SEVD: Synthetic Event-based Vision Dataset for Ego and Fixed Traffic Perception

Manideep Reddy Aliminati*, Bharatesh Chakravarthi*, Aayush Atul Verma,
Arpitsinh Vaghela, Hua Wei, Xuesong Zhou, Yezhou Yang
Arizona State University
{areddy27, bshettah, averma90, avaghel3, hua.wei, xzhou74, yz.yang}@asu.edu

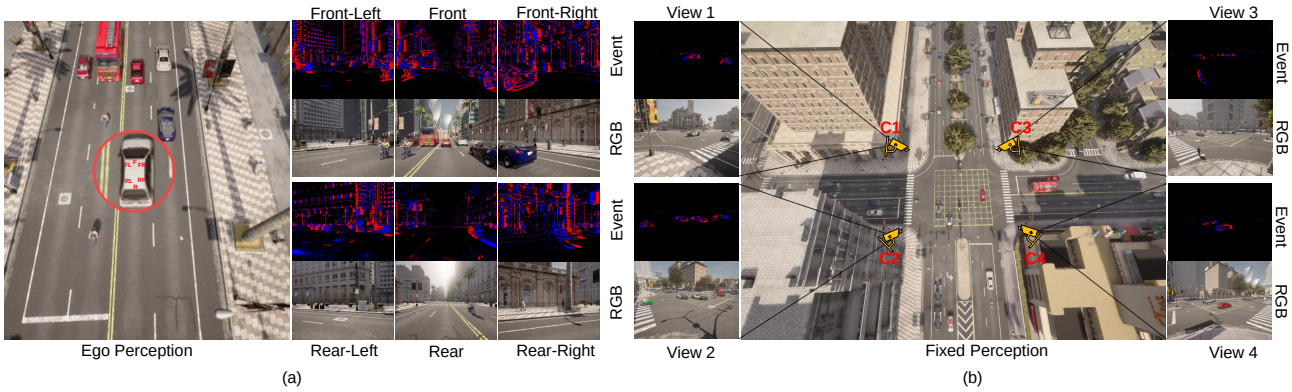

Figure 1. Bird's Eye View of Ego and Fixed Perception Scenario: (a) Shows the six views (Front-Left, Front, Front-Right, Rear-Left, Rear, Rear-Right) from an ego vehicle perception (highlighted in red circle) depicted through event-based and its corresponding RGB frames. (b) Shows four views of an intersection from fixed cameras (C1, C2, C3, C4), with event-based and RGB frames for each view.

Abstract

Recently, event-based vision sensors have gained attention for autonomous driving applications, as conventional RGB cameras face limitations in handling challenging dynamic conditions. However, the availability of real-world and synthetic event-based vision datasets remains limited. In response to this gap, we present SEVD, a first-of-its-kind multi-view ego and fixed perception synthetic event-based dataset using multiple dynamic vision sensors within the CARLA simulator. Data sequences are recorded across diverse lighting (noon, nighttime, twilight) and weather conditions (clear, cloudy, wet, rainy, foggy) with domain shifts (discrete and continuous). SEVD spans urban, suburban, rural, and highway scenes featuring various classes of objects (car, truck, van, bicycle, motorcycle, and pedestrian). Alongside event data, SEVD includes RGB imagery, depth maps, optical flow, semantic, and instance segmentation, facilitating a comprehensive understanding of the scene. Furthermore, we evaluate the dataset using state-of-the-art event-based (RED, RVT) and frame-based (YOLOv8) methods for traffic participant detection tasks and provide baseline benchmarks for assessment. Additionally, we conduct experiments to assess the synthetic event-based dataset's generalization capabilities. The dataset is available at https://eventbasedvision.github.io/SEVD

1. Introduction

In recent years, there has been an increasing focus on neuromorphic or event-based vision due to its ability to excel under high dynamic range conditions, offer high temporal resolution, and consume less power than conventional frame-based vision sensors such as RGB cameras. The event cameras, also known as dynamic vision sensors (DVS), mimic the behavior of biological retinas by continuously sampling incoming light and generating signals only when there is a change in light intensity. This results in an event data stream represented as a sequence of $\langle x, y, t, p \rangle$ tuple, where (x, y) denotes pixel position, t represents time, and p indicates polarity (positive or negative contrast) [17]. Thus, event-based cameras represent

