Research Business Cooperation for Self Driving Cars

28 February 2024

Prof. dr. eng. Sergiu Nedevschi
Technical University of Cluj-Napoca, Romania

Outline

1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

Cluj-Napoca important academic, industrial and business centre

Cluj-Napoca important academic, industrial and business centre

- Capital city of Transylvania
- Rich in history, tradition and beautiful nature
- One of the most important academic, industrial and business centres in Romania
- City of Universities with around 100.000 students
- Important technology hub with more than 1800 software companies
- Bosch Engineering Centre in Automated and Assisted Driving
- Rich cultural life – theaters, operas, museums
- International artistic events: film festivals and two of the largest European music festivals



Technical University of Cluj-Napoca (TUCN)

KEY FIGURES

20.000+ full time students (Bachelor/Master/PhD)

900+ academic staff

750+ non-academic staff

180+ study programs (Bachelor/Master/PhD)

12 faculties and 36 academic departments

80+ research units

2 campuses in Cluj-Napoca and Baia-Mare
4 subsidiaries in Bistrita, Zalau, Satu-Mare, Alba-Iulia

Faculty of Architecture and Urban Planning
Faculty of Automation and Computer Science
Faculty of Automotive, Mechatronics and Mechanical Engineering
Faculty of Building Services Engineering
Faculty of Civil Engineering
Faculty of Electrical Engineering
Faculty of Electronics, Telecommunications and Information Technology
Faculty of Industrial Engineering, Robotics and Production Management
Faculty of Materials and Environmental Engineering
Faculty of Sciences
Faculty of Engineering
Faculty of Humanities

Image Processing and Pattern Recognition Research Centre

- **Set up in 1998** as an independent research team within the Computer Science Department
- Currently **14 academic and research staff and 16 research staff** (PhD students)

Areas of Expertise and Domains of Application

Image processing & pattern recognition

Sensors calibration

Computer vision & artificial intelligence

Monocular, Stereo, LiDAR & RADAR based perception

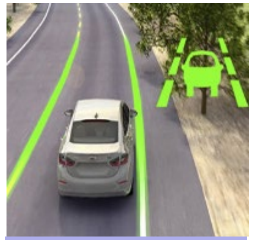
IPPR RC

Machine learning & deep learning for perception


Autonomous forklifts


Autonomous vehicles


Autonomous UAVs


Driving assistance


Medical imaging


Low Earth orbit objects

Partners:

Volkswagen AG, Germany
Robert Bosch GmbH, Germany
Sick AG, Germany
VTT, Finland
AROBS, Romania
IBM Zurich, Switzerland
INRIA Rocquencourt, France
LMA STELaRLab

ETH Zurich, Switzerland
University Joseph Fourier Grenoble, France
University of Braunschweig, Germany
Czech Technical University in Prague

Current Research Projects

SenseMaking **-** Collaborative research project targeted toward the development of a distributed autonomous response for Humanitarian Assistance and Disaster Relief (HADR), in particular, all-domain wildfire response**,** funded by LM (2022-2025)

DeepPerception - Deep Learning Based 3D Perception for Autonomous Driving, PN III-P4-ID-PCE (2022-2025)

FLYSURVEY - Lidar and Photogrammetry Hybrid UAV System for Data Fusion in Construction and Land Monitoring, PD, (2022-2024)

AdaREe - Adaptive Deep Learning for Robust Depth Perception from Monocular Images, PD, (2022-2024)

INOVSAFE - Identification and monitoring of internal and external risk factors for the safety of motor traffic PD, (2022-2024)

EuroDrives - Low level 4D perception (geometry, semantics and temporal fusion) - Bosch – TUCN cooperation in the framework of *European microelectronics value chain for highly automated driving* (“EuroDrives”) (2024-2027)

Research Projects

SEPCA - Visual Semantics and Integrated Control for Autonomous Systems, code PN III-P4-ID-PCCF-2016-0180, (2018-2022)
CAMELEON - Compact image Acquisition and position MEasurement system for targets in the LEOraNge, grant funded by Romanian Ministry of Education and Research, code PN-III-P2-2.1-PED-2019-4819 (2020-2022)
GENDARED - Generic Data Reduction Framework for Space Surveillance, funded by ESA, (2019-2020)
DEEPSENSE - Advanced environmental perception techniques using Deep Learning and probabilistic estimators, code PN-III-P1-1.1-TE-2016-0440, (2018-2020)
VLL – Video based landmarks for localization, funded by Robert Bosch GmbH, (2017-2018)
MULTISPECT - Multispectral environment perception by fusion of 2D and 3D sensorial data from the visible and infrared spectrum, code PN-III-P4-ID-PCE-2016-0727, (2017-2019)
UP Drive - Automated Urban Parking and Driving, H2020 project, (2016-2020)
MULTIFACE - Multifocal System for Real Time Tracking of Dynamic Facial and Body Features, PN-II-RU-TE-2014-4-1746 project, (2015-2017)
R5COP - Reconfigurable ROS-based Resilient Reasoning Robotic Cooperating Systems", FP7 ARTEMIS (2014-2017),
RSM - Road surface measurement and modeling, funded by Robert Bosch GmbH, (2013-2016)
PAN-ROBOTS - Plug and Navigate ROBOTS for smart factories, FP7 project, (2012-2015)

Outline

1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

The „false negative“ / „false positive“ problem

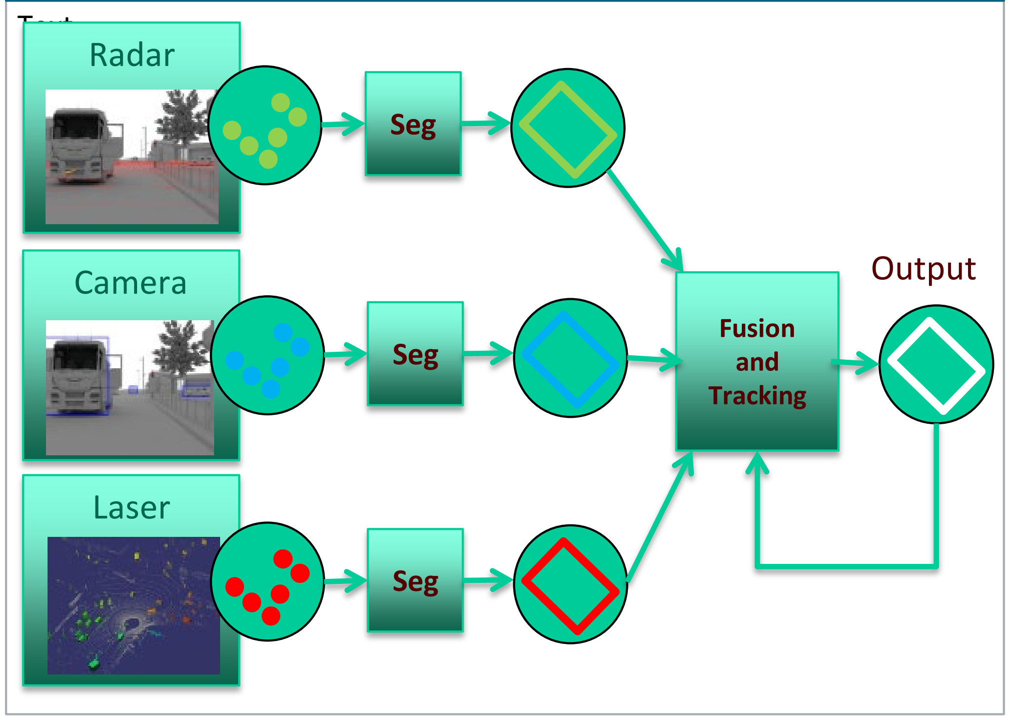
Key solutions for solving the challenge

- **Multiple redundancy at sensory system and algorithmic levels**
  - Different sensor modalities for diverse redundancy (Camera, Stereo, LIDAR, Radar) and 360º coverage for all
  - Different detection algorithms – algorithmic redundancy
  - At least three modalities are needed
- **New and diverse detection algorithms**
  - New deep learning based perception algorithms
  - Low level raw data fusion and enhanced perception algorithms
- **Independent solutions**
  - Use of conventional high-level fusion
- **Joint development of perception, scene understanding, prediction and driving functions**

Sensory Perception for Autonomous Vehicles

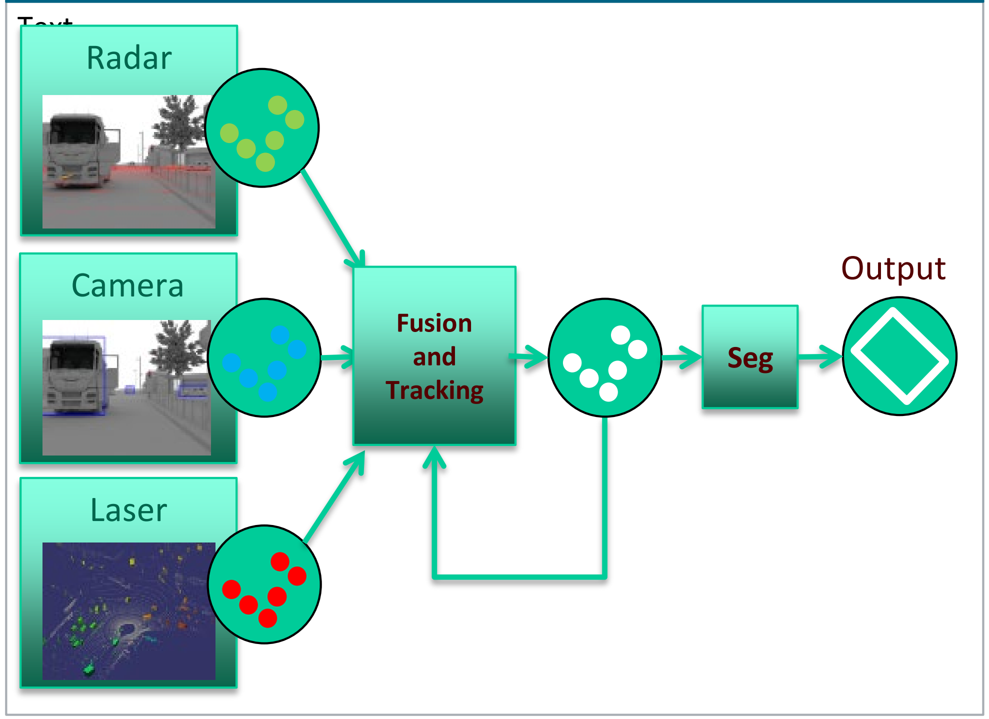
Early Fusion vs. Late Fusion

High-level (late) fusion



Perception based on multiple independent sensors. Robustness against sensor failure. Complex fusion and tracking tasks.

