eTraM: Event-based Traffic Monitoring Dataset

Aayush Atul Verma*, Bharatesh Chakravarthi*, Arpitsinh Vaghela*, Hua Wei, Yezhou Yang
Arizona State University

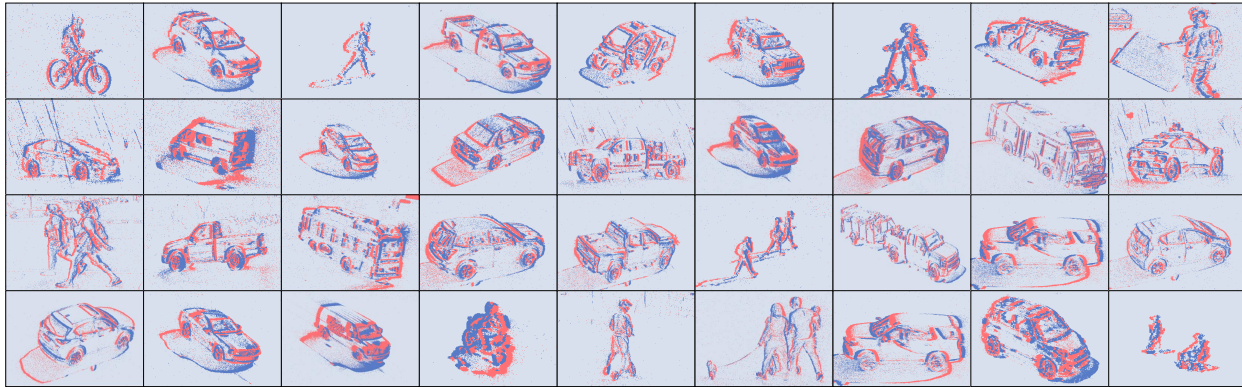{averma90, bshettah, avaghel3, hua.wei, yz.yang}@asu.edu

Figure 1. Unveiling the Dynamic World of Road Traffic: A glimpse into our event-based traffic monitoring dataset featuring diverse traffic participants including pedestrians, various sized vehicles, and micro-mobility that include cycles, wheelchairs, and e-scooters.

Abstract

Event cameras, with their high temporal and dynamic range and minimal memory usage, have found applications in various fields. However, their potential in static traffic monitoring remains largely unexplored. To facilitate this exploration, we present eTraM - a first-of-its-kind, fully event-based traffic monitoring dataset. eTraM offers 10 hr of data from different traffic scenarios in various lighting and weather conditions, providing a comprehensive overview of real-world situations. Providing 2M bounding box annotations, it covers eight distinct classes of traffic participants, ranging from vehicles to pedestrians and micro-mobility. eTraM's utility has been assessed using state-of-the-art methods for traffic participant detection, including RVT, RED, and YOLOv8. We quantitatively evaluate the ability of event-based models to generalize on nighttime and unseen scenes. Our findings substantiate the compelling potential of leveraging event cameras for traffic monitoring, opening new avenues for research and application. eTraM is available at https://eventbasedvision.github.io/eTraM.

1. Introduction

In the dynamic landscape of modern transportation, Intelligent Transportation Systems (ITS) play a pivotal role in enhancing traffic flow [25, 43], route optimization [23, 40], and safety [13, 16]. As an essential task in ITS, traffic participant detection aims to provide information assisting in counting, measuring speed, identifying traffic incidents, and predicting traffic flow. The detection methods must be fast enough to operate in real-time and be insensitive to illumination change and varying weather conditions while utilizing less storage. For instance, in the span of just 1 s, a vehicle may travel over 8.3 m, and a pedestrian could cover over 1.43 m, leading to potential misses in fast-paced traffic scenarios and introducing motion blur concerns [41]. Moreover, nighttime and different weather conditions make the detection task more challenging as many features such as edge, corner, and shadow do not work due to varying illumination. In the face of these challenges, the integration of event cameras into ITS holds great promise for robust traffic participant detection in real-time scenarios.

Event cameras capture an asynchronous and continuous stream of "events" or pixel-level brightness changes instead of traditional static frames at fixed frequencies. Each event is represented by a tuple $\langle x, y, p, t \rangle$ corresponding to an illuminance change greater than a fixed thresh-

---

*Equal contribution

old at pixel location (x, y) and time t, with the polarity $p \in \{+1, -1\}$ indicating whether the illuminance increased or decreased. Event cameras with their exceptional temporal resolution (over $10,000$ fps) and high dynamic range (above 120 dB) [32] have prompted explorations into visual perception, robotics [15, 42], and its various applications in ITS [15, 33, 38].

Existing multimodal traffic datasets captured from various sensors, including RGB cameras, LiDAR, and Radar, have been utilized for several tasks in the context of autonomous vehicles (AV) [8, 10, 18, 37]. However, a largely unexplored yet promising domain lies in the use of event cameras for detection in traffic monitoring. This serves as an inspiration for us to contribute a first-of-its-kind, fully event-based traffic monitoring dataset.