*Equal contribution

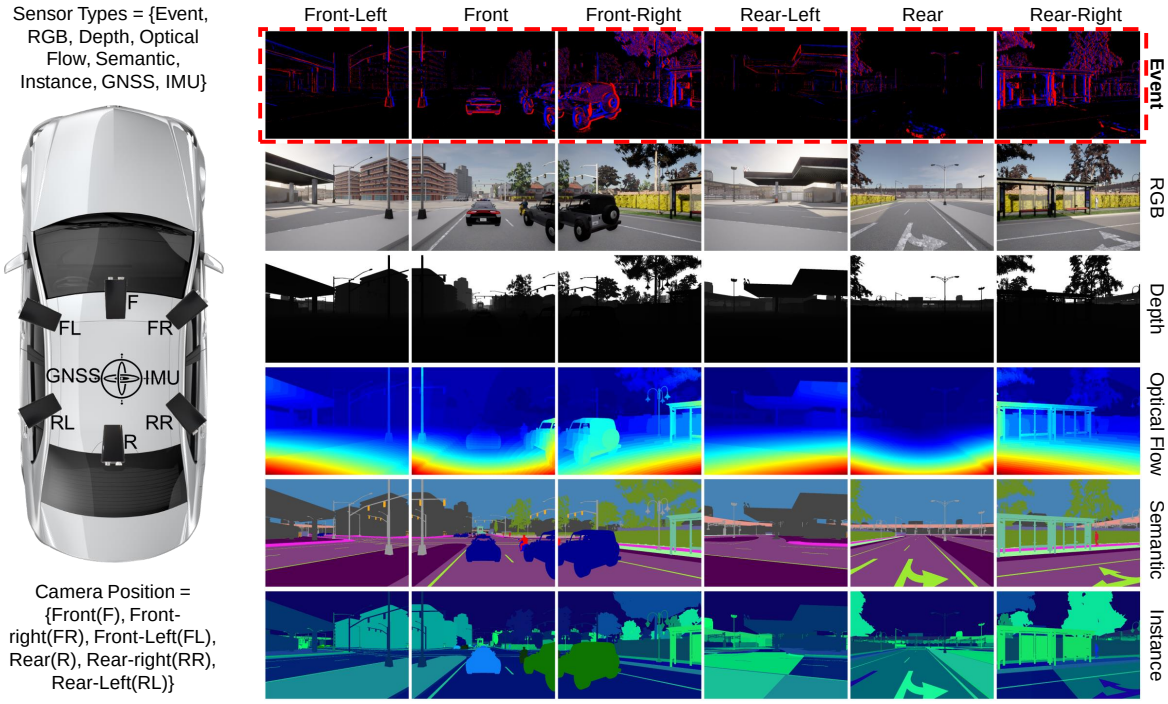

Figure 2. Navigating the Dynamic Landscape of Road Traffic: A glimpse of ego perception data offering six views through event-based vision supported by RGB, depth, optical-flow, semantic, and instance segmentation sensor data generated using CARLA.

a paradigm shift in environmental sensing and perception, capturing minute changes in local pixel-level light intensity and generating asynchronous event streams. This approach revolutionizes how autonomous and traffic ecosystems perceive and respond to their surroundings, enabling precise and efficient real-time processing of dynamic visual data [11, 12, 17].

While event-based sensing represents a novel area, research efforts have been limited in recent years to fully utilize the capabilities of event-based cameras for perception tasks. Notably, researchers have predominantly used event-based cameras like DAVIS346 by iniVation [34] and Prophesee's IMX636 / EVK 4 HD [1] to construct automotive datasets. Additionally, researchers have employed frame-to-event simulators such as ESIM [36] and v2e [22] to generate synthetic event-based data. However, only [22] converts RGB frames of an outdoor scene from MVSEC [44]. This highlights the significant scarcity of readily available synthetic event-based datasets in the field. To bridge this gap and leverage the potential of synthetic data to generate diverse and high-quality vision data tailored for traffic monitoring, we present ***SEVD*** – a **S**ynthetic **E**vent-based **V**ision **D**ataset designed for autonomous driving and traffic monitoring tasks.

SEVD is a multi-view dataset recorded using the CARLA [15] simulator, comprising ego and fixed perception data. Ego perception data is captured through six DVS cameras providing a 360° field-of-view (FoV) from a vehicle (as in Figure 1 (a)). Meanwhile, fixed perception data is recorded from four DVS sensors positioned at specific heights across four locations (as in Figure 1 (b)) at places like intersections, roundabouts, and underpasses, offering multiple views of the same site. Both ego and fixed perception data cover a wide range of environmental conditions. This includes diverse lighting scenarios such as noon, nighttime, and twilight, as well as various weather conditions like clear, cloudy, wet, soft-rainy, hard-rainy, and foggy. Additionally, the dataset spans various scenes, including urban, suburban, highway, and rural settings.

The event cameras are complemented by five different types of sensors, including RGB, depth, optical flow, semantic, and instance segmentation cameras, resulting in a diverse array of data as shown in Figure 2. The dataset also includes GNSS and IMU data to provide additional context for driving scenarios. Annotations for various traffic participants, such as cars, trucks, vans, pedestrians, motorcycles, and bicycles, are provided in both 2D and 3D bounding boxes, following the COCO [28], Pascal VOC [16], and KITTI [20] format respectively. *SEVD* offers raw event streams $\langle x, y, t, p \rangle$ in *.npz* format alongside their corresponding images. This dataset represents a significant advancement as the first-of-its-kind synthetic event-based data providing both ego and fixed perception, featuring a comprehensive range of annotations, extensive recording hours,

