An Energy-Efficient Daily Surveillance System with DVS-CIS Sensor Fusion and Event-based NPU Triggering

Mincheol Cha[*§], Keehyuk Lee[*§], Bobaro Chang[*], Soosung Kim[‡], Daniel Moon[‡], Taeho Lee[*], Tae Sung Kim[†], Hyuk-Jae Lee[*], Hyunsurk Ryu[*‡] Xuan Truong Nguyen[*],
[*]Department of Electrical and Computer Engineering, Seoul National University, South Korea [†]Dept. of Electronic Engineering, Sun Moon University, South Korea [‡]Neuro Reality Vision Corp., South Korea
{mccha, lkh099, bbrchang, taehov, hjlee, truongnx}@capp.snu.ac.kr, ts7kim@sunmoon.ac.kr,
{kimsoosung, dmoon0410, eric.ryu}@nrv.kr

***Abstract*—This study presents a daily surveillance system based on dynamic vision sensors (DVS) and CMOS image sensors (CIS) to enable real-time image recognition with low energy consumption. In such a system, a neural processing unit (NPU) - which executes a DNN model to detect objects on a given CIS image - may consume a lot of energy when always-on. To address the problem, this work introduces a system with a DVS-based region of interest (ROI) detector and an event-based NPU trigger for energy savings. Based on DVS, the ROI detector effectively recognizes scene changes in dynamic environments, e.g., low-light scenes at midnight, which serves as a trigger to invoke the NPU for object detection. Our system prototype was built on a host PC and two Xilinx Zynq+ ZCU106 FPGA boards, one for the DVS-CIS receiver and the other for our NPU. The experimental results demonstrated that Over a 24-hour testing period, our system achieved a 31.5% reduction in energy usage. Operating a YOLOv3-Tiny object detector at 200 MHz, our NPU achieves a latency of just 18 ms, enabling seamless real-time monitoring capabilities.**

***Index Terms*—DVS-CIS Sensor Fusion, NPU, Daily Surveillance, FPGA, Energy Efficiency.**

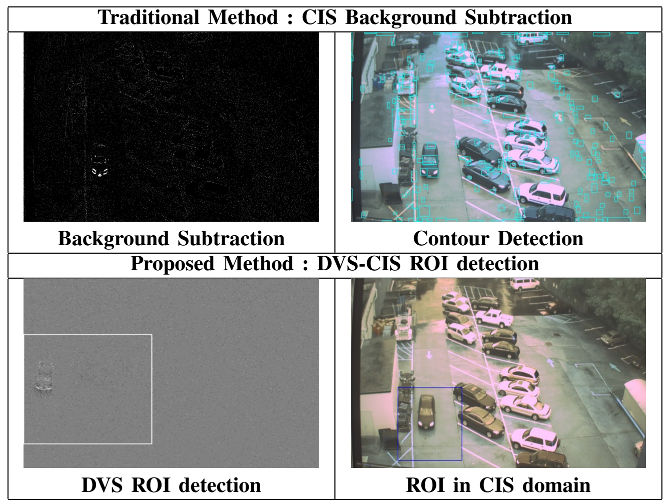


TABLE I: Example of Three ROI Detectors.

I. Introduction

Recent advancements in image recognition systems have transformed fields like autonomous driving, surveillance, and robotics [13], [14]. These are motivated by emerging developments in sensory technology and deep neural networks (DNNs). For example, Dynamic Vision Sensors (DVS) [2], [11], [19], [20] have gained considerable attention due to their complementary capabilities in real-time visual processing. Inspired by the biological visual system, DVS technology has evolved to detect changes in light intensity on a per-pixel basis. DVS can capture fast and dynamic visual information at a submillisecond frame latency [4], [10], [23], enabling DVS to compensate conventional CMOS Image Sensors (CIS) effectively in practical scenarios such as daily surveillance.

