From Blips to Bounding Boxes: The Evolution of Automotive Radar Perception

Michael Ulrich, Corporate Research, Robert Bosch GmbH

130
km/h
km/h
BOSCH

What is Automotive Radar Perception?

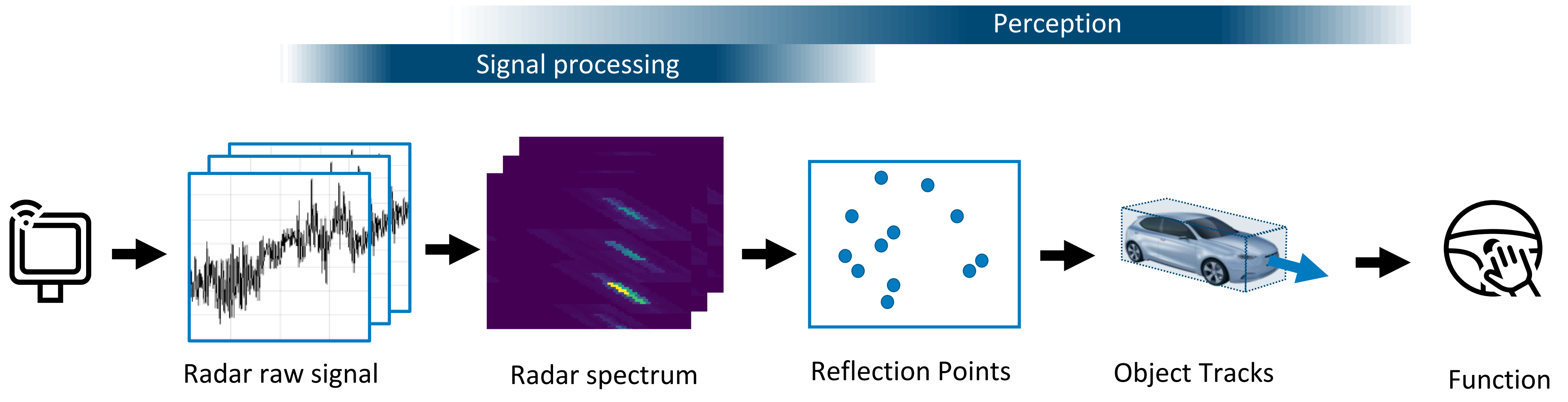


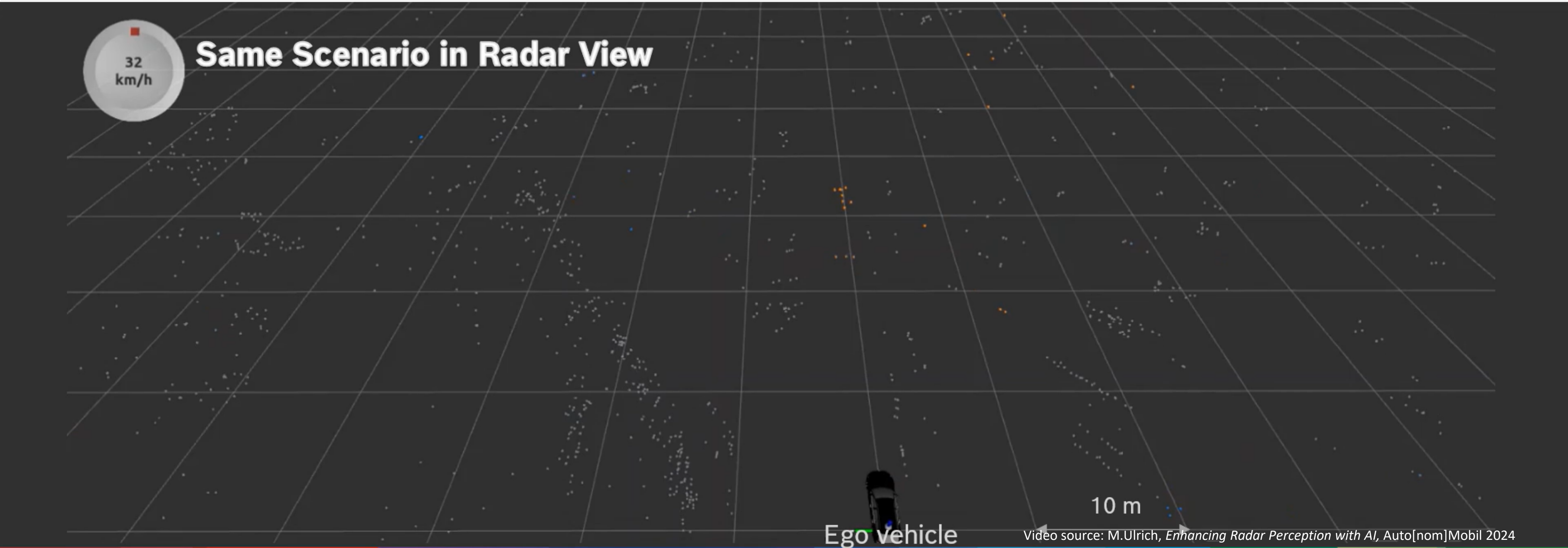
Image source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

WHAT IS GOING ON?

Video source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

Camera View

Radar View



Video source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

Objects

Video source: https://www.ambarella.com/products/automotive-oculii/ https://youtu.be/rCRyoARrcJ4

