Edge Hybrid Neuro-Statistical Learning for Event-based Roadside Visual Motion Detection and Tracking

Cristian Axenie
Technische Hochschule Nürnberg
Keßlerplatz 12, 90489
Nürnberg, Germany
cristian.axenie@th-nuernberg.de

Ertan Halilov
Technische Hochschule Nürnberg
Keßlerplatz 12, 90489
Nürnberg, Germany

Julian Main
Technische Hochschule Nürnberg
Keßlerplatz 12, 90489
Nürnberg, Germany

David Weiss
Technische Hochschule Nürnberg
Keßlerplatz 12, 90489
Nürnberg, Germany

ABSTRACT

The Vision Zero Program's purpose is to reduce traffic-related fatalities and serious injuries while promoting equitable, safe, and healthy mobility for all. Ultimately, the challenge is to detect pedestrians during the day and especially at night in order to implement safety measures. The current study introduces an award-winning low-power solution employing neuromorphic visual sensing and hybrid neuro-statistical processing developed by the Technische Hochschule Nürnberg team for the TinyML Vision Zero San Jose Competition. The solution proposes a novel neuromorphic edge fusion of spiking neural networks and event-based expectation maximization for the detection and tracking of pedestrians and bicyclists. We provide a deployment-ready evaluation of the detection performance along with robustness, energy footprint, and weatherization while emphasizing the advantages of the neuro-statistical edge solution and its city-level scaling capabilities.

Author Keywords

Spiking Neural Networks; Event-based Vision; Expectation Maximization; Neuromorphic Sensors; Neuromorphic Computing; Visual Detection; Visual Tracking; Edge Systems.

CCS Concepts

•Computing methodologies → Machine learning; Object detection; •Computer systems organization → Embedded systems;

NeuroEdge - Efficient AI at the Edge Workshop co-located with HiPEAC'24 Conference, January 18, 2024, Munich, Germany

ACM ISBN 978-1-4503-2138-9.

DOI: 10.1145/1235

INTRODUCTION

For urban and metropolitan areas, population growth is expected to continue in the coming years. Cities and municipalities will continue to strive to provide residents with a sense of security and freedom. Camera-based monitoring systems are currently the most mature solution for urban surveillance, for example, in monitoring road traffic and, of course, detecting abnormal pedestrian motion and activity on the roadside. However, most existing and established algorithms and systems for urban road surveillance require expensive and power-hungry computer hardware. Neuromorphic sensors and computing systems, which are the focus of our award-winning pedestrian detection solution, are a promising alternative. The key advantages of such systems are high energy efficiency, fast and representative (relevant) data acquisition, fast edge processing, improved data and privacy protection, and rationally budgeted use of resources.

"Vision Zero"[1] is a street safety policy that strives for the elimination of traffic fatalities for all transportation modes and has been implemented in multiple locations worldwide. It advocates for cost-effective and accurate technical solutions used to detect pedestrians, bicyclists, or electric scooter riders during the day and especially at night in order to implement various measures for improved safety. Moreover, such systems should provide robust detection across a wide spectrum of weather conditions (e.g., from rainy overcast to clear sky) and also specific time of day variations in luminance (i.e., sunrise or sunset). To sum up, valuable candidate TinyML solutions for Vision Zero need to have an excellent energy footprint, deliver accurate and robust detection, and only use

[1] VisionZero information is available at https://visionzeronetwork.org/about/what-is-vision-zero/

local processing, all this with a budget that allows scaling to the city level [2].