On the other hand, DNNs are evolved and widely adopted for many applications, including image recognition and object detection [1], [17], [21]. DNNs generally come with high computational costs and large memory footprints. Therefore, many research efforts have been made to design high-throughput and cost-effective DNN accelerators, including network-dedicated architectures [8], [12] and generic architectures [6], [15]. For example, [12] proposed a MAC design and dataflow to reduce filter switching activity, improving energy efficiency. To reduce external memory accesses (EMAs), [8] proposed a fusion-like dataflow for some layers in the early stages of YOLO networks that come with large feature maps. [6], [7], [15] selectively use row-based and frame-based weight reuse schemes w.r.t a DNN layer type, significantly reducing EMAs.

Owning to these aforementioned advancements, integrating a DVS-CIS sensory system and a neural processing unit (NPU) appears promising for many daily-life applications. For example, a CCTV or daily surveillance system is generally associated with many sensors and processing units, which highly demands an energy-efficient solution. Specifically, since sensors are statically installed in a daily surveillance system, many approaches are proposed to leverage regions of interest (ROIs) for computation and energy reduction. For example, background subtraction can remove the background, enabling focusing ROIs only. Unfortunately, this may perform poorly

[§]Equal contribution

in many practical scenarios, as shown in Table I. One well-known reason is that CIS integrates light over time, causing motion blur and less precise ROI detection. Meanwhile, DVS registers brightness changes at the pixel level at a high framerate, for instance, 10,000 fps, eliminating motion blur for sharper edges, as shown in Table I. However, DVS lacks color information and resolution compared to CIS.

Inspired by these observations, this work presents an energy-efficient surveillance system with DVS-CIS sensor fusion (DCSF) and an event-based NPU trigger (EBNT). The contributions of this work can be summarized as follows.

- **Architecture and Algorithm:** We proposed an energy-efficient surveillance system with a DVS-CIS stream receiver and an NPU. We proposed an ROI detection algorithm based on DVS streams, enabling identifying ROIs in various practical scenes. The DVS-based ROI detector serves as an event-based trigger to occasionally activate NPU for energy reduction.
- **Evaluation:** We built a prototype system with a host PC and two Xilinx ZCU106 FPGA boards [22], one for the DVS-CIS receiver and the other for NPU.0ver a 24-hour testing period, our system experimentally achieved a result of 31.5% reduction in energy usage.

II. Background and Motivation

A. Case with Always-on NPUs in a Surveillance System

Surveillance systems typically use NPUs to support multiple cameras, but keeping them always on can waste energy. For example, at night, the system must continuously monitor the environment for security, even when scene changes are minimal. As a result, NPUs stay active, consuming power even without detecting movement or target objects.

B. An Opportunity for Energy Saving with DVS-CIS Fusion

DVS technology plays a critical role in modern machine learning applications due to its exceptional dynamic range, high temporal resolution, low power consumption, and frame rates of up to 10,000 fps [11]. As an example, Samsung's 640x480 DVS sensor shows a tiny energy usage of 27mW at 100k events per second (eps), while producing 300 Meps at 50mW energy usage [19], [20].

DVS-CIS fusion appears as a promising energy-efficient solution in a daily surveillance system. For example, DVS with the aforementioned unique characteristics can quickly identify a scene change, even in dynamic and extreme environments such as midnight. Specifically, a DVS-based ROI detector accurately determines if an NPU execution is necessary, saving energy by dynamically and intentionally deactivating NPUs.

C. An Opportunity and a Challenge with DVS-Based ROI Detection

Utilizing DVS for ROI detection is computationally effective. Specifically, DVS introduces highly sparse data since it only captures light changes. For example, leveraging DVS data, algorithm 1 bins events along the row and column directions, yielding a rectangle drawn by horizontal and vertical edges of frequent event density [3]. However, this is tightly optimized for particular patterns, e.g., hands, which may be unable to encompass objects with irregular-shaped boundaries in a surveillance system. More importantly, as the average of events over each frame is used as a threshold, it becomes noise-susceptible in event-sparse and textured environments.

Algorithm 1: Baseline ROI Detection Algorithm [3]

Require: N stacked DVS frames: $frame\_data$
for each column j in frame **do**
  col(j) = number of events that j holds
  if col(j) $>$ avg(col(j)) for $j, j+1, ..., j + roi\_line\_width - 1$ **then**
    if $x\_min = 0$ **then**
      $x\_min \leftarrow j$
    else if $j > x\_max$ **then**
      $x\_max \leftarrow j$
Repeat for each row to obtain (x, y, w, h).