Ambarella™

OCULii™

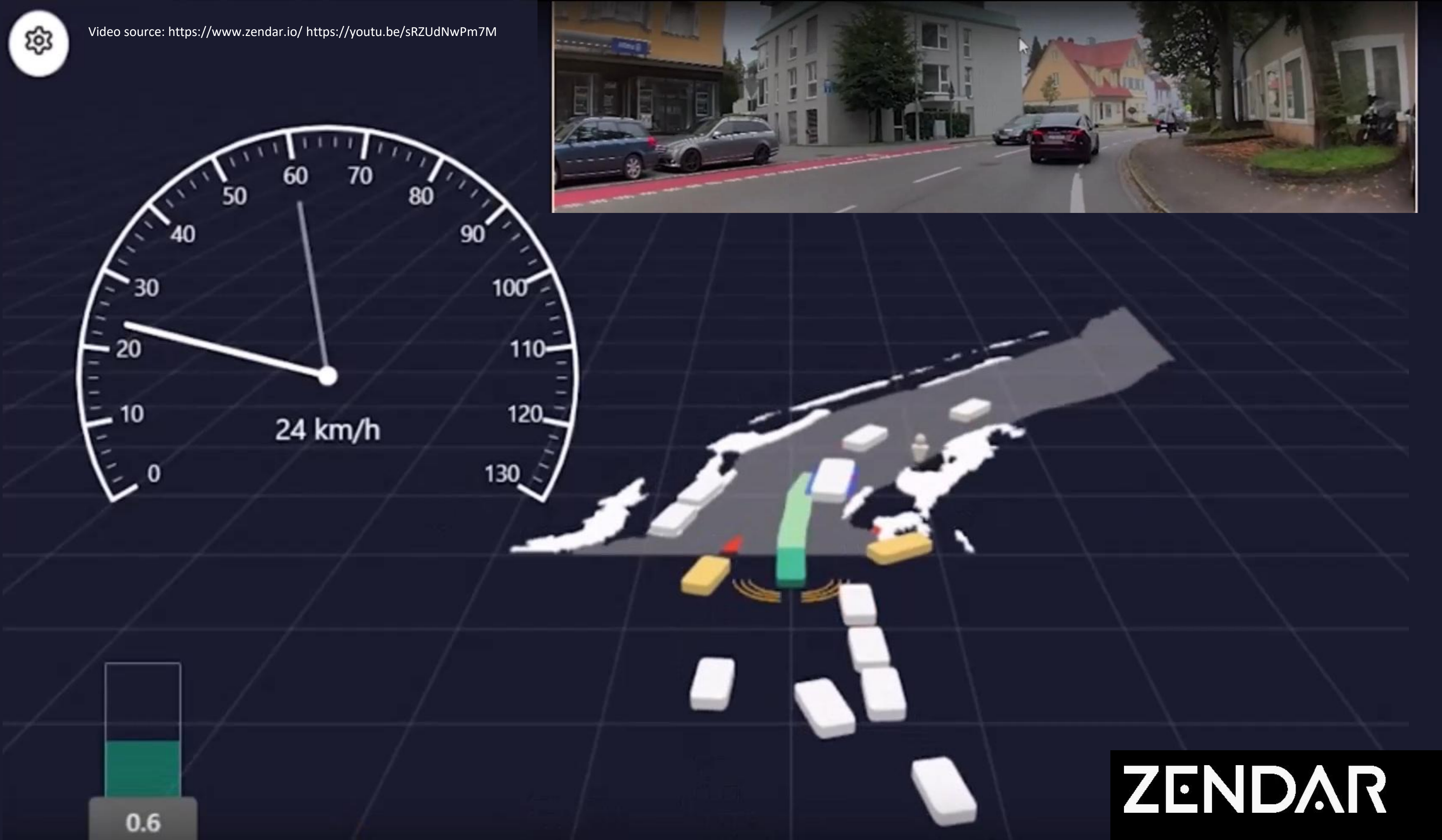
Freespace

Video source: https://youtu.be/t2DfMe9XmPQ

PERCIV.AI

BOSCH

Video source: https://www.zendar.io/ https://youtu.be/sRZUdNwPm7M



What is the difference between radar and lidar?

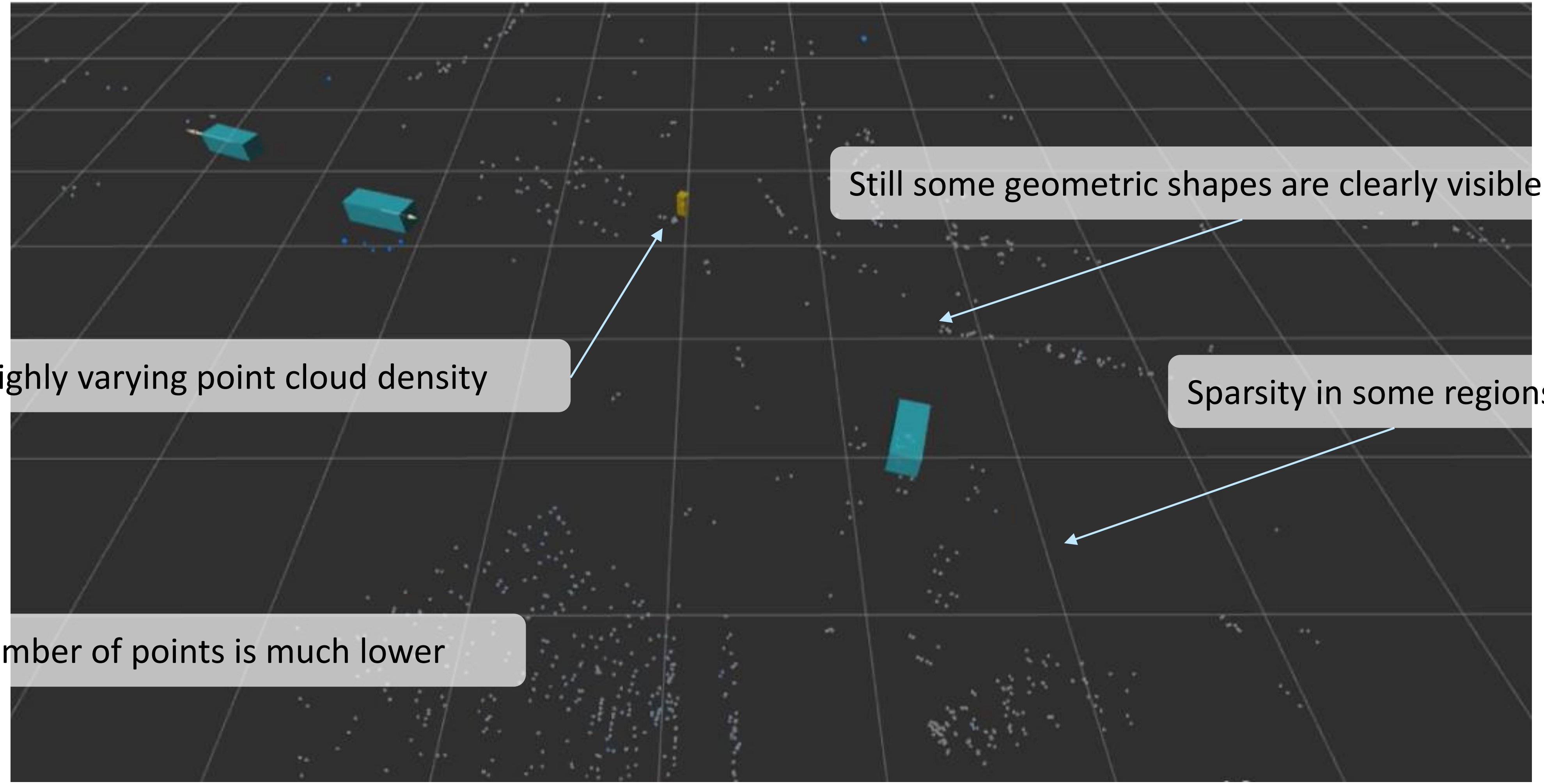
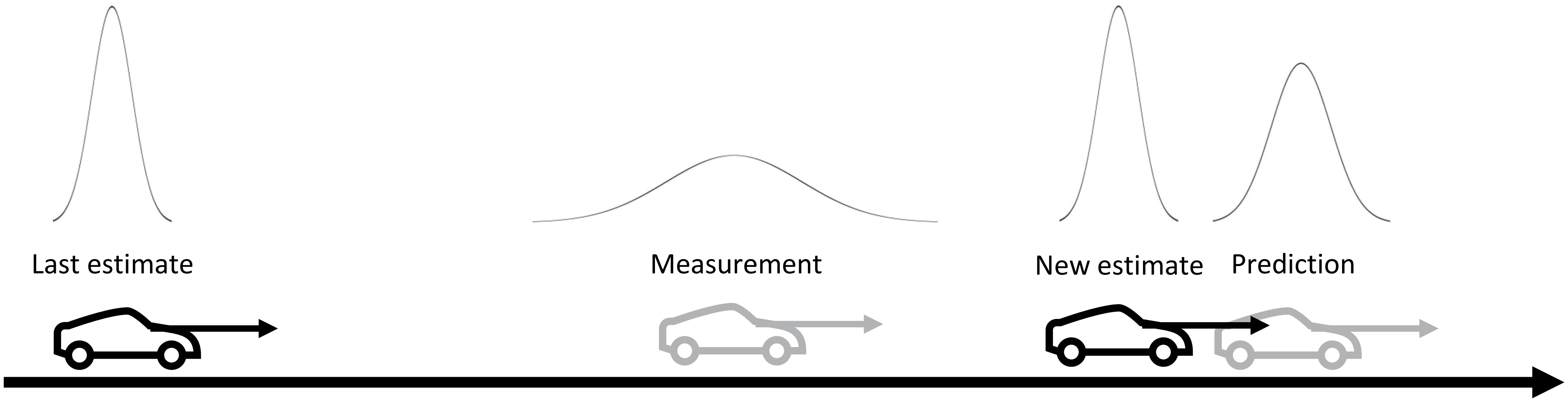


