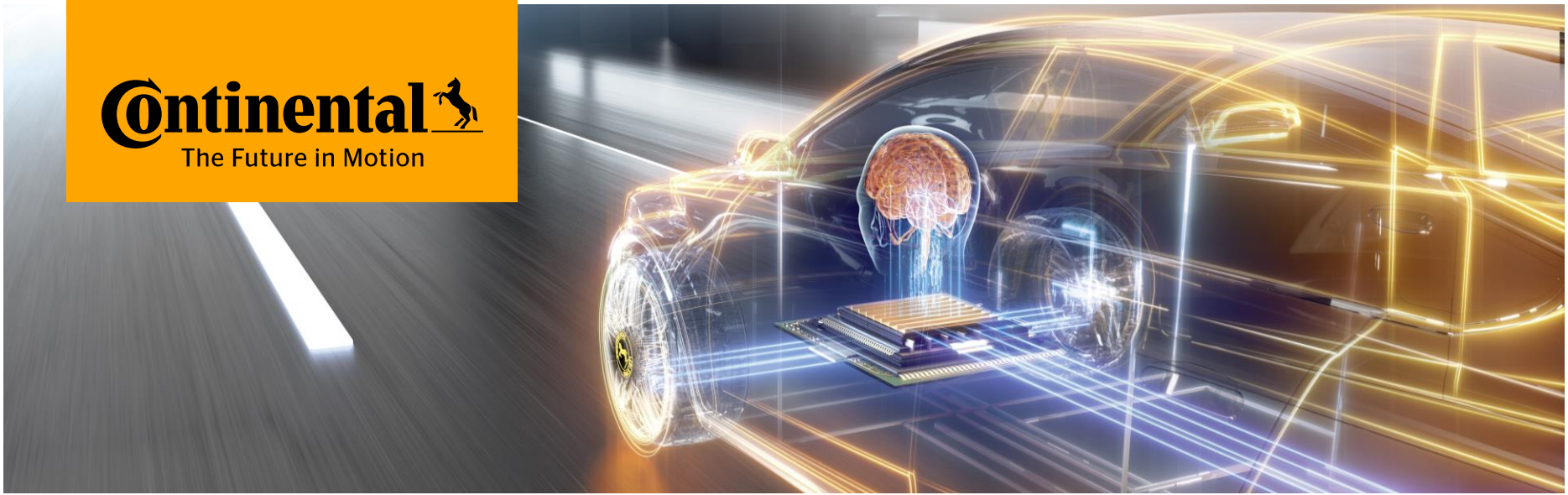


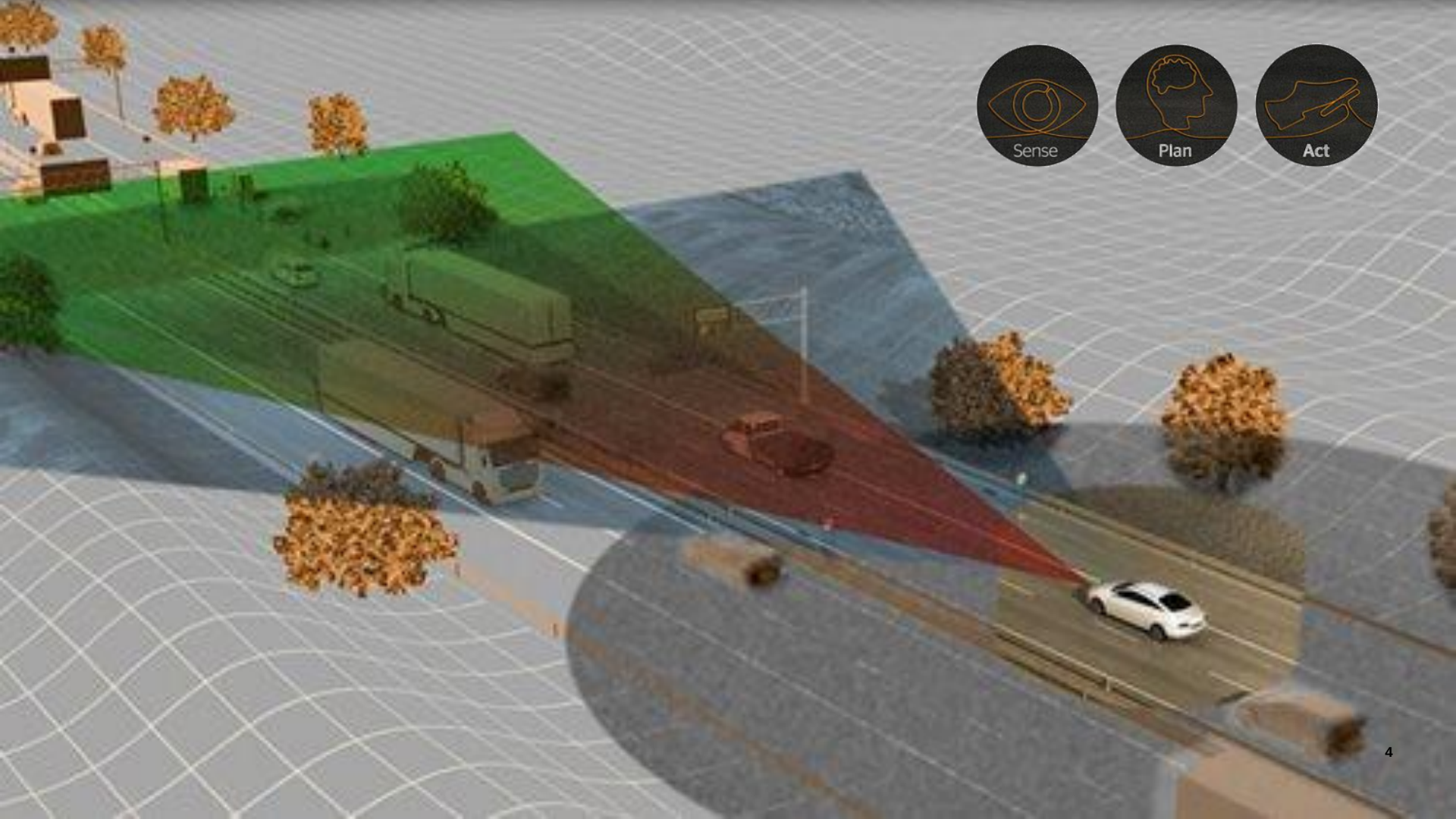
Effective Autonomous Driving with Sensor Fusion

