ML Without Proper Data in Advanced Robotic Navigation – a practical outlook

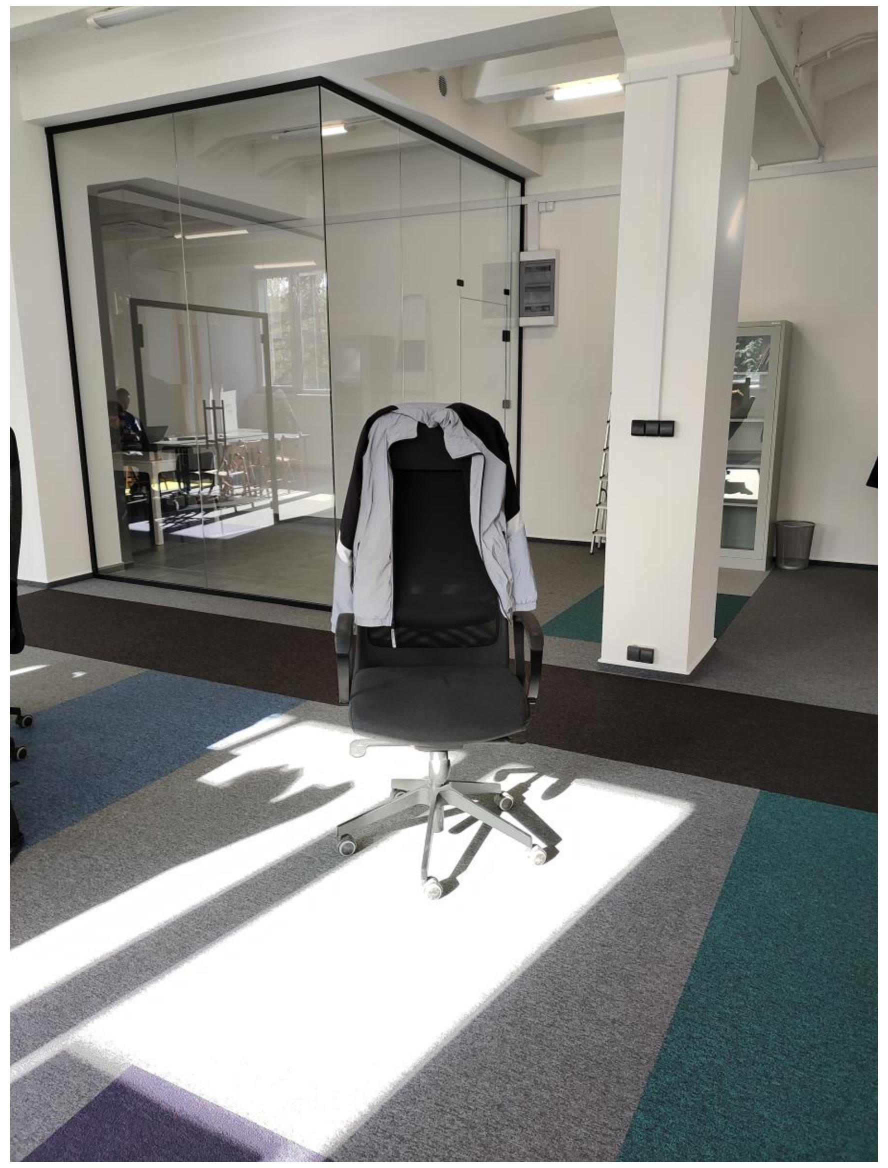
Konrad Cop
Warsaw University of Technology / United Robots

United Robots

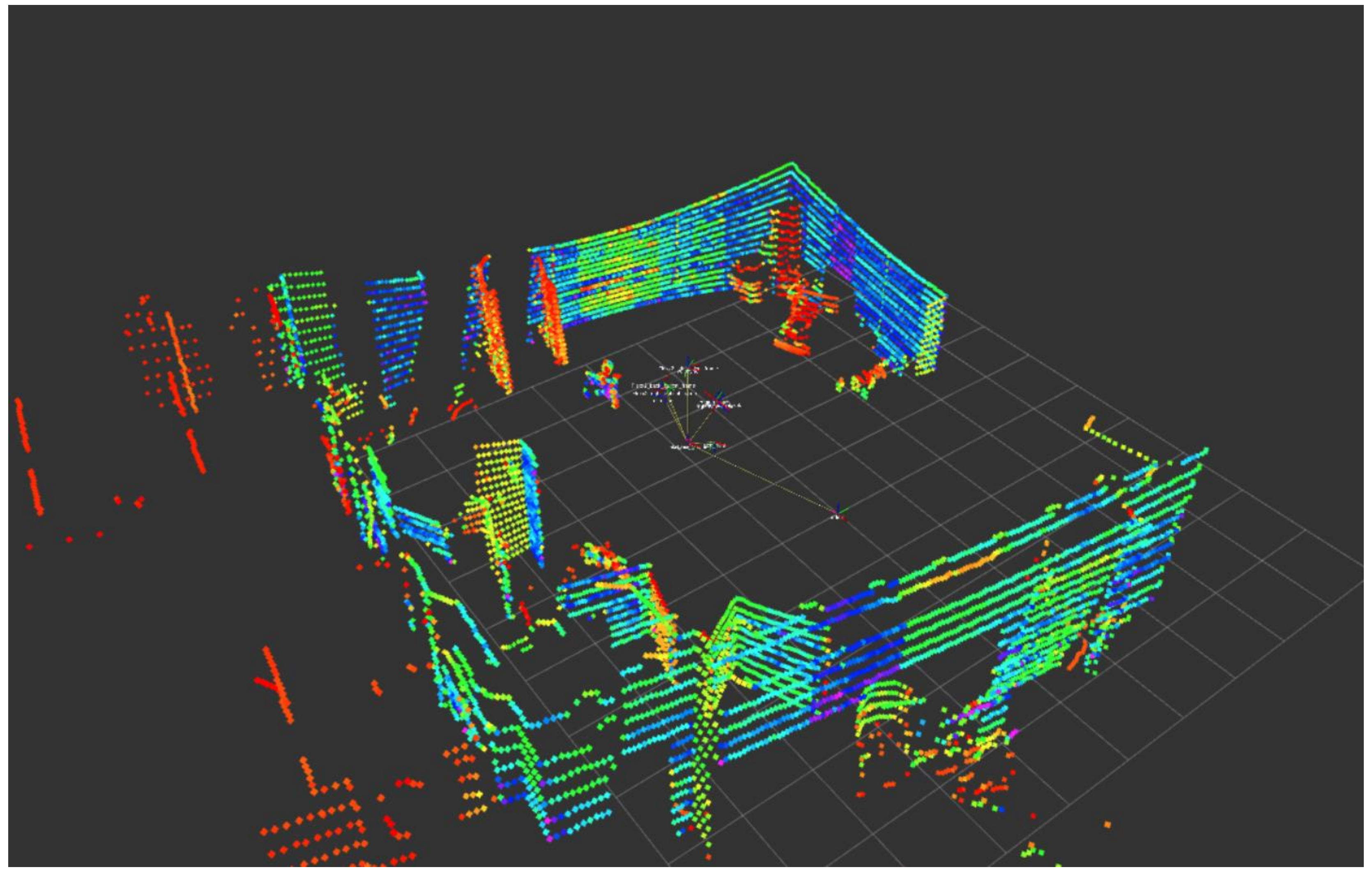
CVLab@PW

Perception with 3D sensors

Depth (ToF)