STATE-OF-THE ART IN PEDESTRIAN DETECTION AND TRACKING

Video-based activity detection systems have a wide range of applications in public space monitoring, such as crowd monitoring [3], public safety [18], traffic monitoring [19], and individual safety [8]. All these approaches share the common goal of automatically detecting and localizing abnormal events to avoid the tedious and time-consuming work of manual detection. Common human detection approaches can be broadly classified into three types: Object Trajectory Methods [1, 13, 24], Global Pattern Methods [25], and Grid Pattern Methods [27, 12, 26, 20, 14, 23]. Object trajectory-based approaches typically split the scene into individual objects before tracking or identifying them. Most global-pattern methods extract low- or mid-level features from the video and evaluate the entire sequence based on features that retain the characteristic aspects of normal behaviour. Techniques based on grid patterns divide frames into smaller regions and examine patterns in each block independently using sparse decomposition or other numerical methods. Although current approaches work well, on the one hand, they rely heavily on complex computer vision technology and require a significant amount of labelled data and computational resources for online and offline model training [15, 11]. Moreover, when collecting data in surveillance situations with cameras, it is inevitable to collect redundant data from the environment. On the other hand, based on the camera images, identification of the captured person is possible, which is problematic in terms of data protection and privacy. Neuromorphic vision sensors, such as the Dynamic Vision Sensor (DVS) [21], are biologically inspired sensors that, in contrast to regular cameras, only report brightness differences of the individual, independent pixels (called events) asynchronously at the moment they occur. Unlike conventional cameras, which capture a complete image at a specific frame rate, DVS offers significant advantages, such as covering a wide dynamic range and providing high temporal resolution in the sub-millisecond range. Moreover, since differences in light intensity caused by moving objects lead to events in the scene, neuromorphic image sensors are natural motion detectors that filter out all redundant information in traffic monitoring systems [9]. However, current neuromorphic algorithms perform poorly compared to conventional models [4]. A first, simple step in developing pedestrian detection and tracking systems that use event-based visual data representation and processing is to adapt existing algorithms for person detection and tracking to event stream processing. Using machine learning or traditional (geometric) machine vision methods, there are several algorithms, such as [4], that use input from a neuromorphic event-based camera to create new features or descriptors [22]. Such systems benefit from the event-based representation, but on the other hand, they also introduce increased computational complexity to the algorithm that "combines" the events to create extended and rich features suitable for the existing algorithms [10]. Unfortunately, these approaches lose one of the most important advantages of event-based processing, namely the very high temporal resolution. Other approaches to event-based processing address "native" event-based algorithms and propose frameworks for aggregating events and manipulating event-based descriptors and features for efficient processing, including detection/recognition [2], tracking [9], encryption [16], and even classification with few shots [17]. The latter describes the current state of the algorithms and their advantages for streaming data [28, 5, 29, 6, 7]. In our project, we propose a new approach in which event streams from the neuromorphic camera are processed simultaneously by a Spiking Neural Network (SNN), used for detection, and a statistical learning algorithm (i.e., Expectation Maximization (EM)), used for tracking. The system is capable of ingesting DVS events and extracting spatiotemporal patterns of their evolution in the perceived roadside scene. The system is able to learn a representation of the sparse input data context (i.e., for the SNN detection) and capture the spatiotemporal signature of the observed scene and the motion it contains (i.e., pedestrian/biker motion detection) on a per-event basis using the event-based EM. In practice, our proposed solution for pedestrian detection is a real-time edge processing system running entirely on an embedded neuromorphic computing platform that accelerates the detection and tracking of pedestrians and bicyclists. For more information about the project and our solution design, architecture, and deployment performance, you can access the TinyML Vision Zero San Jose Competition Final at https://www.youtube.com/watch?v=ZhBCtfalcOk&t=2872s&ab_channel=tinyML.

SOLUTION ARCHITECTURE

Our proposed solution for the TinyML Vision Zero Challenge employed a DVS camera and an embedded Brainchip Akida development board based on a Raspberry PI IO Board carrier. The overall system architecture is depicted in Figure 2 and the data flow describing the functional architecture of the solution is depicted in Figure 1. The pure event-based tracking deployment is based on a custom-designed and coded Expectation Maximization (EM), while the SNN detection is based on minimally aggregated frames of events fed to an Object Detection FOMO (Faster Objects, More Objects) using MobileNetV2 from the Edgepulse Library. The SNN detector was developed in the EdgeImpulse Studio and evaluated along with other models for both CPU and Brainchip Akida targets. We tested both to validate the approach and used it in testing the embedded deployment on the Brainchip Akida Raspberry PI embedded system. For both the SNN detection and the EM tracking, we consider the physics of the pedestrians' motion and extract those specific constraints that characterize human bodies (i.e. reduced skeleton structure) from events data. For the SNN we embed the physics in the loss function whereas the EM embeds it explicitly in the likelihood computation.

Detection using Spiking Neural Networks

The proposed network architecture for the SNN deployment is a custom Brainchip Akida FOMO network. Using the EdgeImpulse Studio[3], this SNN consists of a U-Net, a dense connectivity convolutional neural network, and output classifier layers.

[2] TinyML Challenge information is available at https://www.tinyml.org/event/tinyml-hackathon-2023-pedestrian-detection/