Image source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

The Classical Era:
Taming the Noise with Physics and Filtering

The Core Engine: Kalman Filter Tracking



Kalman Filter: Fuse predicitons with noisy measurements

The Core Engine: Kalman Filter Tracking

Example:
Constant velocity model

$$\underline{\hat{x}} = \begin{bmatrix} p \\ v \end{bmatrix}$$

$$\hat{P}$$

2. Update

$$\underline{\hat{x}} = \underline{\tilde{x}} + K\underline{y}$$

$$\underline{y} = \underline{z} - H\underline{\tilde{x}}$$

$$K = \tilde{P}H^T(H \cdot \tilde{P} \cdot H^T + R)^{-1}$$

Measurement noise

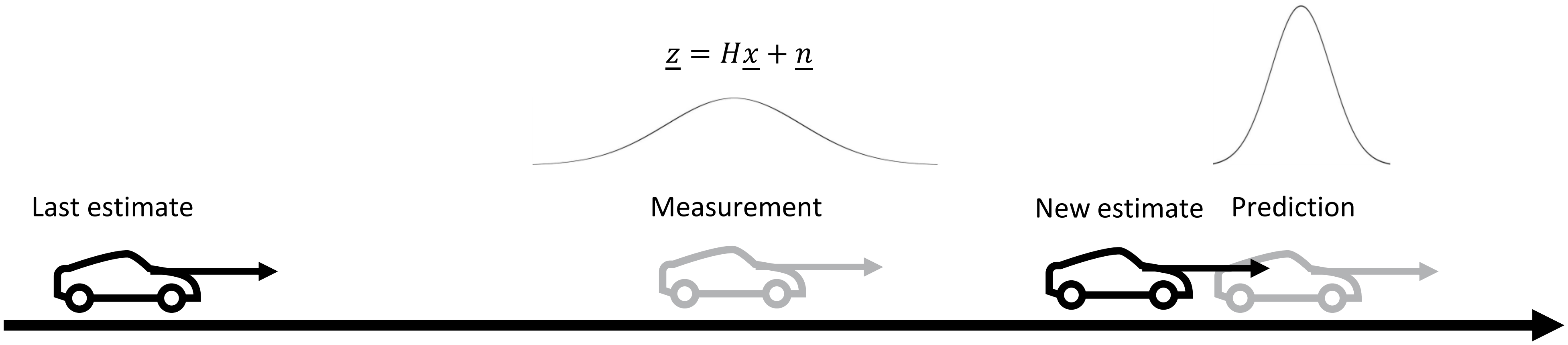
1. Predict

$$\underline{\tilde{x}}(t) = \begin{bmatrix} p + \Delta t \cdot v \\ v \end{bmatrix} = F \cdot \underline{\hat{x}}$$

$$\tilde{P} = F \cdot \hat{P} \cdot F^T + Q$$

State noise

$$\underline{z} = H\underline{x} + \underline{n}$$



New estimate = Variance-weighted mean of prediciton and measurement

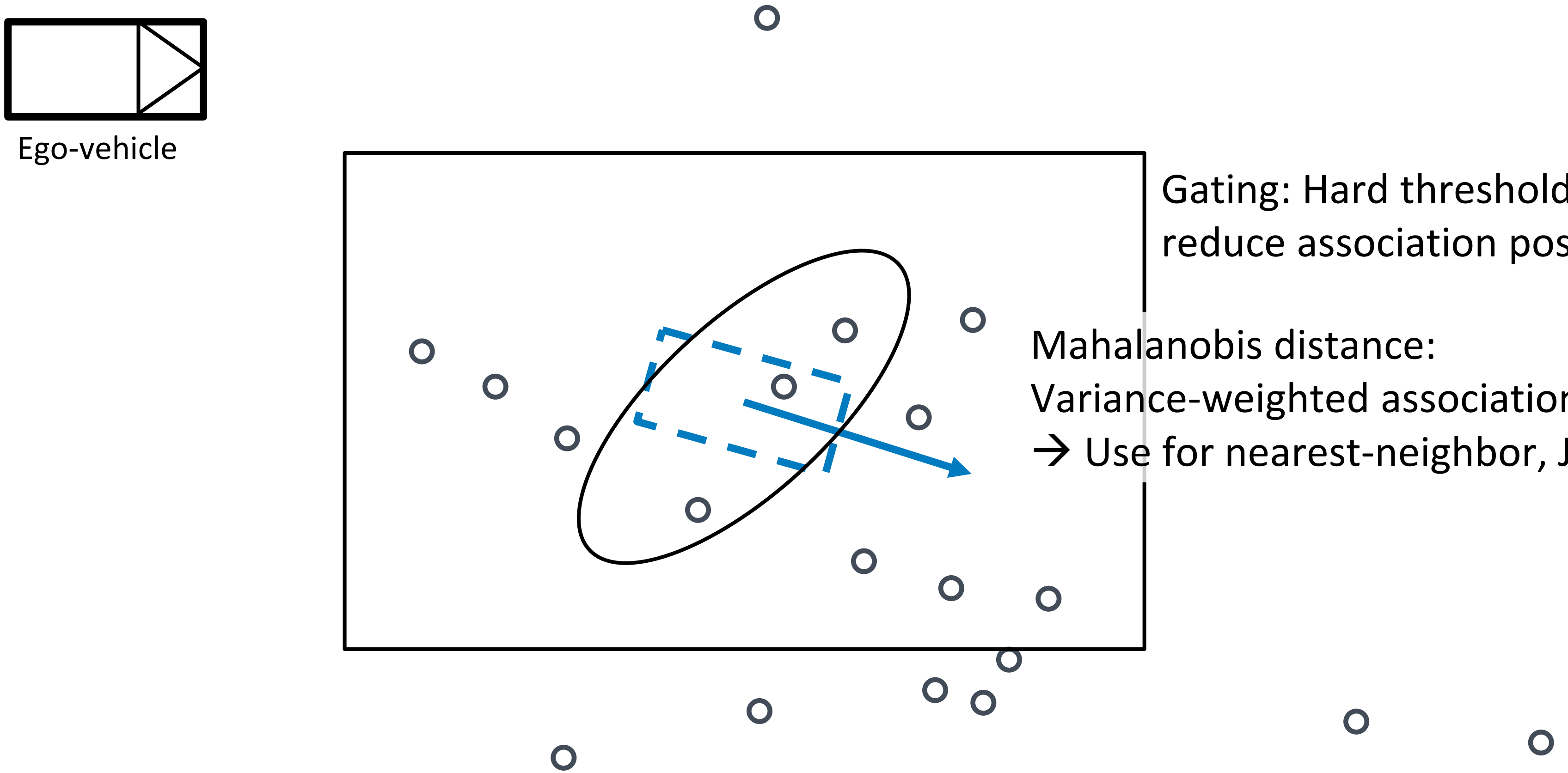
Radar-Specific Refinements of Kalman Filter Tracking

Where do the points belong?

My object
Stationary environment
A new object
Clutter

Public C-SC0 | CR/AAS2 | 2026-02-03

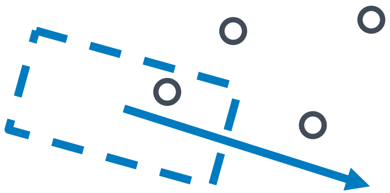
Radar-Specific Refinements of Kalman Filter Tracking



Gating: Hard thresholding to reduce association possibilities

Mahalanobis distance:
Variance-weighted association cost
→ Use for nearest-neighbor, JPDA, …

Radar-Specific Refinements of Kalman Filter Tracking



Multiple reflection points for one objects?