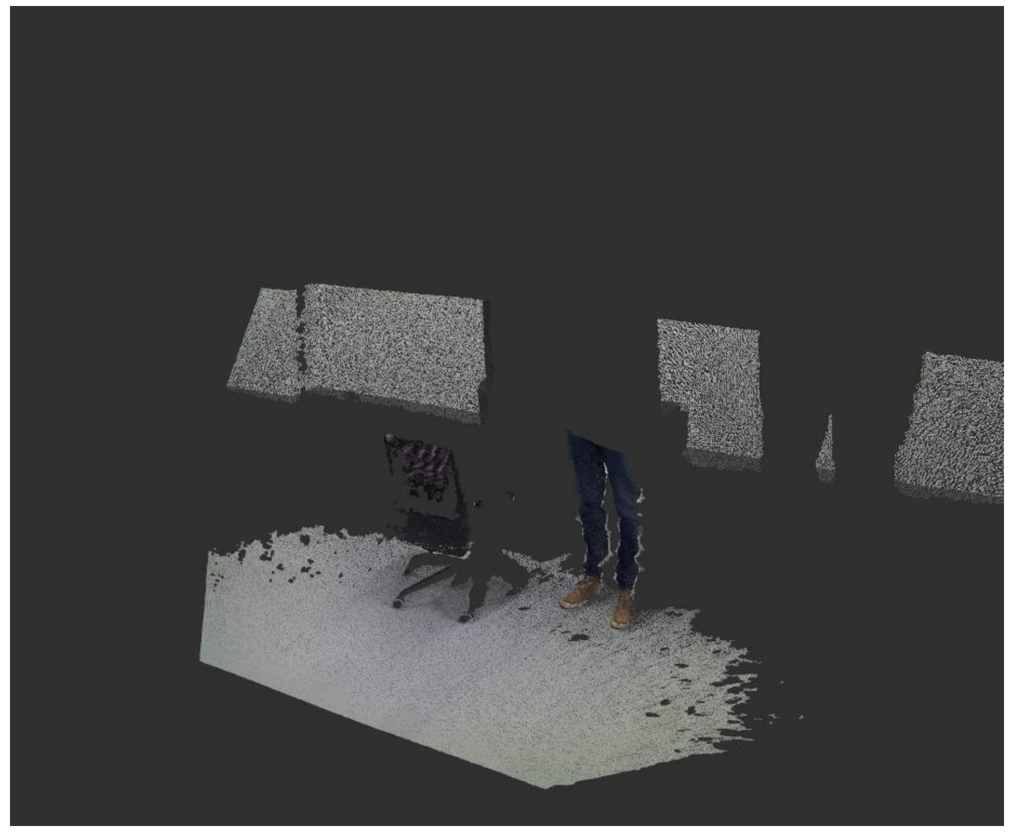
LiDAR

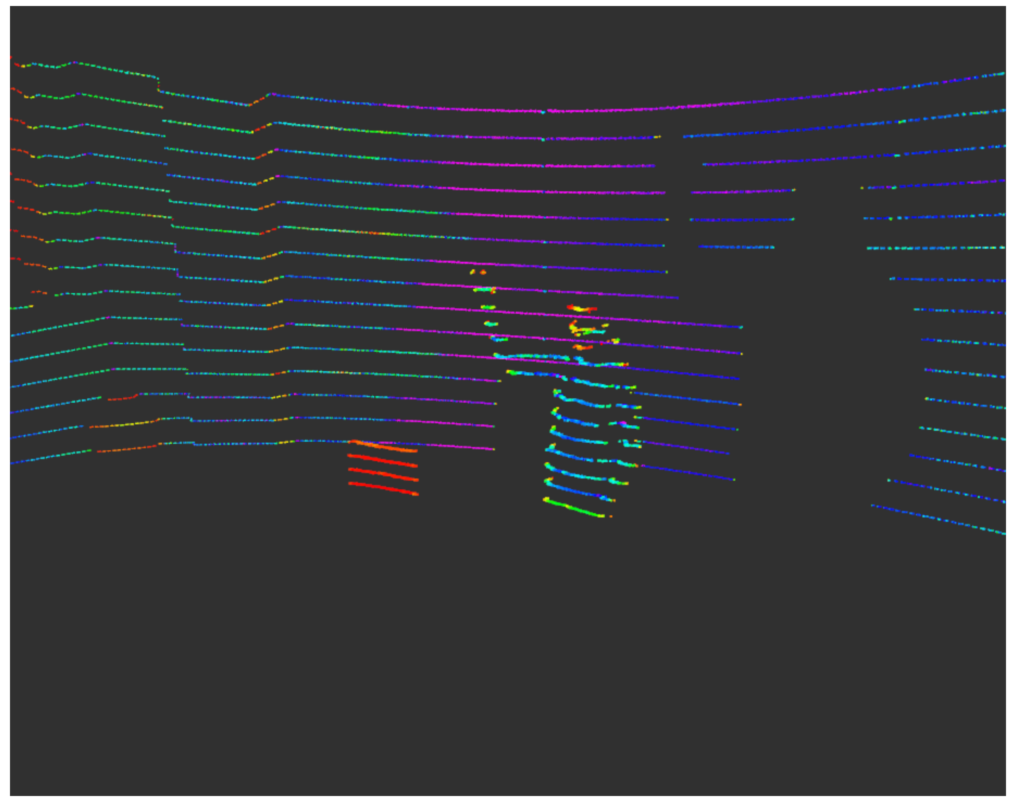


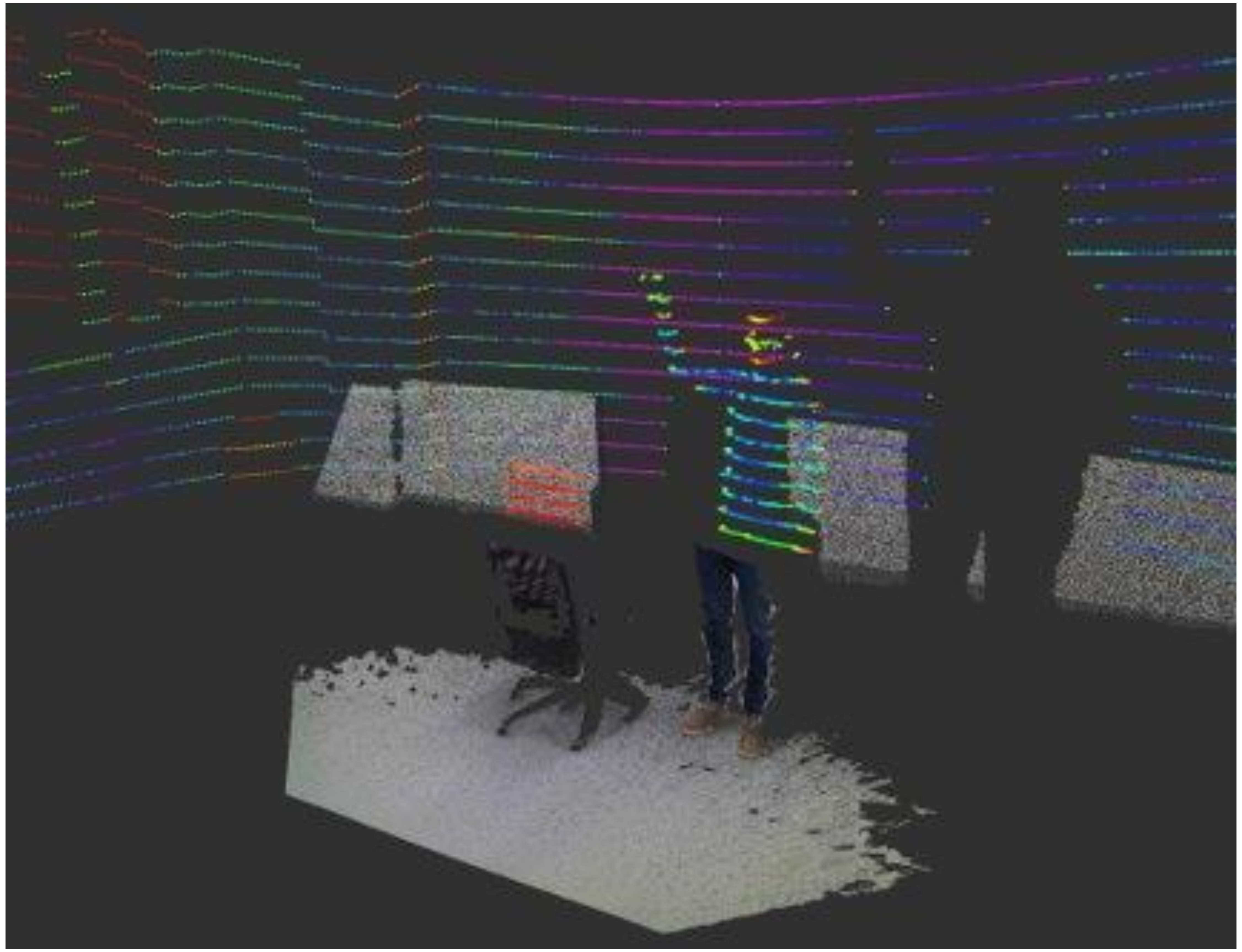
United Robots

CVLab@PW

Perception with 3D sensors

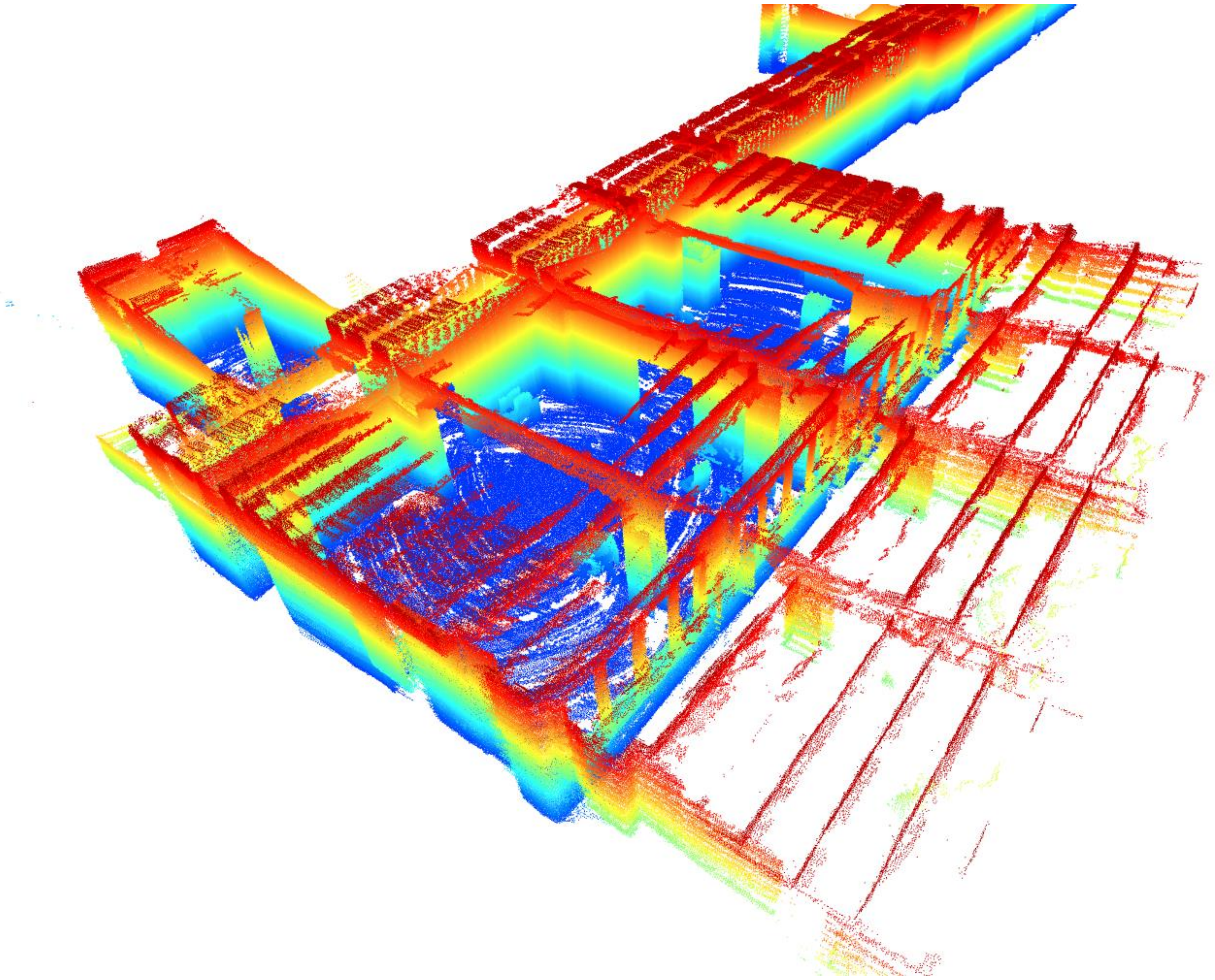






United Robots CVLab@PW

More complex clouds



United Robots

CVLab@PW

Research areas: 3D point clouds

Papers With Code

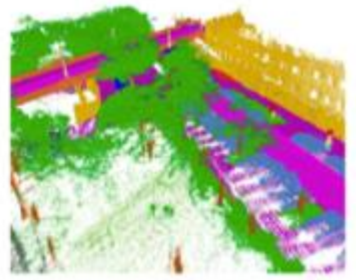
Computer Vision

3D Semantic Segmentation

Edit

150 papers with code • 12 benchmarks • 29 datasets

3D Semantic Segmentation is a computer vision task that involves dividing a 3D point cloud or 3D mesh into semantically meaningful parts or regions. The goal of 3D semantic segmentation is to identify and label different objects and parts within a 3D scene, which can be used for applications such as robotics, autonomous driving, and augmented reality.



Benchmarks

Add a Result

These leaderboards are used to track progress in 3D Semantic Segmentation

Computer Vision

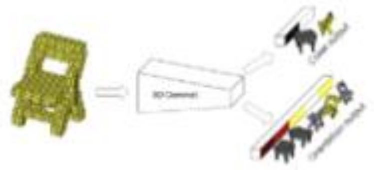
3D Object Classification

Edit

35 papers with code • 3 benchmarks • 6 datasets

3D Object Classification is the task of predicting the class of a 3D object point cloud. It is a voxel level prediction where each voxel is classified into a category. The popular benchmark for this task is the ModelNet dataset. The models for this task are usually evaluated with the Classification Accuracy metric.

Image: Sedaghat et al



Computer Vision

3D Semantic Scene Completion

Edit

23 papers with code • 3 benchmarks • 3 datasets

This task was introduced in "Semantic Scene Completion from a Single Depth Image" (https://arxiv.org/abs/1611.08974) at CVPR 2017 . The target is to infer the dense 3D voxelized semantic scene from an incompleted 3D input (e.g. point cloud, depth map) and an optional RGB image. A recent summary can be found in the paper "3D Semantic Scene Completion: a Survey" (https://arxiv.org/abs/2103.07466), published at IJCV 2021.

Content

Computer Vision

3D Place Recognition

4 papers with code • 1 benchmarks • 1 datasets

Pointcloud-based place recognition and retrieval

Computer Vision

3D Instance Segmentation

Edit

47 papers with code • 8 benchmarks • 13 datasets

Image: OccuSeg

Benchmarks

Add a Result

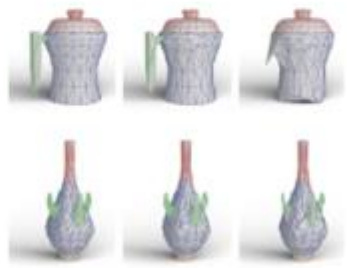
3D Part Segmentation

Edit

61 papers with code • 2 benchmarks • 5 datasets

Segmenting 3D object parts

(Image credit: MeshCNN: A Network with an Edge)



Benchmarks

Add a Result

These leaderboards are used to track progress in 3D Part Segmentation

Computer Vision

Point Cloud Completion

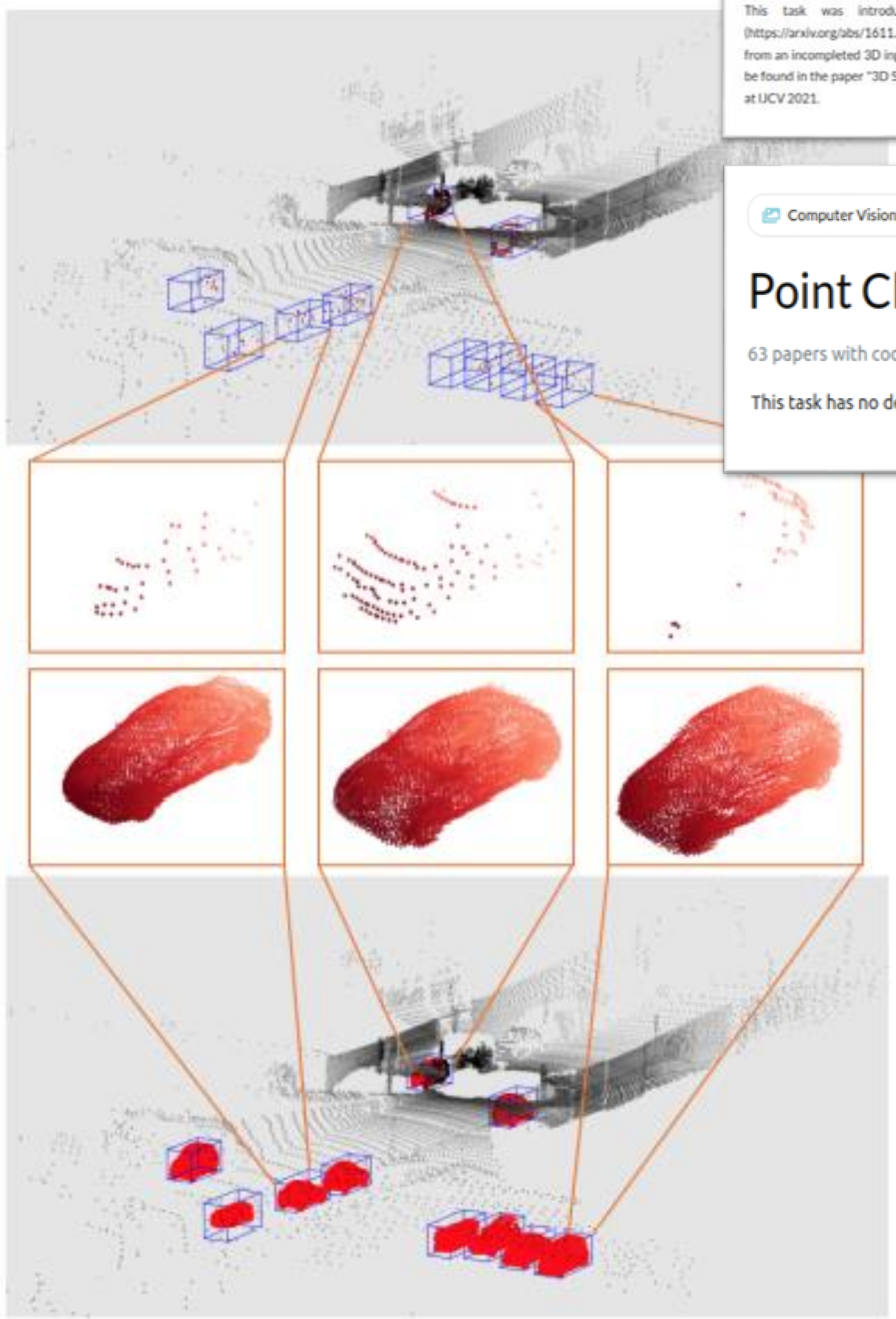
63 papers with code • 3 benchmarks • 4 datasets

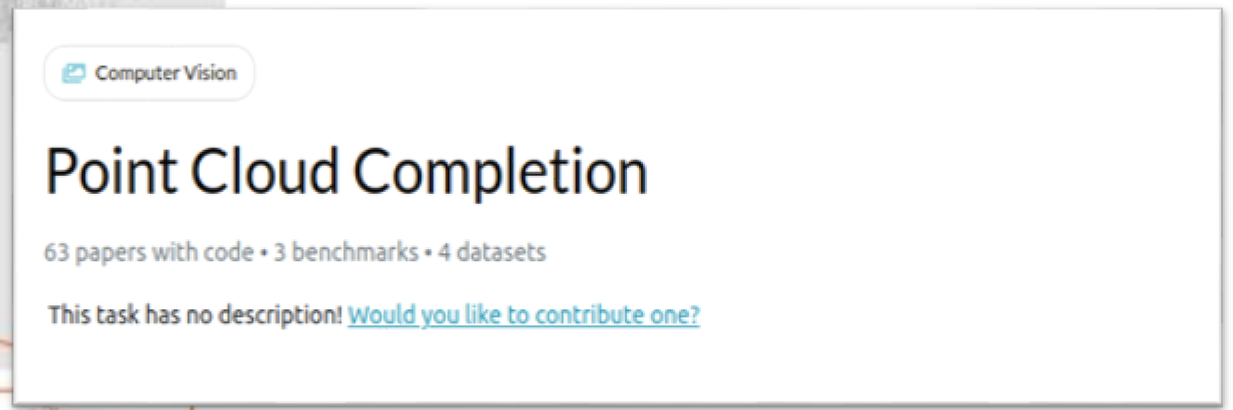
This task has no description! Would you like to contribute one?