[3] Edge Impulse Studio: https://edgeimpulse.com/

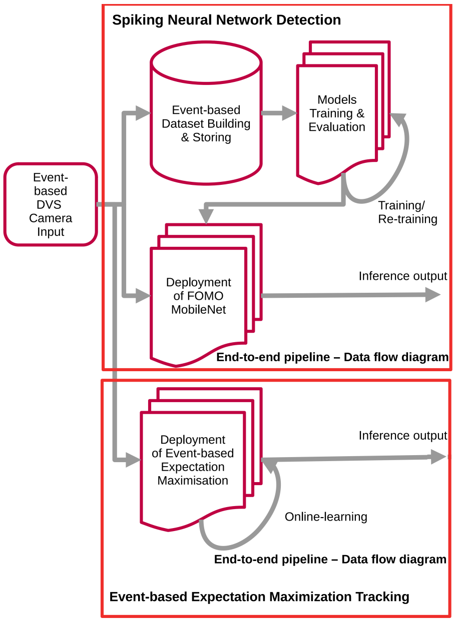

Figure 1. Solution data-flow. The event-based Dynamic Vision Sensor (DVS) data is fed to both the detection module using a Spiking Neural Network and the tracking module using event-based Expectation Maximization. The Spiking Neural Network is a MobileNetv2-based Fast Object More Objects (FOMO) neural network that is trained on aggregated DVS events, is then quantized, and is running on the Brainchip Akida. The event-based Expectation Maximization is a custom and novel implementation fed with single DVS events running on the embedded Raspberry PI board.

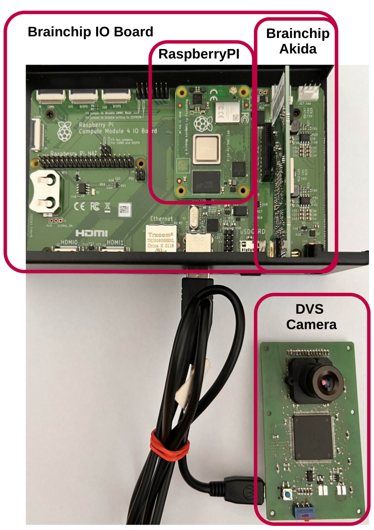

Figure 2. Solution deployment hardware architecture. Sensing is done with DVS and edge processing using an embedded Raspberry PI carrier equipped with a Brainchip Akida neuromorphic computing board.

One reason to choose this architecture is three-fold: 1) the relatively limited amount of training data; 2) the improved pedestrian segmentation and more accurate pedestrian/bike classification, and 3) the suitable mobile deployment on the neuromorphic computing system (i.e., MobileNetV2 is designed for mobile and embedded applications). In our implementation, the SNN efficiency comes from the internal split of the event frames into grids (or heat maps) and the injection of an image classification-like technique to classify the grids independently based on the physics of the tracked traffic participants embedded in the loss function (see Figure 4). The SNN structure in our solution is provided in Figure 3.

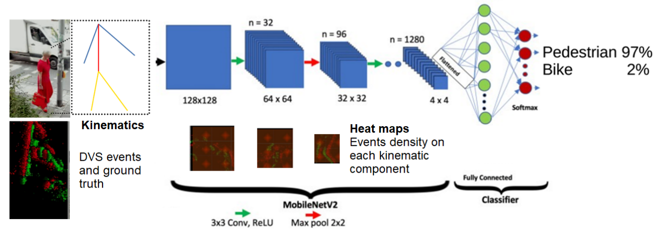

Figure 3. Detection SNN architecture and data flow. Physics constraints are embedded in the system which then uses a modified MobileNetV2 to perform classification. Note that the model is quantized using the MetaTF library from Brainchip before running it natively on the Brainchip Akida neuromorphic chip.

Tracking using Event-based Expectation Maximization

Using the same physics-informed learning principle as infused in the SNN, the event-based EM tracking algorithm uses each incoming event to update the belief of the system by maximizing the log-likelihood of belonging to the skeleton (i.e., to one of the defined segments therein). More precisely, in the event-based EM, the events are coloured to mark their membership to one of the skeleton segments (i.e., blue (head), red (trunk), and yellow (legs)). The algorithm assigns membership to a skeleton segment on a per-event basis. The membership is computed by maximizing the log-likelihood of the event belonging to a Gaussian distribution on a skeleton segment, respectively. More precisely, we embed in the mode only the angles and use the geometric transformations of the pedestrian's lower body (i.e., the kinematics). We also include a prediction model that uses (angular) velocities to distinguish bicyclists from other traffic participants. In the EM algorithm, the expectation step computes the likelihood that the incoming event belongs to one of 4 possible sources (i.e., left pedestrian leg, right pedestrian leg, pedestrian trunk, or noise, respectively), as shown in Figure 4. In the maximization step, the algorithm maximizes the log-likelihood of the model by minimizing the squared error between the events and the tracked motion model.

SOLUTION EVALUATION

In the current section, we introduce the experimental setup with an emphasis on accuracy, latency, robustness, and energy footprint. Please note that the entire experimental codebase used in the study is available at: `https://github.com/th-nuernberg/spider`. The codebase contains both the SNN and the event-based EM modules. For step-by-step guidelines on how to install and operate the software components, we invite the readers to access the end-to-end demo available at `https://www.youtube.com/watch?v=B52wQP1YRVY`.