Robert-Zsolt Kabai
Team Lead, Deep Learning for Point Cloud and Fusion
Machine Learning Methods
ADAS Budapest – Deep Learning Competence Center – Continental

Agenda

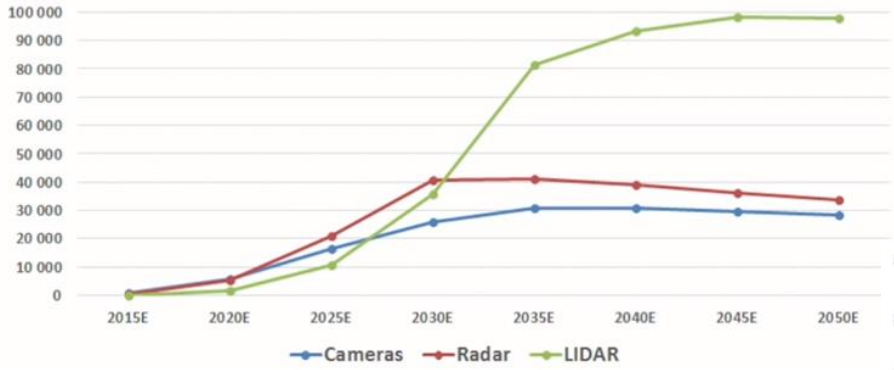
1. Technical challenges of perception
2. AI solutions with Sensor Fusion
3. Budapest Fusion AI developments

Technical challenges of perception

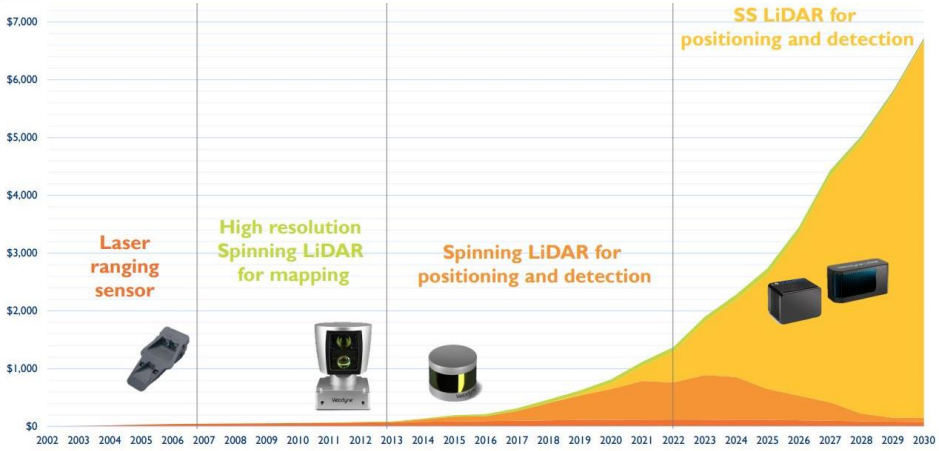

Sense
Plan
Act

Market overview

Sensor Market projections (Goldman Sachs Investment Research)



LIDAR technology market (YOLE)



Key LIDAR requirements for Autonomous Driving

- A cost-effective product ticking all AD requirements could lead the market.

