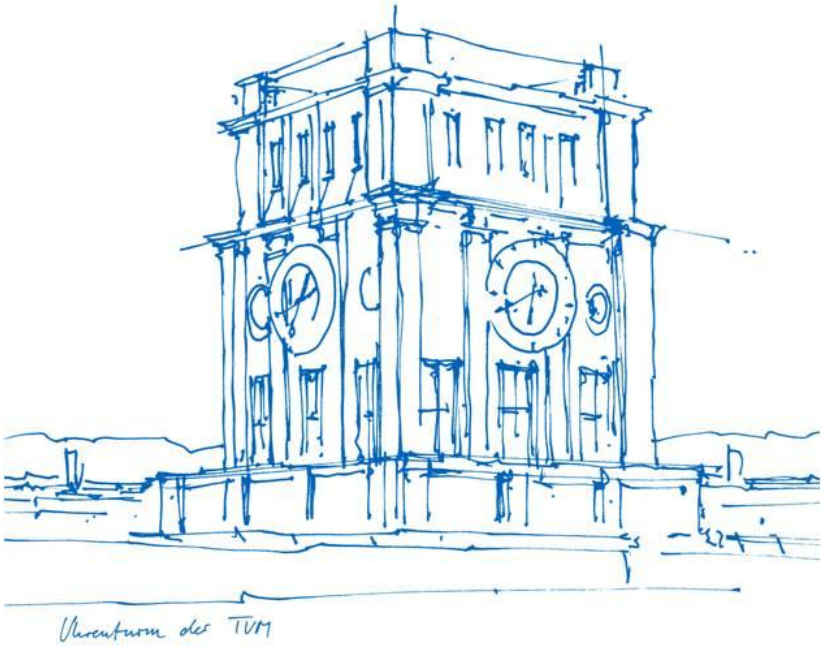
Sensor Modality Fusion

Group members: Nikhitha Radhakrishna Naik and Chenguang Huang

Technische Universität München

Faculty of Informatik

05 November 2019



Contents

1. Introduction
2. Datasets
3. State of the art methods
4. Plan for next week

Introduction

Autonomous driving >> Perception >> Decision making

Sensors:

- ❑ Camera

Pros:
- 360 degree view.

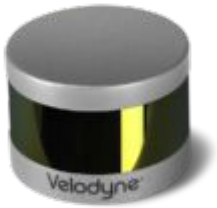
Cons:
- Distance calculation
- Poor in low visibility conditions