In this paper, we present *eTraM* a novel fully event-based traffic perception dataset curated using the state-of-the-art (SOTA) high-resolution Prophesee EVK4 HD event camera [1]. The dataset spans over 10 hr of annotated data from a static perspective that facilitates comprehensive traffic monitoring. We consulted experts from the Institute of Automated Mobility (IAM) [3] and strategically mounted an event camera over selected sites (intersections, roadways, and local streets) to collect traffic data under diverse conditions. The data collection process was systematically conducted across various weather and lighting conditions spanning challenging scenarios such as high glare, overexposure, underexposure, nighttime, twilight, and rainy days. *eTraM* includes over 2M bounding box annotations of traffic participants such as vehicles (cars, trucks, buses, trams), pedestrians, and various micro-mobility (bikes, bicycles, wheelchairs) as shown in Figure 1.

eTraM offers the perspective of a static event camera captured at various scenes, further enhancing its versatility and applicability in real-world scenarios. This approach ensures that *eTraM* captures not only the routine dynamics of traffic but the nuances and challenges presented by a broad spectrum of scenarios and participants as well. We tested the diversity and quality of the dataset through various experiments and evaluated the generalization of event-based methods on nighttime and unseen scenes. *eTraM* stands as a valuable resource, propelling research and innovation in the evolving field of event-based traffic perception in ITS. Our contributions can be summarized as follows:

1. A first-of-its-kind fully event-based dataset from a static perspective encompassing a diverse variety of traffic scenarios (scenes, weather, and lighting conditions) and participants (vehicles, pedestrians, and micro-mobility).

2. Establish baselines using state-of-the-art event-based methods on *eTraM* across various traffic monitoring scenes and lighting conditions.

3. Quantitative evaluation of the generalization capabilities of event-based methods on nighttime and unseen scenarios.

The rest of this paper is structured as follows. Section 2 briefly summarizes the existing event-based dataset contributions. Section 3 presents *eTraM* dataset with details on the acquisition framework, preprocessing, and annotation. In Section 4, we provide baseline results and explore the generalization capabilities of event-based methods on nighttime and unseen scenarios. Section 5 highlights the advantages of utilizing event cameras and the difference between static and ego-motion event datasets.

2. Related Datasets

Early event-based datasets often involved the transformation of frame-based datasets into event streams. A noteworthy example is [27], where MNIST [21] and Caltech-101 [14] datasets were converted to event streams by moving an event camera in front of a screen displaying frames. Later works utilized event simulators [20, 31] to convert widely-used frame-based datasets into their event-based counterparts.

Ego-motion event-based datasets have seen a rise in recent years due to the increased accessibility of event sensors. DDD17 [4], DDD20 [19] pioneered the initial efforts in deploying event cameras for driving scenarios using a 346×260 px DAVIS sensor. These datasets, focused on steering angle prediction, have 12 and 51 hr of driving data, respectively. MVSEC [44] presents a multimodal stereo dataset using 346×260 px DAVIS sensors along with LiDARs, IMUs, and RGB cameras for three-dimensional perception tasks, such as feature tracking, visual odometry, and stereo depth estimation. DSEC [17] further expands these fusion efforts by including 390K annotations for detection tasks on an hour of multimodal stereo data using 640×480 px Prophesee Gen3.1 sensors.

Prophesee [29] introduced two substantial ego-motion datasets [12, 28] for detection. Gen1 Automotive Detection Dataset [12] encompasses a total of 255,781 manually annotated bounding boxes (228,123 cars and 27,658 pedestrians instances) acquired over a span of 39 hr using 304×240 px Prophesee Gen1 sensor. 1 Megapixel Automotive Dataset [28] stands out as the most comprehensive ego-motion event-based detection dataset. It encompasses 15 hr of recorded footage, featuring a resolution of 1280×720 px, with 25M generated bounding boxes. However, they are unable to provide nighttime annotations due to their automated labeling protocol. PEDRo dataset [6] focuses on people detection from a robotics perspective containing 43,259 bounding boxes from 119 recordings with an average duration of 18 s.

Dataset Name	Year	Duration (in hours)	Perspective		Traffic Participants			Lighting			Weather		No. of Bbox	Scenarios
			Ego	Static	VH	PED	MM	Day	Night	Twilight	Clear	Rainy		
DDD17 [4]	2017	12	✓		✓			✓	✓	✓	✓	✓	-	Driving
MVSEC [44]	2018	-	✓		✓			✓	✓		✓		-	Driving, Handheld
DVS Pedestrian [26]	2019	0.1		✓		✓		✓			✓	✓	4.6K	Walking street
DDD20 [19]	2020	51	✓		✓			✓	✓	✓	✓		-	Driving
Gen1 [12]	2020	39	✓		✓	✓		✓			✓		255K	Driving
1 Megapixel [28]	2020	15	✓		✓	✓	✓	✓	✓		✓		25M	Driving
DSEC [17]	2021	1	✓		✓	✓	✓	✓	✓		✓		390K	Driving
DVS-OUTLAB [5]	2021	7		✓			✓	✓			✓	✓	$\sim$ 47K	Playground
PEDRo [6]	2023	0.5	✓			✓		✓	✓	✓	✓	✓	43K	Robotics
eTraM (Ours)	2024	10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	$\sim$ 2M	Intersections, Roadways, Local streets

Table 1. A comprehensive overview of event-based traffic datasets from 2017 to 2024. (VH - Vehicle, PED - Pedestrian, MM - Micromobility)
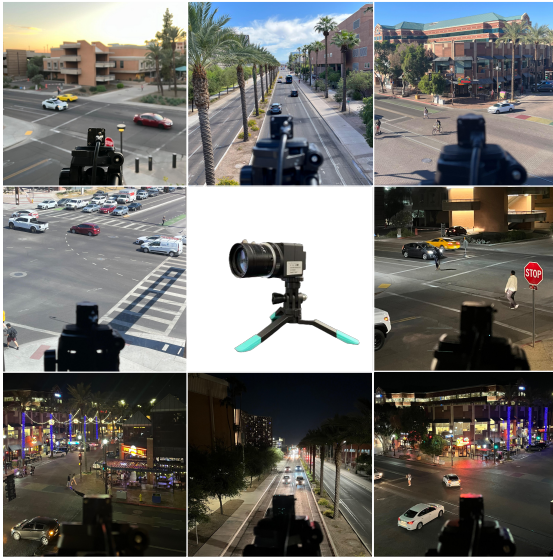

Figure 2. Data Collection Setup: The first four images from the top left display daytime data collection sites, the center image shows the Prophesee EVK4 HD event camera and the last four images depict nighttime data collection sites.

