

LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising

Yuxing Duan¹

¹National Key Lab of Multispectral Information Intelligent Processing Technology

¹Huazhong University of Science and Technology

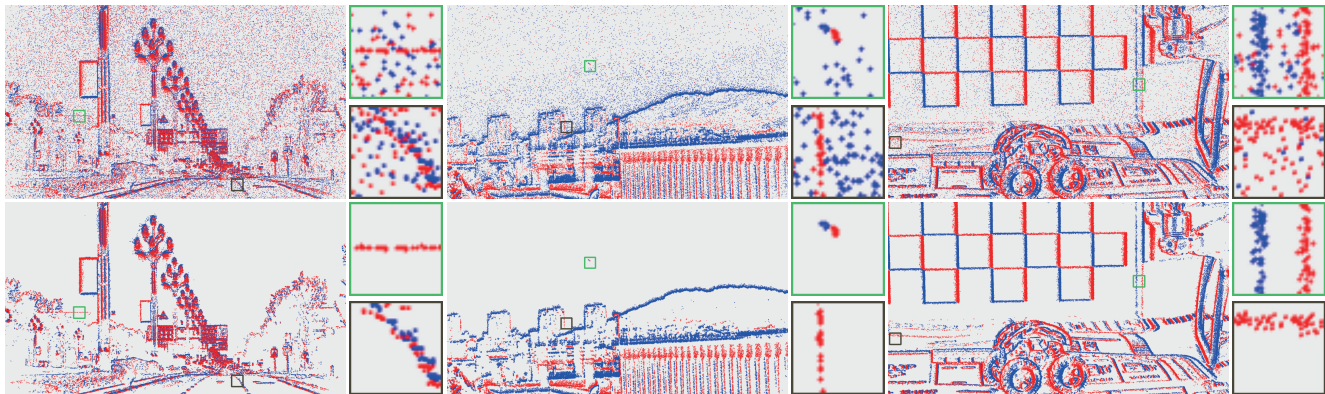


Figure 1. Illustration of the paired raw/denoised event datasets. The proposed real-world event denoising datasets contribute to addressing the inevitable Background Activity (BA) noise encountered by event cameras application under varying camera settings and illumination.

Abstract

Event camera has significant advantages in capturing dynamic scene information while being prone to noise interference, particularly in challenging conditions like low threshold and low illumination. However, most existing research focuses on gentle situations, hindering event camera applications in realistic complex scenarios. To tackle this limitation and advance the field, we construct a new paired real-world event denoising dataset (LED), including 3K sequences with 18K seconds of high-resolution (1200*680) event streams and showing three notable distinctions compared to others: diverse noise levels and scenes, larger-scale with high-resolution, and high-quality GT. Specifically, it contains stepped parameters and varying illumination with diverse scenarios. Moreover, based on the property of noise events inconsistency and signal events consistency, we propose a novel effective denoising framework(DED) using homogeneous dual events to generate the GT with better separating noise from the raw. Furthermore, we design a bio-inspired baseline leveraging Leaky-Integrate-and-Fire (LIF) neurons with dynamic thresholds to realize accurate denoising. The experimental results demonstrate that the remarkable performance of the proposed approach on different datasets. The dataset and code are at <https://github.com/Yee-Sing/led>.

1. Introduction

Event cameras possess unique imaging advantages and are demonstrating tremendous potential for various applications [4, 11, 33, 37]. Unlike traditional frame cameras using integration sampling [46], event camera is differential sampling [29], evading the restriction of exposure period. However, due to this sampling mechanism, random fluctuations in analog signals can easily form noise events. In contrast to the dominance of signal components in images, the noise and signal in event cameras are encoded with same magnitude, easily disturbing inherent structural features, such as lane lines shown in Fig. 1. BA noise detrimentally impact on subsequent tasks like reconstruction [36, 39], motion estimation [40, 41] and event-based object detection [5]. Therefore, event denoising becomes a fundamental issue, notably in the booming event-based vision community.

Paired data is a key issue for event denoising in the deep-learning era, thus an intuitive idea is to construct paired data through simulators [17, 22, 38]. However, the image-based simulation cannot be equivalent to reality because of the domain gap. To obtain the ‘clean part’ from the real-world event, multi-modal data was incorporated to help paired event generation. APS images gradient combined with IMU are used to indicate the event probability within frame period [2]. By jointly estimating the motion parameters with

Datasets	Sequences	Capture/s	Source	Noise level	Illumination	Resolution	Paired
RGB-DAVIS [45]	20	122	Camera	✗	Daytime	180*190	✗
EventNFS [10]	100	4238	Screen Recording	✗	Daytime	224*124	✗
E-MLB [8]	1200	7300	Camera	4	Daytime to Nighttime	346*260	✗
DVSNOISE20 [2]	48	807	Camera	✗	Daytime	346*260	✓
ED-KoGTL [1]	4	15	Camera	✗	Daytime to Nighttime	346*260	✓
DVSCLEAN [12]	144	55	Synthesis	2	Daytime	1280*720	✓
LED	3000	18000	Camera	4	Daytime to Nighttime	1200*680	✓

Table 1. Summary of existing event denoising datasets.

APS images and event streams [9, 45], the temporal mismatch can be alleviated. Constructing a fixed-track setup [1], the events under various illuminations can be referred to APS images in good lighting. Essentially, these methods align with events via transformed multi-mode features. Unfortunately, this heterogeneous aided information cannot fully match with the event, and their effectiveness is unavailable in situations such as low light and motion blur. Hence, few datasets have considered practical scenes where BA noise is easily deteriorated by camera setting and illumination, appearing as complex and nonuniform distributions.

To address this issue, we construct a large-scale, high-quality paired dataset LED for event denoising. LED has three distinct advantages. Firstly, it contains diverse noise levels, covering complex situations under various camera settings and illumination. Secondly, LED includes such as outdoor and indoor scenarios, with abundant objects. Last but not least, LED is collected with the current largest-scale and high-resolution. Moreover, inspired by the multi-sampling images denoising methods [21, 28, 31, 34], we are the first to explore event multi-sampling for denoising. Based on the homogeneous data, we proposed a dual-events denoising framework that can generate higher-quality GT by more accurately separating BA noise from the raw.

The contributions of this paper can be summarized as:

(1) We provide a large-scale real-world paired dataset for event denoising. So far, **LED** is the largest paired real-world event denoising dataset (3000 sequences) and high-resolution (1200*680) with diverse noise levels and scenes.