	Dataset	Year	Event Camera	Perspective		Objects			Lighting			Weather				Multi-view	Other Sensors	Scene	Other Information
				Ego	Fixed	VH	PED	MM	DY	NT	TW	CLR	CDY	RNY	FGY				
Real-world	DDD17 [5]	2017	DAVIS 346B 346×260 px	✓		✓			✓	✓	✓	✓		✓			RGB	Freeway, Highway, City	12 hr, HDF5 format
	MVSEC [44]	2018	DAVIS 346B 346×260 px	✓		✓			✓			✓				✓ (2 event cameras used)	RGB, VLP-16 Lidar, GPS	Indoor, Outdoor	$\sim$ 1 hr, rosbag format
	N-Cars [38]	2018	Prophesee Gen1 304×240 px	✓		✓											-	Urban	1.2 hr DAT format
	DDD20 [21]	2020	DAVIS 346B 346×240 px	✓		✓			✓	✓	✓	✓					RGB	Highway, Urban	39 hr, HDF5 format
	GEN1 [14]	2020	Prophesee Gen1 346×240 px	✓		✓	✓	✓	✓			✓						City, Highway, Suburban, Countryside	10 hr, DAT format 255K bounding boxes
	1Mpx [35]	2020	Prophesee Gen2 1280×720 px	✓		✓	✓	✓	✓			✓					RGB	City, Highway, Suburban, Countryside	14 hr, DAT format, 25M bounding boxes
	DSEC [18]	2021	Prophesee Gen3.1 640×480 px	✓		✓			✓	✓							RGB, Lidar, GNSS	City	$\sim$1 hr
Synthetic	v2e [22]	2021	v2e simulator	✓		✓			✓								-	Indoor, Outdoor	Frame-based from MVSEC converted to the event stream
	SEVD (Ours)	2024	DVS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ (6 Ego and 4 Fixed event cameras used with a 360° FoV)	RGB, Depth, Optical, Semantic, Instance, GNSS, IMU	Urban, Highway, Suburban, Countryside, Intersections	58 hr, npz format, 9M bounding boxes

Table 1. A comprehensive overview of existing real and synthetic event-based automotive traffic perception datasets. (VH: Vehicle, PED: Pedestrian, MM: Micro-mobility, DY: Daytime, NT: Nighttime, TW: Twilight, CLR: Clear, CDY: Cloudy, RNY: Rainy, FGY: Foggy)

and diverse driving conditions. We outline the key contributions of the presented work below:

1. A multi-view (360°) synthetic event-based dataset comprising 27 hr of fixed and 31 hr of ego perception data, with over 9M bounding boxes, recorded across diverse conditions and varying parameters.

2. Establishing benchmark baselines for 2D detection using state-of-the-art event-based and frame-based architectures on *SEVD* across different driving and fixed traffic monitoring scenarios to assess the efficacy of the dataset.

3. A quantitative and qualitative evaluation of synthetic event-based detector's generalization capabilities on real-world data.

The rest of this paper is structured as follows. Section 2 briefly summarizes the existing event-based dataset contributions. Section 3 presents the *SEVD* dataset, detailing the generation framework and annotation process. In Section 4, we provide baselines and explore the generalization capabilities of synthetic event-based detectors to real-world data.

2. Related study

With recent advancements in neuromorphic vision, researchers have curated event-based datasets across various domains [3, 6, 7, 26, 32]. This section provides an overview of event-based automotive traffic perception datasets, focusing on their availability in real or synthetic environments, as summarized in Table 1.

2.1. Event-based real automotive datasets

Real event-based automotive datasets are crucial as they offer valuable insights for advancing autonomous driving research. The DDD17 [5] is the first open dataset offering driving recordings, annotated data for end-to-end learning approaches, and sensor fusion techniques. Meanwhile, the MVSEC [44] dataset addresses the scarcity of labeled data for 3D perception tasks. It offers synchronized stereo pair event-based data captured in diverse scenarios, enabling the development and testing of algorithms for tasks like feature tracking, visual odometry, and stereo depth estimation.

Additionally, N-Cars [38] serves a large dataset for object classification, showcasing improved classification performance in real-time computation for applications like autonomous vehicles and UAV vision. Furthermore, the DDD20 [21] expands DDD17 [5] with an additional 39 hr of data, making it the longest event camera end-to-end driving dataset. The GEN1 [14] automotive dataset offers over 39 hr of recordings captured with a 304×240 px GEN1 sensor. It provides manual annotations for cars and pedestrians, contributing to the advancement of event-based vision tasks such as object detection and classification. The 1 Megapixel Automotive Dataset [35] is a large-scale and high-resolution dataset containing over 14 hr of recordings with 25M bounding boxes of cars, pedestrians, and two-wheelers labeled at high frequency in automotive scenarios.

Lastly, DSEC [18], a unique dataset designed for driving scenarios in challenging illumination conditions. It is the first large-scale stereo dataset with event cameras containing 53 sequences collected in various illumination conditions. This dataset provides ground truth disparity for developing and evaluating event-based stereo algorithms, advancing autonomous driving research.

2.2. Event-based synthetic automotive datasets

In the synthetic event data generation space, prominent simulators include the RPG DAVIS [32], ESIM [36], v2e [22], blinkSim [27], and the CARLA Simulator [15]. The v2e simulator stands out for its ability to generate realistic synthetic DVS events from intensity frames and ensures realism. Notably, v2e utilized the MVSEC [44] dataset to convert RGB to event frames of an outdoor scenario during nighttime, providing vehicle annotations. However, there are no extensive synthetic event-based auto-