Static perception event-based datasets such as DVS-Pedestrian [26] is limited to pedestrian detection. It has 12 sequences recorded using a 346×260 px DAVIS346 camera and contains 4670 labeled instances of pedestrians. DVS-OUTLAB [5] explores the plausibility of using event cameras for long-time monitoring. It consists of recordings from three static 768×640 px CeleX-4 DVS event cameras featuring outdoor urban public areas involving persons, dogs, bicycles, and sports balls as objects of interest.

Table 1 provides an overview of existing event-based datasets. Interest in event cameras has been piqued for intelligent transportation, with many datasets focused on the ego-vehicle perspective, but there is a lack of datasets focusing on traffic monitoring from a static perspective.

3. *eTraM* Dataset

This section details *eTraM*'s acquisition framework, pre-processing techniques, annotation, and statistics, providing insights into its applicability for traffic monitoring.

3.1. Dataset Acquisition Framework

We rely on the Prophesee EVK4 HD event camera [1], notable for its high resolution (1280×720 px), temporal resolution (over $10,000$ fps), dynamic range (above 120 dB), and exceptional low light cutoff (0.08 Lux), to capture high-quality data. The event camera was strategically positioned at approximately 6 m with a pitch angle of about $35°$ to the ground. This configuration is deliberately chosen to maintain consistency with the placement of traffic cameras and to ensure comprehensive coverage of interactions between diverse traffic participants. Figure 2 shows different sites considered for data collection.

eTraM comprises recorded sequences from multiple intersections, roadways, and local streets around Arizona State University, Tempe campus. The sequences were recorded for intervals of $15-30$ min at different times of the day, covering daytime, nighttime, and twilight. The dataset also observes different weather conditions, including sunny, overcast, and rainy. To achieve this, extensive data collection efforts were carried out over a span of 8 months.

3.2. Preprocessing and Annotation

Given the sensitive nature of the event sensor, it was observed that nighttime data tends to exhibit higher levels of noise, primarily attributed to reflections and pointed sources of light from streets and vehicles. To address this challenge and enhance the quality of our data, the recorded sequences are passed through a spatiotemporal filter [7]. This spatiotemporal filter works on the idea that events from real objects should occur closer together in both space and time more often compared to events from random noises [15].
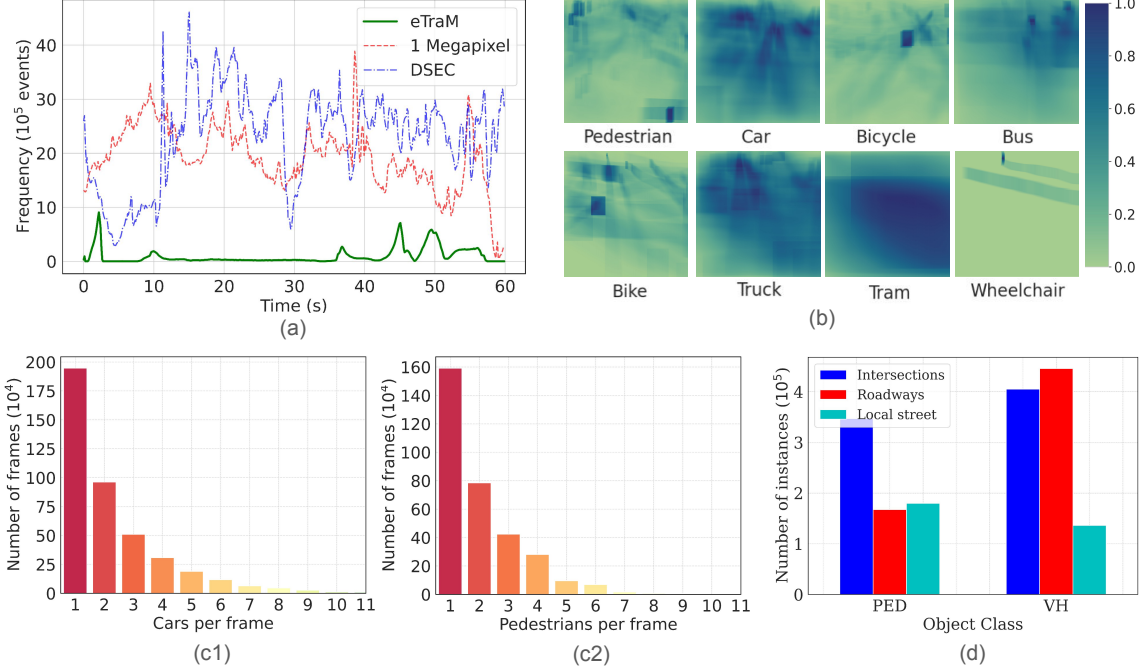
Figure 3. Event-time Distribution and Object Occurrence Statistics of *eTraM*: (a) Histogram of event frequency of *eTraM* (static event dataset) as compared to 1 Megapixel and DSEC (ego-motion event datasets), (b) Shows the object density of various classes across the frame, (c) Power-law distribution of the number of instances within an image for most predominant classes - cars (c1) and pedestrians (c2), and (d) Distribution of two major categories across various traffic sites.