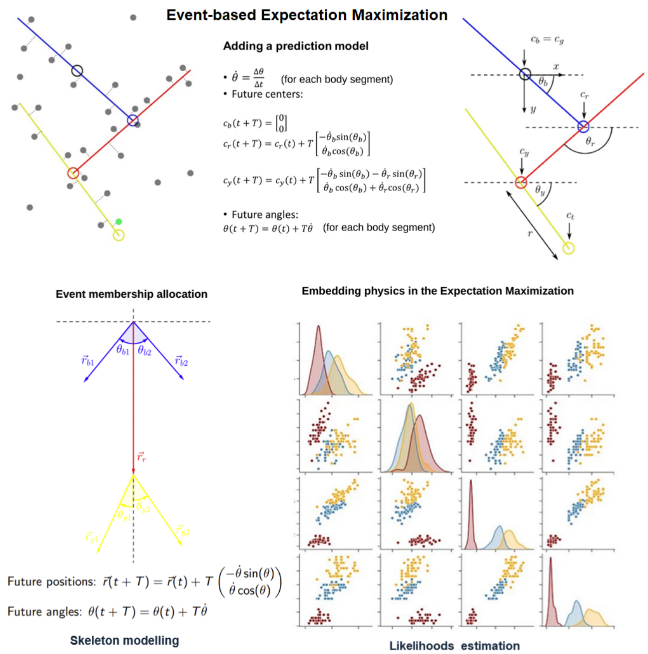

Figure 4. Tracking event-based EM and the kinematics modelling embedded in the system. Mapping motion to the steps of the statistical learning algorithm and visual depiction of the likelihood computation.

Datasets

DVS camera events were used for both the detection and tracking modules of our system. In order to capture ground truth information, we added a traditional CMOS camera on the DVS data acquisition mount with a preset offset capturing the same field of view (i.e., the offset was used in the computation of the tracking metrics evaluation). The frames were only used as a support in the labelling process and not included in the training of the system, the testing, or the evaluation, respectively. We have chosen multiple locations with different properties of the road, the number of pedestrians and bicyclists, road geometries, weather conditions, and time of day. An overview of the datasets' location and sample data is provided in Figure 5. All the acquired and used event-based datasets, along with camera ground truth and sample recordings, are available on Zenodo at: `https://zenodo.org/records/10102416`.

Evaluation of the accuracy

For the accuracy evaluation of both the detection and tracking components of the system, we performed multiple experiments in order to get a statistically valid analysis. For the quantitative evaluation, we executed the following protocol: 1) Read relevant detection and tracking data from each of the experiments (i.e. N = 30 experiments); 2) Perform statistical tests (i.e., a combination of omnibus ANOVA and posthoc pairwise T-test with a significance $p = 0.05$) and adjust the ranking of experiments depending on significance; 3) rank subsets of relevant metrics (i.e., the metrics with a significant difference, e.g., the F1 score for detection and four others tracking specific metrics). Additionally, we also provide the readers with a qualitative evaluation of the detection components (see Figure 6) and the tracking component (see Figure 7). The quantitative evaluation of the detection system using an SNN is provided as a confusion matrix and F1-Score values in Table 1, whereas the quantitative evaluation of the EM tracking system using typical metrics is provided in Table 2.

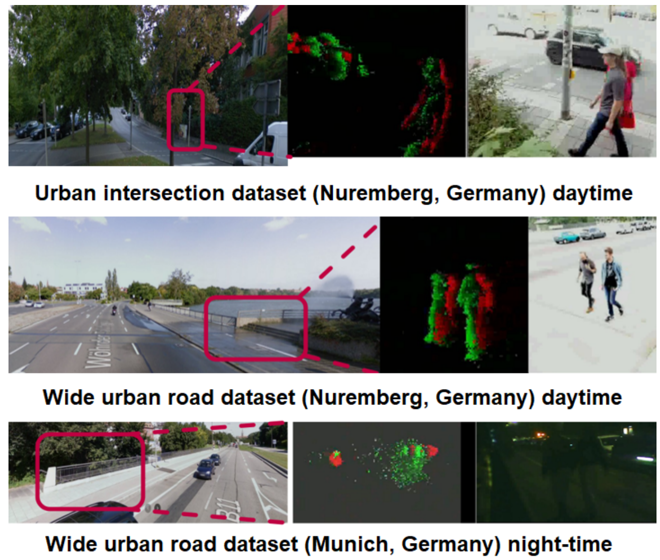

Figure 5. Datasets overview.
***Dataset 1*: 4 lanes (4 per direction) wide street; Google Maps Location `https://goo.gl/maps/JaYGwaTaBHj5H6SL9`; 50 km/h (urban) speed limit, near university campus with pedestrians, people biking, scooting, rolling, etc; Overcast operating environment.**
***Dataset 2*: 8 lanes (4 per direction) wide street; Google Maps Location: `https://goo.gl/maps/jar6AjysZiM2LP5S7`; 50 km/h (urban) speed limit; Near the main train station of the city and a location with people walking/running/jogging and people biking, scooting, rolling, etc.; Ideal operation;**
***Dataset 3*: 6 lanes (3 per direction) wide street on a bridge; Google Maps Location: `https://goo.gl/maps/SEEsmpgmLPcD8fG7A`; 50 km/h (urban) speed limit; Near ring street of Munich and a location with people walking/running/jogging and people biking, scooting, rolling, etc; night time data acquisition.**

