ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ece.uowm.gr

Acceleration and Optimization for Real-Time Multimodal Processing on the Edge

Dimitrios Tsiktsiris

Supervisor: Dr. Minas Dasygenis - Associate Professor

Advisory Committee

Konstantinos Siozios - Associate Professor

Nikolaos Konofaos - Professor

Defence Committee

Nikolaos Asimopoulos - Professor

Nikolaos Ploskas - Associate Professor

Georgios Fragoulis - Professor

Stergios Ganatsios - Professor

Overview

- Introduction
- Problem Definition
- Objectives and Contribution
- Methodology
- Results
- Discussion
- Conclusion

Introduction

Multimodal Edge Processing

Processing data closer to the source (the "edge") using multiple modalities for enhanced accuracy

Advantages

Reduced latency

Improved privacy and security

Increased robustness and accuracy

Challenges

Resource constraints

Dynamic real-world environments

Deployment and management

Research Gaps and Challenges

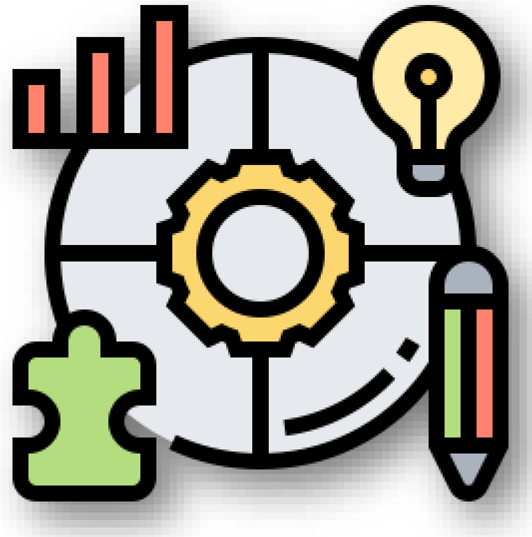
- Edge devices have **constraints** in **processing power** and **memory**, compared to cloud servers.
    - Our solution: Propose **algorithms** and **techniques** specifically **designed for resource-constrained edge devices** and utilizing specialized **hardware accelerators**.
- Real-world conditions are constantly changing, making **single sensor data unreliable**.
    - Our solution**:** Employing sophisticated algorithms to effectively **combine data** from **multiple sensors** and provide a **comprehensive view of the environment**.
- Deployment and Management Complexity
    - Our solution**:** Develop a centralized management platform that can **seamlessly orchestrate** the **deployment**, **configuration**, and **monitoring of edge devices**.

Thesis Objectives and Contributions

This thesis addresses the aforementioned challenges by proposing a new methodology that resulted in an **accelerated multimodal framework** tailored for **edge** computing.

Key objective: Enable the processing of complex data directly at the source leveraging hardware accelerators and optimized AI models and evaluate in real-world applications:

1. Provide **accelerated** computational processing using **hardware designs** for color transformations, edge detection, noise reduction.
2. Introduce **multimodal DL approaches** with **data fusion**, **optimized for edge applications** for abnormal event detection, object detection, proximity assessment, facial identification.
3. Present an extensive **evaluation** on **real-world datasets** and **deployment scenarios** in **public transportation.**



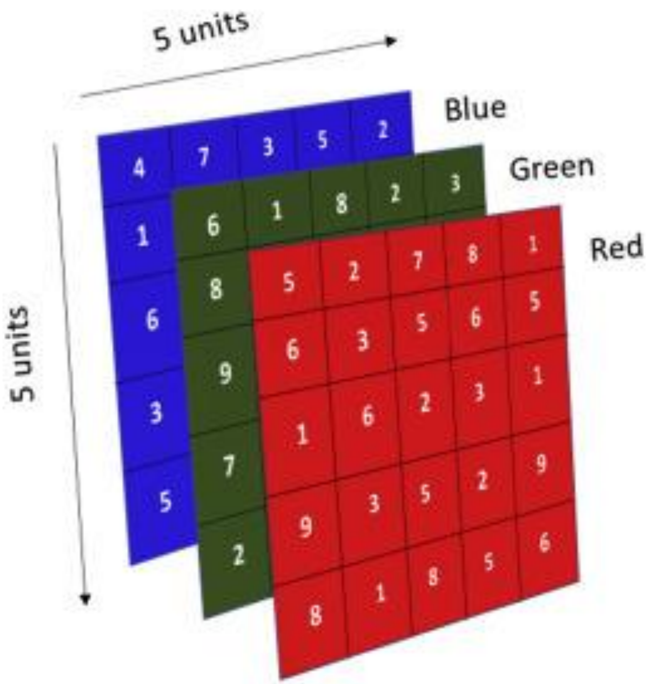
Methodology

ACCELERATION AND OPTIMIZATION FOR REAL-TIME MULTIMODAL PROCESSING ON THE EDGE

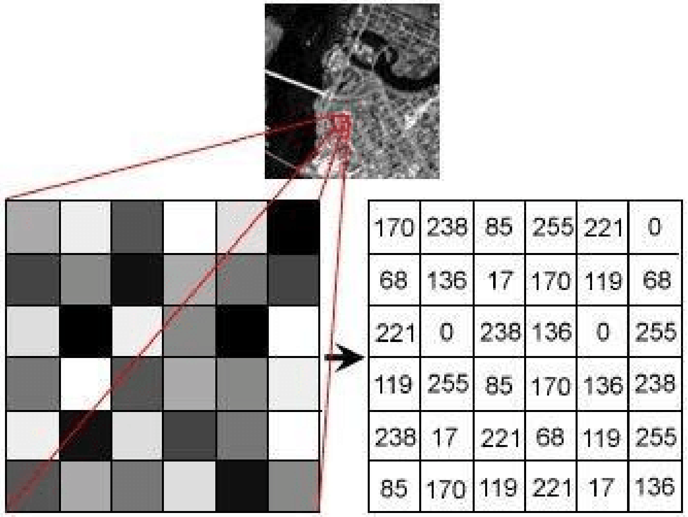
Acceleration using CV Hardware Designs

WHY?

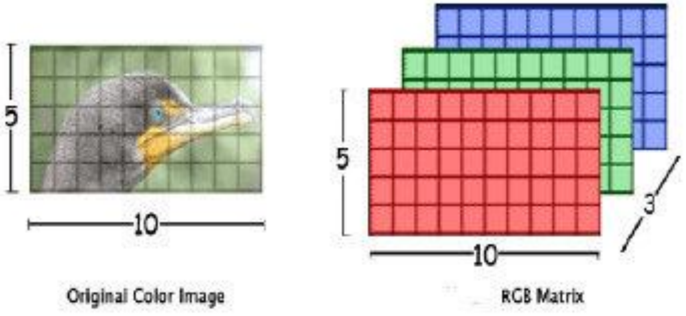
- Digital images are composed of **millions** of tiny dots called **pixels**, each representing a single color.
- Processing each pixel **individually** in a traditional, **sequential processor** is incredibly **time-consuming**, especially for high-resolution images.
- **Parallel processing** enables multiple processing units to work on different parts of the image **simultaneously**, dramatically **accelerating** the overall **processing speed**.



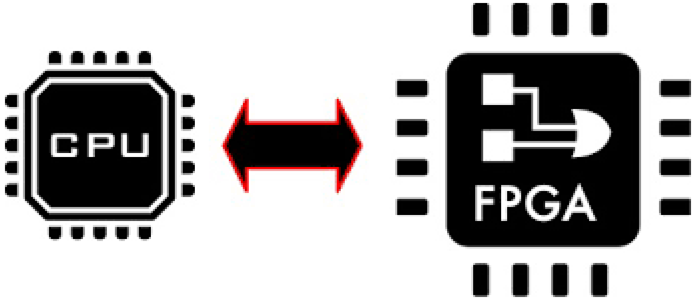
5X5 image matrix



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

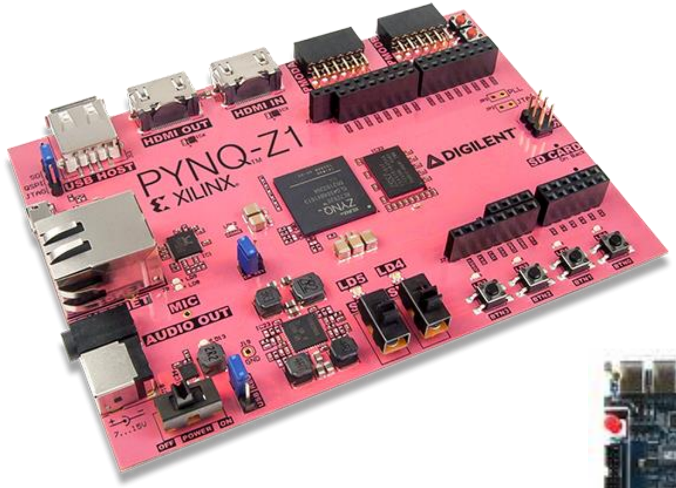


Acceleration using CV Hardware Designs



- Leverage **hardware parallelism** to achieve substantial acceleration in image processing at the edge, bypassing the limitations of sequential processing in traditional **CPU-based** approaches.

- **Design hardware accelerators** to **offload** common **image transformations**:
  - Fast Color Conversion
  - Sobel Edge Detection
  - HLS-Based Edge Detection
  - Noise Reduction using Image Stacking



Sobel Edge Detection

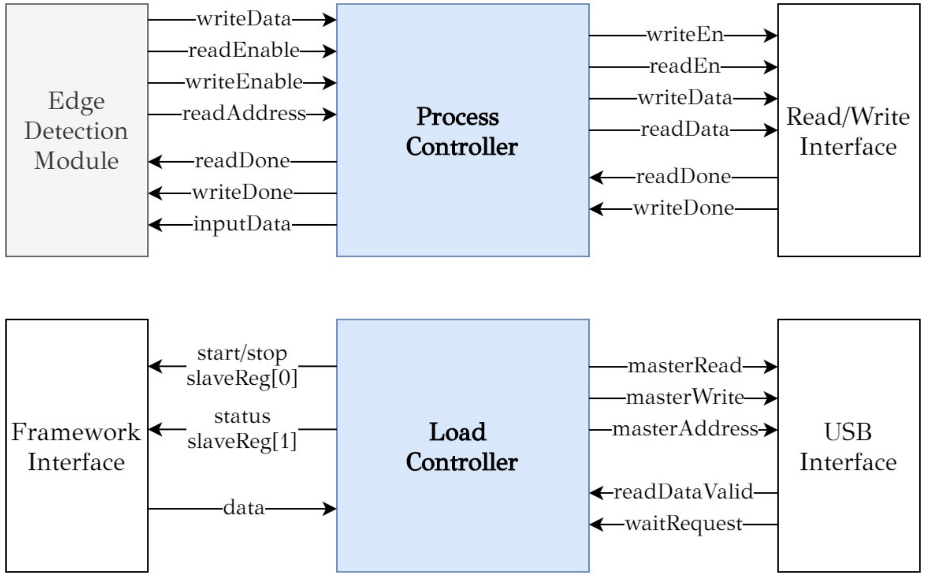
- Calculates the gradient of image intensity at each pixel, indicating the direction and magnitude of the steepest ascent.
- Involves convolving the image with two 3x3 kernels, one for horizontal edges and one for vertical edges.
- Implemented on Altera Cyclone IV EP4CE115 FPGA device.
    - A grayscale module converts color images to grayscale.
    - An edge detection module implements a Sobel Edge Detection algorithm for highlighting edges in grayscale images.
- Achieved 98ms with only 1W average power consumption on 640x480 images.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