Low-level (early) fusion



Low level geometry, semantic and motion fusion generates an enhanced low level representation. Enhanced perception, tracking and scene understanding possibilities.

UP-Drive
Automated Urban Parking and Driving

Autonomous Urban Parking and Driving - UP-Drive

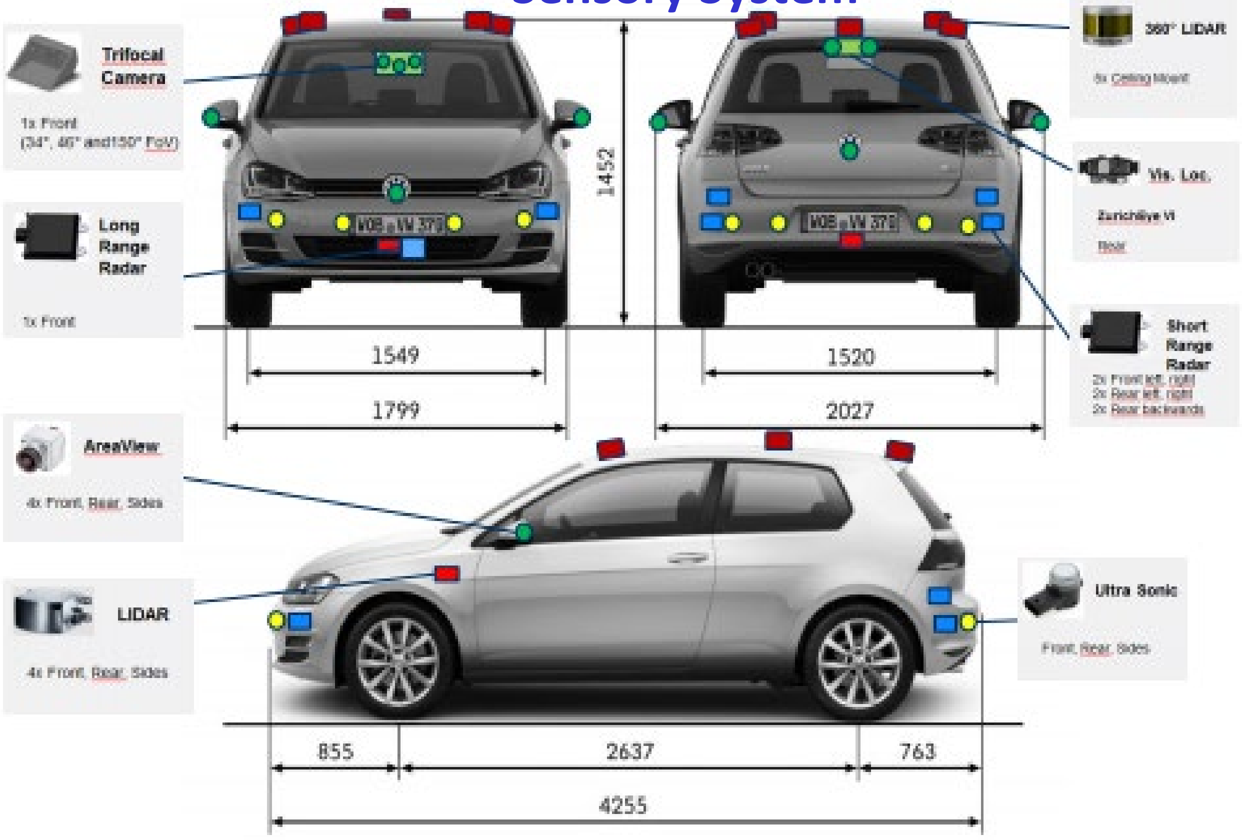
Requirements: - 360° multi-modal perception
- processing cycle > 5 fps
- high accuracy

Constrains: - 30km/h zones
- Mapped area

Overview of the System Hardware Architecture

SETUP GOLF 7

Sensory System

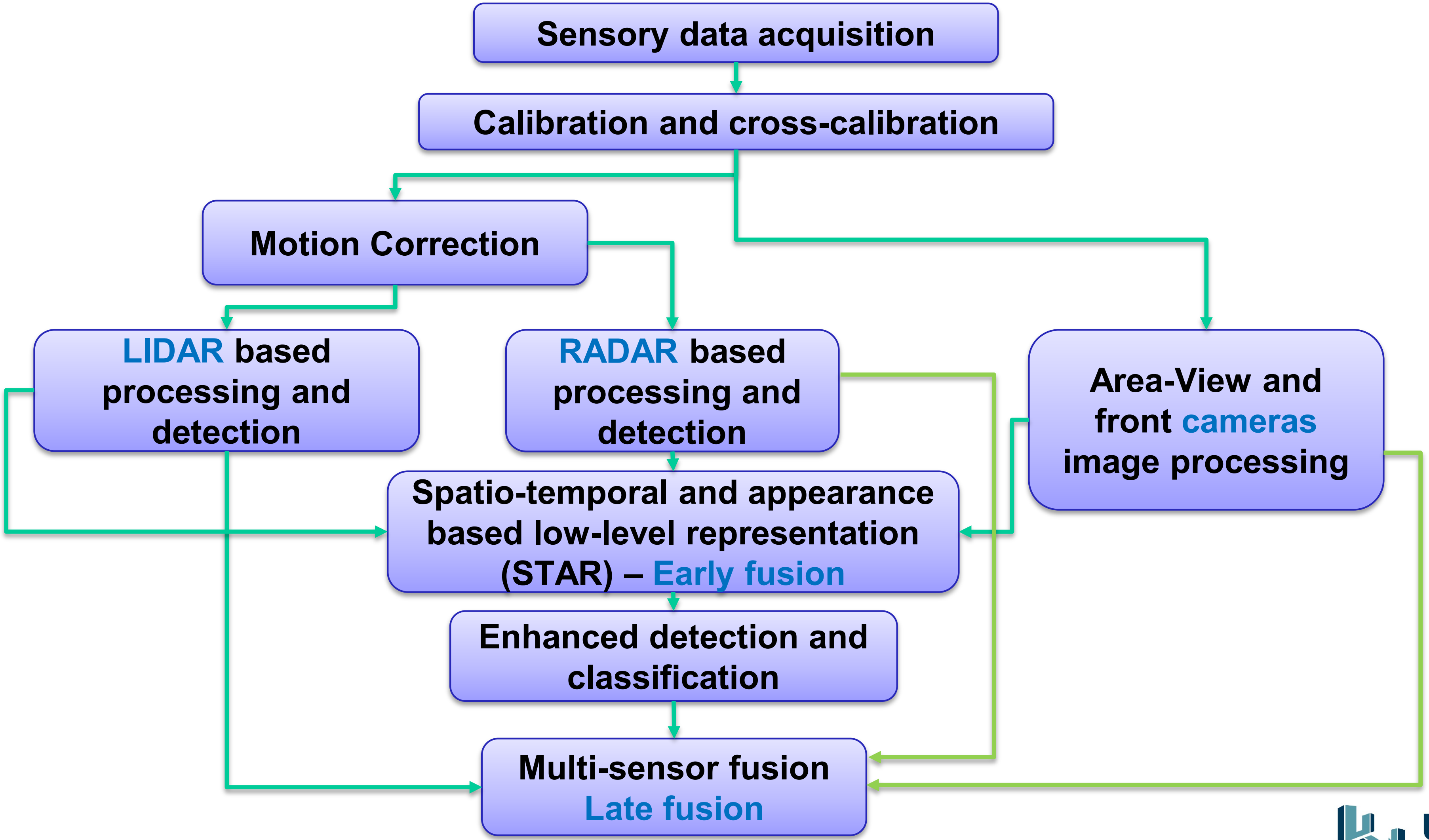


Processing Hardware

12 industrial PCs equipped with Intel Core i7 processors, with 4 cores, operating at 2.6GHz
2 NVIDIA GTX1080 GPUs into two of the industrial PCs (one for proposed perception solution)