(2) We propose a novel dual events denoising framework **DED** based on noise inconsistency. We provide detailed analysis to show that DED can better separate mixed signals and noisy event streams across various camera threshold parameters, light conditions and motion patterns.

(3) We introduce a novel baseline **DTSNN** for event denoising based on the spiking neural network, which utilizes the learnable dynamical spike threshold of the LIF neuron to accurately denoising. Extensive experiments on different real datasets verify the superiority of the proposed method.

2. Related work

Event Denoising Datasets. Currently, the datasets exclusively designed for event denoising with less attention com-

pared to the task-specific ones in the event-based community. In Table 1, we provide an all-sided summary of existing datasets, highlighting the scarcity of real paired datasets.

To address the lack of explicit labels, some paired datasets can be conveniently generated via simulators such as DVSCLEAN [12] and DND21 [20]. However, due to the gap between simulation and real, these simulators fail to fully reflect the actual situation. Other works focus on real paired event denoising data. DVSNOISE20 [2] captures data in 16 pure background scenes. RGB-DAVIS [45] provides 20 indoor/outdoor sequences under good lighting conditions. ED-KoGTL [1] constructs indoor data with four illumination levels under constant trajectory. These paired datasets have achieved favorable results, primarily attributable to the gradient cues extracted from APS images. However, the heterogeneity between images and events limits the availability of diverse lighting conditions, motion patterns, and scenarios for constructing paired data.

Event-based Denoising. Research on event denoising can be categorized into two approaches: filtering-based methods and deep learning methods. The former relies on manual priors to design discriminative models for noise removal, such as utilizing density distinction [15, 20, 30, 50] or motion continuity distinction [3, 44, 47] and motion compensation [6, 45] with CM framework [16] in the spatiotemporal domain. However, the validity of this prior knowledge is sensitive to different signal/noise event distributions, limiting their denoising accuracy. For learning-based methods, such as the CNN-based EDNCNN [2], the PointNet-based AEDNet [12] and [1] GNN-transformer were proposed successively. Additionally, the event denoising and super-resolution tasks are combined in [10].

Multi-Sampling Image Denoising. The multi-sampling principle for image denoising has been extensively studied. Techniques like long exposure time or burst captures benefit the direct acquisition of clear data with increased SNR in static scenes. Further, the multi-sampling was explored to the paradigm of indirectly recovering noise-free data. Not only using paired noisy observations to learn the mapping from different noisy samples to clear counterpart [27, 32], but also employing spatial sub-sampling on a single noisy image to learn the mapping from adjacent noisy samples to GT [23, 25]. Evidently, these denoising approaches imply

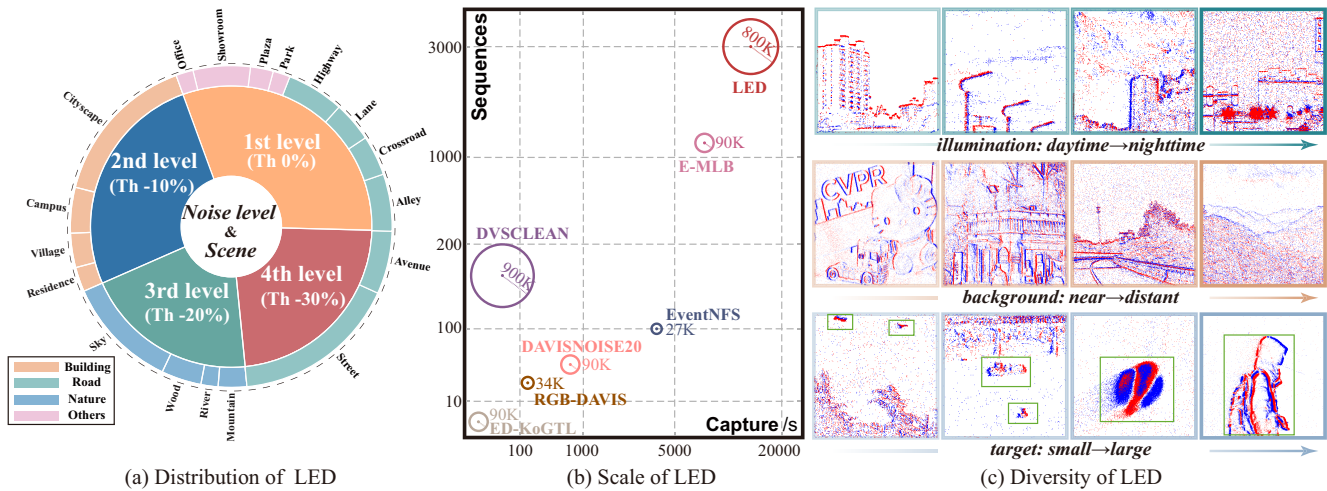


Figure 2. Illustration of the proposed dataset LED. (a) Distribution of noise level and scene of the proposed dataset. (b) Our proposed LED outperforms others in terms of sequences, capture, and resolution (Circles with numbers to indicate). (c) LED collects diverse event streams across various conditions of illumination, depth of field, and target scale.

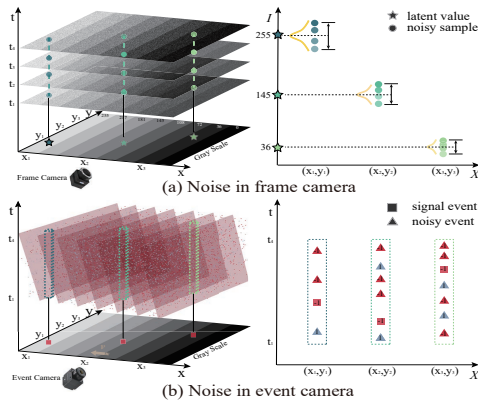


Figure 3. The substantial noise distinctions between frame/event camera. (a) For a stationary grayscale chart, the frame camera sequentially acquires noisy samples, fluctuating around a latent value within a specific distribution. (b) In a horizontal moving case, the event camera outputs binary signal serially, comprising signal events from motion gradient edges and spurious BA noise.

that signal resampling could facilitate suppressing random circuit noise and approximating the expectation value.