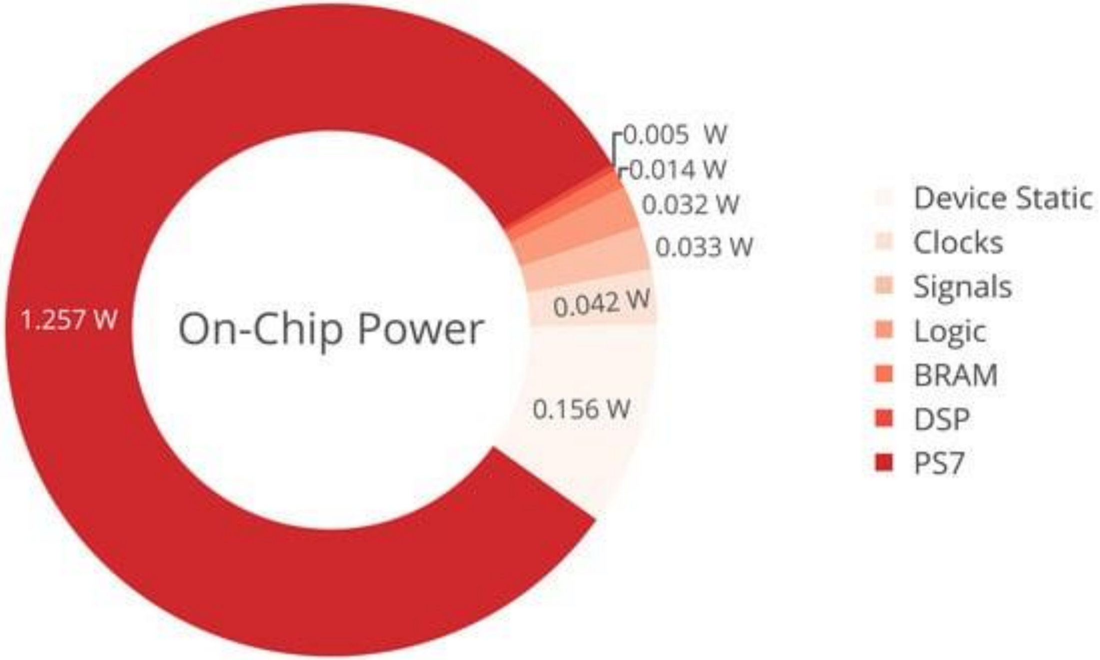
Gy

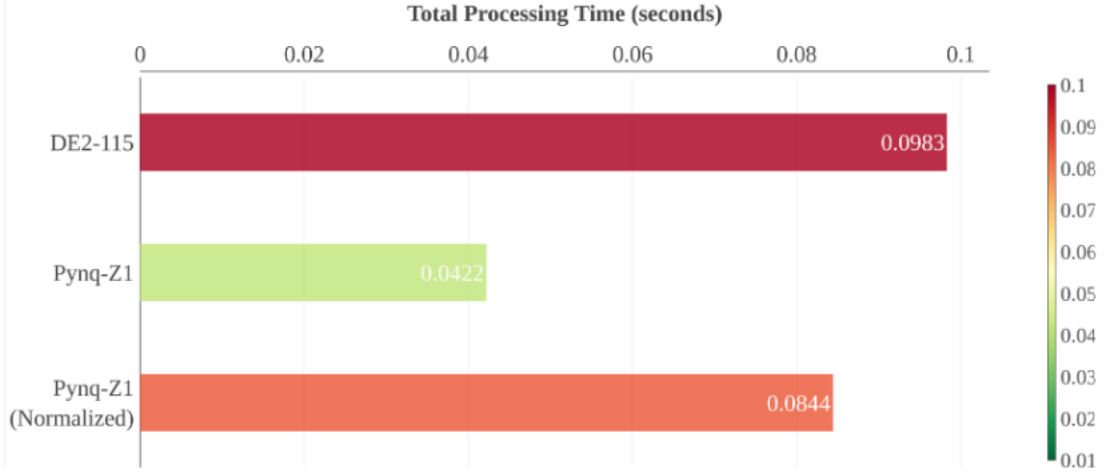


D. Tsiktsiris, *D. Ziouzios, M. Dasygenis, "A Portable Image Processing Accelerator using FPGA", 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), IEEE, 2018.*

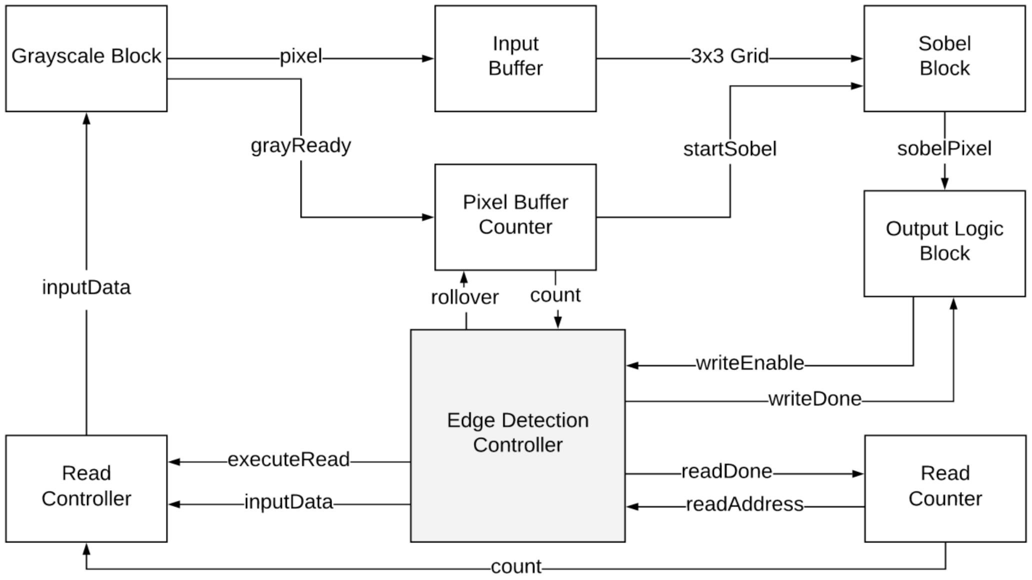
HLS Edge Detection

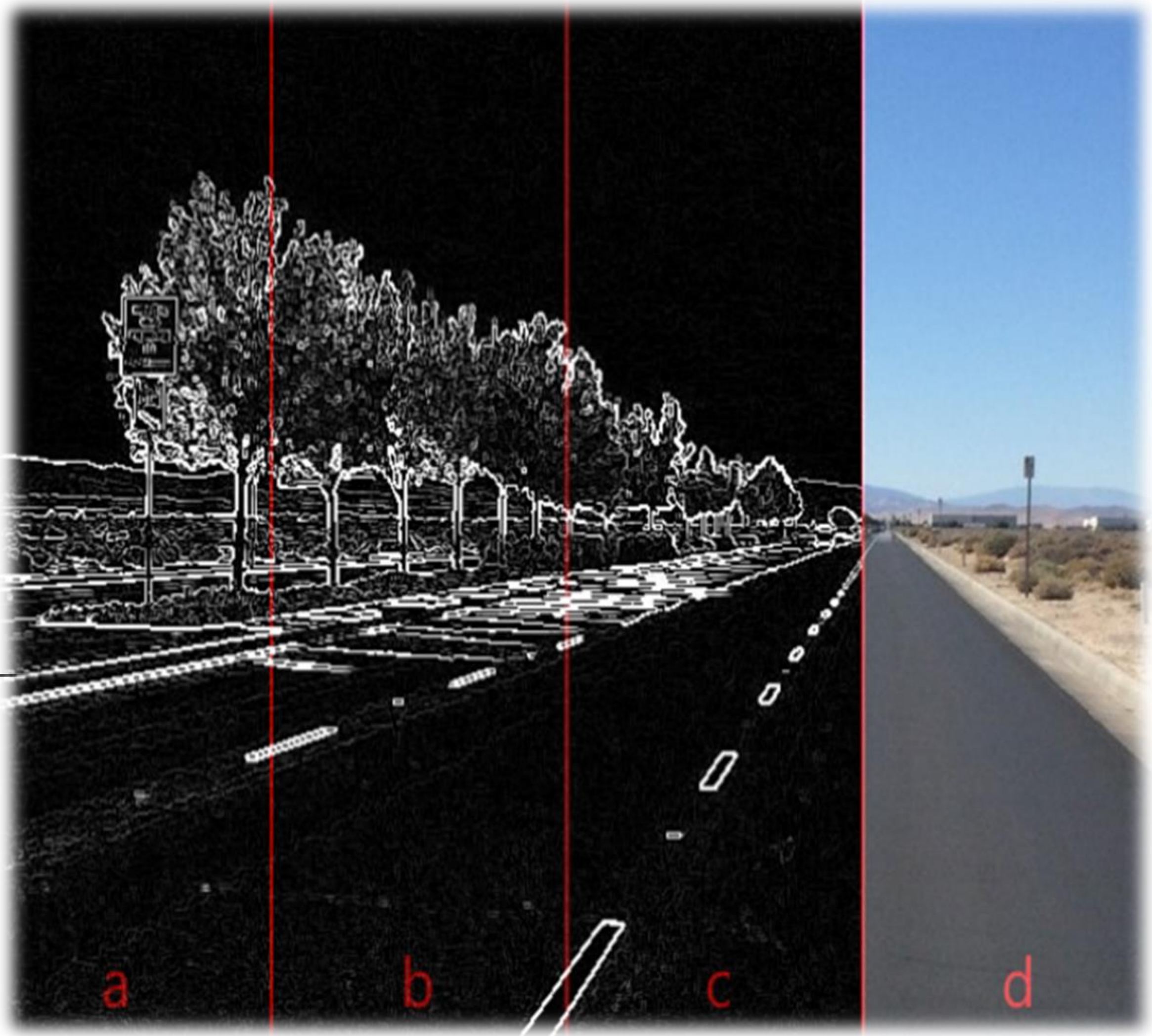
- Implementation with High Level Synthesis using Vivado HLS.
- Sobel Edge Detection on Xilinx PYNQ-Z1
  - Unified grayscale and edge detection modules into a single processing core
  - Int2RGB IP for enhanced image reconstruction performance
  - AXI and DMA engines for component communication and fast data transfer
- Achieved 42ms with 1,5W average power consumption on 640x480 images.
- Compared with bare metal VHDL design the implementation was slightly faster





D. Tsiktsiris, D. Ziouzios, M. Dasygenis. "A high-level synthesis implementation and evaluation of an image processing accelerator", Technologies 7.1 (2018): 4.





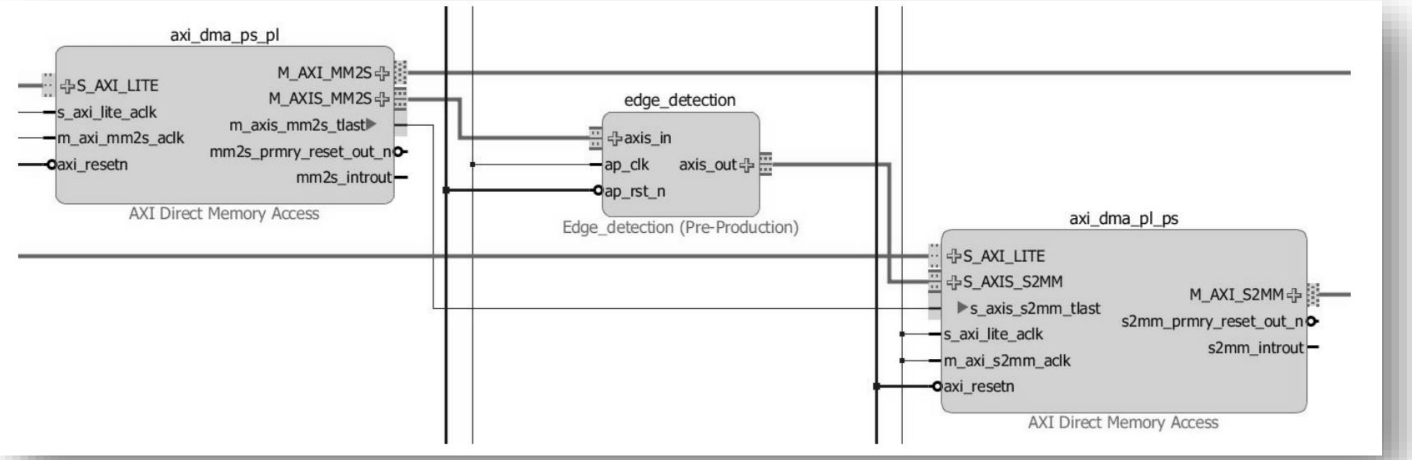
Architecture of the edge detection FPGA module