UP-Drive
Automated Urban Parking and Driving

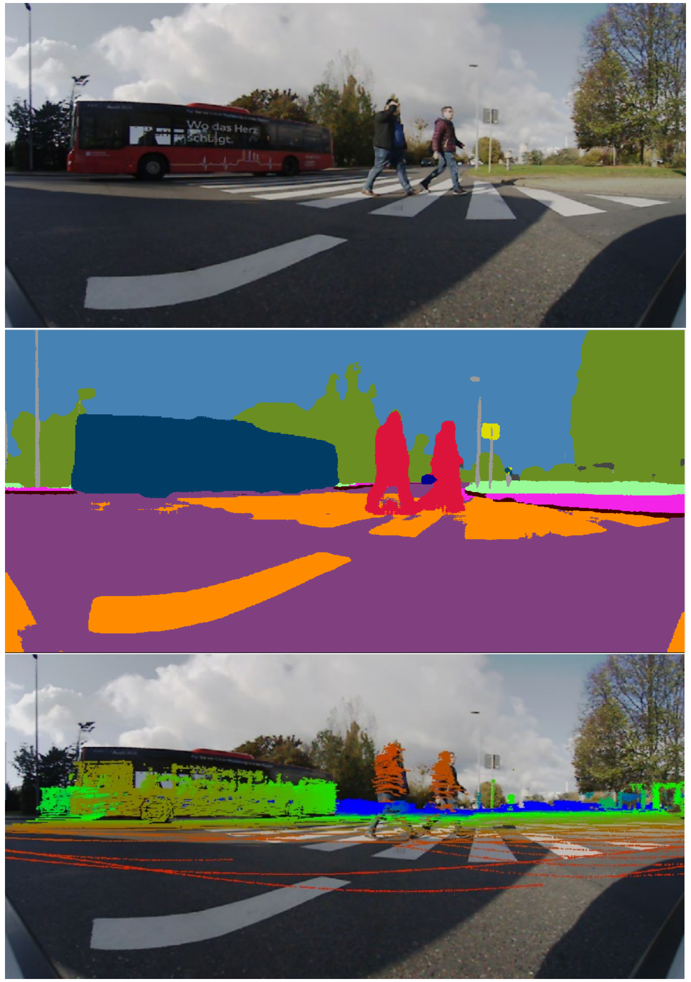
Proposed Solution Based on Low Level Fusion



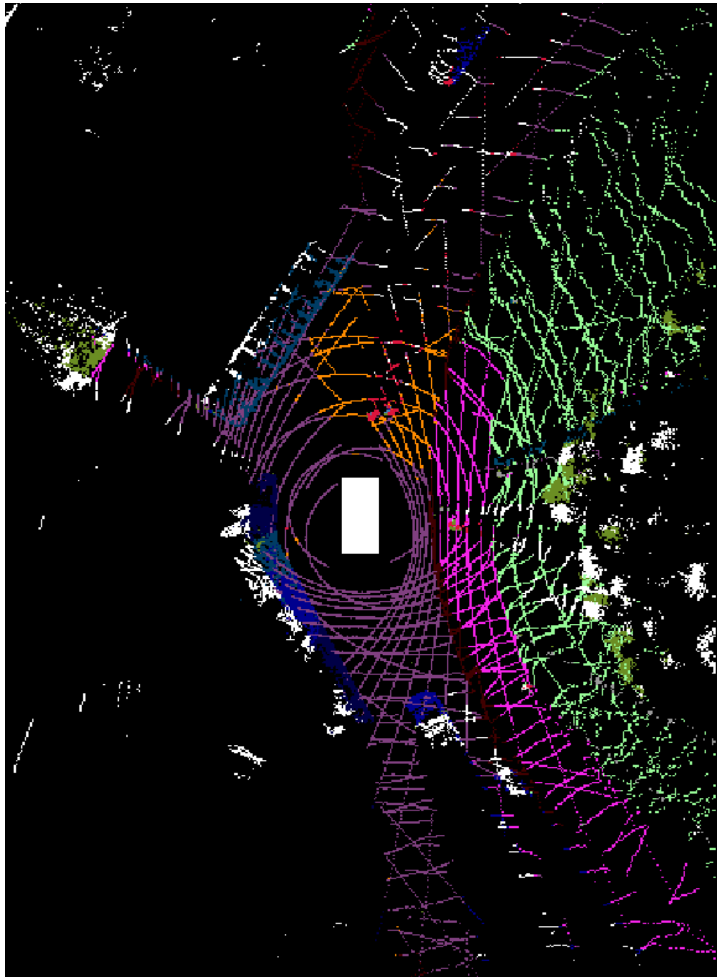
UP-Drive
Automated Urban Parking and Driving

Proposed Solution Based on Low Level Fusion

Semantic and Geometry Low Level Fusion (STAR)



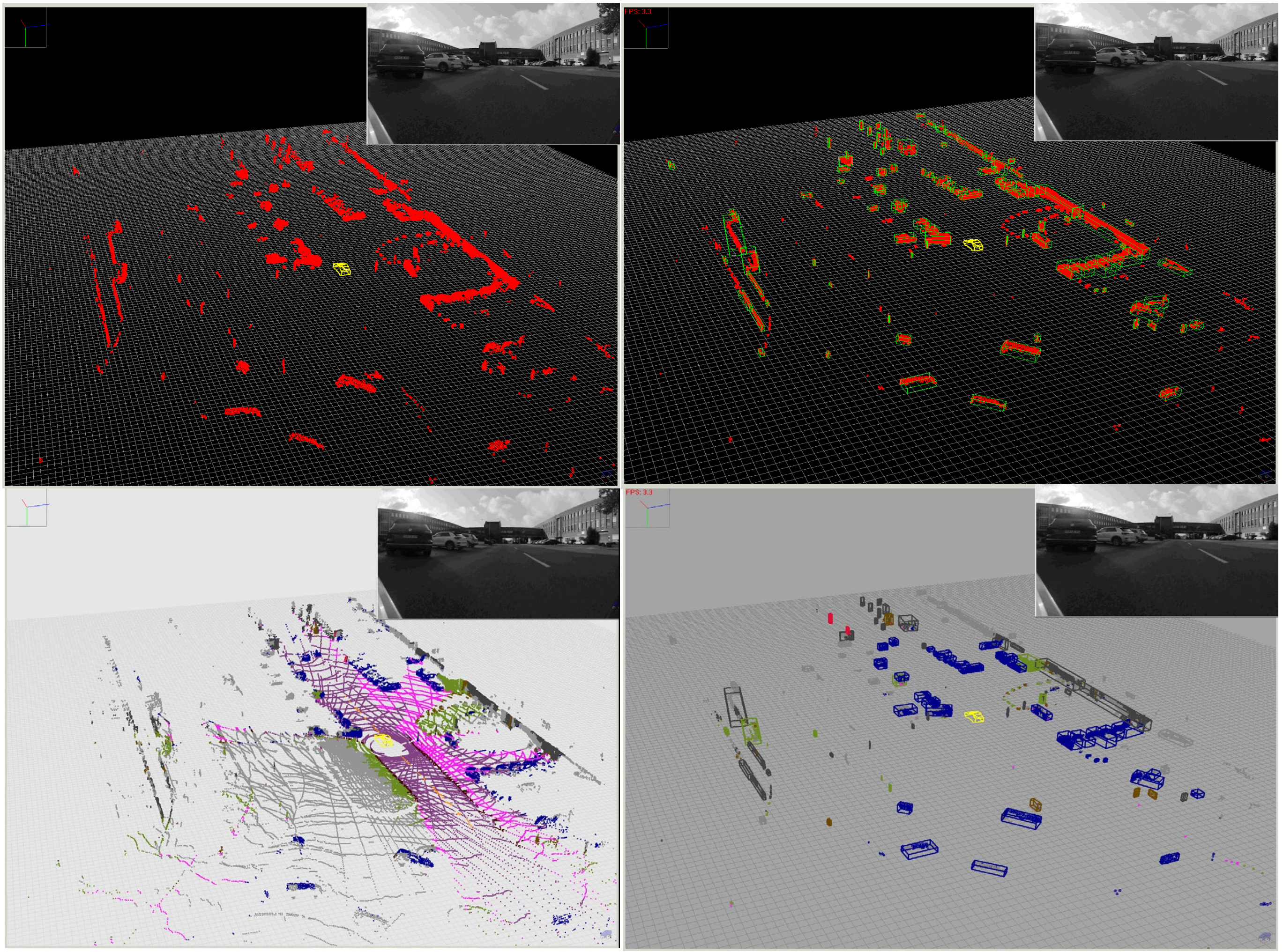
RGB image, semantic segmentation of the image, 3D point cloud projections in the image



