HUE Dataset: High-Resolution Event and Frame Sequences for Low-Light Vision

Burak Ercan[*,1], Onur Eker[*,1,2], Aykut Erdem[3,4], and Erkut Erdem[1]

[1] Hacettepe University, Computer Engineering Department
[2] HAVELSAN Inc.
[3] Koç University, Computer Engineering Department
[4] Koç University, KUIS AI Center

Abstract. Low-light environments pose significant challenges for image enhancement methods. To address these challenges, in this work, we introduce the HUE dataset, a comprehensive collection of high-resolution event and frame sequences captured in diverse and challenging low-light conditions. Our dataset includes 106 sequences, encompassing indoor, cityscape, twilight, night, driving, and controlled scenarios, each carefully recorded to address various illumination levels and dynamic ranges. Utilizing a hybrid RGB and event camera setup. we collect a dataset that combines high-resolution event data with complementary frame data. We employ both qualitative and quantitative evaluations using no-reference metrics to assess state-of-the-art low-light enhancement and event-based image reconstruction methods. Additionally, we evaluate these methods on a downstream object detection task. Our findings reveal that while event-based methods perform well in specific metrics, they may produce false positives in practical applications. This dataset and our comprehensive analysis provide valuable insights for future research in low-light vision and hybrid camera systems.

Keywords: Event-based Vision · Hybrid Camera System · Low-light Image Enhancement

1 Introduction

The growing interest in event-based vision has led to the development of numerous datasets that capture dynamic scenes with high temporal resolution and a wide dynamic range. However, existing datasets often come with certain limitations, such as low resolution, limited diversity of scenes, or a lack of challenging low-light conditions. Given these constraints, our motivation for presenting the HUE (Hacettepe University Event) dataset is threefold. First, our dataset captures events at 1280×720, surpassing the resolution of most available datasets. High-resolution event data is crucial for accurately capturing fine details in complex scenes, enabling more precise analysis and reconstruction tasks. Second, it features a substantial collection of sequences shot in diverse settings. Specifically,

---

[*] These authors contributed equally to this work

the HUE dataset includes 106 sequences recorded in both indoor and outdoor environments, using cameras that are either handheld or mounted on a vehicle. These sequences are captured at various times of the day, such as sunset, twilight, and nighttime, and include a range of camera motions from slow to fast. The scenes also feature both static and dynamic objects, including people, animals, vehicles, buildings, everyday objects, cityscapes, and landscapes. This variety ensures that the dataset can be used to evaluate a wide array of methods under different conditions. Third, the HUE dataset specifically targets challenging low-light scenarios, with sequences captured in conditions where the illuminance on the event sensor is just a few lux. This focus on low-light environments addresses a significant gap in current datasets, which often do not adequately represent such conditions, and is of utmost importance for critical applications like automotive [35], flight control [38], and robotics [50].

Our dataset is collected using a setup consisting of two cameras: one event camera and one frame camera. The event sequences, therefore, include complementary frames as well. We employ a non-coaxial setup where the two cameras do not share the same optical axis and have different optical characteristics. Consequently, the event data and frames collected are not pixel-wise spatially aligned. We share an additional calibration sequence with a flickering chessboard pattern so that camera calibration procedures can be applied to estimate intrinsic and extrinsic camera parameters. Non-coaxial datasets like ours offer the advantage of simpler collection methods, as they do not require additional equipment such as optical beam splitters or specialized cameras like DAVIS [3]. This generic setup is anticipated to become increasingly popular due to its ease of integration, such as the multi-camera systems commonly found in smartphones and robotic systems [43]. Consequently, methods targeting unaligned event and frame data are being developed as well [7], in contrast to earlier works that make the limiting assumption of pixel-wise alignment.

Using the low-light scenes from our dataset, we conduct a comprehensive evaluation of state-of-the-art low-light image enhancement [5,48,51] and event-based image reconstruction [9, 45] and enhancement [22] methods, where we qualitatively and quantitatively compare and contrast their advantages and shortcomings.

In summary, the main contributions of our work are as follows:

– We introduce the HUE dataset, which offers high-resolution event data and complementary frame data captured in diverse and challenging low-light scenarios.
– We provide a comprehensive evaluation of state-of-the-art low-light enhancement and event-based image reconstruction methods using our dataset.
– We highlight the importance of color information in practical applications, particularly in downstream tasks like object detection.

The rest of the paper is organized as follows: Section 2 discusses related work, including comparable low-light image enhancement datasets and event-based vision datasets. Section 3 details our data collection setup and the characteristics of the HUE dataset. Section 4 presents our experimental evaluation of various

methods on the dataset. Finally, Section 5 concludes the paper and discusses future directions. Our dataset, additional calibration sequence, and evaluation scripts are available at `https://ercanburak.github.io/HUE.html`.