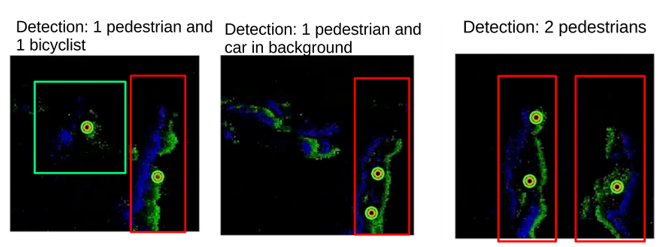

Figure 6. Qualitative evaluation of the detection component. The system uses an SNN fed with aggregated events from the DVS camera to detect the different traffic participants (i.e. bikers, pedestrians) based on heat maps of events. Each detection is marked by two points depicting the physical constraints (i.e. body segments) embedded in the model. Note that the colouring of the events follows the convention: ON events (green) and OFF events (blue), bounding boxes for visualization purposes only.

As for the SNN detection, the relevant metrics are the ones for classification (i.e., confusion matrix and F1-score), for the tracking we employed some specific evaluation metrics. The

Dataset	Background %	Bicyclist %	Pedestrian %
Dataset 1 (daytime)			
Background	99.70	0.26	0.04
Bicyclist	12.10	87.90	0.00
Pedestrian	6.20	0.00	93.80
F1-Score	1.00	0.62	0.77
Dataset 2 (daytime)			
Background	97.50	2.20	0.80
Bicyclist	10.10	89.90	0.00
Pedestrian	3.20	0.00	96.80
F1-Score	1.00	0.70	0.87
Dataset 3 (night)			
Background	99.07	0.30	0.00
Bicyclist	21.10	78.00	0.90
Pedestrian	10.20	13.00	76.80
F1-Score	0.90	0.60	0.70

Table 1. Quantitative evaluation of the SNN detection system. Confusion matrix and the F1 score of the SNN on the validation datasets.

Dataset	Bicyclist	Pedestrian
Dataset 1 (daytime)		
Track Matching Error(%)	13.70	10.10
Tracking Time Delay(s)	0.08	0.03
Tracking Detection Rate(%)	95.00	98.00
Tracking Completeness(s)	0.38	0.25
Dataset 2 (daytime)		
Track Matching Error(%)	11.20	8.21
Tracking Time Delay(s)	0.07	0.02
Tracking Detection Rate(%)	97.00	99.00
Tracking Completeness(s)	0.38	0.25
Dataset 3 (night)		
Track Matching Error(%)	23.30	20.10
Tracking Time Delay(s)	0.09	0.08
Tracking Detection Rate(%)	76.00	79.00
Tracking Completeness(s)	0.84	0.76

Table 2. Quantitative evaluation of the EM tracking system using specific metrics.

Object Tracking Time delay is the estimated delay between the algorithm's detection of a bicyclist or pedestrian and that of the ground truth. Tracking Matching Error is the positional error between the EM prediction interpolated trajectory (i.e., successive likelihood computation of the centre of mass of each body segment) and the ground truth trajectory, and the smaller the number, the better the tracking accuracy. The Tracking Detection Rate is the precision defined as the ratio of Total True Positives/ Total Number of ground truth tracks. Finally, Track Completeness is defined as the time for which the EM prediction trajectory overlaps with the ground truth trajectory divided by the total duration of the ground truth trajectory. An important note to make regarding the tracking validation is that, through its operation principle the event-based camera will not capture the pedestrian/bicyclists which are static (i.e., not relevant). Basically, the end-point of the trajectories, used in the tracking metrics, are chosen based on the ground truth trajectory.

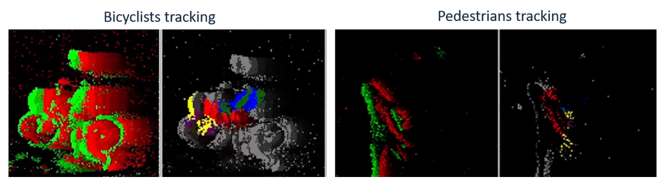

Figure 7. Qualitative evaluation of the tracking component. The system uses an event-based EM algorithm to track the different types of traffic participants (i.e., bikers, pedestrians) based on their physics in a purely event-based fashion. Note that the events are coloured to mark their membership to one of the skeleton segments (i.e. blue (head), red (trunk), and yellow (legs)) and the colouring of the algorithm input events follows the convention: ON events (green) and OFF events (red).