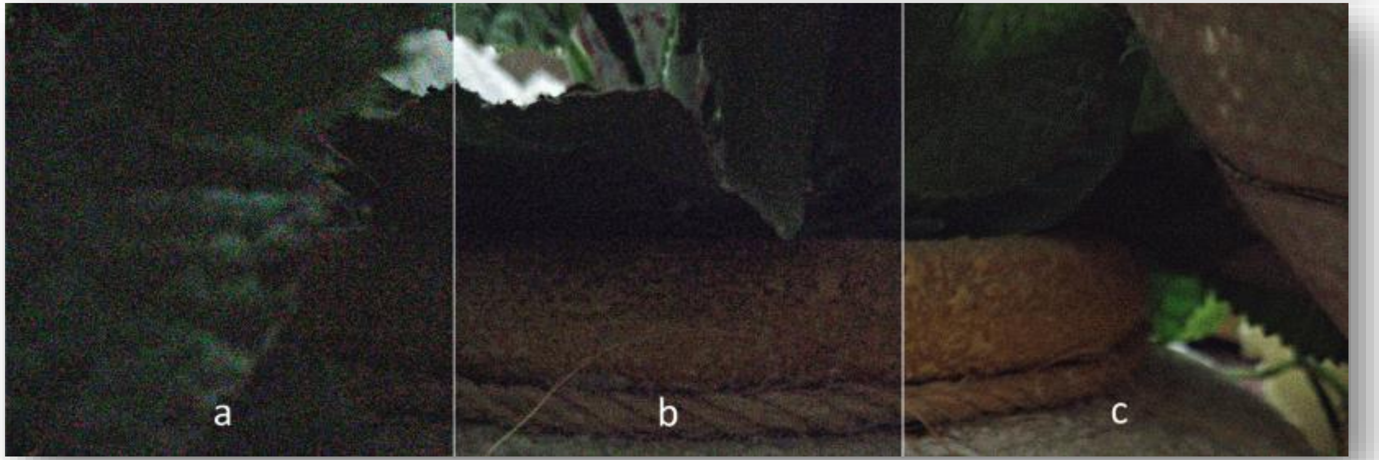
Functional diagram for the FPGA module

D. Tsiktsiris*, D. Ziouzios, M. Dasygenis. “A high-level synthesis implementation and evaluation of an image processing accelerator”, Technologies 7.1 (2018): 4.*

Noise Reduction using Image Stacking

- Implemented on Xilinx Pynq-Z1 board using HLS
  - Buffer of multiple frames
  - Tile matching with pattern-based fuzzy heuristics
  - Tile merging across the buffer using color averaging to reduce noise and enhance dynamic range
- 200ms per 15 frames of 4000×3000 pixel resolution





D. Tsiktsiris, *D. Ziouzios, M. Dasygenis. "HLS Accelerated Noise Reduction Approach Using Image Stacking on Xilinx PYNQ", 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST), IEEE, 2019.*

Accelerated AI Approaches

Part of this research was funded by H2020 AVENUE (Horizon H2020-ART-07-2017)

ACTION RECOGNITION

- Multimodal stream classification
- Pose classification
- Spatiotemporal autoencoder
- Hybrid LSTM classification
- Overhead abnormal event detection

OBJECT DETECTION

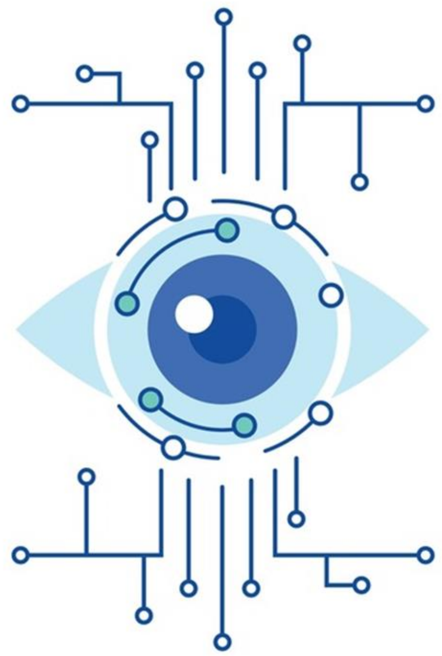
- Overhead Fisheye Object Detection

AUDIO EVENT DETECTION

- Single Event Detection

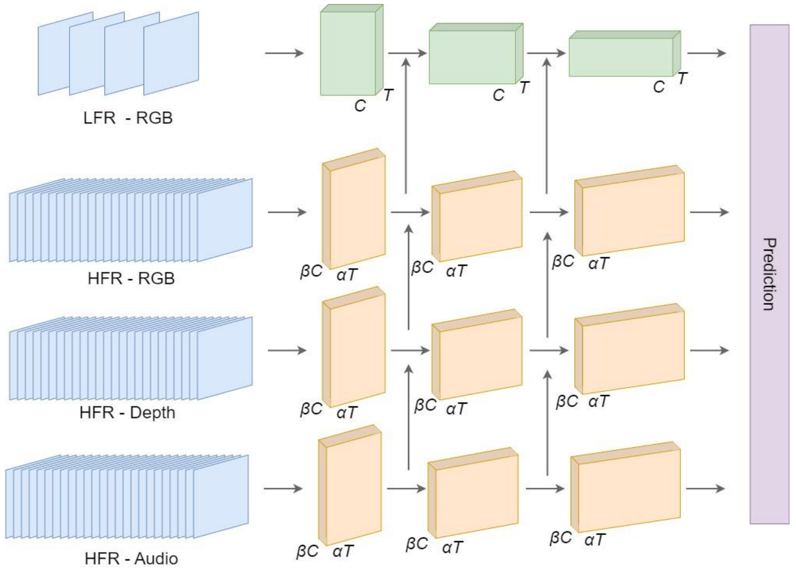
RE-IDENTIFICATION

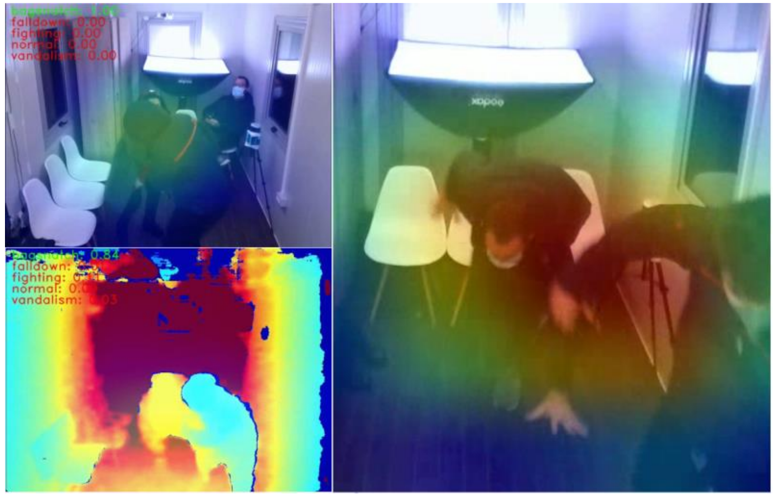
- Siamese facial verification

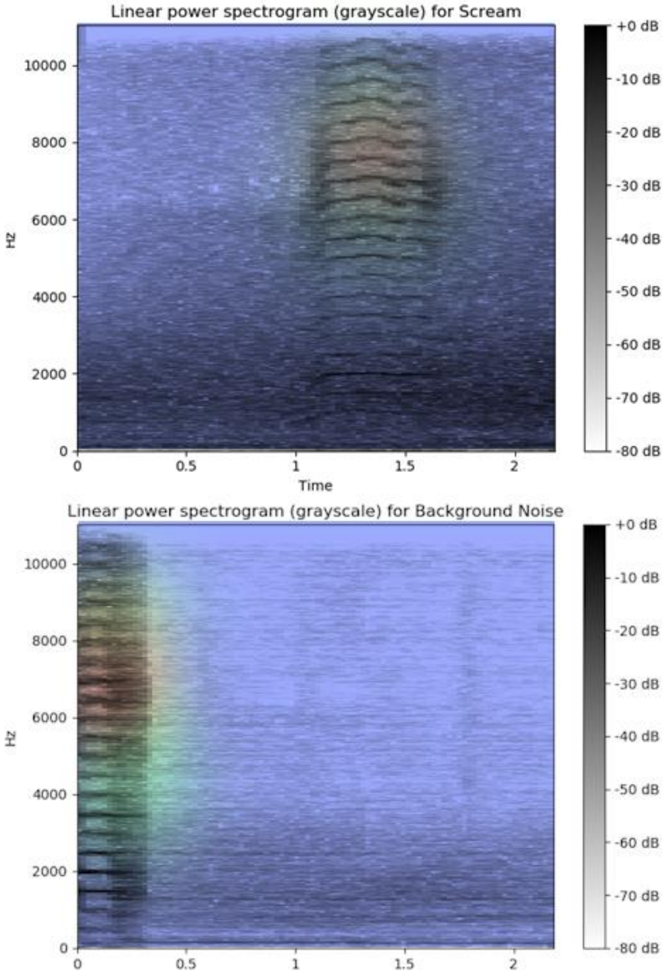


Multimodal Stream Classification

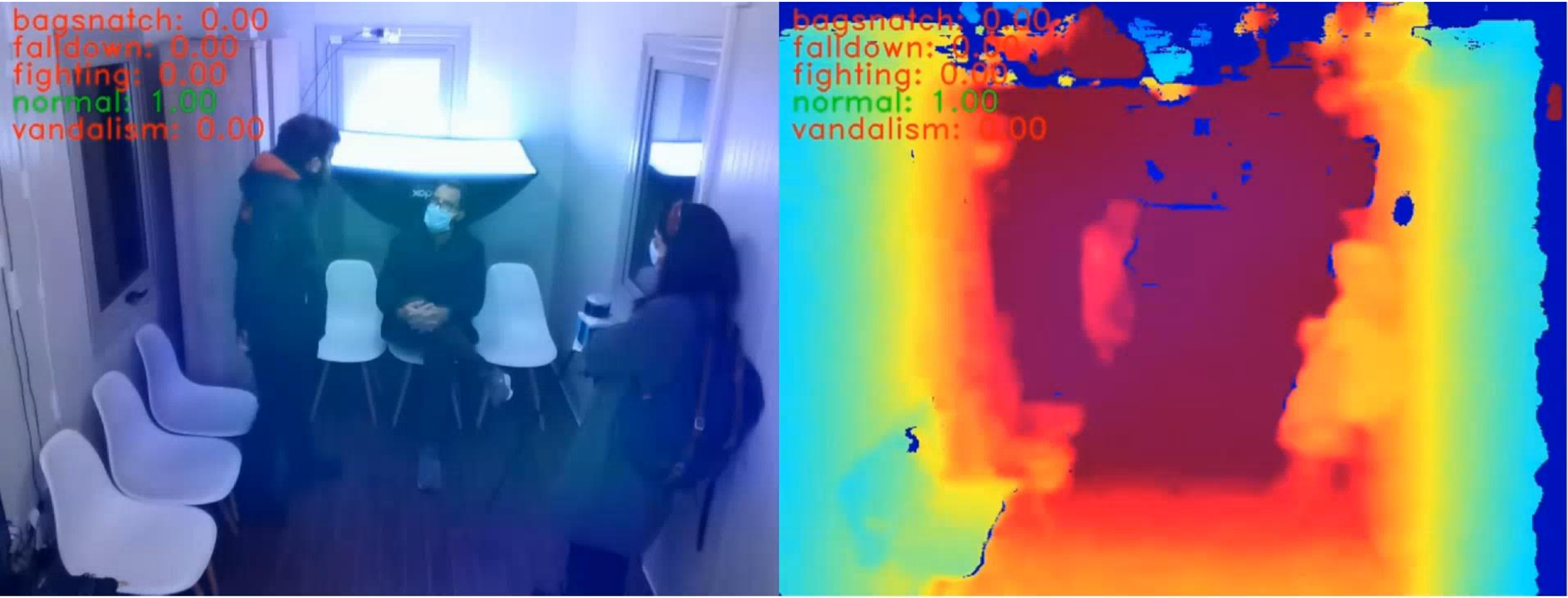
- Abnormal Event Detection using RGB, Depth and Audio (Mel-Spectograms)
- Employs four distinct pathways
  - Low-Frame Rate Path (LFR): Preserves spatial semantics from the RGB modaility.
  - High-Frame Rate Paths (HFR): Focuses on temporal features from RGB, Depth and Audio modalities.
- Utilizes multi-level feature fusion using lateral connections
- Optimized for edge deployment through quantization, CUDA acceleration, layer fusion, and asynchronous data loading
- Achieves 85.1% Top-1 accuracy compared to 81.6% by MoviNet-A6 (RGB-Only)
- Drawbacks: Requires a complex training process, data alignment and synchronization