Objects are not Point Targets

Radar-Specific Refinements of Kalman Filter Tracking

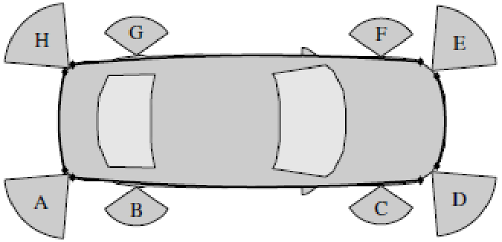


Fig. 3. Vehicle reflection model

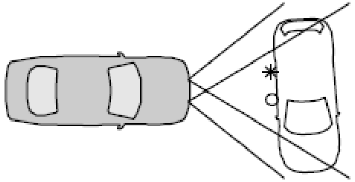
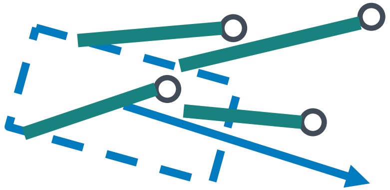


Fig. 4. Target vehicle side

Figure: M. Bühren and B. Yang, *Simulation of Automotive Radar Target Lists using a Novel Approach of Object Representation*, Intelligent Vehicles Symposium 2006



$$\underline{y} = \underline{z} - H\underline{\tilde{x}}$$

$\underline{\hat{x}}$: Must include shape parameters

Reflection model

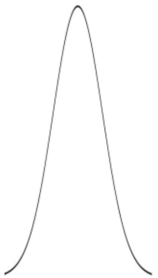
- Originally developed for simulation
- Used for extended object data association

Objects are not Point Targets

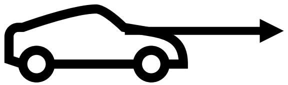
The Deep Learning Era:

End-to-End Perception

Limitations of Classical Methods



$\underline{\hat{x}} = \begin{bmatrix} p \\ v \end{bmatrix}$



Kalman Filter

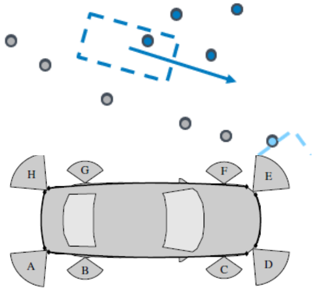
Kalman Filter is <u>optimal</u> for Systems which are:

- Linear
- have i.i.d. Gaussian noise

Radar reality

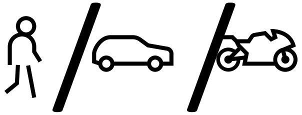
$r = \sqrt{{p_x}^2 + {p_y}^2}$ Non-linear

Q, R Model errors



Data association error

Imperfect reflection model



Imperfect object classification

Public C-SC0 | CR/AAS2 | 2026-02-03

BOSCH

Why Deep Learning for Radar?

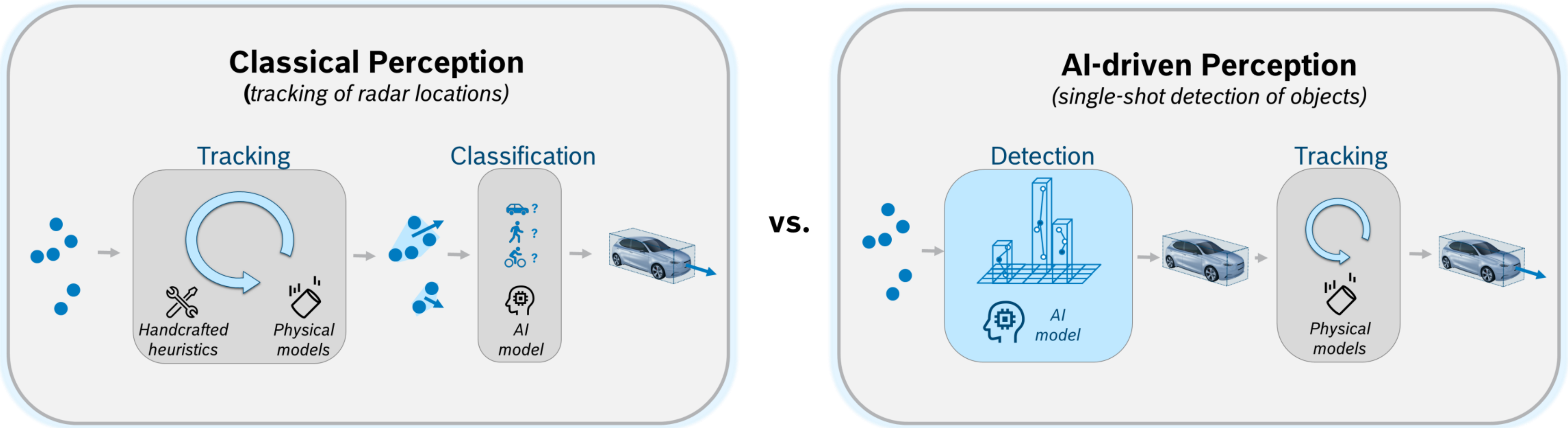


Image source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

Deep Learning for Radar Object Detection

1. Render point cloud to BEV grid
2. ResNet-style backbone
3. Multiple Heads for Perception Outputs

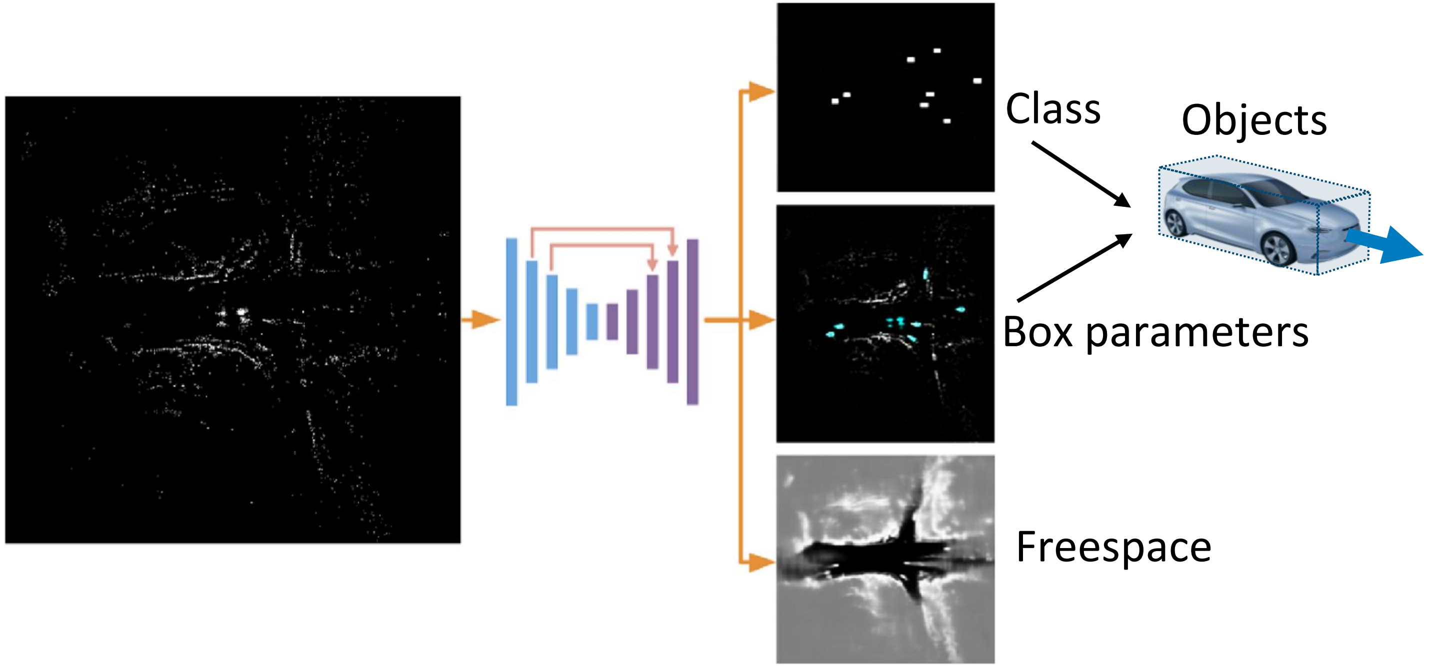
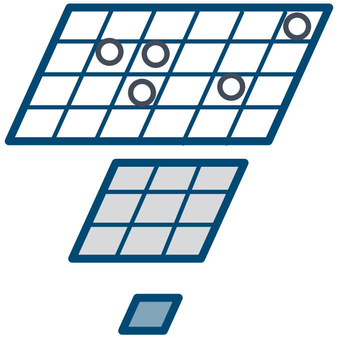
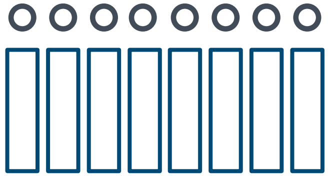


