumination is a key factor degrading the quality of reconstructed images. We model the spatiotemporal disturbance process based on the mechanisms of temporal trailing effects and the spatial non-uniform response of events. The proposed method significantly outperforms the state-of-the-art methods. We believe that our work can contribute to the application of event cameras in real nighttime scenarios.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 62371203. The computation is completed in the HPC Platform of Huazhong University of Science and Technology.

References

- [1] T. Stoffregen, C. Scheerlinck, D. Scaramuzza, T. Drummond, N. Barnes, Li. Kleeman, and R. Mahony. Reducing the sim-to-real gap for event cameras. In *Eur. Conf. Comput. Vis.*, pages 534–549, 2020. [1](#), [2](#), [3](#), [6](#), [7](#)
- [2] K. Lore, A. Akintayo, and S. Sarkar. Llnet: A deep auto-encoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. [1](#), [2](#)
- [3] C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to see in the dark. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3291–3300, 2018.
- [4] R. Wang, Q. Zhang, C. Fu, X. Shen, W. Zheng, and J. Jia. Underexposed photo enhancement using deep illumination estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6849–6857, 2019.
- [5] F. Lv, B. Liu, and F. Lu. Fast enhancement for non-uniform illumination images using light-weight cnns. In *ACM Int. Conf. Multimedia*, pages 1450–1458, 2020.
- [6] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang. Beyond brightening low-light images. *Int. J. Comput. Vis.*, 129:1013–1037, 2021. [2](#), [8](#)
- [7] S. Zheng and G. Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *Winter Conf. Appl. Comput. Vis.*, pages 581–590, 2022.
- [8] C. Chen, Q. Chen, M. Do, and V. Koltun. Seeing motion in the dark. In *Int. Conf. Comput. Vis.*, pages 3185–3194, 2019. [2](#)
- [9] H. Jiang and Y. Zheng. Learning to see moving objects in the dark. In *Int. Conf. Comput. Vis.*, pages 7324–7333, 2019. [2](#)
- [10] D. Triantafyllidou, S. Moran, S. McDonagh, S. Parisot, and G. Slabaugh. Low light video enhancement using synthetic data produced with an intermediate domain mapping. In *Eur. Conf. Comput. Vis.*, pages 103–119, 2020.
- [11] F. Zhang, Y. Li, S.i You, and Y. Fu. Learning temporal consistency for low light video enhancement from single images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4967–4976, 2021.
- [12] R. Wang, X. Xu, C. Fu, J. Lu, B. Yu, and J. Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Int. Conf. Comput. Vis.*, pages 9700–9709, 2021. [2](#)
- [13] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo. Toward fast, flexible, and robust low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5637–5646, 2022. [2](#), [8](#)
- [14] X. Xu, R. Wang, C. Fu, and J. Jia. Snr-aware low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17714–17724, 2022. [1](#), [2](#)
- [15] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 × 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits*, 43(2):566–576, 2008. [1](#)
- [16] C. Posch, D. Matolin, and R. Wohlgenannt. A qvga 143 db dynamic range frame-free pwm image sensor with lossless pixel-level video compression and time-domain cds. *IEEE J. Solid-State Circuits*, 46(1):259–275, 2010. [4](#)
- [17] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, et al. 5.10 a 1280 × 720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86 μm pixels, 1.066 gepps readout, programmable event-rate controller and compressive data-formatting pipeline. In *IEEE Int. Solid-State Circuits Conf.*, pages 112–114, 2020.
- [18] T. Delbruck, C. Li, R. Graca, and B. Mcreynolds. Utility and feasibility of a center surround event camera. In *IEEE Int. Conf. Image Process.*, pages 381–385, 2022.
- [19] G. Taverni, D. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. Bello, and T. Delbruck. Front and back illuminated dynamic and active pixel vision sensors comparison. *IEEE Trans. Circuits Syst. II*, 65(5):677–681, 2018.
- [20] Y. Suh, S. Choi, M. Ito, J. Kim, Y. Lee, J. Seo, H. Jung, D. Yeo, S. Namgung, J. Bong, et al. A 1280 × 960 dynamic vision sensor with a 4.95-μm pixel pitch and motion artifact minimization. In *IEEE Int. Symp. Circuits Syst.*, pages 1–5, 2020. [1](#)
- [21] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger. Interacting maps for fast visual interpretation. In *Proc. Int. Jt. Conf. Neural Netw.*, pages 770–776, 2011. [2](#), [3](#)
- [22] H. Kim, A. Handa, R. Benosman, S. Ieng, and A. Davison. Simultaneous mosaicing and tracking with an event camera. *IEEE J. Solid-State Circuits*, 43:566–576, 2008.