The filtering mechanism discards an event if a threshold amount of events with the same polarity do not occur within a fixed temporal window in the vicinity of its 8-neighborhood spatial coordinates. For any $e = \langle x, y, p, t \rangle$, this condition is represented mathematically in Equation 1,

$$\sum_{t_i=t}^{t+\partial t} \sum_{x_i=x-1}^{x+1} \sum_{y_i=y-1}^{y+1} P(e_i = \langle x_i, y_i, p_i, t_i \rangle, e) > n_{thres} \tag{1}$$

where ∂t represents the temporal window and $P(e_i, e)$ is 1 only when the polarity of e_i and e are the identical. We choose a temporal window of 10 ms with a minimum threshold of 2 neighboring events for our experiments. The specific filter values were determined through experiments detailed in [5] and further validated through a smaller experiment conducted during the training phase. We illustrate the significance of spatiotemporal filtering through visualization in the supplementary material.

Following the denoising stage, events within the stream are partitioned into discrete time bins and consolidated into a single frame, thereby converting the asynchronous event stream into synchronous frames of 30 Hz. These frames are then annotated using CVAT [34], an open-source annotation tool. Our rigorous manual annotation process resulted in the precise identification of over 2M 2D bounding boxes.

3.3. Dataset Statistics

Here, we highlight key characteristics of the collected data and annotations. The dataset encompasses three distinct traffic monitoring scenes with 5 hr of intersection, 3 hr of roadway, and 2 hr of local street data sequences. Data for each scene is collected at multiple locations. For instance, the intersection scene contains data from 2 four-way, three-way, and an uncontrolled intersection. Each location has daytime, nighttime, and twilight data totaling up to 10 hr of data with 5 hr of daytime and nighttime data.

When comparing the event distribution in *eTraM* with other ego-motion event-based traffic datasets like 1 Megapixel Automotive and DSEC for a duration of 60 s, we observe that the number of events in *eTraM* is significantly lesser by a factor of 30, as shown in Figure 3 (a). This is accredited to the event camera being static in *eTraM*, primarily focusing on the dynamic traffic participants in a scene. In contrast, datasets from an ego-motion perspective capture more data due to the surrounding infrastructure's relative motion and lead to dense streams of events. The sparsity of events in *eTraM*, combined with the asynchronous nature of events, leads to low memory utilization, which is particularly advantageous for traffic monitoring infrastructure.

Traffic Site	Lighting	RVT				RED				YOLOv8			
		PED	VH	MM	All	PED	VH	MM	All	PED	VH	MM	All
Intersections	Daytime	0.460	0.813	0.315	**0.722**	0.395	0.593	0.284	0.545	0.167	0.293	0.111	0.190
Roadways		0.430	0.733	0.070	**0.627**	0.347	0.590	0.055	0.551	0.173	0.290	0.004	0.156
Local Streets		0.196	0.938	0.586	0.316	0.208	0.875	0.695	**0.351**	0.124	0.559	0.204	0.296
All Scenes		**0.304**	**0.781**	**0.403**	**0.572**	0.302	0.656	0.251	0.497	0.142	0.309	0.112	0.188
Intersections	Nighttime	0.161	0.465	-	**0.262**	0.149	0.425	-	0.242	0.071	0.375	-	0.149
Roadways		0.310	0.827	-	**0.739**	0.362	0.782	-	0.726	0.004	0.229	-	0.117
Local Streets		0.739	0.868	0.097	**0.829**	0.722	0.831	0.145	0.817	0.198	0.486	0.030	0.239
All Scenes		**0.317**	**0.674**	0.064	**0.523**	0.303	0.660	**0.083**	0.504	0.123	0.322	0.013	0.153
Overall		**0.309**	**0.717**	**0.313**	**0.539**	0.303	0.649	0.197	0.491	0.134	0.314	0.086	0.178

Table 2. Baseline Evaluation: Comprehensive evaluation of state-of-the-art tensor-based methods RVT, RED, and frame-based method YOLOv8 across various traffic sites (Intersections, Roadways, Local Streets) during both daytime and nighttime for PED - Pedestrian, VH - Vehicle, and MM - Micro-mobility.

eTraM contains over 2M instances of 2D bounding box annotations for traffic participant detection. These annotations additionally include object IDs, making it possible to evaluate multi-object tracking, as shown in the supplementary material. The annotation classes encompass a range of traffic participants, from various vehicles (cars, trucks, buses, and trams) to pedestrians and micro-mobility (bikes, bicycles, and wheelchairs).

Figure 3 (b) illustrates the spatial distribution of each class within the frame. We observe a uniform distribution of vehicle instances across the entire frame, while instances belonging to the pedestrian class cover more than 50% of the frame. This safeguards the model from developing a bias for classes in a specific region of the frame.

For accessibility and ease of use, *eTraM* is provided in multiple formats: RAW, DAT, and H5 [30]. Additionally, annotations are available in numpy format. The dataset is split into 70% training, 15% validation, and 15% testing, ensuring each subset has proportional data from each scene. Further statistics of *eTraM* can be seen in Figure 3 (c) & 3 (d). To the best of our knowledge, this stands as a first-of-its-kind event-based dataset for traffic monitoring. Additionally, the inclusion of nighttime data enhances its versatility for a broader range of research applications.