Solid-State Pros/Cons:

+ global shutter	- low resolution
+ high framerate	- range (HFL)
+ cheap	
+ no moving parts	

Scanning (e.g. Velodyne) Pros/Cons:

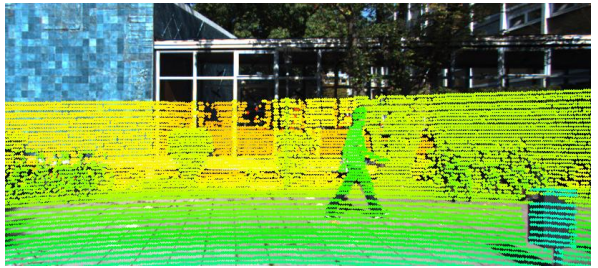
+ high resolution	- rolling shutter
+ range	- low framerate
	- expensive
	- moving parts

Cheaper LIDARs are relatively low resolution

- Functions like Object Detection are challenging with current Solid State LIDAR resolution

Cheaper Solid State LIDAR

Expensive Rotating LIDAR

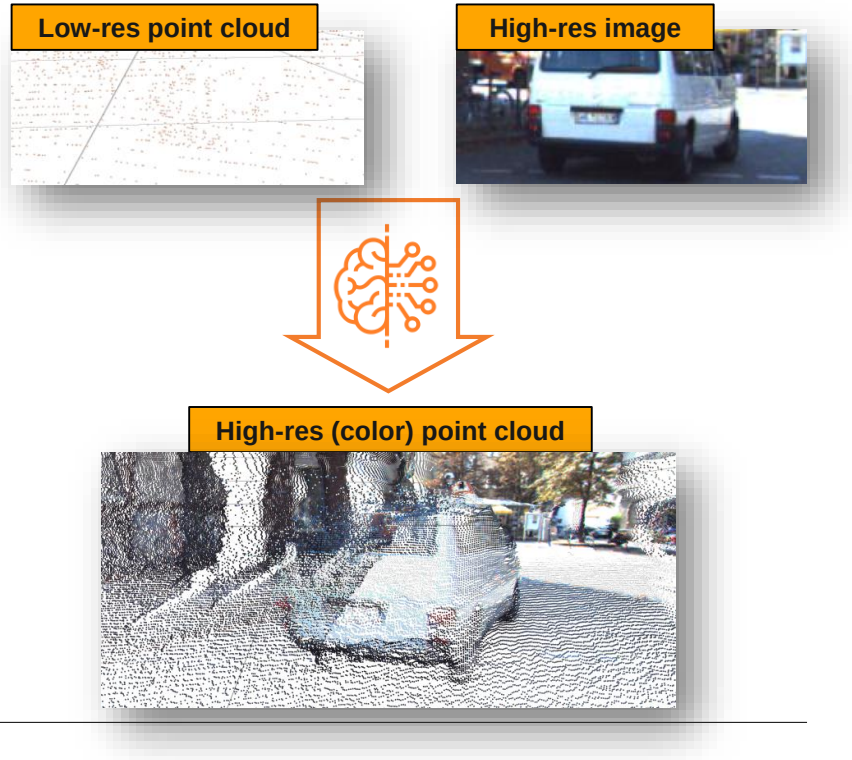


AI solutions with Sensor Fusion

LIDAR resolution upscaling using AI

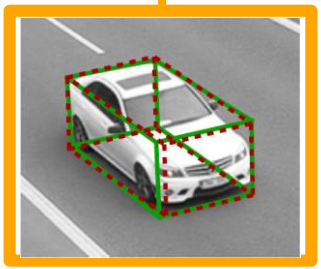
Fusion based 3D Reconstruction

- Sensor Fusion
  - Low resolution LIDAR measurements
  - High resolution camera measurements
- 3D Reconstruction
  - Correct perspective change induced artifacts
    e.g. object occlusion shift
  - Predict 3D position of pixels by high resolution densified ground truth LIDAR fused with image based texture change and object edges boundaries
  - Get reconstructed 3D scene described by high resolution RGB colored LIDAR point cloud



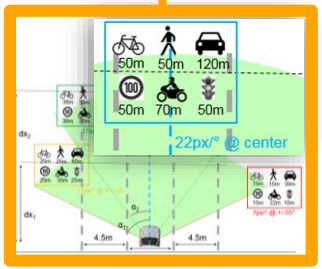
Fusion based 3D Reconstruction opportunities

More precise 3D box fit



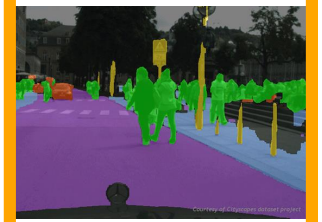
Car + Sign = Truck in OD

Minimum resolution requirement

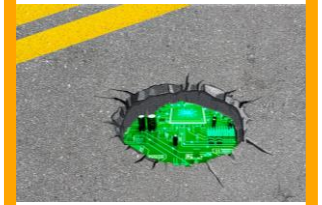


Fusion based 3D Reconstruction opportunities

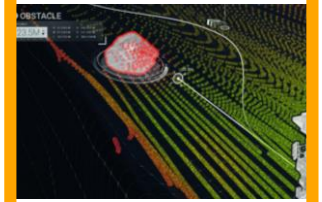
More precise Semantic Segmentation



Drivable Terrain Free Space

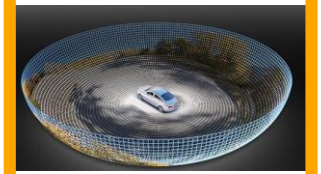


Obstacle detection
Lost Cargo detection
Emergency Breaking
Crossing Traffic
Animals Detection