Figure: A. Popov et al., *NVRadarNet: Real-Time Radar Obstacle and Free Space Detection for Autonomous Driving, ICRA 2023*

Bird‘s-Eye-View Rendering with NVRadarNet

From CNNs to Transformer Models





CNN	Transformer
▪ Representation: Feature grids	▪ Representation: List of tokens (features)
▪ SOTA performance in ~2010-2020	▪ SOTA performance in >2020
▪ Dense representation	▪ Sparse representation
▪ Context is defined by CNN Kernel (window) size → 2D BEV geometrical	▪ Context is defined by attention window size → spatial

Public C-SC0 | CR/AAS2 | 2026-02-03

Converting a Point Cloud to Tokens

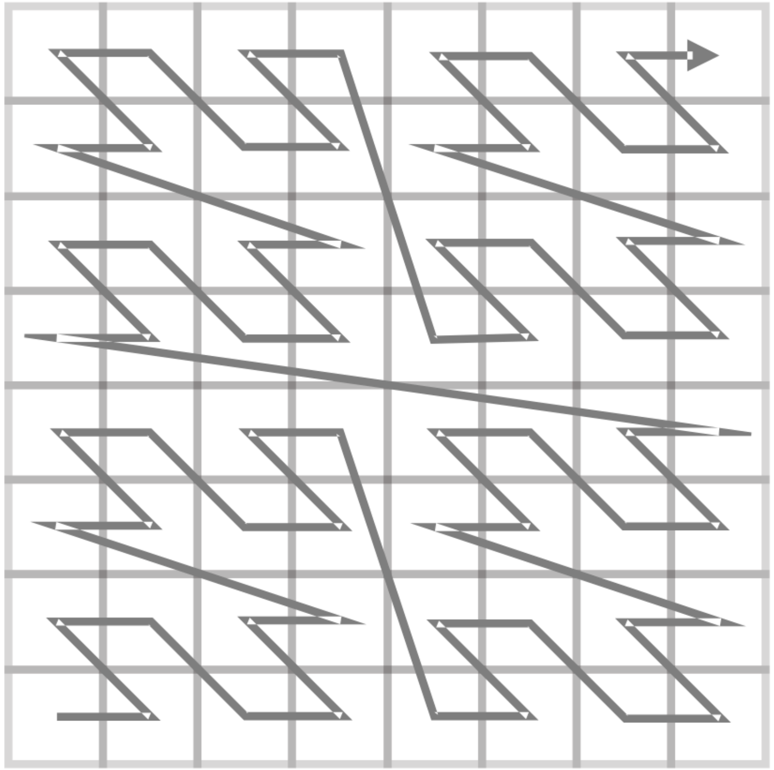


Image source: Wu, Xiaoyang, et al. "Point transformer v3: Simpler faster stronger." *CVPR* 2024.

Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Converting a Point Cloud to Tokens

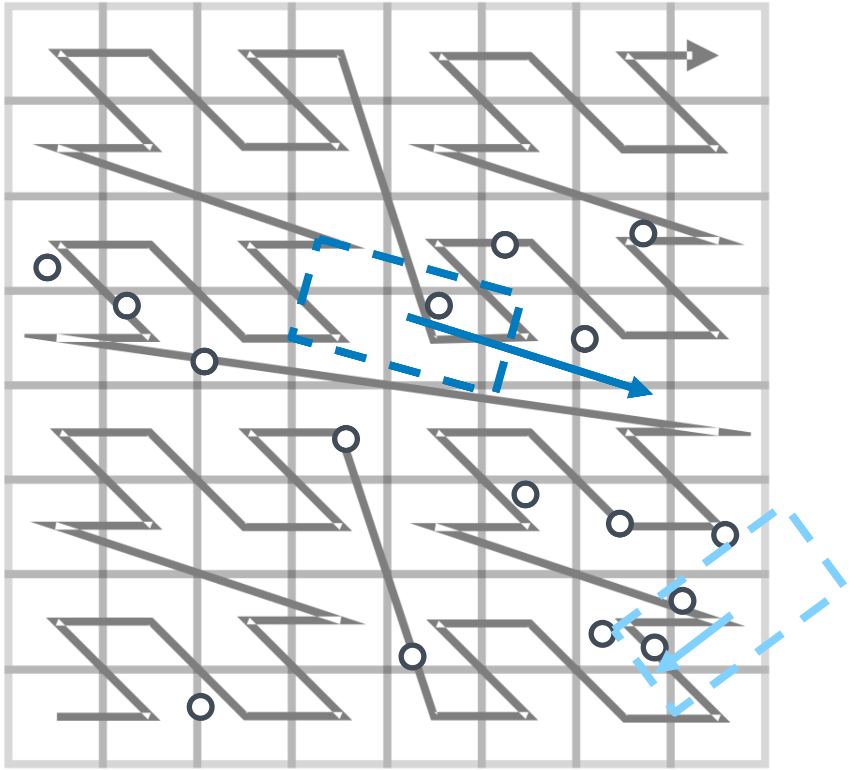


Image source: Wu, Xiaoyang, et al. "Point transformer v3: Simpler faster stronger." *CVPR* 2024.

Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Converting a Point Cloud to Tokens

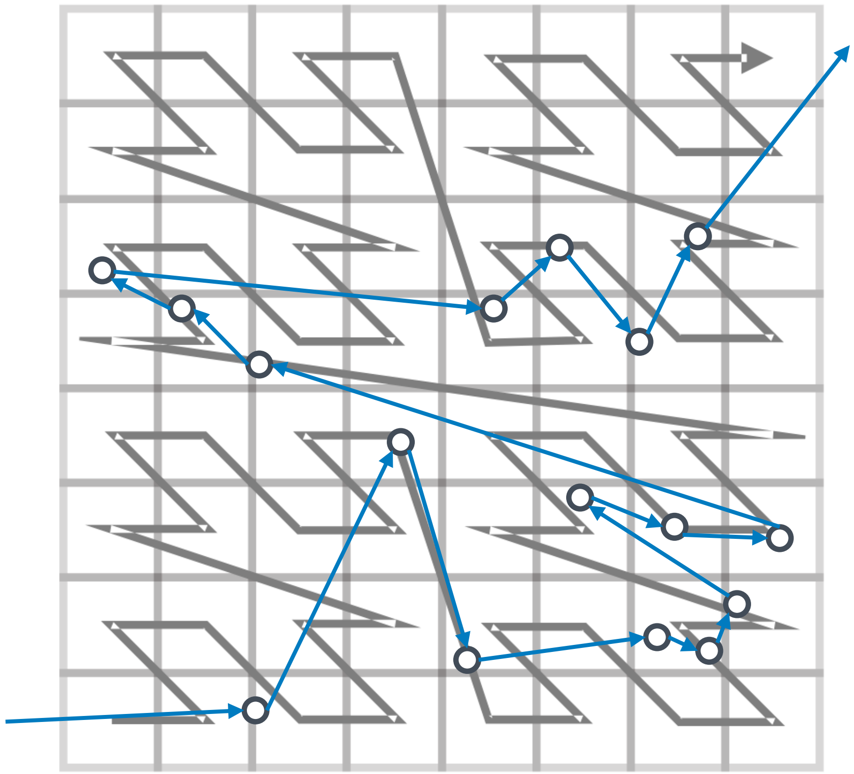
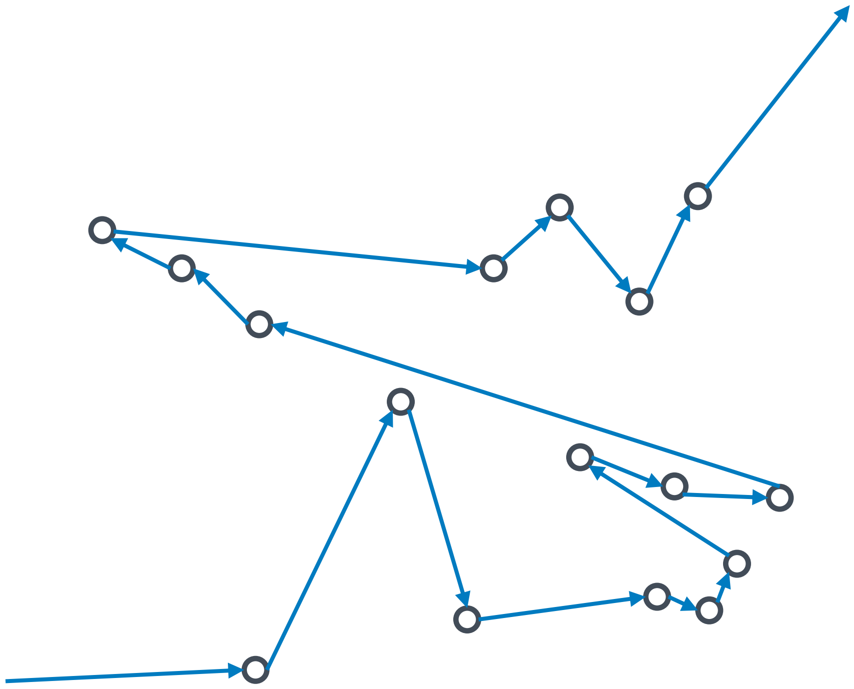


