SELF-DRIVING CARS
PERCEPTION

Photogrammetry & Robotics Lab

Perception for Self-Driving Cars LiDAR-based Perception

Jens Behley

Part of the Course: Techniques for Self-Driving Cars by
C. Stachniss, J. Behley, N. Chebrolu, B. Mersch, L. Peters, I. Bogoslavskyi

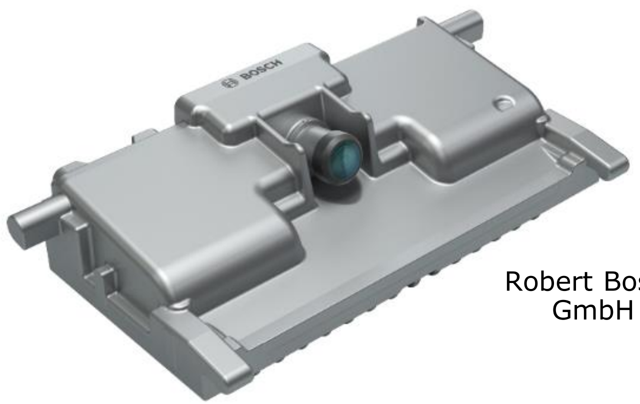
Last Lecture

- Perception stack & sensors
- Common perception tasks
  - Detection
  - Semantic & Panoptic Segmentation
- Challenges in visual perception
- Future directions for visual perception

Perception Suite

- Common Perception Suite
  - (Stereo) Camera
  - LiDAR
  - Radar

Pros & Cons



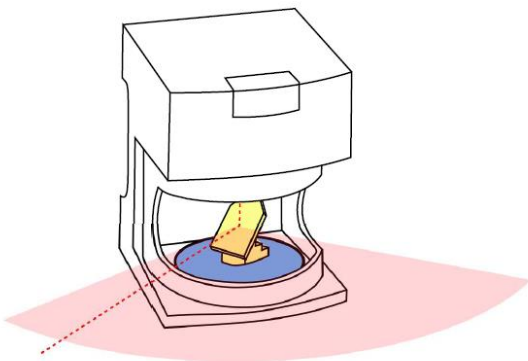
Robert Bosch GmbH

Velodyne Lidar, Inc.

+ color
+ cheap
+ high-resolution
- Affected by illumination

+ Direct 3d representation
+ Unaffected by illumination
- Affected by reflectance properties of objects
- expensive (currently)

LiDAR: Basics



SICK 2D LiDAR

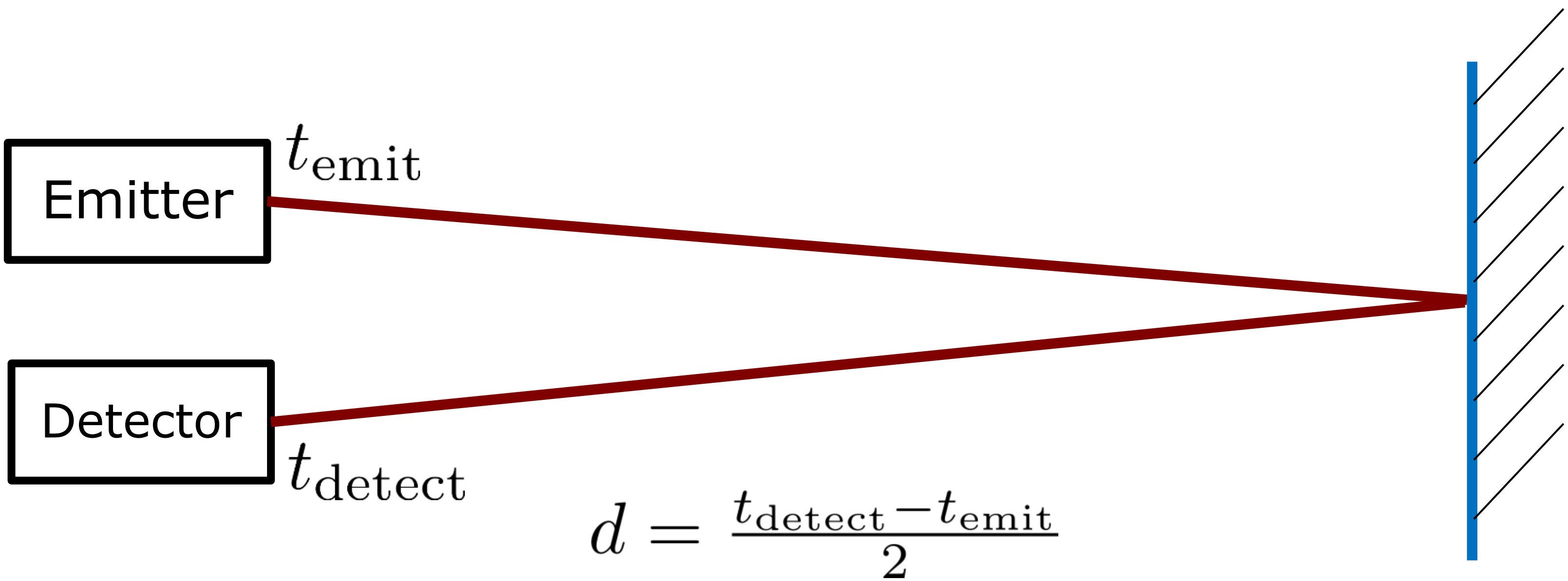
Fraunhofer IAIS – 3DLS-K

Velodyne
Lidar, Inc.

IBEO
automotive

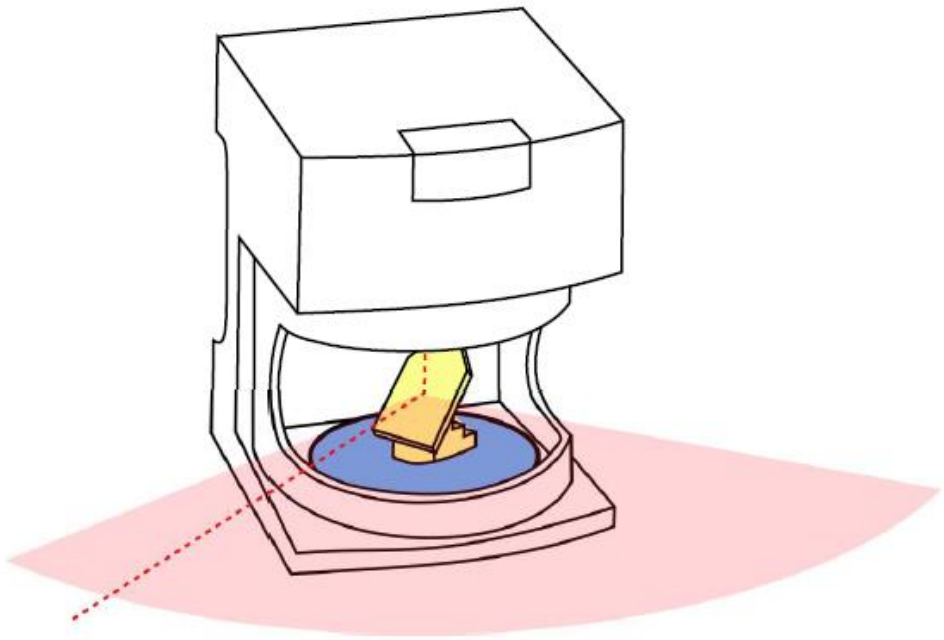
- Main types of LiDAR sensors
  - Mechanical LiDAR (e.g., SICK, Velodyne, Ouster, etc.)
  - Solid-state LiDAR (e.g. IBEO, Livox, etc.)

Measurement Principle

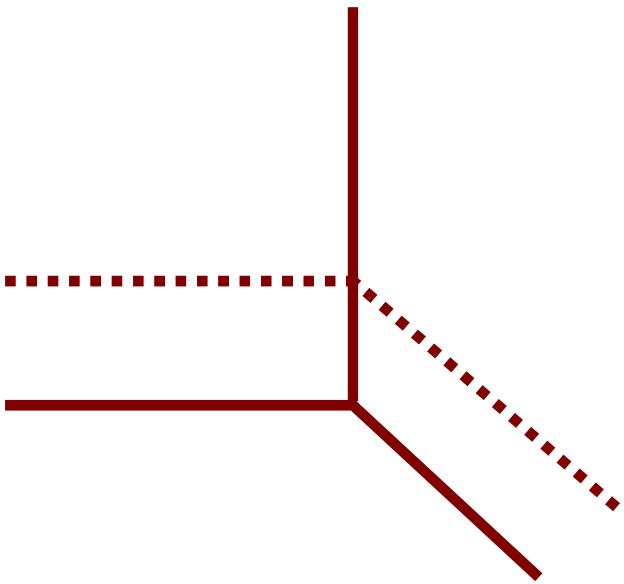


- Time-of-flight of photons (900 nm-1550 nm) from emittance to detection provides distance measurement *d*
- Amount of detected photons corresponds to *remission* or intensity

2D LiDAR in Robotics



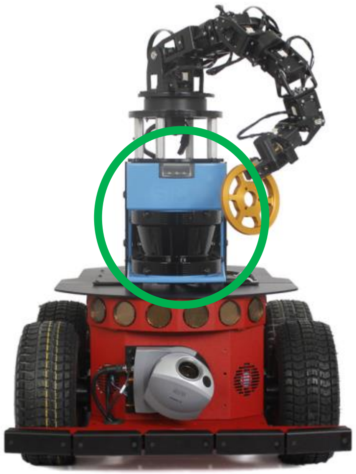
SICK 2D LiDAR



- One beam that is directed using a mirror
- Orientation of mirror measured with optical encoder disc
- Provides “slice” of the environment

2D LiDARs for Indoor Robotics

Rhino at Deutsche Museum Bonn (1998)



Pioneer robot with SICK LMS

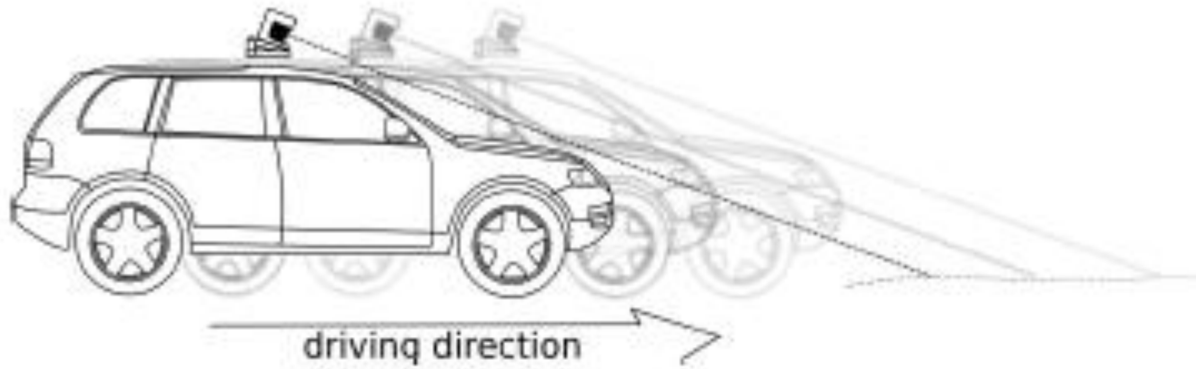
© IEEE

- Many systems use 2D LiDAR for indoor perception and navigation
- For navigation on flat ground
  → slice of measurements often sufficient

[Burgard, 1999]

Towards 3D LiDARs

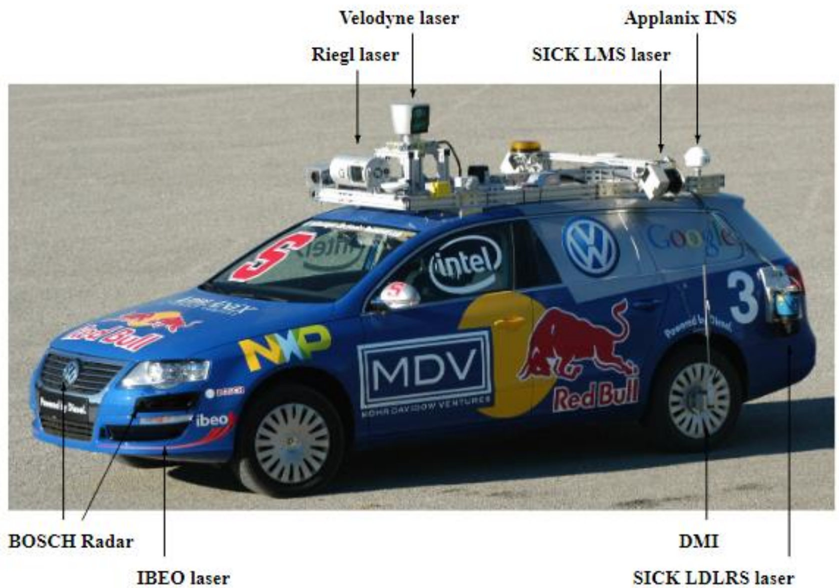
Stanley@DARPA Grand Challenge



- Uneven ground in outdoor environments
- Sweep 2D LiDAR over ground for offroad perception at DARPA Grand Challenge

[Thrun, 2006]