Road Curb Height

Environmental Mapping



Three pillars of successful 3D Reconstruction

Data selection for fusion

- Identify **scenes best suited for 3D Reconstruction**
- Can use **wide range of data** without annotation requirements
- High density, low dynamic object speed to reduce motion artifacts e.g. parking, intersections

Algorithmic ground-truth generation

- **No semantic manual annotation needed!**
- **Scalable process** instead: algorithmic, fusion-tailored, multi-lidar and time-based dense LIDAR ground truth generation
- **Cheap data acquisition** including training-ready ground truth, reusable data from wide range of recordings

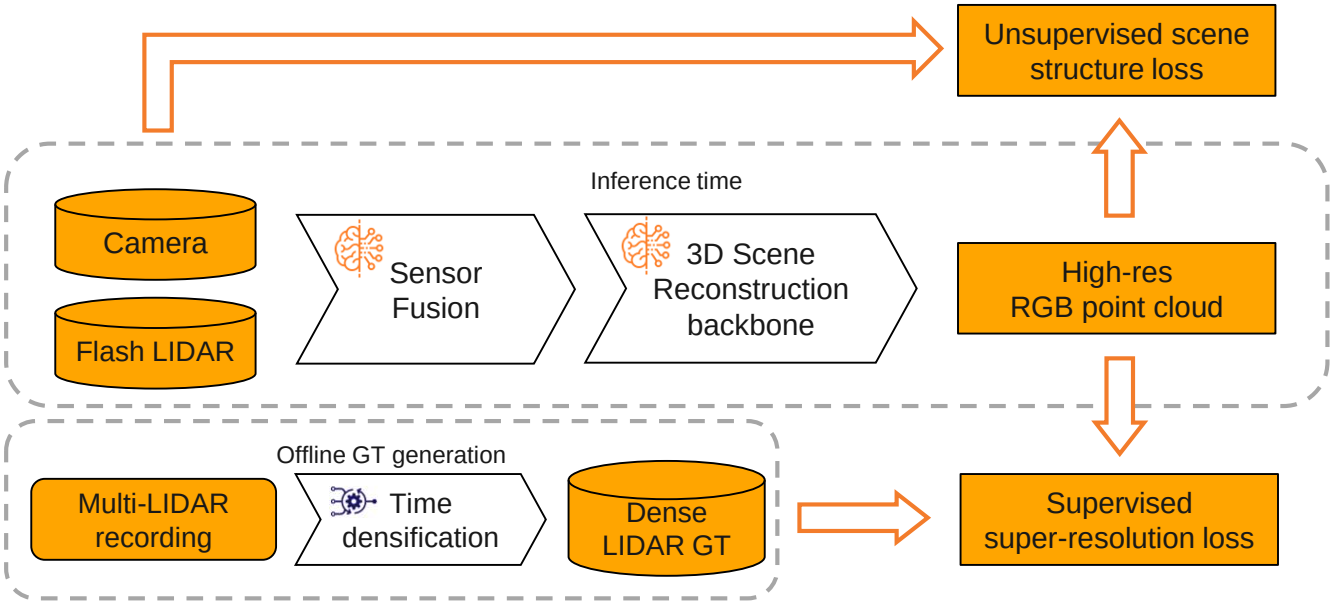
Network Architecture

- **Specialized network architecture** for 3D Reconstruction
- Downstream tasks and 3D Reconstruction ability **trainable separately** with task specific data
- Scene 3D reconstruction reliability and stability can be **validated and proven separately**
- Robust, low-risk, **detached development**; **extra flexibility** for downstream task developer (free to use point cloud or depth map)