Evaluation of the power consumption and the inference latency

The main design goal for our solution was a very low energy footprint while keeping a balance with the accuracy and the minimal inference latency of the edge system. We evaluated, next to accuracy (using the same systematic procedure), the average power consumption and the inference latency. The experimental setup is presented in Figure 8 and the quantitative results in Table 3, respectively.

Evaluation of the deployment robustness

As the solution needed to include weatherization and performance deterioration analysis, we systematically tested in the overall evaluation the performance (i.e. only inference latency) degradation with temperature increase. The overall testing setup is depicted in Figure 8 and the results are compiled in Table 4, respectively.

As reproducing the hardware setup might be difficult for all the readers (i.e., a custom DVS camera and interfacing with the Brainchip Akida Development Board), we kindly invite the readers to view a full system demo available at `https://www.youtube.com/watch?v=B52wQP1YRVY`. This video describes the steps used in the software development, the parametrization and execution of each component (i.e., detection and

Dataset	Power(W)	Latency(ms)
Dataset 1 (daytime)	7.58	14.32
Dataset 2 (daytime)	4.92	8.21
Dataset 3 (night)	5.65	24.62

Table 3. Quantitative evaluation of the system power consumption and inference latency. Average power is computed as "wall power" summing up the DVS sensor and the Raspberry Pi Brainchip Akida board. Average inference latency is measured in software based on the input propagation through the SNN for detection and EM-based tracking. The measurements follow the systematic assessment described for the accuracy analysis.

tracking), the power and weatherization test, and the actual live operation of the system.

ECONOMIC & SOCIAL RELEVANCE OF THE SOLUTION

The societal impact of the project is very high, as it contributes to feasible, rapid, and affordable improvements in roadside safety in large urban areas. With the planned immediate deployment of the technology that leverages and improves existing infrastructure (i.e., gantry mounts), our solution also offers high economic benefits. The three main socially relevant aspects are 1) detection and identification of abnormal pedestrian and bicyclist events along public roads in real-time, regardless of weather conditions (e.g. clouds, sun, rain) or time of day (e.g. dusk, dawn, or night); 2) already known and monitored "hot spots" can now be augmented by the solution, which preserves the privacy of the detected and tracked individuals yet reliably detects their abnormal motions; 3) only salient features of the public road visual scene are extracted while preserving privacy (i.e., no identity reconstruction from DVS events) so that only relevant cues are used in deciding whether the scene is normal or abnormal. Using the Brainchip Akida neuromorphic platform to process event-based vision data on-site reduces system costs by a factor of 20 compared to conventional centralized GPU processing [4] A Brainchip Akida system with a DVS event camera consumes at most 7.6 W when processing camera data on-site. Conventional real-time vision processing systems (connected sometimes also with centralized data centres) require specialized GPUs operating at around 400W per high-speed camera. Hence, the neuromorphic design enables at least a 100-fold power reduction; per expected 100 million devices, this continuously saves 40 GW[5]. In addition, on-site processing reduces transmitted data from about 30MB/sec in conventional systems to