- [23] P. Bardow, A. Davison, and S. Leutenegger. Simultaneous optical flow and intensity estimation from an event camera. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 884–892, 2016.
- [24] Gottfried Munda, Christian Reinbacher, and Thomas Pock. Real-time intensity-image reconstruction for event cameras using manifold regularisation. *Int. J. Comput. Vis.*, 126:1381–1393, 2018. [2](#), [3](#)
- [25] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3857–3866, 2019. [2](#), [3](#), [6](#)
- [26] L. Wang, Y. Ho, K. Yoon, et al. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10081–10090, 2019.
- [27] W. Weng, Y. Zhang, and Z. Xiong. Event-based video reconstruction using transformer. In *Int. Conf. Comput. Vis.*, pages 2563–2572, 2021. [2](#), [6](#), [7](#)
- [28] Y.o Zou, Y. Zheng, T. Takatani, and Y. Fu. Learning to reconstruct high speed and high dynamic range videos from events. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2024–2033, 2021. [2](#), [3](#)
- [29] P. Cadena, Y. Qian, C. Wang, and M. Yang. Spade-e2vid: Spatially-adaptive denormalization for event-based video reconstruction. *IEEE Trans. Image Process.*, 30:2488–2500, 2021. [6](#)
- [30] L. Zhu, X. Wang, Y. Chang, J. Li, T. Huang, and Y. Tian. Event-based video reconstruction via potential-assisted spiking neural network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3594–3604, 2022. [2](#)
- [31] S. Liu and P. Dragotti. Sensing diversity and sparsity models for event generation and video reconstruction from events. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023. [2](#), [3](#)

- [32] L. Wang, T. Kim, and K. Yoon. Eventsr: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8315–8325, 2020. [2](#)
- [33] S. Zhang, Y. Zhang, Z. Jiang, D. Zou, J. Ren, and B. Zhou. Learning to see in the dark with events. In *Eur. Conf. Comput. Vis.*, pages 666–682, 2020. [2](#), [7](#)
- [34] F. Paredes-Vallés and G. de Croon. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3446–3455, 2021. [2](#), [3](#), [4](#), [6](#)
- [35] H. Rebecq, D. Gehrig, and D. Scaramuzza. Esim: an open event camera simulator. In *IEEE Int. Conf. Robot.*, pages 969–982, 2018. [2](#)
- [36] Y.g Hu, S. Liu, and T. Delbruck. v2e: From video frames to realistic dvs events. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1312–1321, 2021. [2](#), [4](#), [7](#)
- [37] A. Zhu, Z. Wang, K. Khant, and K. Daniilidis. Eventgan: Leveraging large scale image datasets for event cameras. In *IEEE Int. Conf. Comput.*, pages 1–11, 2021.
- [38] S. Lin, Y. Ma, Z. Guo, and B. Wen. Dvs-voltmeter: Stochastic process-based event simulator for dynamic vision sensors. In *Eur. Conf. Comput. Vis.*, pages 578–593, 2022. [2](#)
- [39] J. Han, C. Zhou, P. Duan, Y. Tang, C. Xu, C. Xu, T. Huang, and B. Shi. Neuromorphic camera guided high dynamic range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1730–1739, 2020. [2](#)
- [40] S. Tulyakov, A. Bochicchio, D. Gehrig, S. Georgoulis, Y. Li, and D. Scaramuzza. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17755–17764, 2022. [2](#)
- [41] D. Coltuc, P. Bolon, and J. Chassery. Exact histogram specification. *IEEE Trans. Image Process.*, 15(5):1143–1152, 2006. [2](#)
- [42] H. Ibrahim and N. Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum.*, 53(4):1752–1758, 2007.
- [43] T. Arici, S. Dikbas, and Y. Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. Image Process.*, 18(9):1921–1935, 2009.
- [44] T. Celik and T. Tjahjadi. Contextual and variational contrast enhancement. *IEEE Trans. Image Process.*, 20(12):3431–3441, 2011. [2](#)
- [45] D. Jobson, Z. Rahman, and G. Woodell. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.*, 6(3):451–462, 1997. [2](#)
- [46] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X.o Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2782–2790, 2016.
- [47] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 26(2):982–993, 2016. [2](#)
- [48] C. Zheng, D. Shi, and W. Shi. Adaptive unfolding total variation network for low-light image enhancement. In *Int. Conf. Comput. Vis.*, pages 4439–4448, 2021. [2](#)
- [49] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P Srinivasan, and J. Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16190–16199, 2022.