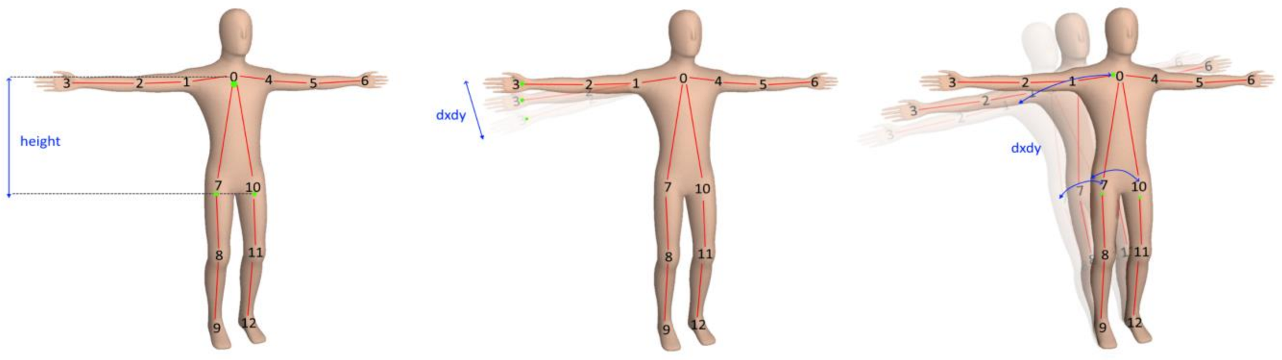


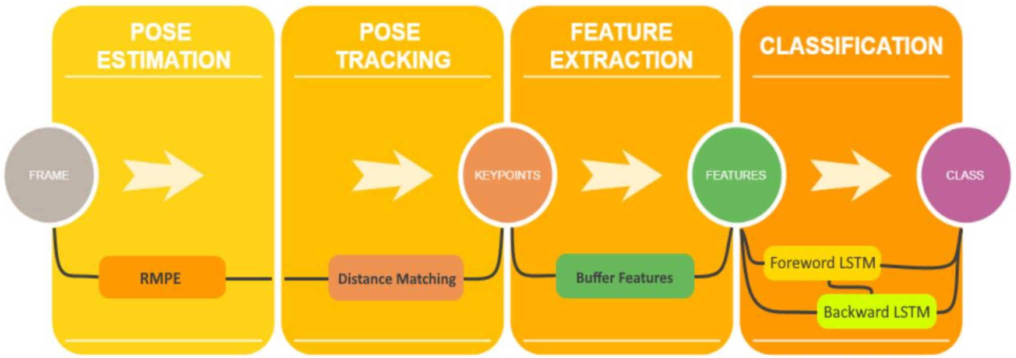
*D. **Tsiktsiris**, A. Lalas, M. Dasygenis, K. Votis, "Multimodal Abnormal Event Detection in Public Transportation", IEEE Access, 2024.*

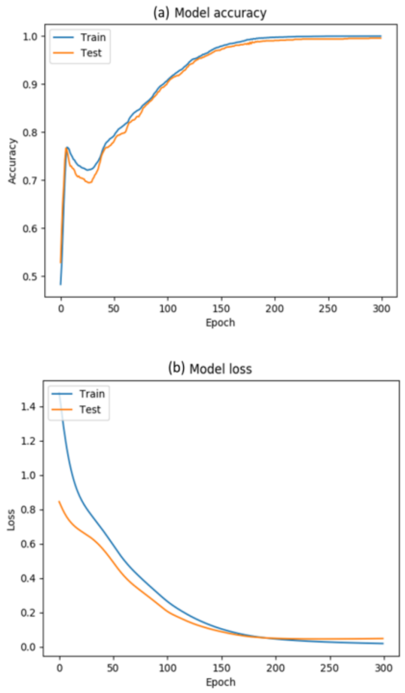

bagsnatch: 0.00
falldown: 0.00
fighting: 0.00
normal: 1.00
vandalism: 0.00
bagsnatch: 0.00
falldown: 0.00
fighting: 0.00
normal: 1.00
vandalism: 0.00

Pose Classification

- Identifies abnormal behavior based on human pose trajectories
- Four-stage pipeline with pose estimation, tracking, feature extraction and classification.
- Utilizes LSTM network for classification based on extracted features from human poses
- Demonstrated 99.6% accuracy in classifying normal and abnormal behavior
- Drawbacks: Occlusion issues, Unbalanced datasets





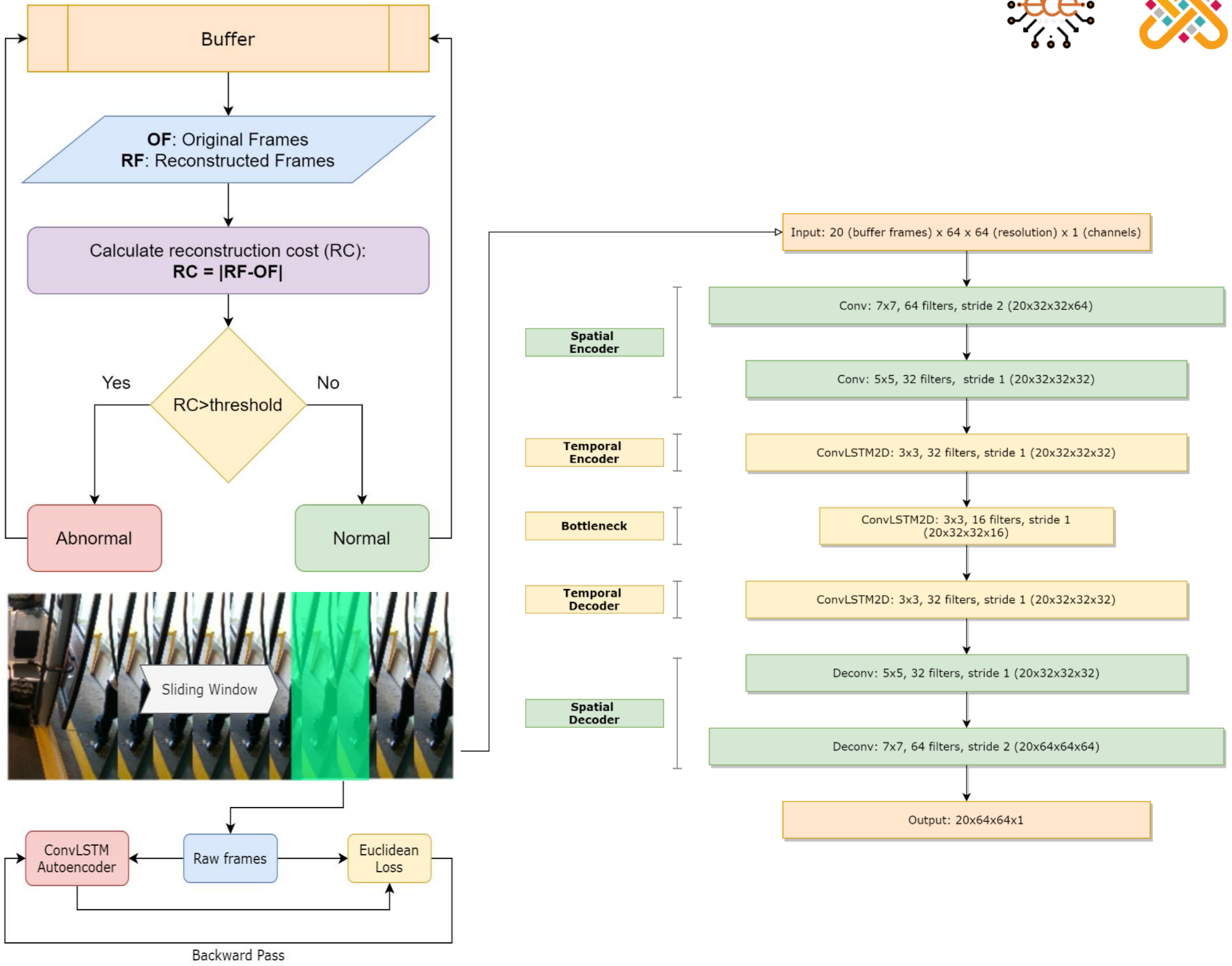


***D. Tsiktsiris**, N. Dimitriou, A. Lalas, M. Dasygenis, K. Votis, "Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures", Sensors 20.17 (2020): 4943.*

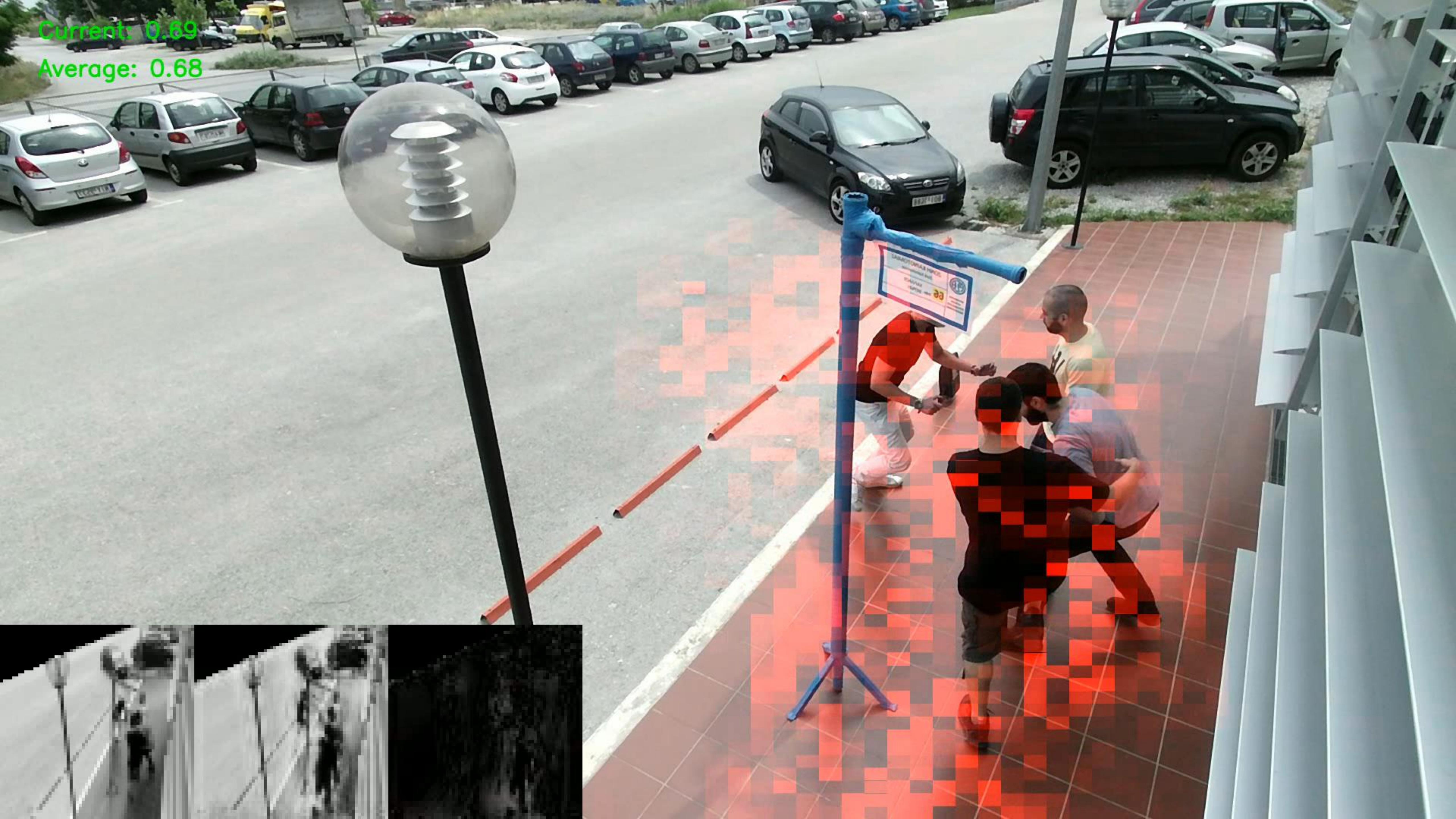

P2 normal
P1 normal
P: 2

Spatiotemporal AutoEncoder

- Trained on normal video volumes to learn regular patterns and detect deviations
- Utilizes spatial convolution and a temporal encoder-decoder (LSTM)
- Bottleneck compression forces the network to learn a more compact and efficient representation of the temporal information.
- The difference between the reconstructed and original frames (reconstruction cost, RC) exceeds a threshold, the video buffer is classified as abnormal.
- Drawbacks: Binary only classification using thresholds in regularity score