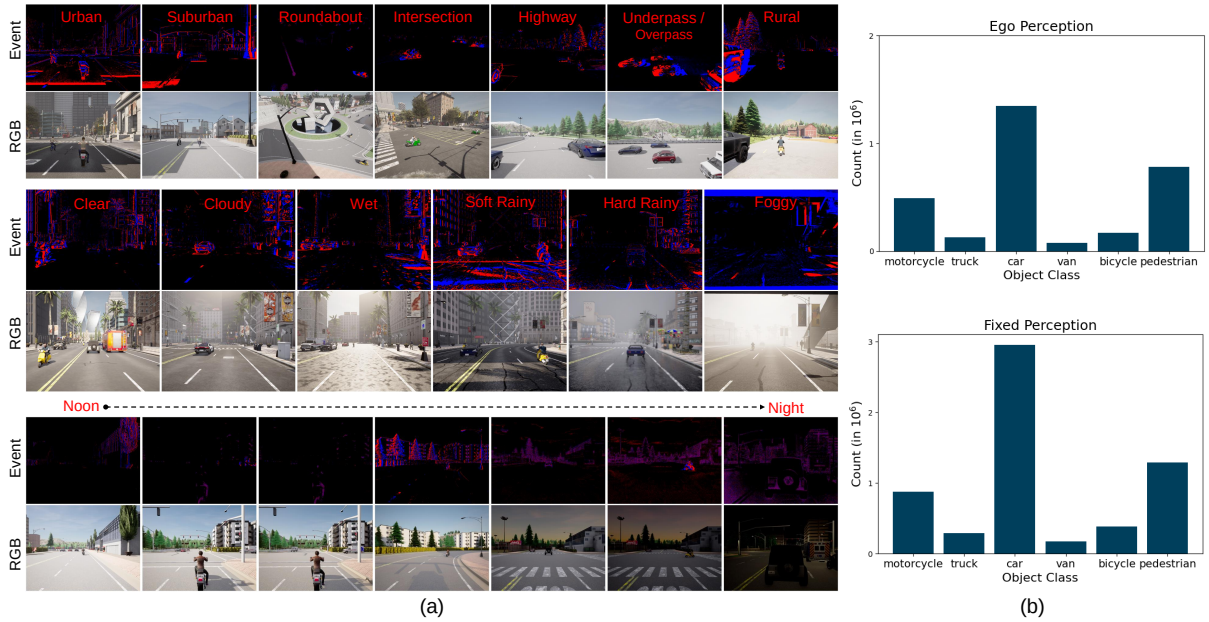

Figure 3. (a) Visual Representation (event and RGB) of *SEVD* Dataset Features: The scene diversity (top row) considered during data generation from the urban to the rural, weather variability (middle row) captured in the dataset, ranging from clear skies to foggy scenarios, and the dynamic conditions (bottom row) showcasing sequences with continuously shifting parameters, mirroring real-world driving scenarios. (b) Dataset Distribution: shows the distribution of instances for each class in both ego and fixed perception scenarios.

motive traffic perception datasets to the best of our knowledge. Unlike the aforementioned simulators, which primarily transform existing frame-based datasets into their event counterparts, CARLA generates event data by uniformly sampling between two consecutive synchronous frames in dynamic traffic conditions. Interestingly, despite its potential, the DVS capabilities of the CARLA simulator remain largely unexplored.

3. SEVD - Dataset overview

This section provides an overview of the sensor suite and multi-camera configuration employed for capturing both ego and fixed perception data. We discuss the data generation pipeline implemented within the CARLA simulator environment, describe the labeling protocol, and a breakdown of the generated data.

3.1. Sensor setup

The sensor suite comprises a strategically positioned array of sensors of each type (event, RGB, depth, optical flow, semantic, and instance), tailored for both ego and fixed scenarios. In ego scenarios, the cameras offer coverage from front to rear, including front-right, front-left, rear-right, and rear-left perspectives, each with overlapping FoV providing a comprehensive 360° view. Notably, the rear camera features a wider 110° FoV, while the others have a 70° FoV, following the approach used in nuScenes [8]. For fixed-perception scenarios, four cameras, each offering a 90° FoV, are used. This strategic setup enhances our data collection strategy, contributing to a diverse range of data across various conditions.

The DVS camera within the CARLA environment is configured with a dynamic range of 140 dB, a temporal resolution in the order of microseconds, and a resolution of 1280×960 px. It generates a continuous stream of events represented by $\langle x, y, t, p \rangle$, where an event is triggered at pixel coordinates (x, y) and timestamp t when the logarithmic intensity change L exceeds a predefined constant threshold C, as defined by the equation below [10]:

$$L(x, y, t) - L(x, y, t - \delta t) = \mathrm{p} \times C \tag{1}$$

$t - \delta t$ is the time when the last event at that pixel was triggered and p represents the polarity of the event: $p = +1$ indicates an increment in brightness, and $p = -1$ indicates a decrement, with $C = 0.3$.

Furthermore, the RGB, depth, optical flow, semantic, and instance segmentation cameras are all configured to operate at a resolution of 1280×960 px. Depth information is encoded into grayscale images with floating-point values between [0,1] and with a 1 mm resolution. The optical flow, crucial for motion-centric algorithms, is provided in UV map format. For each frame, we provide panoptic segmentation labels, encompassing both instance and semantic segmentation, across the 23 classes defined in the

