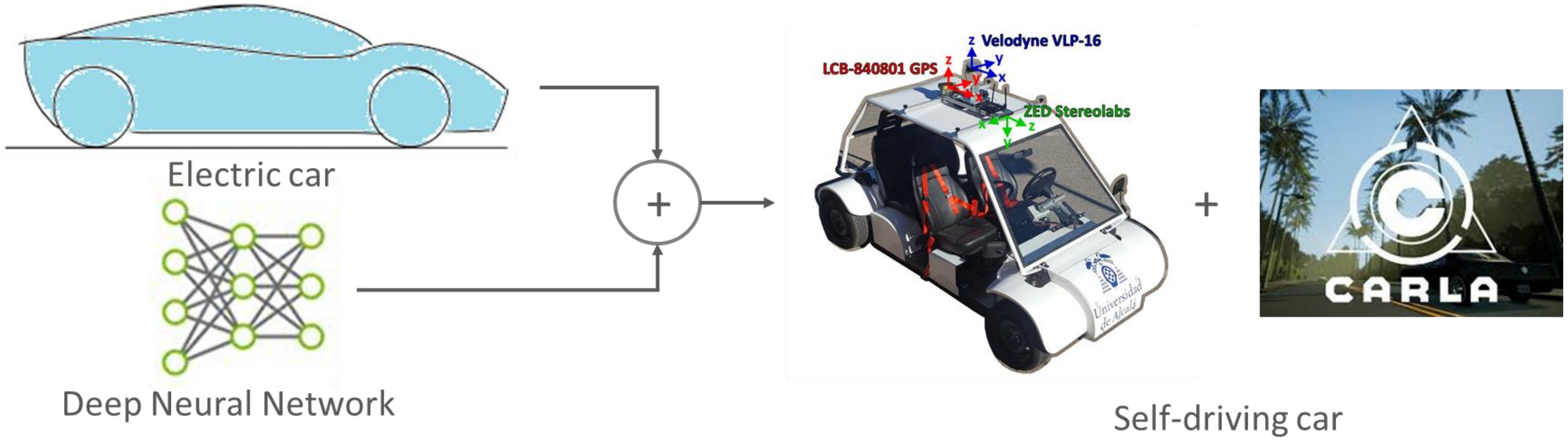
Perception for intelligent vehicles in urban environments

Luis M. Bergasa

Robesafe Lab. University of Alcalá. Spain



Universidad de Alcalá

Comunidad de Madrid

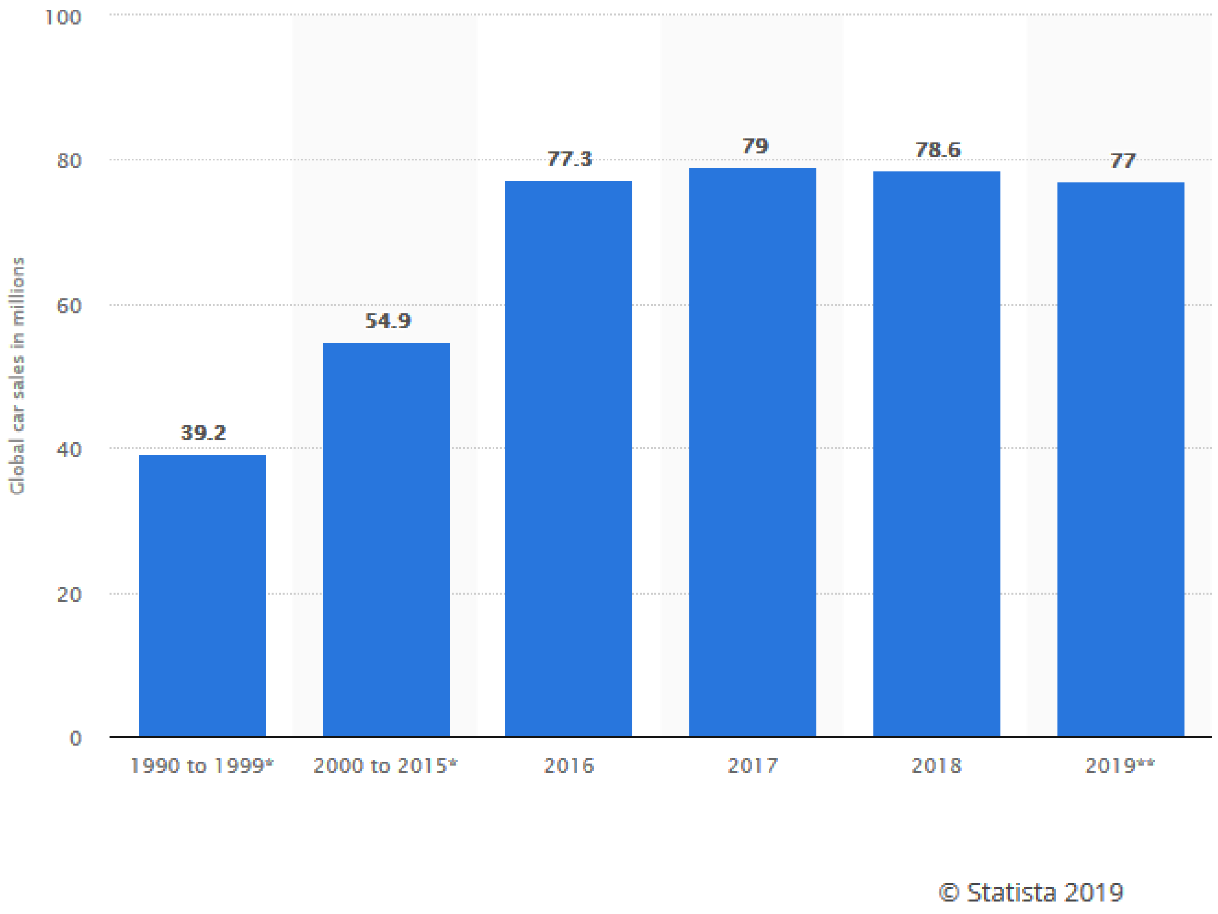
RoboCity2030.org

Madrid Robotics Digital Innovation Hub

RobeSafe

Introduction

There are millions of vehicles driving around the world



Source: https://www.statista.com/statistics/200002/international-car-sales-since-1990/