United Robots

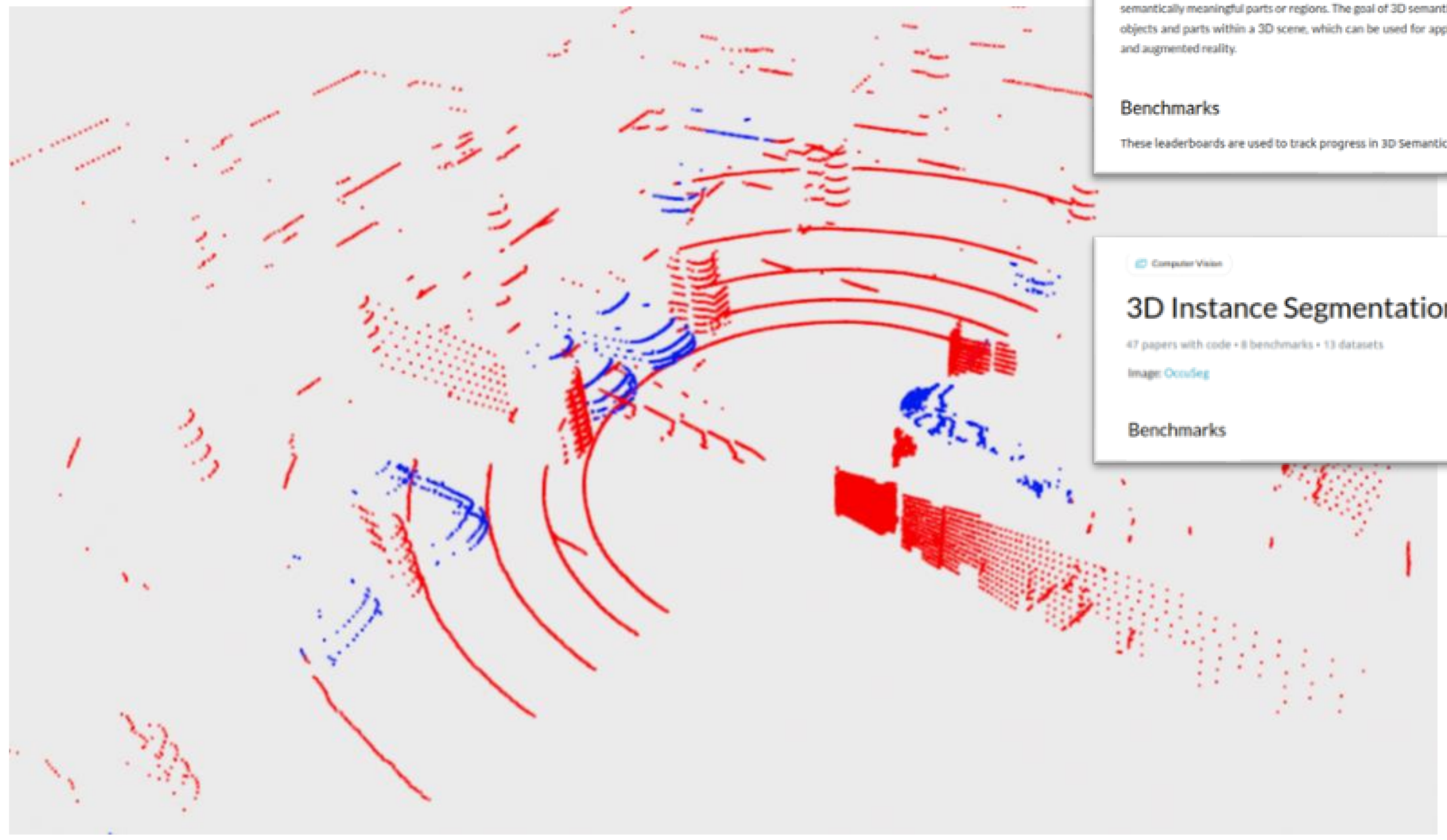
CVLab@PW

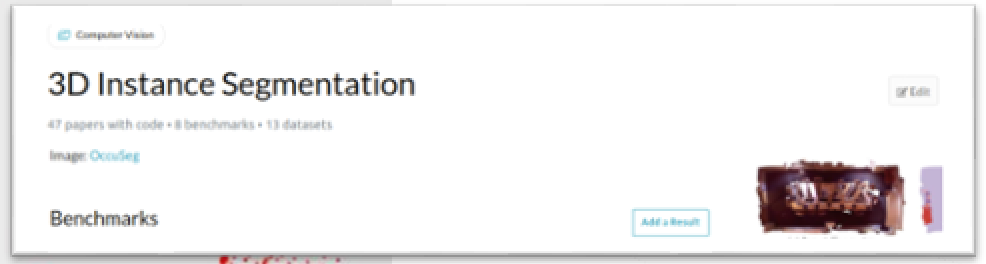
Exemplary applications in robotics



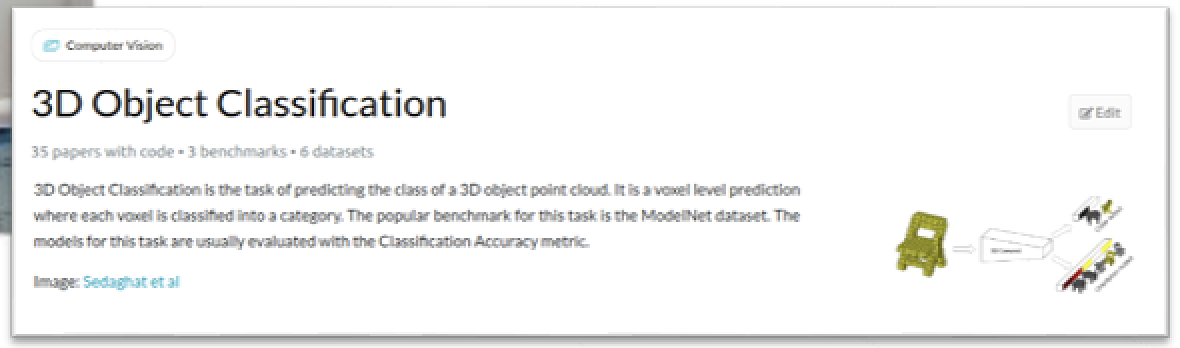


Yuan, W., Khot, T., Held, D., Mertz, C. and Hebert, M., 2018, September. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)* (pp. 728-737). IEEE.





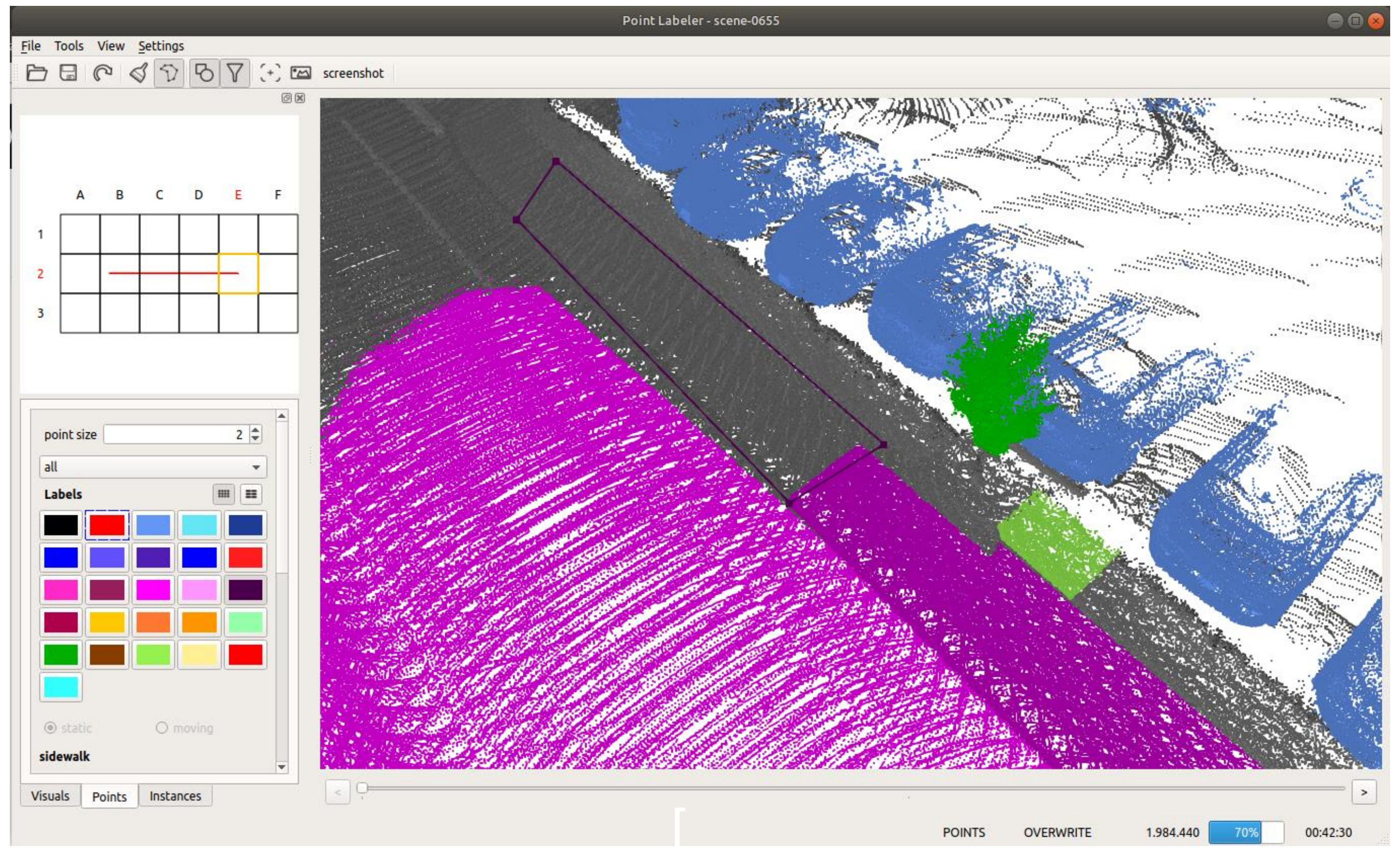
Source: Soltan, S.; Oleinikov, A.; Demirci, M.F.; Shintemirov, A. Deep Learning-Based Object Classification and Position Estimation Pipeline for Potential Use in Robotized Pick-and-Place Operations. *Robotics* **2020**, *9*, 63. https://doi.org/10.3390/robotics9030063



Challenge ?

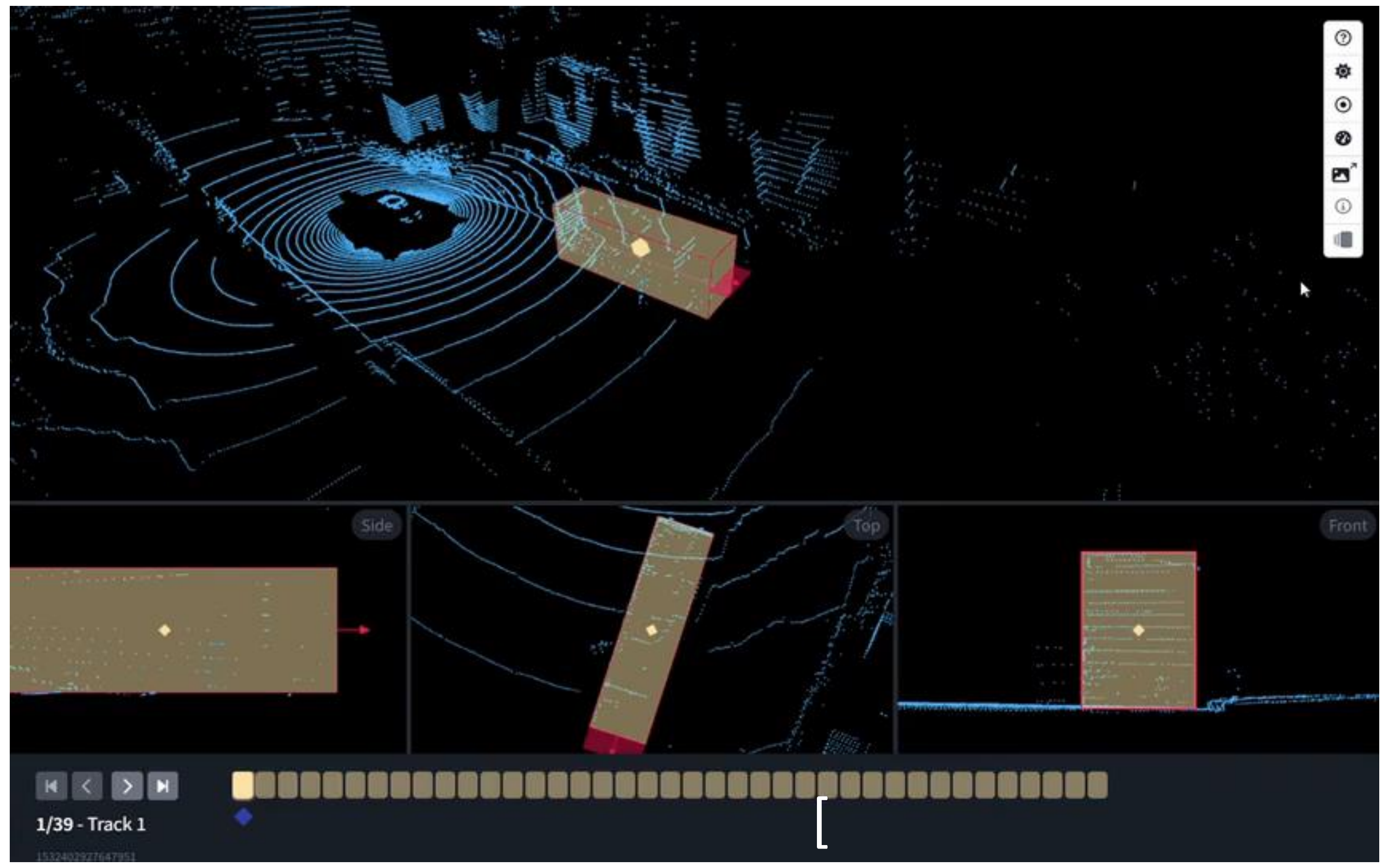
3D point clouds annotation is time consuming

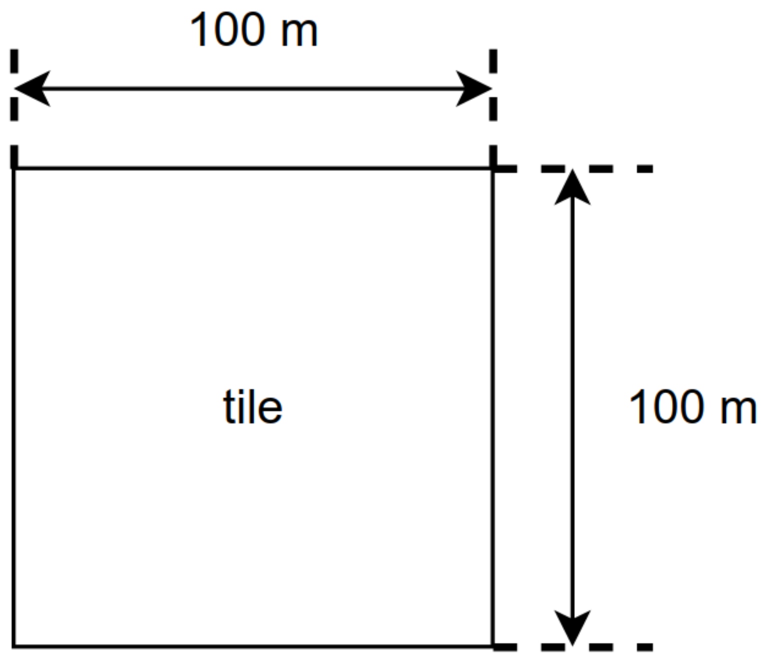
Semantic KITTI Point Labeler (open source)



We provided regular feedback to the annotators to improve the quality and accuracy of labels. Nevertheless, a single annotator also verified the labels in a second pass, *i.e.*, corrected inconsistencies and added missing labels. In summary, the whole dataset comprises 518 tiles and over 1 400 hours of labeling effort have been invested with additional 10 – 60 minutes verification and correction per tile, resulting in a total of over 1 700 hours.