3D LiDARs at Urban Challenge



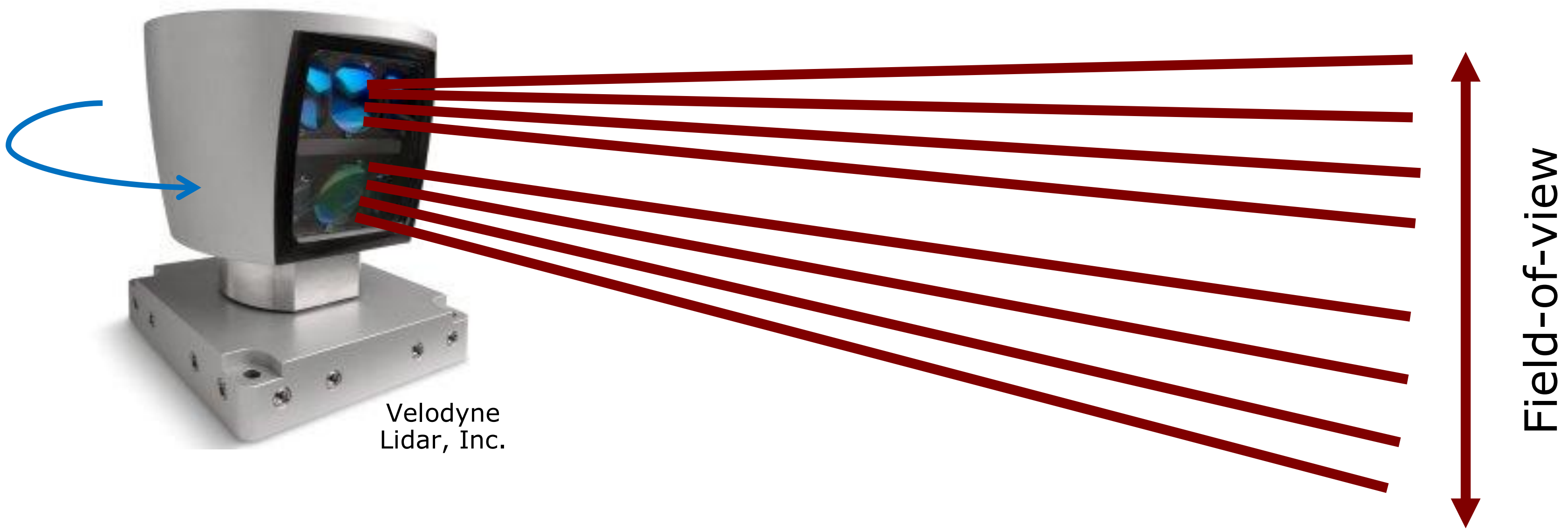
Stanford's Junior

CMU's Boss

- Success of LiDARs at Grand Challenge inspired development of rotating multi-beam LiDARs
- Almost all cars featured a full 3D LiDAR at DARPA Urban Challenge

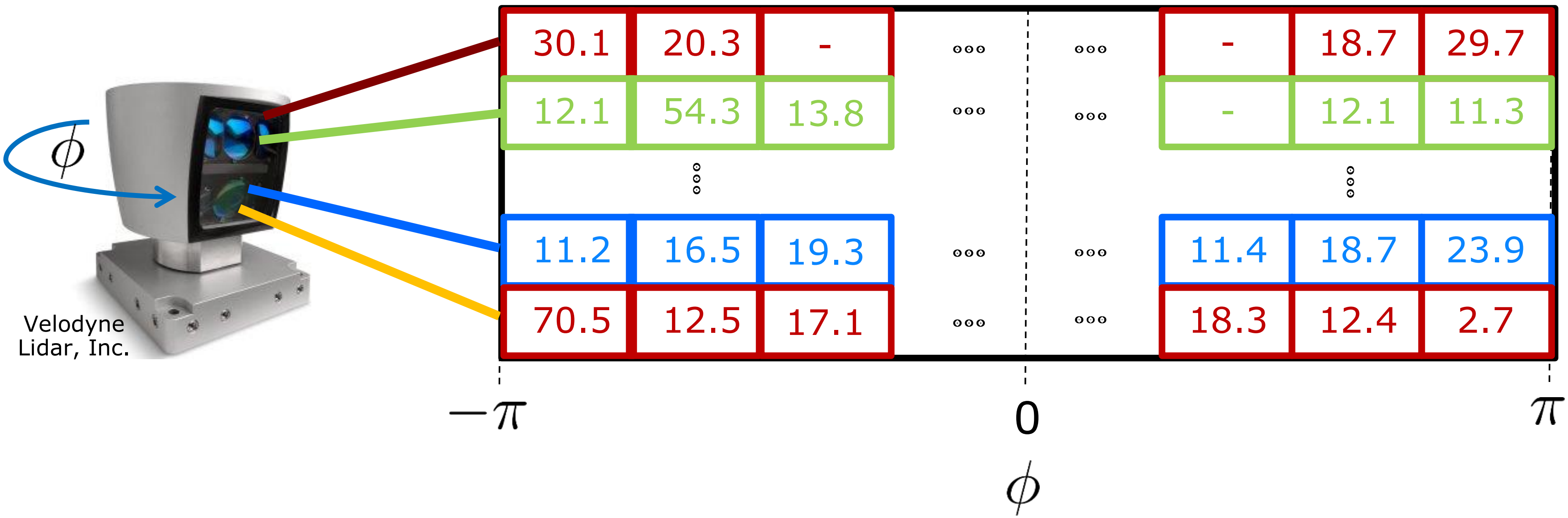
[Montemerlo, 2008] [Urmson, 2008]

Multi-beam LiDARs



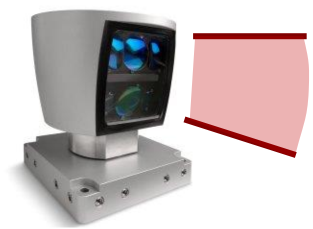
- Rotating multi-beam LiDARs provide 360° degree horizontal field-of-view

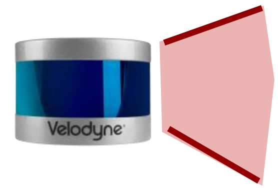
Range Image

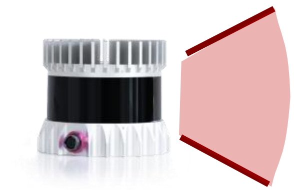


- Rotating LiDAR provides range image with ranges for each beam and turn angles ϕ
- Common range image widths: 1024, 2048

Mechanical LiDARs: Examples







Model	Velodyne HDL-64E	Velodyne VLP-16	Velodyne VLS-128	Ouster OS1-128
Year	2007	2014	2019	2019
#Beams	64	16	128	128
Vert. FoV	24.7	30.0	40.0	45.0
Freq. (Hz)	5-20	5-20	5-20	10/20
Price	~$75,000	~$8,000	?	~$18,000

Images and specifications from manufacturers.

Solid-state LiDAR sensors

LIVOX

IBEO automotive

Velodyne Lidar Inc.

- Fixed laser diodes reduces effects of wear of mechanical parts
- Use electromechanical and optical effects to direct laser beam

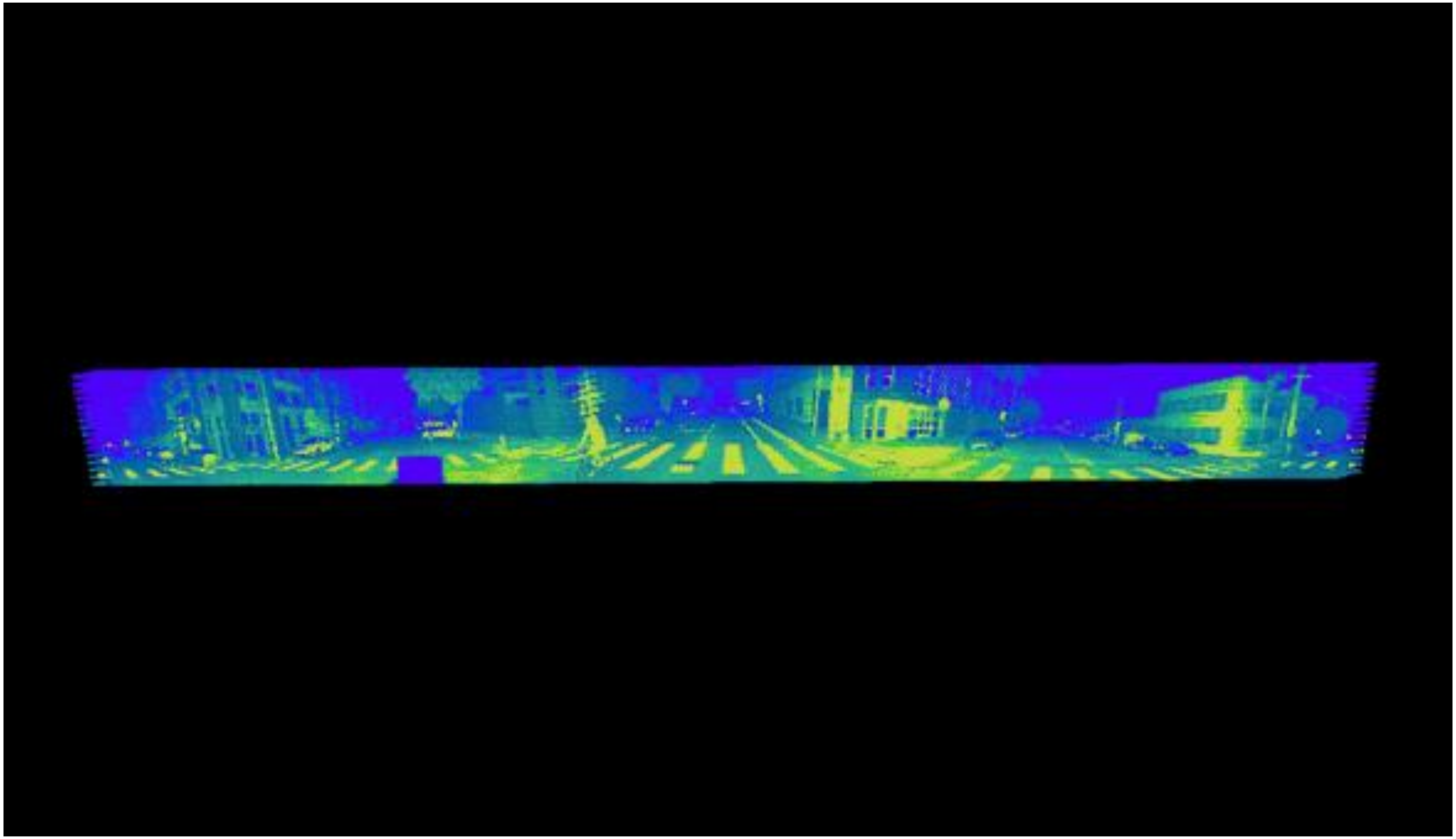
Solid-state LiDAR: Examples

Model	Livox Aviva	Velodyne Velarray H800	Ibeo NEXT
Vert. FoV	4.5	16.0	30.0/5.6
Horiz. FoV	70.4	120.0	60.0/11.2
Price	$1,599	($500)	?

- Restricted horizontal field-of-view
- Much smaller sensors and lower price

Images and specifications from manufacturers.

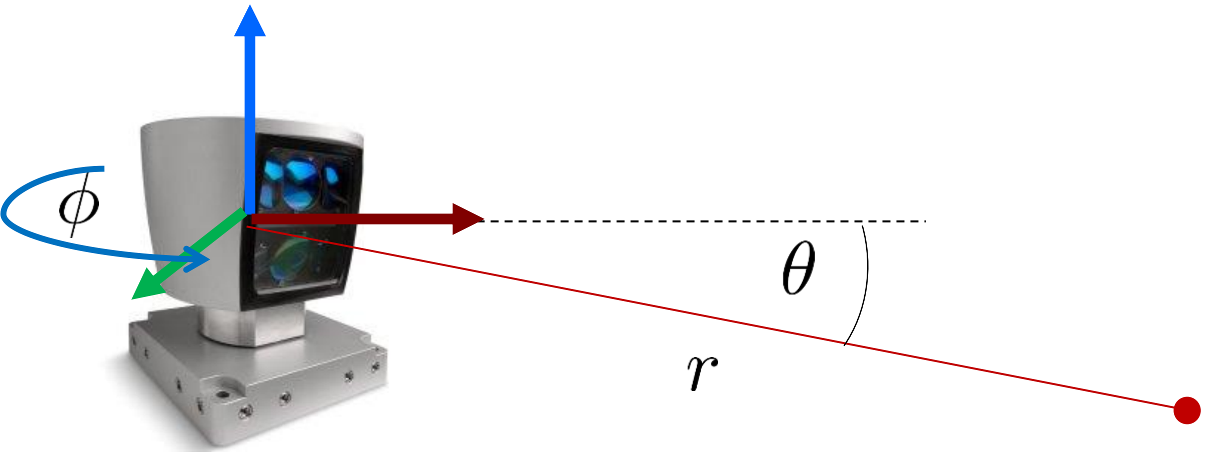
Range Image to Point Cloud



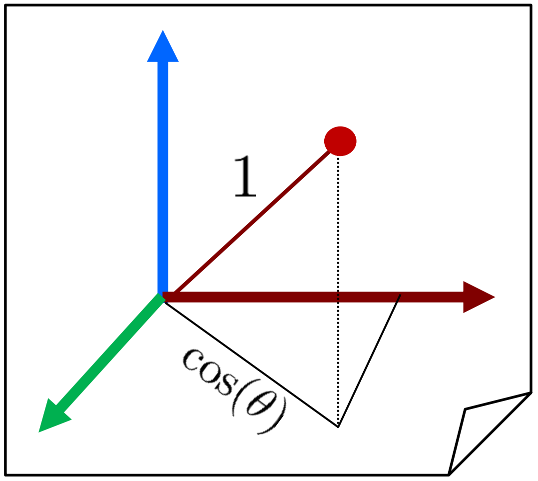
- Angles + range can be converted into 3d points