2   Related Work

2.1   Datasets for Event-Guided Low-light Enhancement

Collecting paired low-light/normal-light RGB and event data is challenging, resulting in very few real-world datasets. Jiang *et al.* [16] introduced the LIE dataset, the first real-world dataset designed specifically for this purpose. It includes event streams and underexposed frames captured in various indoor and outdoor settings, mostly below 2 lux, using the DAVIS346 event camera for accurate photometric data. The dataset was collected by adjusting the camera's light intake in static scenes, triggering events with light changes indoors and different exposure times outdoors. Liang *et al.* [22] evaluated their methods on real-world data using a custom hybrid camera system. This system combines an industrial camera (FLIR Chameleon 3 Color, 1920×1080 at 20 FPS) and an event camera (DAVIS346, 346×260) using a beam splitter mounted in front of the two cameras with 50% optical splitting. The EvLight framework, proposed by Liang *et al.* [21], introduces a new dataset with over 30K pairs of spatially and temporally aligned images and events under various lighting conditions. This dataset was captured using a robotic arm to ensure precise alignment. The robotic system follows a predefined path, and the DAVIS346 event camera operates with fixed parameters like exposure time. Initially, paired image and event sequences are captured under normal lighting conditions. An ND8 filter is then applied to the camera lens to capture low-light sequences while maintaining consistent camera settings. There are also event datasets that include night sequences without specifically targeting the task of event-guided low-light enhancement. However, these suffer from limitations such as very few night sequences and low resolution [34, 54], lack of complementary frames [29], or limited diversity of the scenes [12].

Table 1 highlights the advantages of our dataset for event-guided low-light enhancement compared to existing datasets targeting this task. One key benefit is our dataset's much higher resolution, with events captured at 1280×720 and frames at 1456×1088. This provides more detailed and high-quality data for both events and frames. In contrast, other datasets, such as those by Jiang *et*

Table 1: Real-world datasets collected for event-guided low-light enhancement.

Dataset	Lux	Event Res.	Frame Res.	Seqs.	GT	Release	Scene	Camera
[16]	0.1-5	346×260	346×260	206	✓	✗	Static	Static
[21]	-	346×260	346×260	91	✓	✓	Static	Dynamic
[22]	-	346×260	1920×1280	-	✗	✗	-	-
Ours	0-24	1280×720	1456×1088	106	✗	✓	Dynamic	Dynamic

al. [16] and Liang *et al.* [21], use much lower resolutions of 346×260 for both events and frames. Even though Liang *et al.* [22] achieves a frame resolution of 1920 × 1280, their event resolution is still much lower. Moreover, our dataset includes dynamic scenes captured with moving cameras, offering a wider range of real-world applications. This is a significant improvement over other methods that typically use static scenes or stationary cameras.

2.2   Low-light Enhancement

Enhancing images and videos captured under low-light conditions is a significant challenge in computer vision. Traditional low-light enhancement techniques have largely focused on histogram equalization [1,14,30,31] and Retinex theory-based methods [4,13,17,40,42] to improve image quality. Recently, deep learning approaches have provided more robust and adaptive solutions. Additionally, event-guided methods have further expanded the potential for low-light enhancement by utilizing the high dynamic range and temporal resolution of event cameras, overcoming the limitations of conventional frame-based methods.

Low-light image enhancement. The first deep learning-based LLIE method, LLNet [24], employs a variant of stacked-sparse denoising autoencoder to simultaneously brighten and denoise low-light images. Lv *et al.* [25] introduced an end-to-end multi-branch enhancement network (MBLLEN) that includes a feature extraction module, an enhancement module, and a fusion module to extract effective feature representations. Xu *et al.* [48] proposed using the signal-to-noise ratio (SNR) as prior information for enhancing low-light images, employing convolutional neural networks (CNNs) to encode local information in high SNR regions and transformers to capture relationships with distant pixels in noisy low SNR regions. Wu *et al.* [46] addressed color consistency issues in enhanced dark images by using a semantic segmentation method to produce and utilize semantic priors, resulting in higher color consistency. Wei *et al.* [44] combined Retinex theory with CNNs to develop RetinexNet, which effectively enhances low-light images. Cai *et al.* [5] proposed the Retinex-based Illumination-Guided Transformer (IGT), which uses illumination representations to guide self-attention computation and enhance interactions between regions of different exposure levels.

Low-light video enhancement. Chen *et al.* [6] proposed a siamese network to ensure generalization capability for processing videos of dynamic scenes. Wang *et al.* [41] formulated an end-to-end framework for enhancing underexposed videos, emphasizing a self-supervised strategy for noise reduction and illumination enhancement based on Retinex theory. To enhance moving objects in dark videos, Haiyang *et al.* [15] proposed a 3D U-Net-based network to map RAW low-light videos to normal-light videos. Zhang *et al.* [51] introduced a method that enforces temporal stability in low-light video enhancement by inferring motion fields from static images and synthesizing short-range video sequences. Fu *et al.* [11] addressed camera motion and spatial alignment by proposing a Retinex-based Light Adjustable Network (LAN) for iterative refinement and adaptive illumination adjustment.

Event-guided low-light enhancement. Liu *et al.* [23] introduced a low-light video enhancement approach using synthetic event guidance, featuring an Event and Image Fusion Transform (EIFT) module for event-image fusion and an Event-Guided Dual Branch (EGDB) module for low-light enhancement, encompassing event synthesis, event and image fusion, and low-light enhancement. Jiang *et al.* [16] utilized a residual fusion module to integrate features from event and frame data, addressing the domain gap with a feature pyramid structure and multi-level encoder for high-order semantic information extraction. Liang *et al.* [22] proposed a novel approach with a multimodal coherence modeling module to establish coherence between frames and events, and a temporal coherence propagation module for coherence-aware aggregation. The Ev-Light framework [21] featured an event-guided low-light enhancement approach with SNR-guided regional feature selection to fuse high-SNR image regions with events from low-SNR regions, and a holistic-regional fusion branch to integrate structural and textural information from both events and images.