Top view of the semantically enhanced 3D point cloud

Proposed Solution Based on Low Level Fusion

Enhanced Detection and Classification



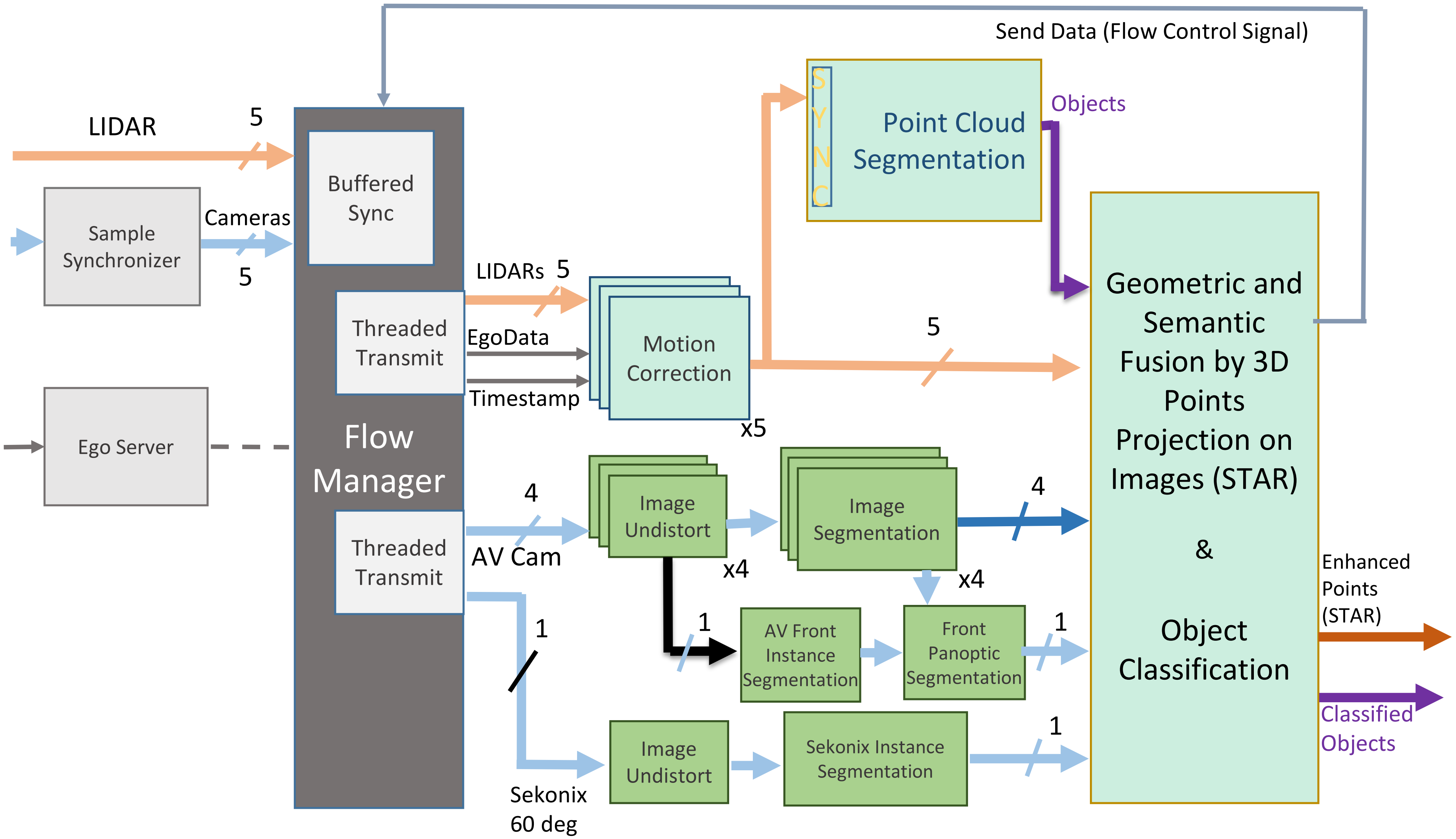
1.3D point cloud from fusion of multiple LiDARs and RADARs data

2 3D point cloud conventional segmentation in cuboids

3. Semantically enhanced 3D point cloud

4. Cuboids refinement and classification

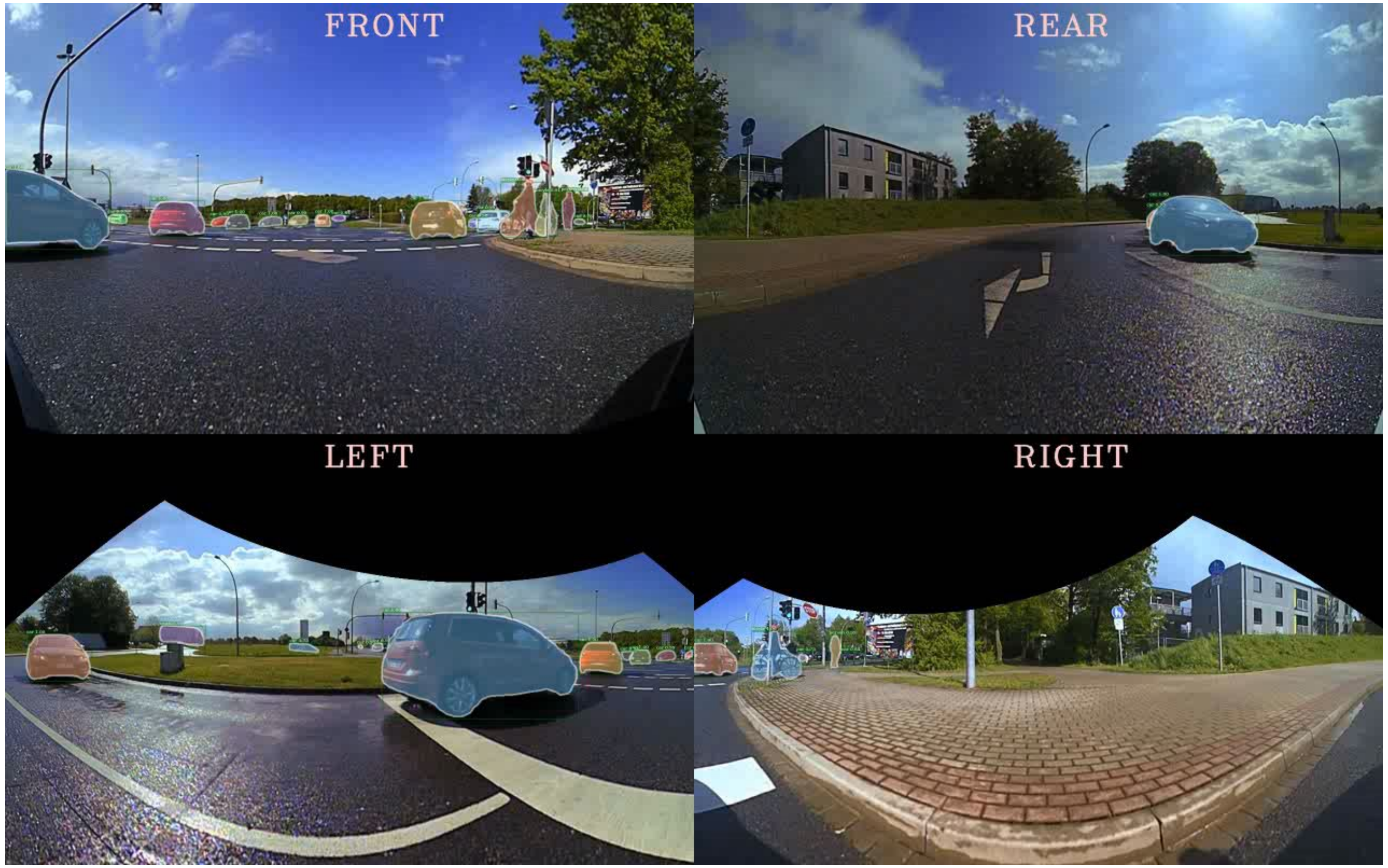
Overview of the ADTF Implementation Architecture



Area View Fisheye Images Semantic Segmentation

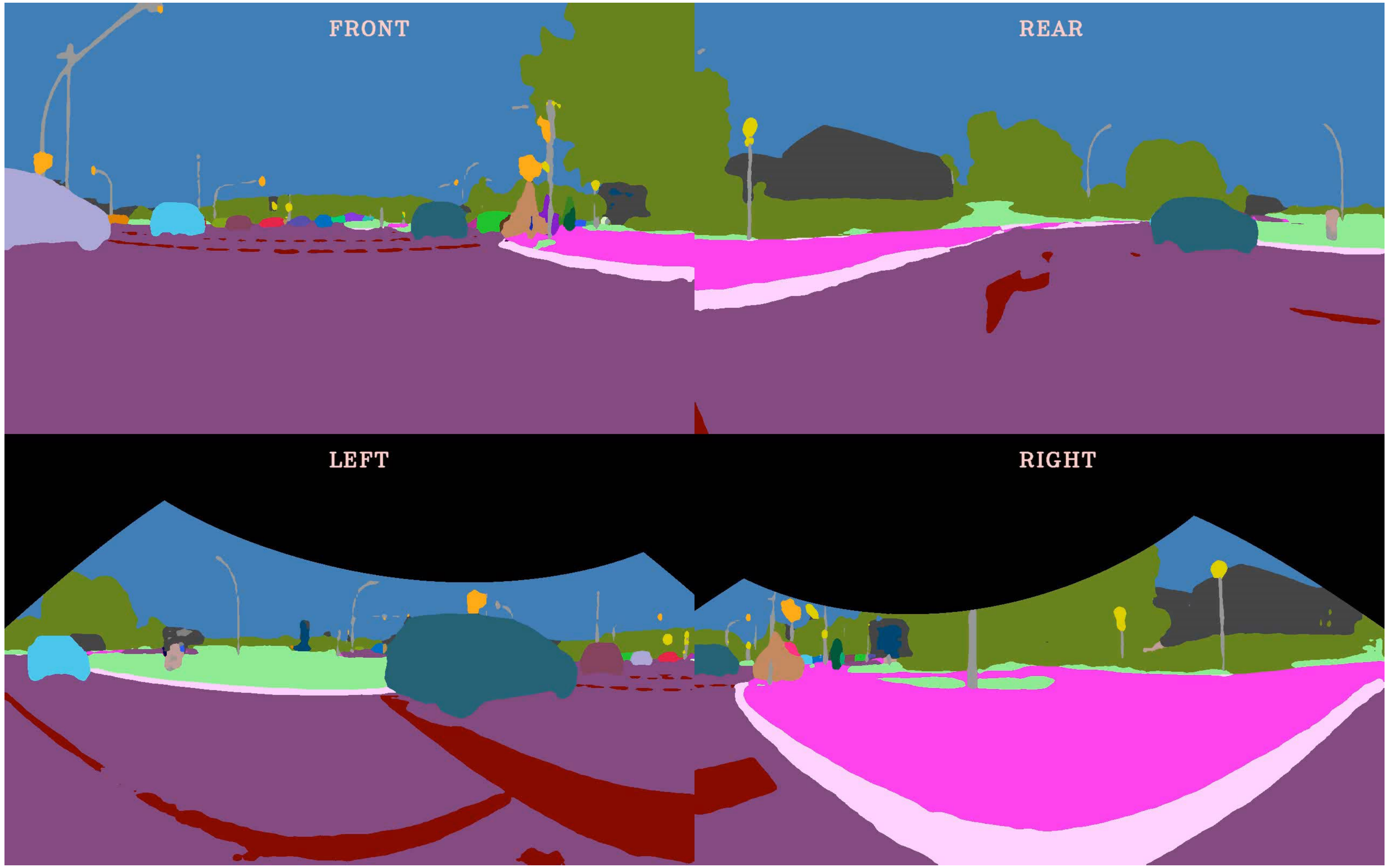
800 x 1280, FV=180∘, ERFNet, 9ms with Tensor RT INT - 8