- ❑ Radar

Pros:
- Speed and location of the object
- Works in low visibility

Cons:
- Differentiate between different vehicles

- ❑ Lidar



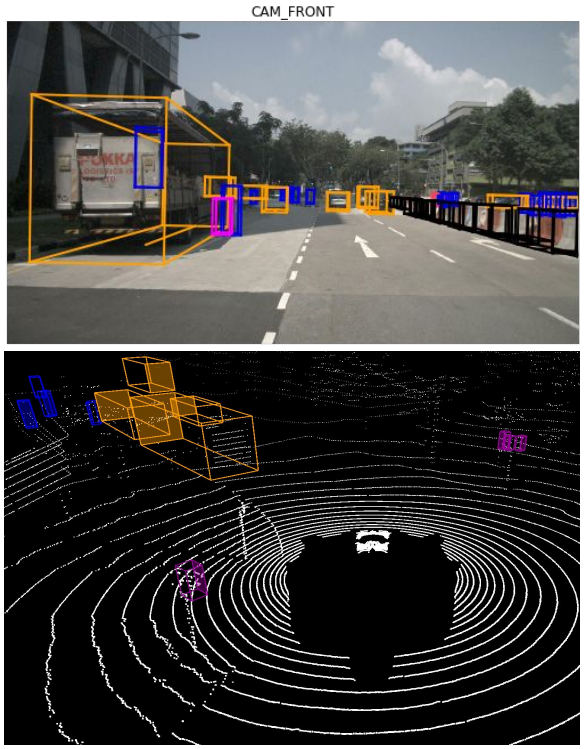
Pros:
- 3D view
- Shape and depth

Cons:
- Expensive
- Limited range

Introduction

Application of perception: 3D Object Detection

1. Find the position of the bounding box in 3D
2. Classify the contents inside the bounding box



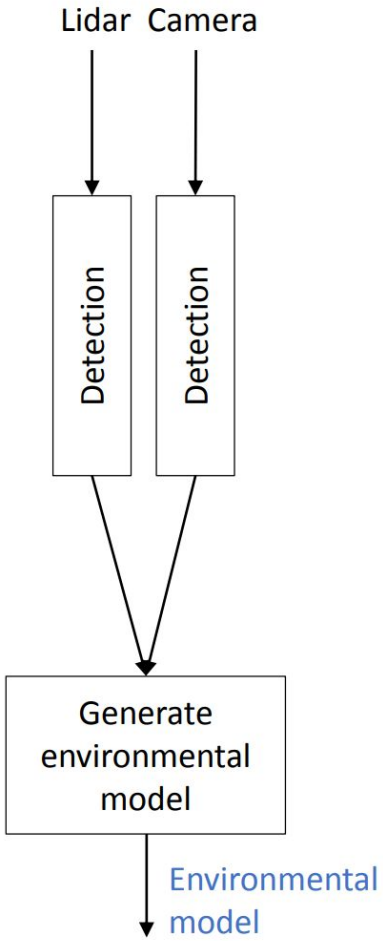
Data used to achieve these goals:

1. Monocular image
2. Lidar point cloud
3. Both image and point cloud

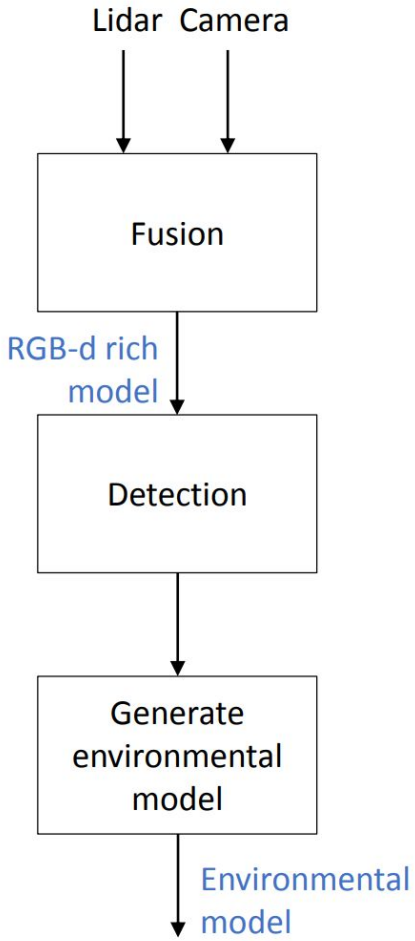
Question: When we use more than one sensor's data, how do we fuse them?

Introduction : Sensor Fusion

Object level fusion



Low-level fusion



- Low-level fusion advantages
  - Expanded content of information
  - Super resolution for sparse sensors
  - Joint probabilities at the pixel level
  - Best overall detection performance
  - Prone to sensor failure ?