2.3   Event-based Image Reconstruction

Reconstructing intensity images from event data is a key task in event-based vision, bridging image and event domains. This allows visualization of events, application of frame-based methods to event data [28, 32], and making use of relations between these two domains in contexts like unsupervised domain adaptation [37, 47, 53]. While earlier works rely on some limiting assumptions such as known or restricted camera movement, static scenes, or brightness constancy [2, 19, 20], recent deep learning based methods trained on large synthetic datasets show impressive results without such assumptions [9, 33, 36, 45]. Zhang *et al.* [52] specifically target event data of low-light scenes and propose an unsupervised domain adaptation network to generate intensity images as if captured in daylight. However, their codes or models are not publicly shared. We use two recent open-source event-based image reconstruction methods for our experimental analysis, ET-Net [45] and HyperE2VID [9], which can successfully reconstruct fine details in dark scenes, where traditional frame cameras suffer from underexposure or noise. The details of these two methods are presented in Section 4.

3   Dataset

3.1   Data Collection Setup

Different types of optical setups can be used to collect complementary event and frame data. One option is to use a camera that integrates both event-based and frame-based data collection mechanisms within the same pixel array, such as the Dynamic and Active Pixel Vision Sensor (DAVIS) [3]. This setup provides synchronized and pixel-wise aligned events and frames. However, it has two significant drawbacks: the shutter activity associated with each frame introduces

Fig. 1: A picture showing our data collection setup containing a PROPHESEE Gen4M event camera paired with an Allied Vision Alvium compact traditional CMOS camera.

noise into the event data, which is absent in sensors dedicated solely to event detection, and the most advanced model of this hybrid camera, the DAVIS346, only supports a maximum resolution of 346×260 pixels. The second option involves using a beam splitter, an optical device featuring a $50/50$ mirror that reflects half of the incoming light and transmits the other half. This arrangement allows frame and event sensors to share an aligned field of view. Despite its advantages, this setup requires additional equipment, which incurs extra costs and occupies more space. The third option, which we have chosen, is a stereo hybrid event-frame camera setup, where separate event and frame cameras are mounted side by side. This setup offers simplicity, as it does not require a beam splitter or a specialized low-resolution camera like the DAVIS. The details of our data collection setup are presented in the following.

Hardware and Optics Our setup utilizes a PROPHESEE Gen4M event camera [10] and an Allied Vision Alvium compact CMOS camera. The event camera features a Sony IMX636 event-based sensor, which is in a $1/2.5''$ format with pixel dimensions of 4.86 $\mu m \times 4.86$ μm. It offers a resolution of 1280×720 and a dynamic range exceeding 120 dB. The event camera is equipped with a Soyo SFA 0820-5M lens, having a fixed focal length of 8 mm, a maximum aperture of $f/2.0$, a minimum focus distance of 0.1 meters, and a horizontal field of view of approximately 38°. The RGB camera contains a Sony IMX273 global shutter sensor, which is in a $1/2.9''$ format with pixel dimensions of 3.45 $\mu m \times 3.45$ μm. It provides a resolution of 1456×1088, a dynamic range of 75 dB, and a 12-bit analog-to-digital converter. The RGB camera is paired with a Tamron M118FM06 lens, having a fixed focal length of 6 mm, a maximum aperture of $f/1.4$, a minimum focus distance of 0.1 meters, and a horizontal field of view of approximately 45°.

The cameras are positioned with a baseline distance of approximately 2 cm between their optical axes, and their relative positions are kept fixed in an optical setup. To achieve this, a custom mechanical part designed specifically for the cameras was produced using a 3D printer. A picture of our setup is presented in Fig. 1. To capture objects at varying distances in focus, both lenses have been set to a narrow aperture of $f/8$, providing a wide depth of field. The focusing

distances of both lenses were then adjusted to the hyperfocal distance, ensuring sharp capture of all objects beyond a certain distance, such as 0.5 meters. For time synchronization, the RGB and event cameras are directly connected using a dedicated cable, linking one of the general-purpose input/output (GPIO) lines of the RGB camera to one of the external trigger lines of the event camera. The software details of time synchronization are described in the following section.

Camera Settings and Software The RGB camera is configured to record at a frame rate of 25 frames per second, with an exposure time set to minimize motion blur in dynamic scenes. Specifically, the exposure time is set to a maximum of 35 milliseconds. The digital gain value of the RGB camera is adjusted based on the scene's brightness—higher for dark scenes and lower for bright scenes—to manage image brightness and noise levels. To avoid excessive noise, we prefer slightly underexposed images in low-light scenarios. The PROPHESEE Gen4 event sensor has six basic operating parameters: `bias_diff`, `bias_diff_on`, `bias_diff_off`, `bias_hpf`, `bias_fo`, and `bias_refr`. These parameters control the sensor's sensitivity to light intensity changes, the cutoff frequencies of event filters, and the refractory period of each pixel. Adjusting these settings affects the sensor's overall sensitivity, temporal precision, background noise, and event delay. To maintain a balance among these characteristics, we have decided to keep the event sensor's parameters at their factory default settings.