Map Type	Lighting Conditions	RED (EVENT)				RVT (EVENT)				YOLOv8 (RGB)			
		Car	Pedestrian	Motorcycle	All Classes	Car	Pedestrian	Motorcycle	All Classes	Car	Pedestrian	Motorcycle	All Classes
Intersection	**Noon**	0.530	0.386	0.415	0.387	0.537	0.810	0.551	0.552	0.801	0.439	0.697	0.659
Roundabout		0.865	0.439	0.770	0.728	0.841	-	0.846	0.854	0.987	0.475	0.957	0.886
Underpass / Overpass		0.779	-	-	0.756	0.955	-	-	0.944	0.823	-	-	0.828
All Maps		0.680	0.398	0.575	0.520	0.679	0.809	0.694	0.683	0.838	0.441	0.778	0.717
Intersection	**Night**	0.451	0.331	0.144	0.310	0.469	0.650	0.228	0.477	0.737	0.328	0.613	0.619
Roundabout		0.488	0.235	0.472	0.548	0.547	-	0.566	0.626	0.747	0.220	0.797	0.738
Underpass / Overpass		0.693	-	-	0.473	0.357	-	-	0.315	0.774	-	-	0.769
All Maps		0.467	0.313	0.229	0.398	0.477	0.644	0.328	0.496	0.736	0.303	0.671	0.667
Intersection	**Twilight**	0.479	0.323	0.398	0.370	0.574	0.720	0.601	0.600	0.658	0.334	0.610	0.577
Roundabout		0.902	0.454	0.742	0.760	0.884	-	0.888	0.896	0.987	0.583	0.938	0.903
Underpass / Overpass		0.777	-	-	0.669	0.961	-	-	0.942	0.840	-	-	0.793
All Maps		0.571	0.346	0.439	0.474	0.650	0.720	0.630	0.670	0.708	0.362	0.648	0.647

Table 2. Fixed perception Baseline Evaluation: Results of tensor-based methods RVT and RED, along with the frame-based approach YOLOv8, across different map types (Intersections, Roundabout, Underpass, and Overpass) in fixed perception scenarios. RVT and RED evaluations are conducted on event data, while YOLOv8 is evaluated on RGB data.

Cityscapes [13] annotation scheme. Additionally, GNSS and IMU sensors are incorporated to track the position and orientation of the vehicle precisely in ego perception scenarios.

3.2. Data generation and annotation

We leverage the CARLA simulator's robust data generation pipeline to record traffic data across diverse scenes (see Figure 3 (a), top row), illumination, weather conditions (see Figure 3 (a), middle row)), and varying traffic densities. Each data generation iteration initializes the simulation with custom settings and generates traffic. The data sequences are generated under discrete environmental weather conditions, with each set of sequences collected using different domain parameters and initial states. Additionally, we generate sequences with continuously shifting conditions (see Figure 3 (a), bottom row), where a shift is generated by interpolating the state between the initial and final frame. To ensure precise synchronization among sensors, the simulation operates in synchronous mode with a time-step of 0.1 seconds, corresponding to 10 frames per second (fps). Following the simulation phase, we employ CARLA's bounding box API to generate annotations in both 2D and 3D bounding box formats, such as COCO, Pascal VOC, and KITTI. Additionally, we offer annotations for optical flow and dense depth. We provide code for custom data generation at https://github.com/eventbasedvision/SEVD

3.3. Recording and statistics

SEVD offers a diverse range of recordings featuring various combinations of scenes, weather, and lighting conditions. For instance, in an ego perception scenario, recordings may feature an ego vehicle navigating through a suburban environment under a continuous domain shift, transitioning from noon to night in clear weather conditions as depicted in Figure 3 (a) bottom row. Another example of a fixed perception may feature a twilight setting with a soft rain over a four-way intersection. Each recording spans durations of 2 to 30 min. We provide a total of 27 hr of fixed and 31 hr of ego perception event data collectively. Similarly, we offer an equal volume of data from other sensor types, resulting in a cumulative 162 hr of fixed and 186 hr of ego perception data.

SEVD comprises extensive annotations, including 2D and 3D bounding boxes for six categories (car, truck, bus, bicycle, motorcycle, and pedestrian) of traffic participants, totaling approximately 9M bounding boxes, with cars being the most prevalent category as illustrated in Figure 3 (b). To facilitate model training, the dataset is segmented into subsets containing 70% train, 15% validation, and 15% test data, ensuring proportional representation across various combinations.

4. Experiments

SEVD is a comprehensive dataset, offering both ego and fixed perception multi-view multi-sensor data, particularly emphasizing event-based vision for autonomous driving and traffic monitoring. In our experiments, we focus on modeling event data for the 2D detection task across various scenes, weather conditions, and lighting scenarios. To establish baselines, we train state-of-the-art event-based detectors, including Recurrent Vision Transformers (RVT) [19], Recurrent Event-camera Detector (RED) [35]. Additionally, we evaluate frame-based (RGB) data using the You Only Look Once (YOLOv8) [24] detector. While our primary focus lies in event data, *SEVD* also provides other modality data to support broader research endeavors. Additionally, we conduct quantitative and qualitative evaluations for event-based detectors using real-world event data, offering insights into their generalization performance in dynamic scenarios. For all experiments, the evaluation metrics are based on Mean Average Precision (mAP) at a 50% Intersection over Union (IoU) threshold.