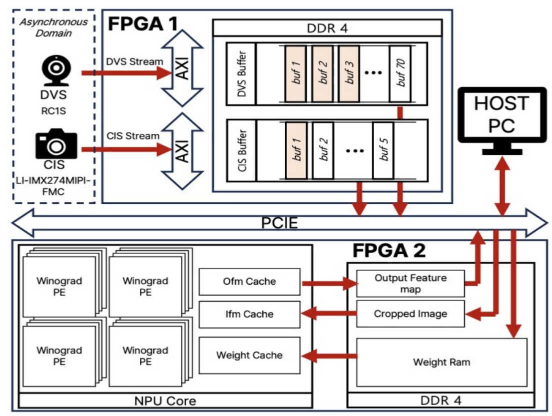


Fig. 1: Overall Sensor Fusion Model Architecture.

Inspired by the observations, this work proposes an energy-efficient surveillance system with DVS-CIS sensor fusion (DCSF) and an event-based NPU trigger (EBNT), as explained in the following section.

III. Proposed System Architecture

A. Sensor Fusion Model Architecture

The proposed sensor fusion system processes real-time data streams from DVS and CIS using a two-FPGA architecture, as shown in Fig. 1. FPGA 1 is responsible for data acquisition and buffering, with asynchronous DVS and CIS streams being processed independently through separate AXI channels. The DVS stream sends high-frequency frame data transactions of up to 2,000fps, which is accommodated by 70 dedicated buffers in the DVS Buffer module. In contrast, the CIS stream uses a lower frame rate of 60fps and is stored in 5 CIS Buffers. Both sets of buffered data are stored in shared DDR4 memory and then transferred via PCIe to the host PC where further processing takes place before being sent to FPGA2.

FPGA 2 houses the NPU Core, which utilizes Winograd Processing Elements (PEs) that perform Winograd Convolution algorithm to accelerate convolution layers [9]. Integration of the Winograd PE reduces power consumption of the NPU, since Winograd Convolution reduces the number of required multipliers while maintaining throughput. Such power efficiency makes the architecture well-suited for real-time low power sensor fusion and object detection applications. The detection results are stored in DRAM and can be retrieved by the Host PC through PCIe for display.

B. Advanced DVS ROI Detection for Irregular Objects

To delineate the ROI around event occurrences in the DVS, we propose a linear algorithm (Algorithm 2) that counts events in stacked DVS frames, applying a penalty to intervening non-events. This approach identifies a pixel range with active events per row, filtering out noise to define a minimal ROI encompassing all such rows. Unlike the baseline (Algorithm 1 [3]), our method remains robust, even for cases with minimal percentage of event area regardless of target shape or size.

C. Energy and Accuracy Optimization with DVS-CIS System

This section mathematically explains how a DVS-CIS sensor fusion system reduces the total energy consumption. For a given NPU implementation π, and fixed-size inputs from the DVS as $\mathbf{X}_{\text{DVS}}$ and CIS as $\mathbf{X}_{\text{CIS}}$, we represent the energy consumption $\mathbf{E}$ as follows:

$$\mathbf{E}(\pi, \mathbf{X}_{\text{CIS}}) = \mathbf{P}(\pi, \mathbf{X}_{\text{CIS}}) + \mathbf{P}(\mathbf{X}_{\text{CIS}}) \tag{1}$$

$$\begin{aligned}\mathbf{E}(\pi, \mathbf{X}_{\text{CIS, DVS}}) = {} & \mathbf{P}(\pi, \mathbf{X}_{\text{CIS}}) * \mathbf{f}(\pi, \mathbf{X}_{\text{CIS, DVS}}) \\ & + \mathbf{P}(\pi) * (1 - \mathbf{f}(\pi, \mathbf{X}_{\text{CIS, DVS}})) \\ & + \mathbf{P}(\mathbf{X}_{\text{CIS}}) + \mathbf{P}(\mathbf{X}_{\text{DVS}})\end{aligned} \tag{2}$$

where $\mathbf{f}(\pi, \mathbf{X})$ represent the activation frequency of π given input $\mathbf{X}$, $\mathbf{P}(\pi, \mathbf{X})$ represents the hardware power consumption in Watts of running input $\mathbf{X}$. Additionally, $\mathbf{P}(\pi)$, $\mathbf{P}(\mathbf{X})$ represents the static power used from π and the static power to achieve the input $\mathbf{X}$. According to Eqs 1, 2, the overall power can be suppressed by preventing the NPU from running when no significant ROI exists, i.e. when $\mathbf{f}(\pi, \mathbf{X}_{\text{CIS, DVS}})$ is low.