3. The High-quality Paired LED Dataset

3.1. Dataset Statistics and Features

The primary factors affecting BA noise in event cameras can be categorized into intrinsic camera settings and extrinsic lighting conditions. By setting different camera thresholds, we can selectively record events with varying noise levels. Generally, a lower threshold means higher sensitivity but also increases the likelihood of BA noise. In contrast to fixed parameters in other datasets, we use four level thresholds to collect events, starting from the default value of 0% and gradually decreasing to -10%, -20%, and -30%,

as shown in Fig. 2 (a). To cope with the challenges posed by complex lighting conditions, data collection is from daytime to nighttime, encompassing a transition from high to low illumination and a mixture of natural light with artificial light sources. Thus, the impact of external conditions on the event stream distribution can also be controlled.

Due to the difficulty of constructing real paired datasets, the development of existing datasets remains limited, as shown in Table 1. In this work, we used a vehicle-mounted or pan-tilt equipped with event cameras featuring a 16mm lens. Totally, we collected approximately 5 hours of event stream, including 3K sequences. The composition of noise levels and scenes is depicted in Fig. 2 (a), with roughly equal amounts collected during daytime and nighttime. Nearly 20 typical scenes were captured, including urban buildings, traffic roads, natural scenery, indoor exhibits, etc.

Our dataset goes beyond just the number of sequences and data volume, primarily aiming to consider the intricate influences of relevant factors on noise. Thus, we endow LED with diversity to reflect the impact of objects and environments on the event, which helps cover the cases of densely and non-uniformly distributed noisy situations in reality both spatially and temporally. In Fig. 2 (c), we exhibit that LED diversity is not only embodied in the richness of noise and illumination levels but also features such as depth of field and scene semantics. Meanwhile, high-resolution imaging aids the event cameras in capturing spatial information more precisely. Unfortunately, most existing ones with low resolution may hinder advanced visual tasks. Therefore, we collected data with 1200*680 resolution(after cropping), achieving high spatiotemporal resolution. Additionally, LED includes abundant objects of various scales, including aerial targets, pedestrians, and vehicles, making it well-suited for downstream tasks.

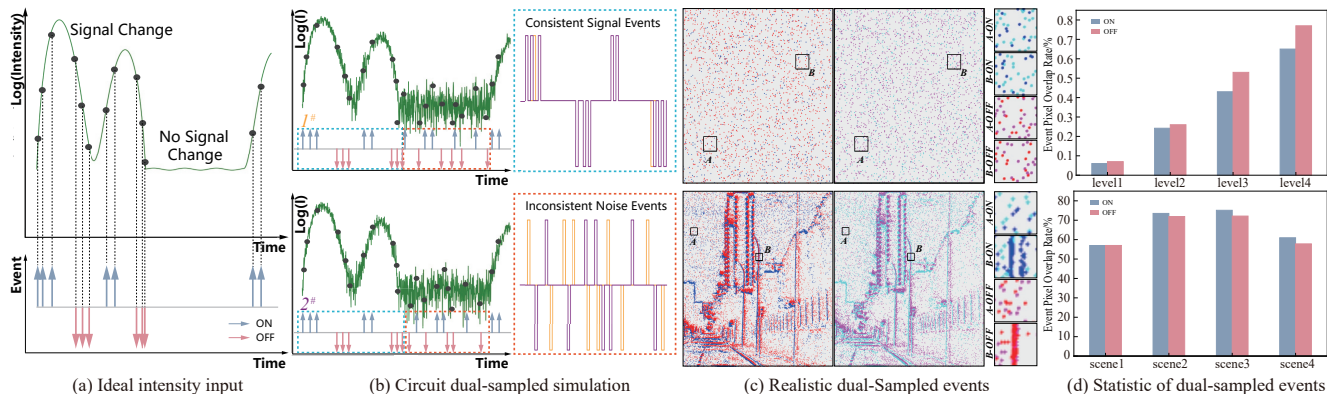


Figure 4. Illustration of event camera dual-sampling analysis: (a) Given an ideal intensity input, which includes varying and steady stages, generates a series of signal events. (b) The twice samplings of the same input in the circuit model, both generate additional BA noise in previously steady stages, resulting in consistency discrepancies between noise events and signal events. (c) The visualization results of the actual dual-sampled events cumulative frame demonstrate the misalignment of inconsistent dual-sampled noise events and the alignment of coexisting signal events (two color groups indicating the respective polarities of the dual events). (d) The statistical results of the two tests also prove a low overlap rate between dual-sampled event pixels and a much higher one where signal events are present.

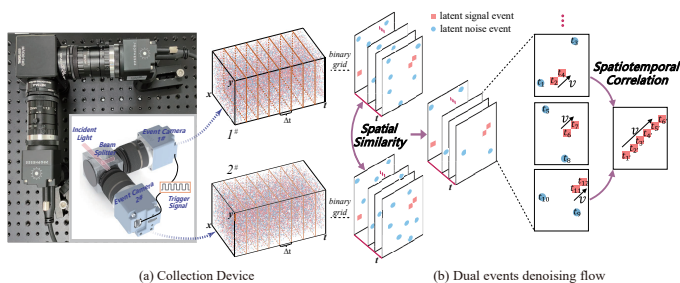


Figure 5. Overview of the dual events denoising framework. (a) Our collection device consists of two identical EVK4s forming a co-axial system with a 1:1 beamsplitter. (b) We first perform spatial similarity processing to retain the consistent parts, followed by sequentially spatiotemporal correlation constraints to remove the residual small amount of noise event from the previous step.

3.2. Dual Events Denoising Framework

Given a noisy event stream, the key lies in properly obtaining paired GT. In this section, we explore utilizing multi-sampling to achieve event stream denoising.

Background. Multi-sampling denoising methods are commonly used in image denoising. For frame images, averaging the results of temporal multi-samples can approximate the true value for each pixel owing to multi-samples smoothing effect on random noise components, as shown in Fig. 3 (a). For event stream, due to their dynamic sampling and the binary values, temporal multi-samples may contain both signal events and BA noise, as illustrated in Fig. 3 (b), directly taking the average of multi-samples cannot yield true signal value. Although the noise forms differ between the two modalities, they ultimately originate from the inherent noise in the circuitry analog signal, while resampling fundamentally helps suppress the random noise source.

Analysis. This concept inspired us to utilize multi-sampling

for event denoising, particularly when considering the binary state of events (presence or absence) where just two samplings are required to potentially identify noise. Thus, we initially investigated the noise inconsistency at the pixel level. Based on the DVS working principle [29], we build a circuit model in the Cadence [42] platform to simulate the dual-sampling with twice identical intensity input. The noisy analog signal converted from input generates BA noise at the originally steady stage compared to the ideal case in Fig. 4 (a). Clearly, the two sets of noise exhibit inconsistency, while the signals demonstrate high consistency, causing distinct alignment differences between the two cases as shown in the close-up of Fig. 4 (b).