3D point cloud labeling tool by Segments.ai (proprietary)





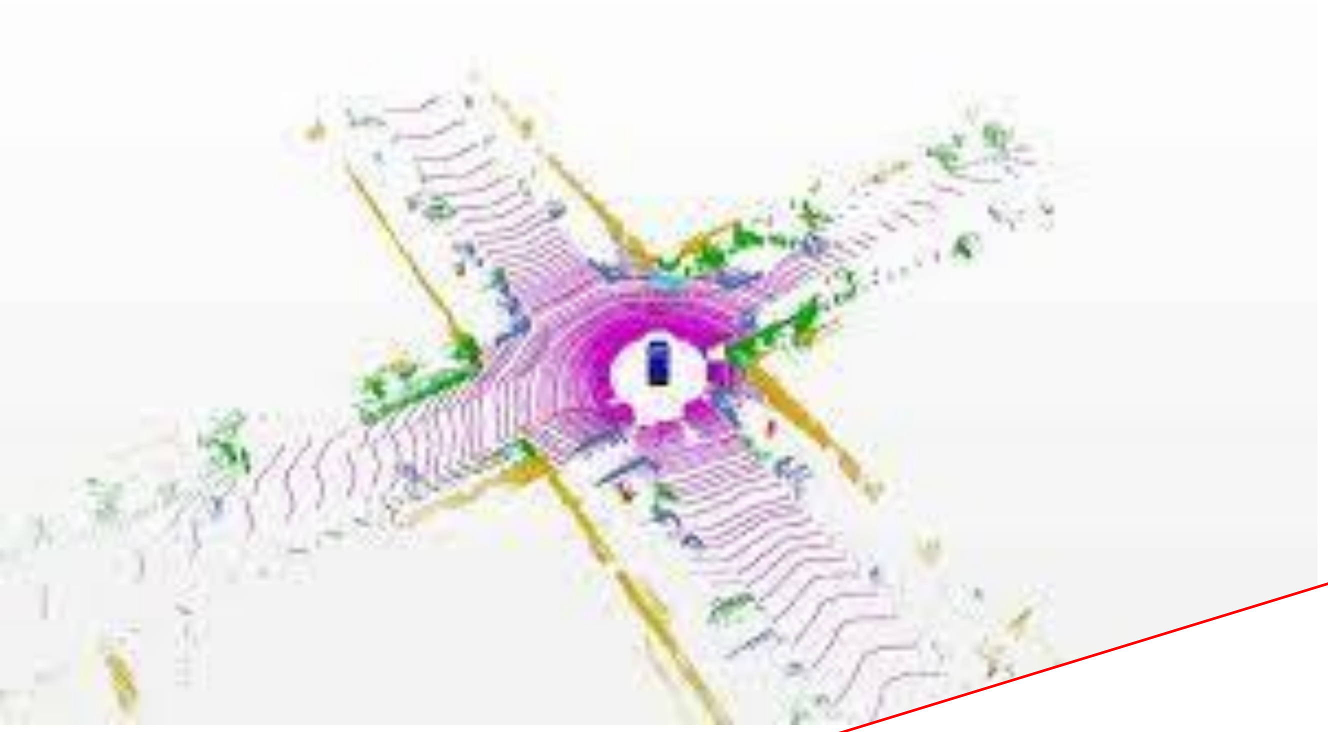
~ 50km trajectory

United Robots

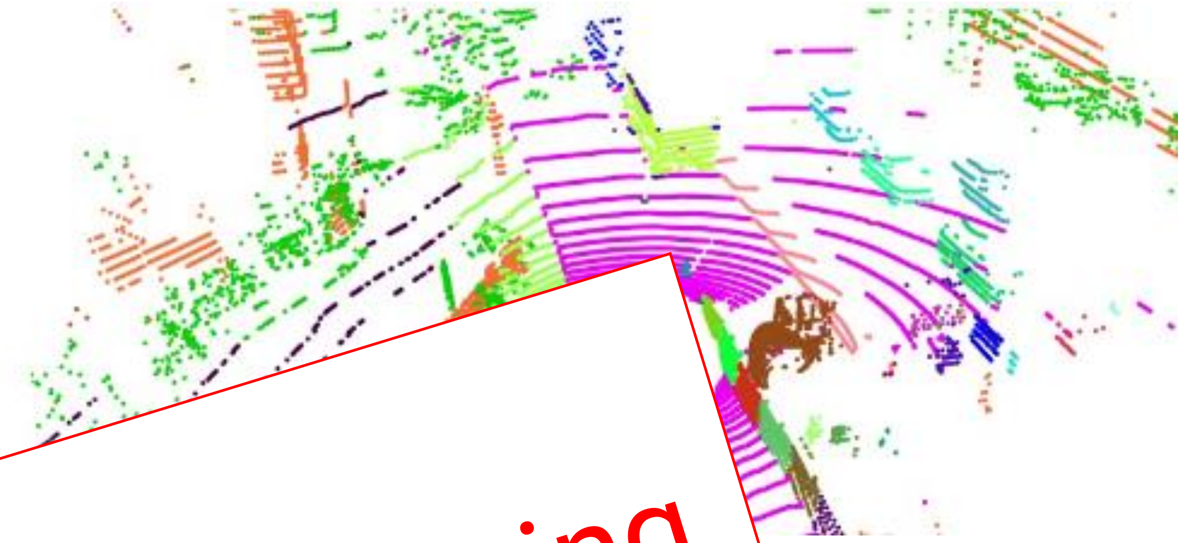
CVLab@PW

Available open source data

KITTI

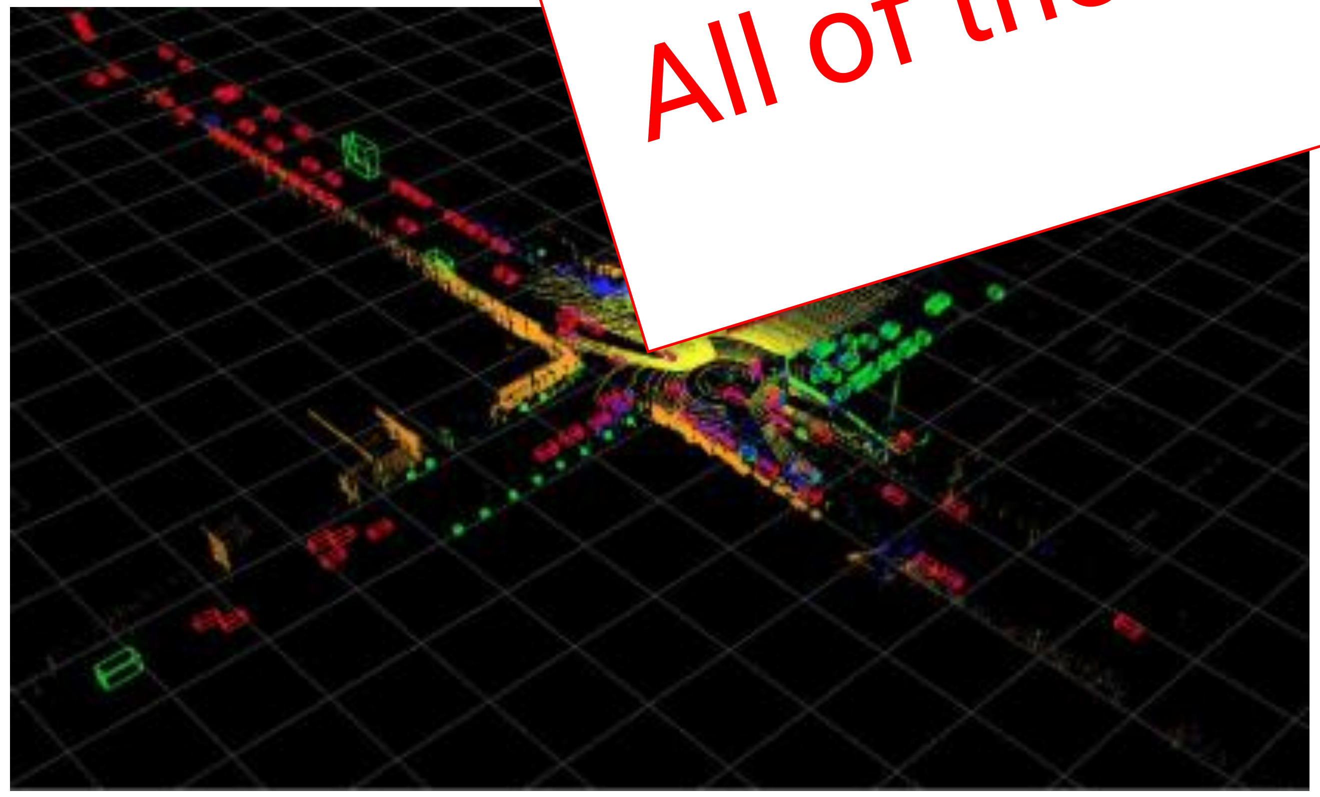


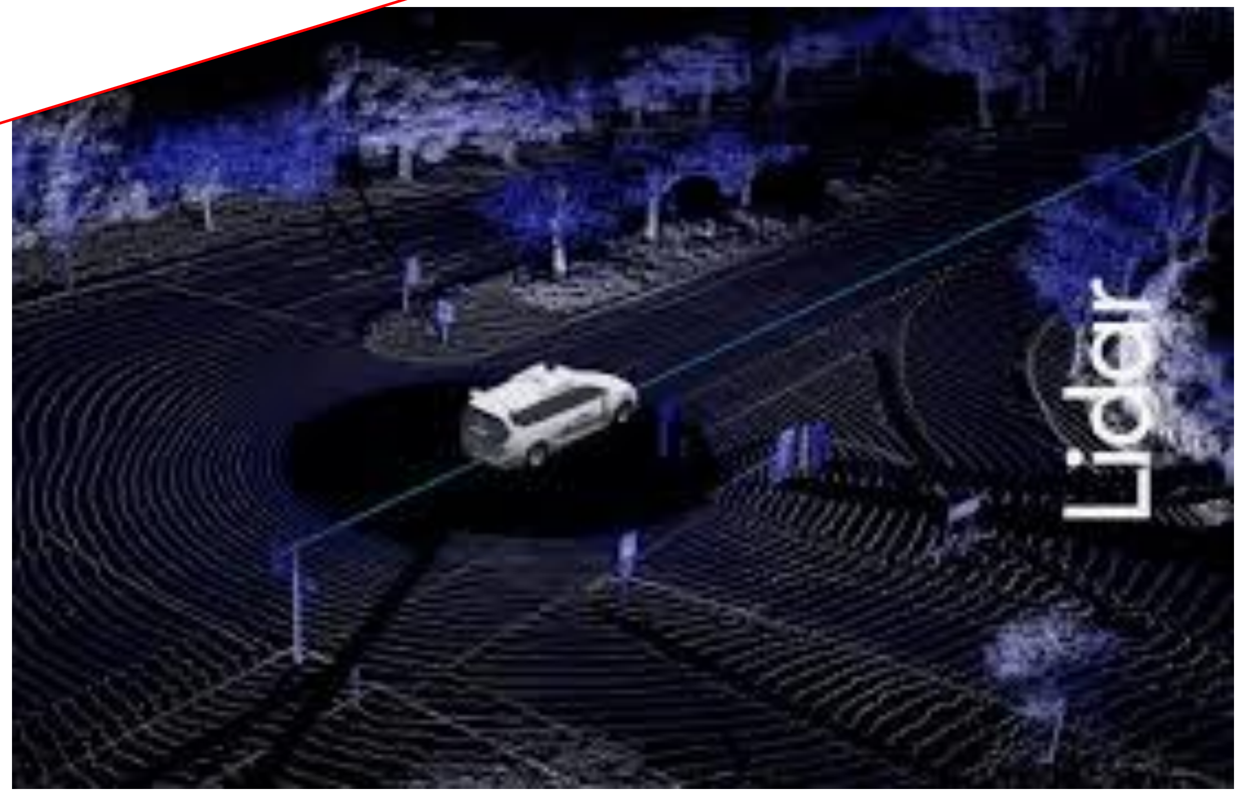
nuScenes



All of them for urban driving

Pandaset

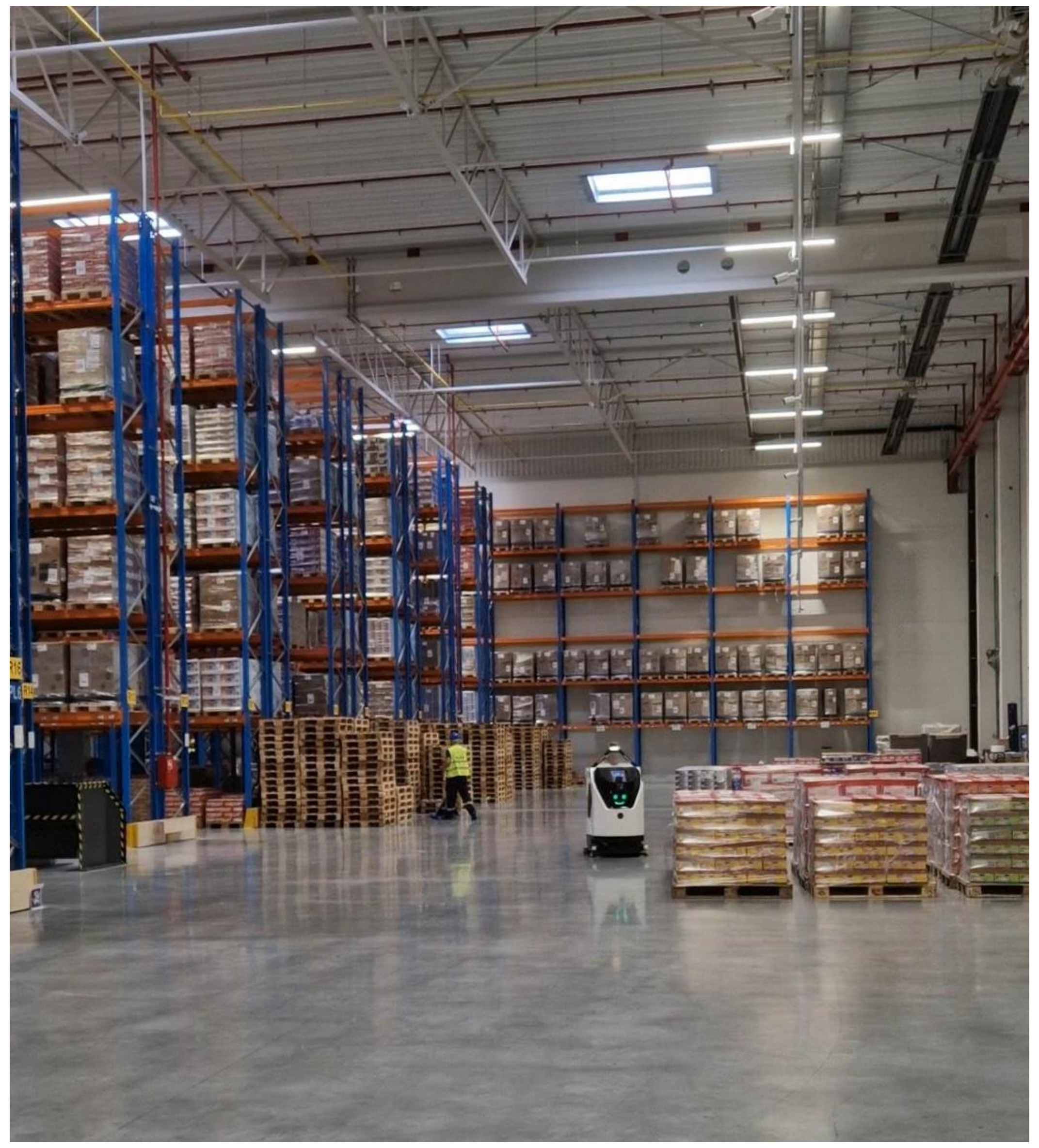


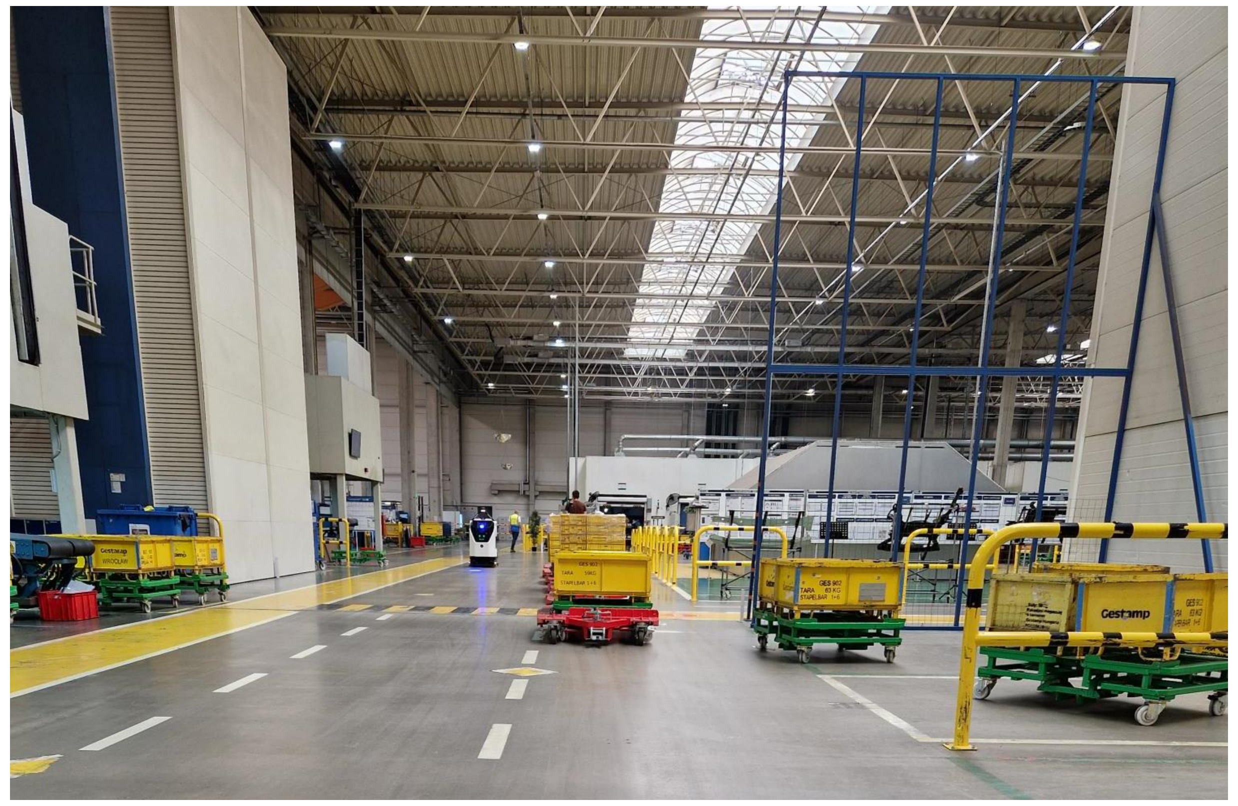


United Robots