Algorithm 2: Proposed ROI Detection Algorithm

Require: N stacked DVS frames : $frame\_data$
Initialize variables: (x, y, w, h)
for each row i in frame **do**
  for each column j in frame **do**
    if $frame\_data[i][j]$ has events **then**
      $cur\_score \leftarrow cur\_score + roi\_event\_score$
    else
      $cur\_score \leftarrow cur\_score - 1$
    if $cur\_score < 0$ **then**
      $cur\_score \leftarrow 0, cur\_x \leftarrow j$
    if $cur\_score > max\_score$ **then**
      $max\_score \leftarrow cur\_score$
      $x\_range \leftarrow [cur\_x, j]$
  if $max\_score \geq roi\_min\_score$ **then**
    $[x\_min, x\_max] \leftarrow \bigcup x\_range$
    if so for roi_line_width consecutive rows **then**
      (x, y, w, h) covers $[x\_min, x\_max]$

IV. Evaluation

A. Methodology and Experiment Setup

System Implementation. We implemented our prototype system with a host PC and two Xilinx ZCU106 FPGA boards [22]. The first board is used for a DVS-CIS streaming receiver, which is connected to two sensor boards: (1) a custom DVS sensor [18] and (2) a LI-IMX274MIPI-FMC CIS camera [5]. For visualization, a host PC can receive and display DVS and CIS frames streamed from the FPGA. The second board is used for our NPU. Both boards are connected to the host PC via PCIe. The system setting is shown in Fig. 2.

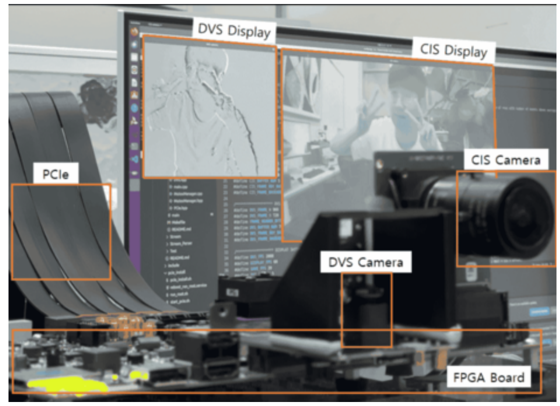


Fig. 2: The Prototype System Setup.

DVS-CIS Stream Receiver. Following the specifications [5], [18], the CIS sensor is configured at the FHD resolution at 60 fps and the DVS sensor is configured at the 960x720 resolution at 2,000 fps. The operating clock frequency of the core is set to 300 MHz The bus data width is set to 64 bits, resulting in peak DMA bandwidth of 19.2 Gb/s. The DVS-CIS receiver is lightweight, consuming only 69.5K LUTs, 86.6K FFs, 134.5 BRAM blocks of 36Kb, 16 URAMs, and 38 DSPs.