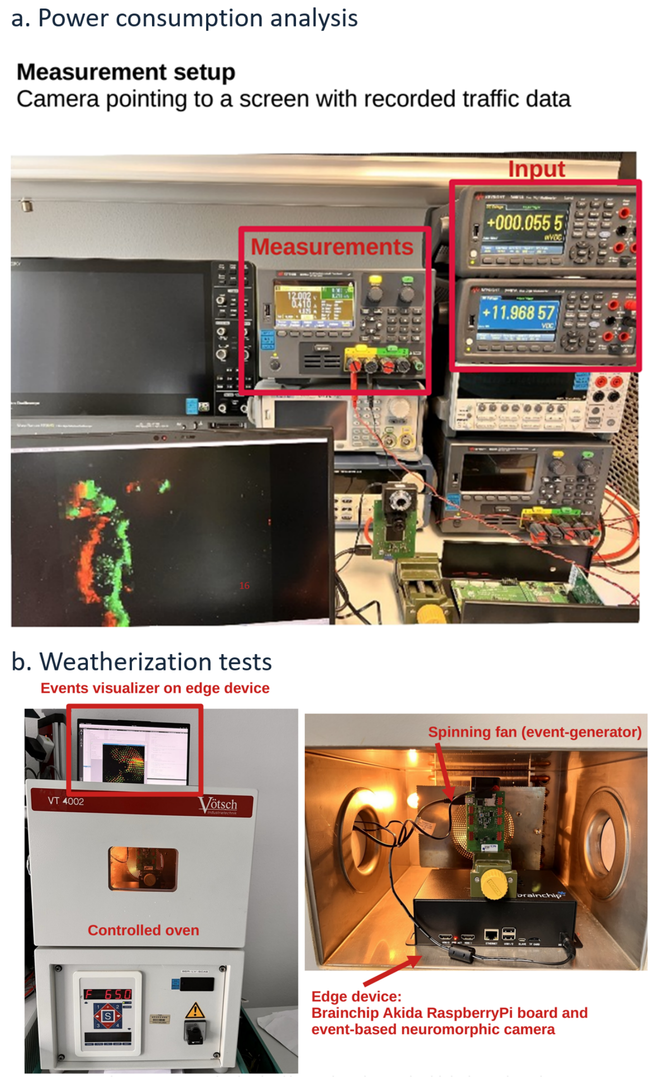

Figure 8. Weatherization and power consumption setup. a) Power consumption analysis: in order to test the robustness of the end-to-end system we performed various power measurements in the closed loop. We fed recorded ground truth to the DVS and measured how the power and performance changed in the presence of voltage drops or changes. b) Weatherization tests: we have tested the performance variation when increasing the operating temperature from outdoor (6'C), to room temperature (21'C) to up to a heated oven (65'C). Instead of the screen, the DVS camera captures a very high number of events generated by the oven fan and then further propagated for processing in order to check inference time/latency variation under extreme weather conditions.

300 B/sec, a 10,000-fold reduction[6]. Event cameras detect changes around 33 times faster than conventional cameras and can be used at night and in poor weather conditions due to their wide dynamic range of up to 140 dB.

CONCLUSIONS AND OUTLOOK

In this paper, we describe an award-winning solution for neuromorphic edge processing for roadside monitoring. Our so-

[4] In large quantities: event-based camera costs $50, the custom neuromorphic board developed by Brainchip with RaspberryPI costs $300; compared to GPU cost and a price of $1000 per high-speed camera.

[5] 100×106 units × (400 W - 4 W) = 39.6 GW

[6] If we assume 25 bytes per object (8 for timestamp, 16 for camera coordinates, 1 for class), 10–12 objects are detected per second: 25×12 = 300 B

Dataset	Temperature('C)	Latency(ms)
Dataset 1 (daytime)		
outdoor	6	16.72
lab	21	14.32
oven	65	18.94
Dataset 2 (daytime)		
outdoor	6	12.09
lab	21	8.21
oven	65	18.53
Dataset 3 (night)		
outdoor	6	30.56
lab	21	24.62
oven	65	34.78

Table 4. Quantitative evaluation of the system power consumption and inference latency. Average power is computed as "wall power", summing up the DVS sensor and the Raspberry Pi Brainchip Akida board. Average inference latency is measured in software based on the input propagation through the SNN for detection and EM-based tracking. The measurements follow the systematic assessment described for the accuracy analysis and are swept over three temperature ranges.

lution is a creative endeavour in which event streams from a neuromorphic camera are processed simultaneously by an SNN, used for detection, and an event-based EM, used for the tracking of pedestrians and bicyclists. By exploiting the underlying spatiotemporal patterns of the event-based data in the perceived roadside scene, the system learns a representation of the sparse input data context based on the imposed physics of motion. Our analysis, experiments, and preliminary deployment demonstrate that our solution is a promising candidate for roadside real-time neuromorphic edge detection and tracking, delivering robust and accurate performance, a very good energy footprint, and supporting city-level scaling.

REFERENCES

[1] Sk Arif Ahmed, Debi Prosad Dogra, Samarjit Kar, and Partha Pratim Roy. 2018. Trajectory-based surveillance analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 7 (2018), 1985–1997.

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

[3] Muhammed V Anees and Santhosh G Kumar. 2018. Deep learning framework for density estimation of crowd videos. In *2018 8th International Symposium on Embedded Computing and System Design (ISED)*. IEEE, 16–20.

[4] Lakshmi Annamalai, Anirban Chakraborty, and Chetan Singh Thakur. 2019. Evan: Neuromorphic event-based anomaly detection. *arXiv preprint arXiv:1911.09722* (2019).

[5] Lakshmi Annamalai, Vignesh Ramanathan, and Chetan Singh Thakur. 2022. Event-LSTM: An unsupervised and asynchronous learning-based representation for event-based data. *IEEE Robotics and Automation Letters* 7, 2 (2022), 4678–4685.

[6] Cristian Axenie, Radu Tudoran, Stefano Bortoli, Mohamad Al Hajj Hassan, and Goetz Brasche. 2019. NARPCA: Neural Accumulate-Retract PCA for Low-Latency High-Throughput Processing on Datastreams. In *Artificial Neural Networks and Machine Learning–ICANN 2019: Theoretical Neural Computation: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part I 28*. Springer, 253–266.