Animation from Ouster

Range Image to Point Cloud

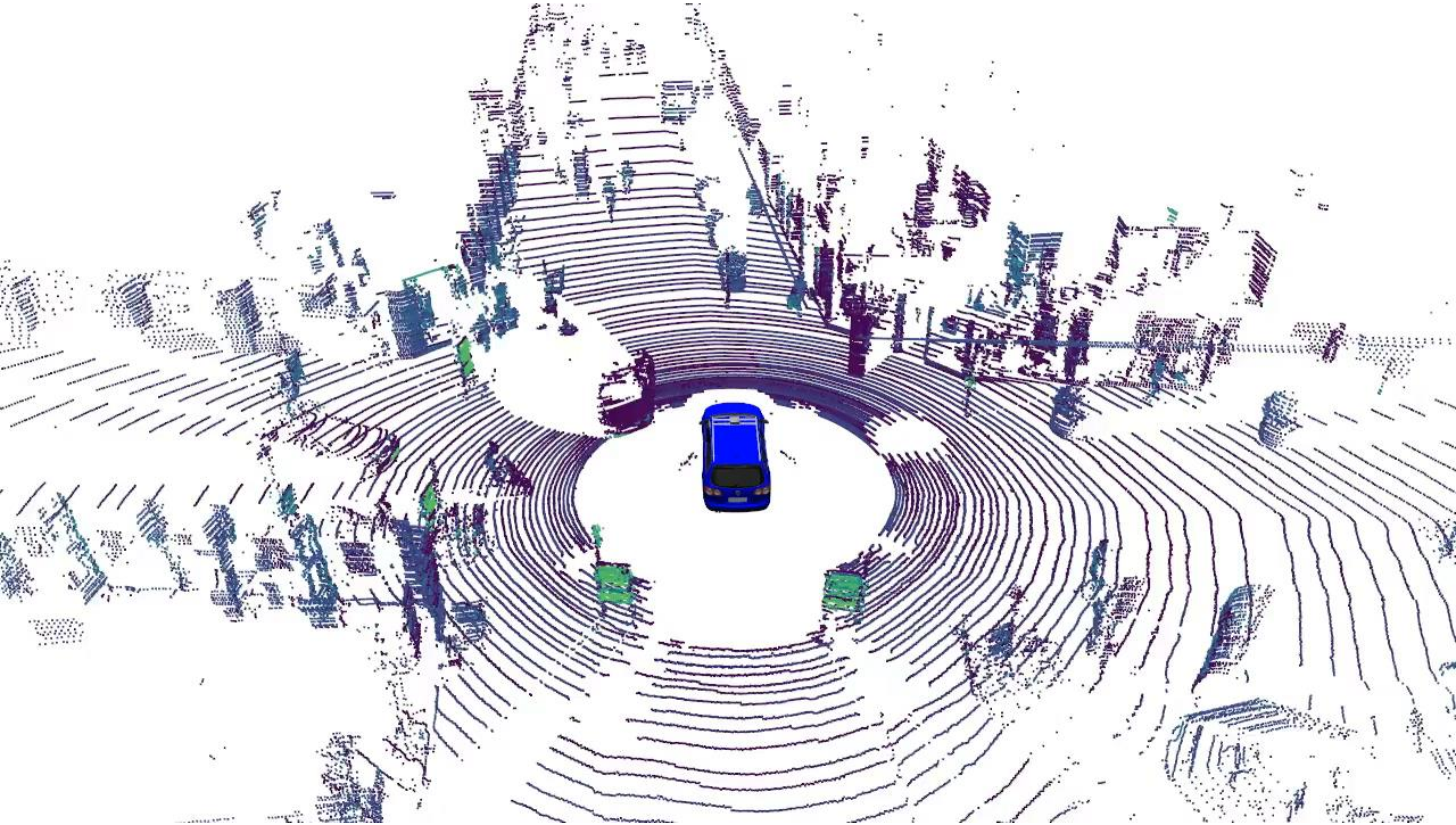


$$
\begin{aligned}
x &= r\cos(\theta)\cos(\phi) \\
y &= r\cos(\theta)\sin(\phi) \\
z &= r\sin(\theta)
\end{aligned}
$$



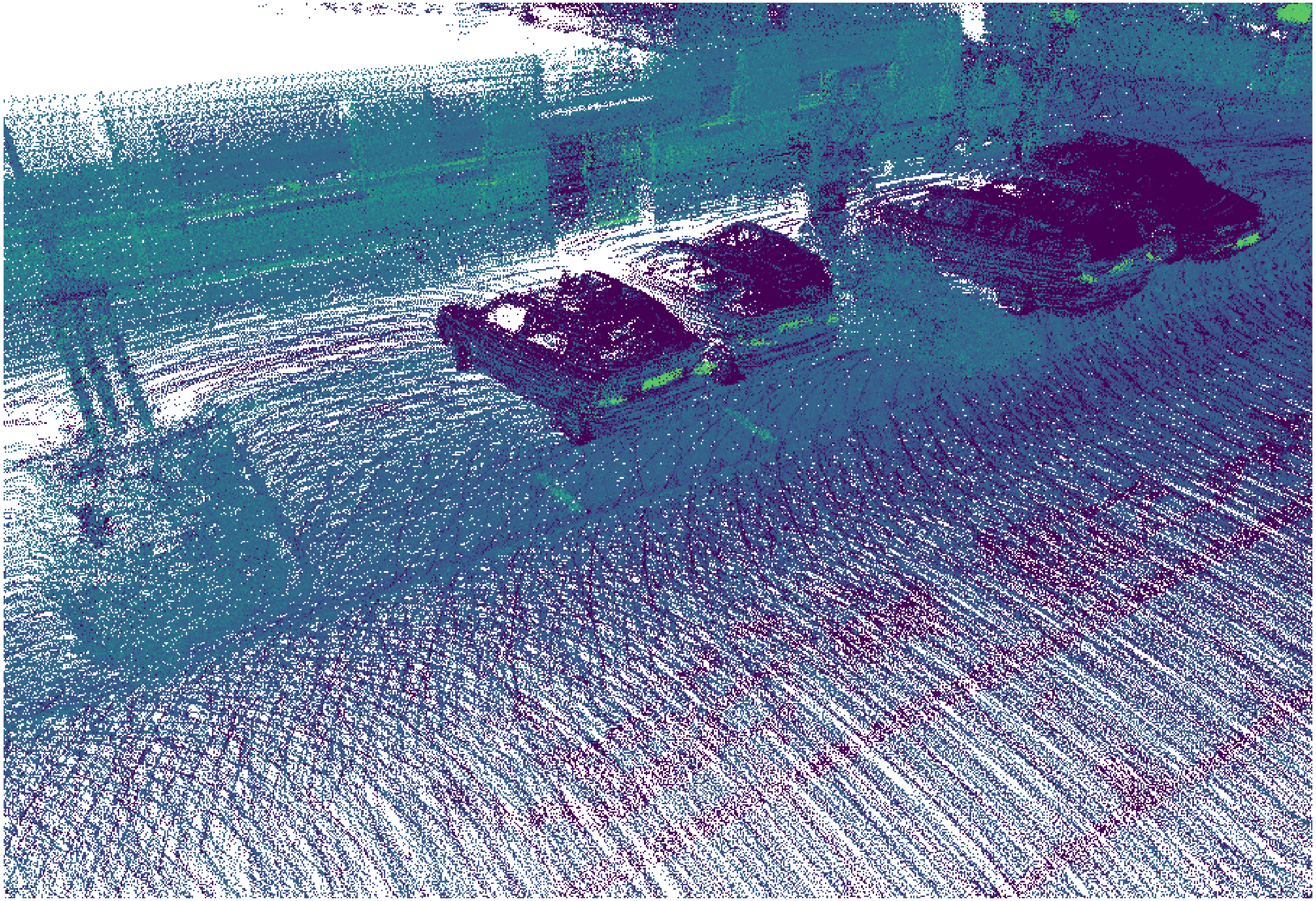
- Angles + range can be converted into 3d points
- **Spherical coordinates** to Euclidean coordinates of azimuth ϕ and inclination θ

Point Clouds: Velodyne HDL-64



KITTI tracking sequence 19

Spatial Aggregation



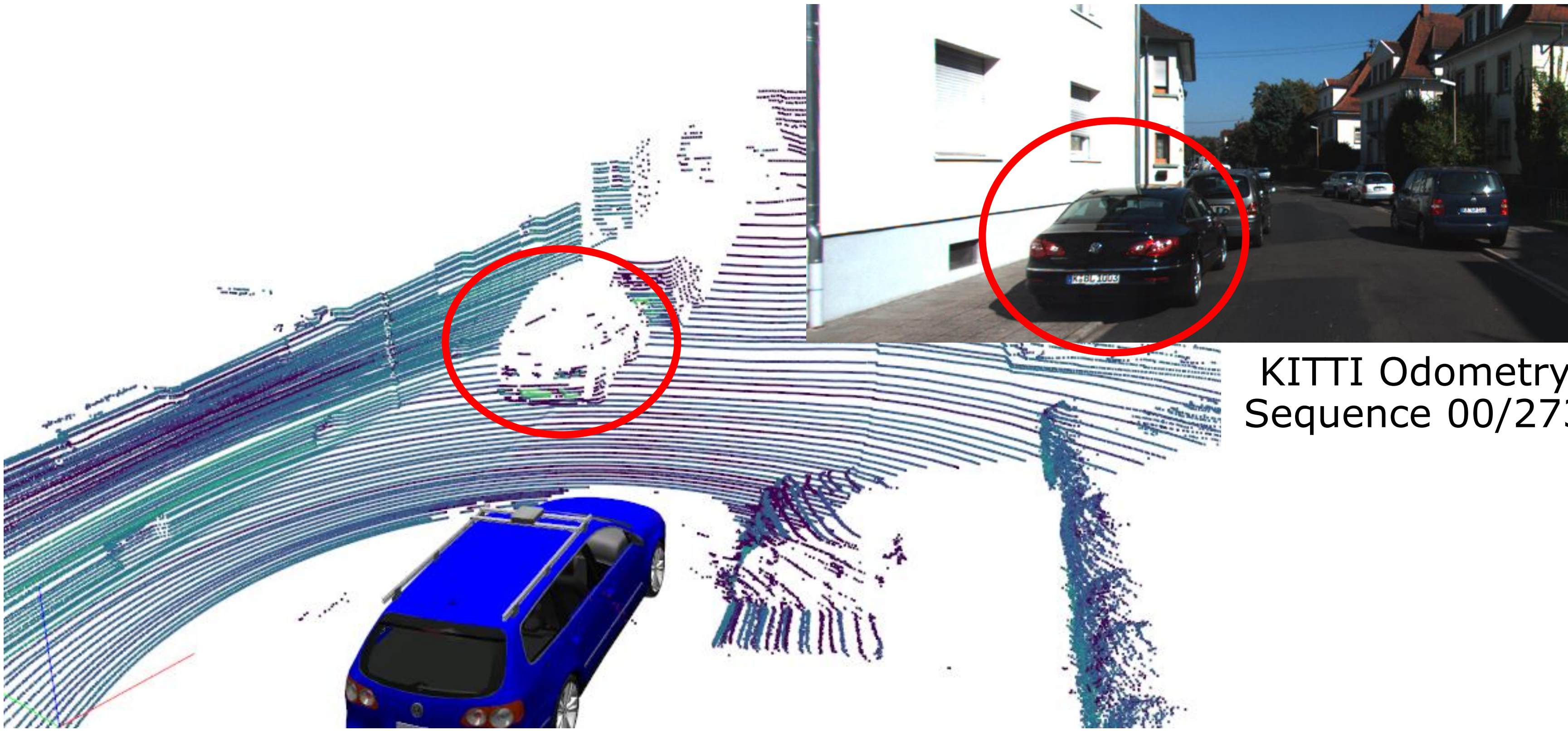
- 3D point clouds can directly aggregated into 3D world representation (with given poses)

Challenges: Distance-dependent Sparsity

5 m | 10 m | 15 m

- Objects at increasing distances are covered with decreasing number of measurements

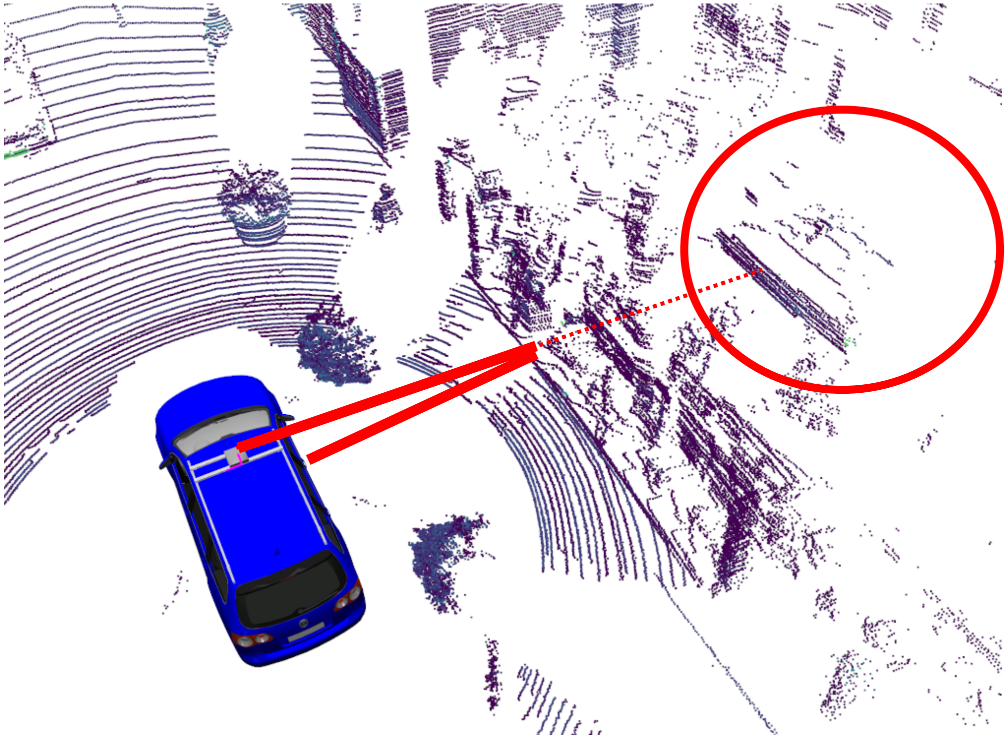
Challenges: Remission



KITTI Odometry
Sequence 00/273

- LiDAR beam gets absorbed by black objects
- Sparse returns on black cars

Challenges: Reflections



- Total reflection of LiDAR beams can lead to "false positives"
- False measurements can introduce wrong detections

Early LiDAR Perception

- Classical approaches based on features
  - Histogram descriptors capturing radius/k-nearest neighborhood of a point
  - **Example:** Spin Images, NARF, etc.

