The Fourteenth International Conference on Smart Cities, Systems, Devices and Technologies (SMART 20205)

April 06, 2025 to April 10, 2025 - Valencia, Spain

On-the-Edge Inference Enabled Vision System for Smart Cities

Carmelo Scribano, Ignacio Sanudo Olmedo, Micaela Verucchi, Danda Pani Paudel, Marko Bertogna, Luc Van Gool

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Institute for Computer Science, Artificial Intelligence and Technology

Dr. Carmelo Scribano

Contact: carmelo.scribano@unimore.it

Background:

- Ph.D in Mathematics from University of Modena and Reggio Emilia (UNIMORE), Modena, Italy (2024).
- Visiting Researcher at **INSAIT** (Sofia, Bulgaria) from september 2024 to March 2025 (6 Months).

Research Interests:

- Improving Inference performance of Deep Learning Models for inference on Embedded Devices.
- Computer Vision applications for Automotive.

dAIEDGE Project

- 3-year **Horizon Europe** (GA No. 101120726) initiative (2023–2026) with 36 partners across 15 countries.

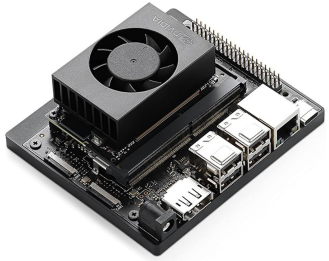
Key Objectives:

- Strengthen Europe’s cutting-edge AI ecosystem by pioneering distributed & **edge AI** solutions
- Develop new paradigms, algorithms & architectures for hybrid, distributed AI
- Create a dynamic network connecting leading research centres, digital innovation hubs, and industry partners

Edge Inference: Hardware

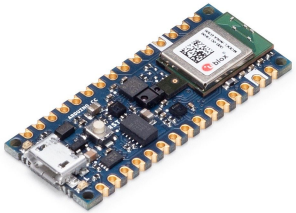
Server Grade GPU (H200)

- 30.000$ (per chip)
- TDP 200W
- 141Gb VRAM
- 2.958 TOPS INT8



Edge AI (ORIN NANO)

- 250$
- TDP 7-25W
- 8GB VRAM
- 67 TOPS INT8



Tiny ML (STM32)

- 10$
- 165 mW
- 1MB SRAM
- <1 GOPS

INSAIT and **UNIMORE** in **dAIEDGE**

- **HIPERT** (UNIMORE Spin-Off) is leader of the **Smart-City** use case.
- **INSAIT** is supporting the Smart-City use case (among the others) implementing cutting edge multi-task learning for edge device.
    - This approach integrates various computer vision tasks (classification, detection, and segmentation) enabling multiple functions to operate efficiently with reduced memory consumption.

This work represents the first step in the collaboration between Unimore and INSAIT.

Smart-City Use case: **MASA**

Cities need intelligent infrastructure to support autonomous vehicles and responsive urban planning: Traffic flow monitoring, Incident detection (e.g., collisions, illegal turns), Pedestrian and vehicle tracking

The **Modena Automotive Smart Area (MASA)** is the testbed for the smart city use case:

- Italy's first open-air urban laboratory dedicated to experimentation of Autonomous Driving, vehicle-to-any (V2X) connectivity and Smart City technology.
- 3km^2-wide area of urban territory, adjacent to a transportation hub (train station, bus stops), equipped with **cameras**, sensors and private **communication networks** (4G, 5G soon).

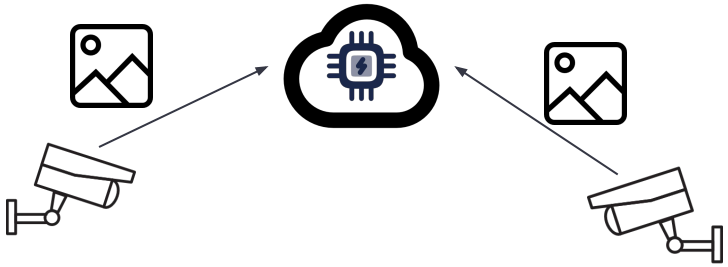
MASA
Modena Automotive Smart Area

Smart City Use case: Edge Inference

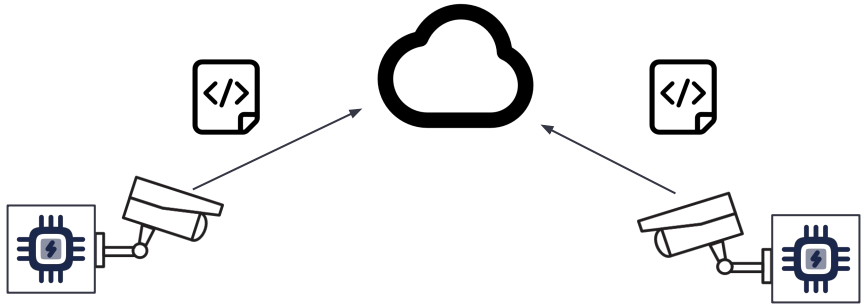
On-edge inference ensures:

- Low latency response (no need for cloud roundtrip)
- Reduced bandwidth usage
- Privacy preservation, with no Image data being transmitted

Cloud Inference Infrastructure



Edge Inference Infrastructure



HAura: Edge Computing for Smart City

HAura is the Road Side Unit (**RSU**) being developed by HIPERT/UNIMORE

Key Features:

- Built around NVidia Orin Nano SoC.
- Dual RGB Camera.
- WiFi, Ethernet, 4G and 5G connectivity.
- Over-The-Air (OTA) Upgradable
- Powerful Computer Vision Sack, running entirely **on the edge**

Fig1. HAura hardware installed at MASA

HAura Technology Stack

HAura process detections entirely **on-edge**, transmitting only metadata.

- JSON structure to encode trackID, position and localization.
- Transport over MQTT Protocol

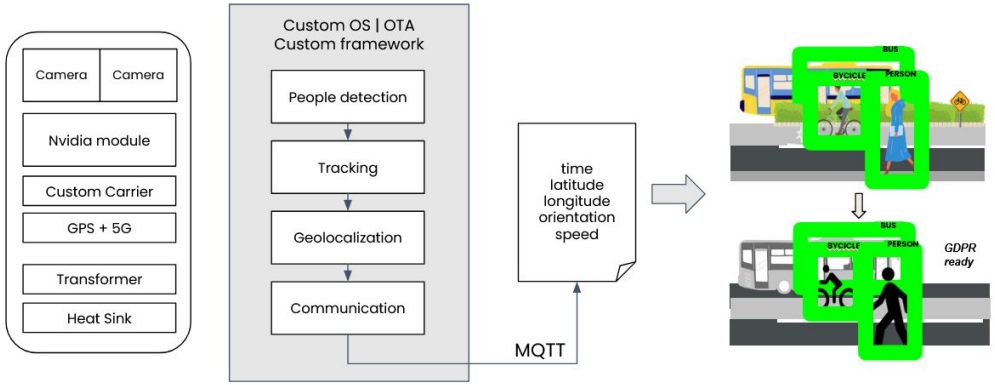


Fig2. HAura execution pipelineFig1. HAura hardware installed at MASA

```
{
  "camIdx": 0,
  "nObjects": 1,
  "objects": [
    {
      "latitude": 45.06582260131836,
      "longitude": 7.662070274353027,
      "speed": 0.0,
      "orientation": 0,
      "id": 1089,
      "cl": 2
    }
  ]
}
```

Fig3. Sample HAura metadata

HAura Vision Stack (V1)

- Self-Diagnostic: Lightweight DNN monitor camera feed for occlusion and dirt.
- Object Detection: YOLO-V4 [1] object detector with 6 classes (person, car, bike, bicycle, truck, bus).
- Multi-Object-Tracking: Based on BYTETRACK [2]
- GeoTracking: Reprojecting 2D detection to GPS coordinates leveraging camera extrinsics

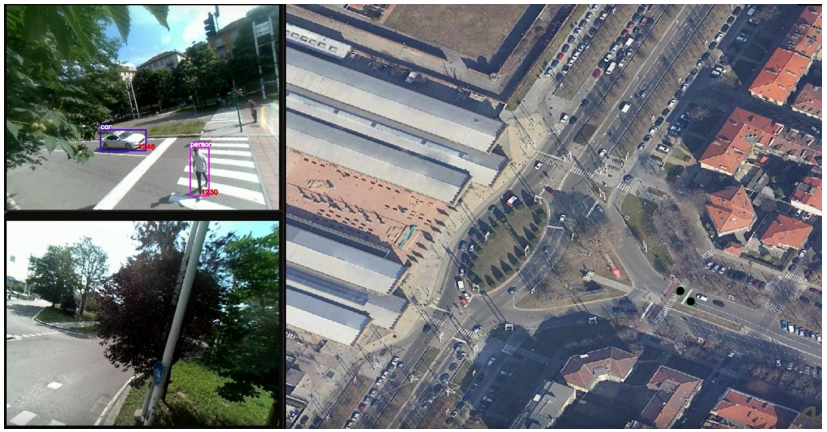


Fig4. HAura perception output

HAura Agregator

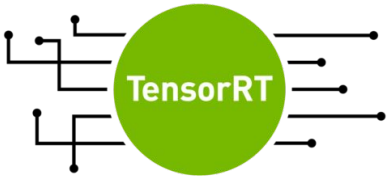
The metadata produced is sent to an **Aggregator** server.