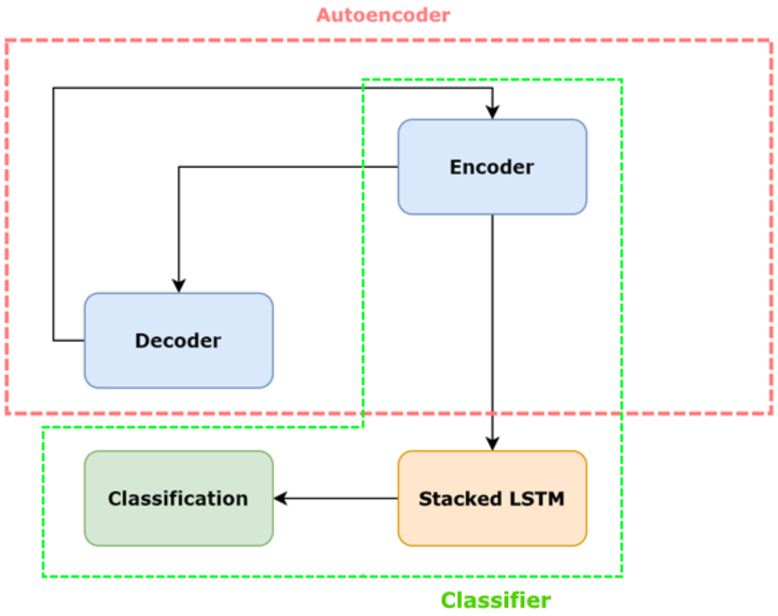
D. Tsiktsiris, *N. Dimitriou, A. Lalas, M. Dasygenis, K. Votis, "Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures", Sensors 20.17 (2020): 4943.*

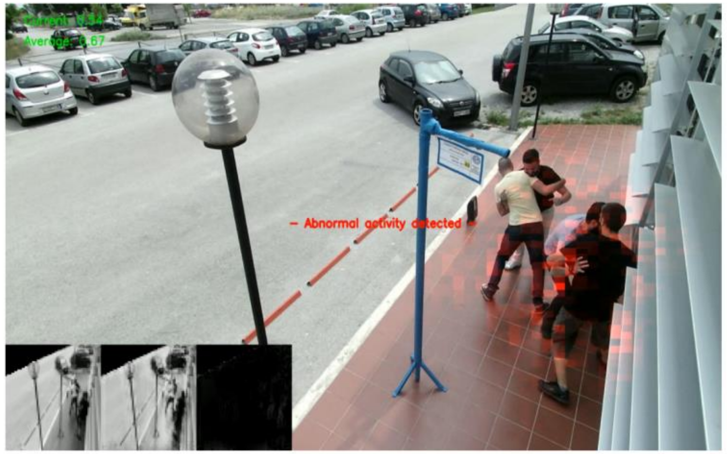

Current: 0.69
Average: 0.68

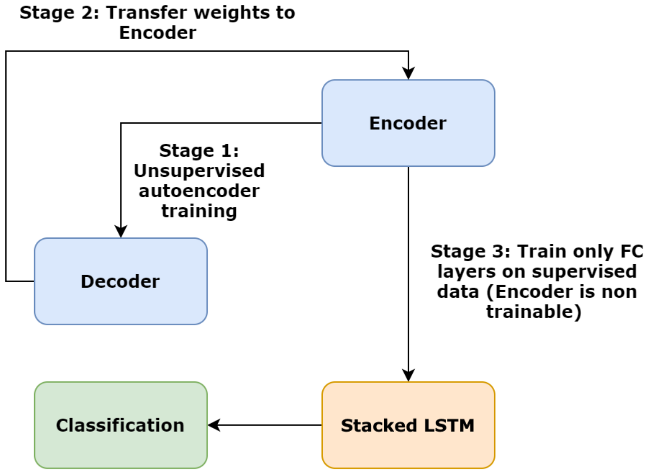

regularity score Sr(t)
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
frame t

Hybrid LSTM Classification

- Extends the previous method by incorporating a semi-supervised approach.
- Utilizes the encoder from the spatiotemporal autoencoder as a high recaller for anomalies
- Combines the encoder with an LSTM classifier for false positive reduction
- Achieves high recall and precision in anomaly detection
- Drawbacks: Complex two-stage training



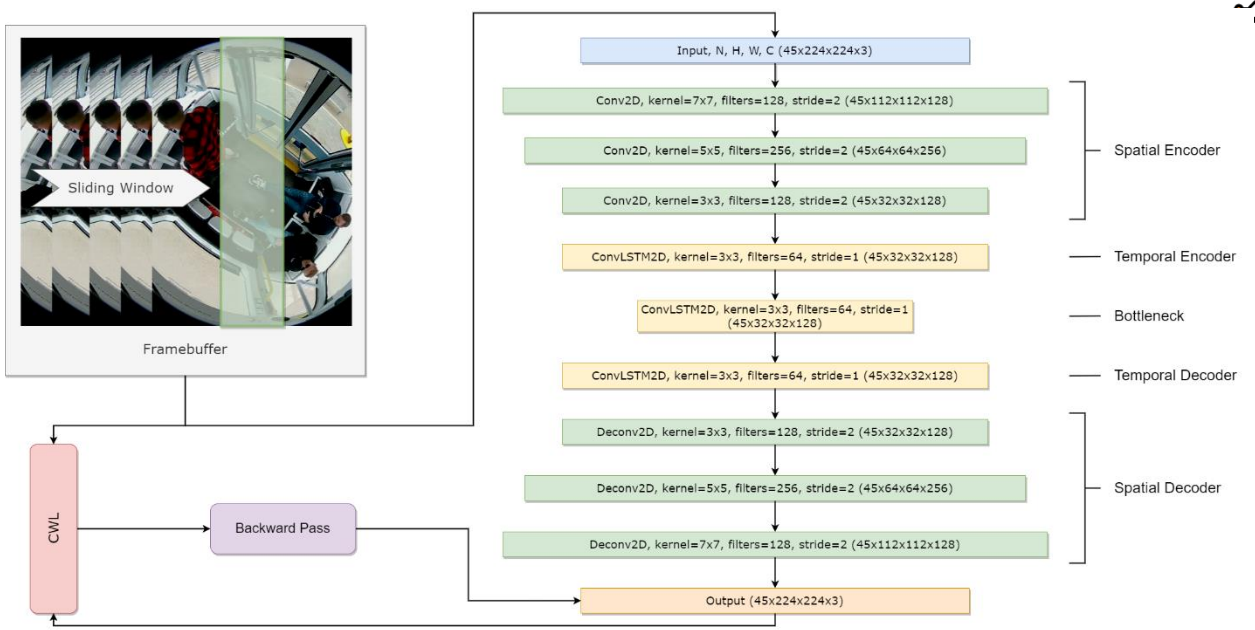


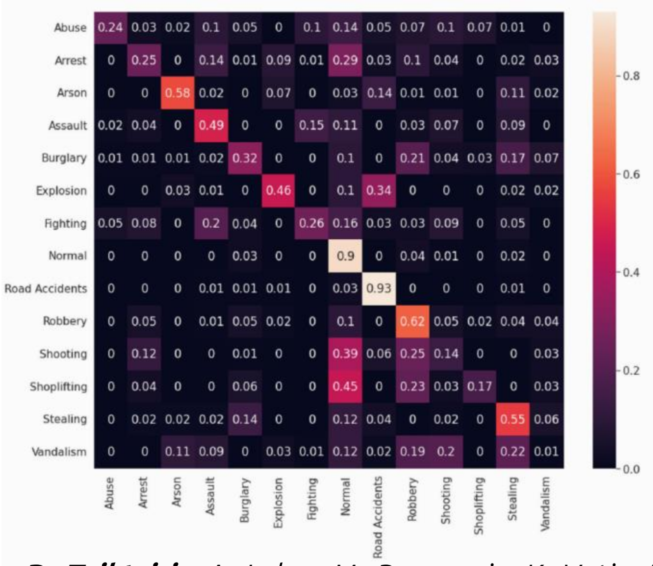


D. Tsiktsiris, *N. Dimitriou, A. Lalas, M. Dasygenis, K. Votis, "Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures", Sensors 20.17 (2020): 4943.*

Overhead Abnormal Event Detection

- Utilizes CAE with fisheye overhead perspective for occlusion free anomaly detection
- Designed to capture important features while removing noise and redundancy
- Employs a Center Weighted and Scale Invariant Loss function for accurate reconstruction of the central image area
- Trained in two phases:
  - Unsupervised learning for capturing regular patterns
  - Supervised learning for anomaly classification
- Achieved 88% Top-1 accuracy vs 86% Chen et al. [18]

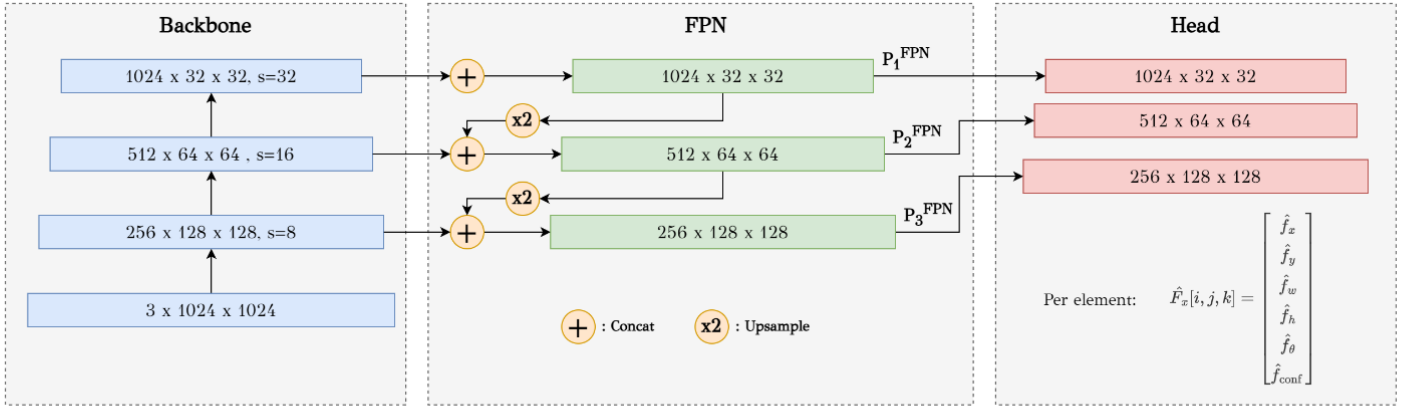


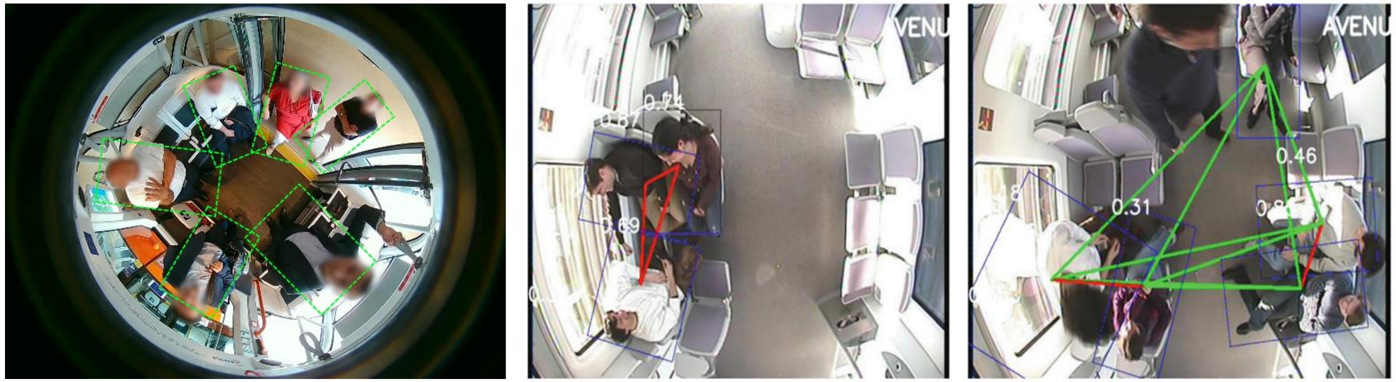


***D. Tsiktsiris**, A. Lalas, M. Dasygenis, K. Votis, "Enhancing the Safety of Autonomous Vehicles: Semi-Supervised Anomaly Detection With Overhead Fisheye Perspective", IEEE Access, 2024.*