To further explore the inter-class differences between noise and signal events in real-world dual-sampled events stream, we established a collection device as shown in Fig. 5 (a), to construct genuinely spatiotemporal dual-sampling. Two experimental conditions to validate the aforesaid circuit simulation phenomenon are a stationary camera and a moving one respectively. In the former, pure BA noise events was obtained, while the latter captured mixed events containing both noise and signal. In static tests, although at the highest noise level (-30% threshold), both two sets of events exhibit significant spatiotemporal disparity. As observed in the right zoomed-in patches at the upper row in Fig. 4 (c), the global spatial misalignment of dual BA noises events was evident. The overlap rate of the pixels triggering event is also to be less than 1%, as shown in the upper-row histogram of Fig. 4 (d). Not unexpectedly, in dynamic tests, regions with BA noise events remain inconsistency, while the coexistent signal events showing spatiotemporal alignment as depicted in lower-row counterpart, leading to a dramatic increase in overlap rate compared to the pure BA noise cases, as shown in the bottom histogram of Fig. 4 (d).

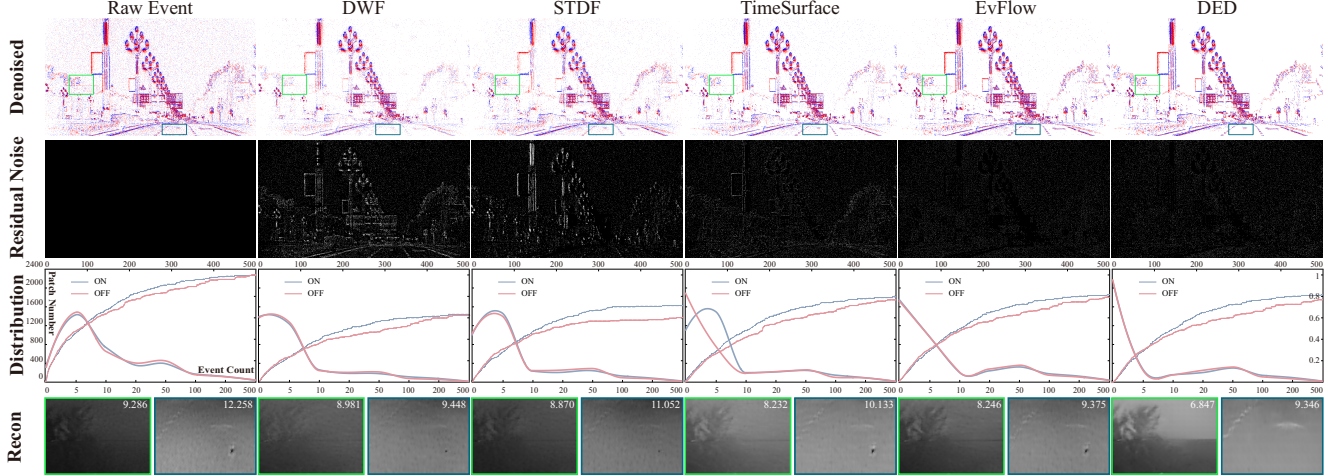


Figure 6. Analysis of different event denoising results on our dataset. From left to right, the first column is the raw events, and the remaining five columns represent different methods, namely DWF, STDF, TimeSurface, EvFlow, and the proposed DED. From top to bottom, the first row shows the denoised results, the second row is the residual noise, the third row denotes the statistical distribution of the denoised results and the last row represents the intensity image reconstruction corresponding to the zoomed region denoised events.

Formulation. These results support the insight that the consistency discrepancies between dual-sampled events could naturally help distinguish noise. Therefore, we develop a denoising framework with the dual events stream, called DED. In DED framework, given the spatiotemporal synchronized dual event streams Y_1 and Y_2 , they are composed of common latent signal X and respective noise streams N_1 and N_2 . The noise model can be defined as follows:

$$Y_1 = X + N_1, \quad Y_2 = X + N_2. \quad (1)$$

Because of the consistency of signal events and the inconsistency of noise events in Y_1 and Y_2 , after binary grid operation of event stream according to certain temporal window Δt , we can perform spatial similarity to process them:

$$X^* = \|\|X + N_1\| \circ \|\|X + N_2\|\|_1, \quad (2)$$

where \circ denotes the Hadamard product used to obtain the similar part X^* of two groups of raw event binary frames. Due to the probability of a few noises occurring simultaneously in dual-sampling, X^* may still contain some much-isolated noise compared to the raw. As shown in Fig. 5 (b), for finer denoising, we further utilize the spatiotemporal correlation of the signal event stream to remove the residual noise from the previous step. Specifically, we accumulate the events from the relevant spatiotemporal range in corresponding X^* within several consecutive Δt to form a spatiotemporal relationship set. The presence of these few isolated noise events, which are randomly triggered in the spatiotemporal domain, disrupts the inherent spatiotemporal correlation of signal reflecting a certain motion model. Therefore, to leverage this spatiotemporal continuity, we transform it into the following minimization problem:

$$\min \frac{1}{N-1} \sum_{i=1}^N |e_{\mathbf{x}}^{t_{i+1}} - e_{\mathbf{x}}^{t_i}|, \quad t_i \in n\Delta t, \quad \mathbf{x} \in \Omega, \Omega \in X^*, \quad (3)$$

where $e_{\mathbf{x}}^{t_i}$ denotes the i_{th} occurring event according to the sorted timestamp, \mathbf{x} means the event spatial coordinates, n is the number of selected consecutive time windows, Ω represents the candidate spatial neighborhood, and N is total events number in Ω . Additionally, the aforesaid processing is all performed on the two channels of event polarity.

3.3. Evaluation and Discussion

The quality of GT is critical for a real paired event dataset. However, it is challenging to objectively evaluate the denoising results since directly obtaining clean data is unpractical. To fairly assess the denoising effects of different methods, we conducted a comprehensive evaluation.