We developed software to complement our hardware setup, running on a computer and communicating with the RGB and event cameras via USB interfaces. This software configures the cameras, controls data acquisition, and stores the acquired events and frames. The image frames from the RGB camera and event streams from the event camera are saved to the computer's permanent memory. For time synchronization, the GPIO line of the RGB camera is programmed to be at a high voltage level (logic 1) during exposure and at a low voltage level (logic 0) at other times. The event camera continuously monitors the voltage level on its external trigger line and records the transitions with high temporal precision. Consequently, the start and end times of each RGB camera frame exposure are timestamped using the event camera's high-resolution clock, enabling precise synchronization of data from both cameras. The frames captured by the RGB camera are saved as 8-bit 3-channel color images. The event camera records brightness change events, including pixel position, polarity (increase or decrease in brightness), and a high-precision timestamp. Additionally, the amount of light falling on the event sensor, measured in lux, is recorded with each frame.

3.2   Collected Dataset

Our dataset includes a diverse array of sequences captured in various challenging scenarios. Comprising 106 sequences with an average duration of approximately 17 seconds, the dataset totals around 30 minutes of footage. This includes over 44,000 frames and approximately 27 billion events. To manage its size and variability, we have divided the dataset into six categories based on the scene char-

Table 2: Types of scenes in our HUE Dataset.

Category	Seqs.	Duration	Lux	Description
Indoor	23	6 mins	0-3	Indoor environments dimly lit via natural or artificial light sources.
City	11	4 mins	0-14	Capturing cityscapes through the window of a mid-rise building.
Twilight	23	10 mins	0-24	Outdoors during twilight, featuring natural and urban elements.
Night	16	3 mins	~ 0	Taken in urban environments during night, with moving people, cars, etc.
Driving	16	6 mins	~ 0	From the windshield of a car driving around the city in twilight and night.
Controlled	17	2 mins	0-2	Same scene captured under varying lighting levels and camera settings.

acteristics. Below, we explain each category in detail, while Table 2 presents an overview and Fig. 2 displays frames and event visualizations.

HUE-Indoor: This category includes 23 sequences captured in dimly lit indoor environments, with sensor illuminance levels of just a few lux. This setting allows the evaluation of methods under low-light conditions within indoor environments. 14 of these sequences are lit with natural light filtering through windows, while the remaining sequences are illuminated by artificial light sources. Approximately one-third of the sequences contain dynamic scenes. While some sequences are filmed in large indoor halls and corridors featuring distant objects, most focus on closer subjects. HUE-Indoor is particularly valuable for assessing image enhancement performance on objects with fine details, such as small text and textured regions, under low-light conditions.

HUE-City: This category comprises 11 sequences recorded from the window of a mid-rise building, capturing cityscapes. Seven sequences are recorded during twilight hours, while two are captured at night. In these sequences, the camera setup moves slowly, and the scenes are predominantly static, although occasional moving objects such as vehicles or birds are present. Common elements like roads and buildings are mostly distant from the camera, with scenes extending towards the horizon. This setup tests the spatial resolution limits of the event camera and evaluates each method's capability to reconstruct fine details, as the scenes contain many objects and textures represented by just a few pixels on the sensor. The remaining two sequences are taken during the daytime, capturing both indoor and outdoor scenes. These sequences test the performance under high-dynamic range, since the ratio between the illuminance levels of highly lit and low-lit parts of the scene often varies by orders of tens or hundreds.

HUE-Twilight: This portion of our dataset includes 23 sequences taken outdoors during the twilight hours of the evening when the sun is below the horizon

Fig. 2: Sample scenes from different categories of our proposed HUE dataset.

at varying degrees. More specifically, we define twilight hours as the period when the solar elevation angle (the angle of the sun's geometric center relative to the horizon) is between 0 and −12 degrees, encompassing both civil twilight and nautical twilight. The primary source of illumination is sunlight scattering in the atmosphere, supplemented occasionally by artificial lights such as vehicle headlamps and streetlights. The average illuminance on the event sensor is just a few lux for most of these scenes. Approximately one third of the sequences feature static scenes, while the remaining majority are dynamic, capturing movement within the environment. The sequences feature a mix of natural and urban elements, ranging from lakes and forests to vehicles and buildings.

HUE-Night: This subset contains 16 sequences, all of which are taken in urban environments during the night, *i.e.* when the solar elevation angle is below −18 degrees. This leads to significantly darker conditions, with sensor illumination always being below 1 lux. For these sequences, artificial lights such as streetlights and vehicle headlamps become the primary source of illumination. Half of the sequences feature dynamic elements such as moving people and vehicles. Containing some of the darkest sequences in HUE, this subset is particularly valuable for assessing performance in very low light conditions.

HUE-Driving: This category includes 16 driving sequences, where the camera setup is mounted inside of a vehicle's front windshield, monitoring ahead through this window. Throughout these recordings, the vehicle travels through various street settings in twilight and nighttime, capturing elements such as other cars, pedestrians, motorbikes, parked vehicles, gas stations, tunnels, interchanges, and

roundabouts. This subset is important for evaluating performance in dynamic scenes and under challenging lighting conditions, such as dark roads and rapidly moving headlights.