4. Dataset Evaluation

In this section, we present the results of experiments evaluating *eTraM* for real-world detection scenarios. We first train and evaluate models to establish baselines and assess their performance in various scenarios (Section 4.1). We further conduct experiments to test the ability of these models to generalize on nighttime conditions and unseen scenes (Section 4.2). For all experiments, we evaluate *eTraM* on SOTA tensor-based object detection meth-ods, specifically Recurrent Vision Transformers (RVT), Recurrent Event-camera Detector (RED), and a frame-based method, You Only Look Once (YOLOv8). This comparison helps us understand how SOTA utilizing dense tensor representation, performs compared to the conventional frame-based method. For evaluation, we focus on the three broad categories of traffic participants, vehicles (VH), pedestrians (PED), and micro-mobility (MM), and report the mean Average Precision at an intersection-over-union (IoU) threshold of 50% (AP50).

4.1. Baseline Evaluation

To assess the performance of detection methods on *eTraM*, we train them on 7 hr of data and evaluate them on 1.5 hr of validation and 1.5 hr of test data. RVT and RED were trained from scratch over 3 days on A100, while YOLOv8 was trained for 2 days. We conducted separate evaluations to provide insights into how each model performs in different scenes and lighting conditions. This allows us to understand how these models handle diverse and changing environmental contexts.

Several key observations emerged from the evaluation results of tensor-based models shown in Table 2. Notably, the performance of vehicle detection consistently outperforms that of pedestrians across all scenes and models. The performance in the micro-mobility category exhibits significant variance, likely due to the smaller number of instances and the broad range of subjects captured within this category. During the daytime, both vehicle and pedestrian detection results are at par in intersections and roadways. In local streets, the performance of vehicle detection improves compared to other scenes due to fewer instances and reduced occlusion. Conversely, we observe a decline in the performance of pedestrian detection due to higher pedestrian densities leading to more inter-class occlusions.
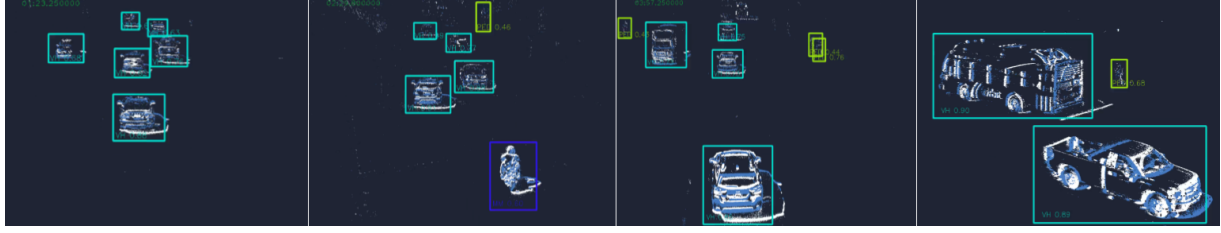
Figure 4. Qualitative Results: Showcasing detection of vehicles (cyan), pedestrians (yellow), and micro-mobility (blue) on *eTraM*.

During nighttime, the performance of vehicle detection in local streets and roadways remains consistent. Interestingly, pedestrian detection results improve significantly on local streets at night. This can be attributed to factors such as reduced pedestrian-to-pedestrian occlusion, instances being closer to the camera, and additional visual features like shadows that become more prominent at night due to street lights. Due to noise from various light sources, a drop in performance is observed at intersections during nighttime. Despite this, the performance during nighttime is at par with daytime scenarios, with an increase in performance on vehicle detection on roadways at night. This increase could be due to reduced inter-class occlusion as we observe fewer vehicles in nighttime conditions.

Furthermore, tensor-based methods such as RED and RVT, which incorporate temporal information through Recurrent Neural Networks in their architecture, consistently outperform the conventional frame-based method of YOLOv8. Figure 4 shows an example of detection on *eTraM*. The baseline results highlight the challenges and strengths of various traffic monitoring scenarios and categories, particularly showcasing the effectiveness of event-based models in nighttime conditions.

4.2. Generalization Evaluation

A key characteristic required for real-world deployment is that models demonstrate transferability to unseen scenarios. Since event cameras are invariant to absolute illuminance levels, we expect event-based models to also demonstrate transferability to nighttime data. We conduct experiments where we control the train and test set to evaluate their transferability in the following sections quantitatively.

4.2.1 Generalization on Night time

In [28], qualitative assessments were conducted to compare the generalization capabilities of event-based models against frame-based models in nighttime scenarios. We aim to quantitatively assess how well event-based models RVT and RED, trained on daytime data, can perform in nighttime conditions. We conducted a controlled experiment where we trained two models for each architecture, one on 2 hr of only daytime data and another with 2 hr of daytime along

with 45 min of nighttime data. Both models were evaluated on previously unseen nighttime data.

Train Set	RVT		RED	
	VH	PED	VH	PED
Day	0.566	0.166	0.374	0.354
Day+Night	0.761	0.254	0.673	0.422

Table 3. Evaluation of generalization capabilities of RVT and RED on night time data for PED - pedestrian and VH - vehicle class for models trained on only daytime and a combination of day and nighttime data.

The summarized results can be found in Table 3. We observe a consistent trend across every object class and architecture where models trained on data with nighttime sequences consistently outperform models trained solely on daytime sequences. Despite the expectation that event cameras would exhibit proficient performance in nighttime scenarios, these observations reveal that relying solely on daytime event data may not be sufficient. These models fail to attain performance levels comparable to those achieved by models trained on nighttime data with the model. This suggests that the unique challenges posed by nighttime conditions necessitate explicit training with relevant data to ensure optimal model performance. Upon closer inspection, we find that nighttime data has environmental interferences and distinct variations of noise. This could be inherent in nighttime data due to the heightened sensitivity of event cameras, which is a factor in this discrepancy.