- [50] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5901–5910, 2022. [2](#), [8](#)
- [51] S. Liu, M.g Gao, V.y John, Z. Liu, and E. Blasch. Deep learning thermal image translation for night vision perception. *ACM Trans. Intell. Syst. Technol.*, 12(1):1–18, 2020. [2](#)
- [52] F. Luo, Y. Li, G. Zeng, P. Peng, G. Wang, and Y. Li. Thermal infrared image colorization for nighttime driving scenes with top-down guided attention. *IEEE Trans. Intell. Transp. Syst.*, 23(9):15808–15823, 2022. [2](#)
- [53] M. Ashiba, M. Tolba, A. El-Fishawy, and F. El-Samie. Hybrid enhancement of infrared night vision imaging system. *Multimed. Tools Appl.*, 79:6085–6108, 2020. [2](#)
- [54] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou. Llvip: A visible-infrared paired dataset for low-light vision. In *Int. Conf. Comput. Vis.*, pages 3496–3504, 2021. [2](#)
- [55] L. Liu, J. An, J. Liu, S. Yuan, X. Chen, W. Zhou, H. Li, Y. Wang, and Q. Tian. Low-light video enhancement with synthetic event guidance. In *AAAI Conf. Artif. Intell.*, volume 37, pages 1692–1700, 2023. [2](#)
- [56] J. Liang, Y. Yang, B. Li, P. Duan, Y. Xu, and B. Shi. Coherent event guided low-light video enhancement. In *Int. Conf. Comput. Vis.*, pages 10615–10625, 2023.
- [57] Yu Jiang, Yuehang Wang, Siqi Li, Yongji Zhang, Minghao Zhao, and Yue Gao. Event-based low-illumination image enhancement. *IEEE Trans. Multimedia*, 2023.
- [58] Q. Wang, H. Jin, H. Su, and Z. Xiao. Event-guided attention network for low light image enhancement. In *Proc. Int. Jt. Conf. Neural Netw.*, pages 1–8, 2023. [2](#)
- [59] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 989–997, 2019. [5](#)
- [60] D. Gehrig, A. Loquercio, K. Derpanis, and D. Scaramuzza. End-to-end learning of representations for asynchronous event-based data. In *Int. Conf. Comput. Vis.*, pages 5633–5643, 2019. [6](#)
- [61] B. Crawford. Visual adaptation in relation to brief conditioning stimuli. *Proc. Royal Soc. B P Roy Soc B-Biol Sci*, 134(875):283–302, 1947. [6](#)
- [62] K. R Boff, L. Kaufman, and J. Thomas. *Handbook of perception and human performance*, volume 1. 1986. [6](#)
- [63] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In *Int. Conf. Comput. Vis.*, pages 0–0, 2019. [6](#)
- [64] X. Shi, Z. Chen, H. Wang, D. Yeung, Wa. Wong, and Wa. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inform. Process. Syst.*, 28, 2015. [6](#)

- [65] Y. Wang, M. Long, J. Wang, Z. Gao, and P. Yu. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Adv. Neural Inform. Process. Syst.*, 30, 2017. 6
- [66] R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 586–595, 2018. 7
- [67] W. Lai, J. Huang, O. Wang, E. Shechtman, E. Yumer, and M. Yang. Learning blind video temporal consistency. In *Eur. Conf. Comput. Vis.*, pages 170–185, 2018. 7
- [68] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 2017. 7
- [69] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7
- [70] M.s Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robot. Autom.*, 6(3):4947–4954, 2021. 7
- [71] A Zhu, D Thakur, T Özaslan, B Pfrommer, V Kumar, and K Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robot. Autom.*, 3(3):2032–2039, 2018. 7
- [72] Ling Gao, Yuxuan Liang, Jiaqi Yang, Shaoxun Wu, Chenyu Wang, Jiaben Chen, and Laurent Kneip. Vector: A versatile event-centric benchmark for multi-sensor slam. *IEEE Robot. Autom.*, 7(3):8217–8224, 2022. 7
- [73] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 7
- [74] S. Wang, J. Zheng, H. Hu, and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.*, 22(9):3538–3548, 2013. 7, 8
- [75] A. Mittal, R. Soundararajan, and A. Bovik. Making a “completely blind” image quality analyzer. *IEEE Sign. Process. Letters*, 20(3):209–212, 2012. 8
- [76] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang. Perceptual quality assessment of smartphone photography. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3677–3686, 2020. 8