Overhead Object Detection

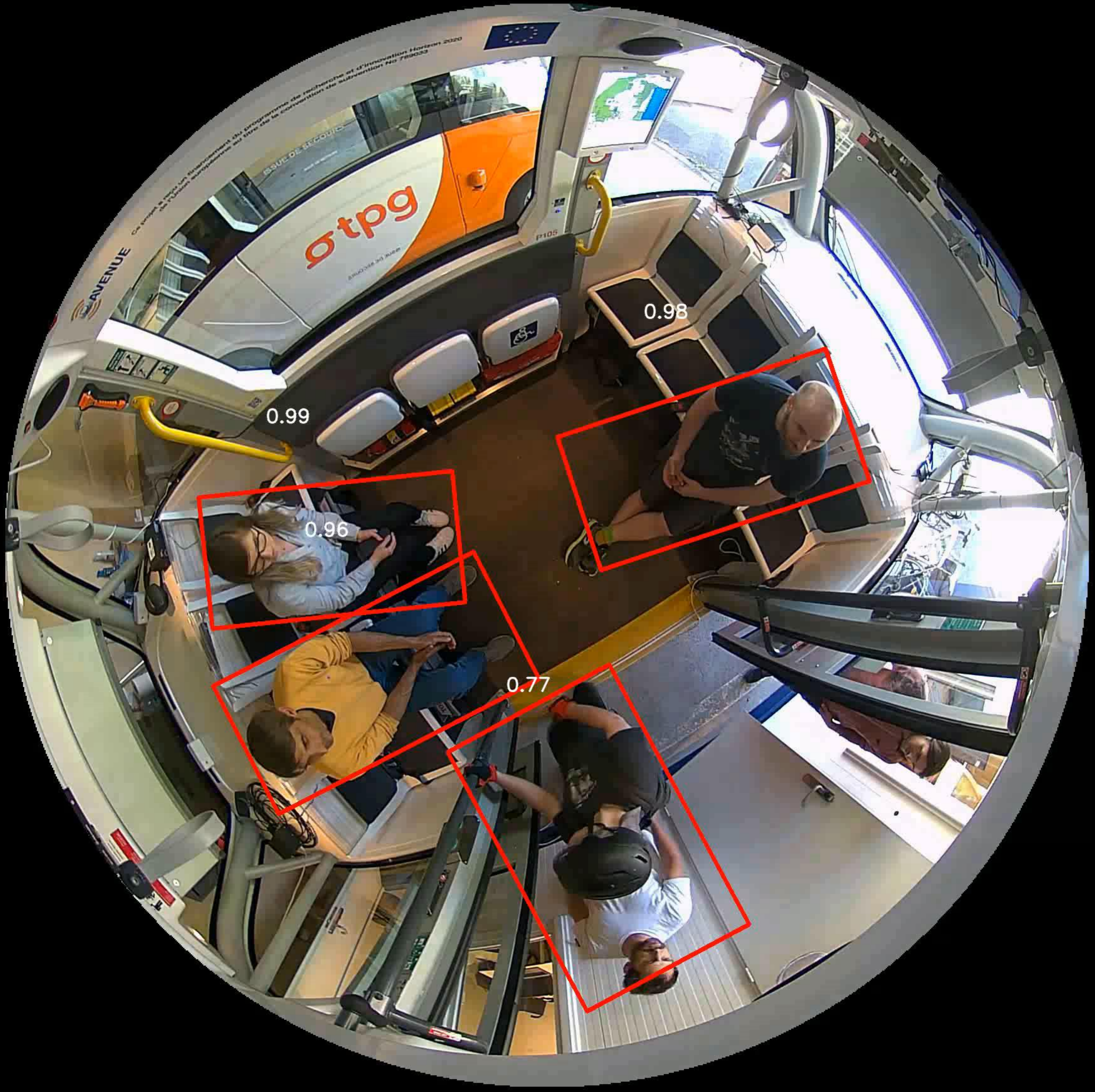
- Object detection in overhead fisheye images
- Employs a three-stage architecture:
  - Backbone network (feature extractor)
  - Feature pyramid network (FPN)
  - Bounding box regression network
- Utilizes rotation-aware bounding box regression for accurate object localization and orientation in fisheye images
- Achieves 92.3% accuracy in overhead passenger detection vs 87% by [32]





D. Tsiktsiris, A. Lalas, M. Dasygenis, K. Votis, "Improving Passenger Detection With Overhead Fisheye Imaging", IEEE Access, 2024.

D. Tsiktsiris, A. Lalas, M. Dasygenis, K. Votis, D. Tzovaras, "An efficient method for addressing covid-19 proximity related issues in autonomous shuttles public transportation". Cham: Springer International Publishing, 2022.

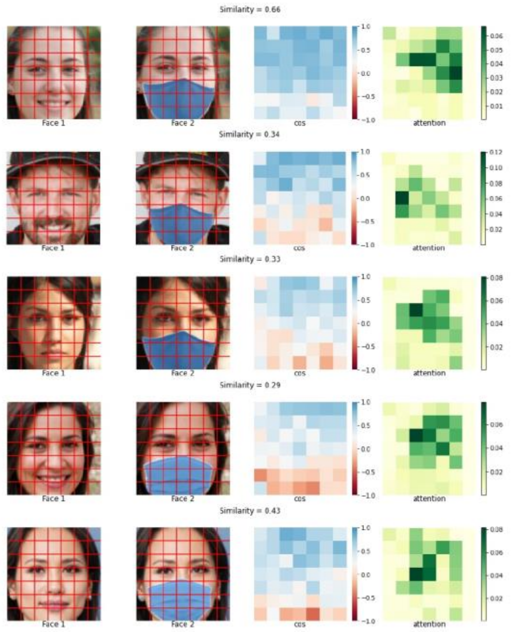


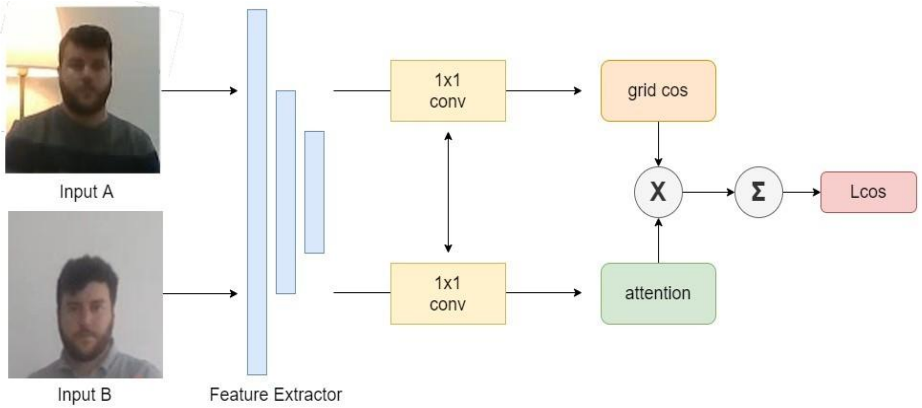
tpg
AVENUE
0.99
0.98
0.96
0.77

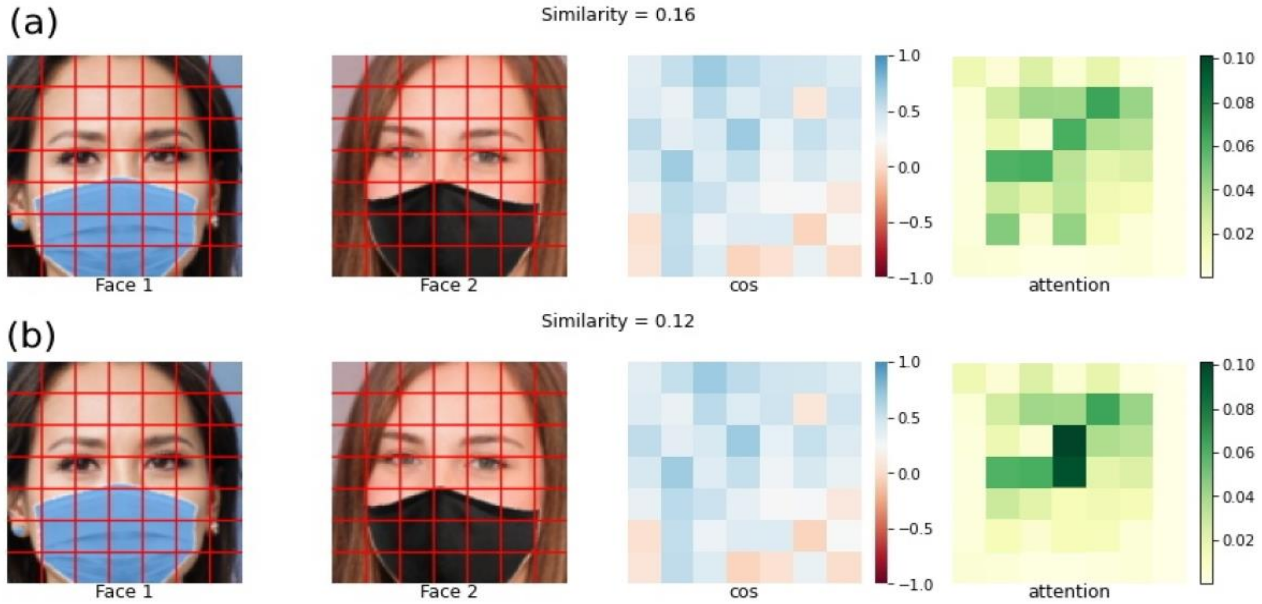

In-vehicle security
AVENUE
Enhance the sense of security and trust
Smart Feedback
Video Modality Status
No abnormal events are detected
Event History
Found Christian Zinckernagel in this vehicle!
Found Sophie Green in this vehicle!
Found Sophie Green in this vehicle!
Found Sophie Green in this vehicle!
Found Sophie Green in this vehicle!
2.52
0.72
0.85
0.62
Temperature
26.16°C
Smoking
No
CO2
679 ppm
PM25
7 ppm
holo-vehicle-1, Slagelse, Denmark

Facial Identification

- Employs one-shot learning using a Siamese neural network for face verification.
  - Network learn to minimize distance for positive pairs and maximize distance for negative pairs.
  - Given two new face images, compute their embeddings (representations) in the learned space.
- Augmented existing datasets to include face masks for improved performance in real-world scenarios
- Achieved high accuracy of 99.35% in identifying individuals and 74.52% accuracy on datasets with face masks