While event-based models are unable to generalize on nighttime data out of the box, they have at par results on daytime and nighttime data when trained on the combined dataset (refer Table 2). This highlights the need for labeled nighttime data to allow event-based models to augment current frame-based systems.

4.2.2 Generalization on Unseen Scenes

To substantiate the ability of event-based detectors to generalize to previously unencountered traffic scenes, we train our model on a subset of the dataset. We evaluate the ability by testing each architecture on two independent test sets,

Figure 5. Demonstrating Effectiveness of Event Camera for Traffic Scenarios: Yellow circle (top row) tracks a car that halts at a stop sign with a lack of motion captured in the third frame, red circle (bottom row) tracks a car that violates the stop sign where motion is continuously captured in every frame. Additionally, a green arrow (top row) shows a car traveling at a high speed, resulting in a high event density.

| Test | RVT | | RED | |
Set	VH	PED	VH	PED
Held In	0.449	0.316	0.556	0.521
Held Out	0.628	0.529	0.572	0.509

Table 4. Evaluation of generalization capabilities of RED and RVT on unseen traffic scenarios for PED - pedestrian and VH - vehicle tested on held in and held out test set.
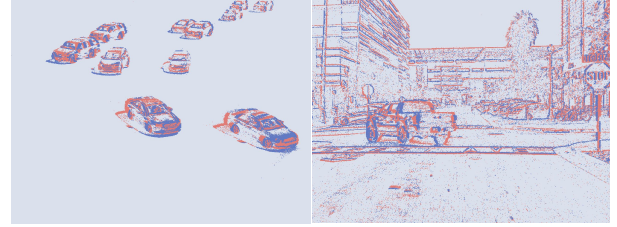

Figure 6. Qualitative Comparison: Events captured from a static event camera (left) show enhanced visibility of moving vehicles compared to an ego-motion event camera (right).

one "held in" that contains sequences from intersections that our model has seen during the training phase and the other on the "held out" test set with data from an intersection that is skipped during training.

As depicted in Table 4, we observe a comparable performance of the models across both major categories. On evaluating the model on the "held out" test set, representing an entirely new and unseen traffic scene, the model's generalization capability is evident. These values are at par with the performance on the "held in" test set. This similarity in results highlights the model's ability to seamlessly extend its learning to previously unseen traffic scenes, validating its transferability across changing environments. This generalization to unseen intersections is pivotal for real-world deployment.

5. Discussion

5.1. Event Cameras in Traffic Monitoring

In this section, we discuss the advantages of using event cameras to augment traditional frame camera systems for traffic monitoring.

Traditional cameras take continuous snapshots at a fixed frequency, potentially capturing individuals' identities and sensitive information. In comparison, event cameras register events with no color and texture information, significantly reducing the probability of gathering sensitive in-

formation of any individual [5, 9, 28]. Further, the advantages of event cameras extend to their performance in low light conditions and their robustness towards mitigating motion blur. As we demonstrate through our baseline evaluations in Table 2, event cameras display equally superior performance in nighttime conditions while maintaining a high temporal resolution that aids it in substantially reducing motion blur [32]. An essential requirement of traffic monitoring is the need for resource-efficient sensors. Since event cameras are designed to operate on a sparse data stream generated by significant visual changes, they exhibit lower memory requirements and power consumption compared to traditional frame cameras [2, 35]. This makes event cameras sustainable and cost-effective for continuous monitoring over extended periods.

We further highlight the capabilities brought by event cameras in traffic monitoring scenarios through two practical applications, as illustrated in Figure 5. Firstly, in detecting stop sign violations, event cameras excel in capturing instantaneous changes in the visual scene, enabling precise detection of vehicles coming to a halt. The number of events associated with the object makes it easy to discern whether a moving vehicle is slowing down or has stopped. Secondly, due to the high temporal resolution of the event camera and the event density corresponding to an object, it is straight-
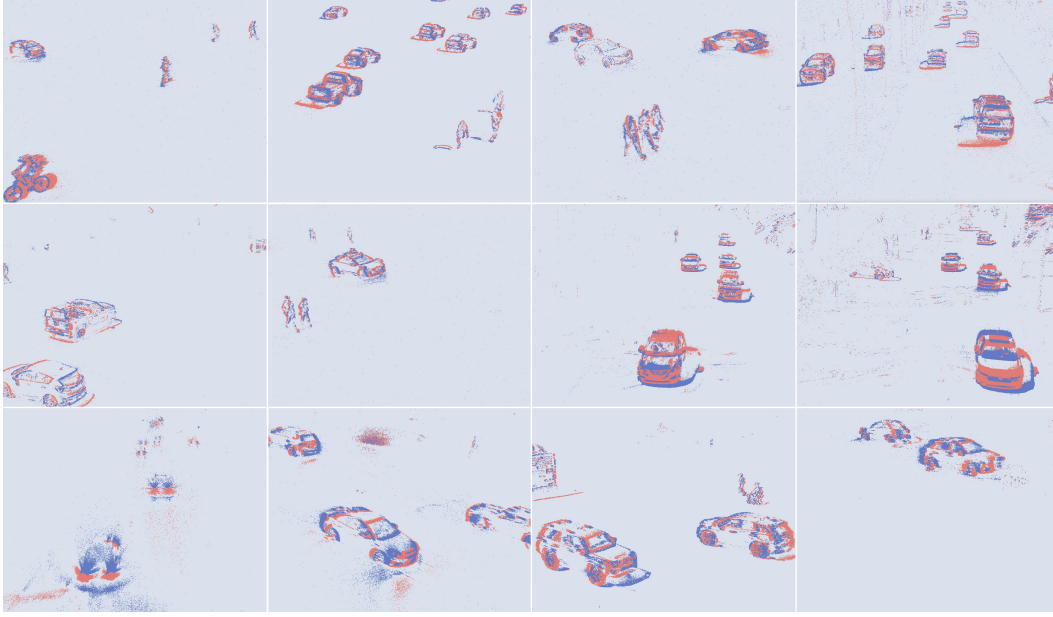
Figure 7. Traffic Site Diversity in *eTraM*: We show the various instances encapsulating the interactions amongst multiple traffic participants captured from a static roadside perspective with daytime (first row), twilight (second row), and nighttime (last row) showing increasing sensor noise (top to bottom) due to light sources such as headlights and street lights.