HUE-Controlled: In this category, we present 17 sequences capturing the same scene, under varying lighting levels and camera settings. More specifically, we record a toy zebra on a desk indoors from a close distance. The recordings are made at night, with the only light source being a dimmable halogen lamp, indirectly illuminating the subject. Throughout these sequences, we adjust the lamp to produce ten different illuminance levels. We record two sequences for each of the first seven illuminance levels, by setting two different digital gain values for the RGB camera. For each of the last three illuminance levels, which are the darkest, we record a single sequence with a gain value of 48, which is the maximum allowed by the RGB camera. This category allows us to make more controlled comparisons under various illuminance levels.

4   Experiments

4.1   Tested Approaches

We evaluate and compare several state-of-the-art methods belonging to four different categories both quantitatively and qualitatively on our HUE dataset: **(I)** two RGB image-based methods, RetinexFormer [5] and SNR-Net [48], **(II)** an RGB video-based method, namely StableLLVE [51], **(III)** two event-based image reconstruction methods, ET-Net [45] and HyperE2VID [9], and **(IV)** a hybrid event and RGB video-based method, EvLowLight [22].

RetinexFormer [5] introduces a one-stage Retinex-based Framework (ORF), revising the traditional Retinex model with perturbation terms for reflectance and illumination to handle corruptions. ORF estimates illumination to brighten images and employs a corruption restorer to reduce noise, artifacts, and color distortions. Additionally, an Illumination-Guided Transformer (IGT) with Illumination-Guided Multi-head Self-Attention (IG-MSA) models long-range dependencies. IGT is integrated into ORF, forming the complete RetinexFormer method.

SNR-Net [48] presents an SNR-Aware framework that dynamically adjusts enhancement based on local signal-to-noise ratio (SNR) estimations. It distinguishes between regions with low and high SNR, applying long-range operations to heavily corrupted areas and short-range operations to clearer regions. It includes a noise estimation module and an SNR-aware transformer, selectively using tokens with sufficient SNR to dynamically enhance image quality.

StableLLVE [51] aims to ensure temporal consistency in low-light video enhancement using static images. The core idea is to predict motion fields from a single image to synthesize short-range video sequences, addressing temporal stability challenges. By using optical flow to simulate motion, this approach applies

temporal consistency even when training on static images.

ET-Net [45] combines the strengths of CNNs and Transformers in a hybrid network designed for event-based video reconstruction. It leverages CNNs for capturing fine-scale local analysis and Transformers for global context modeling. ET-Net introduces a Token Pyramid Aggregation (TPA) module to integrate multi-scale token information, enhancing the relation of internal and intersected semantic concepts. This addresses the limitations of CNNs, particularly in modeling long-range dependencies essential for high-quality video reconstruction from event data.

HyperE2VID [9] introduces a dynamic neural network architecture for event-based video reconstruction. It utilizes hypernetworks to generate per-pixel adaptive filters, informed by a context fusion module that combines data from event voxel grids and previously reconstructed intensity images. HyperE2VID employs a recurrent encoder-decoder backbone with dynamic convolutions applied at the decoder, adapting to spatial variations and enhancing both static and dynamic parts of the scene.

EvLowLight [22] integrates event data with frame data to enhance low-light videos. It leverages the unique ability of event cameras to record brightness changes at extremely high temporal resolution and with a high dynamic range. EvLowLight features a multimodal coherence modeling module that compensates for differences between events and frames, ensuring accurate alignment under challenging low-light conditions. A temporal coherence propagation module further refines the output by sampling features from corresponding points in consecutive events and frames, reducing noise and enhancing the signal-to-noise ratio of the final video output.

4.2   Evaluation Results

All comparison results were generated using the official codes, pre-trained models, and default parameters provided for each method. In particular, for Retinex-Former and SNR-Net, we used pre-trained models on the LOL-v1 dataset [44]. StableLLVE was evaluated with the pre-trained model on their custom synthetic dataset. For ET-Net and HyperE2VID, we utilize official pre-trained models trained on a synthetic training set as described in [36] and employ EVREAL evaluation framework [8] to generate reconstructions. EvLowLight was tested using the model trained on their proposed synthetic dataset.

To quantitatively evaluate the quality of the low-light enhancement and event-based image reconstruction methods, we used no-reference metrics since our dataset lacks paired normal-light ground truth images. These metrics provide perceptual quality scores by directly processing the input images without needing ground truth references. We employed four no-reference metrics: BRISQUE [26], NIQE [27], MANIQA [49], and MUSIQ [18]. BRISQUE and NIQE are traditional metrics that use hand-crafted features to measure adherence to natural

Table 3: Quantitative results of the methods on no-reference metrics.