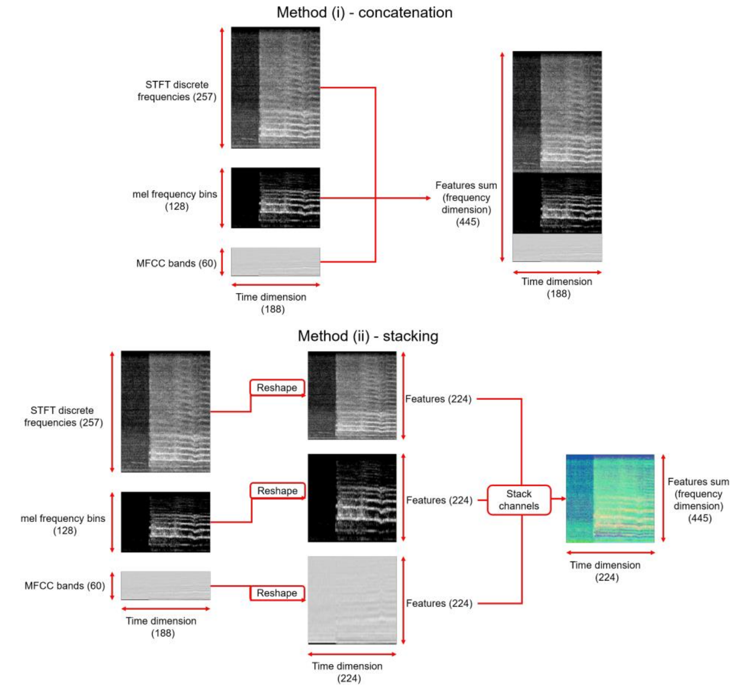


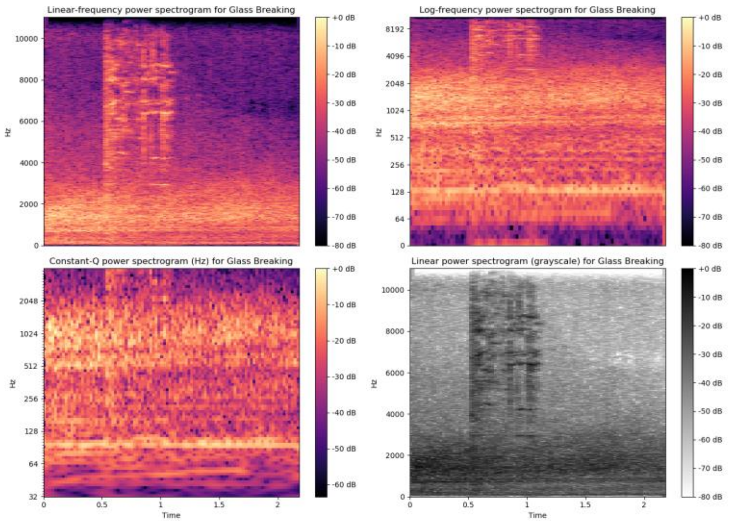


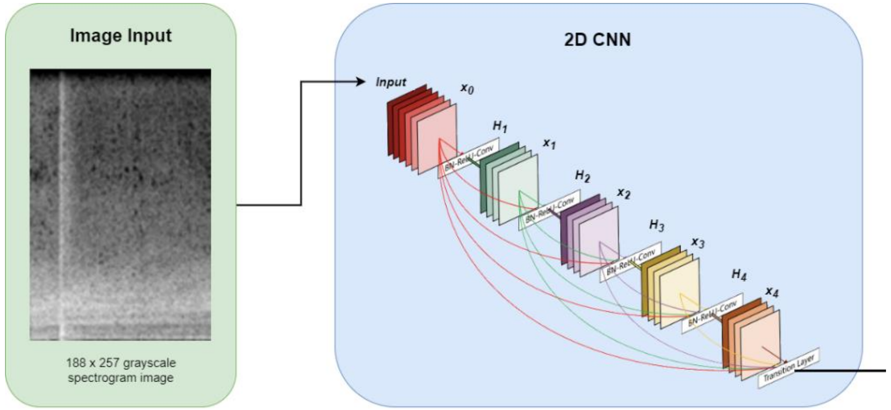
***D. Tsiktsiris**, A. Lalas, M. Dasygenis, K. Votis, D. Tzovaras, "Enhanced Security Framework for Enabling Facial Recognition in Autonomous Shuttles Public Transportation During COVID-19". Cham: Springer International Publishing, 2021.*

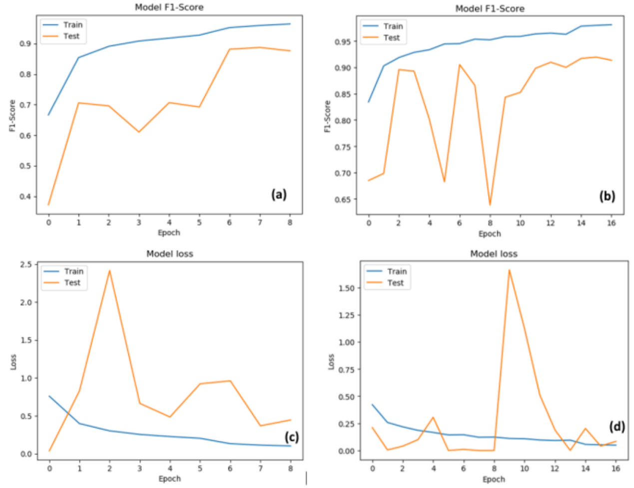
Audio Classification

- Utilizes a DenseNet-121 CNN
- Investigated performance under various SNR settings
- Demonstrated high accuracy in classifying different audio events (gunshot, glass breaking, scream)
- Explored the generalizability of the network across different SNR environments









D. Tsiktsiris, *Vafeiadis, A., Lalas, A., Dasygenis, M., Votis, K., & Tzovaras, D. (2022). A novel image and audio-based artificial intelligence service for security applications in autonomous vehicles. Transportation Research Procedia, 62, 294-301.*

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

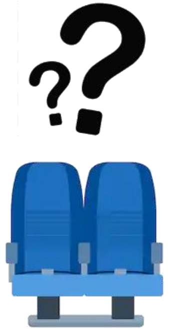
Enhancing Safety in Public Transportation

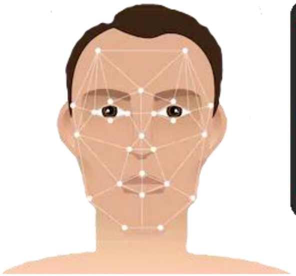
IN-CABIN MONITORING SERVICES FOR AUTONOMOUS VEHICLES AND INFRASTRUCTURE

Part of this research was funded by H2020 AVENUE (Horizon H2020-ART-07-2017)

Problem Definition

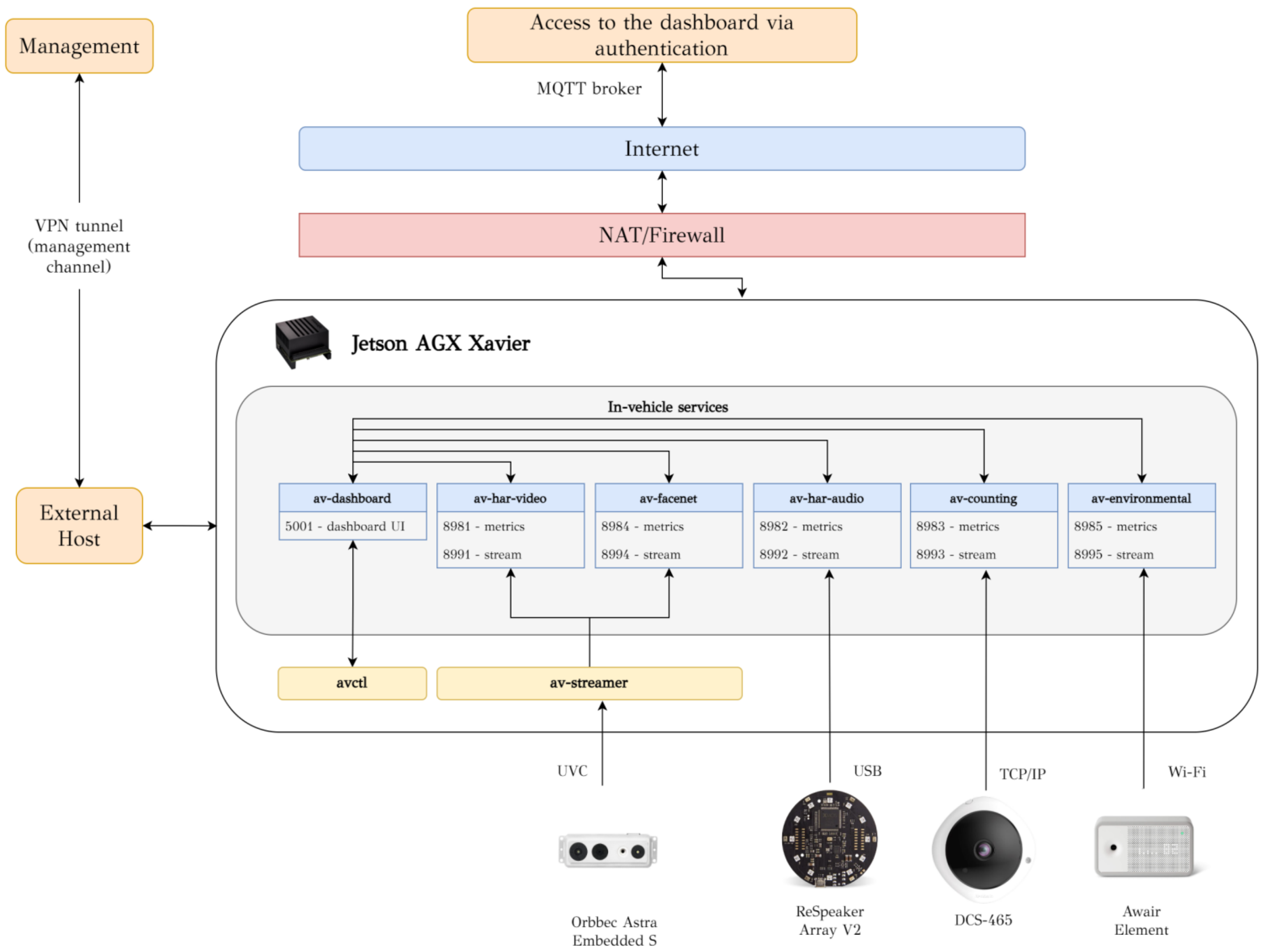
- Autonomous vehicles offer convenience but raise concerns about **passenger safety and security in the absence of a human driver**:
  - Passenger are concerned about their **safety** inside the vehicle in response to internal threats or emergencies
- **Objective:** Boost passenger confidence, enabling independent use of public transportation
  - Detect and notify about **abnormal events** (aggression, abnormal incidents, vandalism, petty crimes)
  - Provide accurate **passenger counting** for resource optimization and safety
  - Implement **facial recognition** for personalized monitoring and security





In-Cabin Monitoring System

- Modular design with specialized sub-modules for real-time:
  - Audio and video data processing
  - Passenger counting, proximity assessment and facial recognition
  - Behavior analysis
- Utilizes a common foundation of hardware and sensors
- Efficient operation under 20 Watts in total using the Jetson AGX Xavier
- 5G connectivity for metadata transmission and management through VPN

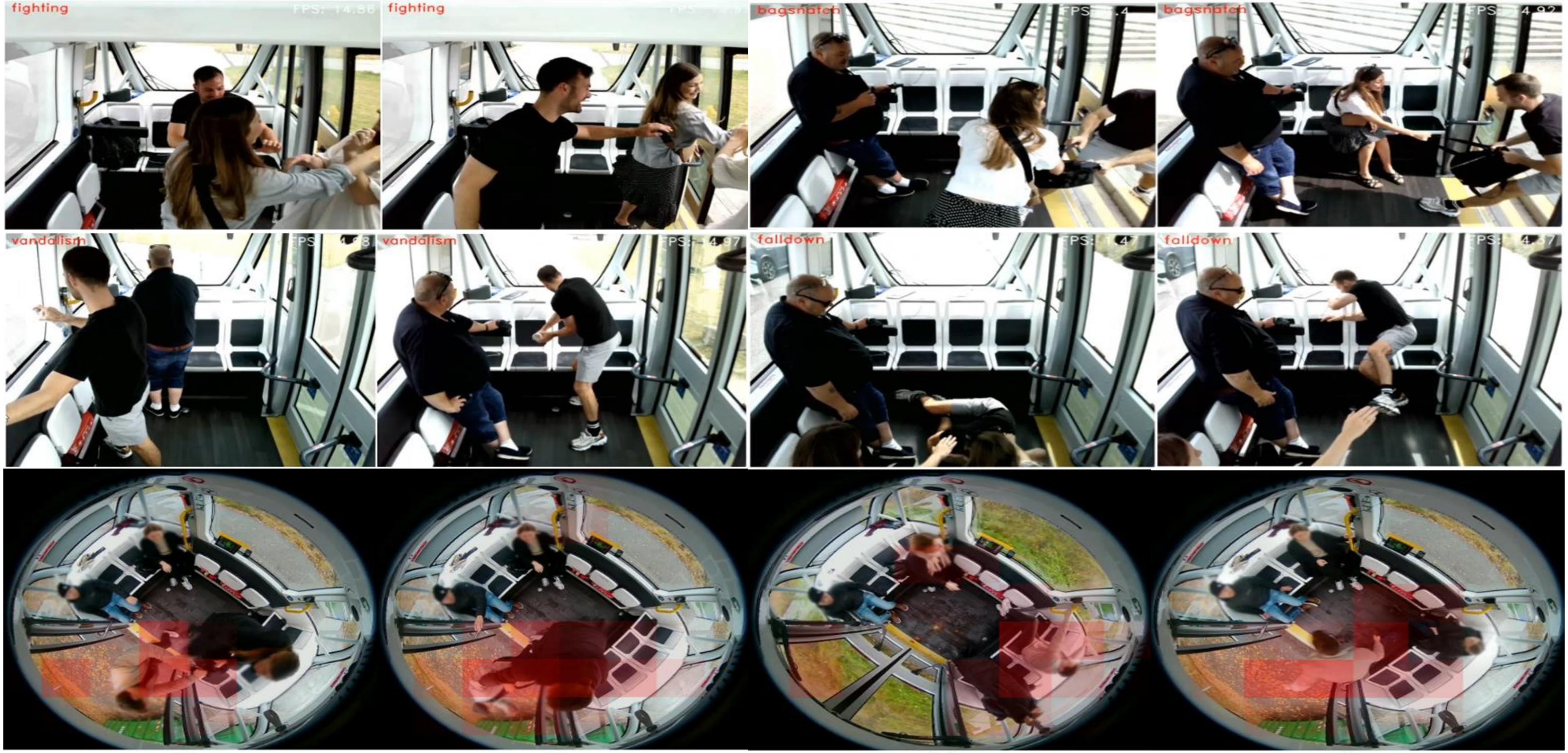


Real-world Installation

Real-world installation on NAVYA and Ruter autonomous minibuses in the context of the European H2020 projects AVENUE, SHOW, ULTIMO, AUTOTRUST