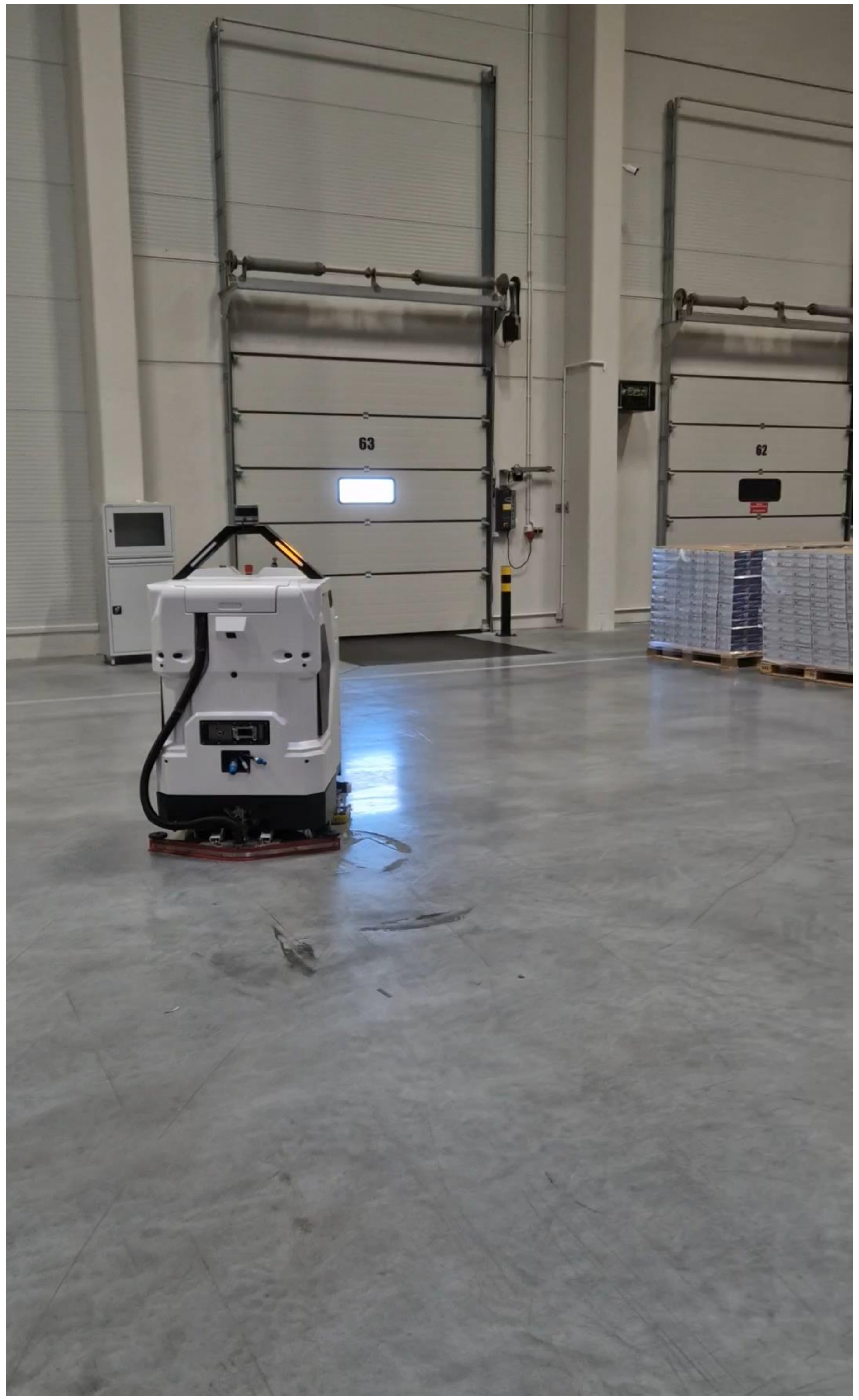
CVLab@PW

What about other environments?





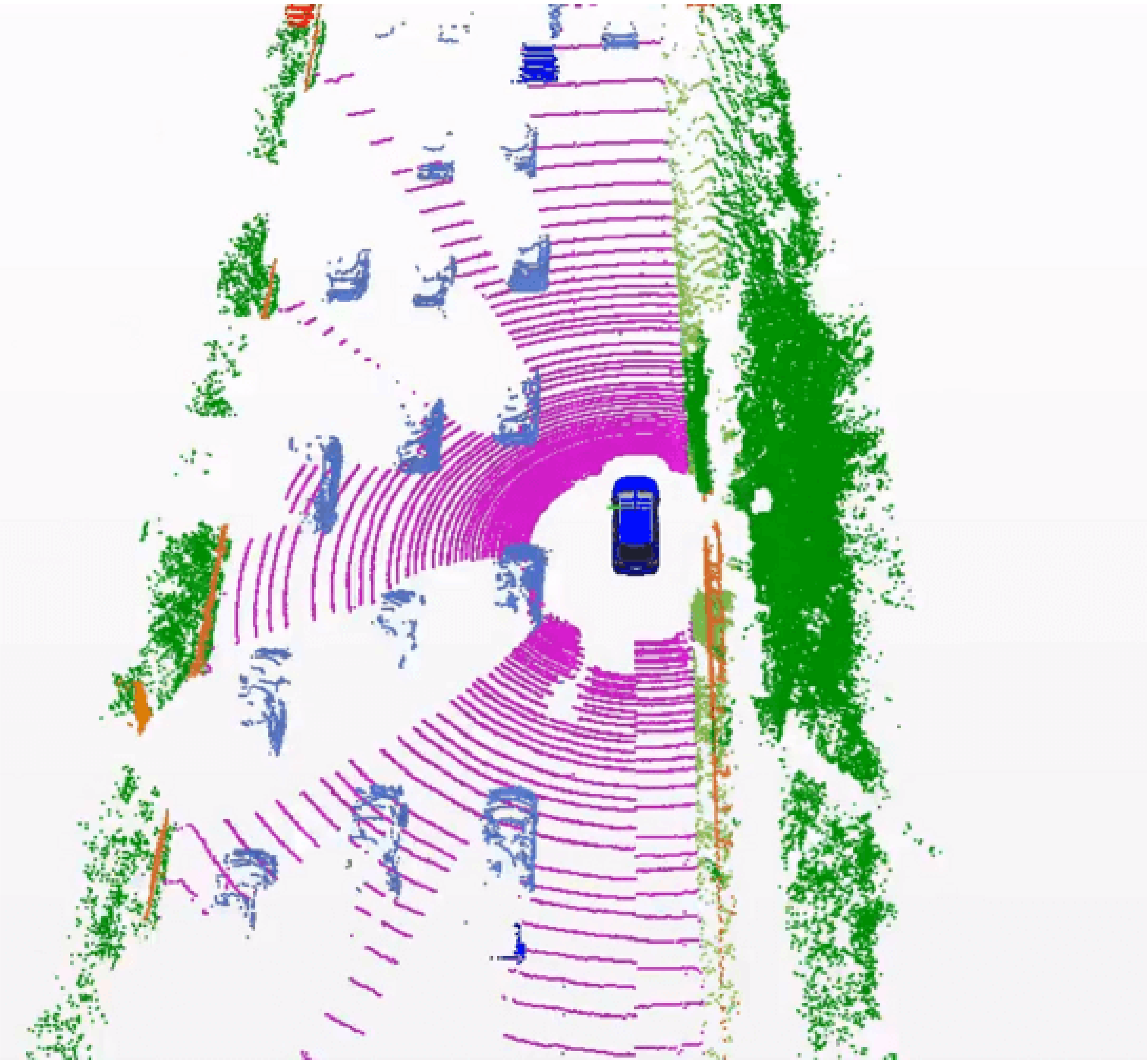
Different perspective: robot navigation use case



United Robots

CVLab@PW

Target: segment out only objects of interest



SphereFormer used for the following tasks

Lai, X., Chen, Y., Lu, F., Liu, J. and Jia, J., 2023. Spherical transformer for lidar-based 3d recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 17545-17555).