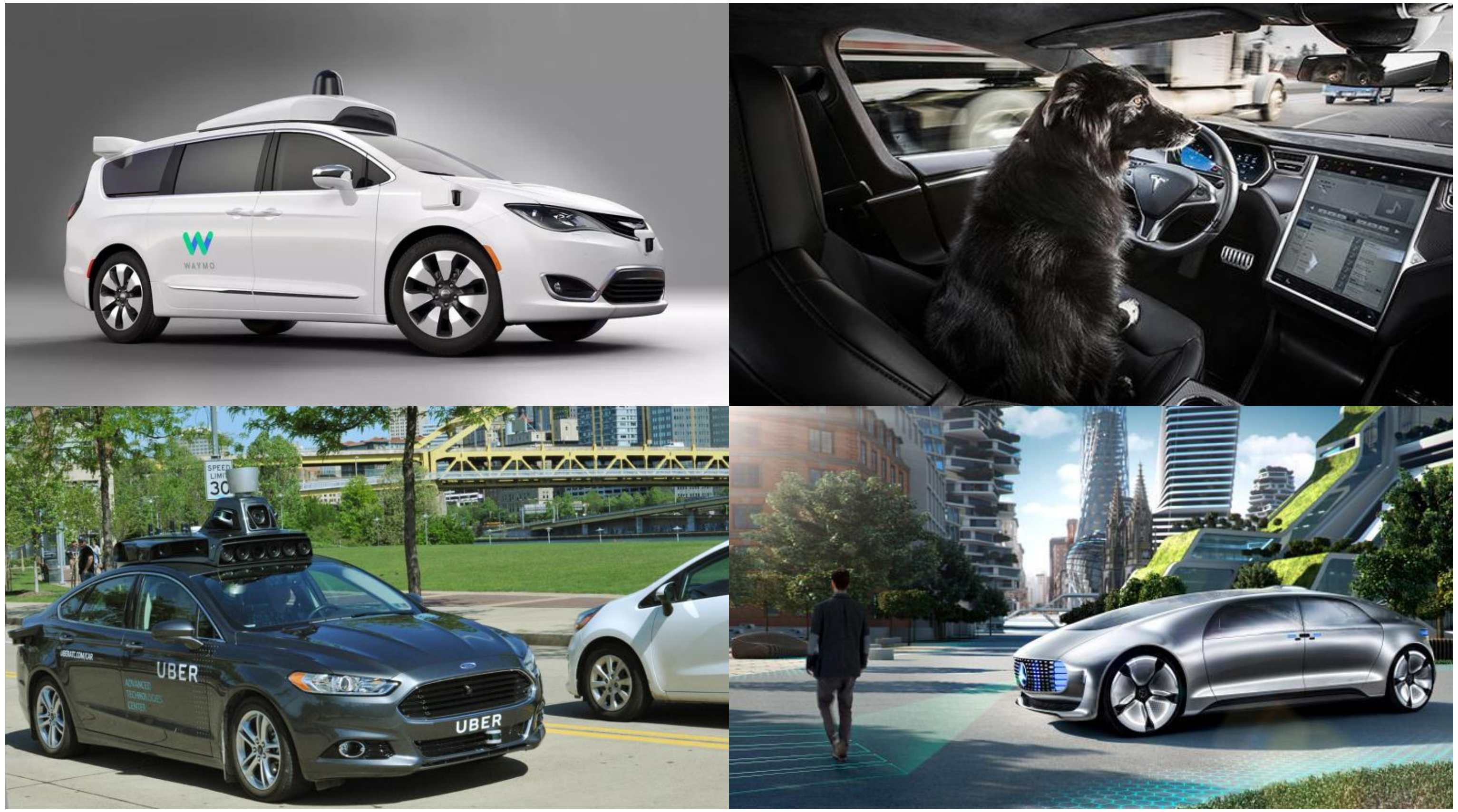
Introduction

1.3 million die every year in manual driving

Source: https://extranet.who.int/roadsafety/death-on-the-roads/

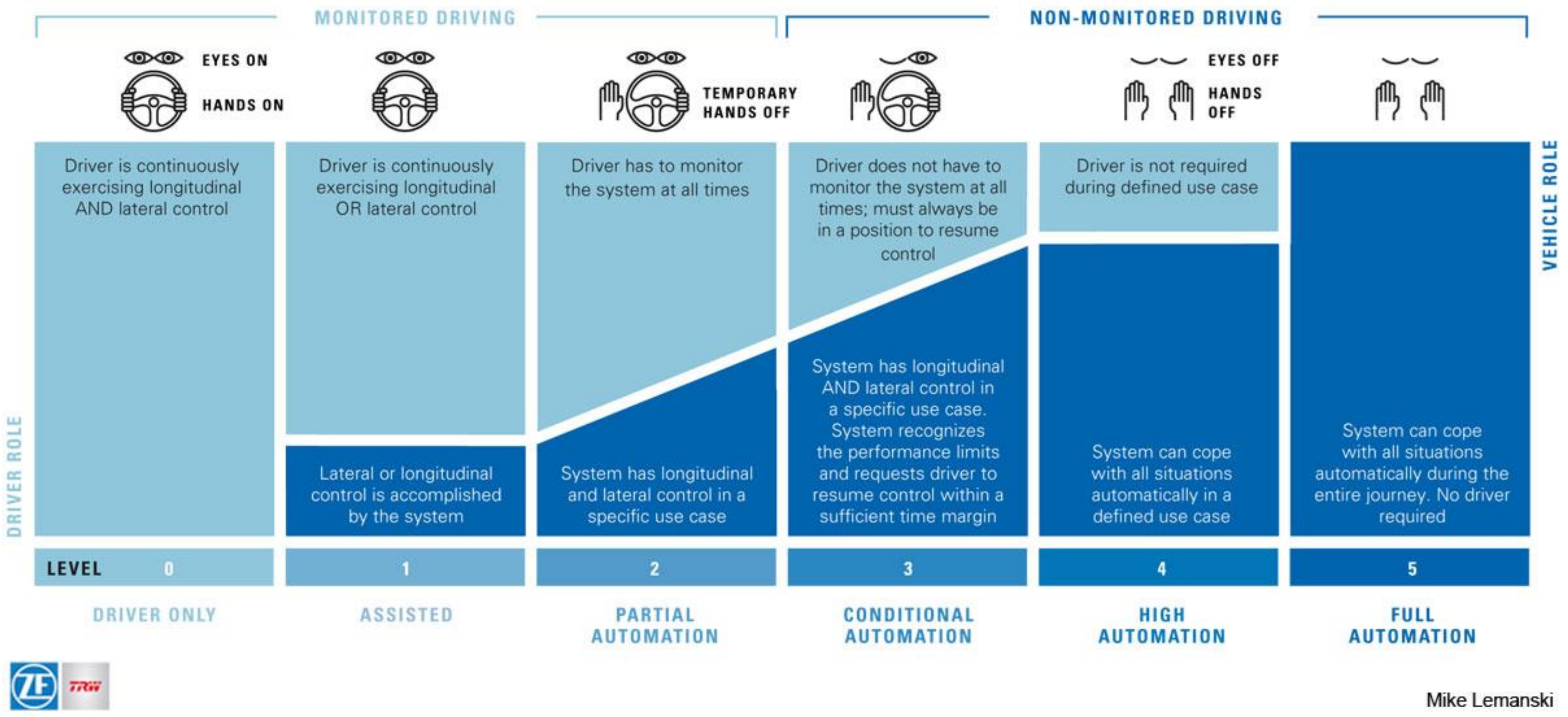
Introduction

Self-driving cars: Safe lives, increase mobility and improve efficiency

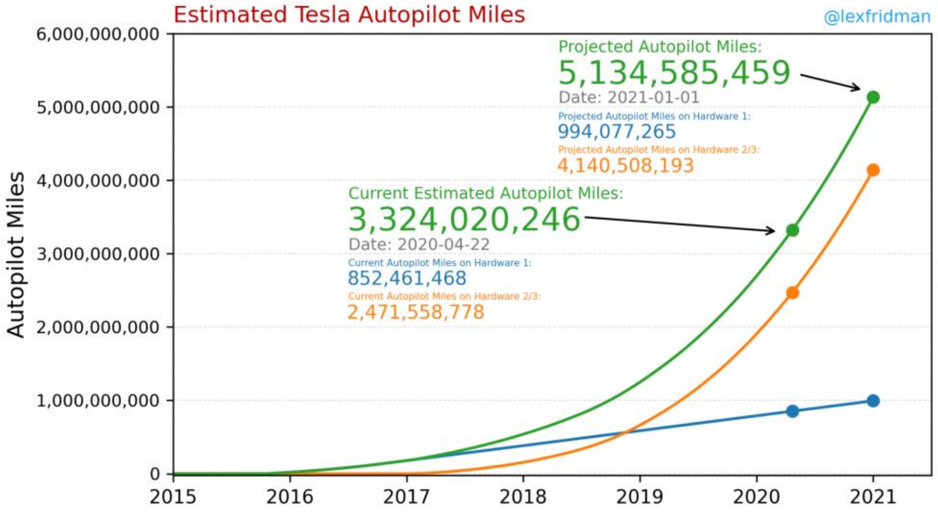


Introduction

Self-driving cars: **Levels of Automation** (SAE J3016)



Introduction



Example L2 System:

Tesla Autopilot

3+ billion miles

Example L4 System:

Waymo

20+ million miles

Introduction

Predictions about Fully Autonomous Vehicle

Elon Musk, CEO of Tesla

- **2017:** "My guess is that in probably 10 years it will be very unusual for cars to be built that are not fully autonomous.“
- **2018:** “The upcoming autonomous coast-to-coast drive will showcase a major leap forward for our self-driving technology.”

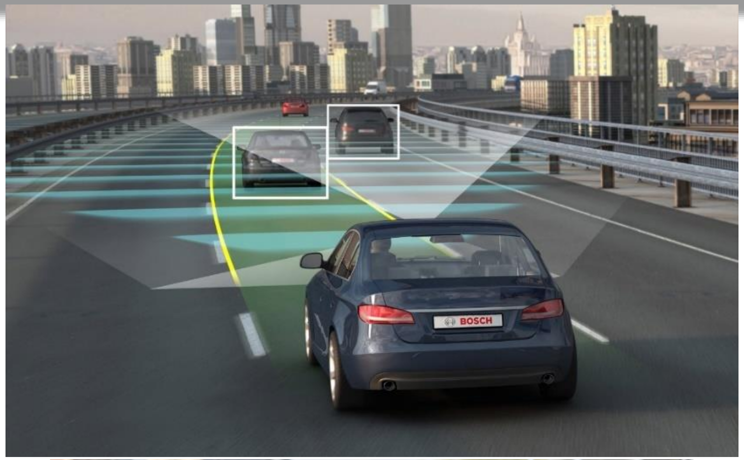
Rodney Brooks, Professor at the MIT

- **=2031:** A major city bans manually driven cars from a non-trivial portion of a city.
- **>2045:** Majority of US cities ban manually driven cars with drivers from a non-trivial portion of a city.

Introduction

Problem: **Complexity**

- Varied environments
- Diverse conditions
- With other traffic agents
- Following “human” laws



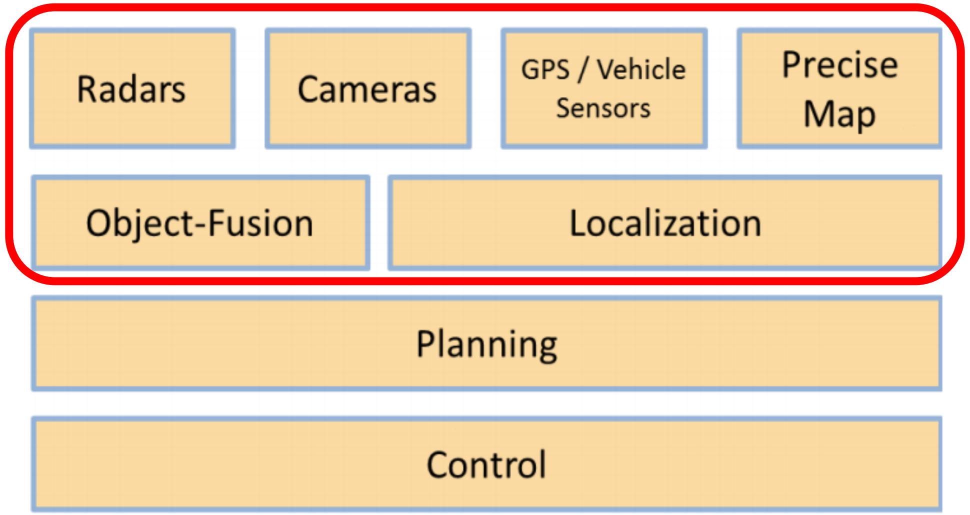
- **Pipeline & Modules of a Self-driving Car**

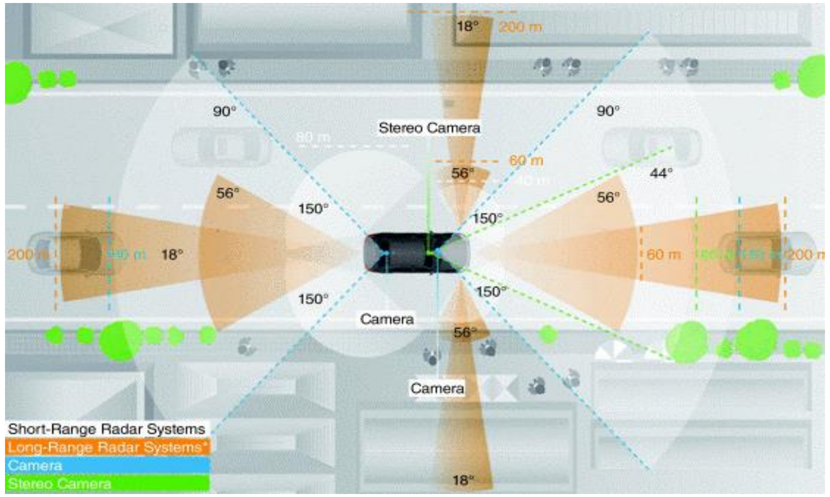
Mapping → Perception → Localization → Planning → Control