- Approaches based on Conditional Random Fields (CRFs), etc. until ~2014
  - Graph representation: nodes are points, edges represent neighboring points (k-NN or radius neighborhood)
  - Node and edge potentials to minimize joint objective over whole graph

[Behley, 2012] [Triebel, 2006]

Neural Networks for Perception

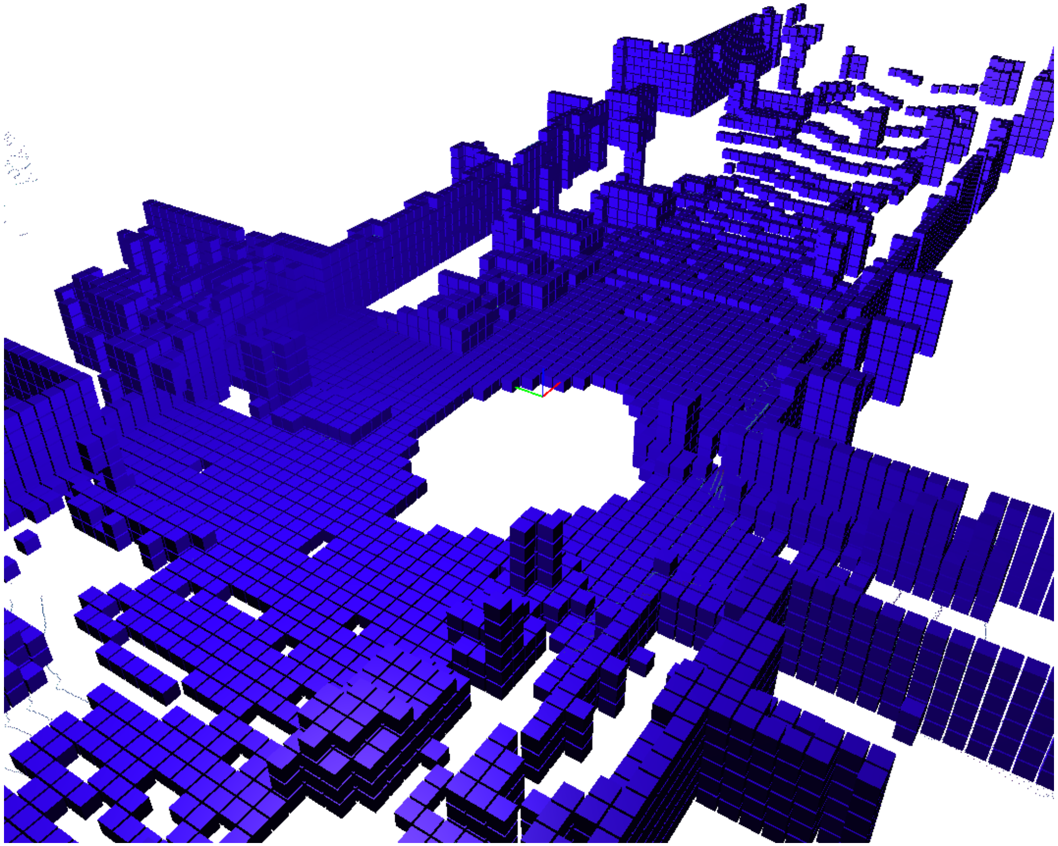
- Deep learning for LiDAR perception the dominant paradigm
- Convolutions are straight-forward, natural basic building block for image-based deep learning

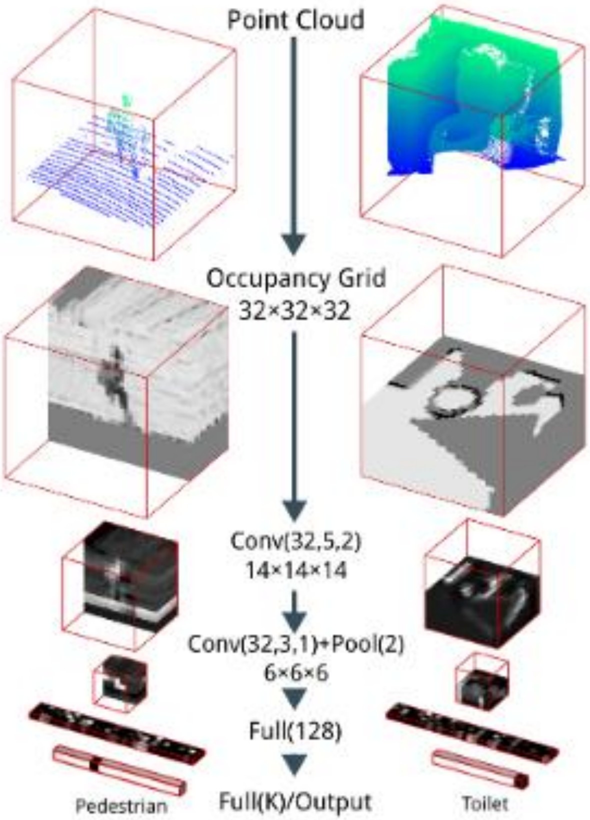
- However: **Unclear** what is the best way to extract features from point clouds

Common Representations and Operations on 3D Data

Representation	Operation
Voxel Grid	3D Convolution
Sparse Voxel Grid	Sparse Convolutions
Projection to Image	2D Convolution
Graphs	Graph Convolution
Point Clouds	Point Convolution

Voxel Grids

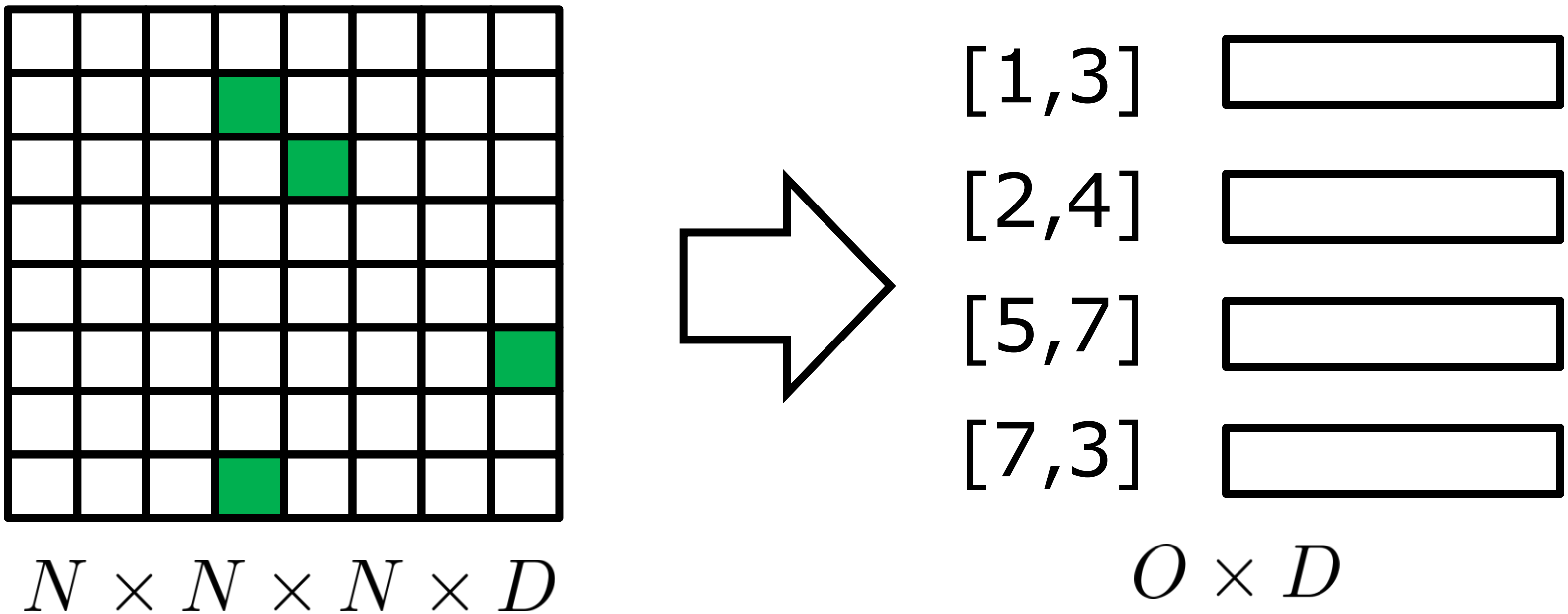




- Convert point cloud into **voxels**
- 3D convolutions on dense voxel grids

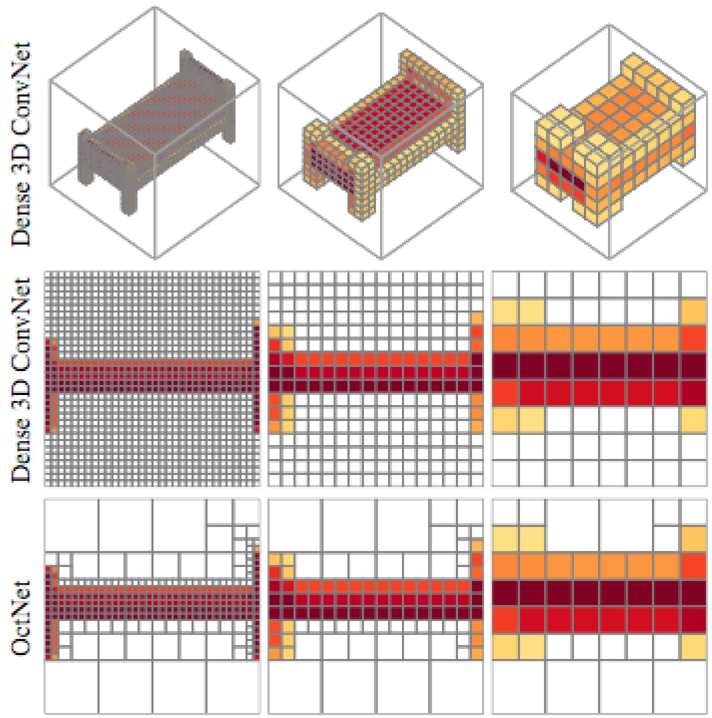
[Maturana, 2015] Right figure from [Maturana, 2015]

Sparse Convolutions



- **Problem:** Dense voxel grids need a lot of memory, even when most voxels are "empty"
- **Solution:** Represent only **occupied** voxels and use sparse convolution