- HIPERT provide **Smart Traffick Monitoring (STM)** functionality, which include detection and tracking of road users (Cars, Bikes, Pedestrians..).
- Third-Parties will be able to implement innovative applications on top of STM aggregated metadata.

Fig5. Haura STM interface

Edge Inference in Practice

HAura stack leverage TensorRT, NVIDIA's high performance inference framework

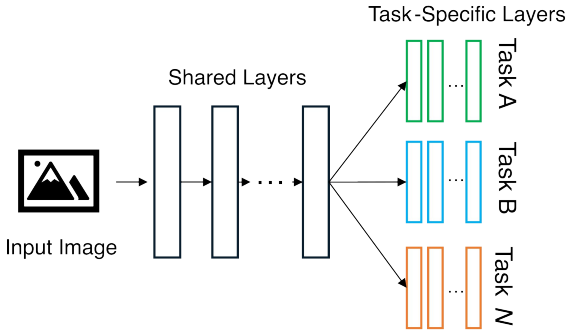


- Deploying powerful vision models ad the edge require minimizing computational cost:
  - Reduced precision computation with quantization
  - Structured pruning (i.e, removing layers) or unstructured pruning (i.e, removing weights).
  - Specialized architectural choice (e.g, Multi-Task Learning).

- **INSAIT** and **UNIMORE** are collaborating on developing cutting edge approaches to deploy next-generation Vision Foundation Models at the edge.

Next-Gen Vision Stack (V2)

A novel **Multi-Task** perception model is being developed by **INSAIT**



Key Features:

- Powerful foundation backbone based on DINO-V2 [3].
- Including tasks of Object Detection, Panoptic Segmentation, Depth Estimation and Human Pose Estimation (ICCV’25 Submission).
- Novel technique to reduce computational footprint of DINO-V2 Backbone (ICCV’25 Submission).

Multi-Task Perception Model

(INSAIT) “AHMAD: Adaptive Hybrid Multi-task Vision Learning with Assisted Distillation”

- A Generalist Multitasking Vision framework for five vision tasks. Built on past work [4].
- SOTA results for panoptic and semantic segmentation for COCO-val [5].
- Competitive results for keypoint detection and closed-vocabulary object detection.

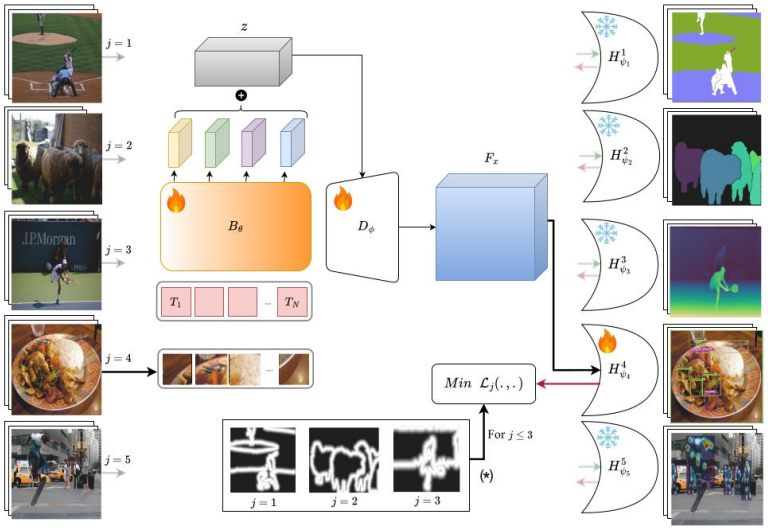


Fig6. AHMAD overview

References

dAIEDGE Project: https://daiedge.eu/

Modena Automotive Smart Area: https://www.automotivesmartarea.it/?lang=en

[1]: Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection."

[2]: Zhang, Yifu, et al. "Bytetrack: Multi-object tracking by associating every detection box." European conference on computer vision.

[3]: Oquab, Maxime, et al. "Dinov2: Learning robust visual features without supervision."

[4]: Prisadnikov, Nedyalko, et al. "A Simple and Generalist Approach for Panoptic Segmentation."

[5]: Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context."

Thank you!