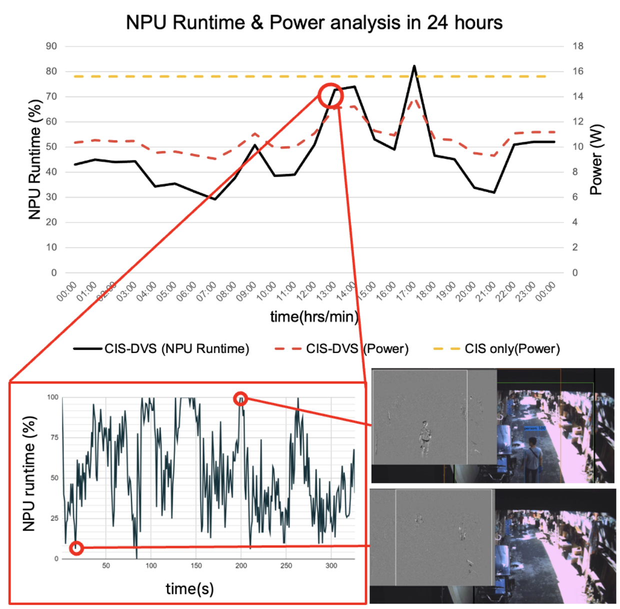


Fig. 3: Runtime and Power Performance Analysis

Neural Processing Unit. The NPU core is implemented using Verilog HDL and deployed on a Xilinx Zynq+ ZCU106 board operating at a frequency of 200 MHz. The NPU used 1,053 multipliers for processing elements, utilizing 60% of the DSPs on the ZCU106. It used 105 blocks of 36K BRAM 136.0K LUTs, and 198.0K FFs. The NPU achieves a latency of 18 ms for a YOLOv3-Tiny model, resulting in a throughput of 362.01 GOPS at 200 MHz. It consumes 2.317 W, achieving a power efficiency of 156.24 GOPS/W.

B. Quantitative Evaluation Results

1) System-Level Energy Reduction: The power consumption of each component in the DVS-CIS NPU system is shown in Table II. The overall power usage can be reduced by limiting NPU operation when the DVS identifies no ROI.

TABLE II: Power Consumption of System Components

	$\mathbf{P}(\pi, \mathbf{X}_{\text{CIS}})$	$\mathbf{P}(\pi)$	$\mathbf{P}(\mathbf{X}_{\text{CIS}})$	$\mathbf{P}(\mathbf{X}_{\text{DVS}})$
Power (W)	9.218	0.939	5.702	0.689

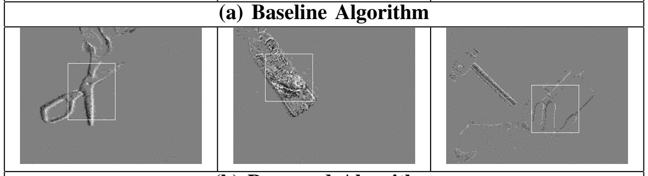
Figure 3 shows our prototype system in operation, capturing a street view in the Philippines. Over a 24-hour period, we recorded variations in the NPU runtime and power consumption, and compared these measurements to a baseline system without DVS, in which the NPU remains continuously active.

As illustrated in Figure 3, the NPU operates in the inferring state for 46% of the time. Based on Eqs. 1 and 2, and using the values from Table II, the power consumption for the baseline and our proposed system are $\mathbf{E}(\pi, \mathbf{X}_{\text{CIS}}) = 14.92\,\text{W}$ and $\mathbf{E}(\pi, \mathbf{X}_{\text{CIS,DVS}}) = 11.13\,\text{W}$, respectively which corresponds to an average energy reduction of 25.4%.

C. Visual Evaluation Results

The VIRAT dataset [16] was used for evaluation.

Scissors	Bottle	Ring Stand
(a) Baseline Algorithm		



(b) Proposed Algorithm		

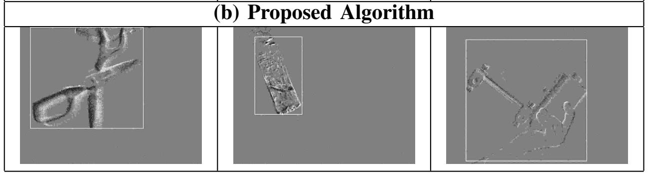


TABLE III: ROI Detection Comparison.

(a)CIS only	(b)DVS only	(c)DVS-CIS Combined
No detection + Refined ROI = Object Detected		

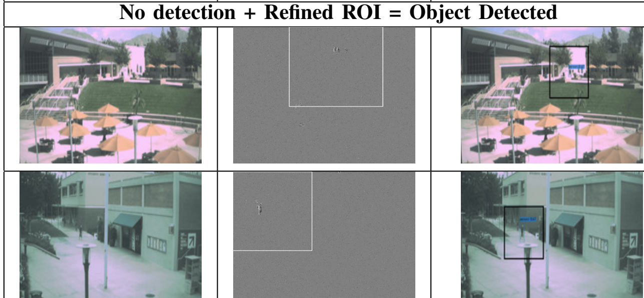


TABLE IV: Small Object (People) Detection Comparison.