Budapest Fusion AI developments

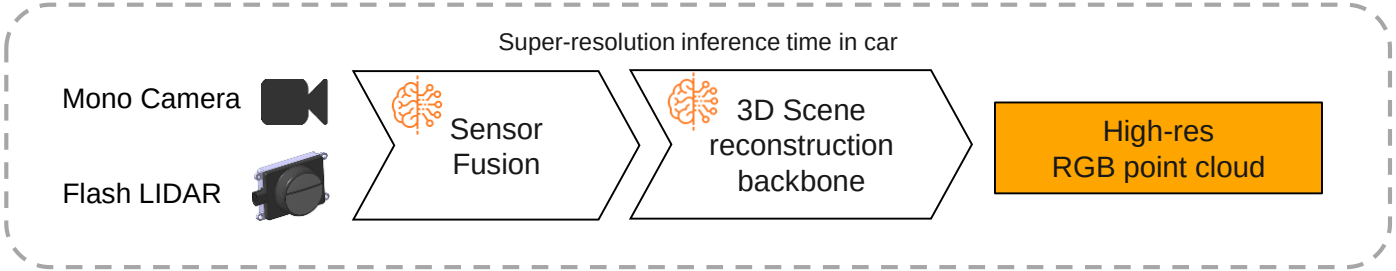
LIDAR-Camera Fusion - 3D Reconstruction

Concept details



LIDAR-Camera Fusion - 3D Reconstruction

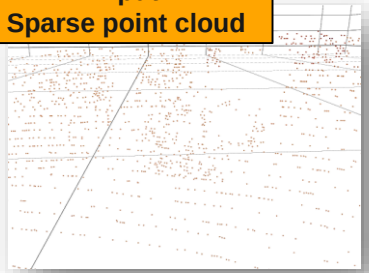
Inference time setup



LIDAR-Camera Fusion - 3D Reconstruction

3D view of point clouds

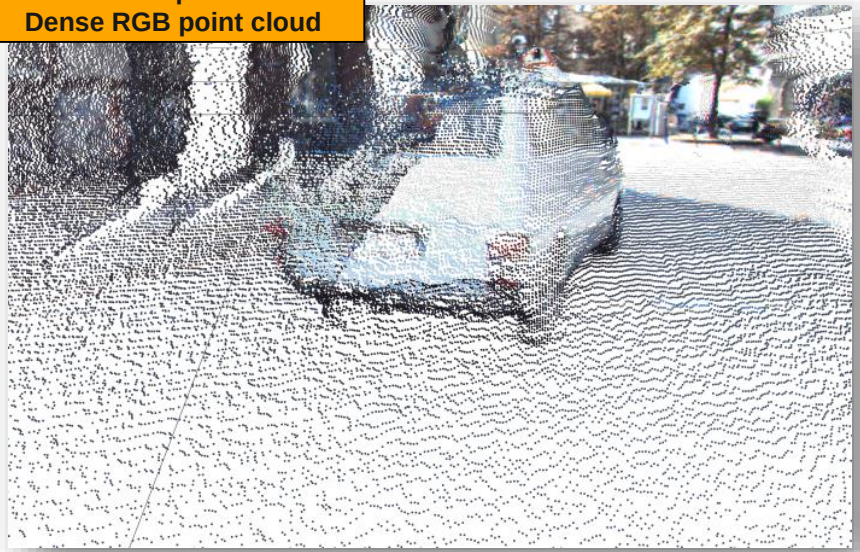
Input
Sparse point cloud



Input
RGB image

Fusion based
3D Reconstruction

Output
Dense RGB point cloud



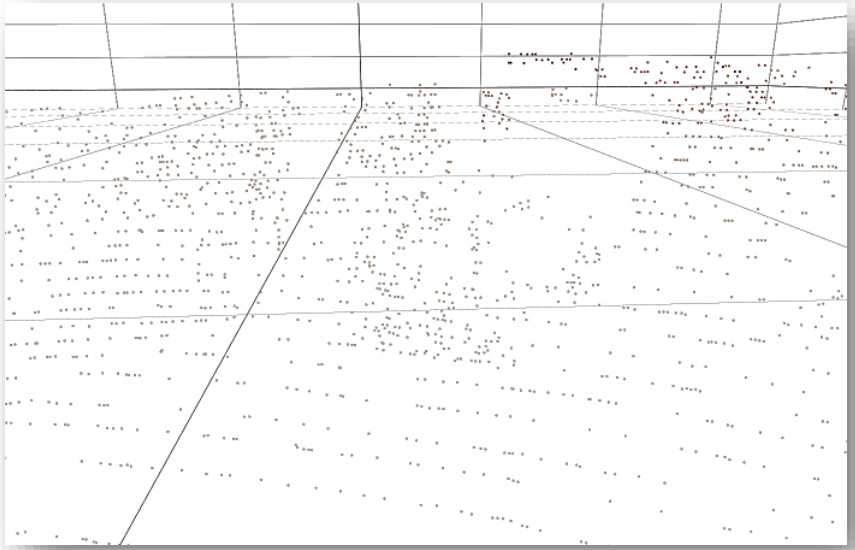
LIDAR-Camera Fusion - 3D Reconstruction