↑

Decision Making

Perception

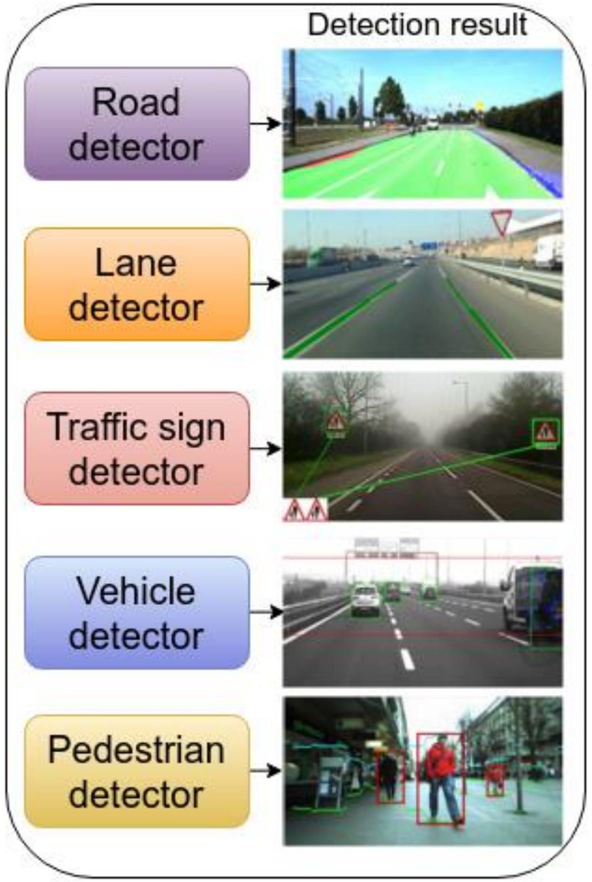


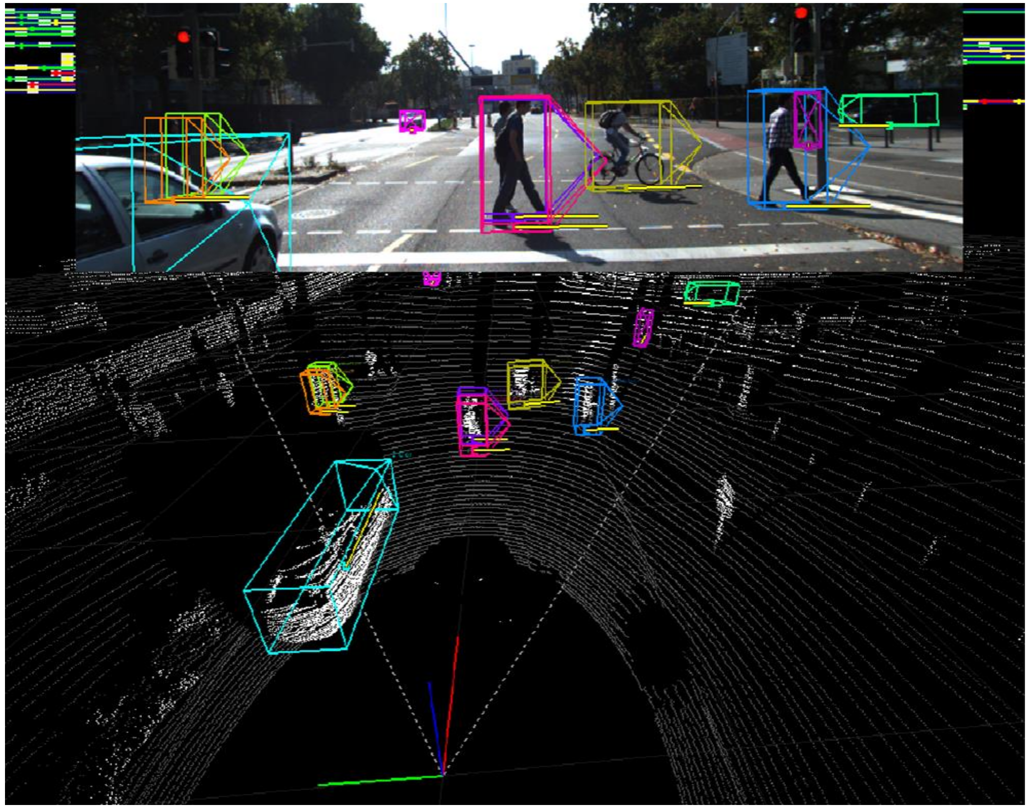


- Perception: multiple sensors for understanding both **static** and **dynamic** environments

Perception

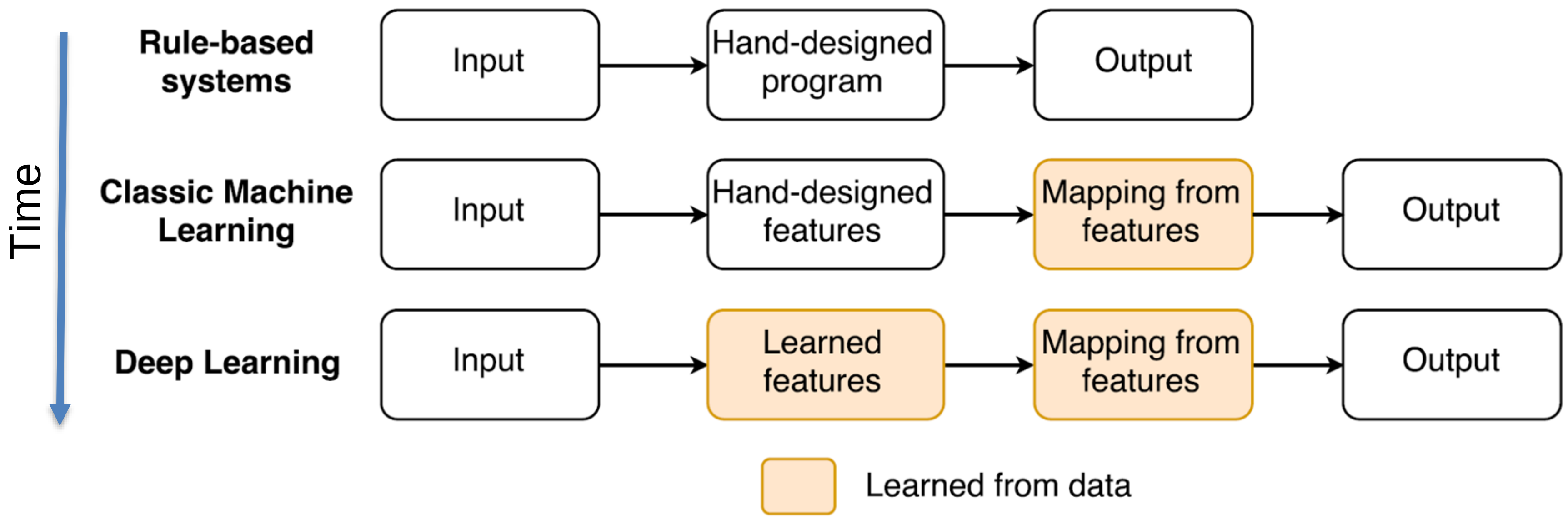
From many years ago **methods** for **scene understanding** have been **hand-designed** and based in models





Perception

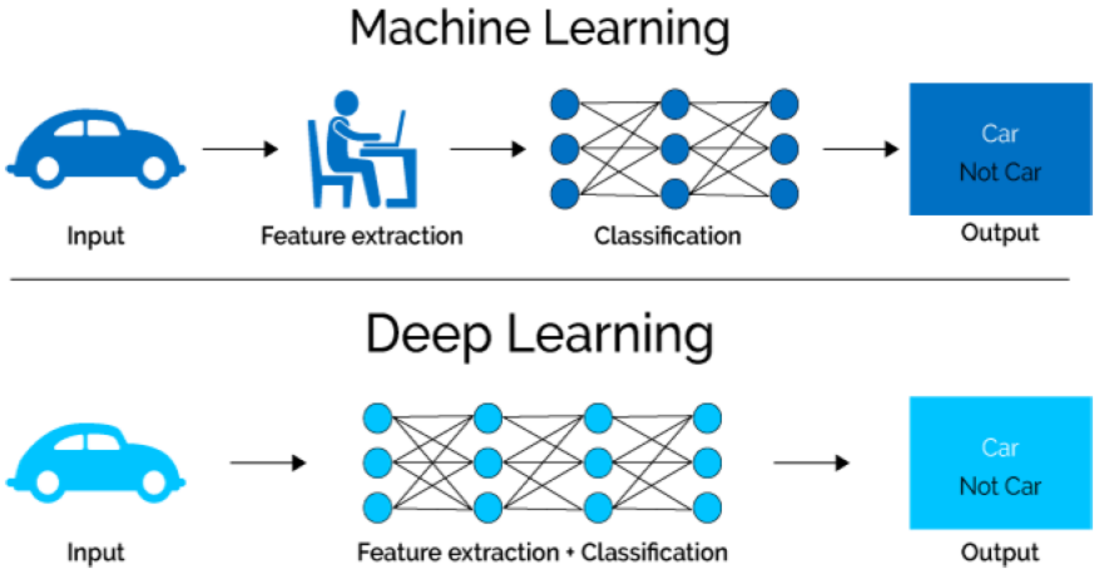
3 approaches in time:

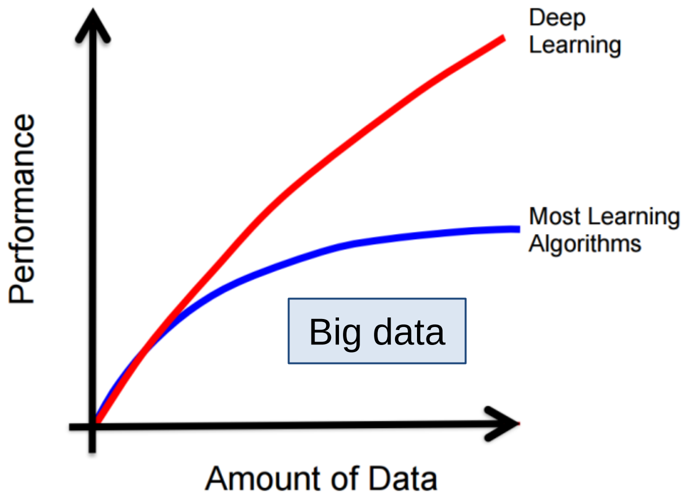


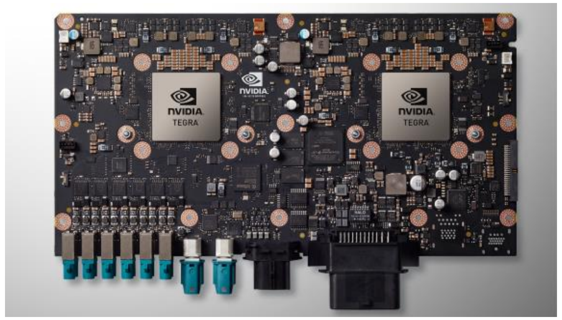
From **Model-driven** to **Data-driven** algorithms

Deep Learning

Why Deep Learning? Scalable Machine Learning and **Parallelizable** processing







Specific HW
CPU -> GPU -> TPU

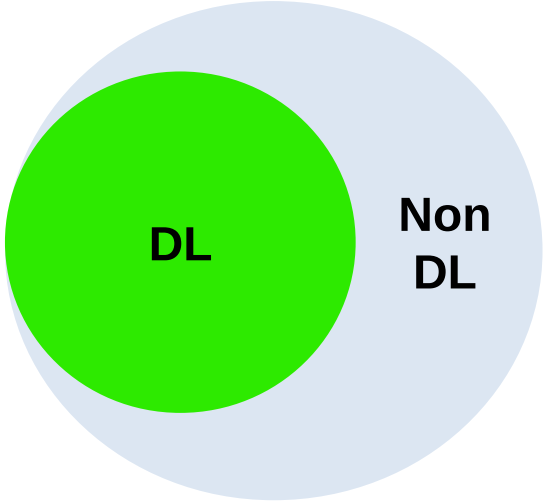
Deep Learning

Self-driving real cars difficulties

- Scene understanding is hard
- Sometimes DL has unintended consequences
- Real-time and consumption constrains
- Safety is “the first”

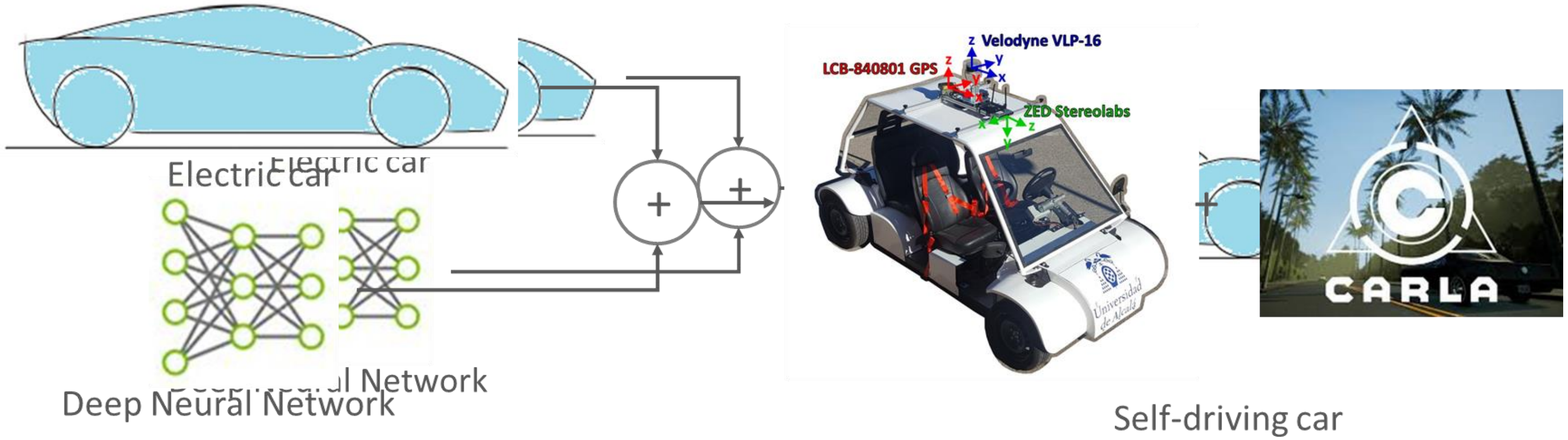
Automakers have started to use DL with caution

- Factories for collecting data for end-to-end learning
- Hybrid-systems
  - When DL is not confident or consistent the system leverages expert domain knowledge



Our approach

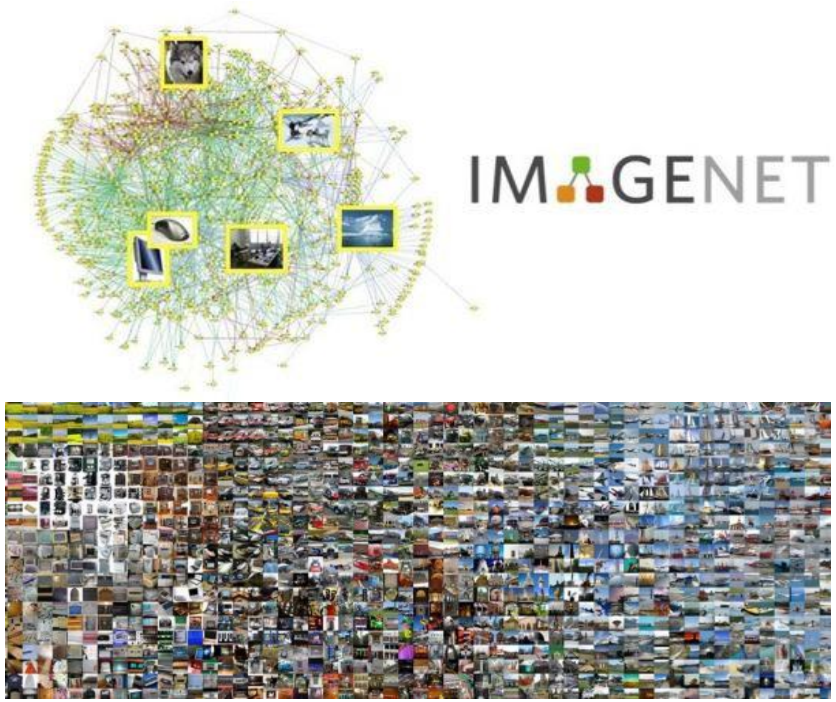
To contribute as university researchers **to apply DL techniques to the intelligent vehicles**

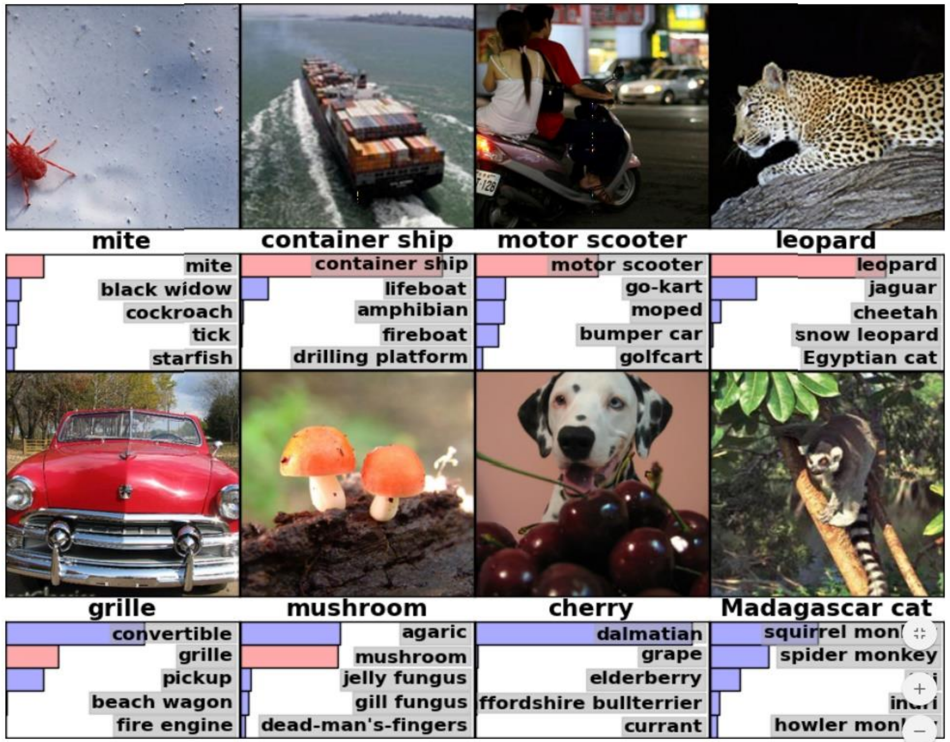


- **Goals**:
  - Take advantage of the open-source datasets, frameworks, DL models and simulators
  - Run current models and contribute with new ones
  - Test our contributions in simulation and in a real car

Semantic Segmentation

- We started in 2014 with the boom of ImageNet for image classification:

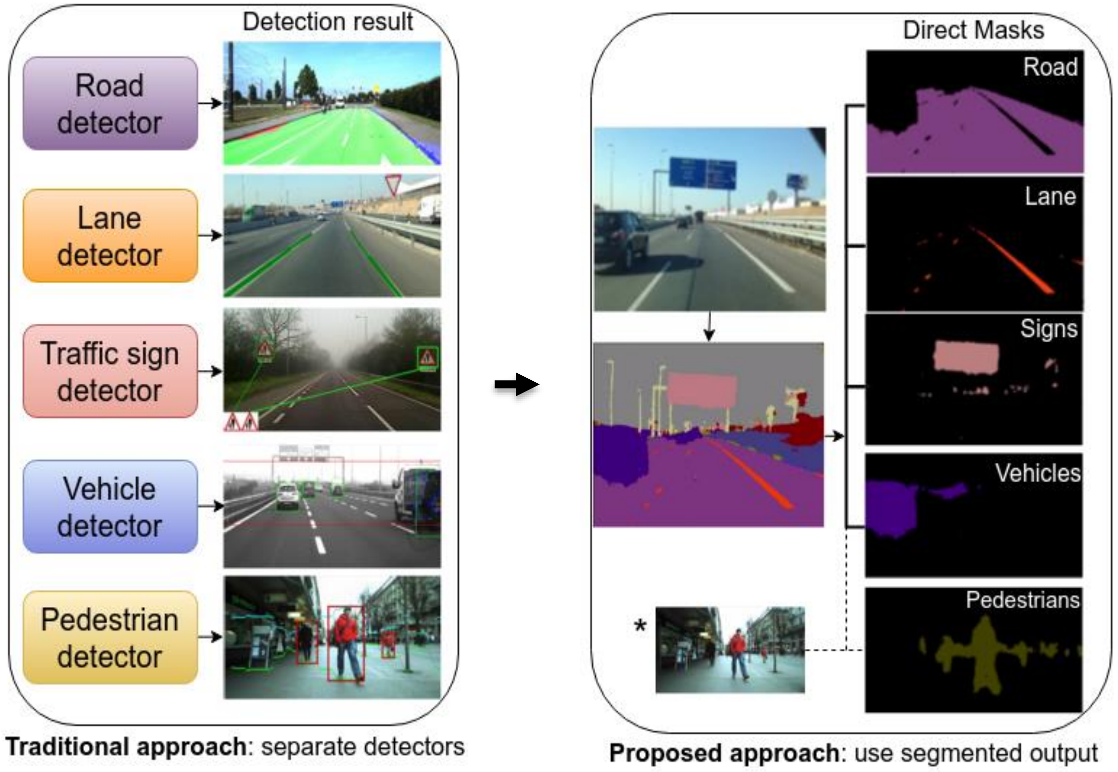




- Known architectures: AlexNet, Inception, VGG, ResNet, …
- Results highly overcame traditional methods

Semantic Segmentation

- But soon jumped to SS as a way to unify perception task in self-driving cars (**holistic approach**)



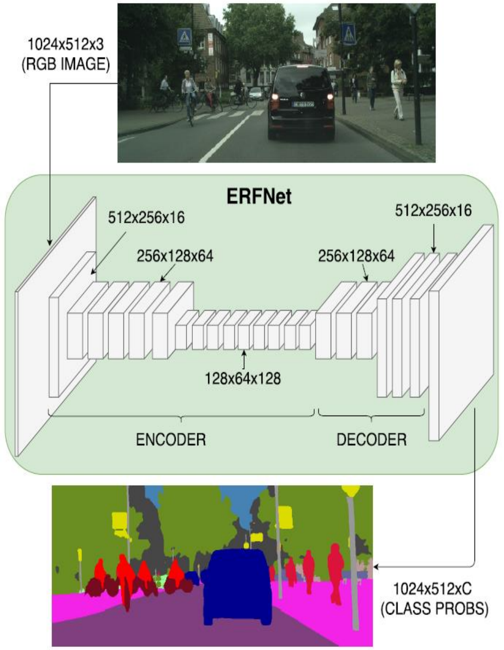
Classifying object categories at the pixel-level

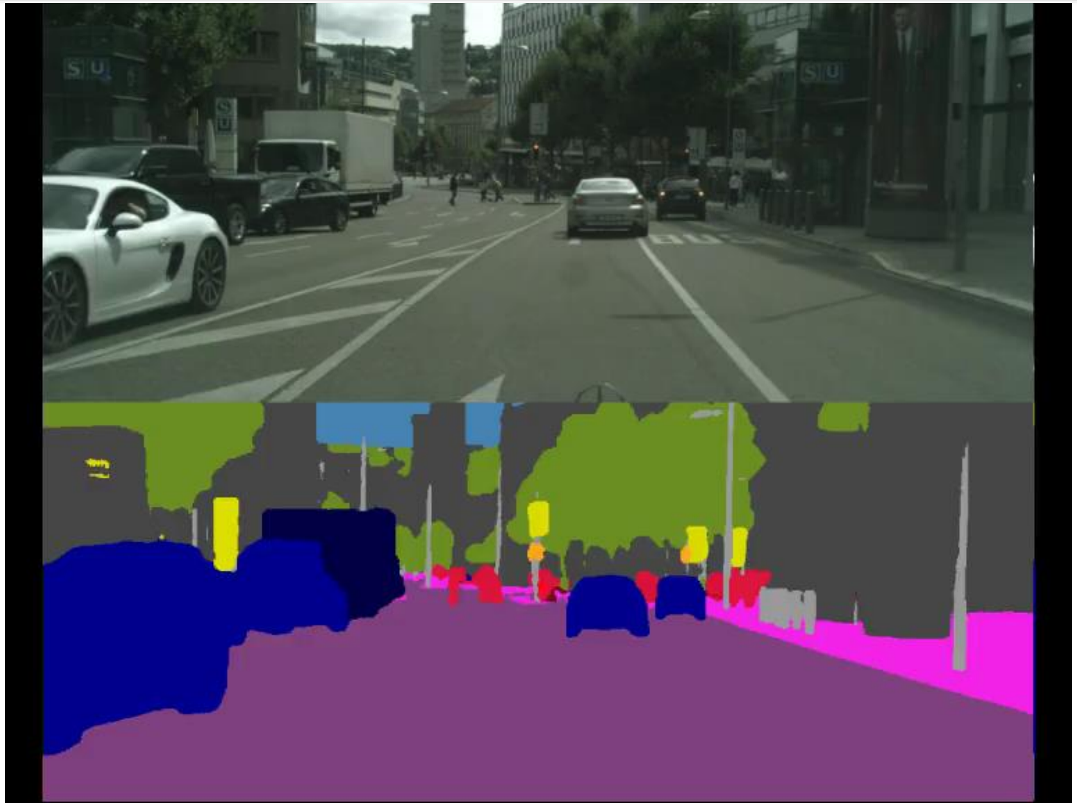
"*Can We Unify Monocular Detectors for Autonomous Driving by using the Pixel-Wise Semantic Segmentation of CNNs?*", E. Romera, L. M. Bergasa and R. Arroyo, in **IEEE IV-WS 2016**

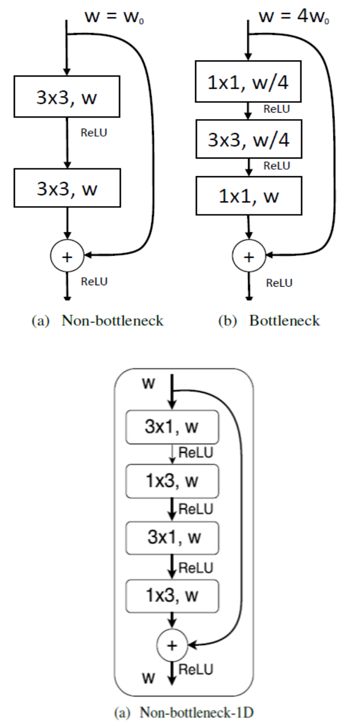
Semantic Segmentation: ERFNet

- **SS problems in 2016**:
  - Current architectures were very costly to process
  - Important concern in IV → Computational resources
- We proposed an architecture:
  - **Accurate**: similar accuracy to top-performing nets
  - **Efficient**: low processing times / memory
- **ERFNet: Efficient Residual Factorized Network**
- **Encoder-decoder net** that combines:
  - Residual layers
  - Filter factorization (1D kernels)

Semantic Segmentation: ERFNet







[Romera et al, 2017] "*Efficient ConvNet for Real-time Semantic Segmentation",* E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, **IEEE Intelligent Vehicles Symposium (IV) 2017,** pp. 1789-1794, Redondo Beach (USA), June 2017. **[Best Student Paper Award]**

[Romera et al, 2018] "ERFNet: *Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation",* E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, **IEEE Transactions on Intelligent Transportation Systems (T-ITS), January 2018. [IEEE T ITS Main publication (GSM)]**

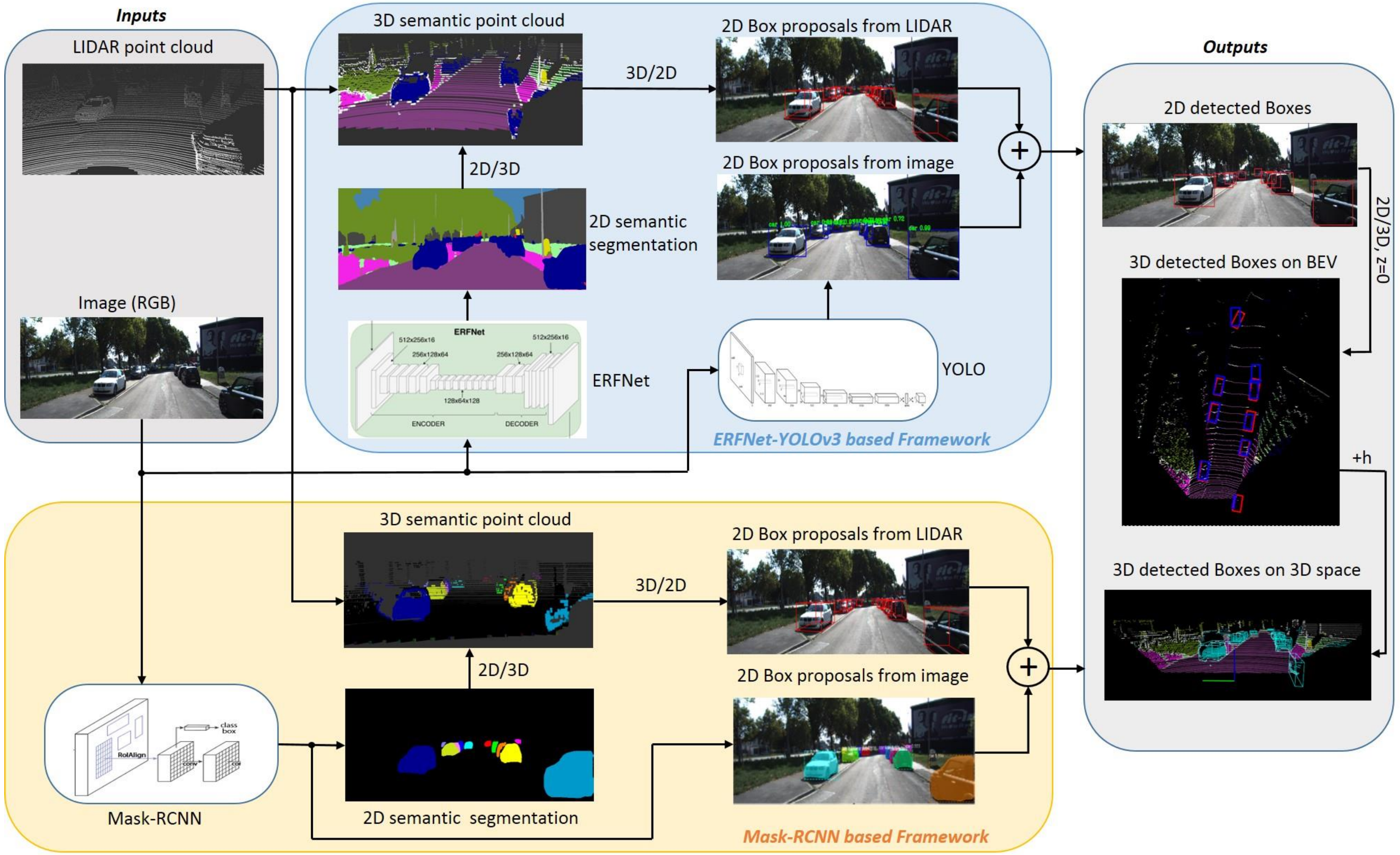
GitHub repository: https://github.com/Eromera/erfnet