Area View Fisheye Images Instance Segmentation



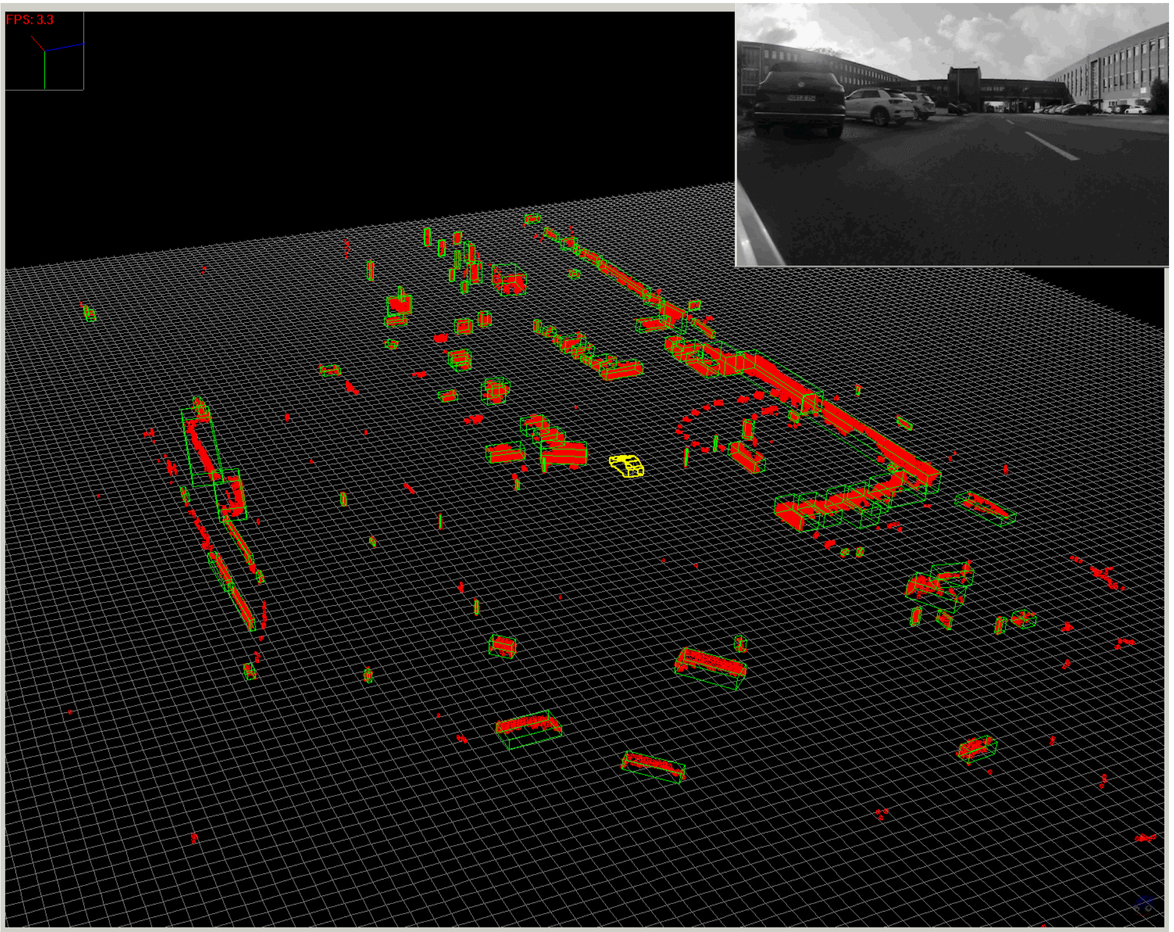
800 x 1280, FV=180°, RetinaNet, 66 ms with Tensor RT, fp - 32

Area View Fisheye Images Panoptic Segmentation

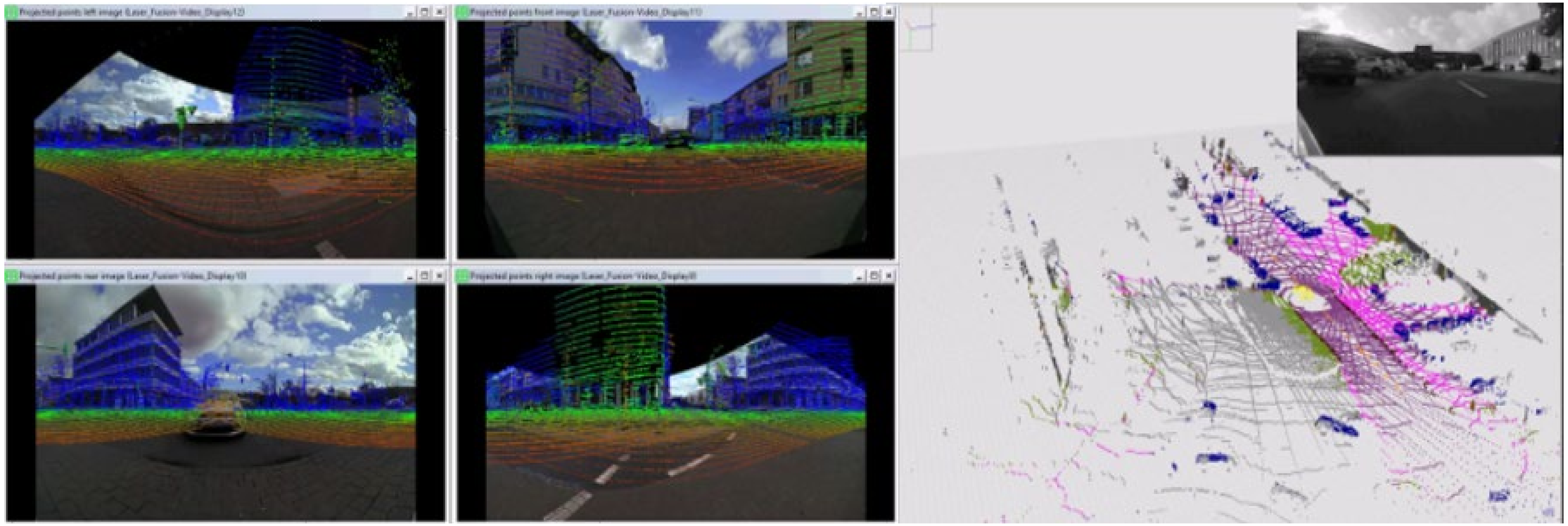


800 x 1280, FV=180°, RetinaNet, 66 ms with Tensor RT, fp - 32

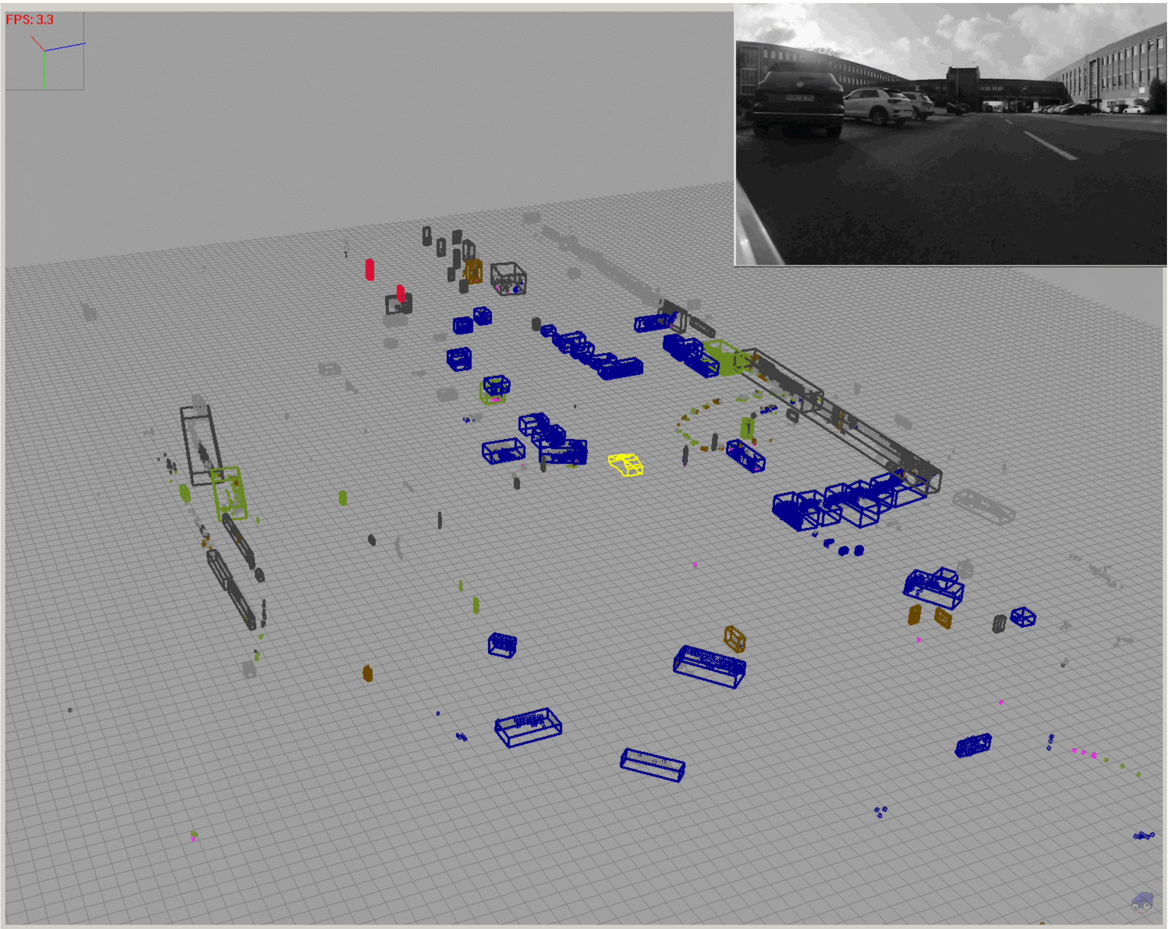
Point Cloud Based 3D Object Detection



Semantically Enhanced Point Cloud (STAR)