CIS only	DVS only	DVS-CIS Combined
No detection + Refined ROI = Object Detected		

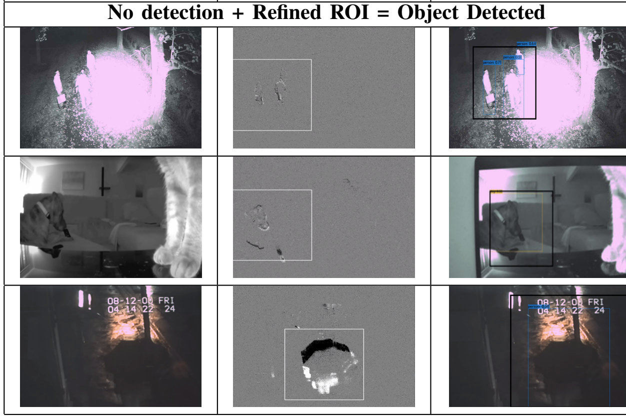


TABLE V: Object Detection in a Low Light Environment.

1) ROI Boundary Refinement for Irregular Objects: Table III presents the experiment results with two ROI algorithms on three types of objects: scissors, a water bottle, and a ring stand. Compared to the baseline, the proposed algorithm identifies dense horizontal ranges of events to effectly cover the targets.

2) Enhanced Detection Accuracy via ROI Scaling: Table IV(a) demonstrates the detection of a small object (a bottle) using only the CIS, where the object is not detected. In contrast, Table IV(b) and (c) illustrates the improved detection capability in the DVS-CIS system, emphasizing the benefits of sensor fusion in enhancing accuracy via ROI scaling.

Table V presents an experiment result under various low-light scenarios. The first row, showing a high-intensity reflection situation, highlights the failure of the CIS-only setup to detect any objects, whereas the DVS-CIS setup not only detects but also refines the object's ROI, resulting in clear object identification. The subsequent rows illustrate that the DVS-CIS combined setup consistently improves object visibility and identification accuracy.

V. Conclusion

We present an energy-efficient surveillance system with DVS-CIS fusion and real-time DNN-based objection detection. Future work will focus on multi-sensor applications such as depth perception and low-latency obstacle detection for autonomous driving.

Acknowledgment

This work was supported in part by the Ministry of Trade, Industry & Energy (MOTIE, Korea) through the Technology Innovation Program (Industrial Strategic Technology Development Program No. 20014490, "Development of Technology for Commercializing Lv.4 Self-driving Computing Platform Based on Centralized Architecture") and in part by the Ministry of SMEs and Startups (No. S3305034). This work was supported in part by the Information Technology Research Center (ITRC) support program (IITP-2024-2020-0-01461) and the Artificial Intelligence Semiconductor Support Program for Nurturing the Best Talents (IITP-2023-RS-2023-00256081).

References

[1] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020.

[2] Tobi Delbrück, Bernabe Linares-Barranco, Eugenio Culurciello, and Christoph Posch. Activity-driven, event-based vision sensors. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 2426–2429, 2010.

[3] Laura Duarte, Mohammad Safeea, and Pedro Neto. Event-based tracking of human hands. *Sensor Review*, 41(4):382–389, September 2021.

[4] Daniel Gehrig and Davide Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, 629:1034–1040, 05 2024.

[5] Leopard Imaging Inc. Li-imx274mipi-fmc datasheet, 2024. Rev. 1.4.

[6] Jicheon Kim, Chunmyung Park, Eunjae Hyun, Xuan Truong Nguyen, and Hyuk-Jae Lee. A Highly-Scalable Deep-Learning Accelerator With a Cost-Effective Chip-to-Chip Adapter and a C2C-Communication-Aware Scheduler. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 14(3):455–468, September 2024.

[7] Jicheon Kim, Chunmyung Park, Eunjae Hyun, Xuan Truong Nguyen, and Hyuk-Jae Lee. A Scalable Multi-Chip YOLO Accelerator With a Lightweight Inter-Chip Adapter. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2024.