Methods	NIQE ↓	BRISQUE ↓	MANIQA ↑	MUSIQ ↑
RetinexFormer [5]	<u>4.56</u>	**16.34**	0.17	24.54
SNR-Net [48]	11.07	58.63	0.18	23.71
StableLLVE [51]	5.52	20.26	0.17	21.04
ET-Net [45]	**4.21**	25.62	**0.31**	**46.00**
HyperE2VID [9]	4.86	<u>21.14</u>	<u>0.28</u>	<u>40.29</u>
EvLowLight [22]	6.26	47.59	0.18	37.78

scene statistics, considering distortions such as blur, noise, and compression. Conversely, MANIQA and MUSIQ are deep-learning-based methods designed to assess perceptual image quality end-to-end, specifically focusing on distortions typically observed in the outputs of neural network-based image restoration algorithms.

Table 3 presents the quantitative results of the evaluated low-light enhancement and event-based image reconstruction methods on our HUE dataset. It is important to note that these results were obtained without aligning the RGB data with the event data. Hence, the methods were evaluated on slightly different scenes with different resolutions, which may contribute to the variability in their performance.

In terms of BRISQUE score, RetinexFormer achieves the best results, closely followed by the event-based image reconstruction method HyperE2VID. For NIQE, ET-Net leads with the best scores, followed by RetinexFormer, while SNR-Net records the lowest scores for both BRISQUE and NIQE metrics. When considering MANIQA and MUSIQ, the event-based image reconstruction methods ET-Net and HyperE2VID dominate with the highest scores. EvLowLight, a hybrid event and image method, ranks third in MUSIQ, while it shares the third position in MANIQA with the RGB image-only method SNR-Net. StableLLVE shows the lowest score on MUSIQ, and along with RetinexFormer, it has the lowest scores on MANIQA.

From these results, it is evident that event-based image reconstruction methods generally outperform others in NIQE, MANIQA and MUSIQ metrics, while BRISQUE favors RetinexFormer. Grayscale reconstructions produced by event-based image reconstruction methods are perceived as having less noise and richer texture, contributing to their higher scores in these metrics. Moreover, EvLowLight consistently scores better than other RGB-based methods across metrics except NIQE and BRISQUE. It should be noted that the no-reference metrics used were not specifically designed for low-light enhancement tasks, making their use as direct performance indicators somewhat questionable. These metrics do not adequately account for color and semantic information in their score assessments.

Fig. 3 presents qualitative results from sample scenes for each subset of our dataset. The visual results reveal that while ET-Net and HyperE2VID effectively capture texture and structural information, their reconstructions lack color fi-
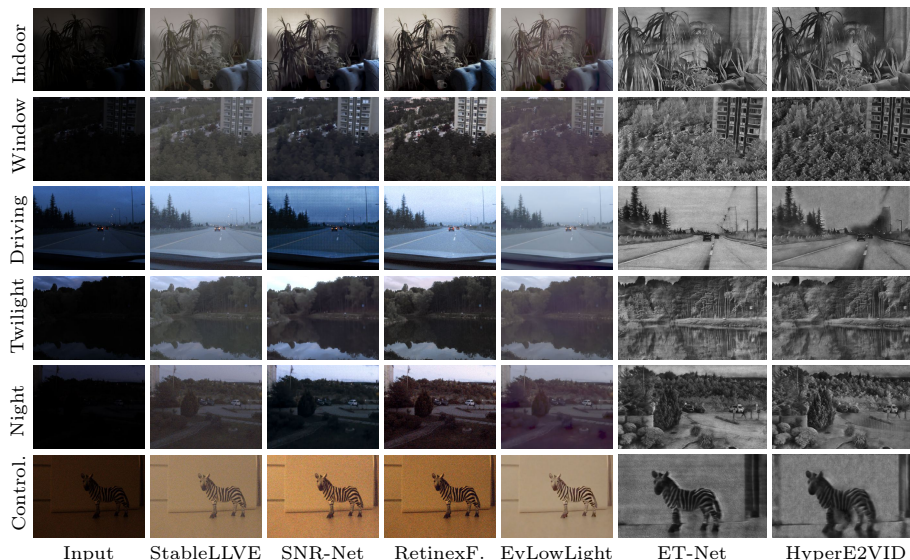
Fig. 3: Qualitative comparison of the low-light enhancement and event-based image reconstruction methods on sample scenes from different subsets of our HUE dataset.

delity. Additionally, the high noise levels in low-light conditions are evident in the generated images. EvLowLight demonstrates superior noise reduction compared to RGB-based image and video methods, yet it struggles to maintain accurate color representation. This suggests that the inclusion of event information in EvLowLight significantly aids in noise reduction, resulting in smoother images. Conversely, image and video-based methods, though richer in color, perform less effectively in denoising. In conclusion, based on both qualitative and quantitative results, no single method consistently outperforms the others across all evaluation metrics.