Image source: Wu, Xiaoyang, et al. "Point transformer v3: Simpler faster stronger." *CVPR* 2024.

Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Converting a Point Cloud to Tokens

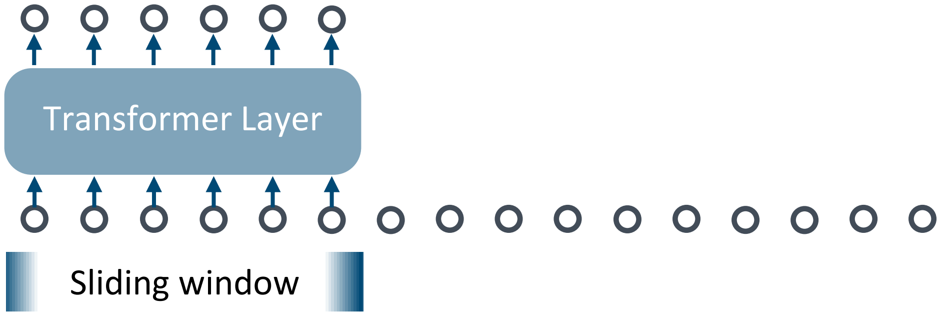


Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Converting a Point Cloud to Tokens

Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

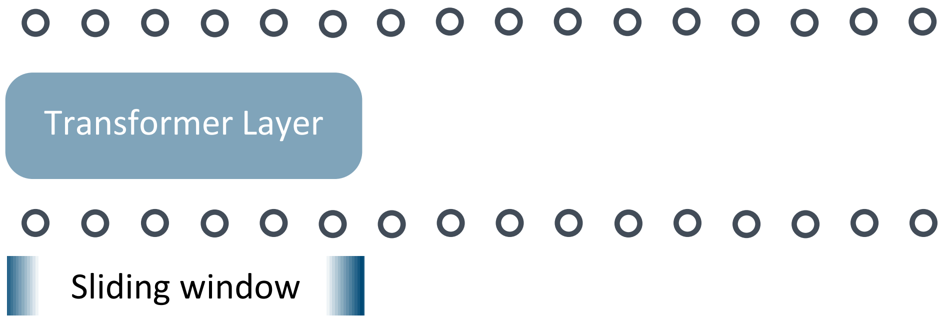
Converting a Point Cloud to Tokens



Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

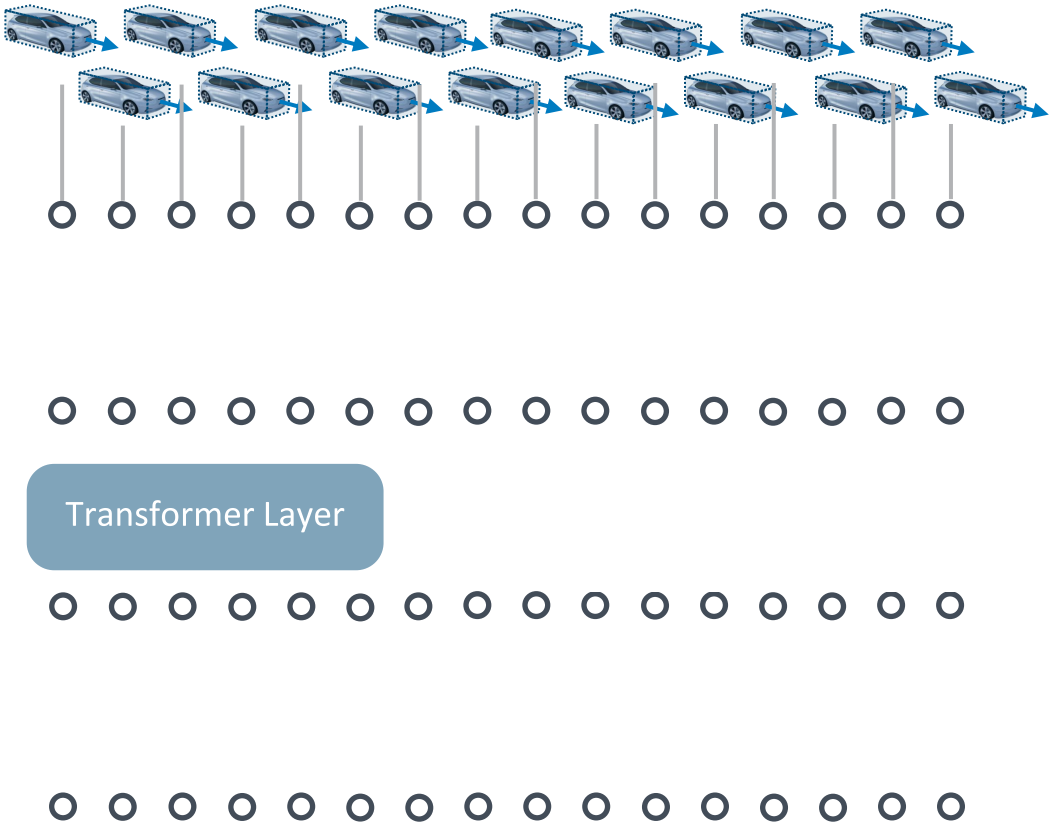
BOSCH

Converting a Point Cloud to Tokens



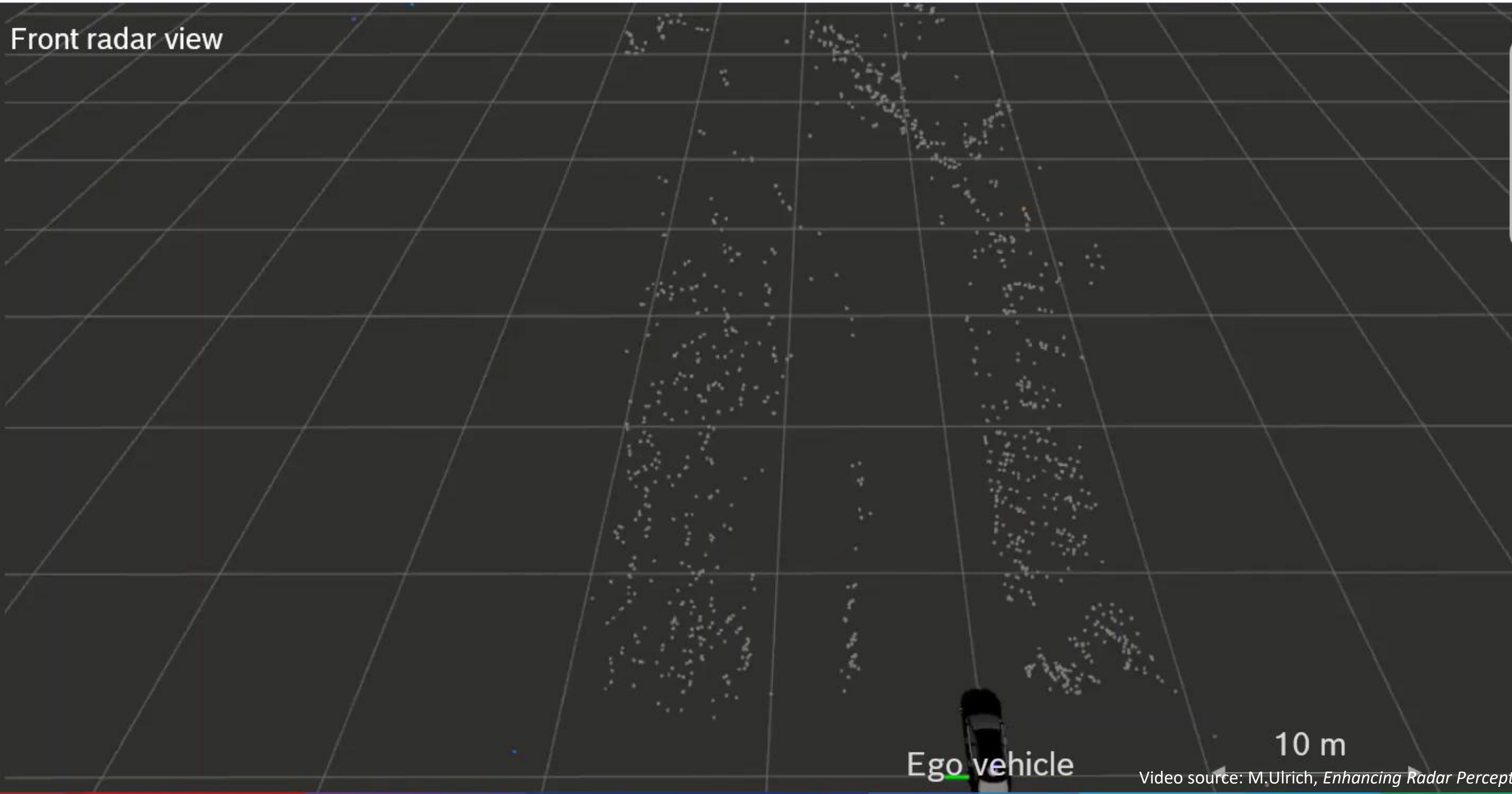
Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Converting a Point Cloud to Tokens



Sequence Models: Modern, Efficient Tokenization with Point Transformer v3

Classical Perception



Radar points