4.1. Baseline evaluation

To evaluate detector performance on *SEVD*, we train them using 10 hr of data and assess them on separate sets

Map Type	Lighting Conditions	RED (EVENT)				RVT (EVENT)				YOLOv8 (RGB)			
		Car	Pedestrian	Motorcycle	All Classes	Car	Pedestrian	Motorcycle	All Classes	Car	Pedestrian	Motorcycle	All Classes
Urban	**Noon**	0.524	0.124	0.095	0.167	0.598	0.149	0.268	0.444	0.961	0.63	0.806	0.881
Suburban		0.375	0.057	0.430	0.239	0.523	0.338	0.714	0.618	0.887	0.376	0.815	0.744
Rural		0.418	0.068	0.226	0.143	0.488	0.473	0.396	0.471	0.878	0.46	0.584	0.552
Highway		0.248	-	-	0.101	0.500	-	-	0.485	0.881	-	-	0.870
All Maps		0.348	0.098	0.270	0.159	0.514	0.243	0.565	0.515	0.892	0.536	0.744	0.753
Urban	**Night**	0.085	0.080	0.053	0.048	0.175	0.228	0.290	0.223	0.607	0.642	0.478	0.668
Suburban		0.409	0.159	0.212	0.202	0.735	0.054	0.439	0.534	0.836	0.641	0.847	0.825
Rural		0.115	0.076	0.030	0.044	0.190	0.129	0.010	0.145	0.727	0.353	0.396	0.493
Highway		0.046	-	-	0.034	0.158	-	-	0.127	0.559	-	-	0.456
All Maps		0.166	0.074	0.088	0.085	0.284	0.175	0.229	0.260	0.701	0.472	0.525	0.649
Urban	**Twilight**	0.307	0.121	0.330	0.306	0.499	0.073	0.637	0.573	0.909	0.46	0.765	0.721
Suburban		0.323	0.161	0.136	0.164	0.390	0.279	0.387	0.294	0.772	0.671	0.764	0.713
Rural		0.256	0.151	0.289	0.186	0.265	0.366	0.258	0.311	0.775	0.616	0.544	0.597
Highway		0.502	-	-	0.267	0.566	-	-	0.590	0.878	-	-	0.924
All Maps		0.381	0.128	0.287	0.211	0.485	0.249	0.569	0.485	0.854	0.559	0.712	0.720

Table 3. Ego Perception Baseline Evaluation: Results of tensor-based methods RVT and RED, along with the frame-based approach YOLOv8, across different map types (Urban, Suburban, and Rural) in ego-driving scenarios. RVT and RED evaluations are conducted on event data, while YOLOv8 is evaluated on RGB data captured from the front view of the ego-vehicle.

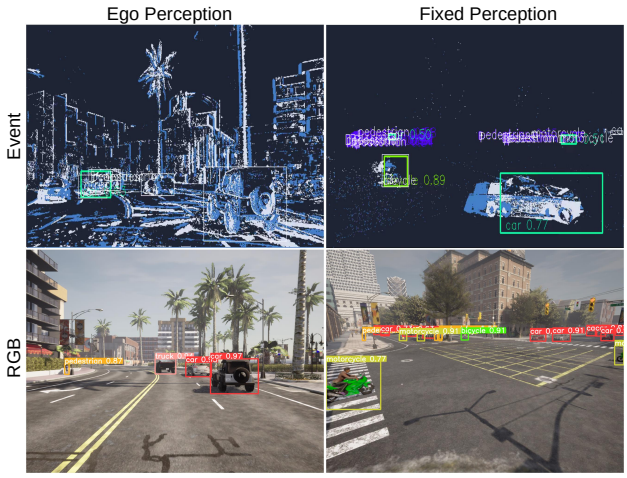

Figure 4. Qualitative Results: Showcasing event-based and frame-based detection of different classes of objects in ego (left column) and fixed (right column) perception scenarios.

Train Set	Synthetic			Real world		
	Car	Pedestrian	All Classes	Car	Pedestrian	All Classes
Fixed	0.537	0.810	0.552	0.384	0.217	0.391

Table 4. Generalization assessment of RVT model, trained on synthetic data and evaluated on real-world fixed perception data.

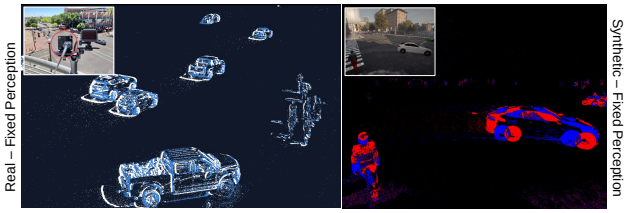

Figure 5. Real-World Fixed Perception Event-Data Acquisition: Data captured at an intersection using the high-resolution Prophesee EVK4 HD event camera (left), similar to a setting used in CARLA (right).

comprising 2.5 hr for validation and 2.5 hr for testing. RVT and RED were trained from scratch over 4 days using an A100 GPU, while YOLOv8 underwent training for 3.5 days on an A2000 GPU. We conducted distinct evaluations to discern how each model operates across various scenes, weather conditions, and lighting scenarios, encompassing both ego and fixed perception settings. This approach enables us to gain insights into the models' adaptability to diverse environmental contexts. Several key observations emerged from the evaluation of fixed and ego perception using tensor-based and frame-based models, as depicted in Table 2 and Table 3.