3D Semantic Segmentation

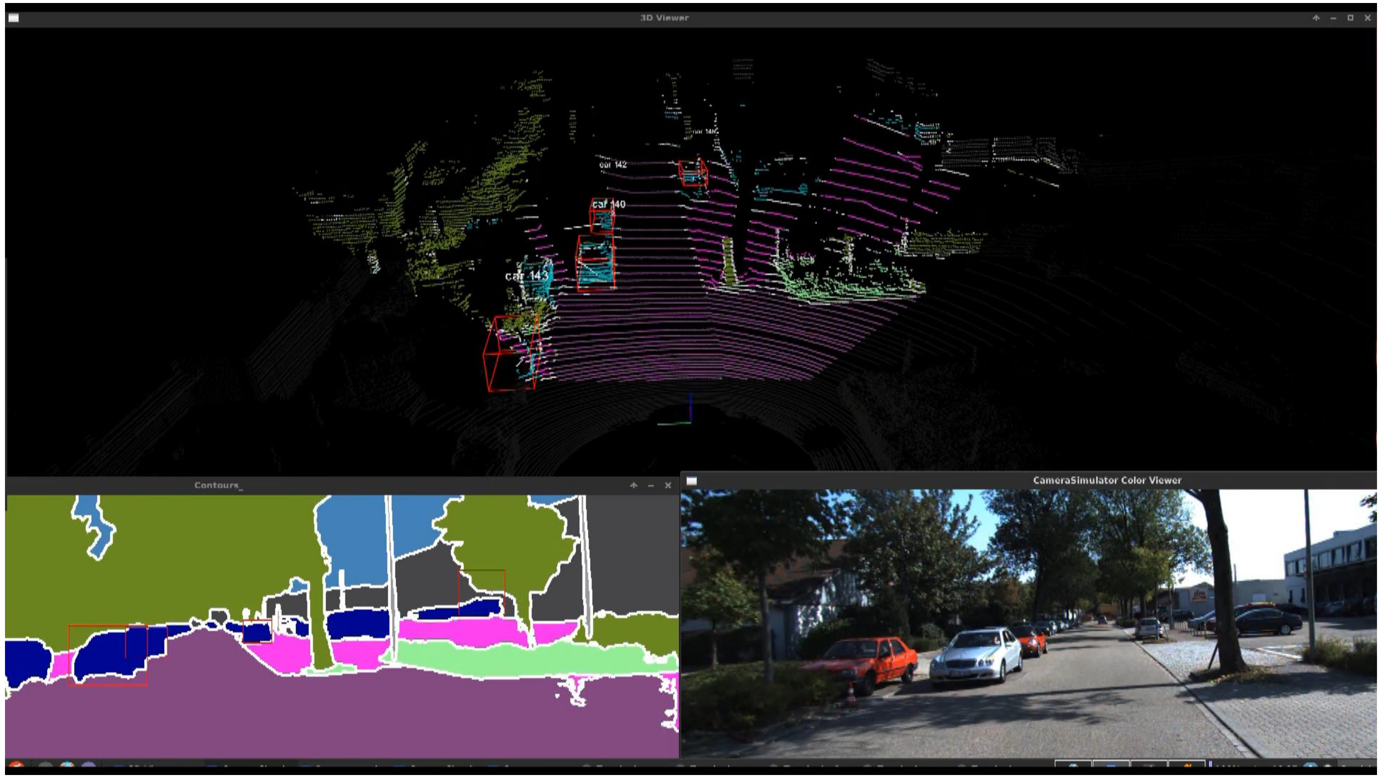
- **2D SS lacks space awareness**
- **3D information is needed** to map surrounding objects to navigate safely
- 3 options to **transfer 2D SS to 3D:**
  - Bird's view
  - Stereo Depth
  - **Geometric 2D/3D projection.** Fusion with LiDAR
- Tested in our autonomous vehicle
  - ZED camera: 1920x1080 @ 60fps via USB3

3D Multi-Object Detection

3. Fusion camera & LiDAR -> 3D Multi-Object Detection



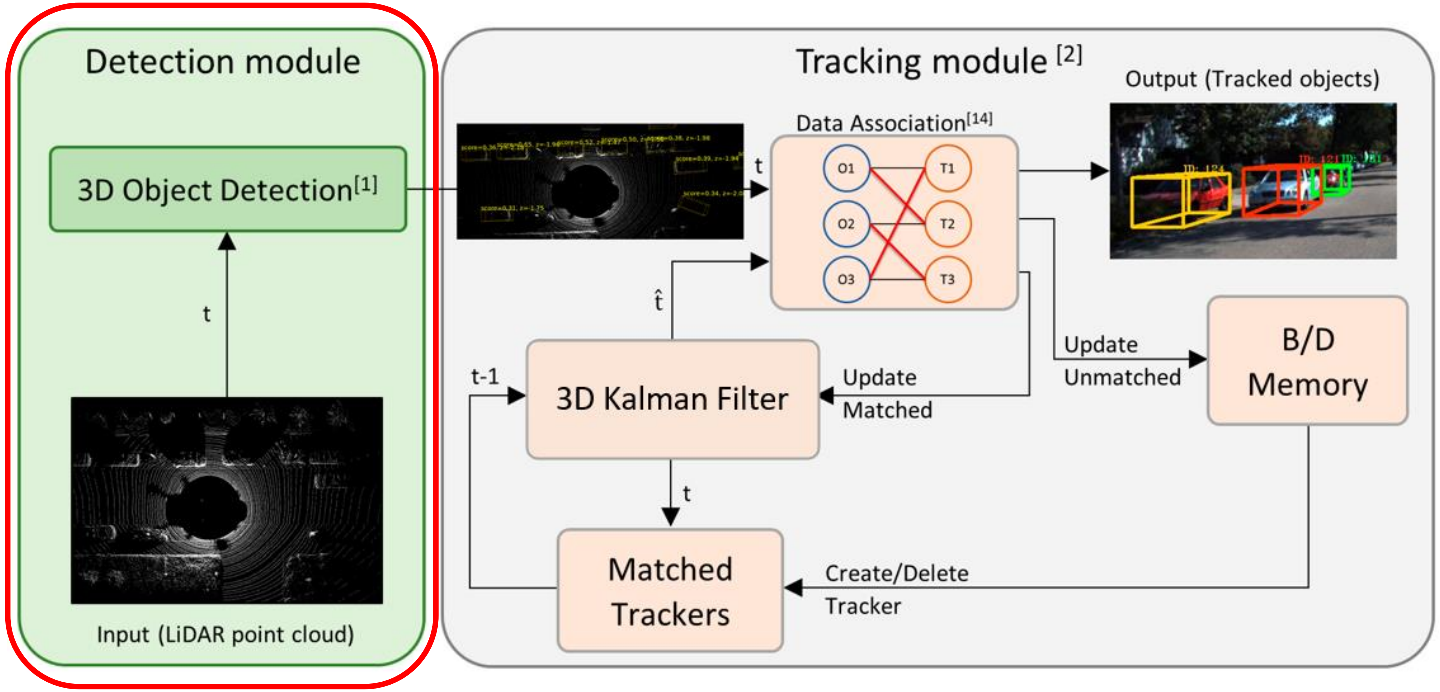
3D Multi-Object Detection



[Barea et al, 2018] “Vehicle Detection and Localization using 3D LIDAR Point Cloud and Image Semantic Segmentation”, *R. Barea, C. Pérez-de-Rivas, L.M. Bergasa, E. López-Guillén, E. Romera, E. Molinos, M. Ocaña, J. López,* **IEEE Conference on Intelligent Transportation Systems (ITSC 2018),** pp. 3481-3486, Maui, Hawaii, USA, November 2018.

[Barea et al, 2019] “Integrating well-known CNNs for Multi-Sensor 3D Vehicle Detection in Real Autonomous Driving Environments”, *R. Barea, L.M. Bergasa, E. Romera, E. López-Guillén, O. Pérez, M. Tradacete, J. López,* **IEEE Conference on Intelligent Transportation Systems (ITSC 2019),** pp.1425- 1431, Auckland, New Zealand, October 2019.