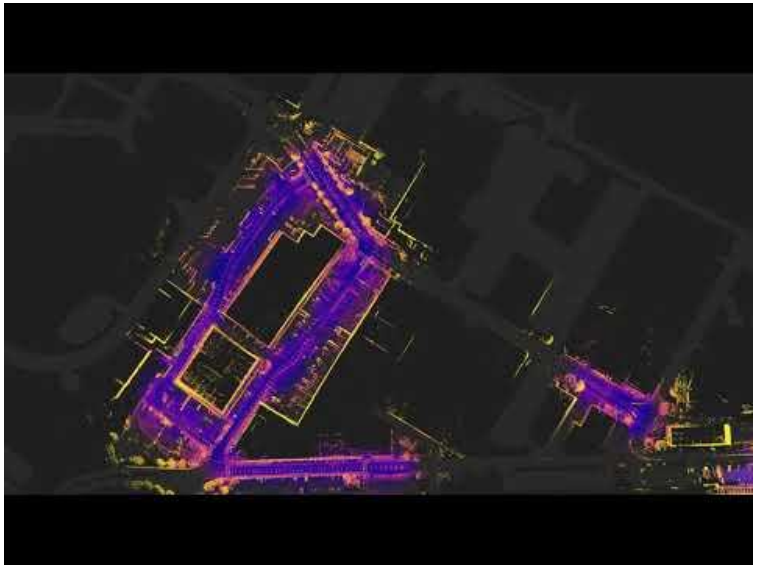
Datasets

NUSCENES

700 scenes for training,

150 scenes for evaluation,

150 for testing

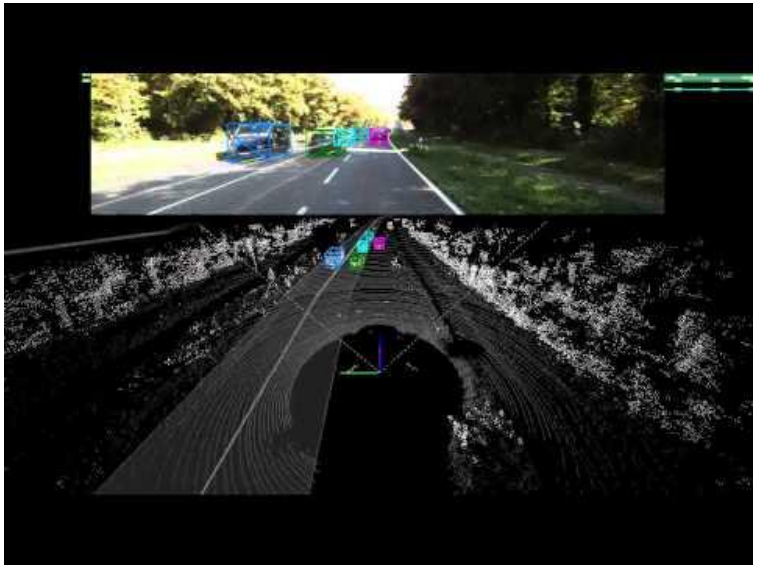


KITTI

7481 training images

7518 testing images

and corresponding point cloud



State of the art methods

3D Object Detection with point cloud:

1. PointRCNN(CVPR, 2019)

Method	Modality	Car (IoU=0.7)			Pedestrian (IoU=0.5)			Cyclist (IoU=0.5)		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
MV3D [4]	RGB + LiDAR	71.09	62.35	55.12	-	-	-	-	-	-
UberATG-ContFuse [17]	RGB + LiDAR	82.54	66.22	64.04	-	-	-	-	-	-
AVOD-FPN [14]	RGB + LiDAR	81.94	71.88	66.38	50.80	42.81	**40.88**	64.00	52.18	46.61
F-PointNet [25]	RGB + LiDAR	81.20	70.39	62.19	**51.21**	**44.89**	40.23	71.96	56.77	50.39
VoxelNet [43]	LiDAR	77.47	65.11	57.73	39.48	33.69	31.51	61.22	48.36	44.37
SECOND [40]	LiDAR	83.13	73.66	66.20	51.07	42.56	37.29	70.51	53.85	46.90
Ours	LiDAR	**85.94**	**75.76**	**68.32**	49.43	41.78	38.63	**73.93**	**59.60**	**53.59**

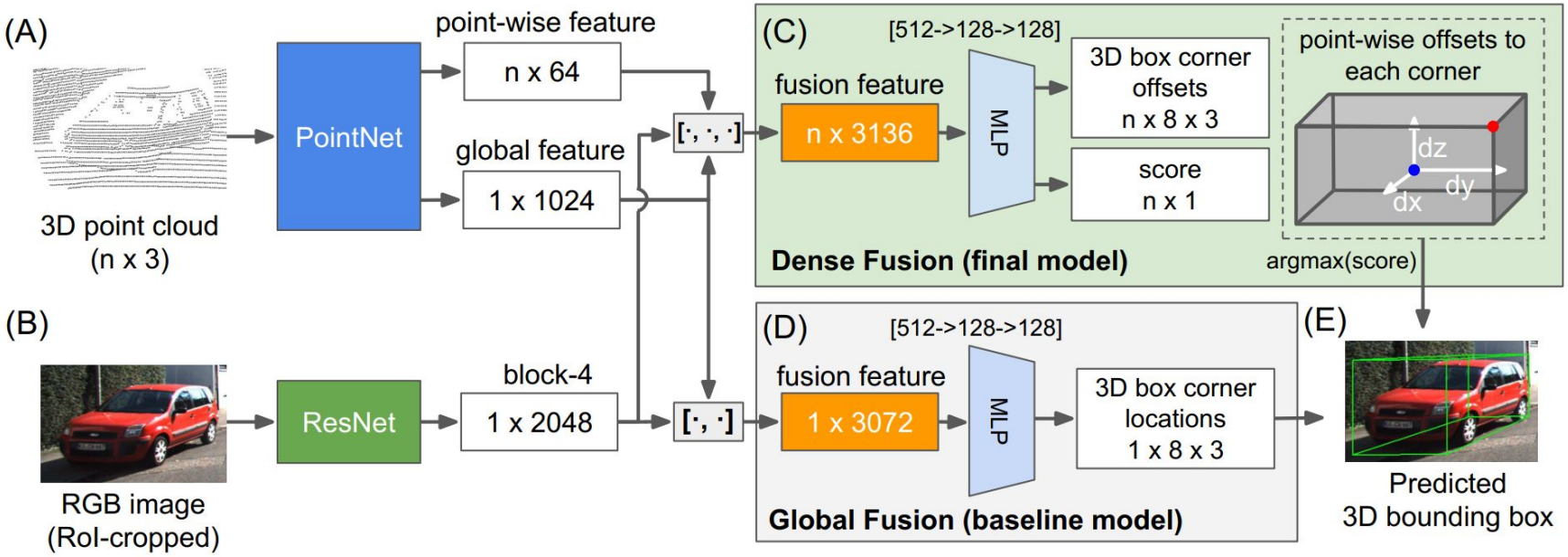
Table 1. Performance comparison of 3D object detection with previous methods on KITTI *test* split by submitting to official test server. The evaluation metric is Average Precision(AP) with IoU threshold 0.7 for car and 0.5 for pedestrian/cyclist.