Tracked object

Video source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

The Future of Radar Perception and the Fusion Frontier

What is Automotive Radar Perception?

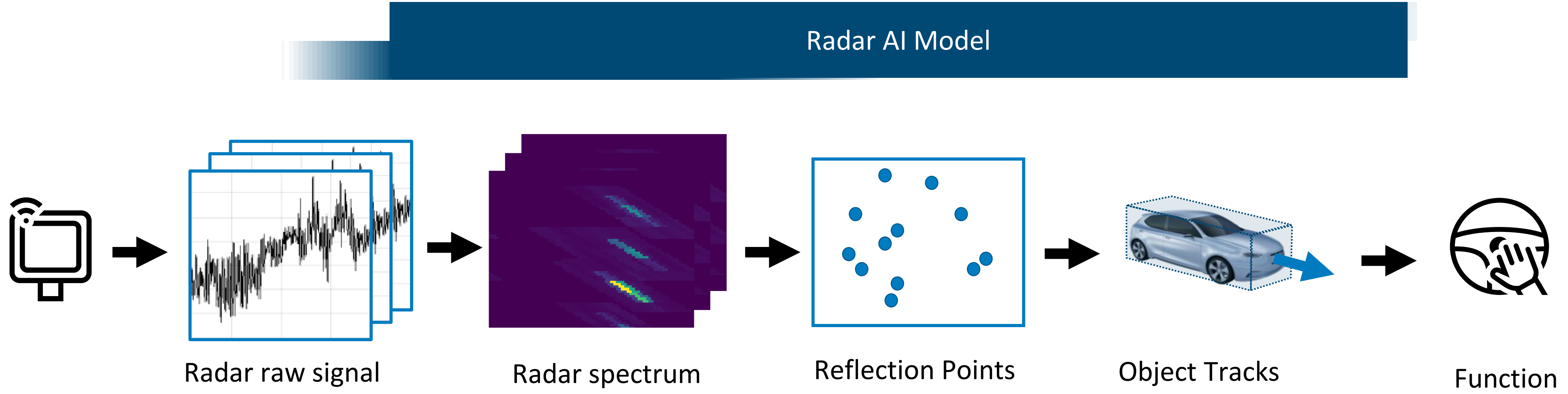
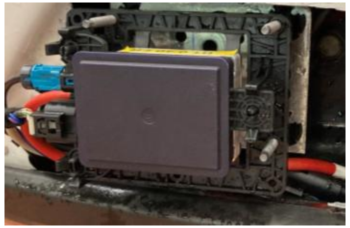


Image source: M.Ulrich, *Enhancing Radar Perception with AI,* Auto[nom]Mobil 2024

Radar AI Enables Streamlined Processing and New Features



Based on
Bosch standard
radar

Image source and video: M.Baus, *A Futuristic Vision on a Radar System for automated driving,* AutoSense China 2025

A single, end-to-end, model for perception and signal processing

Perception is More than Radar

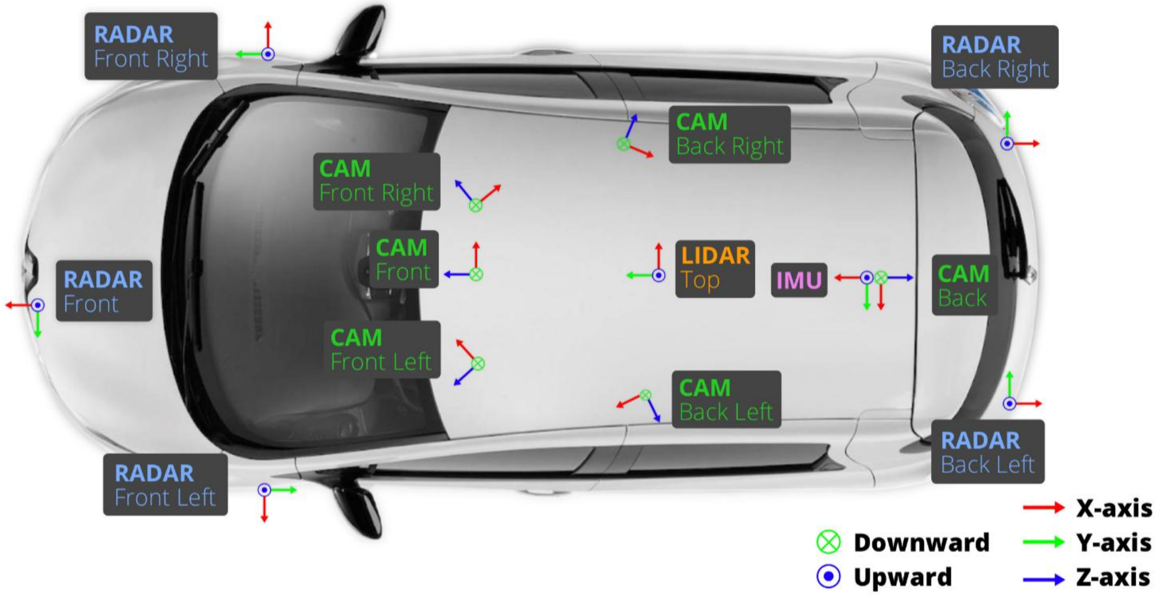


Image source: H.Caesar etal, *nuScenes: A multimodal dataset for autonomous driving*, CVPR 2020

How to combine the data of different sensors and modalities?