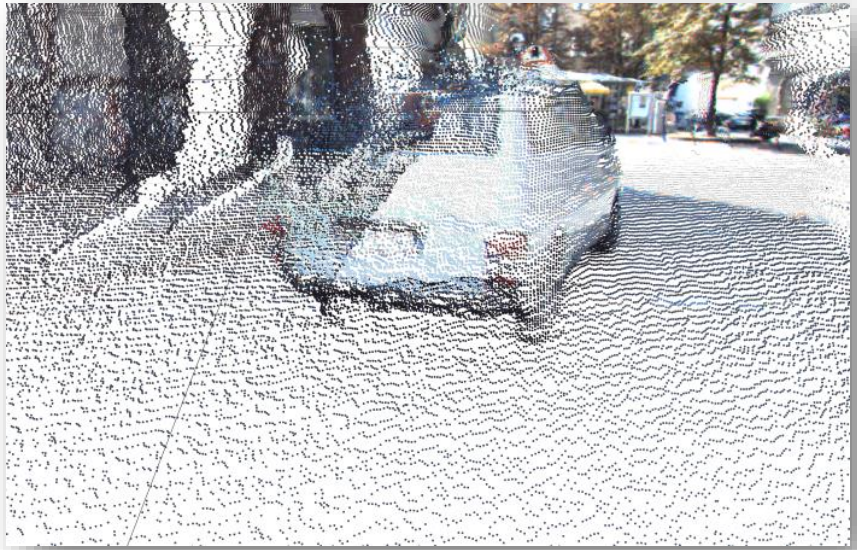
3D view of point clouds – a closer look

Input
Sparse point cloud

Fusion based
3D Reconstruction

Output
Dense RGB point cloud

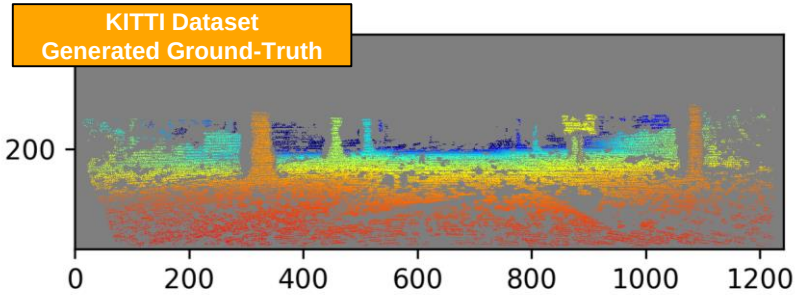


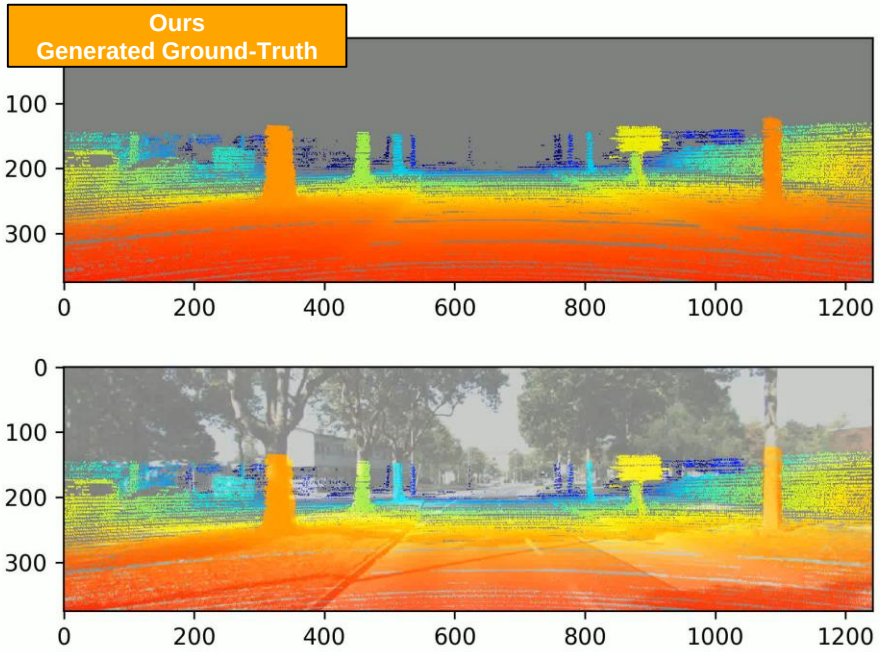


LIDAR Ground-truth generat

Demo

- Proprietary Ground Truth densification algorithm
- More accurate and much more dense compared to public KITTI version





Demo

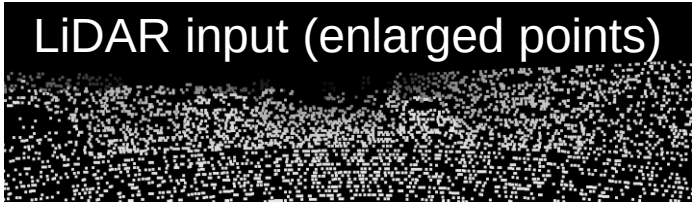
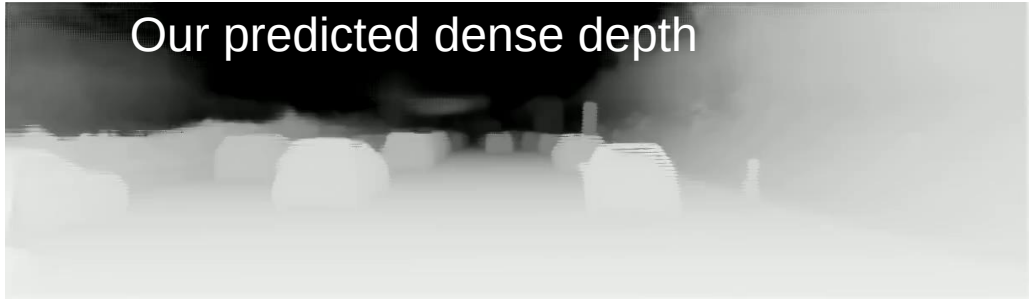

LiDAR input (enlarged points)