For the fixed perception scenario, we utilize all four views from the sensor setup (refer to Figure 2 (b)). Similar performance is noted among RED and RVT for car class detection, with RVT outperforming RED for pedestrian and motorcycle classes across all scenes and lighting conditions. Overall, RVT exhibits better detection performance across all classes and lighting conditions compared to RED. For frame-based detection, YOLOv8 demonstrates competitive performance in detecting cars across different lighting conditions but lags in pedestrian detection. For ego perception scenarios, we utilize front view data of the ego vehicle. RVT outperforms RED in terms of all class detection across all scenes and lighting. Consistent detection performance is observed for the car class in ego scenarios. For frame-based detection, cars and motorcycles exhibit better performance than pedestrian detection. Notably, detectors show similar performance in noon and twilight conditions, with a decline observed in nighttime conditions for both fixed and ego perception settings. Figure 4 shows qualitative detection results for ego and fixed perception.

(a) (b)

Figure 6. Qualitative Results: Sample of detection instances in ego perception, showing (a) incorrect and (b) correct detections, highlighting the variability in model performance.

4.2. Generalization on real-world data

To assess the synthetic data-trained model's generalization capabilities on real-world data, we conduct quantitative and qualitative experiments on synthetic event-based detector models. For this purpose, we opt for RVT due to its superior performance compared to RED.

In our quantitative evaluation, we assess the fixed-perception detector model over real-world fixed event-based data, presenting the results in Table 4. The real-world event data was collected at an intersection near a university campus using the high-resolution Prophesee EVK4 HD event camera [1]. The event camera was strategically positioned at approximately 6 m with a pitch angle of about 35° to the ground, as shown in Figure 5 (left). This configuration mirrors that of DVS sensors in the CARLA simulation environment. RVT exhibits a relatively decent performance transitioning to real-world scenarios across all classes of objects, with a drop in performance for the pedestrian class. We present qualitative results for ego perception, utilizing 1 Megapixel Automotive dataset from Prophesee [35] as depicted in Figure 6. We intend to quantitatively evaluate real-world ego and fixed [40] perception scenarios as part of extended work.

5. Discussion

In this section, we highlight the advantages of the *SEVD* dataset across various domains. The high temporal resolution of event-based cameras significantly enhances performance in numerous tasks such as detection [25,35], tracking [29, 42, 43], ReID [2, 9], trajectory prediction [31], optical flow [33], feature tracking [30, 37], and SLAM [23, 41]. This is particularly beneficial for autonomous driving applications, where real-time response is crucial as we offer dataset over 9M bounding box annotations (2D and 3D) in different formats (COCO, Pascal VOC, and KITTI), along with tracking IDs, facilitating tasks like 2D/3D detection, tracking, re-identification, and trajectory prediction.

SEVD, being a multi-view synthetic vision-based ego and fixed perception dataset, plays a vital role in supporting existing methods that rely on data fusion from different sensor modalities to overcome the limitations of single sensor types and enhance detection performance. Further, multi-view helps to overcome challenges, such as occlusion, limited perception horizon due to a restricted field of view, and low-point density at distant regions, posed by a single viewpoints. [4].

Moreover, our dataset extends beyond ego-motion perception to include data from infrastructure perception, supporting research in Vehicle-to-Infrastructure (V2I) communication. This inclusion allows for the exploration of cooperative perception systems, where information sharing between vehicles and infrastructure enhances situational awareness and navigation safety [39]. Such advancements can expedite the development of innovative solutions for V2I communication and collaborative autonomous systems, ultimately fostering safer and more efficient transportation networks.

6. Conclusion

The *SEVD* dataset marks a notable advancement in synthetic event-based dataset for autonomous driving and traffic monitoring, offering 27 hr of fixed and 31 hr of ego perception event data, complemented by an equal amount of data from other sensor types. In total, the dataset encompasses a substantial 162 hr of fixed and 186 hr of ego perception data, offering a comprehensive view of diverse environmental conditions, including various lighting and weather scenarios across different scenes. This rich diversity of scenarios within *SEVD* provides researchers with ample opportunities for exploration. The dataset's extensive annotation framework, featuring over 9M bounding boxes for various traffic participants in both 2D and 3D formats, along with raw event streams and their corresponding data from other sensor types, facilitates a deeper understanding of synthetic vision data and enables more effective algorithm development and evaluation. Furthermore, We report baselines for 2D detection tasks using state-of-the-art event-based detectors and detection performance for frame-based data. We believe *SEVD* serves as a valuable resource for researchers and practitioners in the field, supporting advancements in event-based vision technology and contributing to the development of safer and more efficient transportation systems.

References

[1] Event Camera Evaluation Kit 4 HD IMX636 Prophesee-Sony. 2, 7

[2] Shafiq Ahmad, Pietro Morerio, and Alessio Del Bue. Person re-identification without identification via event anonymization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11132–11141, 2023. 7

[3] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, et al. A low power, fully event-based gesture recognition system. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7243–7252, 2017. 3

[4] Eduardo Arnold, Mehrdad Dianati, Robert de Temple, and Saber Fallah. Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):1852–1864, 2020. 7

[5] Jonathan Binas, Daniel Neil, Shih-Chii Liu, and Tobi Delbruck. Ddd17: End-to-end davis driving dataset. *arXiv preprint arXiv:1711.01458*, 2017. 3

[6] Tobias Bolten, Regina Pohle-Frohlich, and Klaus D. Tonnies. Dvs-outlab: A neuromorphic event-based long time monitoring dataset for real-world outdoor scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1348–1357, June 2021. 3

[7] Chiara Boretti, Philippe Bich, Fabio Pareschi, Luciano Prono, Riccardo Rovatti, and Gianluca Setti. Pedro: An event-based dataset for person detection in robotics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4064–4069, 2023. 3