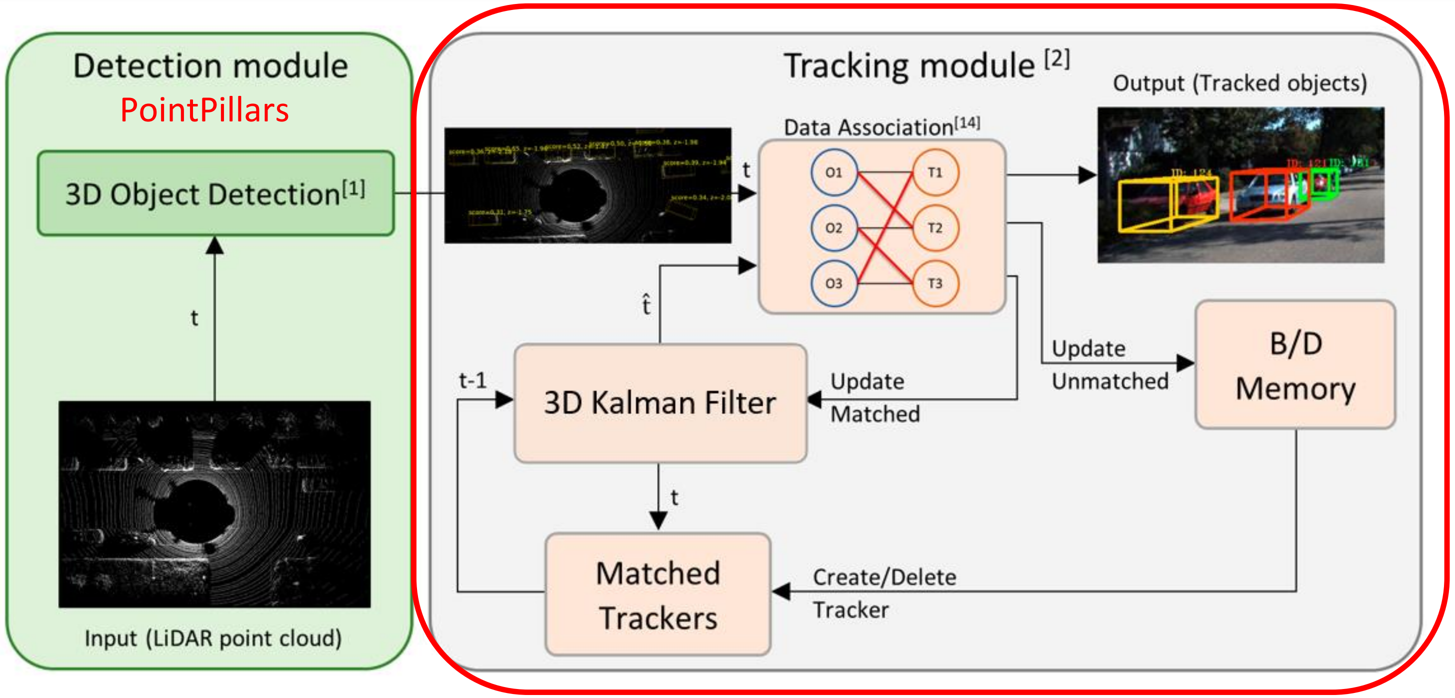
3D Multi-Object Detection and Tracking



3D Object Detection results in the KITTI validation set

Method	Type of input	Frequency (Hz)	Car (AP)	Pedestrian (AP)	Cyclist (AP)
PointPillars Lang et al. (2019)	Voxel based	**41.7**	**86.46**	**57.75**	**80.057**
SECOND Yan et al. (2018)		19.8	88.61	56.55	80.59
PointRCNN Shi et al. (2018)	Point based	6.3	88.94	61.89	85.01
PointRCNN-IoU Shi et al. (2018)		6.3	89.01	62.69	87.48
Part-A^2-Free Shi et al. (2020b)		5.6	89.12	**70.31**	87.65
Part-A^2-Anchor Shi et al. (2020b)		7.5	89.56	65.69	85.50
PV-RCNN Shi et al. (2020a)	Combination	4.6	**90.35**	63.12	**88.34**

3D Multi-Object Detection and Tracking



Tracking based on AB3DMOT (online version)

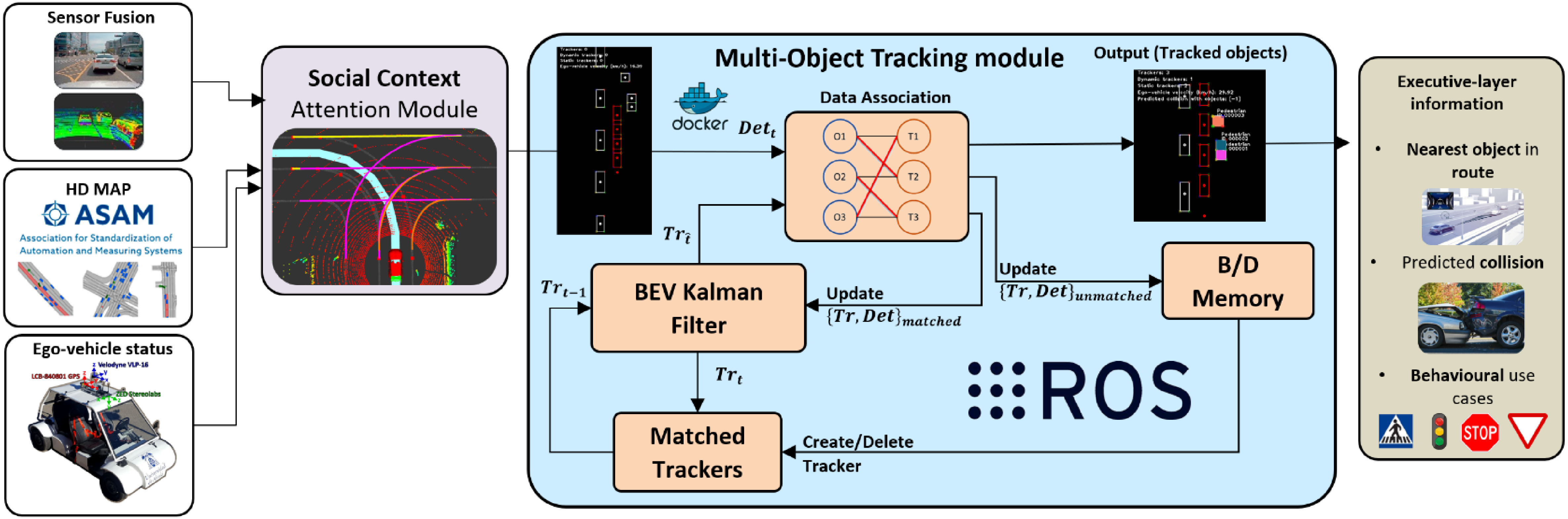
- **BEV transformation** of the detected bounding boxes
- **3D Kalman Filter** predicts the state of trajectories in BEV
- The detections at frame t and predicted trajectories are matched using the **Hungarian algorithm**
- Matched trajectories are updated to obtain **update trajectories** at frame t
- **Unmatched trajectories** are used to delete disappeared trajectories (Death) or create new ones (Birth)
- **Updated trackers** (matched + births) are passed to the following frame

3D Multi-Object Detection and Tracking

[Gómez-Huelamo et al, 2021] "360º Real-Time and Power-efficient 3D DAMOT for Autonomous Driving applications", Carlos Gómez-Huélamo, Javier Del Egido, Luis M. Bergasa, Rafael Barea, Elena López-Guillén, Miguel E. Ortíz-Humani, Miguel Antunes, in **Multimedia Tools and Applications**. 2021. In revision.

[Egido et al, 2020] "360º Real-Time 3D Multi-Object Tracking validated in KITTI dataset and CARLA simulator", Javier Del Egido, Carlos Gómez-Huélamo, Luis M. Bergasa, Rafael Barea, Elena López-Guillén, Javier Araluce, Rodrigo Gutiérrez-Moreno, Miguel Antunes, in **Workshop of Physical Agents** (**WAF**). Alcalá de Henares, Spain, November 2020

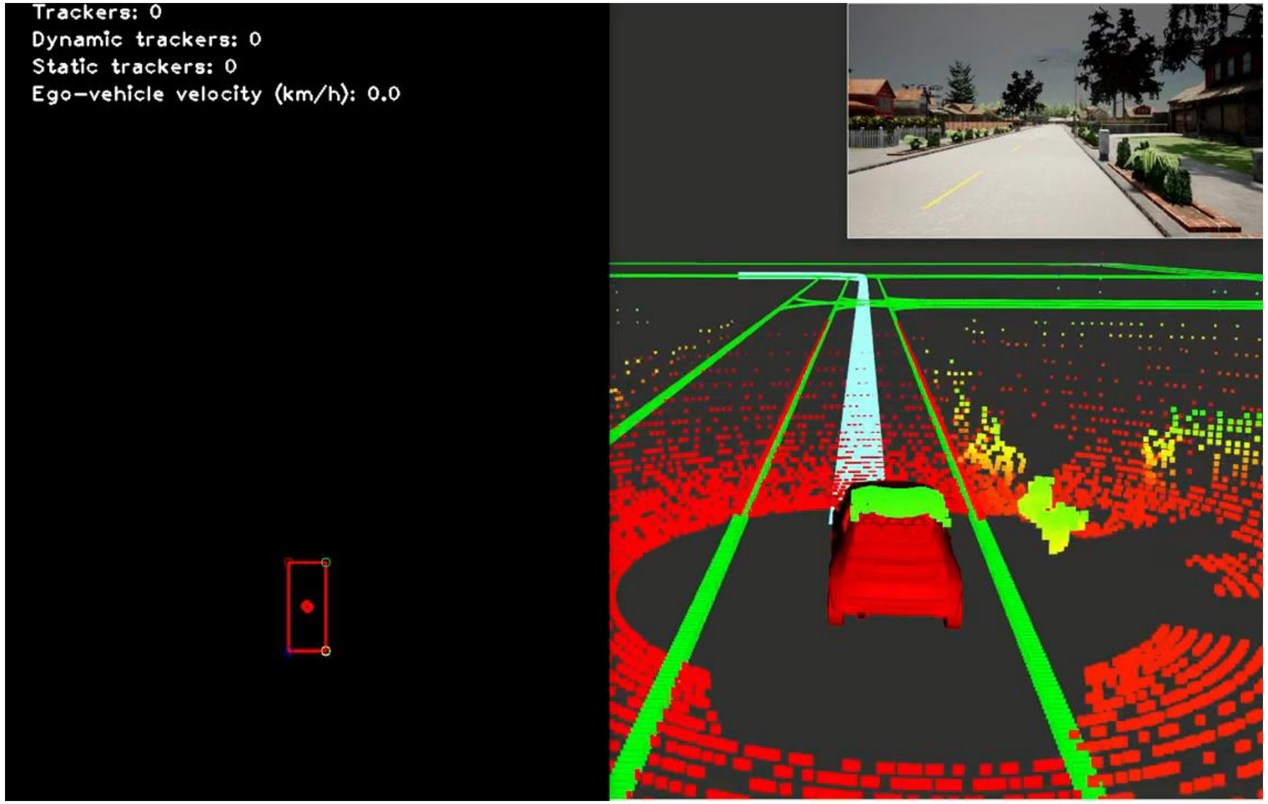
Behavior Prediction: SmartMOT



SmartMOT: Behaviour prediction in multi-agent and dynamic environments

- Semantic information of **HDmap** and **ego-vehicle status** is added
- **Social Context Attention Module** (SCAM) fed by Sensors, HDmap (lanes of interest) the ego-vehicle status (velocity)
- **MOT** module **the same** that the previous one
- **Output**: **predicted collisions** and **actions** to be taken