State of the art methods

3D Object Detection with low level sensor fusion:

2. PointFusion(CVPR, 2018)

extract features of point cloud and image respectively and concatenate them.

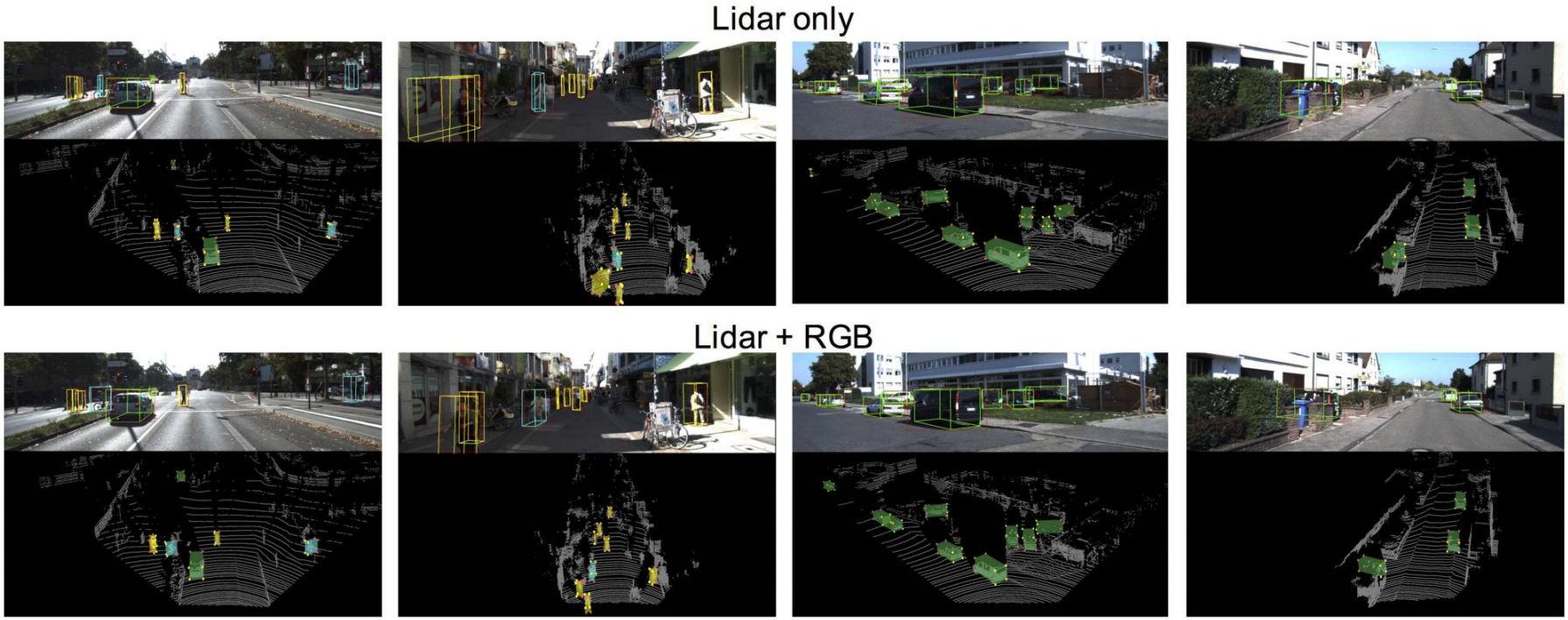


State of the art methods

3D Object Detection with low level sensor fusion:

2. PointFusion(CVPR, 2018)

more precise bounding box size and orientation

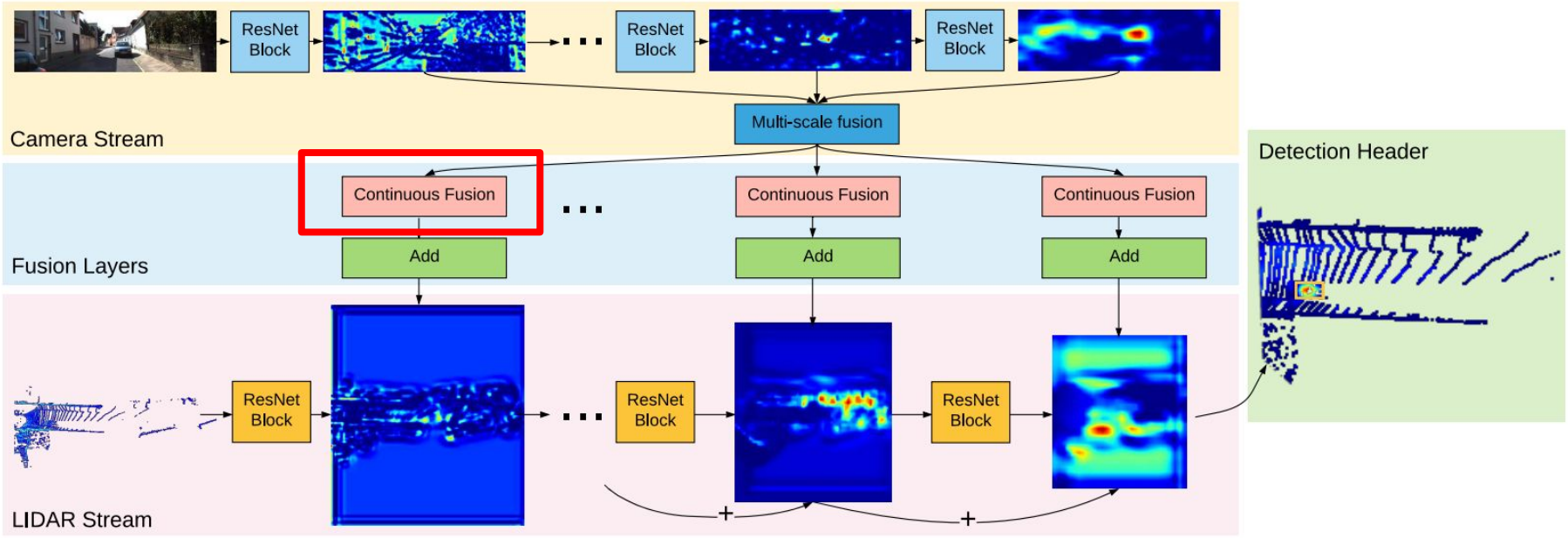


State of the art methods

3D Object Detection with low level sensor fusion:

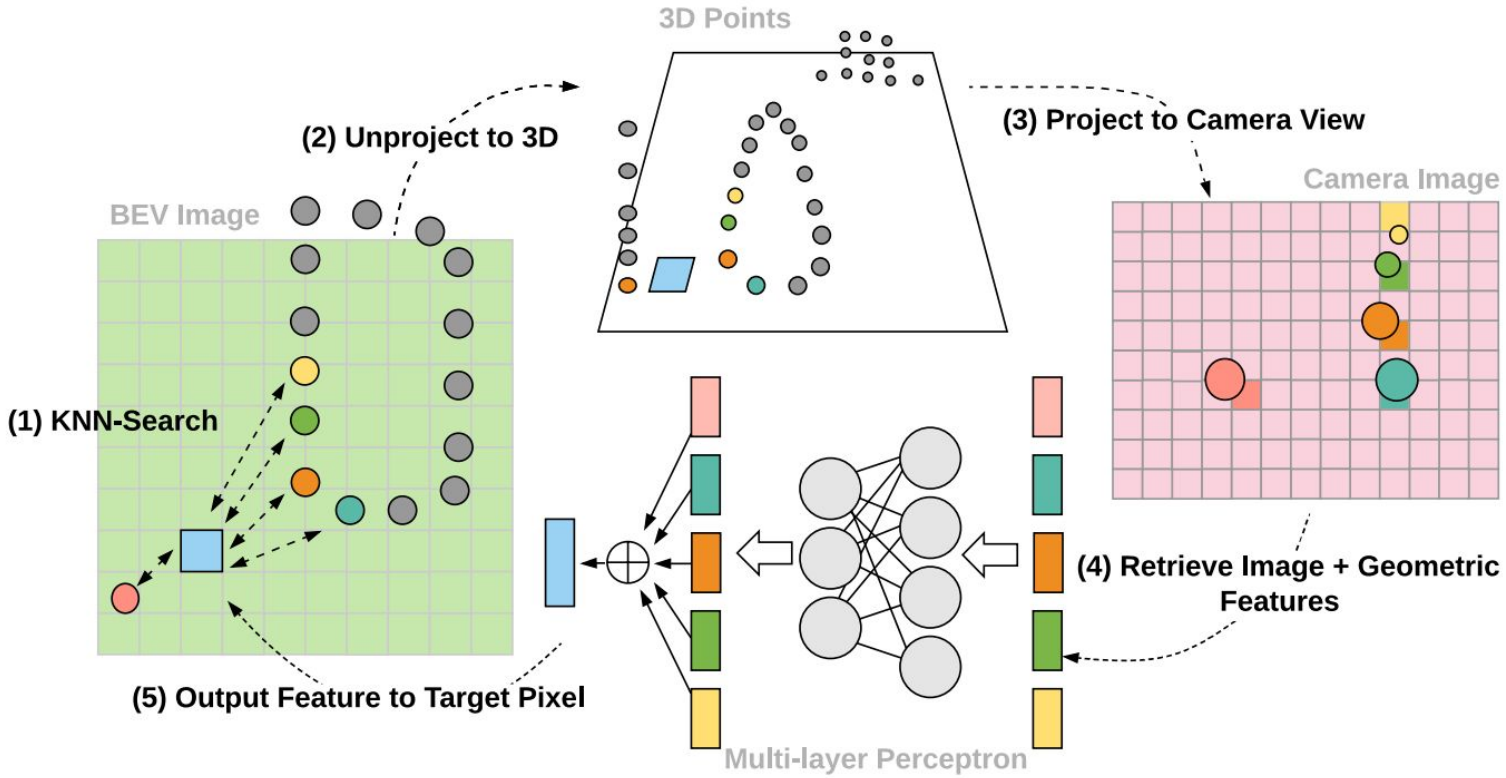
3. Deep Continuous Fusion for Multi-Sensor 3D Object Detection(ECCV,2018)