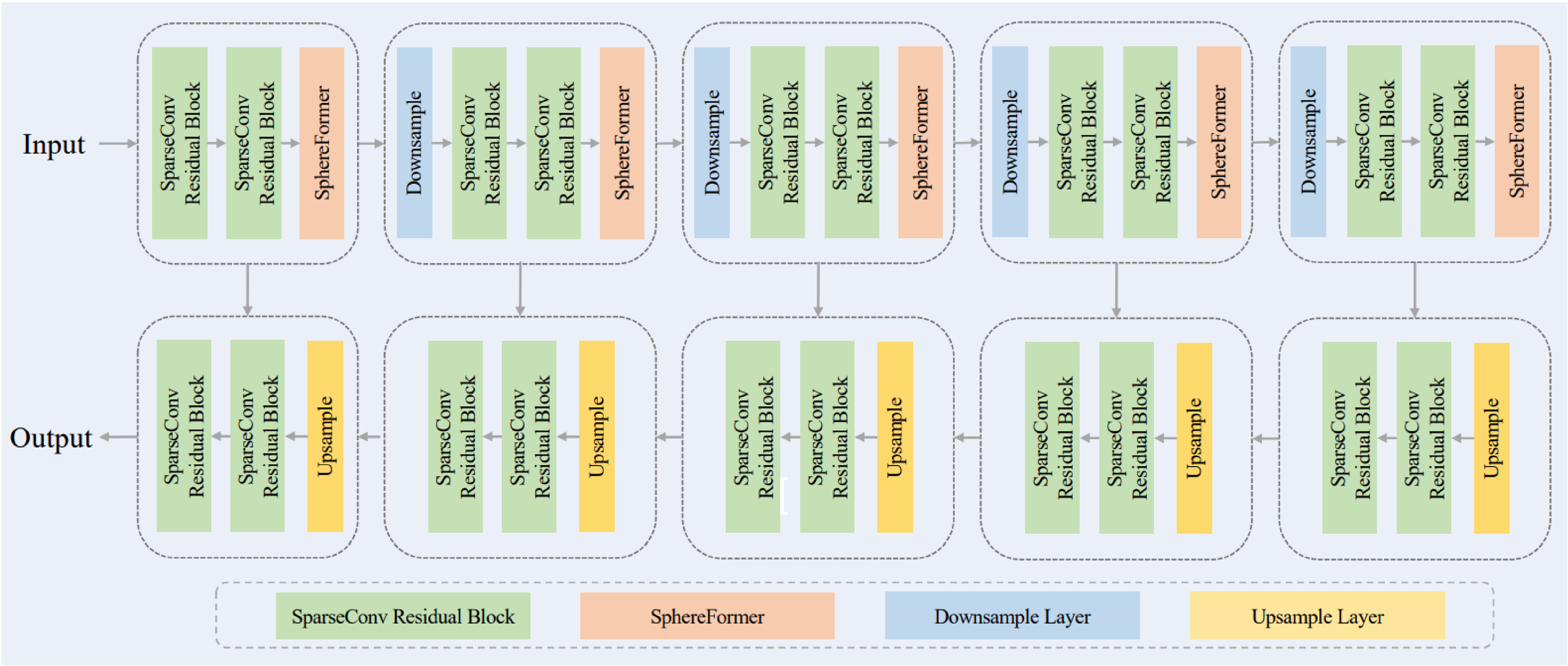
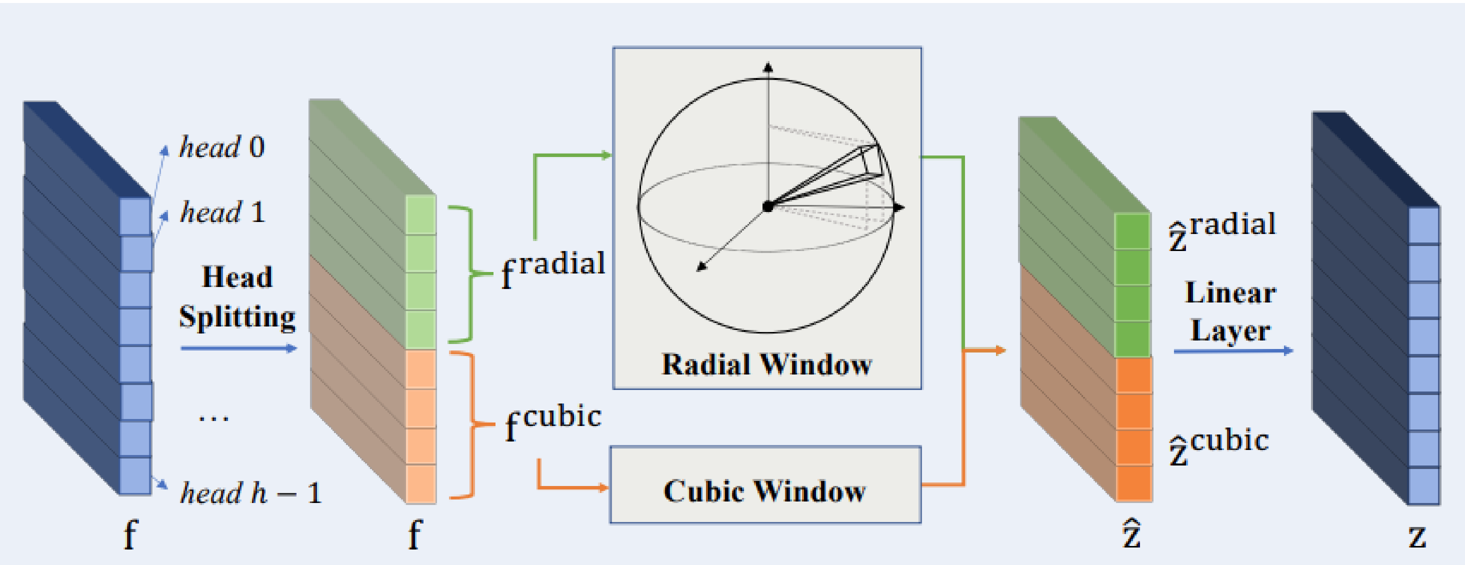


Figure 1. The framework structure. Our proposed module (*i.e.*, SphereFormer) is inserted into the end of each encoding stage.

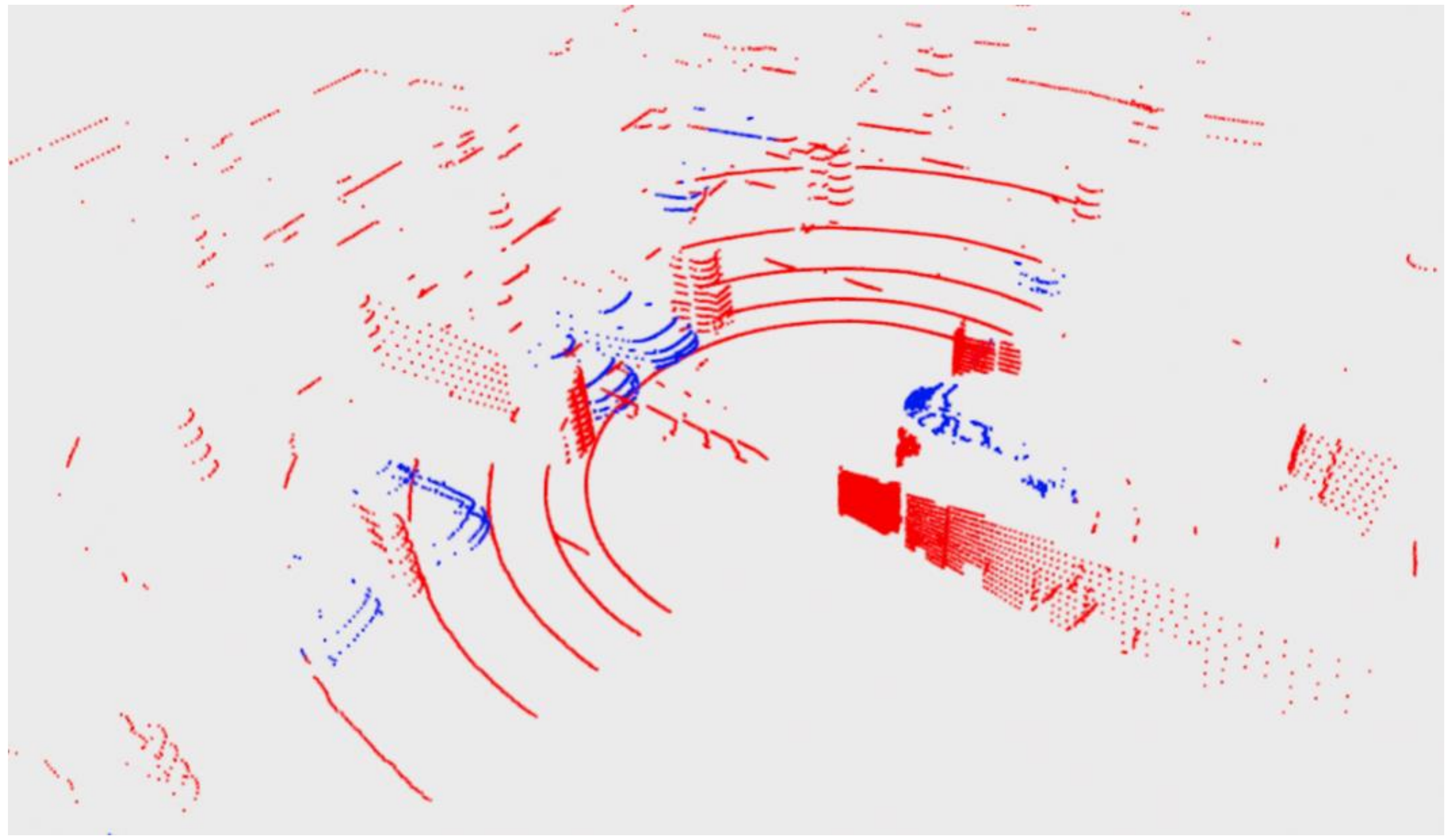


United Robots

CVLab@PW

Proof of Concept – underground parking use case

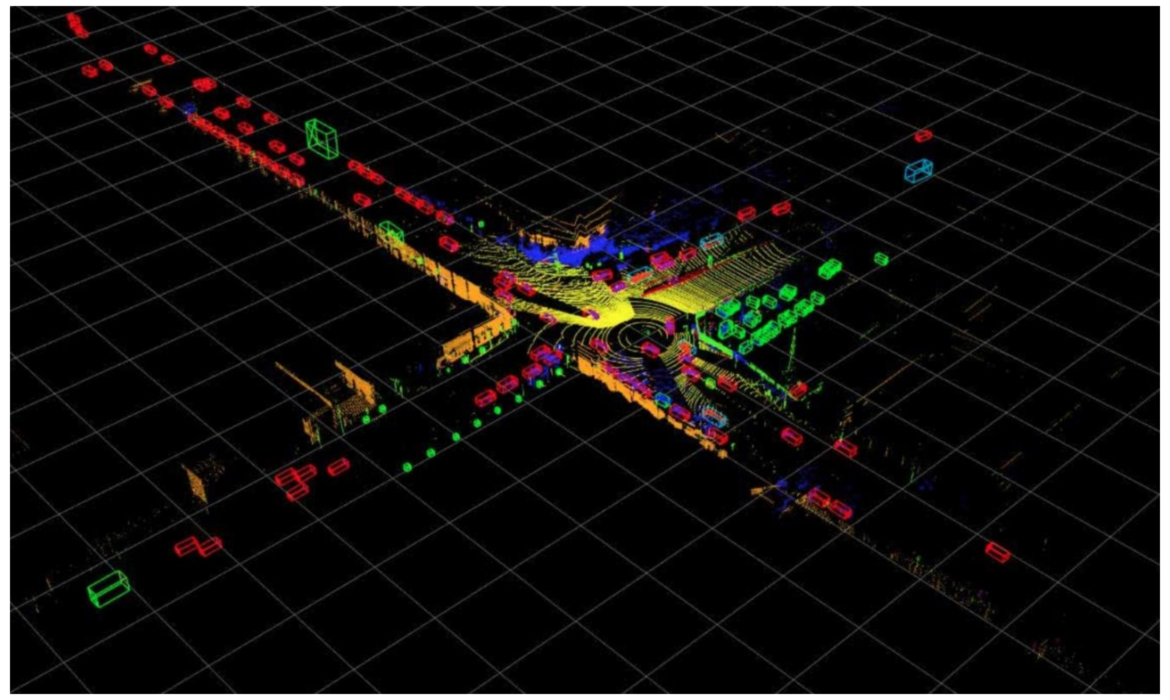
1. Segmenting cars instances from single LiDAR scans
2. Binary segmentation: cars vs everything else
3. No ground truth for training available from actual dataset