forward to detect sudden acceleration or erratic behavior of fast-moving traffic participants. This capability is critical for identifying potentially hazardous situations on the road, such as near misses, and allowing for prompt intervention or alert generation.

5.2. Static and Ego Event-based Datasets

Next, we discuss the difference between static and ego-motion event-based datasets, emphasizing the significance of the former for traffic monitoring. In static datasets events are more likely to be associated with relevant traffic participants rather than extraneous background changes seen in ego-motion data, enhancing the visibility of these participants. This enhanced visibility is illustrated in Figure 6, which provides a qualitative comparison between events captured from a static and an ego-motion event camera. As shown, dynamic traffic participants from the static camera are better visible than those from the ego-motion camera. Another distinction is the field of view provided by static roadside datasets. Positioned at an elevated height, these cameras capture a broader scene and include far-away objects, while ego-motion datasets are constrained to the vehicle's immediate surroundings. The elevated perspective of static datasets proves advantageous, enabling the detection of traffic participants at a distance and providing a comprehensive view of the traffic environment, as seen in Figure 7. This would be beneficial to a broad spectrum of applications, like traffic scene estimation [22, 24, 39], road safety monitoring [36] and traffic signal control [11].

6. Conclusion

We present *eTraM*, a large-scale manually annotated event-based dataset for traffic monitoring. Our meticulously curated dataset, captured using the cutting-edge Prophesee EVK4 HD high-resolution event camera, provides new traffic detection and tracking opportunities from a static perspective. With over 10 hr of annotated event data spanning various scenes and lighting conditions, our dataset offers insights into the complex dynamics of traffic scenarios. The comprehensive annotations include over 2M bounding boxes of various traffic participants, from vehicles to pedestrians and micro-mobility. Further, nighttime generalization results highlight the value of labeled nighttime data, enabling event-based models to achieve daytime-equivalent performance on nighttime data. We demonstrate the ability of event-based models to generalize effectively to unseen scenarios, emphasizing its potential in real-world traffic monitoring. As the field of event-based sensing continues to advance, *eTraM* holds the potential to serve as an invaluable resource in driving its development and presenting new opportunities in intelligent transportation systems.

7. Acknowledgment

References

[1] Event Camera Evaluation Kit 4 HD IMX636 Prophesee-Sony. 2, 3

[2] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, Jeff Kusnitz, Michael Debole, Steve Esser, Tobi Delbruck, Myron Flickner, and Dharmendra Modha. A low power, fully event-based gesture recognition system. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7388–7397, 2017. 7

[3] Arizona Commerce Authority. IAM - Arizona Institute of Automated Mobility. 2

[4] Jonathan Binas, Daniel Neil, Shih-Chii Liu, and Tobi Delbruck. Ddd17: End-to-end davis driving dataset, 2017. 2, 3

[5] Tobias Bolten, Regina Pohle-Fröhlich, and Klaus D. Tönnies. Dvs-outlab: A neuromorphic event-based long time monitoring dataset for real-world outdoor scenarios. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1348–1357, 2021. 3, 4, 7

[6] 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. 2, 3

[7] Tobias Brosch, Stephan Tschechne, and Heiko Neumann. On event-based optical flow detection. *Frontiers in neuroscience*, 9:137, 2015. 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 (CVPR)*, June 2020. 2

[9] 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. 7

[10] 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

[11] Longchao Da, Minchiuan Gao, Hao Mei, and Hua Wei. Prompt to transfer: Sim-to-real transfer for traffic signal control with prompt learning. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI'24)*, 2024. 8

[12] Pierre de Tournemire, Davide Nitti, Etienne Perot, Davide Migliore, and Amos Sironi. A large scale event-based detection dataset for automotive, 2020. 2, 3

[13] Wenlu Du, Junyi Ye, Jingyi Gu, Jing Li, Hua Wei, and Guiling Wang. Safelight: A reinforcement learning method toward collision-free traffic signal control. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 14801–14810, 2023. 1

[14] Li Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006. 2

[15] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *CoRR*, abs/1904.08405, 2019. 2, 3

[16] Marichelo Garcia-Venegas, Diego Alberto Mercado-Ravell, Luis A Pinedo-Sanchez, and Carlos A Carballo-Monsivais. On the safety of vulnerable road users by cyclist detection and tracking. *Machine Vision and Applications*, 32(5):109, 2021. 1

[17] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 2021. 2, 3

[18] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 2

[19] 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, 2020. 2, 3

[20] 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 (CVPR) Workshops*, pages 1312–1321, June 2021. 2