[7] Cristian Axenie, Radu Tudoran, Stefano Bortoli, Mohamad Al Hajj Hassan, Alexander Wieder, and Goetz Brasche. 2020. SPICE: Streaming PCA fault identification and classification engine in predictive maintenance. In *Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part II*. Springer, 333–344.

[8] Bo-Wei Chen, Chen-Yu Chen, and Jhing-Fa Wang. 2013. Smart homecare surveillance system: Behavior identification based on state-transition support vector machines and sound directivity pattern analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 43, 6 (2013), 1279–1289.

[9] Guang Chen, Peigen Liu, Zhengfa Liu, Huajin Tang, Lin Hong, Jinhu Dong, Jörg Conradt, and Alois Knoll. 2020. Neuroaed: Towards efficient abnormal event detection in visual surveillance with neuromorphic vision sensor. *IEEE Transactions on Information Forensics and Security* 16 (2020), 923–936.

[10] Yunhua Chen, Weijie Bai, Qingkun Huang, and Jinsheng Xiao. 2021. Efficient Motion Symbol Detection and Multikernel Learning for AER Object Recognition. *IEEE Transactions on Cognitive and Developmental Systems* 14, 4 (2021), 1544–1552.

[11] Yong Shean Chong and Yong Haur Tay. 2017. Abnormal event detection in videos using spatiotemporal autoencoder. In *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part II 14*. Springer, 189–196.

[12] Yang Cong, Junsong Yuan, and Ji Liu. 2013. Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition* 46, 7 (2013), 1851–1864.

[13] Serhan Coşar, Giuseppe Donatiello, Vania Bogorny, Carolina Garate, Luis Otavio Alvares, and François

Brémond. 2016. Toward abnormal trajectory and event detection in video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 3 (2016), 683–695.

[14] Xiaoguang Cui, Yuan Tian, Lubin Weng, and Yiping Yang. 2014. Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition. In *Fifth International Conference on Graphic and Image Processing (ICGIP 2013)*, Vol. 9069. SPIE, 142–148.

[15] Allison Del Giorno, J Andrew Bagnell, and Martial Hebert. 2016. A discriminative framework for anomaly detection in large videos. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*. Springer, 334–349.

[16] Bowen Du, Weiqi Li, Zeju Wang, Manxin Xu, Tianchen Gao, Jiajie Li, and Hongkai Wen. 2021. Event encryption for neuromorphic vision sensors: Framework, algorithm, and evaluation. *Sensors* 21, 13 (2021), 4320.

[17] Anik Goyal and Soma Biswas. 2021. Few-shot classification without forgetting of event-camera data. In *Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing*. 1–8.

[18] Chuanping Hu, Gengjian Xue, Lin Mei, Li Qi, Jie Shao, Yanfeng Shang, and Jian Wang. 2017. Building an intelligent video and image analysis evaluation platform for public security. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–6.

[19] Razie Kaviani, Parvin Ahmadi, and Iman Gholampour. 2014. Incorporating fully sparse topic models for abnormality detection in traffic videos. In *2014 4th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 586–591.

[20] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. 2013. Anomaly detection and localization in crowded scenes. *IEEE transactions on pattern analysis and machine intelligence* 36, 1 (2013), 18–32.

[21] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. 2008. A 128x128 120 dB 15us latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits* 43, 2 (2008), 566–576.

[22] Peigen Liu, Guang Chen, Zhijun Li, Huajin Tang, and Alois Knoll. 2022. Learning Local Event-based Descriptor for Patch-based Stereo Matching. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 412–418.

[23] Xuan Mo, Vishal Monga, Raja Bala, and Zhigang Fan. 2013. Adaptive sparse representations for video anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology* 24, 4 (2013), 631–645.

[24] Luis Patino, James Ferryman, and Csaba Beleznai. 2015. Abnormal behaviour detection on queue analysis from stereo cameras. In *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–6.

[25] Oluwatoyin P Popoola and Kejun Wang. 2012. Video-based abnormal human behavior recognition—A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 865–878.

[26] Venkatesh Saligrama and Zhu Chen. 2012. Video anomaly detection based on local statistical aggregates. In *2012 IEEE Conference on computer vision and pattern recognition*. IEEE, 2112–2119.

[27] Hela Sfar and Amel Bouzeghoub. 2018. Activity recognition for anomalous situations detection. *IRBM* 39, 6 (2018), 400–406.

[28] Evangelos Stromatias, Miguel Soto, Teresa Serrano-Gotarredona, and Bernabé Linares-Barranco. 2017. An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. *Frontiers in neuroscience* 11 (2017), 350.

[29] Aaron Voelker, Ivana Kajić, and Chris Eliasmith. 2019. Legendre memory units: Continuous-time representation in recurrent neural networks. *Advances in neural information processing systems* 32 (2019).