Fig. 6 presents comparative results of representative event denoising methods could be used for GT generation: spatiotemporal density filter-based (DWF[20], STDF[15]), smoothness optimization-based (TimeSurface[26]), and motion estimation-based (EvFlow[44]). The first and second rows display the denoised and residual noise visualization, respectively. It can be observed that DED effectively eliminates almost all the scattered BA noise with good preservation of the inherent structured feature, while the others exhibit residual noise and varying degrees of signal event damage. Statistical analysis on the global distribution of the denoised event is shown in the third row. Compared to others, DED demonstrates a higher number of blank patches while maintaining a high event preservation rate, which better reflects the overall sparsity and local concentration of the ideal event stream. Moreover, better denoising often leads to higher-quality reconstruction. In the last row, the intensity images reconstructed by E2VID [39] indicate that DED accurately restores nighttime scene information, such as leaves, building contours, and road signs

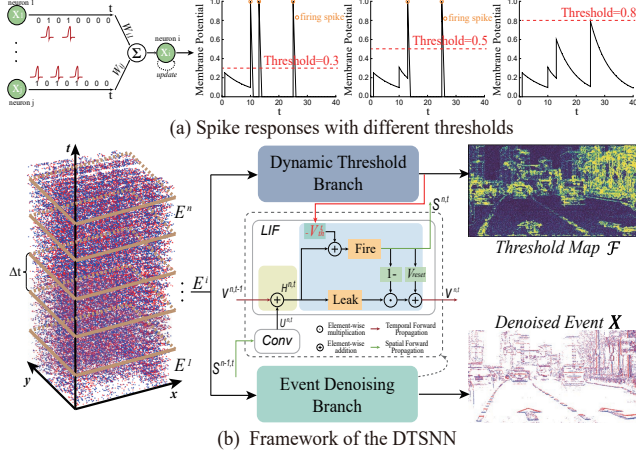


Figure 7. An overview of the DTSNN model. (a) The dynamic threshold mechanism of LIF neuron. (b) The model with two branches consisting of DTB and EDB. Please refer to our supplementary material for more details about network architecture.

with lower NIQE [35], which further indirectly validates the accuracy of denoised results and also confirms the superior effect of the processed data on reconstruction.

4. DTSNN for Event Denoising

Compared to artificial neural networks(ANN), spiking neural network (SNN) have lower precision but are inherently suited for processing event-driven data due to their and information transfer mode temporal dynamics, gradually gaining prominence in event-based tasks [48, 49, 51, 52].

Dynamic Threshold Mechanism. Adjustable spiking neuron offers enhanced biological plausibility [7, 13, 43]. Indeed, the neuron firing threshold, similar to the event camera threshold parameter, functions as controlling the output. Intuitively, the dynamic threshold (DT) mechanism can adapt to more complicated situations since a high threshold could suppress noisy input when noise dominates, while a low one helps to sensitively preserve the expected signals. Inspired by this, we transformed the fixed threshold into dynamic ones based on the LIF neuron model [19] to mimic this anisotropic biomechanism. The tendency of decreasing spike quantity of the postsynaptic neurons responses with an increasing threshold as illustrated in Fig. 7(a).

DTSNN Model. Therefore, we propose a fully SNN (all synaptic operations are SNN-based) incorporating learnable DT for event denoising, namely DTSNN. As illustrated in Fig. 7 (b), DTSNN consists of a dynamic threshold branch (DTB) and event denoising branch (EDB). The former dynamically generates a threshold map based on the successive event input, indicating the approximate spatial regions of signal or noise. This map is then passed to the latter, controlling the spiking process of LIF neurons in the last layer,

which is depicted in Fig. 7 (b) and can be formulated as:

$$\begin{cases} H^{n,t} = V^{n,t-1} + U^{n,t} \\ S^{n,t} = Hea(H^{n,t} - \mathcal{F}(V_{th}^t)) \\ V^{n,t} = V_{reset}S^{n,t} + \frac{1}{\tau}H^{n,t} \odot (1 - S^{n,t}) \end{cases} \quad (4)$$

where n and t denote the layer number and time step respectively. $H^{n,t}$ is the membrane potential which is produced by coupling the spatial feature $U^{n,t}$ and the temporal input $V^{n,t-1}$. The DT map $\mathcal{F}(V_{th}^t)$ determines whether the output spiking matrix $S^{n,t}$ should be fired or stay as zero, formulating the final denoised event matrix \mathbf{X} . $Hea(x) = 1$ represents the Heaviside step function when $x \geq 0$, otherwise $Hea(x) = 0$, and \odot means a element-wise multiplication.

Notably, both two branches receive the same consecutive data as input, which is a binary event frame within a certain time window of Δt . The denoising branch is supervised by the GT of signal events within Δt , while the threshold map label of the dynamic threshold branch comes from the signal events over a longer temporal period.

5. Experiments

5.1. Experiments Setup

Implementation Details. We use a combined loss consisting of the L1-norm and BCE to train the proposed network. Our models are implemented with the open-source framework SpikingJelly [14], using NVIDIA A100 GPU. The Adam optimizer is employed with a batch size of 8 and a learning rate of 0.002. A fixed threshold of 0.5 is chosen for the rest of the neurons except at the last membrane potential layer. Besides, the reset value V_{reset} and the membrane time constant τ of all LIF neurons are set to 0 and 2, respectively. **Datasets.** We conduct quantitative experiments on existing paired datasets. Due to the enormous scale of billions of events in LED, training them all would be extremely time-consuming and resource-intensive. Therefore, 600 sequences were randomly selected for training with each consisting of consecutive 10 segments, and 60 sequences were randomly chosen for testing. To further evaluate the performance of the discrepancy of different datasets, we qualitatively test on typical public datasets DSEC [18] and E-MLB [8]. For event-based denoising methods, besides the methods in Sec.3.3 we select the representative supervised denoising methods including EDnCNN [2] and AEDNet[12]. **Evaluation Methods.** We use an index of event denoising accuracy $DA = \frac{1}{2}(\frac{TP}{GP} + \frac{TN}{GN})$ to measures the denoising performance on LED, where TP, TN, GP and GN are the number of true signal, true noise, total signal, and total noise respectively, forming two parts: signal retain (SR) and noise removal (NR). The metrics on other paired datasets are adopted from their proposed ones. Moreover, visualization results was qualitatively evaluated on other datasets.