Semantically Enhanced Point Cloud (STAR)



Experiments and Online Evaluation

UP-Drive
Automated Urban Parking and Driving

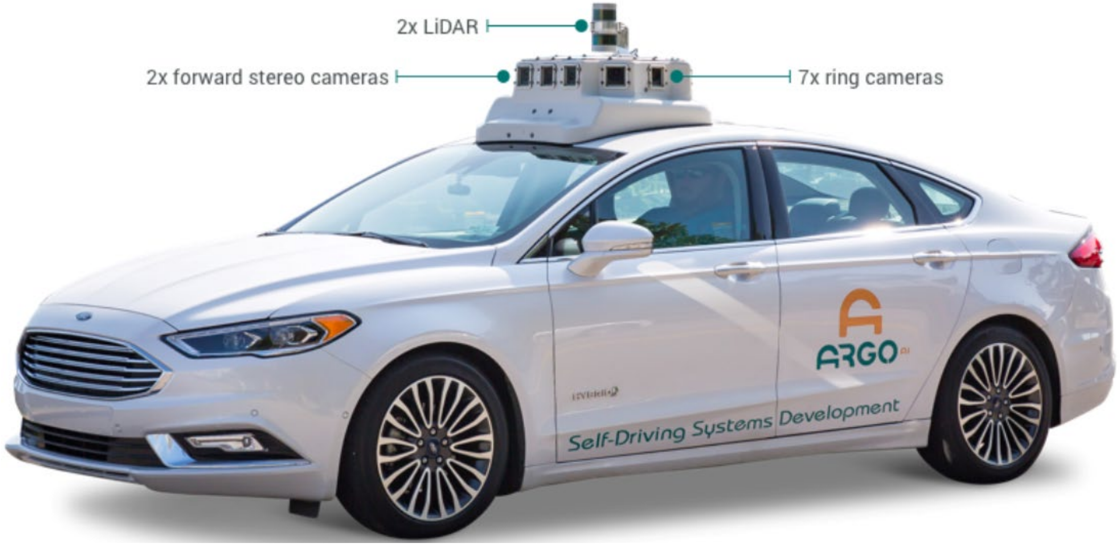
Outline

1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

Lessons Learned

- ➢ **New sensors** would be necessary with less synchronization problems, faster answer, better **360°** and depth coverage
- ➢ Including the **sequence / video processing** approaches **to increase the robustness based on temporal integration**, **to allow dynamics perception** (ego-motion, tracking) and **self-supervised approaches**
- ➢ **Semantic and geometry fusion** in static and dynamic conditions
- ➢ Low-level **fusion together with independent perception solutions** provide the best result

- **Sensor setup should be optimized** for low-level fusion
  - ➢ By placing LiDAR and camera sensors as close to each other as possible
  - ➢ Synchronizing all sensors
  - ➢ Roof-top box (e.g. Argo) seems to be a solution



Lessons Learned

Camera:

- **Video semantic, instance and panoptic segmentation**
  - Increased robustness, but also dynamic features perception (**ego-motion and tracking**)
- Structure from motion
- **Self-supervised depth estimation**
- **Direct fusion of semantic to geometry**
- Multi-camera approaches providing **360° coverage:**
  - area-view system with 4 fisheye cameras for **near range**
  - narrow field of view cameras in different configurations **for far range**
- **FLIR cameras for night vision**

Lessons Learned

LiDAR:

Traditional versus solid state LiDARs

- Traditional LiDARs provide depth information with HFV of 360 degree (10 frames/s)
- **New generation of LiDARs provide supplementary color and infrared information**
- Solid state LiDARs have reduced horizontal field of view but **higher frame rate** (40 frames/s)

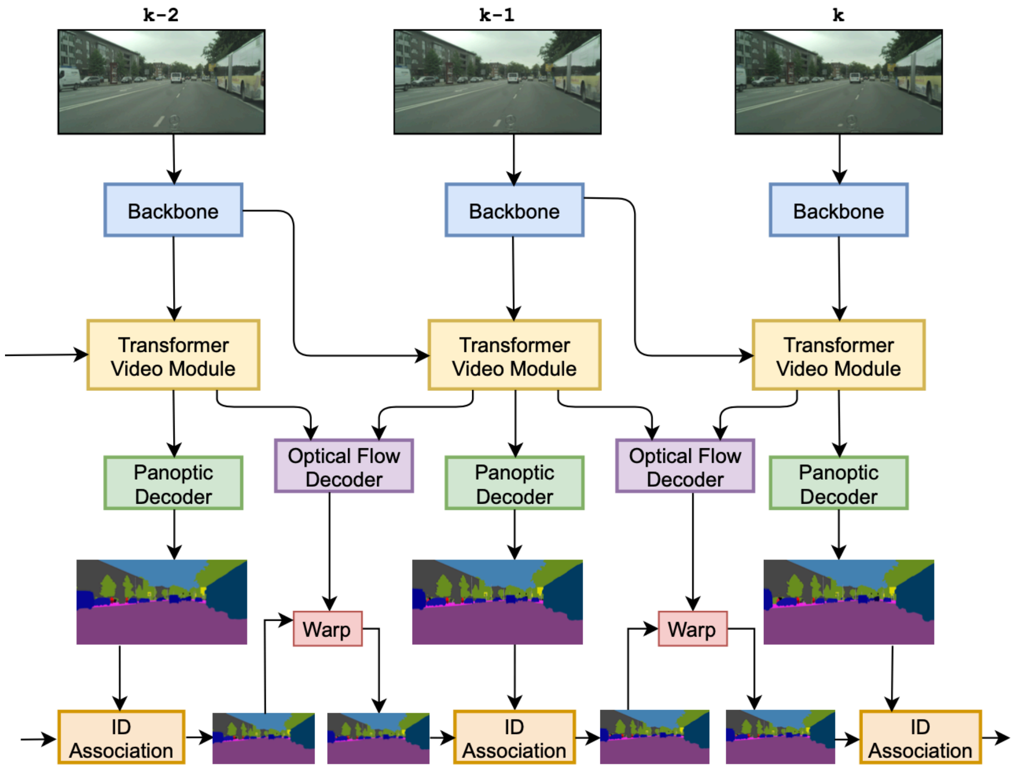
- Deep Learning base processing of 3D point cloud sequences
- Low level fusion with semantic information coming from cameras

Outline

1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

Time-Space Transformers for Video Panoptic Segmentation - New video panoptic network: semantic, instance & panoptic segmentation as well as instance tracking.

- Extend the bottom-up image panoptic segmentation network Panoptic DeepLab to video
- A novel Transformer **video module** refines the current backbone features by processing the sequence of spatio-temporal features
- **Multiple heads:** semantic segmentation, instance center prediction, offset regression and optical flow estimation
- **Tracking** based on **mask propagation with optical flow** and instance ID association between warped and predicted instance masks
- **Optical flow trained in an unsupervised regime** with the photometric loss
- **Three different designs of the attention mechanism**: space self-attention, global time-space attention and local time-space attention.
- **Weakly supervised learning** (only 1 in 5 images are labelled).
- **Extensive experiments on the Cityscapes-VPS dataset** and demonstrate that the proposed modules increase accuracy and temporal consistency



High level overview of the VPS-Transformer network

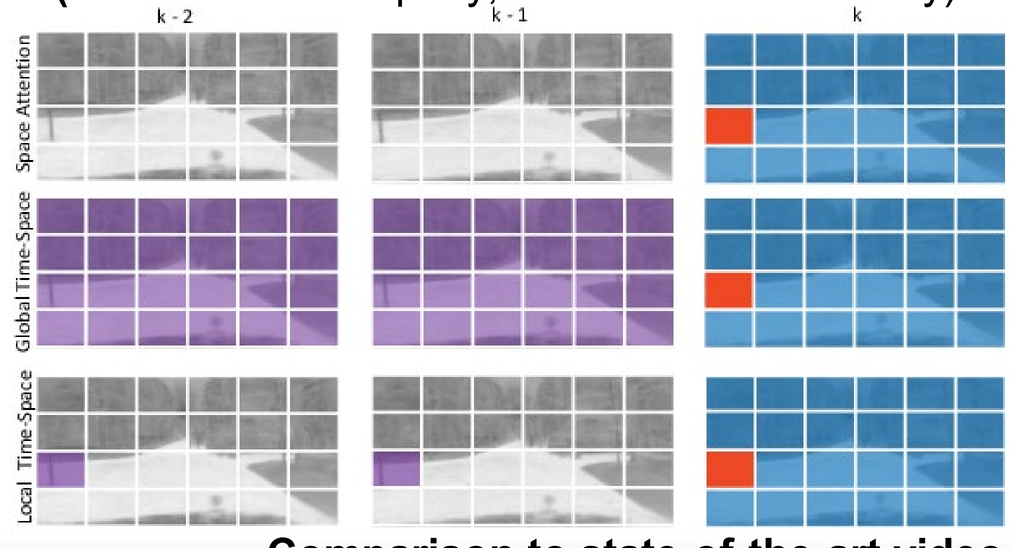
A. Petrovai, S. Nedevschi, Time-Space Transformers for Video Panoptic Segmentation, *Proceedings of WACV*, Hawaii, USA, 3-5 January 2022, pp. 925-934