Octree-based Voxelization

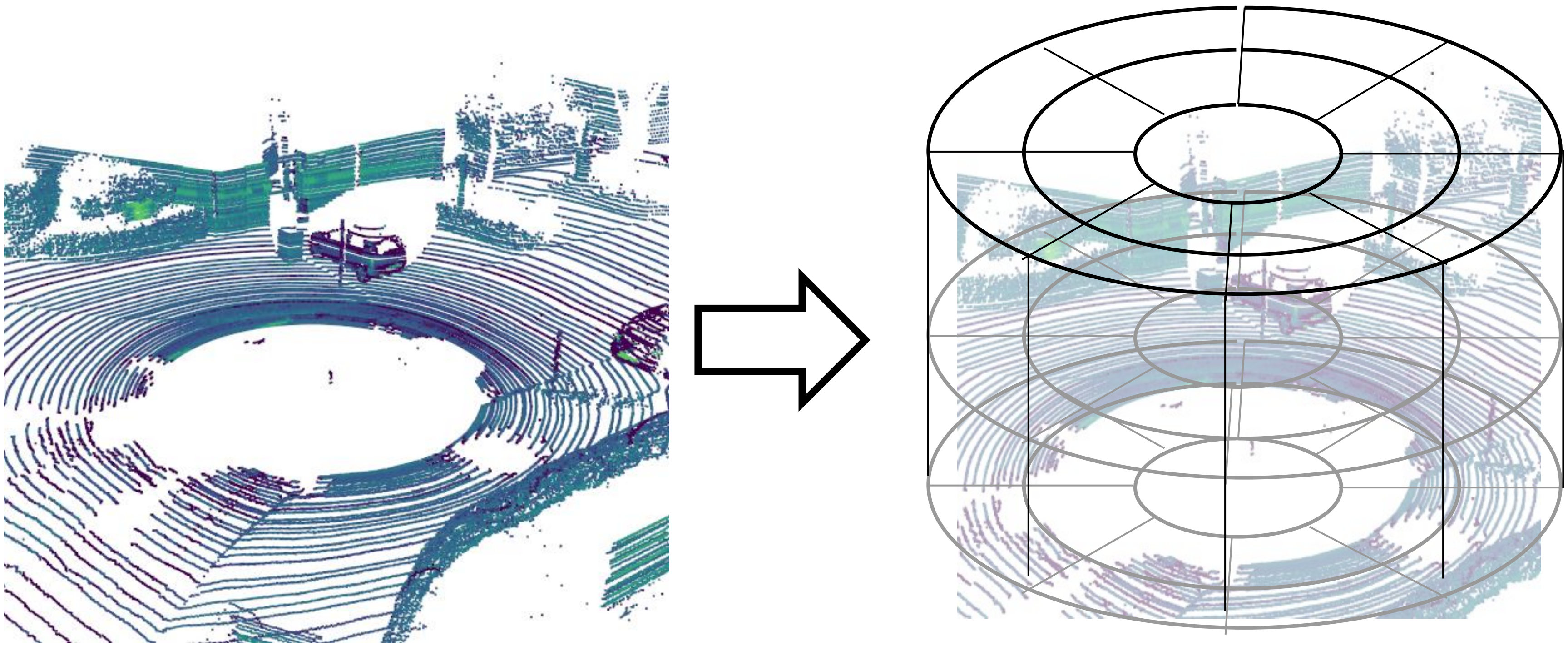


- Other approaches employed Octrees that avoid to explicitly represent empty space

[Riegler, 2017]

Images from [Riegler, 2017]

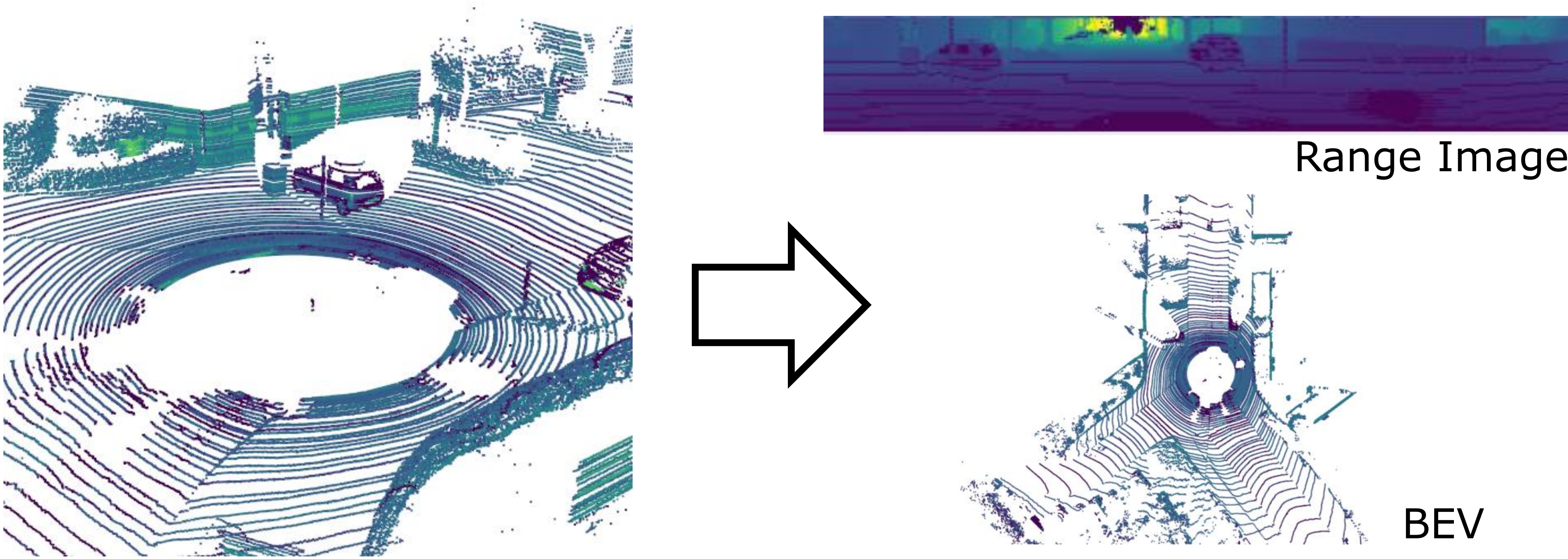
Cylindrical Voxel Grids



- **Idea:** Exploit geometry of scanning method
- Concentric voxels arranged in polar grid

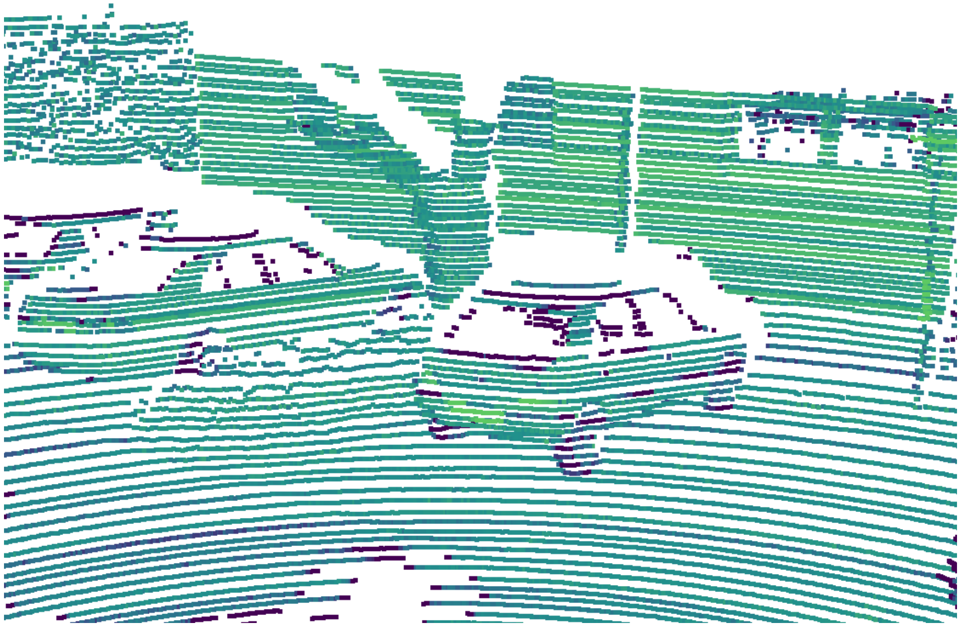
[Zhu, 2021]

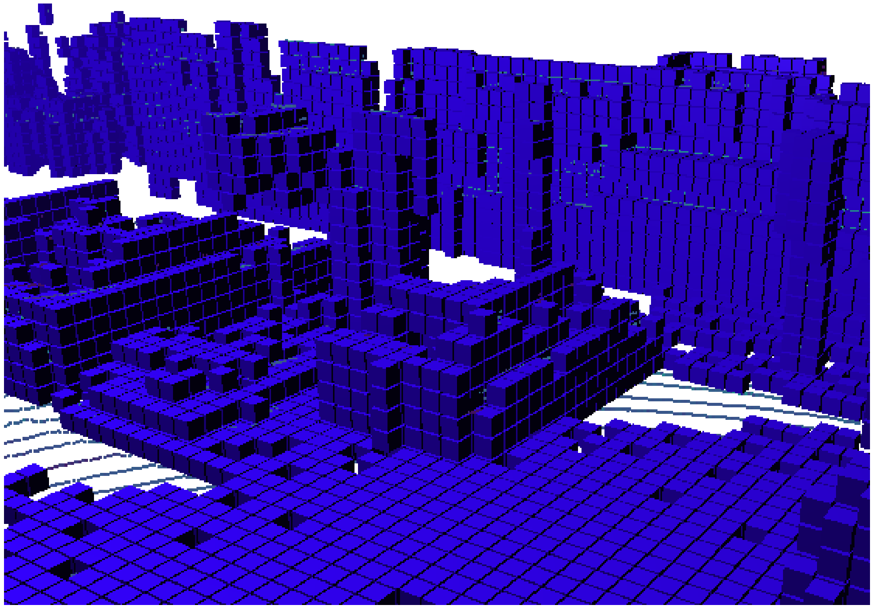
Projective Approaches



- Project point cloud to 2D image projection
  - Range Image (spherical/cylindrical projection)
  - Bird’s eye view (BEV)
- Leverage building blocks and ideas from 2D image-based CNNs

Why point clouds?

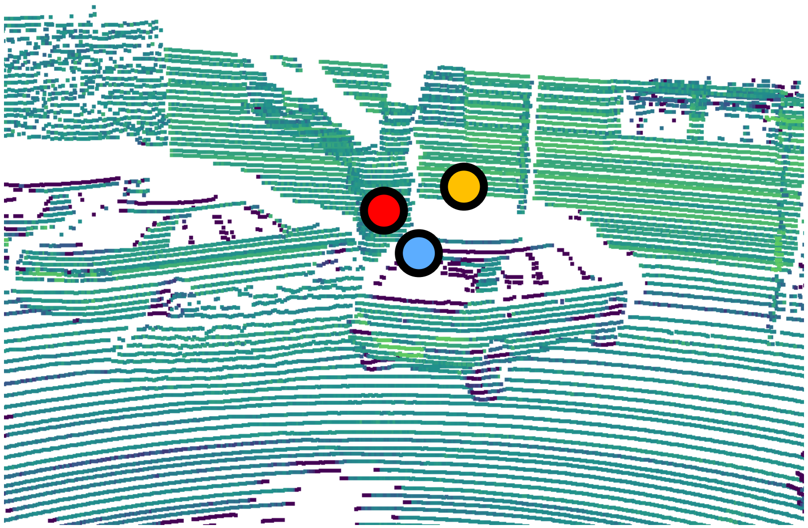


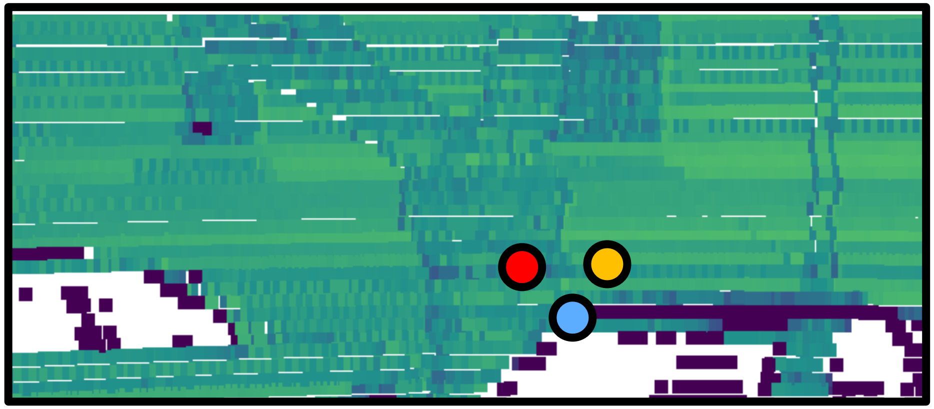


Problems with voxel-based approach:

- Loss of fine details at larger voxel sizes
- “Holes” in voxel grids at smaller sizes

Why point clouds?