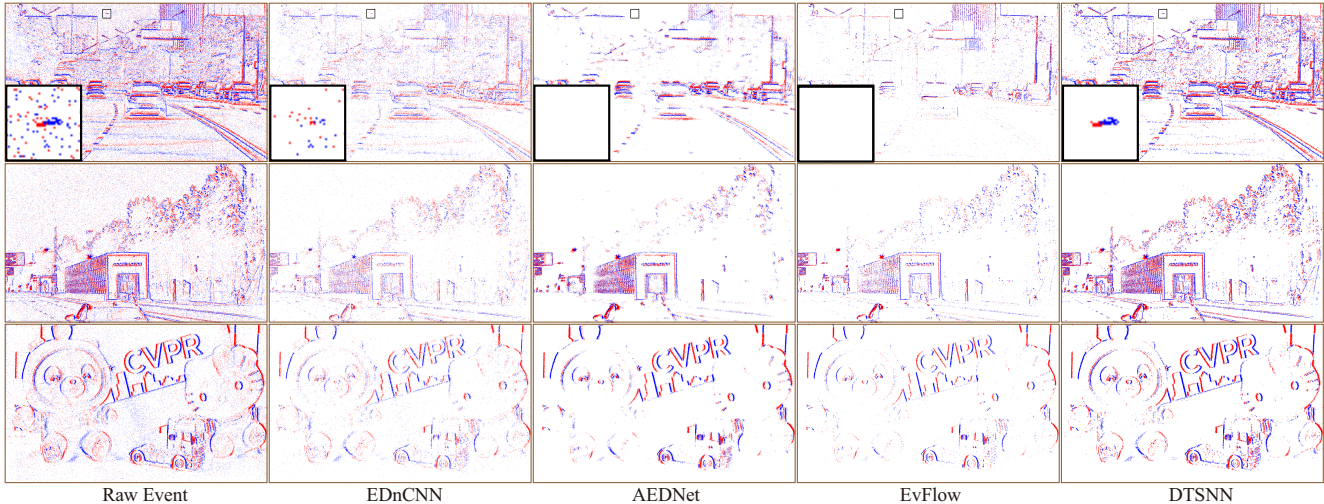


Figure 8. Visual comparisons on LED. Comparing with state-of-the-arts, DTSNN achieves excellent denoising results and it is capable of removing noise events scattered around in sky, ground, and structural contours, simultaneously better preserve scene details.

Datasets	LED			DVS CLEAN	DVS NOISE20	Average
	SR \uparrow	NR \uparrow	DA \uparrow	SNR \uparrow	RPMD \downarrow	
Knoise[24]	16.4	98.9	57.6	25.21	17.64	1.33 ms
DWF[20]	42.6	83.9	63.3	26.96	31.39	3.67 ms
STDF[15]	37.4	99.1	68.3	19.30	30.23	<u>2.04 ms</u>
TS[26]	30.6	98.3	64.5	13.98	15.25	5.76 ms
EvFlow[44]	52.3	96.4	74.4	23.74	22.50	150.5 ms
EDnCNN[2]	80.0	82.0	81.0	20.29	22.25	251.4 ms
AEDNet[12]	<u>81.2</u>	83.6	<u>82.4</u>	<u>25.58</u>	18.51	708.1 ms
DTSNN	86.0	86.5	86.2	29.26	<u>16.13</u>	8.20 ms

Table 2. Quantitative results comparison on different datasets. We mark the **best** and second best.

5.2. Quantitative Evaluation

The main quantitative results are presented in Table 2. In summary, the proposed method achieves the best overall results on three datasets, followed by AEDNet. Particularly, DTSNN balances signal retention and noise removal effectively, demonstrating the highest denoising accuracy. Meanwhile, our model also achieved faster inference speed on average among them. Notably, the runtime from 120 * 120 events indicates that the efficiency of grid-based event representation and processing is significantly higher than the manner based on a single-event level, although the latter better preserves the asynchronous property of the event.

5.3. Qualitative Evaluation

Evaluation on LED Dataset. To further evaluate the denoising performance, we compare with the qualitative results of typical methods on LED. As shown in Fig. 8, DTSNN outperforms other methods by achieving visually pleasing results, which not only effectively removes noise events in areas such as the sky, ground and building con-

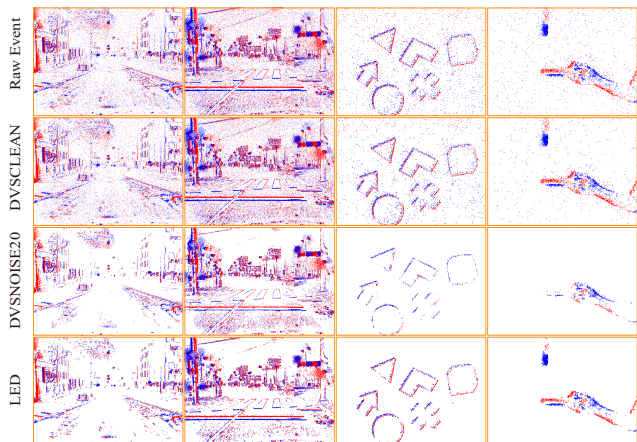


Figure 9. Illustration of the LED diversity. We train DTSNN on different datasets: DVSCLEAN, DVSNOISE20 and LED, and test on real unpaired datasets (from left to right: DSEC→E-MLB). The model trained on LED has achieved better denoising results.

tours, but also preserve the structural features of the scene and objects well. For example, the zoomed region of first row in Fig. 8 highlights that our method is the only one successfully preserving the information of small airborne target after denoising. Similarly, our method also stands out in retain lane marking located in the bottom-left of second row.

Evaluation on Public Datasets. To evaluate the generalization across different datasets, we also train DTSNN on synthetic event denoising dataset DVSCLEAN [12] and real one DVSNOISE20 [2] respectively, testing on DSEC [18] and E-MLB [8], the former is resolution of 640*480 and the latter is resolution of 346*260. As shown in Fig. 9, the model trained on DVSCLEAN performs poorly due to the huge domain gap between. DVSNOISE20 excels at removing noise effectively but inevitably loses some details. The

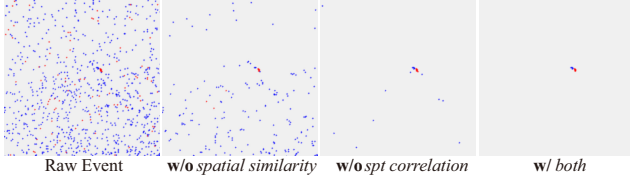


Figure 10. Ablation study of the dual events denoising framework. The first column represents the noisy raw and the remaining three columns represent the flow without spatial similarity, without spatiotemporal correlation constraint and complete DED

Architecture	Denoising Accuracy	Energy Ratio
ANN	84.5	18×
SNN	86.0	1×

Table 3. Comparison between ANN and SNN.

model trained on LED achieves satisfactory results across various scenes in these datasets which simultaneously removes noise without apparent signal loss, strongly supporting the great generalization of the proposed LED dataset.