[8] Minsik Kim, Kyoungseok Oh, Youngmock Cho, Hojin Seo, Xuan Truong Nguyen, and Hyuk-Jae Lee. A Low-Latency FPGA Accelerator for YOLOv3-Tiny With Flexible Layerwise Mapping and Dataflow. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 71(3):1158–1171, 2024.

[9] Andrew Lavin and Scott Gray. Fast Algorithms for Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4013–4021, 2016.

[10] Hyeongi Lee and Hyoseok Hwang. Ev-ReconNet: Visual Place Recognition Using Event Camera With Spiking Neural Networks. *IEEE Sensors Journal*, 23(17):20390–20399, 2023.

[11] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008.

[12] Kyeongjong Lim, Gyuri Kim, Taehyung Park, Xuan Truong Nguyen, and Hyuk-Jae Lee. An Energy-Efficient YOLO Accelerator Optimizing Filter Switching Activity. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2472–2476, 2022.

[13] Arnav Vaibhav Malawade, Trier Mortlock, and Mohammad Abdullah Al Faruque. EcoFusion: energy-aware adaptive sensor fusion for efficient autonomous vehicle perception. In *Proceedings of the 59th ACM/IEEE Design Automation Conference*, DAC '22, page 481–486, New York, NY, USA, 2022. Association for Computing Machinery.

[14] Javier Mendez, Miguel Molina, Noel Rodriguez, Manuel P Cuellar, and Diego P Morales. Camera-LiDAR multi-level sensor fusion for target detection at the network edge. *Sensors*, 21(12):3992, 2021.

[15] Duy Thanh Nguyen, Hyeonseung Je, Tuan Nghia Nguyen, Soojung Ryu, Kyujoong Lee, and Hyuk-Jae Lee. ShortcutFusion: From Tensorflow to FPGA-Based Accelerator With a Reuse-Aware Memory Allocation for Shortcut Data. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(6):2477–2489, 2022.

[16] Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, J. K. Aggarwal, Hyungtae Lee, Larry Davis, Eran Swears, Xioyang Wang, Qiang Ji, Kishore Reddy, Mubarak Shah, Carl Vondrick, Hamed Pirsiavash, Deva Ramanan, Jenny Yuen, Antonio Torralba, Bi Song, Anesco Fong, Amit Roy-Chowdhury, and Mita Desai. A large-scale benchmark dataset for event recognition in surveillance video. In *CVPR 2011*, pages 3153–3160, 2011.

[17] Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. 2018.

[18] Hyunsurk Eric Ryu. Industrial DVS design; key features and applications. In *Conf. on Computer Vision and Pattern Recognition*, volume 3, 2019.

[19] Bongki Son, Yunjae Suh, Sungho Kim, Heejae Jung, Jun-Seok Kim, Changwoo Shin, Keunju Park, Kyoobin Lee, Jinman Park, Jooyeon Woo, Yohan Roh, Hyunku Lee, Yibing Wang, Ilia Ovsiannikov, and Hyunsurk Ryu. 4.1 A 640×480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 66–67, 2017.

[20] Yunjae Suh, Seungnam Choi, Masamichi Ito, Jeongseok Kim, Youngho Lee, Jongseok Seo, Heejae Jung, Dong-Hee Yeo, Seol Namgung, Jongwoo Bong, Sehoon Yoo, Seung-Hun Shin, Doowon Kwon, Pilkyu Kang, Seokho Kim, Hoonjoo Na, Kihyun Hwang, Changwoo Shin, Jun-Seok Kim, Paul K. J. Park, Joonseok Kim, Hyunsurk Ryu, and Yongin Park. A 1280×960 Dynamic Vision Sensor with a 4.95-m Pixel Pitch and Motion Artifact Minimization. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2020.

[21] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7464–7475, 2023.

[22] Xilinx Inc. *ZCU106 Evaluation Board User Guide (UG1244)*, October 2019.

[23] Junwei Zhao, Shiliang Zhang, Zhaofei Yu, and Tiejun Huang. SpiReco: Fast and Efficient Recognition of High-Speed Moving Objects With Spike Camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7):5856–5867, 2024.