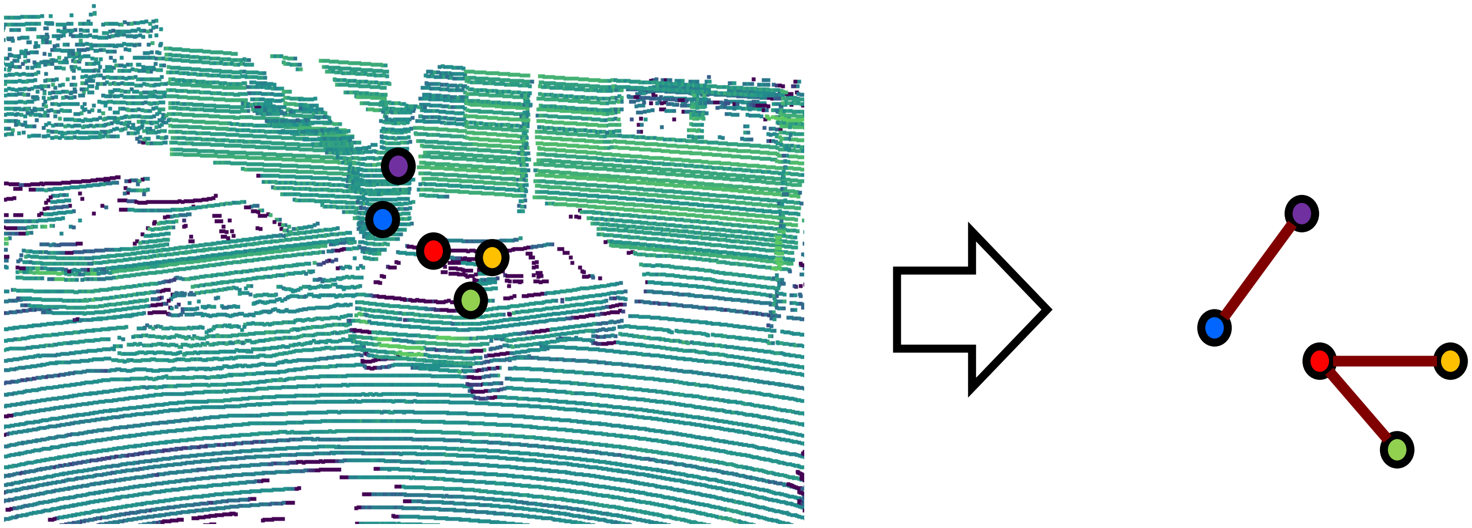




Problem with projective approach:

- Neighborhood in projections usually does not correspond to spatial neighborhood

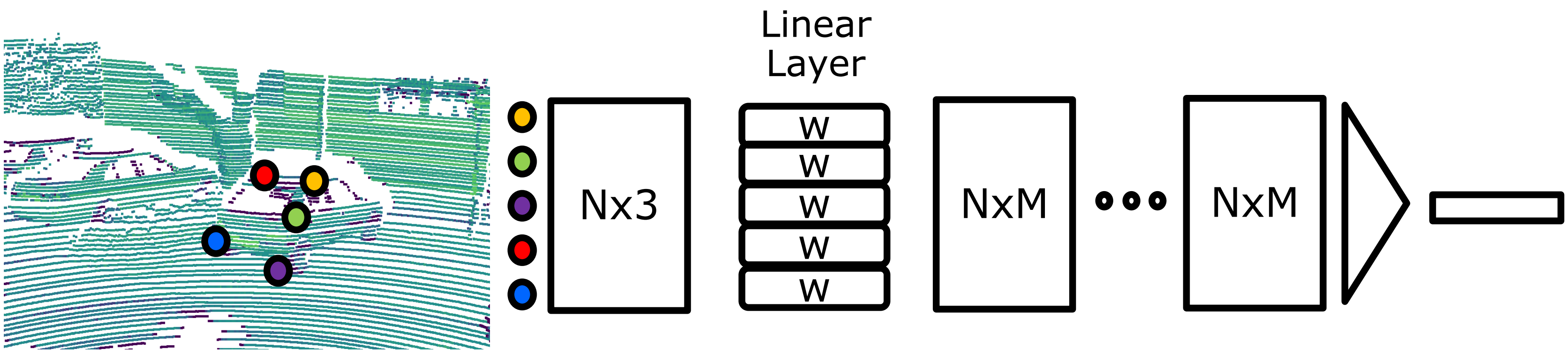
Graph-based Representation



- Represent point cloud as graph
- Edges in graph structure are defined via spatial neighborhood
  - K-nearest neighbors or radius neighborhood
- Feature aggregation on graph structure

[Wang, 2019]

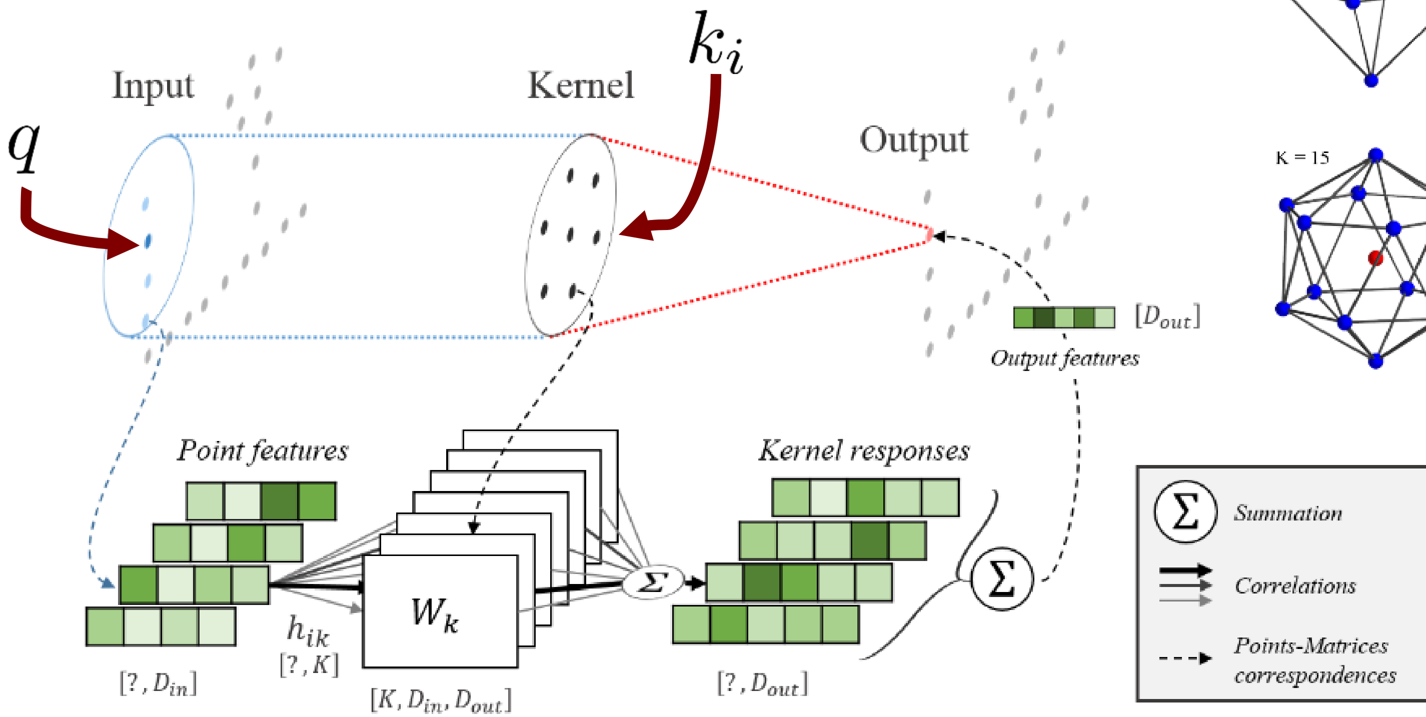
PointNet

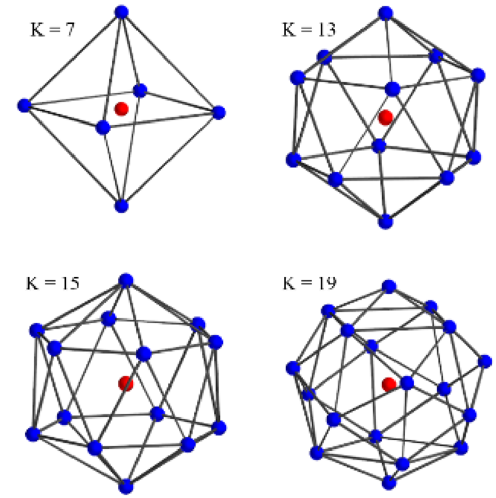


- Use linear projections on each point with non-linearity
- Use pooling operations (avg. or max pooling) to make operation order-invariant

[Qi, 2017]

KPConv



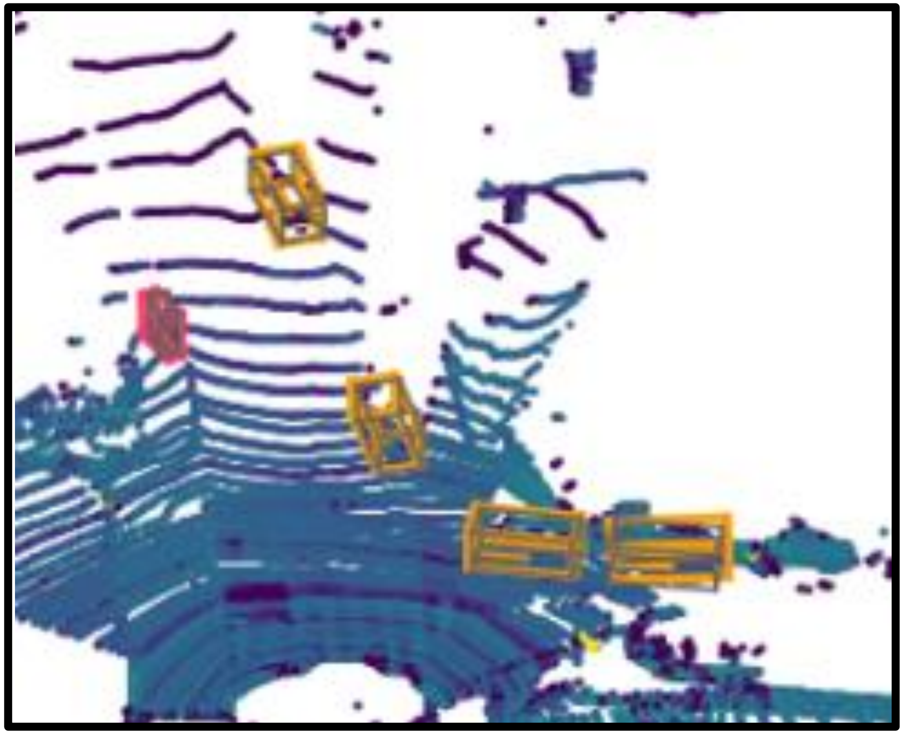


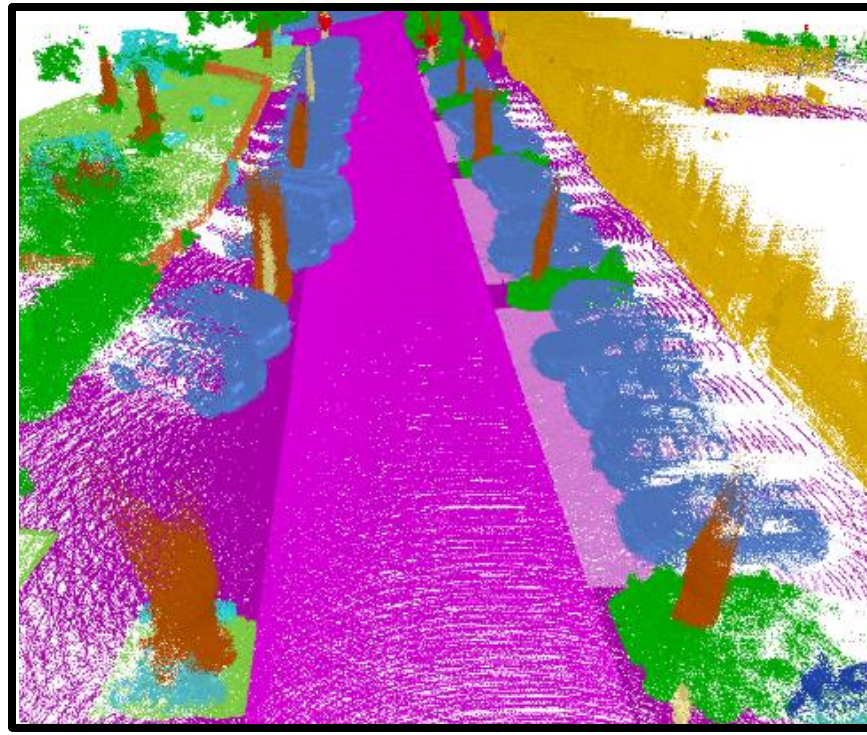
- Define fixed location for kernel points k_i in relation to query point q
- Distance to kernel points determine influence of corresponding weight

[Thomas, 2019]

Images from [Thomas, 2019]