4.3   Analysis on Downstream Task

We conducted further evaluations to assess the effectiveness of each method on a downstream task: object detection. Specifically, we used the YOLOv7 [39] object detector to detect vehicles in the enhanced images produced by low-light enhancement and event-based image reconstruction methods. Fig. 4 shows example results from the object detector. Despite the high scores of event-based image reconstruction methods on no-reference metrics, qualitative results reveal that these methods often produce numerous false-positive detections. In contrast, methods such as RetinexFormer accurately detected all vehicles in the scene. These findings demonstrate the importance of color information in downstream tasks like object detection. While event-based methods perform well in certain metrics, they may fall short in practical applications where accurate color representation is crucial.
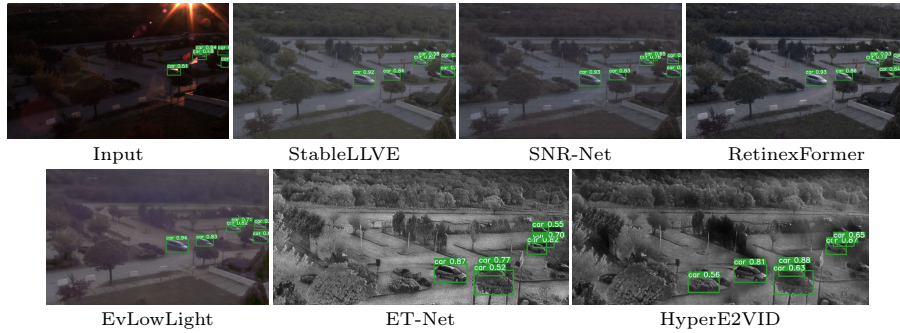
Fig. 4: Object detection performance on input RGB image, low-light enhancement and event-based image reconstruction methods.

5   Conclusion

In this paper, we introduced the HUE dataset, a comprehensive collection of high-resolution event and frame sequences captured in a wide range of low-light conditions. Our dataset is distinguished by its high resolution, diverse scenarios, and specific focus on challenging low-light environments. It includes sequences recorded both indoors and outdoors, at various times of day, and with differing camera motions and scene dynamics, enabling extensive evaluation of low-light enhancement and event-based image reconstruction methods.

We assessed several state-of-the-art methods across four categories: RGB image-based, RGB video-based, event-based, and hybrid event-RGB methods. Our quantitative analysis using no-reference metrics revealed that event-based image reconstruction methods generally performed well in metrics like NIQE, MANIQA, and MUSIQ, while RGB-based methods such as RetinexFormer performed best in BRISQUE. However, qualitative evaluations and a downstream object detection task emphasized the importance of accurate color information, with RGB-based methods showing superior performance in detecting objects without false positives.

These findings highlight the necessity of considering both quantitative metrics and practical application outcomes when evaluating low-light enhancement techniques. While event-based methods are powerful in certain metrics, they may require further refinement for tasks involving color fidelity and object detection. Conversely, hybrid methods like EvLowLight show the potential benefits of integrating event data to reduce noise and improve image smoothness. The HUE dataset, coupled with our detailed evaluations, provides a valuable resource for advancing research in low-light vision and hybrid camera systems. We anticipate that our dataset and findings will inspire future research in this field, ultimately leading to more robust and versatile low-light imaging solutions.

Acknowledgments

References

1. Arici, T., Dikbas, S., Altunbasak, Y.: A histogram modification framework and its application for image contrast enhancement. IEEE Transactions on image processing **18**(9), 1921–1935 (2009)
2. Bardow, P., Davison, A.J., Leutenegger, S.: Simultaneous optical flow and intensity estimation from an event camera. In: CVPR. pp. 884–892 (2016)
3. Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.: A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. IEEE J. Solid-State Circuits **49**(10), 2333–2341 (2014)
4. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: Proceedings of the IEEE international conference on computer vision. pp. 4000–4009 (2017)
5. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12504–12513 (2023)
6. Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: Proceedings of the IEEE/CVF International conference on computer vision. pp. 3185–3194 (2019)
7. Cho, H., Jeong, Y., Kim, T., Yoon, K.J.: Non-coaxial event-guided motion deblurring with spatial alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12492–12503 (2023)
8. Ercan, B., Eker, O., Erdem, A., Erdem, E.: EVREAL: Towards a comprehensive benchmark and analysis suite for event-based video reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3943–3952 (2023)
9. Ercan, B., Eker, O., Saglam, C., Erdem, A., Erdem, E.: HyperE2VID: Improving event-based video reconstruction via hypernetworks. IEEE Transactions on Image Processing **33**, 1826–1837 (2024). `https://doi.org/10.1109/TIP.2024.3372460`
10. Finateu, T., Niwa, A., Matolin, D., Tsuchimoto, K., Mascheroni, A., Reynaud, E., Mostafalu, P., Brady, F., Chotard, L., LeGoff, F., et al.: 5.10 a 1280× 720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86 μm pixels, 1.066 geps readout, programmable event-rate controller and compressive data-formatting pipeline. In: ISSCC. pp. 112–114 (2020)
11. Fu, H., Zheng, W., Wang, X., Wang, J., Zhang, H., Ma, H.: Dancing in the dark: A benchmark towards general low-light video enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12877–12886 (2023)
12. Gehrig, M., Aarents, W., Gehrig, D., Scaramuzza, D.: DSEC: A stereo event camera dataset for driving scenarios. IEEE Robotics and Automation Letters **6**(3), 4947–4954 (2021)
13. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing **26**(2), 982–993 (2016)
14. Ibrahim, H., Kong, N.S.P.: Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics **53**(4), 1752–1758 (2007)
15. Jiang, H., Zheng, Y.: Learning to see moving objects in the dark. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7324–7333 (2019)