[8] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[9] Chengzhi Cao, Xueyang Fu, Hongjian Liu, Yukun Huang, Kunyu Wang, Jiebo Luo, and Zheng-Jun Zha. Event-guided person re-identification via sparse-dense complementary learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17990–17999, 2023. 7

[10] CARLA Contributors. CARLA - DVS camera, Accessed 2024. Accessed on: February 15, 2024. 4

[11] Bharatesh Chakravarthi, M Manoj Kumar, and BN Pavan Kumar. Event-based sensing for improved traffic detection and tracking in intelligent transport systems toward sustainable mobility. In *International Conference on Interdisciplinary Approaches in Civil Engineering for Sustainable Development*, pages 83–95. Springer, 2023. 2

[12] Guang Chen, Hu Cao, Jorg Conradt, Huajin Tang, Florian Rohrbein, and Alois Knoll. Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception. *IEEE Signal Processing Magazine*, 37(4):34–49, 2020. 2

[13] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset. In *CVPR Workshop on the Future of Datasets in Vision*, volume 2, page 1, 2015. 5

[14] Pierre De Tournemire, Davide Nitti, Etienne Perot, Davide Migliore, and Amos Sironi. A large scale event-based detection dataset for automotive. *arXiv preprint arXiv:2001.08499*, 2020. 3

[15] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2, 3

[16] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111:98–136, 2015. 2

[17] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 1, 2

[18] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954, 2021. 3

[19] Mathias Gehrig and Davide Scaramuzza. Recurrent vision transformers for object detection with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13884–13893, 2023. 5

[20] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. 2

[21] Yuhuang Hu, Jonathan Binas, Daniel Neil, Shih-Chii Liu, and Tobi Delbruck. Ddd20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020. 3

[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. 2, 3

[23] Jianhao Jiao, Huaiyang Huang, Liang Li, Zhijian He, Yilong Zhu, and Ming Liu. Comparing representations in tracking for event camera-based slam. In *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, pages 1369–1376, 2021. 7

[24] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, Jan. 2023. 5

[25] Alexander Kugele, Thomas Pfeil, Michael Pfeiffer, and Elisabetta Chicca. How many events make an object? improving single-frame object detection on the 1 mpx dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3912–3921, 2023. 7

[26] Bin Li, Hu Cao, Zhongnan Qu, Yingbai Hu, Zhenke Wang, and Zichen Liang. Event-based robotic grasping detection with neuromorphic vision sensor and event-grasping dataset. *Frontiers in neurorobotics*, 14:51, 2020. 3

[27] Yijin Li, Zhaoyang Huang, Shuo Chen, Xiaoyu Shi, Hongsheng Li, Hujun Bao, Zhaopeng Cui, and Guofeng Zhang. Blinkflow: A dataset to push the limits of event-based optical flow estimation. *arXiv preprint arXiv:2303.07716*, 2023. 3

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 2

[29] Jiawei Lu and Xuesong Simon Zhou. Virtual track networks: A hierarchical modeling framework and open-source tools for simplified and efficient connected and automated mobility (cam) system design based on general modeling network specification (gmns). *Transportation Research Part C: Emerging Technologies*, 153:104223, 2023. 7

[30] Nico Messikommer, Carter Fang, Mathias Gehrig, and Davide Scaramuzza. Data-driven feature tracking for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5642–5651, 2023. 7

[31] Marco Monforte, Luna Gava, Massimiliano Iacono, Arren Glover, and Chiara Bartolozzi. Fast trajectory end-point prediction with event cameras for reactive robot control. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4035–4043, 2023. 7

[32] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 36(2):142–149, 2017. 3

[33] Jun Nagata and Yusuke Sekikawa. Tangentially elongated gaussian belief propagation for event-based incremental optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21940–21949, 2023. 7

[34] Neuromorphic vision systems. iniVation. https://inivation.com/, 2024. Accessed: February 8, 2024. 2

[35] 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. 3, 5, 7

[36] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *Conference on robot learning*, pages 969–982. PMLR, 2018. 2, 3

[37] Hochang Seok and Jongwoo Lim. Robust feature tracking in dvs event stream using bezier mapping. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1658–1667, 2020. 7

[38] Amos Sironi, Manuele Brambilla, Nicolas Bourdis, Xavier Lagorce, and Ryad Benosman. Hats: Histograms of averaged time surfaces for robust event-based object classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1731–1740, 2018. 3

[39] A Stevens and J Hopkin. Benefits and deployment opportunities for vehicle/roadside cooperative its. 2012. 7

[40] Aayush Atul Verma, Bharatesh Chakravarthi, Arpitsinh Vaghela, Hua Wei, and Yezhou Yang. etram: Event-based traffic monitoring dataset. *arXiv preprint arXiv:2403.19976*, 2024. 7

[41] Gongyu Yang, Qilin Ye, Wanjun He, Lifeng Zhou, Xinyu Chen, Lei Yu, Wen Yang, Shoushun Chen, and Wei Li. Live demonstration: Real-time vi-slam with high-resolution event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 7

[42] 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. 7

[43] Jiqing Zhang, Xin Yang, Yingkai Fu, Xiaopeng Wei, Baocai Yin, and Bo Dong. Object tracking by jointly exploiting frame and event domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13043–13052, 2021. 7

[44] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018. 2, 3