LiDAR-based Perception



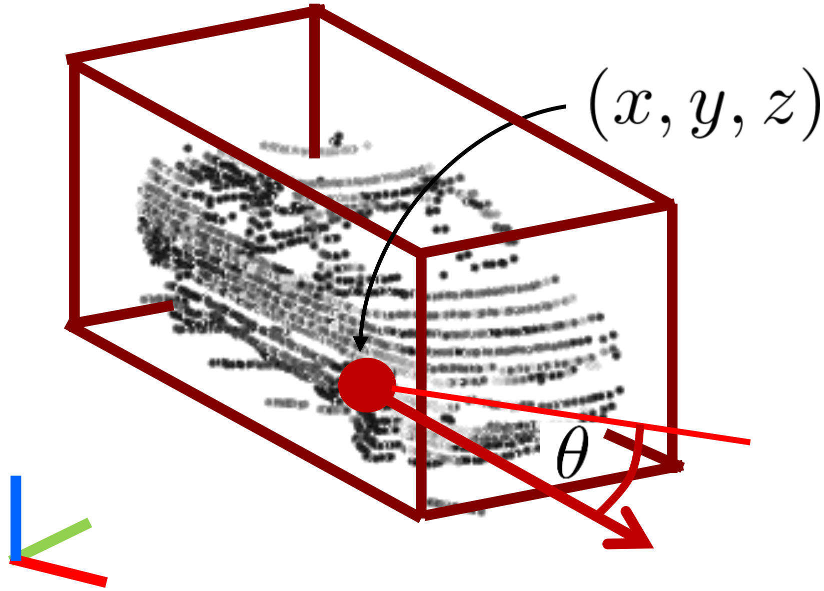


3D Object Detection

Semantic Segmentation

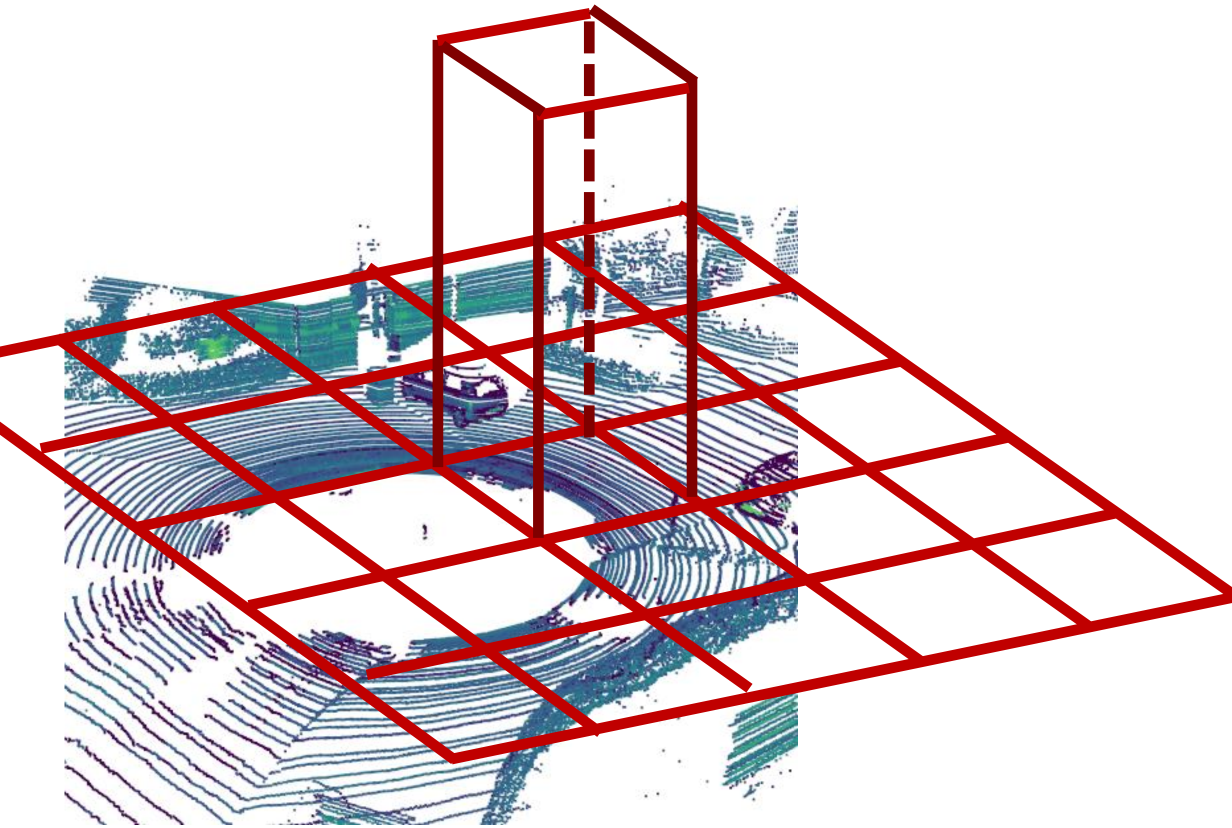
Panoptic Segmentation

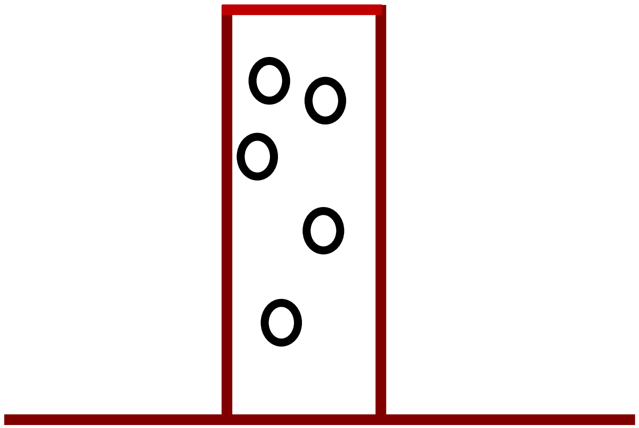
3D Object Detection



- **Task:** oriented bounding boxes of objects
- Bounding box parameterized by
  - location (x, y, z) (center or ground bounding box)
  - orientation θ (rotation angle around z)
  - size by width, height, length

PointPillars (1)

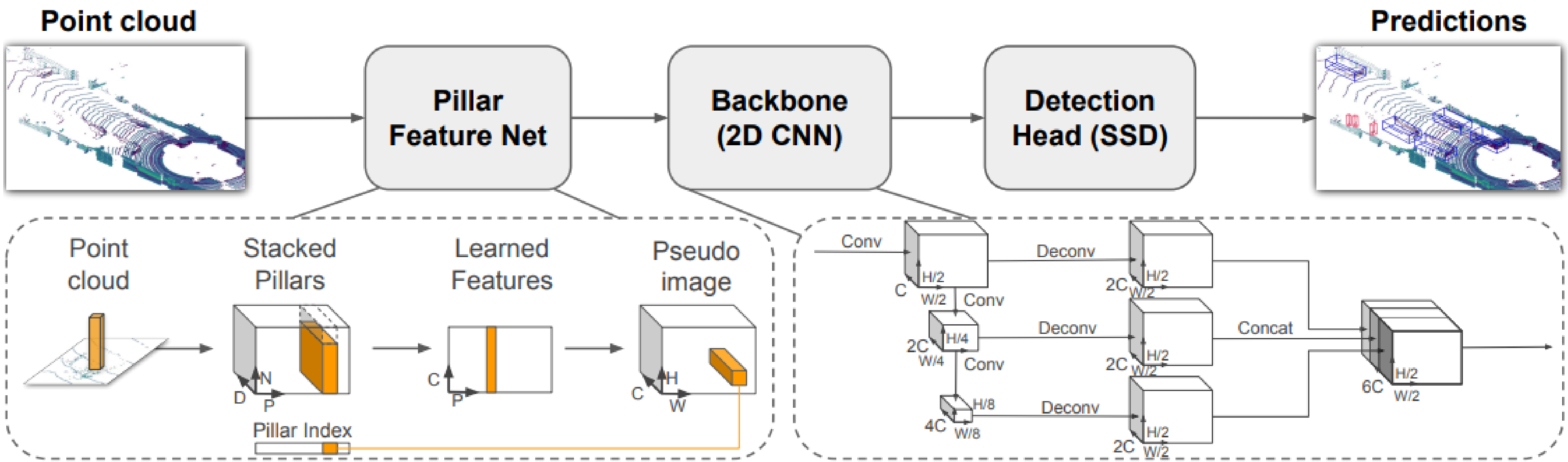




- Represent point cloud by pillars (2.5D)
- Each pillar aggregates points at a 2D location in a grid

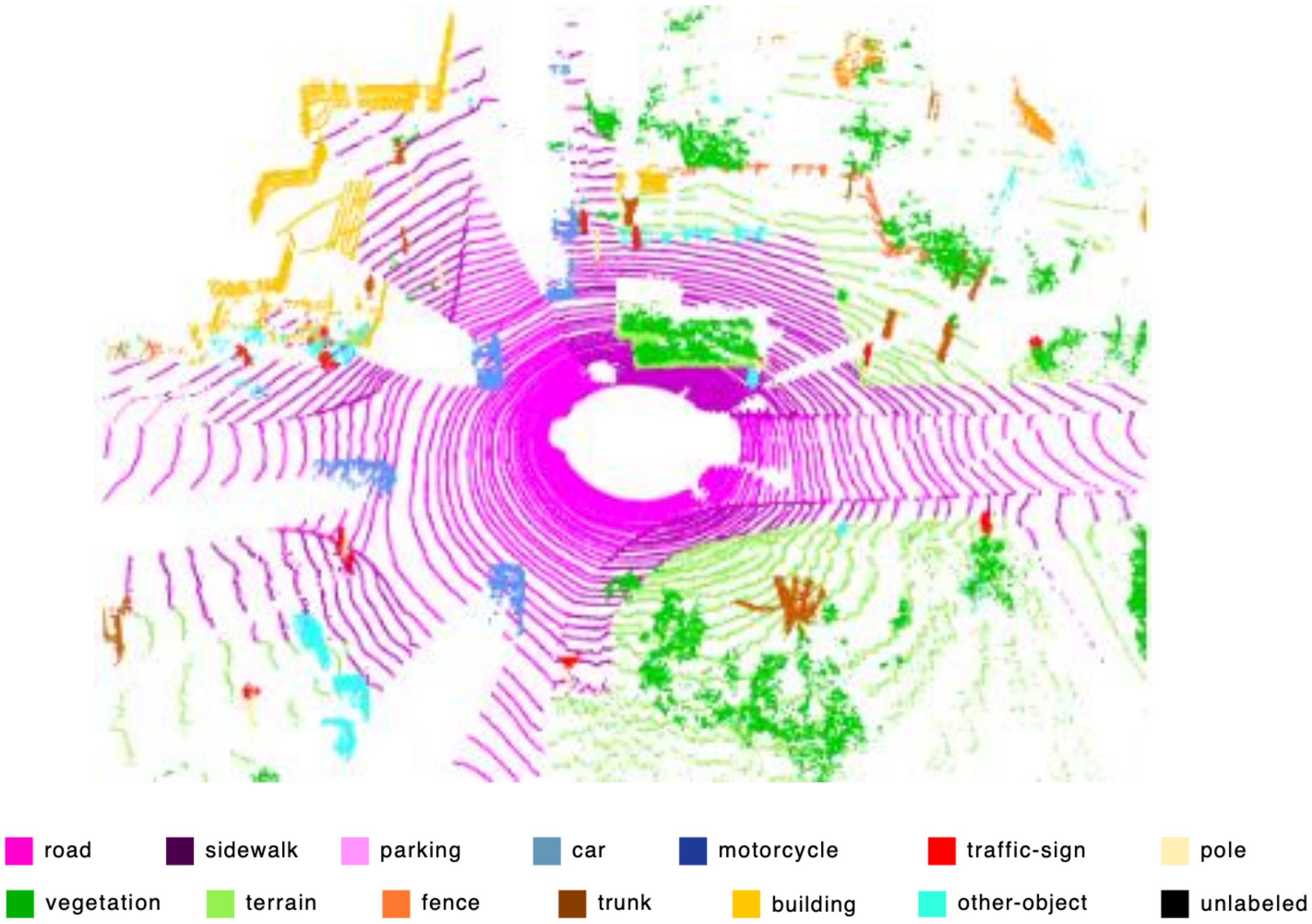
[Lang, 2019]

PointPillars (2)



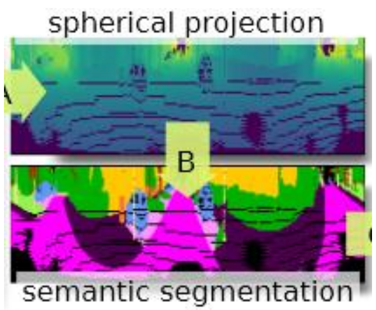
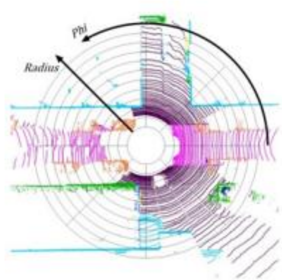
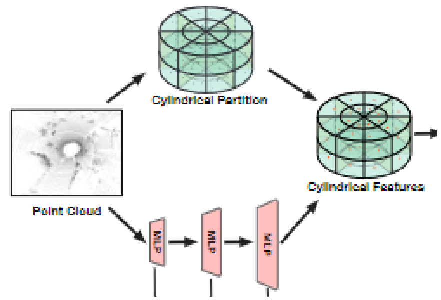
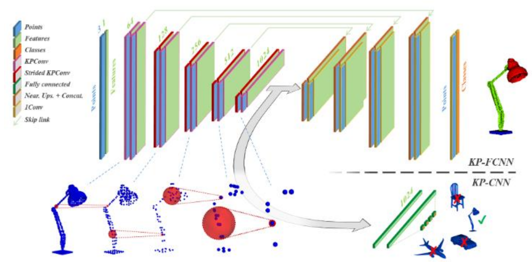
- Convert points in pillar to pillar features
- Use sparse convolutions on pillars to extract features
- Apply common detection approach on 2D BEV representation

LiDAR-based Segmentation



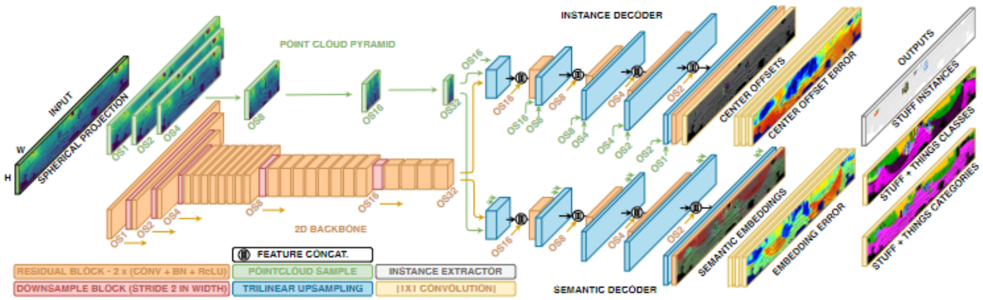
- **Task:** Assign each point a class label