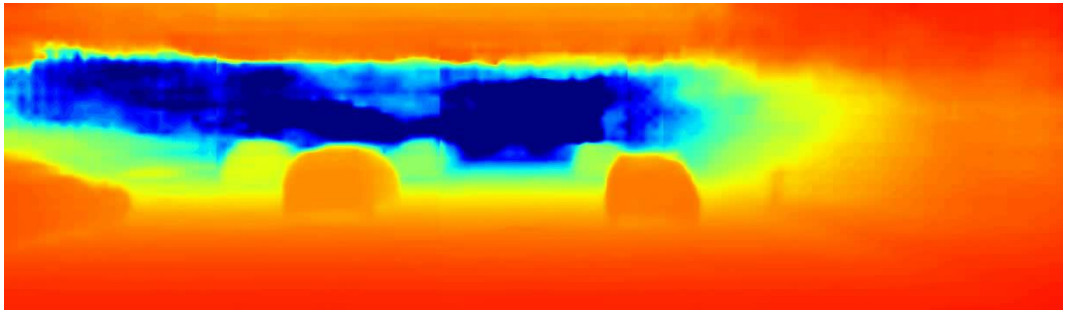

Image input

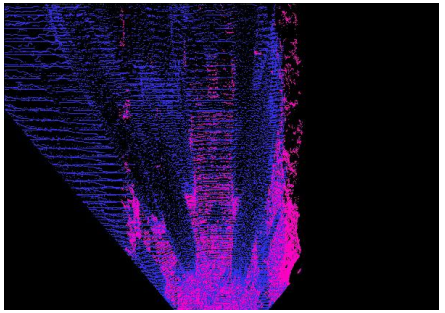

Our predicted dense depth

Demo – Comparison with Vision-only models

Vision-only Supervised	
RMSE	100%

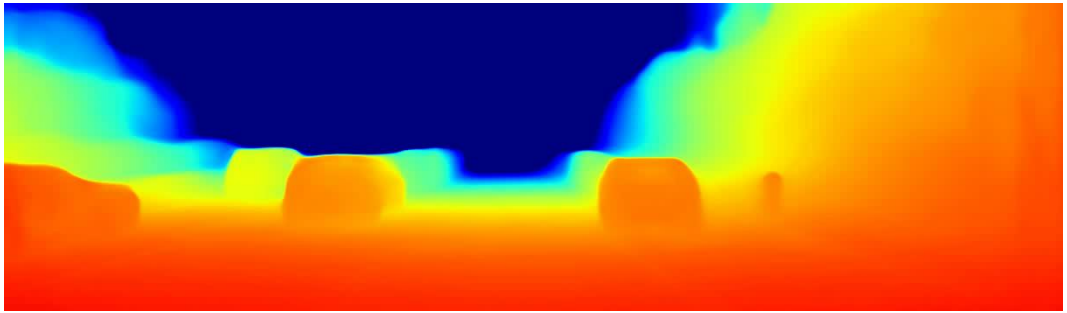
[DORN, Fu et al. 2018]

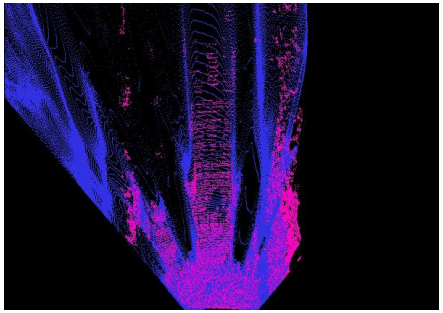




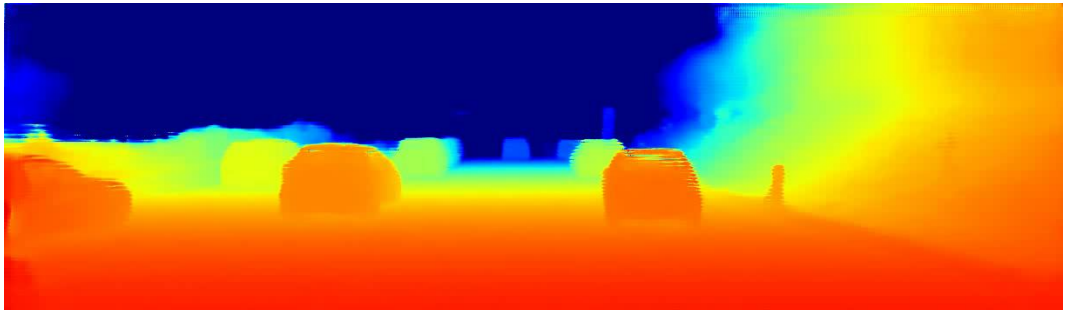
Vision-only Structure From Motion	
RMSE	126%

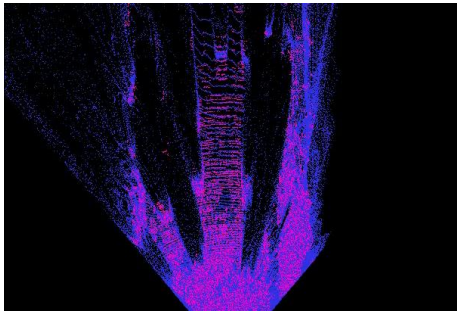
[SC-SfM, Bian et al. 2020]