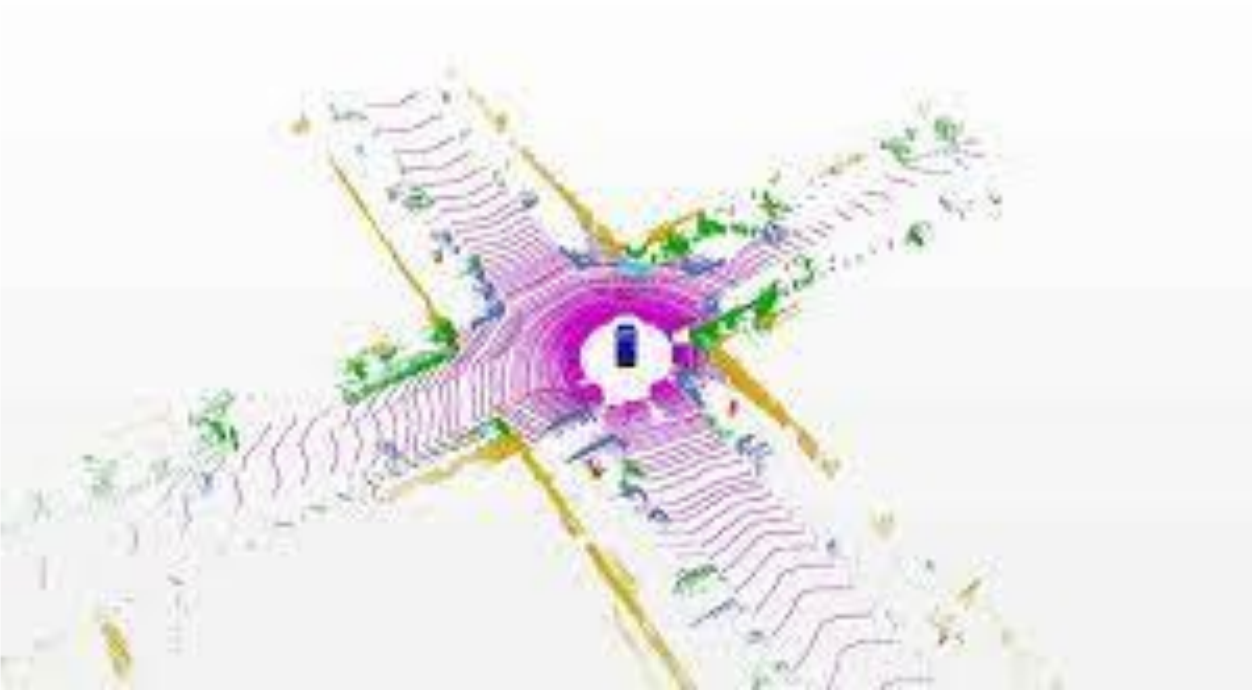


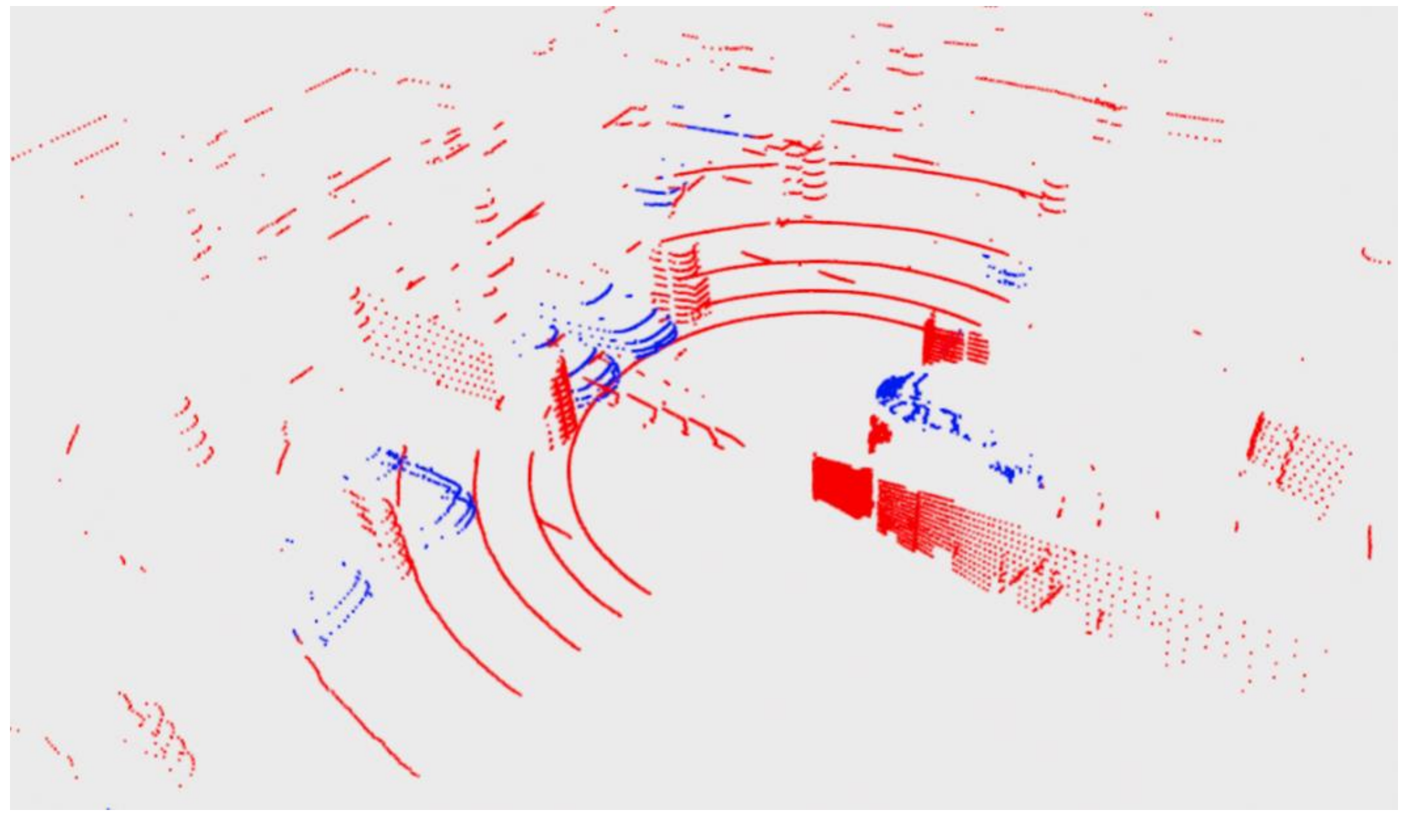
United Robots

CVLab@PW

Approach 1: directly use avaliable open source datasets







Accuracy

Train \ Test	Original KITTI	Original Pandaset	Our dataset
Original KITTI	99.3	87.3	96.1
Original Pandaset	99.6	99.76	90.4

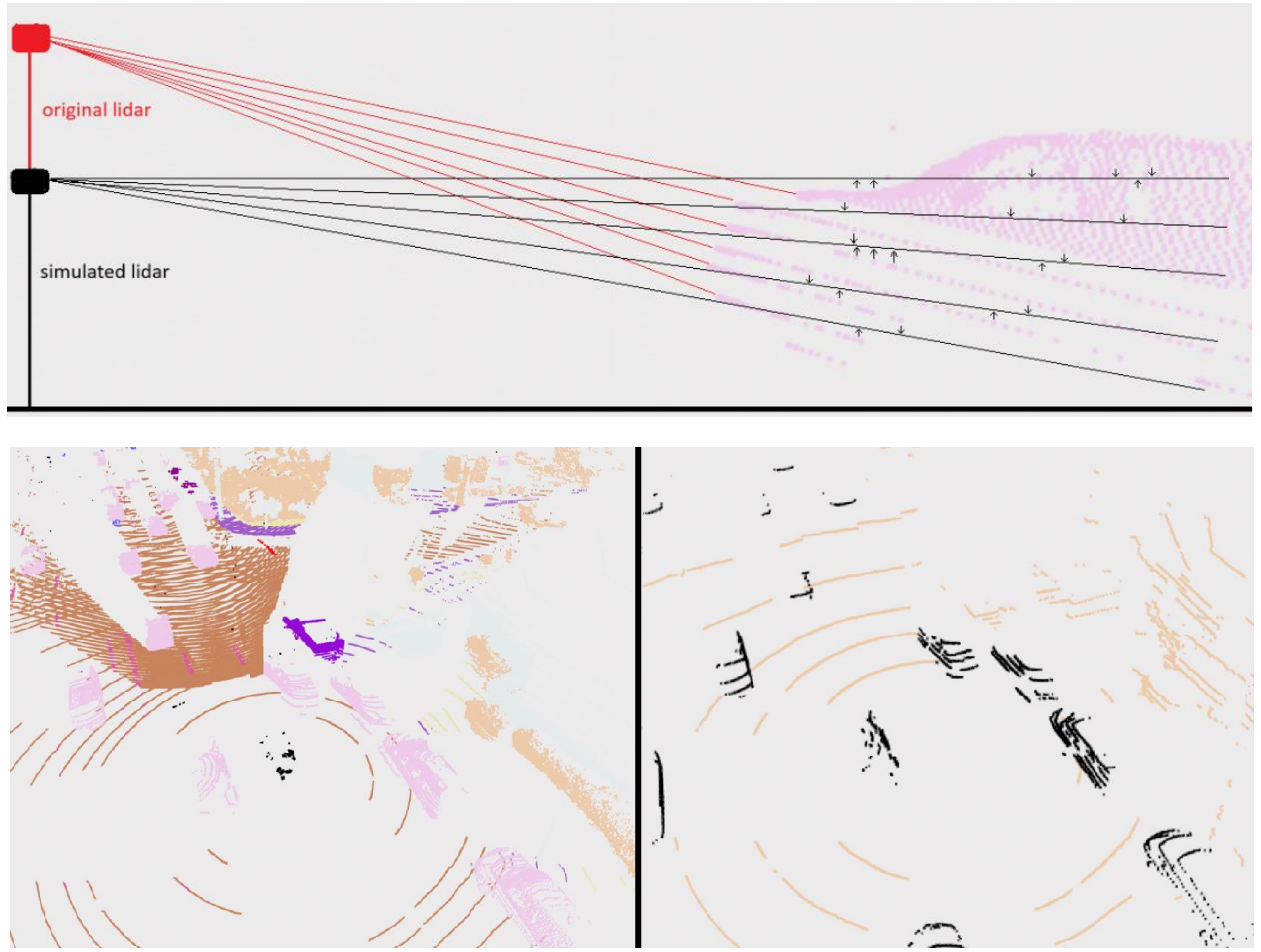
IoU

Train \ Test	Original KITTI	Original Pandaset	Our dataset
Original KITTI	97	86.2	41.4
Original Pandaset	85.9	98	75.9

United Robots

CVLab@PW

Approach 2: preprocess data before using



Preprocessing to align training data with the density of UR data

Accuracy

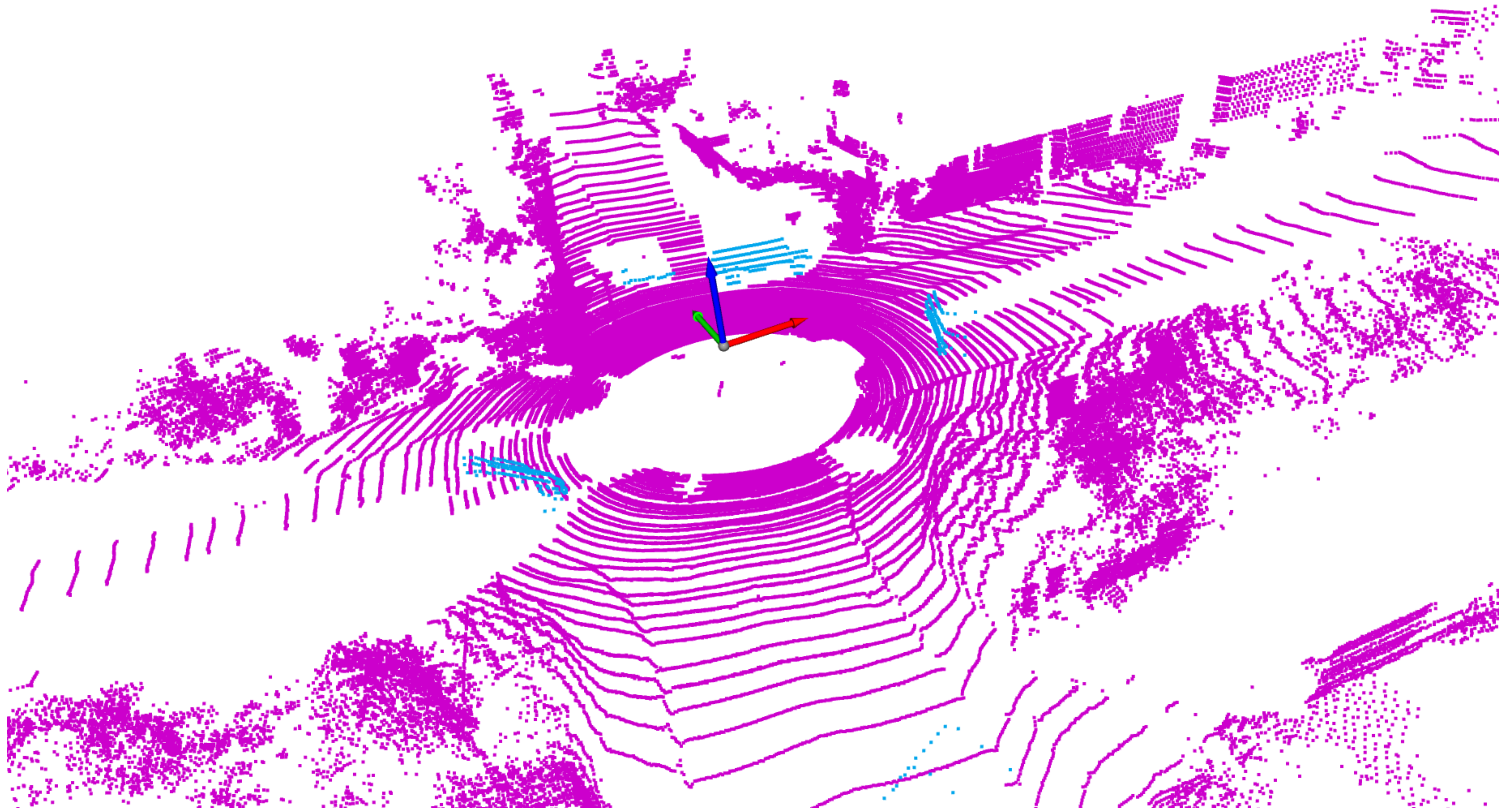
Test Train	Original KITTI	Processed KITTI	Original Pandaset	Processed Pandaset	Our dataset
Original KITTI	99.3	86.7	87.3	78.6	96.1
Processed KITTI	6.3	97.3	16.5	84.3	87.6
Original Pandaset	99.6	87.9	99.76	85.2	90.4
Processed Pandaset	38.8	90.5	44.3	98.3	86.2

IoU

Test Train	Original KITTI	Processed KITTI	Original Pandaset	Processed Pandaset	Our dataset
Original KITTI	97	36.8	86.2	66.2	41.4
Processed KITTI	5.4	88.2	16.5	82	76.3
Original Pandaset	85.9	62.9	98	79.9	75.9
Processed Pandaset	0.1	56.7	35.1	89.8	69.6

United Robots CVLab@PW

Approach 3: insert object instances from other dataset



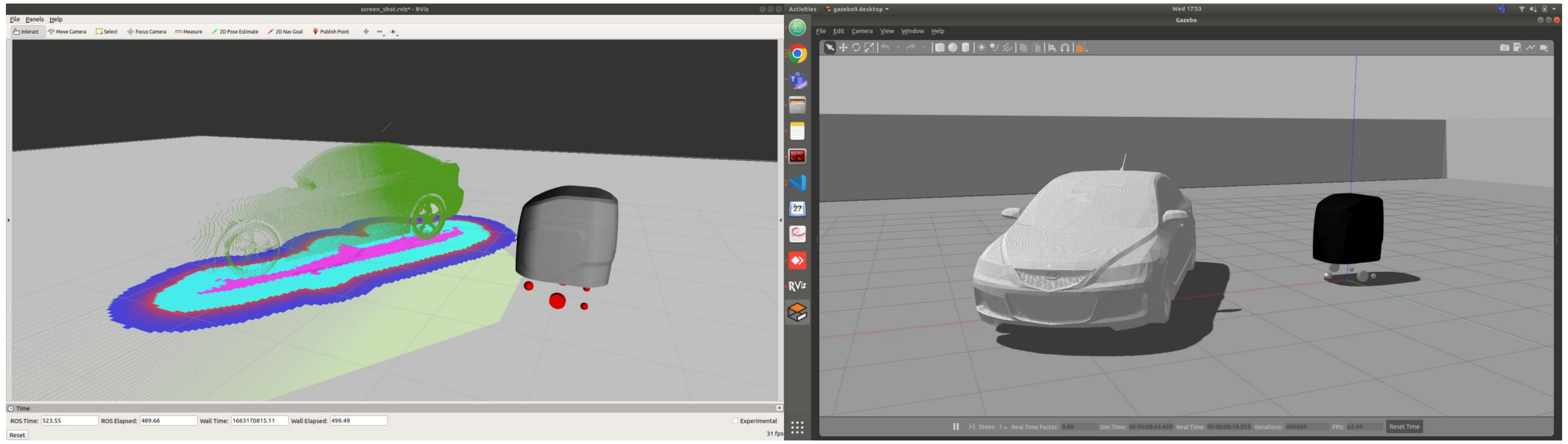
Inserting object instances from one dataset into background from another.