Common Approaches

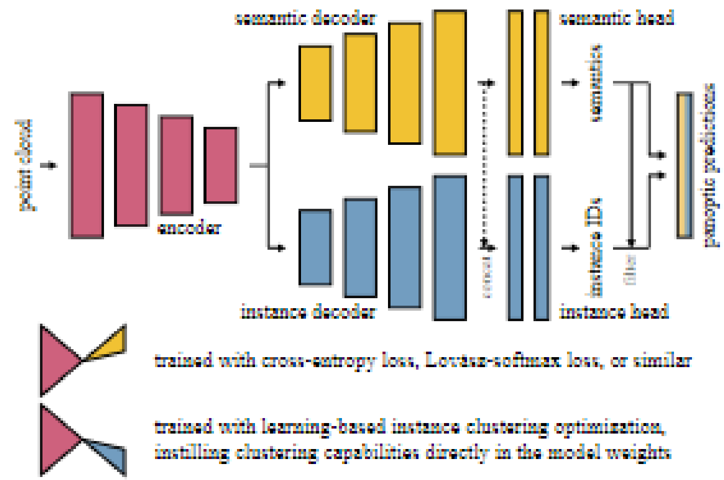
Projective Approaches		Voxel-based Approaches	Point-based Approaches
			
[Milioto, 2019]	[Zhang, 2020]	[Zhu, 2021]	[Thomas, 2019]

- Very similar to semantic segmentation in images, all approaches follow encoder-decoder architecture
- Main difference: Representation of point cloud

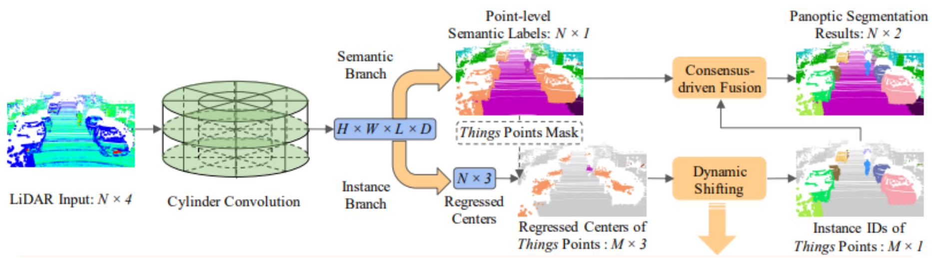
LiDAR Panoptic Segmentation



Range Image [Milioto, 2020]



Point clouds [Gasperini, 2021]

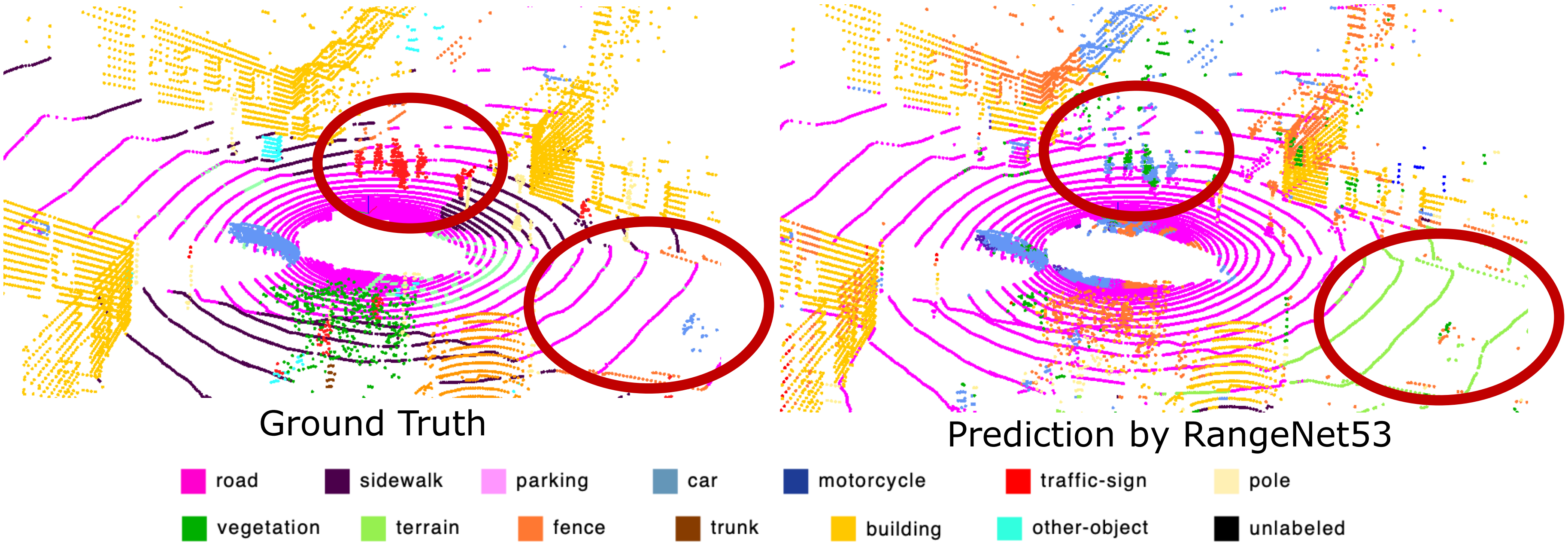


Cylindrical Voxels [Hong, 2021]

- Most approaches use **bottom-up** instance segmentation with **separate decoders** for semantic and instance prediction

[Langer, 2020]

Challenge: Domain Transfer



- **Common Problem:** Approaches learned on one sensor do not work well on others
- Different sampling of environment, different number of beams, etc.

Approaches for Domain Transfer

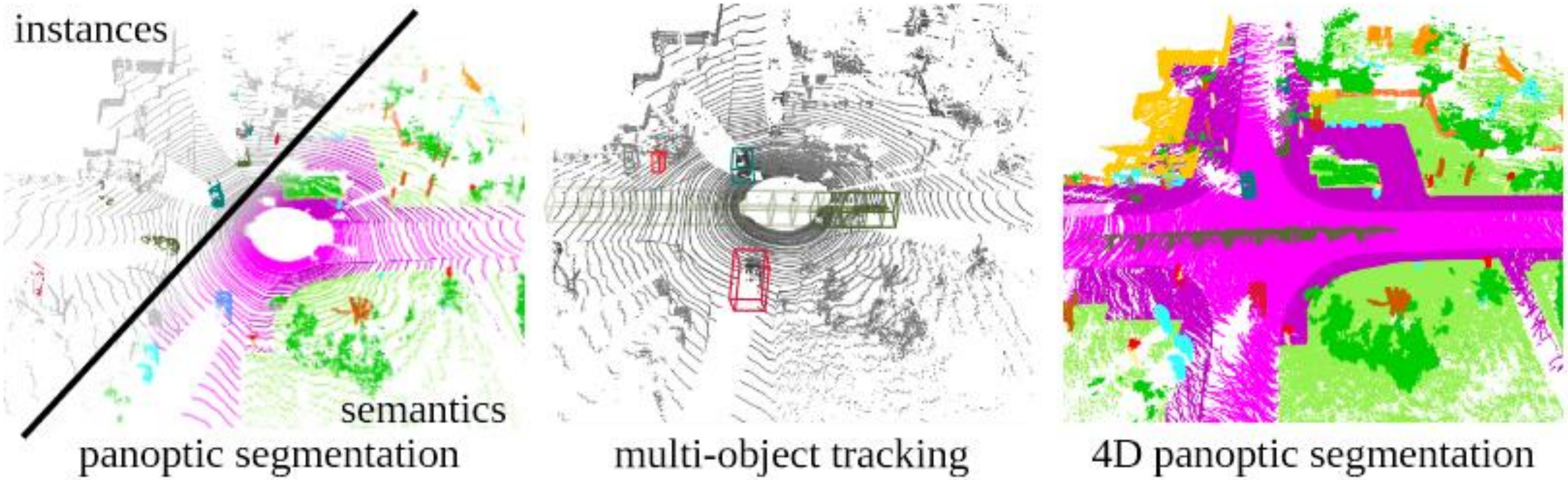
Aim: Avoid to label too much additional data

- Unsupervised Domain Adaption
  - Adapt already trained approach to new situation

- Self-supervised Representation Learning
  - Pre-train network on target domain and use few labeled scans for fine-tuning

[Langer, 2020] [Xie, 2020]

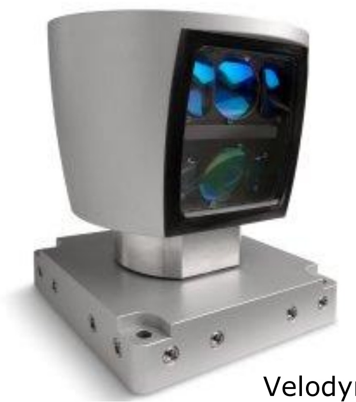
4D Panoptic LiDAR Segmentation



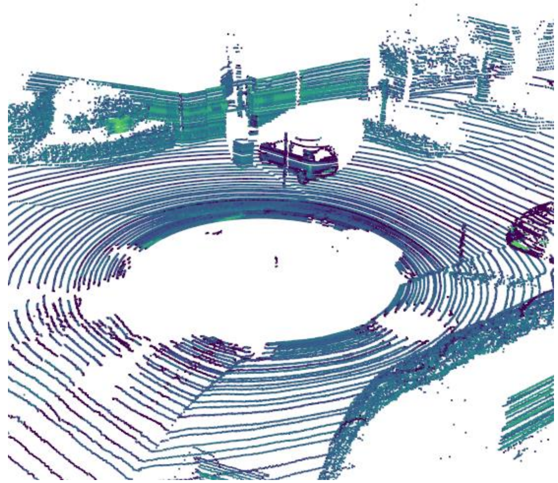
- Combine panoptic segmentation and tracking
- Estimate temporally consistent instance associations

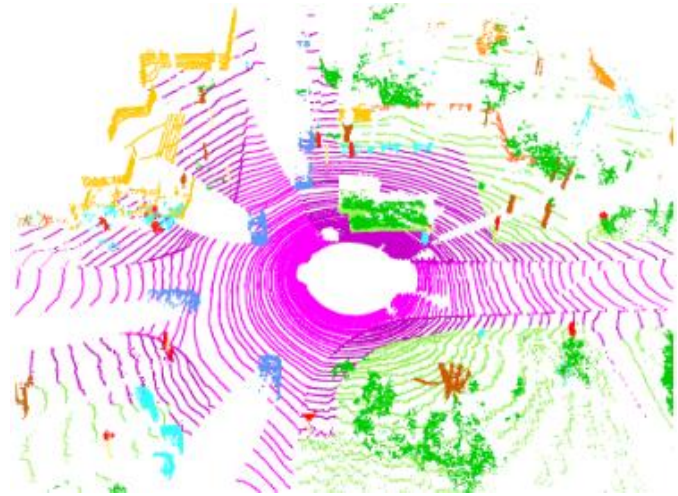
[Aygün, 2021]

Summary



Velodyne
Lidar, Inc.





- LiDARs are viable sensors in the perception stack of most autonomous cars
- Representation & operations for feature extraction on point clouds
- Approaches for LiDAR-based perception: Detection, Semantic/Panoptic Segmentation, etc.

References

[Aygün, 2021] Aygün et al. 4D Panoptic LiDAR Segmentation, CVPR, 2021.

[Behley, 2012] Behley et al. Performance of Histogram Descriptors for the Classification of 3D Laser Range Data in Urban Environments, ICRA, 2012.

[Burgard, 1998] Burgard et al. The Interactive Museum Tour-Guide Robot, AAAI, 1998.

[Gasperini, 2021] Gasperini et al. Panoster: End-to-end panoptic segmentation of LiDAR point clouds, IEEE RA-L, vol. 6(2), 2021.

[Hong, 2021] Hong et al. LiDAR-based Panoptic Segmentation via Dynamic Shifting Network, CVPR, 2021.

[Langer, 2020] Langer et al. Domain Transfer for Semantic Segmentation of LiDAR Data using Deep Neural Networks, IROS, 2020.

[Maturana, 2015] Maturana et al. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition, IROS, 2015.

[Milioto, 2019] Milioto et al. RangNet++: Fast and Accurate LiDAR Semantic Segmentation, IROS, 2019.

[Milioto, 2020] Milioto et al. LiDAR Panoptic Segmentation for Autonomous Driving, IROS, 2020.

[Montemerlo, 2008] Montemerlo et al. Junior: The Stanford Entry in the Urban Challenge, JFR, vol. 25(9), pp. 569-597, 2008

[Qi, 2017] Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR, 2017.

[Riegler, 2017] Riegler et al. OctNet: Learning Deep 3D Representations at High Resolution, CVPR, 2017.

[Urmson, 2008] Urmson et al. Autonomous Driving in the Urban Environments: Boss and the Urban Challenge, JFR, vol. 25(8), pp. 425-466, 2008.

[Thrun, 2006] Thrun et al. Stanley: The Robot that Won the DARPA Grand Challenge, JFR, vol. 23(9), pp. 661-692, 2006.

[Thomas, 2019] Thomas et al.

[Triebel, 2006] Triebel et al. Robust 3D Scan Point Classification using Associative Markov Networks, ICRA, 2006.

[Wang, 2019] Wang et al. Dynamic Graph CNN for Learning on Point Clouds, TOG, vol.38(5), 2019.

[Xie, 2020] Xie et al. PointContrast: Unsupervised Pre-training for 3D Point Cloud Understanding, ECCV, 2020.

[Zhang, 2020] Zhang et al. PolarNet: An Improved Grid Representation for Online LiDAR Point Cloud Semantic Segmentation, CVPR, 2020.

[Zhu, 2021] Zhu et al. Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation, CVPR, 2021.