5.4. Ablation Study and Discussion

Effectiveness of DED Framework. The DED framework aims to fully exploit the property of noise inconsistency /signal consistency in dual events and the self-correlation in a single event stream. As shown in Fig. 10, the small target signal is overwhelmed by noise events, utilizing solely spatiotemporal correlation constraints can only remove partial noise events because some of them also satisfy this relationship. Relying on the spatial similarity of dual event streams effectively disassembles the mutual relationship in the dense noise event group. Therefore, combining both procedures can achieve a more refined denoising result.

Effectiveness of Spiking Neuron. Intuitively, the inherent spatiotemporal sparsity of event streams seamlessly aligns with SNN because of the temporal information ability of spiking neurons. To compare the performance differences between SNN and ANN architectures in the event denoising tasks, the evaluation was conducted on the same structure. As shown in Table 3, the SNN based on LIF neuron achieves better denoising accuracy while maintaining 18× lower power consumption advantage compared to its ANN version (the spiking neurons replaced with ReLU).

Effectiveness of DT Module. To evaluate the effects of the proposed learnable threshold mechanism of DTSNN, we train the network with various fixed threshold neurons and dynamic threshold neurons, respectively. As can be observed from Table 4 and Fig. 11 (a), the denoising accuracy improves after introducing the DT module due to the network has the tendency to increasing spiking probability for signal pixels according to threshold prediction, and the Fig. 11 (b) illustrates DT module could assist in locating the approximate signal or noise region to retain more signal events from structural area compared to the FT model.

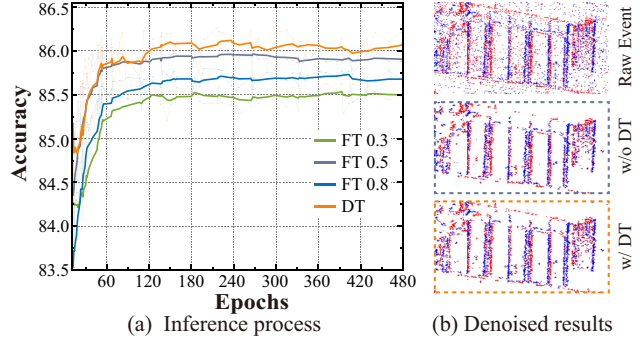


Figure 11. Illustration of the DT model Effectiveness. (a) Inference accuracy of SNN with various fixed thresholds and with DT. (b) Denoised results without DT and with DT, respectively.

Type	FTSNN (0.3 th)	FTSNN (0.5 th)	FTSNN (0.8 th)	DTSNN
Accuracy	85.5	86.0	85.7	86.2

Table 4. Inference accuracy with proposed methods implemented.

Discussion. Essentially, true events comes from intensity changes. When artificial light source variations exist, the light source itself, its coverage area and reflection from ground texture would result in relevant events. These signal events exhibit noise-like features with discontinuous structure, while the proposed DED framework considers them as signal event that should be preserved. Thus, with the supervision of such signal labels, the proposed model also tend to retain such events arise from lamp as shown in left two columns of Fig. 9 despite their visually unappealing effect.

Limitation. The DED framework may fail in extreme cases where similar part are dominated by noise events. However, such globally anomalous high-frequency noise is rare in practice. Moreover, the proposed model struggles to capture those insignificant signals attaching to structural periphery, limiting signal retain performance in metrics.

6. Conclusion

In this paper, we construct a new large-scale high-quality paired real event denoising dataset LED, which provides diverse noise levels and scenes which aims to cover various internal and external conditions. Based on the consistency difference of noise and signal events, the proposed DED framework could effectively generate high-quality GT and detailed analysis confirms our better results than others. In addition, we propose a novel baseline DTSNN for event denoising featuring with dynamic threshold mechanism of LIF neuron. The experimental results demonstrate the superiority of the proposed dataset and denoising method.

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

References

- [1] Yusra Alkendi, Rana Azzam, Abdulla Ayyad, Sajid Javed, Lakmal Seneviratne, and Yahya Zweiri. Neuromorphic camera denoising using graph neural network-driven transformers. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. [2](#)
- [2] R Baldwin, Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. Event probability mask (epm) and event denoising convolutional neural network (edncnn) for neuromorphic cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1701–1710, 2020. [1](#), [2](#), [6](#), [7](#)
- [3] Ryad Benosman, Charles Clercq, Xavier Lagorce, Sio-Hoi Ieng, and Chiara Bartolozzi. Event-based visual flow. *IEEE transactions on neural networks and learning systems*, 25(2):407–417, 2013. [2](#)
- [4] Clément Cabriel, Tual Monfort, Christian G Specht, and Ignacio Izeddin. Event-based vision sensor for fast and dense single-molecule localization microscopy. *Nature Photonics*, pages 1–9, 2023. [1](#)
- [5] Haosheng Chen, David Suter, Qiangqiang Wu, and Hanzhi Wang. End-to-end learning of object motion estimation from retinal events for event-based object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10534–10541, 2020. [1](#)
- [6] Jinze Chen, Yang Wang, Yang Cao, Feng Wu, and Zheng-Jun Zha. Progressivemotionseg: Mutually reinforced framework for event-based motion segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 303–311, 2022. [2](#)
- [7] Jianchuan Ding, Bo Dong, Felix Heide, Yufei Ding, Yunduo Zhou, Baocai Yin, and Xin Yang. Biologically inspired dynamic thresholds for spiking neural networks. *Advances in Neural Information Processing Systems*, 35:6090–6103, 2022. [6](#)
- [8] Saizhe Ding, Jinze Chen, Yang Wang, Yu Kang, Weiguo Song, Jie Cheng, and Yang Cao. E-mlb: Multilevel benchmark for event-based camera denoising. *IEEE Transactions on Multimedia*, 2023. [2](#), [6](#), [7](#)
- [9] Peiqi Duan, Zihao W Wang, Boxin Shi, Oliver Cossairt, Tiejun Huang, and Aggelos K Katsaggelos. Guided event filtering: Synergy between intensity images and neuromorphic events for high performance imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8261–8275, 2021. [2](#)
- [10] Peiqi Duan, Zihao W Wang, Xinyu Zhou, Yi Ma, and Boxin Shi. Eventzoom: Learning to denoise and super resolve neuromorphic events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12824–12833, 2021. [2](#)
- [11] Davide Falanga, Kevin Kleber, and Davide Scaramuzza. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 5(40):eaaz9712, 2020. [1](#)
- [12] Huachen Fang, Jinjian Wu, Leida Li, Junhui Hou, Weisheng Dong, and Guangming Shi. Aednet: Asynchronous event denoising with spatial-temporal correlation among irregular data. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1427–1435, 2022. [2](#), [6](#), [7](#)
- [13] Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2661–2671, 2021. [6](#)
- [14] Wei Fang, Yanqi Chen, Jianhao Ding, Zhaofei Yu, Timothée Masquelier, Ding Chen, Liwei Huang, Huihui Zhou, Guoqi Li, and Yonghong Tian. Spikingjelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40):eadi1480, 2023. [6](#)
- [15] Yang Feng, Hengyi Lv, Hailong Liu, Yisa Zhang, Yuyao Xiao, and Chengshan Han. Event density based denoising method for dynamic vision sensor. *Applied Sciences*, 10(6):2024, 2020. [2](#), [5](#), [7](#)
- [16] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3867–3876, 2018. [2](#)
- [17] Daniel Gehrig, Mathias Gehrig, Javier Hidalgo-Carrió, and Davide Scaramuzza. Video to events: Recycling video datasets for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3586–3595, 2020. [1](#)
- [18] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 2021. [6](#), [7](#)
- [19] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014. [6](#)
- [20] Shasha Guo and Tobi Delbruck. Low cost and latency event camera background activity denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):785–795, 2022. [2](#), [5](#), [7](#)
- [21] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. [2](#)
- [22] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1312–1321, 2021. [1](#)
- [23] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14781–14790, 2021. [2](#)
- [24] Alireza Khodamoradi and Ryan Kastner. $o(n)o(n)$ -space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Transactions on Emerging Topics in Computing*, 9(1):15–23, 2018. [7](#)