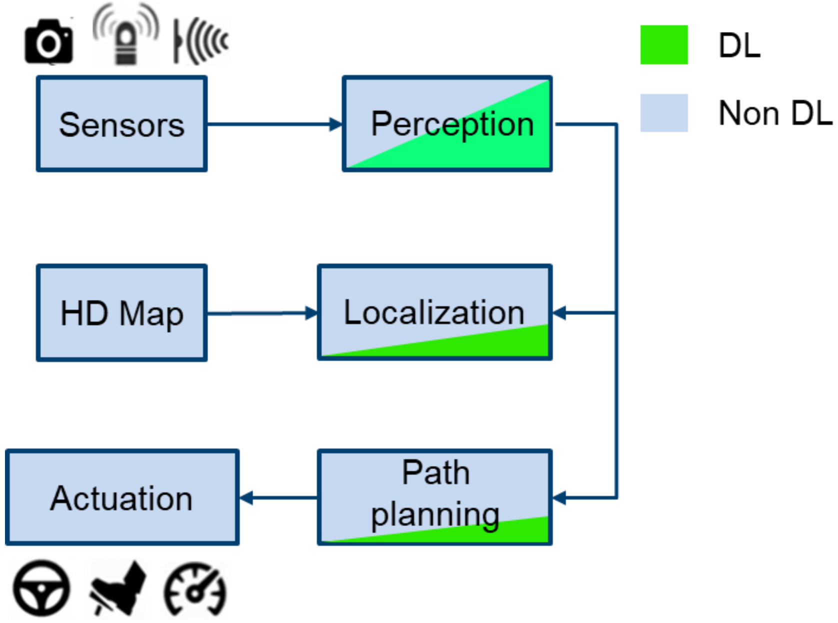
Behavior Prediction: SmartMOT



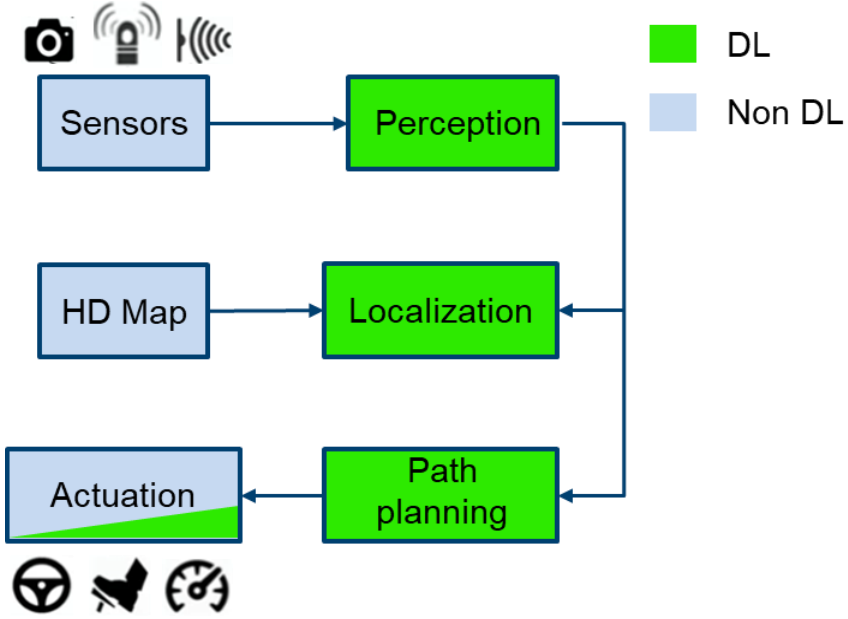
[Gómez-Huélamo et al, 2021] "SmartMOT: Exploiting the fusion of HD Maps and Multi-Object Tracking for Real-Time Motion Prediction in Intelligent Vehicles applications", Carlos Gómez-Huélamo, Luis M. Bergasa, Rodrigo Gutiérrez, Felipe Arango, Alejandro Díaz, in **IEEE Intelligent Vehicles Symposium** (**IV**), Nagoya, Japan, July 2021. Accepted for publication

[Gutierrez et al, 2021] "Validation Method of a Self-Driving Architecture for Unexpected Pedestrian Scenario in CARLA Simulator", Rodrigo Gutiérrez, Felipe Arango, Carlos Gómez-Huélamo, Luis M. Bergasa, Rafael Barea, Javier Araluce, in **IEEE Intelligent Vehicles Symposium** (**IV**), Nagoya, Japan, July 2021. Accepted for publication

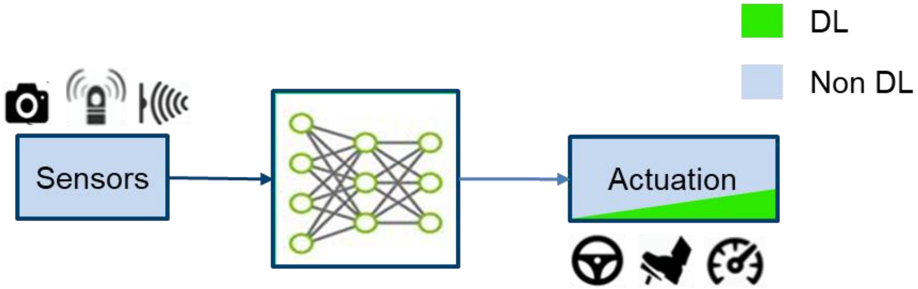
DL & Self-driving Challenges



Classical Navigation architecture

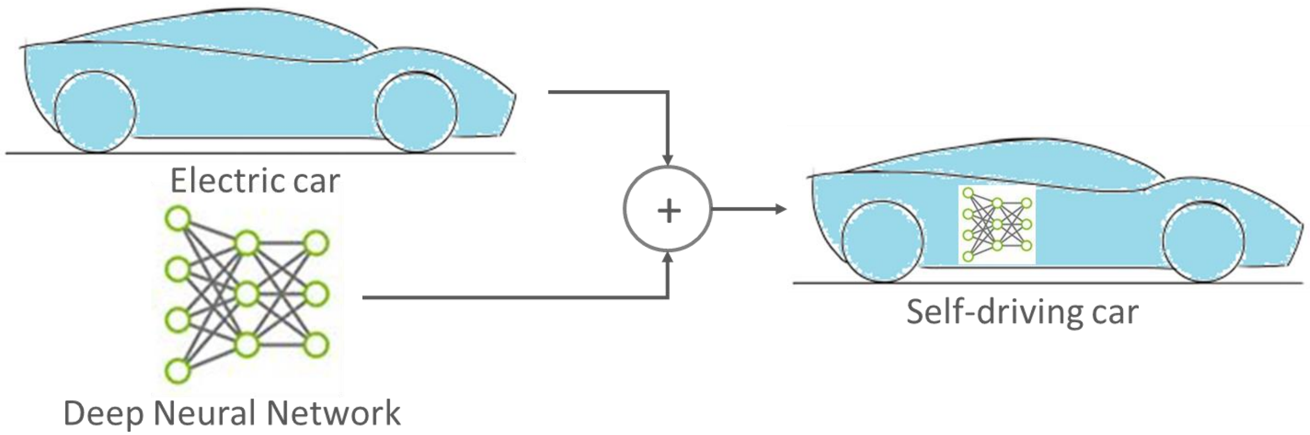


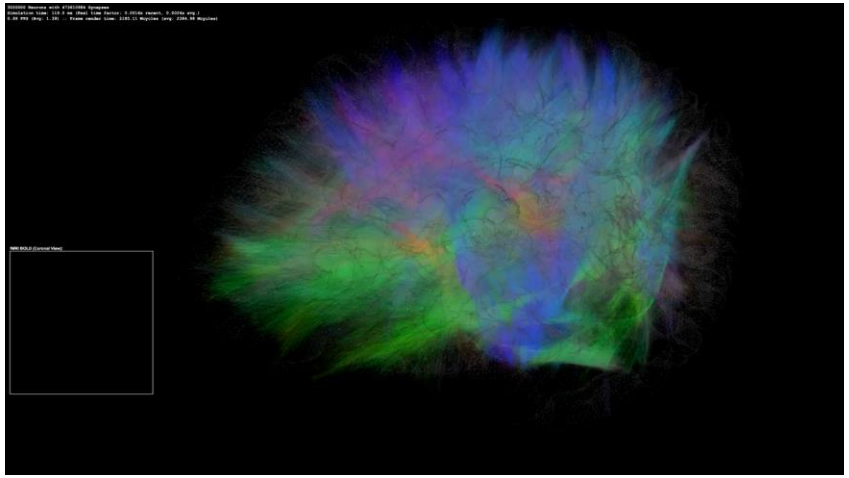
Advanced Navigation architecture



End-to-End Deep Learning Navigation architecture

DL & Self-driving Challenges





"The future of the DL depends on some graduate student who is deeply suspicious of everything I have said."

Geoffrey Hinton
"Godfather of Deep Learning"

THANKS!

luism.bergasa@uah.es

www.robesafe.uah.es/personal/bergasa/

Universidad de Alcalá

Comunidad de Madrid

UNIÓN EUROPEA Fondos Estructurales Invertimos en su futuro

RoboCity2030.org

Madrid Robotics Digital Innovation Hub

RobeSafe