16. Jiang, Y., Wang, Y., Li, S., Zhang, Y., Zhao, M., Gao, Y.: Event-based low-illumination image enhancement. IEEE Transactions on Multimedia (2023)
17. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE transactions on image processing **6**(3), 451–462 (1997)
18. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: Musiq: Multi-scale image quality transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5148–5157 (2021)
19. Kim, H., Handa, A., Benosman, R., Ieng, S., Davison, A.: Simultaneous mosaicing and tracking with an event camera. In: BMVC (2014)
20. Kim, H., Leutenegger, S., Davison, A.J.: Real-time 3d reconstruction and 6-dof tracking with an event camera. In: ECCV. pp. 349–364 (2016)
21. Liang, G., Chen, K., Li, H., Lu, Y., Wang, L.: Towards robust event-guided low-light image enhancement: A large-scale real-world event-image dataset and novel approach. arXiv preprint arXiv:2404.00834 (2024)
22. Liang, J., Yang, Y., Li, B., Duan, P., Xu, Y., Shi, B.: Coherent event guided low-light video enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10615–10625 (2023)
23. Liu, L., An, J., Liu, J., Yuan, S., Chen, X., Zhou, W., Li, H., Wang, Y.F., Tian, Q.: Low-light video enhancement with synthetic event guidance. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1692–1700 (2023)
24. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61**, 650–662 (2017)
25. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: BMVC. vol. 220, p. 4. Northumbria University (2018)
26. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing **21**(12), 4695–4708 (2012)
27. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012)
28. Muglikar, M., Gehrig, M., Gehrig, D., Scaramuzza, D.: How to calibrate your event camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1403–1409 (2021)
29. Perot, E., De Tournemire, P., Nitti, D., Masci, J., Sironi, A.: Learning to detect objects with a 1 megapixel event camera. Advances in Neural Information Processing Systems **33**, 16639–16652 (2020)
30. Pizer, S.M.: Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In: Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia. vol. 337, p. 1 (1990)
31. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing **39**(3), 355–368 (1987)
32. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: Events-to-video: Bringing modern computer vision to event cameras. In: CVPR. pp. 3857–3866 (2019)
33. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: High speed and high dynamic range video with an event camera. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(6), 1964–1980 (2019)
34. Scheerlinck, C., Barnes, N., Mahony, R.: Continuous-time intensity estimation using event cameras. In: ACCV. pp. 308–324 (2018)

35. Shariff, W., Dilmaghani, M.S., Kielty, P., Moustafa, M., Lemley, J., Corcoran, P.: Event cameras in automotive sensing: A review. IEEE Access (2024)

36. Stoffregen, T., Scheerlinck, C., Scaramuzza, D., Drummond, T., Barnes, N., Kleeman, L., Mahony, R.: Reducing the sim-to-real gap for event cameras. In: ECCV (2020)

37. Sun, Z., Messikommer, N., Gehrig, D., Scaramuzza, D.: Ess: Learning event-based semantic segmentation from still images. In: European Conference on Computer Vision. pp. 341–357. Springer (2022)

38. Vitale, A., Renner, A., Nauer, C., Scaramuzza, D., Sandamirskaya, Y.: Event-driven vision and control for uavs on a neuromorphic chip. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 103–109. IEEE (2021)

39. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7464–7475 (2023)

40. Wang, L., Xiao, L., Liu, H., Wei, Z.: Variational bayesian method for retinex. IEEE Transactions on Image Processing **23**(8), 3381–3396 (2014)

41. Wang, R., Xu, X., Fu, C.W., Lu, J., Yu, B., Jia, J.: Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9700–9709 (2021)

42. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE transactions on image processing **22**(9), 3538–3548 (2013)

43. Wang, Z., Pan, L., Ng, Y., Zhuang, Z., Mahony, R.: Stereo hybrid event-frame (shef) cameras for 3d perception. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 9758–9764. IEEE (2021)

44. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)

45. Weng, W., Zhang, Y., Xiong, Z.: Event-based video reconstruction using transformer. In: ICCV. pp. 2563–2572 (2021)

46. Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1662–1671 (2023)

47. Xie, C., Gao, W., Guo, R.: Cross-modal learning for event-based semantic segmentation via attention soft alignment. IEEE Robotics and Automation Letters (2024)

48. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022)

49. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: MANIQA: Multi-dimension attention network for no-reference image quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1191–1200 (2022)

50. Yang, Y., Bartolozzi, C., Zhang, H.H., Nawrocki, R.A.: Neuromorphic electronics for robotic perception, navigation and control: A survey. Engineering Applications of Artificial Intelligence **126**, 106838 (2023)

51. Zhang, F., Li, Y., You, S., Fu, Y.: Learning temporal consistency for low light video enhancement from single images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4967–4976 (2021)
52. Zhang, S., Zhang, Y., Jiang, Z., Zou, D., Ren, J., Zhou, B.: Learning to see in the dark with events. In: ECCV. vol. 1, p. 2 (2020)
53. Zheng, X., Wang, L.: Eventdance: Unsupervised source-free cross-modal adaptation for event-based object recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17448–17458 (2024)
54. Zhu, A.Z., Thakur, D., Özaslan, T., Pfrommer, B., Kumar, V., Daniilidis, K.: The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. IEEE Robot. Autom. Lett. **3**(3), 2032–2039 (2018)