[21] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2

[22] Xiaoliang Lei, Hao Mei, Bin Shi, and Hua Wei. Modeling network-level traffic flow transitions on sparse data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 835–845, 2022. 8

[23] Siyuan Liu, Yisong Yue, and Ramayya Krishnan. Adaptive collective routing using gaussian process dynamic congestion models. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 704–712, 2013. 1

[24] Duo Lu, Eric Eaton, Matt Weg, Wei Wang, Steven Como, Jeffrey Wishart, Hongbin Yu, and Yezhou Yang. Carom air-vehicle localization and traffic scene reconstruction from aerial videos. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10666–10673. IEEE, 2023. 8

[25] Hao Mei, Xiaoliang Lei, Longchao Da, Bin Shi, and Hua Wei. Libsignal: An open library for traffic signal control. *Machine Learning*, pages 1–37, 2023. 1

[26] Shu Miao, Guang Chen, Xiangyu Ning, Yang Zi, Kejia Ren, Zhenshan Bing, and Alois Knoll. Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection. *Frontiers in Neurorobotics*, 13, 2019. 3

[27] Garrick Orchard, Ajinkya Jayawant, Gregory Cohen, and Nitish Thakor. Converting static image datasets to spiking neuromorphic datasets using saccades, 2015. 2

[28] Etienne Perot, Pierre de Tournemire, Davide Nitti, Jonathan Masci, and Amos Sironi. Learning to detect objects with a 1 megapixel event camera. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 16639–16652. Curran Associates, Inc., 2020. 2, 3, 6, 7

[29] Prophesee. Prophesee - Metavision for Machines. 2

[30] Prophesee S.A. Prophesee Documentation - File Formats. © Copyright Prophesee S.A - All Rights Reserved. 5

[31] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto, editors, *Proceedings of The 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pages 969–982. PMLR, 29–31 Oct 2018. 2

[32] 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, 2021. 2, 7

[33] J.P. Rodríguez-Gomez, A. Gómez Eguíluz, J.R. Martínez-de Dios, and A. Ollero. Asynchronous event-based clustering and tracking for intrusion monitoring in uas. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8518–8524, 2020. 2

[34] Boris Sekachev, Nikita Manovich, Maxim Zhiltsov, Andrey Zhavoronkov, Dmitry Kalinin, Ben Hoff, TOsmanov, Dmitry Kruchinin, Artyom Zankevich, DmitriySidnev, Maksim Markelov, Johannes222, Mathis Chenuet, a andre, telenachos, Aleksandr Melnikov, Jijoong Kim, Liron Ilouz, Nikita Glazov, Priya4607, Rush Tehrani, Seungwon Jeong, Vladimir Skubriev, Sebastian Yonekura, vugia truong, zliang7, lizhming, and Tritin Truong. opencv/cvat: v1.1.0, Aug. 2020. 4

[35] Rafael Serrano-Gotarredona, Matthias Oster, Patrick Lichtsteiner, Alejandro Linares-Barranco, Rafael Paz-Vicente, Francisco Gomez-Rodriguez, Luis Camunas-Mesa, Raphael Berner, Manuel Rivas-Perez, Tobi Delbruck, Shih-Chii Liu, Rodney Douglas, Philipp Hafliger, Gabriel Jimenez-Moreno, Anton Civit Ballcels, Teresa Serrano-Gotarredona, Antonio J. Acosta-Jimenez, and BernabÉ Linares-Barranco. Caviar: A 45k neuron, 5m synapse, 12g connects/s aer hardware sensory–processing– learning–actuating system for high-speed visual object recognition and tracking. *IEEE Transactions on Neural Networks*, 20(9):1417–1438, 2009. 7

[36] Anshuman Srinivasan, Yoga Mahartayasa, Varun Chandra Jammula, Duo Lu, Steven Como, Jeffrey Wishart, Yezhou Yang, and Hongbin Yu. Infrastructure-based lidar monitoring for assessing automated driving safety. Technical report, SAE Technical Paper, 2022. 8

[37] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang,

[38] Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. *CoRR*, abs/1912.04838, 2019. 2

[38] Abhishek Tomy, Anshul Paigwar, Khushdeep S. Mann, Alessandro Renzaglia, and Christian Laugier. Fusing event-based and rgb camera for robust object detection in adverse conditions. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 933–939, 2022. 2

[39] Hua Wei, Chacha Chen, Chang Liu, Guanjie Zheng, and Zhenhui Li. Learning to simulate on sparse trajectory data. In *Machine Learning and Knowledge Discovery in Databases: Applied Data Science Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part IV*, pages 530–545. Springer, 2021. 8

[40] Hua Wei, Dongkuan Xu, Junjie Liang, and Zhenhui Jessie Li. How do we move: Modeling human movement with system dynamics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4445–4452, 2021. 1

[41] Zhuang Zhang, Lijun Zhang, Dejian Meng, Luying Huang, Wei Xiao, and Wei Tian. Vehicle kinematics-based image augmentation against motion blur for object detectors. Technical report, SAE Technical Paper, 2023. 1

[42] Xu Zheng, Yexin Liu, Yunfan Lu, Tongyan Hua, Tianbo Pan, Weiming Zhang, Dacheng Tao, and Lin Wang. Deep learning for event-based vision: A comprehensive survey and benchmarks, 2023. 2

[43] Fan Zhou, Liang Li, Ting Zhong, Goce Trajcevski, Kunpeng Zhang, and Jiahao Wang. Enhancing urban flow maps via neural odes. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence,{IJCAI} 2020*, 2020. 1

[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