Fundamental Building Blocks of a Fusion System

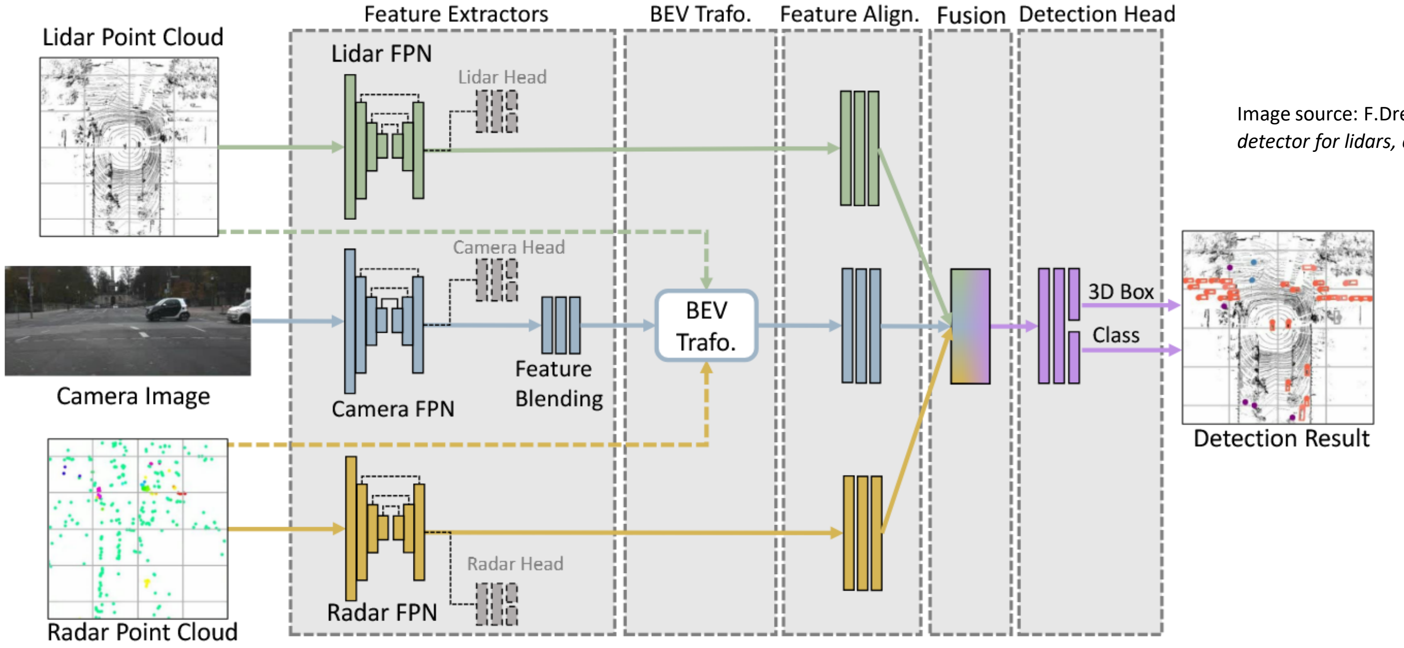


Image source: F.Drews etal, *DeepFusion: A robust and modular 3D object detector for lidars, cameras and radars, IROS*, 2022

Per sensor:
Feature extractor

Transform to common representation

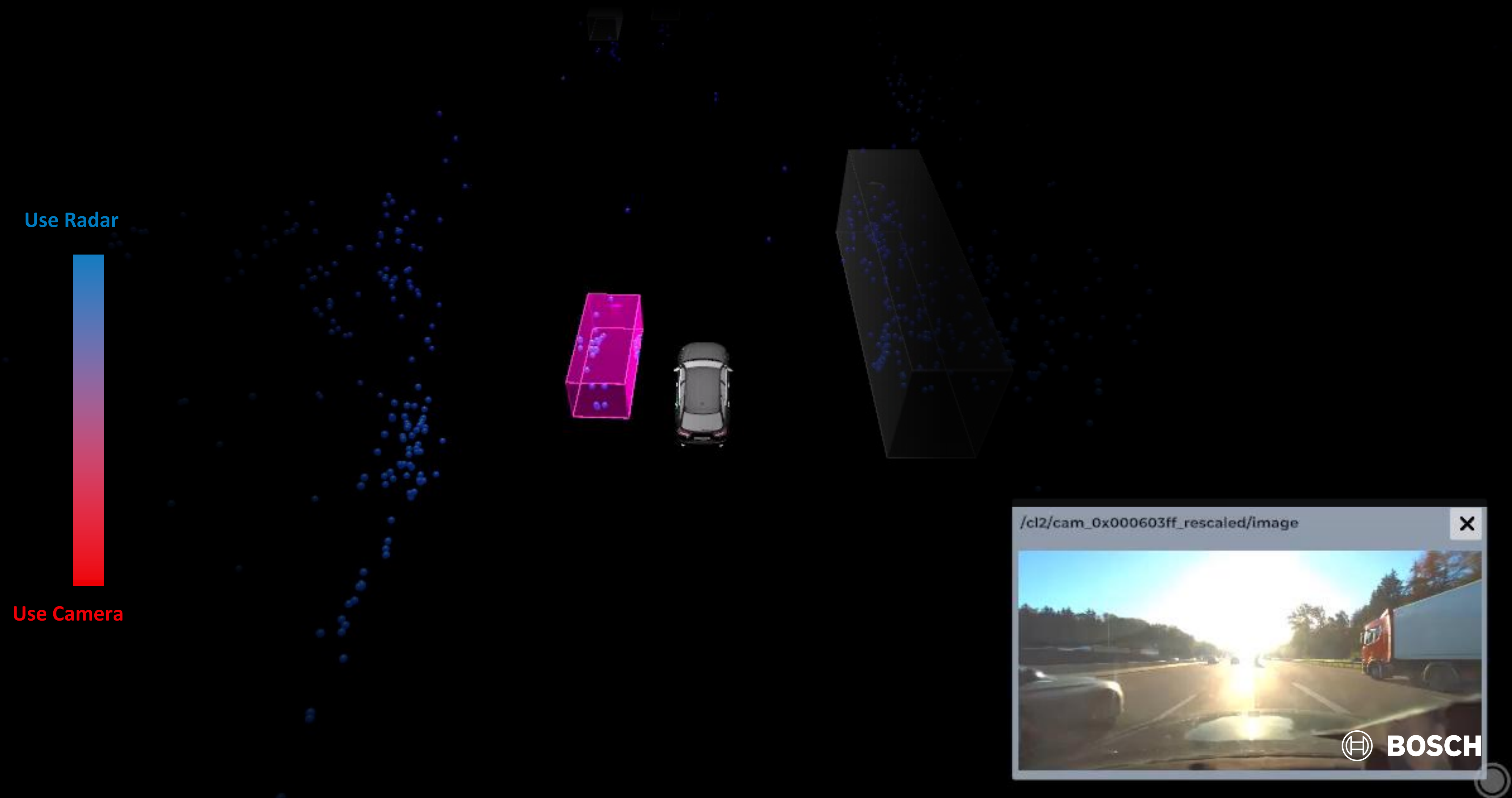
Combine features and process together

Video source: G.Gläser, *KI im Produkt – Wie Bosch die Umfeldwahrnehmung eines Autos verbessert –*, 2025

BOSCH

Video source: G.Gläser, *KI im Produkt – Wie Bosch die Umfeldwahrnehmung eines Autos verbessert –*, 2025

Video source: G.Gläser, *KI im Produkt – Wie Bosch die Umfeldwahrnehmung eines Autos verbessert –*, 2025



Summary

The Evolution of Automotive Radar Perception

- Radar is essential: All-weather sensing, cost and robustness
- Kalman Filtering:
  - A robust baseline with physics (motion models) and statistics (filtering)
  - Key concepts: Prediction/Update, data association, extended object tracking
- Deep Learning Perception:
  - End-to-end, data-driven method → Moving from handcrafted models to learned representations
  - Key concepts: CNNs and the bird’s-eye-view, transformers and point clouds as a sequence
- Beyond Point-Clouds:
  - Integrated perception and signal processing: Efficiency and performance gains
  - Radar-camera fusion: Combining the best of two worlds