Ours	
RMSE	28%





3D Object Detection enhanced with Fusion

PATENT

Mono Camera

Flash LIDAR

Sensor Fusion

3D Scene reconstruction backbone

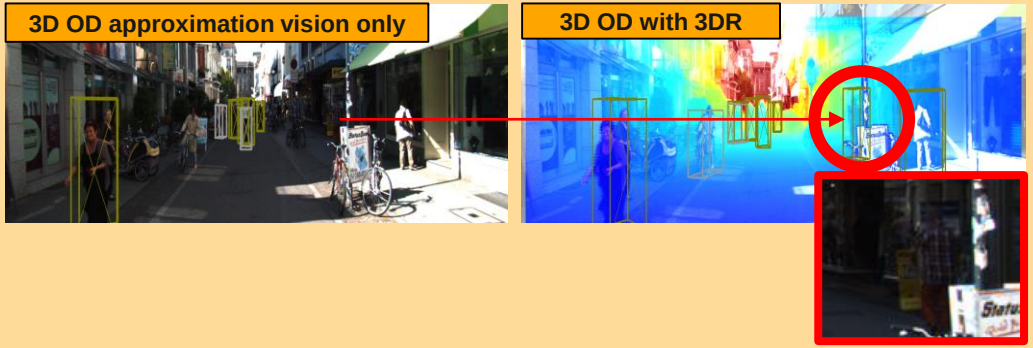
High-res RGB Point Cloud

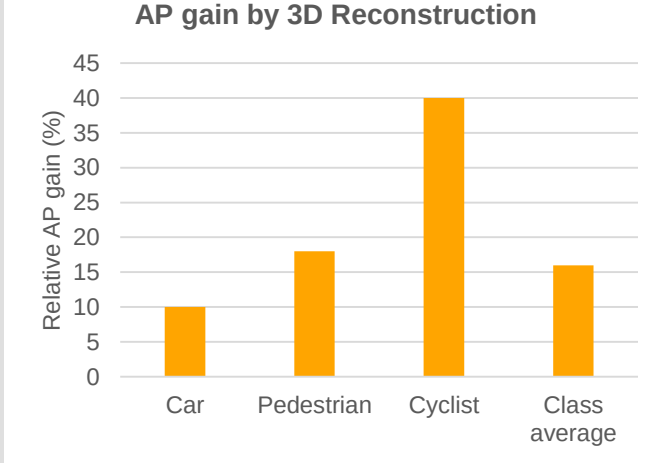
Object Detection backbone

+15% avg. AP gain

Detected Objects

Saving cyclists and pedestrians



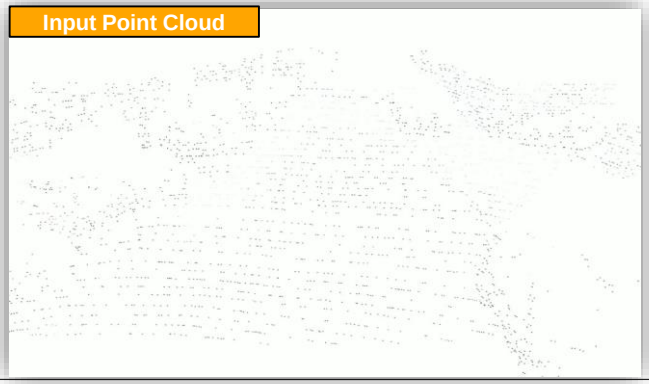


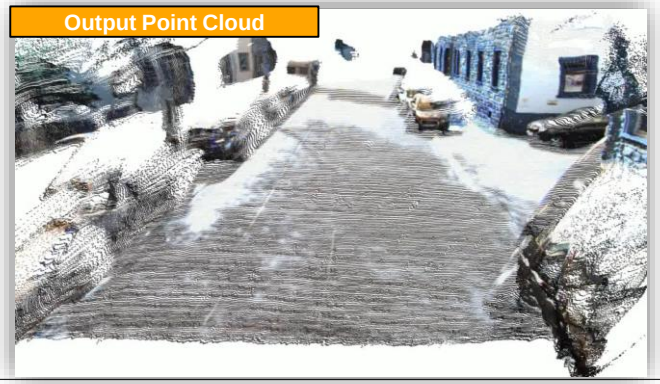
Summary

Scope:
- Creating up to *~100x resolution enhancement* for LIDARs using AI, with LIDAR-Camera low level Fusion and 3D Reconstruction, creating a cost-effective superior resolution LIDAR

Algorithmic benefits:
- *Enhance accuracy and robustness* of downstream tasks (e.g. Object Detection, Semantic Segmentation etc.) by using the advantages of both sensors to correct the drawbacks of both

Product benefits:
- *Superior performance for driving functions*
  e.g. Drivable Terrain, Free Space, Obstacle Detection,
  Emergency Braking, Road Curb Height, Environmental Mapping





Safe and Dynamic Driving towards Vision Zero

Questions?

SensePlanAct