Video Panoptic Segmentation

Time-Space Transformers for Video Panoptic Segmentation

The three attention schemes

(Current frame = query; Past frames = memory)



Space Attention:
A query location attends to all tokens from the query only

Global Time-Space Attention:
A query location attends to all locations from the memory and all locations from the query

Local Time-Space Attention:
A query location attends to all tokens from the query and to memory tokens located at the same spatial position

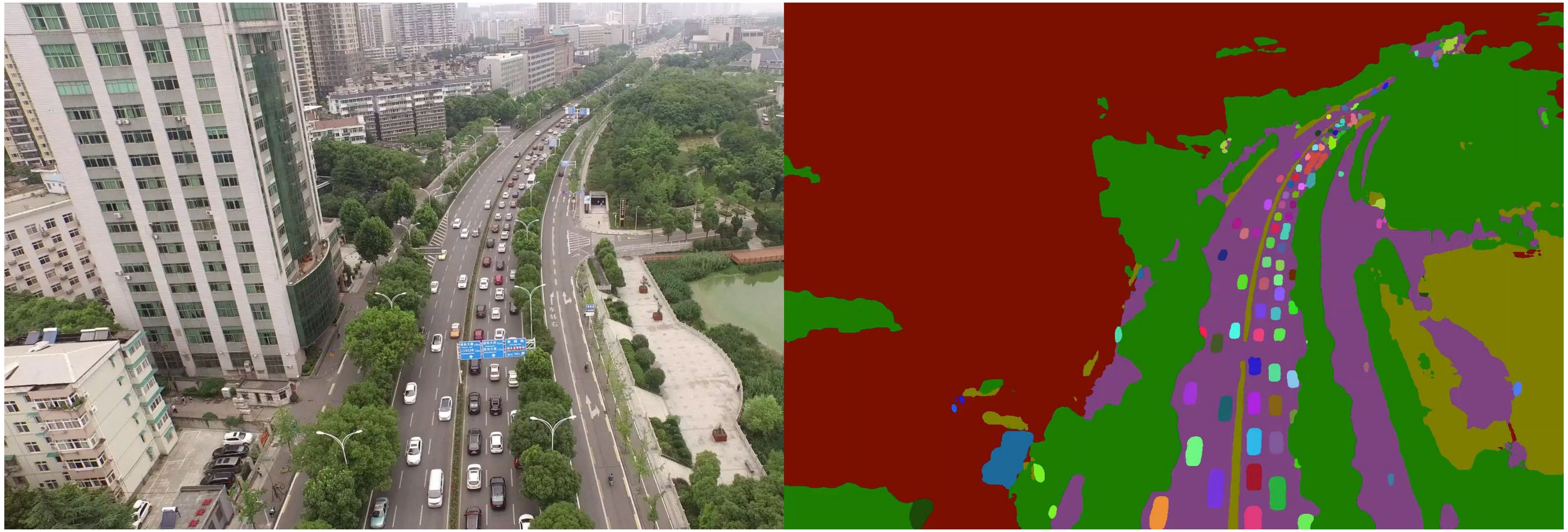
Comparison to state-of-the-art video panoptic

Method	Backbone	PQ	VPQ	Time (ms)
VPSNet [21]	ResNet50	62.7	56.1	770
ViP-DeepLab + MV [33]	WR-41	67.9	59.9	> 400
Baseline - Panoptic DeepLab [8]	ResNet50	63.0	52.0	86
VPS-Transformer (ours)	ResNet50	64.8	57.3	112
Baseline - Panoptic DeepLab [8]	HRNet-W48	66.1	55.1	168
VPS-Transformer (ours)	HRNet-W48	67.6	59.8	185

B + Transformer Video Module S = 1 + Tracking

Video Panoptic Segmentation

Experiments on UAVid data set



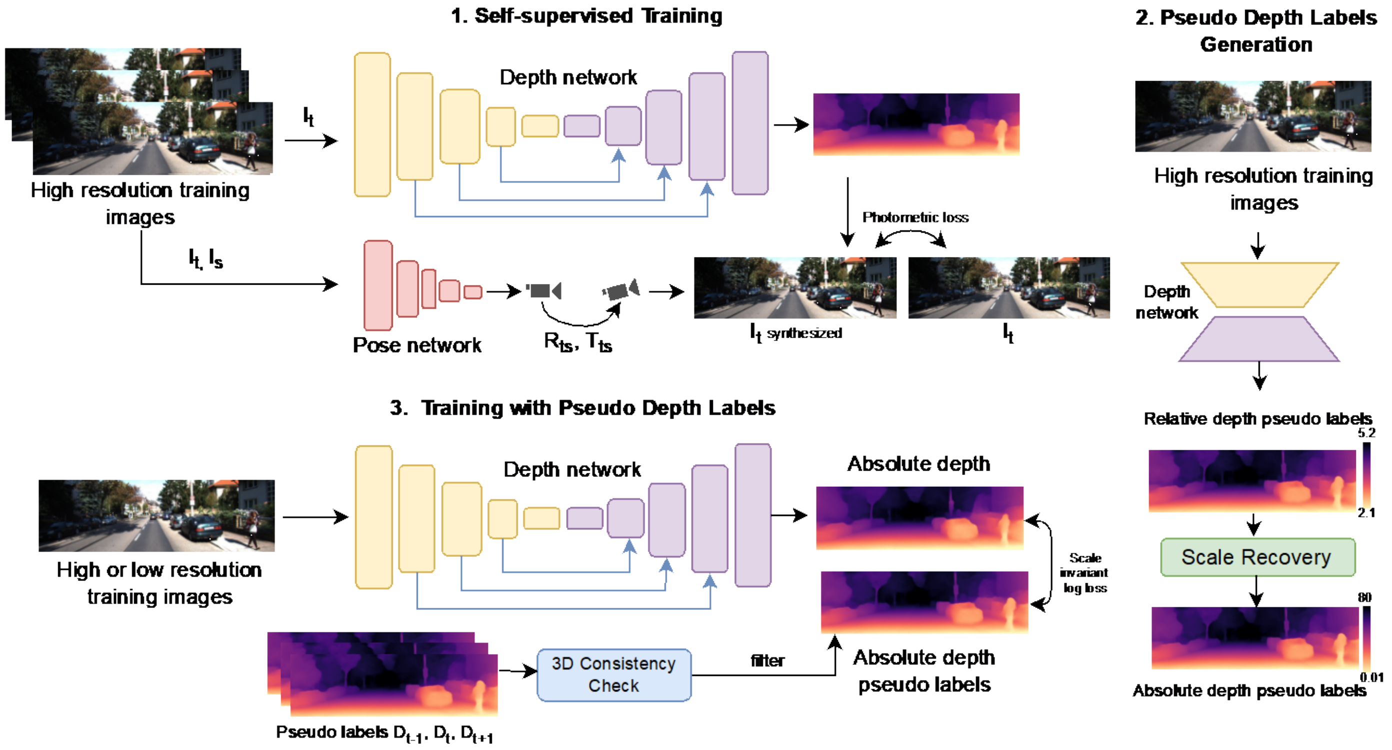
Distillation Framework for Self-supervised Depth Estimation

Exploiting Pseudo Labels in a Self-Supervised Learning Framework for Improved MDE

First stage
- Self-supervised teacher
- High-resolution training images
- Depth network with ResNet-50 backbone
- Auto-masking for stationary pixels and minimum reprojection loss for occlusions

Second stage
- High-resolution pseudo depth labels
- Scale pseudo-labels from relative depth to absolute depth → scale-aware and inter-frame scale consistent pseudo-labels
- Scale recovery from depth prediction using the scale between an estimated camera height and a real-camera height

Third stage:
- Supervised student
- High or low resolution training images
- Depth network with ResNet-50 or ResNet-18 backbone
- Remove errors with a 3D consistency check
- Regression loss



A. Petrovai and S. Nedevschi, "Exploiting Pseudo Labels in a Self-Supervised Learning Framework for Improved Monocular Depth Estimation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR)*, New Orleans, USA, June 2022

Distillation Framework for Self-supervised Depth Estimation

Exploiting Pseudo Labels in a Self-Supervised Learning Framework for Improved MDE