- [25] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2129–2137, 2019. 2
- [26] Xavier Lagorce, Garrick Orchard, Francesco Galluppi, Bertram E. Shi, and Ryad B. Benosman. Hots: A hierarchy of event-based time-surfaces for pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7):1346–1359, 2017. 5, 7
- [27] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 2
- [28] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T Barron, Dillon Sharlet, Ryan Geiss, et al. Handheld mobile photography in very low light. *ACM Trans. Graph.*, 38(6):164–1, 2019. 2
- [29] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128×128 120 db $15 \mu\text{s}$ latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576, 2008. 1, 4
- [30] Hongjie Liu, Christian Brandli, Chenghan Li, Shih-Chii Liu, and Tobi Delbruck. Design of a spatiotemporal correlation filter for event-based sensors. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 722–725. IEEE, 2015. 2
- [31] Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, and Jian Sun. Fast burst images denoising. *ACM Transactions on Graphics (TOG)*, 33(6):1–9, 2014. 2
- [32] Ali Maleky, Shayan Kousha, Michael S Brown, and Marcus A Brubaker. Noise2noise2flow: realistic camera noise modeling without clean images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17632–17641, 2022. 2
- [33] Rohit Mangalwedhekar, Nivedita Singh, Chetan Singh Thakur, Chandra Sekhar Seelamantula, Mini Jose, and Deepak Nair. Achieving nanoscale precision using neuromorphic localization microscopy. *Nature Nanotechnology*, pages 1–10, 2023. 1
- [34] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2502–2510, 2018. 2
- [35] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6
- [36] Liyuan Pan, Richard Hartley, Cedric Scheerlinck, Miaomiao Liu, Xin Yu, and Yuchao Dai. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2519–2533, 2020. 1
- [37] Etienne Perot, Pierre De Tournemire, Davide Nitti, Jonathan Masci, and Amos Sironi. Learning to detect objects with a 1 megapixel event camera. *Advances in Neural Information Processing Systems*, 33:16639–16652, 2020. 1
- [38] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Conference on robot learning*, pages 969–982, 2018. 1
- [39] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1964–1980, 2019. 1, 5
- [40] Timo Stoffregen and Lindsay Kleeman. Event cameras, contrast maximization and reward functions: An analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12300–12308, 2019. 1
- [41] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7244–7253, 2019. 1
- [42] Cadence Design Systems. Software downloads-cadence design systems. https://www.cadence.com/en_US/home/support/software-downloads.html. 4
- [43] Siqi Wang, Tee Hiang Cheng, and Meng-Hiot Lim. Ltmd: Learning improvement of spiking neural networks with learnable thresholding neurons and moderate dropout. *Advances in Neural Information Processing Systems*, 35:28350–28362, 2022. 6
- [44] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, and Hongkai Wen. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6358–6367, 2019. 2, 5, 7
- [45] Zihao W Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1609–1619, 2020. 2
- [46] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 1
- [47] Jinjian Wu, Chuanwei Ma, Leida Li, Weisheng Dong, and Guangming Shi. Probabilistic undirected graph based denoising method for dynamic vision sensor. *IEEE Transactions on Multimedia*, 23:1148–1159, 2020. 2
- [48] Man Yao, Huanhuan Gao, Guangshe Zhao, Dingheng Wang, Yihan Lin, Zhaoxu Yang, and Guoqi Li. Temporal-wise attention spiking neural networks for event streams classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10221–10230, 2021. 6
- [49] Jiqing Zhang, Bo Dong, Haiwei Zhang, Jianchuan Ding, Felix Heide, Baocai Yin, and Xin Yang. Spiking transformers for event-based single object tracking. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 8801–8810, 2022. 6
- [50] Pei Zhang, Zhou Ge, Li Song, and Edmund Y L. Neuromorphic imaging with density-based spatiotemporal denoising. *IEEE Transactions on Computational Imaging*, 2023. 2

- [51] Xiang Zhang, Wei Liao, Lei Yu, Wen Yang, and Gui-Song Xia. Event-based synthetic aperture imaging with a hybrid network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14235–14244, 2021. 6
- [52] Lin Zhu, Xiao Wang, Yi Chang, Jianing Li, Tiejun Huang, and Yonghong Tian. Event-based video reconstruction via potential-assisted spiking neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3594–3604, 2022. 6