	Accuracy KITTI	IoU KITTI	Accuracy UR	IoU UR
Processed KITTI	97.3	88.2	87.6	76.3
Enhanced KITTI	98.1	91.1	90.7	62.7

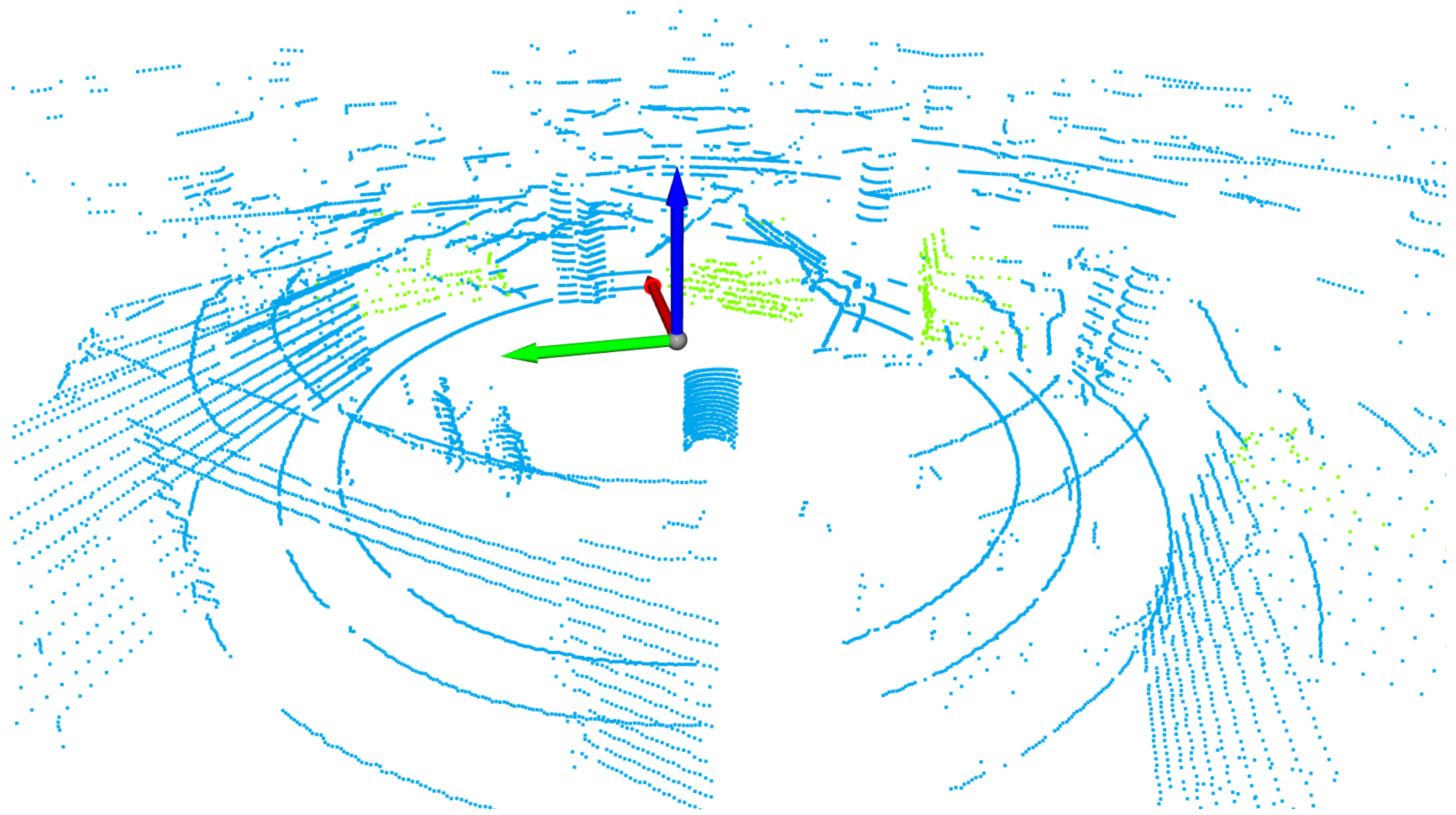
United Robots

CVLab@PW

Approach 4: naive synthetic instances



Simulating sensor returns using mesh models of cars



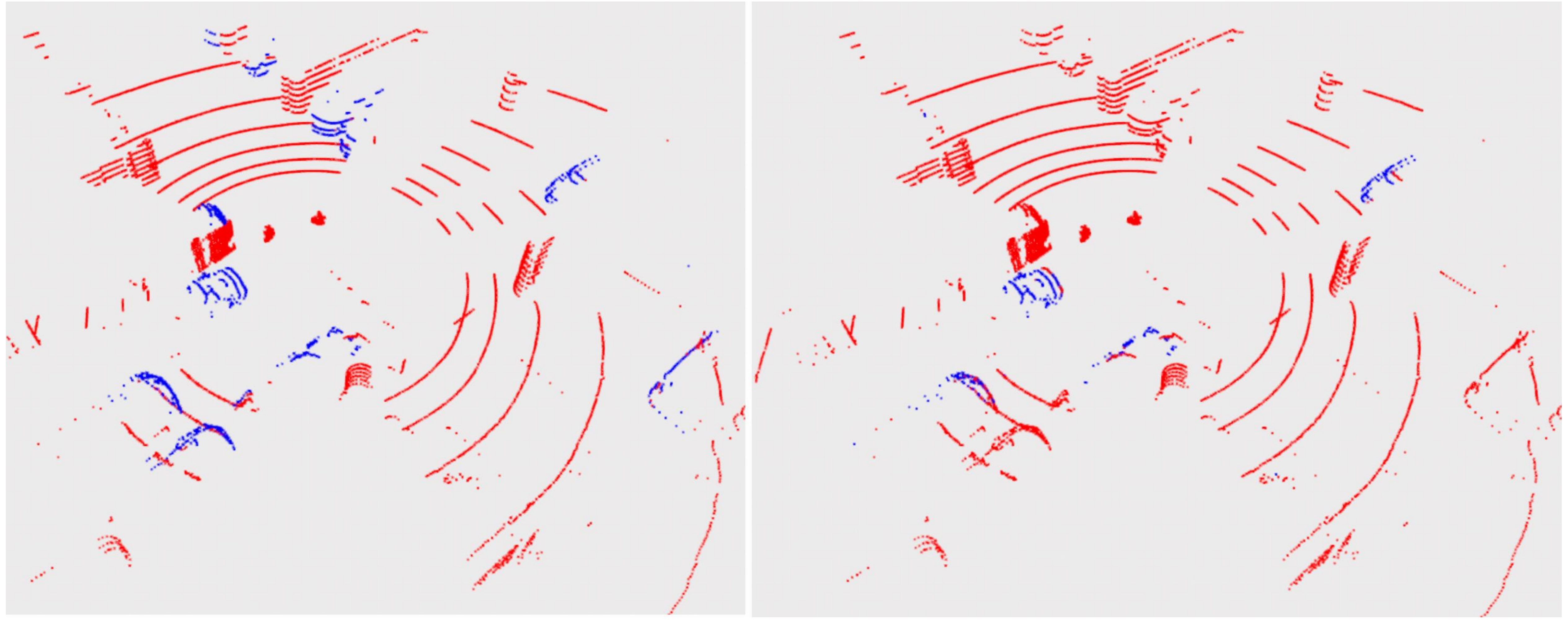
	Accuracy	IoU
Processed KITTI	87.6	76.3
Original Pandaset	90.4	75.9
Synthetic	54.7	51.9

United Robots

CVLab@PW

Approach 5: augmented synthetic instances

1. Noise as normal distribution of fluctuations of each point
2. Dropout of random number of points
3. Shift up/down, left/right of complete point cloud



Results of segmentation after synthetic data augmentation. Ground truth (left), segmentation result (right).

	Accuracy	IoU
Processed KITTI	87.6	76.3
Original Pandaset	90.4	75.9
Synthetic	54.7	51.9
Synthetic (noise augmentations)	58.3	52.8
Synthetic (full augmentation)	63.3	56.1

United Robots

CVLab@PW

Open point: advanced modelling of noise

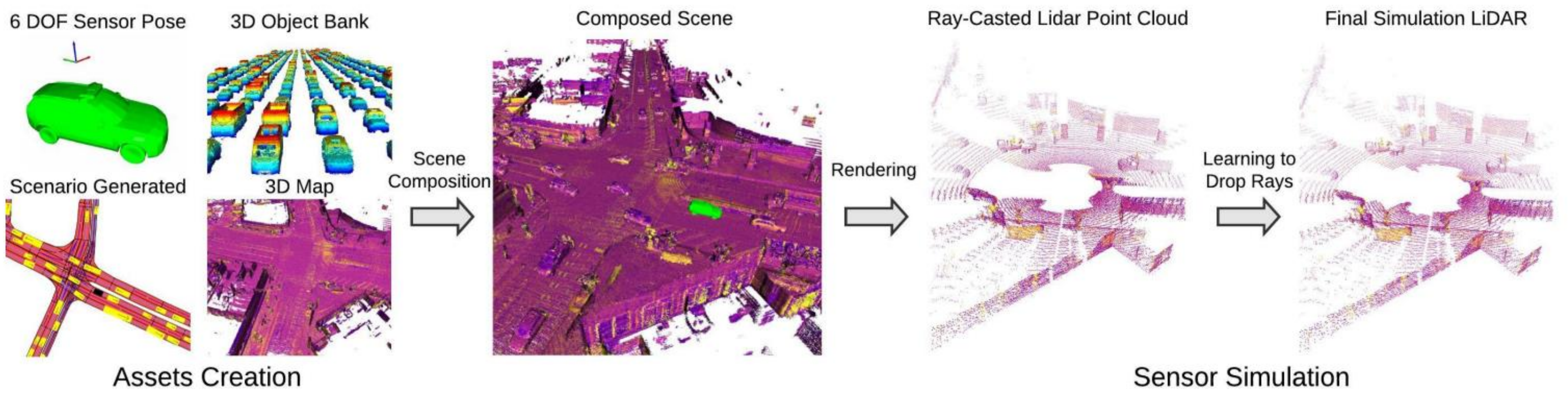


Figure 2: **LiDARsim Overview Architecture:** We first create the assets from real data, and then compose them into a scene and simulate the sensor with physics and machine learning.

Train Set	Overall	Vehicle	Background
CARLA[47] (Baseline)	0.65	0.36	0.94
LiDARsim (Ours)	0.89	0.79	0.98
SemanticKITTI (Oracle)	0.90	0.81	0.99

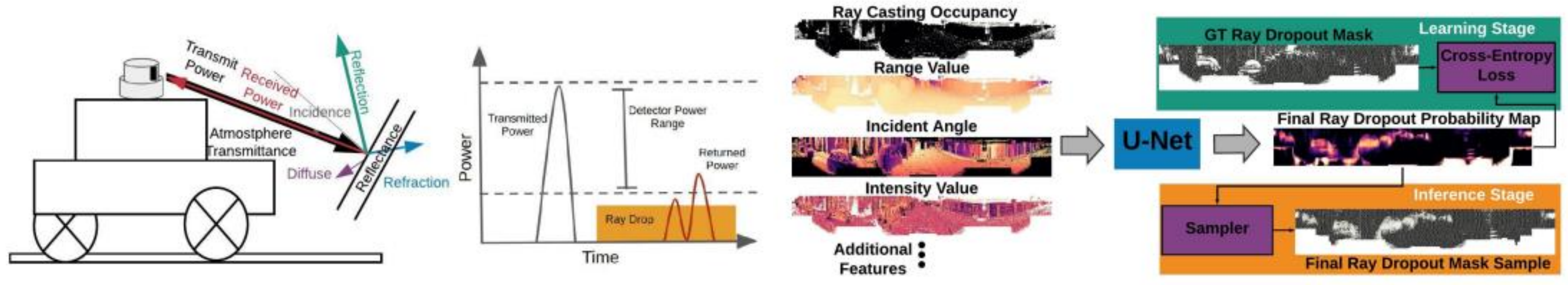


Figure 6: **Left**: Raydrop physics explained: Multiple real-world factors and sensor biases determine if the signal is detected by LiDAR receiver. **Right**: Raydrop network: Using ML and real data to approximate the raydropping process.

Manivasagam, Sivabalan, et al. "Lidarsim: Realistic lidar simulation by leveraging the real world." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

United Robots

CVLab@PW

The actual challenge

~~Advanced algorithms for perception~~

Lack of sufficient training data

Prohibitively time-consuming data tagging

United Robots

CVLab@PW

Thanks for your attention!

Reach me out for more topics related to ML in robotics

United Robots

CVLab@PW