Depth-aware Video Panoptic Segmentation

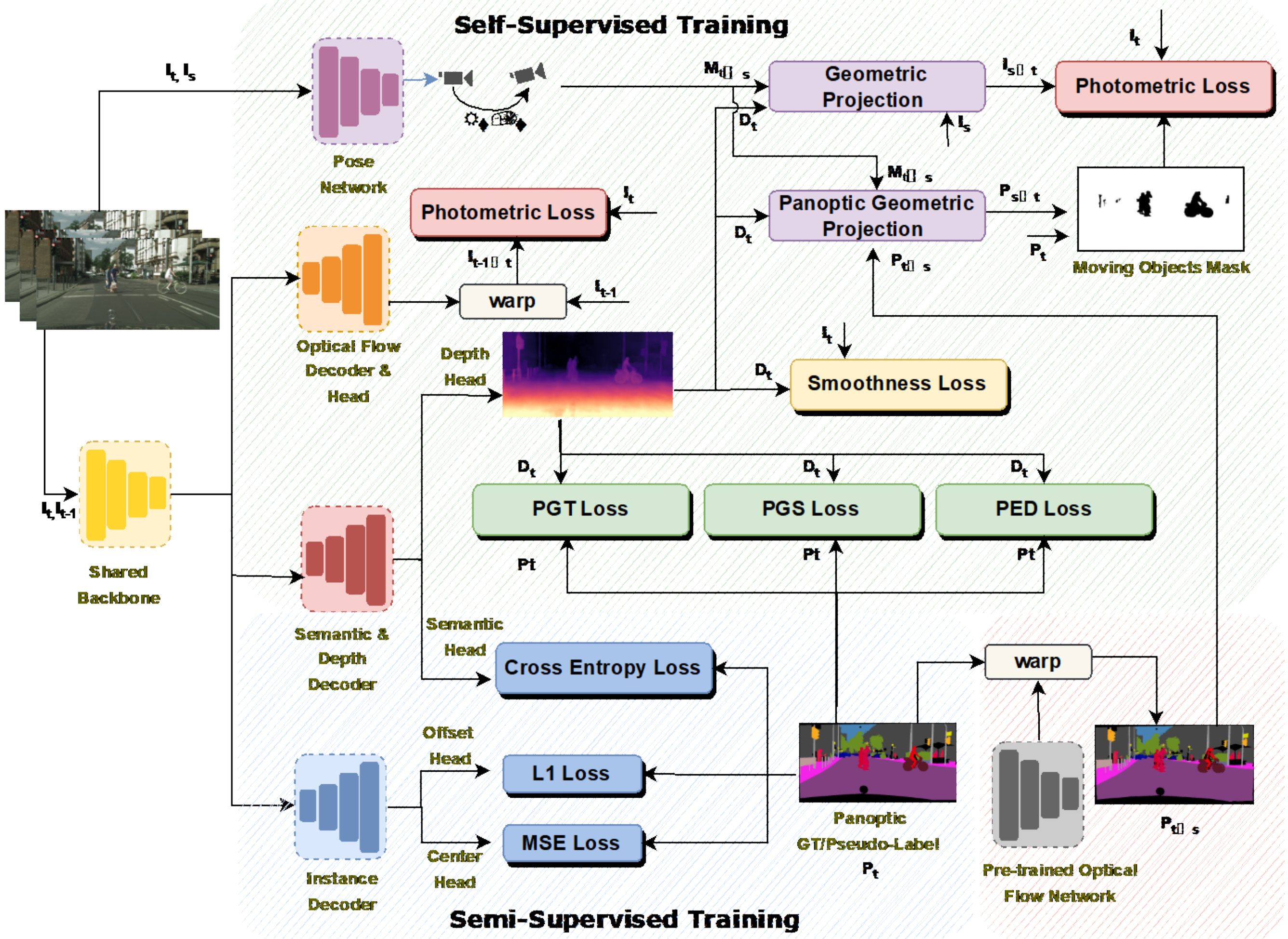
A Self-Supervised Monocular Depth Estimation Approach to Depth-aware Video Panoptic Segmentation

Multi-task network

- Self-supervised depth estimation
- Semi-supervised video panoptic segmentation
- Leverage large amounts of unlabeled data for improved results

Depth improvement with panoptic guidance

- Moving objects masks from video panoptic segmentation
- Panoptic-guided losses



A. Petrovai and S. Nedevschi, "MonoDVPS: A Self-Supervised Monocular Depth Estimation Approach to Depth-aware Video Panoptic Segmentation", accepted for publication in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), USA, January 2023

Depth-aware Video Panoptic Segmentation

A Self-Supervised Monocular Depth Estimation Approach to Depth-aware Video Panoptic Segmentation

Undergoing researches

- Solutions for the use of computer vision foundation models for perception tasks
- Domain adaptation techniques to facilitate the use of synthetic data
- Optimization of the models
- Explainablity of the proposed solutions
  - -based on stabilization and validation
  - -neuro-symbolic approaches

Outline

1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

Current status

Significant expertise in the domain of perception, scene understanding, risk assessment
Multiple generation of perception solutions developed

Looking for automotive customers:

- **Volkswagen**: demonstrators for new autonomous driving solutions
- **Bosch, Continental**: advanced driving assistance and autonomous driving components and systems
- **NPX Semiconductors**: design of specialized automotive processors, accelerators, software

Looking for SMEs:

- Innovative application for UGV or robots in different applicative domains

Looking for publicly funded projects:

- Horizon Europe: perception, scene understanding, risk assessment, planning for mobile systems

Outline

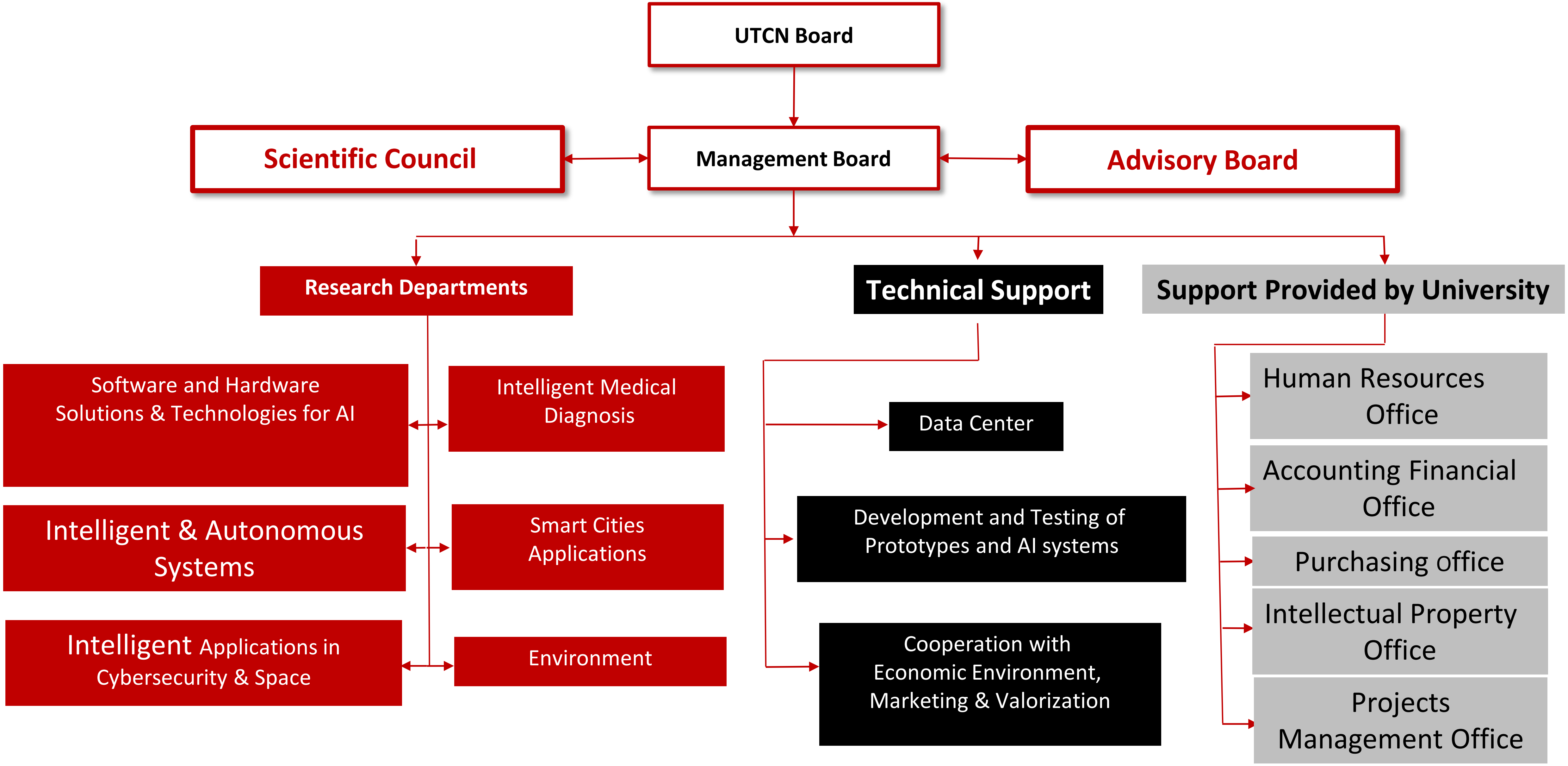
1. Introduction
2. Background
3. Lesson learned
4. New developments
5. Current status
6. Needed help

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE @ TUCN (AIRI @ TUCN)

Activity	Realization stage	Value (EURO)
Technical project for construction permit	✓ Done	
Execution project	✓ Done	
Execution auction	✓ Done	
Execution	2024 -2026	~25.000.000
Research equipment acquisition	2024-2026	~10.000.000

Building Parameters	
Footprint	2,028 m^2
Built area	13.850 m^2
Structure	2U + G + 5L + R
Equipment	
1.	DGX Super POD with DGX H100
2.	DGX H100 DL servers
3.	Software/hardware development systems
4.	Robot/drone/automated vehicle components

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE @ TUCN (AIRI @ TUCN)



Outline

The National Plan for Research, Development and Innovation 2022-2027

- **Improvement of Research Infrastructure**
- **Fundamental and applicative research**
- **Cooperation of research entities with SMEs and Companies**

European Regional Development Fund (ERDF)

- **Smart Growth, Digitization and Financial Instruments (PoCIDIF) Program**

  RO AI HUB - 65 mil. Euro
- **Regional operational programs**

National Recovery and Resilience Plan

- **Improvement of Educational and Research Infrastructure**
- **Updating of the educational programs and introduction of new programs focused on AI at bachelor, master and PhD level**
- **Support for SMEs**

Horizon Europe

- **DIGITAL Innovation HUBs 18 DIH in RO offering assistance to SMEs**
- **Research projects**
- **Research Networks**

TECHNICAL UNIVERSITY

OF CLUJ-NAPOCA

ROMANIA

Sergiu.Nedevschi@cs.utcluj.ro