(a) Inverter
(b) Installation area – panel
(c) Overhead fisheye camera (D-Link DCS 4625)
(d) Location of the NVIDIA Jetson AGX Xavier embedded system
(e) Monitoring screen for passengers
(f) Power-supply cables hidden behind the vehicle's panel

Real-world Evaluation of Abnormal Events Detection



(a) (b) (c) (d)

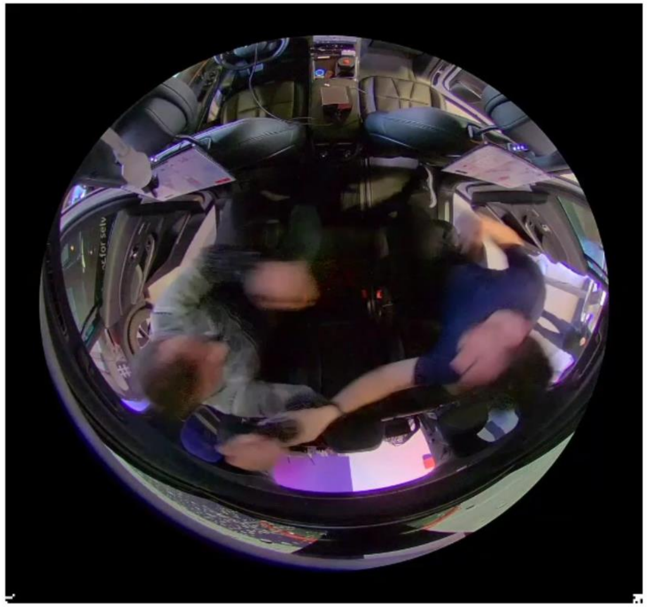
Dashboard

OBU Maintenance Dashboard

ULTIMO

Augmented Security & Trust

Passenger Counting



Video

Detected fighting (54%)

bagsnatch	14.80
fighting	54.17
normal	15.49
falldown	3.44
vandalism	12.09

Audio

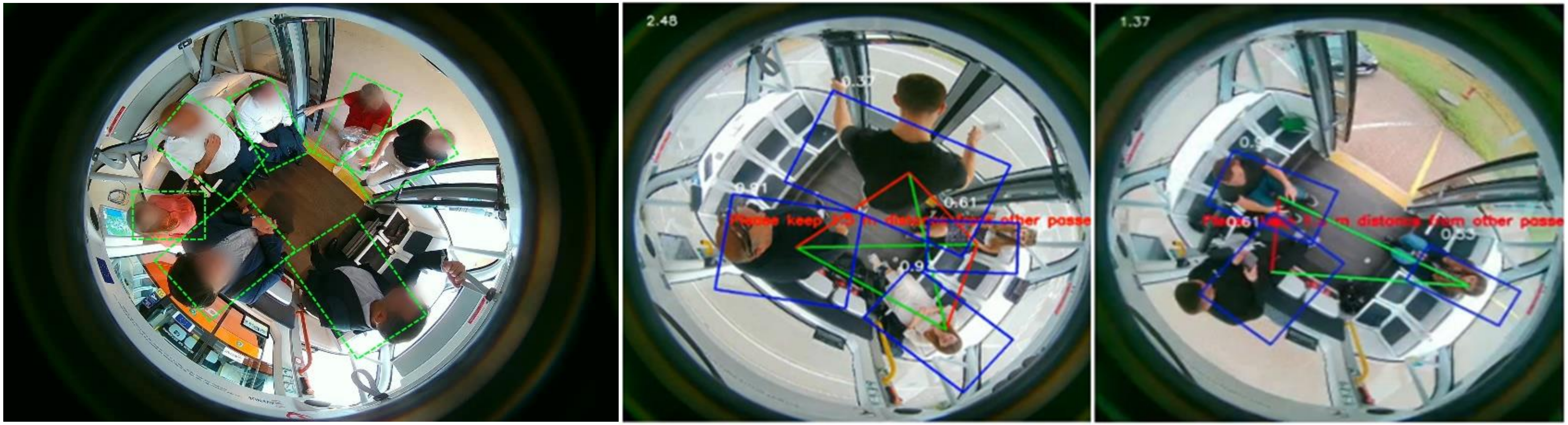
No abnormal events

gunshot	15.05
screaming	15.05
speech	39.81
noise	15.05
glassbreaking	15.05

Detected Passengers: 0

n_pax	0

Real-time Passenger Proximity Detection



All passengers are correctly detected, even the two ones that are partially occluded

Green lines represent a safe distance, while red lines an unsafe one

Real-World Face Identification Evaluation

Identification of the passengers inside the AV

Discussion

- This dissertation has successfully demonstrated the potential of accelerated multimodal AI frameworks for **enhancing edge computing capabilities across various sectors**.
- The integration of multiple data modalities, hardware acceleration techniques, and software optimization strategies has enabled the development of **efficient systems** for **real-time applications**.
- The framework's practical value has been showcased through its successful implementation and **evaluation in real-world use cases** in public transportation.

Future Work

- Explore advanced **privacy-preserving** techniques to address data privacy and security concerns
- Develop more **robust and adaptable** AI models capable of handling greater data diversity and complexity
- Implement **distributed training and inference** mechanisms for enhanced scalability in distributed edge computing environments
- Integrate the framework with emerging technologies, such as **AR/VR**, for novel applications in various domains

Conclusion

- This dissertation has made a substantial contribution to the field of AI and edge computing by presenting an **accelerated multimodal framework** that effectively addresses the challenges of **processing diverse data types in resource-constrained environments**.
- The framework's adaptability, scalability, and real-time processing capabilities introduce new intelligent systems capable of operating **autonomously**, **adapting** to dynamic environments, and ultimately **enhancing human lives**.
- This research serves as a **foundation** for future innovations, inspiring further exploration and development in the rapidly evolving field of **AI and edge computing**.

PhD Publications

Journals

[J1] **D. Tsiktsiris**, A. Lalas, M. Dasygenis and K. Votis, “Multimodal Abnormal Event Detection in Public Transportation”, IEEE Access, 2024 **(Q1)**.

[J2] **D. Tsiktsiris**, A. Lalas, M. Dasygenis and K. Votis, “Enhancing the Safety of Autonomous Vehicles: Semi-Supervised Anomaly Detection with Overhead Fisheye Perspective”, IEEE Access, 2024 **(Q1).**

[J3] **D. Tsiktsiris**, A. Lalas, M. Dasygenis and K. Votis, “Improving Passenger Detection with Overhead Fisheye Imaging”, IEEE Access, 2024 **(Q1)**.

[J4] T. Sanida, **D. Tsiktsiris**, A. Sideris, and M. Dasygenis. “A heterogeneous implementation for plant disease identification using deep learning”. In Multimedia Tools and Applications, vol. 81, p. 15041–15059, 2022 **(Q1)**.

[J5] T. Sanida, A. Sideris, **D. Tsiktsiris**, and M. Dasygenis. “Lightweight neural network for COVID-19 detection from chest X-ray images implemented on an embedded system”. In Technologies, vol. 10, p. 37, 2022 **(Q1)**.

[J6] **D. Tsiktsiris**, N. Dimitriou, A. Lalas, M. Dasygenis, K. Votis, and D. Tzovaras. “Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures”. In Sensors, vol. 20, p. 4943, 2020 **(Q1)**.

[J7] D. Ziouzios, **D. Tsiktsiris**, N. Baras, and M. Dasygenis. “A distributed architecture for smart recycling using machine learning”. In Future Internet, vol. 12, p. 141, 2020 **(Q2)**.

[J8] **D. Tsiktsiris**, D. Ziouzios, and M. Dasygenis. “A High-Level synthesis implementation and evaluation of an image processing accelerator”. In Technologies, vol. 7, p. 4, 2018 **(Q1)**

Conferences

[C1] **D. Tsiktsiris**, N. Dimitriou, Z. Kolias, S. Skourti, P. Girssas, A. Lalas, K. Votis, and D. Tzovaras. “A Framework for Contextual Recommendations Using Instance Segmentation”. In International Conference on Human-Computer Interaction, July, 2023, pp. 395-408, Cham: Springer Nature Switzerland.

[C2] **D. Tsiktsiris**, A. Vafeiadis, A. Lalas, M. Dasygenis, K. Votis, and D. Tzovaras. “A novel image and audio-based artificial intelligence service for security applications in autonomous vehicles”. In Transportation Research Procedia, vol. 62, p. 294–301,2022.

[C3] **D. Tsiktsiris**, A. Lalas, M. Dasygenis, K. Votis, and D. Tzovaras. “An Efficient Method for Addressing COVID-19 Proximity Related Issues in Autonomous Shuttles Public Transportation”. In Artificial Intelligence Applications and Innovations: 18th IFIP WG 12.5 International Conference, AIAI 2022, Hersonissos, Crete, Greece, June 17–20, 2022, Proceedings, Part I, 2022.

[C4] A. Sideris, T. Sanida**, D. Tsiktsiris**, and M. Dasygenis. “Image Hashing Based on SHA-3 Implemented on FPGA”. In Recent Advances in Manufacturing Modelling and Optimization: Select Proceedings of RAM 2021, Springer, 2022, p. 521–530.

[C5] **D. Tsiktsiris**, T. Sanida, A. Sideris, and M. Dasygenis. “Accelerated Defective Product Inspection on the Edge Using Deep Learning”. In Recent Advances in Manufacturing Modelling and Optimization: Select Proceedings of RAM 2021, Springer, 2022, p. 185–191.

[C6] T. Sanida**, D. Tsiktsiris**, A. Sideris, and M. Dasygenis. “A Heterogeneous Lightweight Network for Plant Disease Classification”. In 2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2021.

[C7] **D. Tsiktsiris**, A. Lalas, M. Dasygenis. K. Votis, and D. Tzovaras. “Enhanced Security Framework for Enabling Facial Recognition in Autonomous Shuttles Public Transportation During COVID-19”. In Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, Hersonissos, Crete, Greece, June 25–27, 2021, Proceedings,2021.

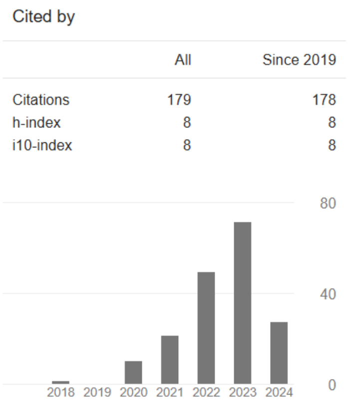
[C8] D. Ziouzios**, D. Tsiktsiris**, N. Baras, S. Bibi, and M. Dasygenis. “A generator tool for Carry Look-ahead Adders (CLA)”. In SHS Web of Conferences, 2021.

[C9] **D. Tsiktsiris**, K. Kechagias, M. Dasygenis, and P. Angelidis. “Accelerated seven segment optical character recognition algorithm”. In 2019 Panhellenic Conference on Electronics & Telecommunications (PACET), 2019.

[C10] **D. Tsiktsiris**, D. Ziouzios, and M. Dasygenis. “HLS Accelerated Noise Reduction Approach Using Image Stacking on Xilinx PYNQ”. In 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2019.

[C11] **D. Tsiktsiris**, D. Ziouzios, and M. Dasygenis. “A portable image processing accelerator using FPGA”. In 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), 2018.

[C12] A. Sideris, T. Sanida, **D. Tsiktsiris**, and M. Dasygenis. “Acceleration of Image Processing with SHA-3 (Keccak) Algorithm using FPGA”.



Thank You!
Questions?