project image pixel into BEV(Bird Eye View) and adapt Continuous Fusion



State of the art methods

Continuous Fusion Layer

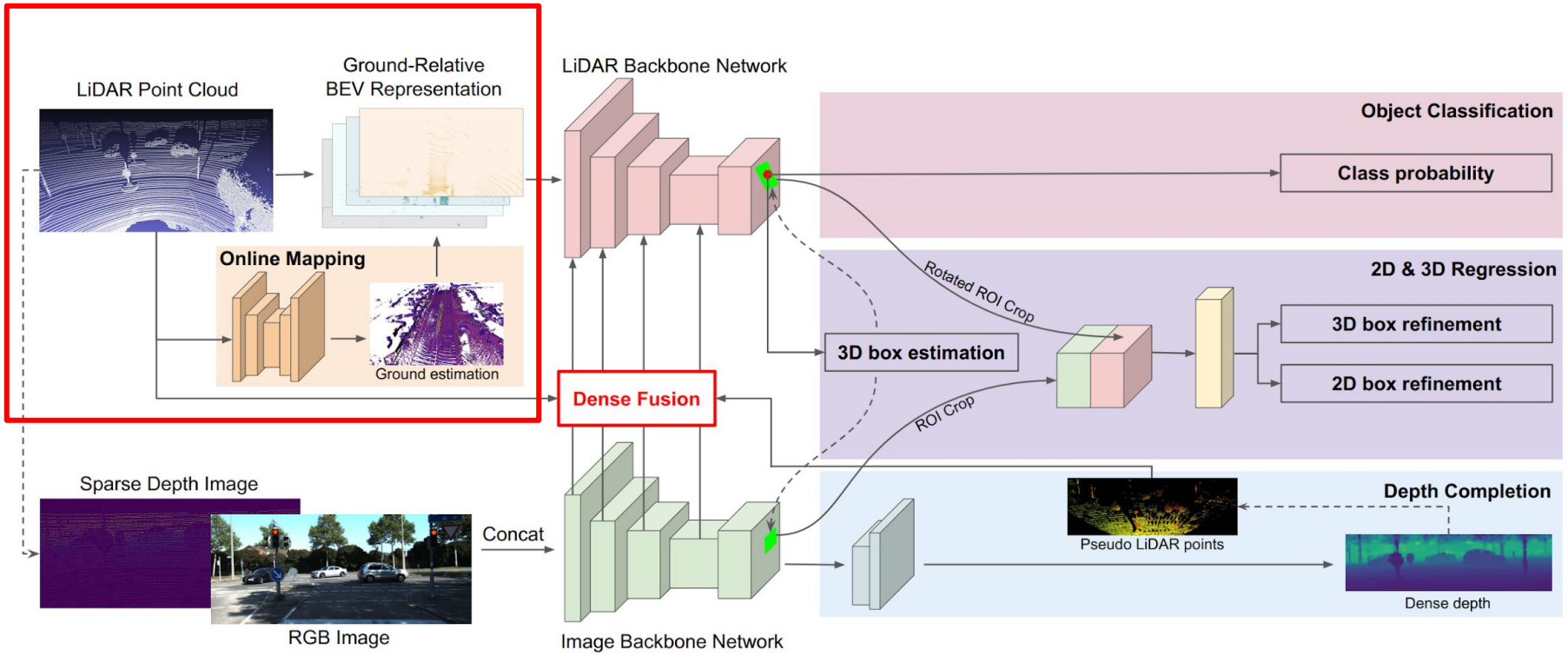


State of the art methods

3D Object Detection with low level sensor fusion:

4. Multi-Task Multi-Sensor Fusion for 3D Object Detection(CVPR,2019)

convert lidar points' height to relative height with respect to ground

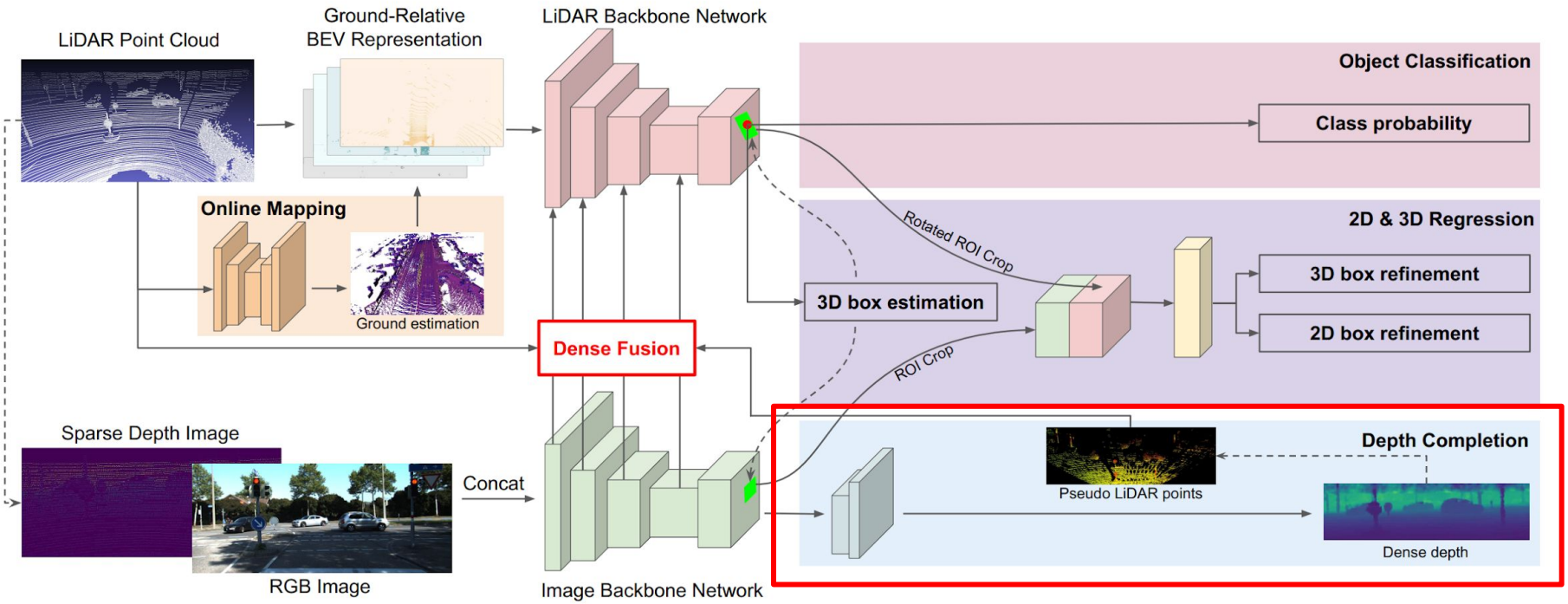


State of the art methods

3D Object Detection with low level sensor fusion:

4. Multi-Task Multi-Sensor Fusion for 3D Object Detection(CVPR,2019)

use KITTI depth completion dataset to create a dense point cloud

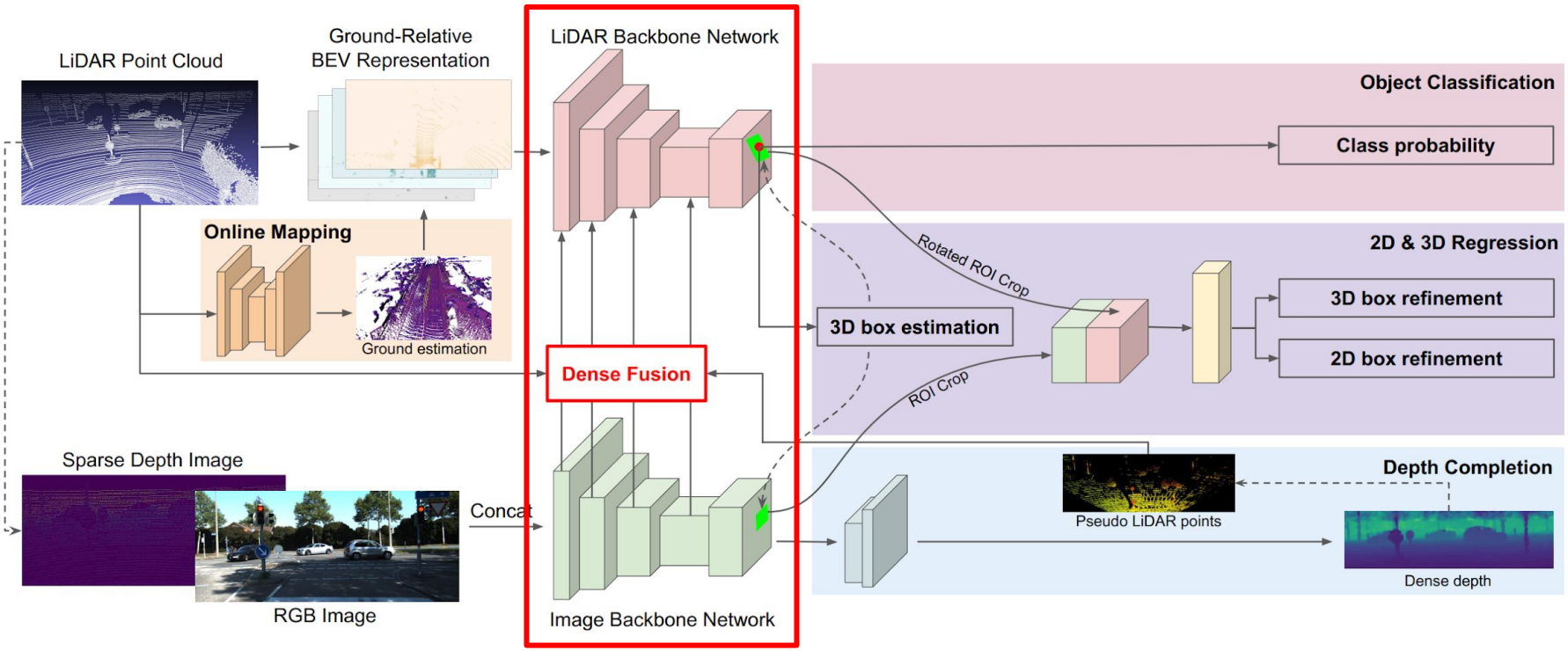


State of the art methods

3D Object Detection with low level sensor fusion:

4. Multi-Task Multi-Sensor Fusion for 3D Object Detection(CVPR,2019)

use dense point cloud to fuse

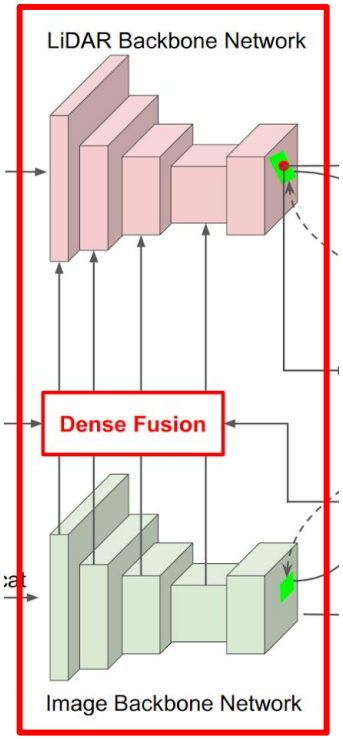


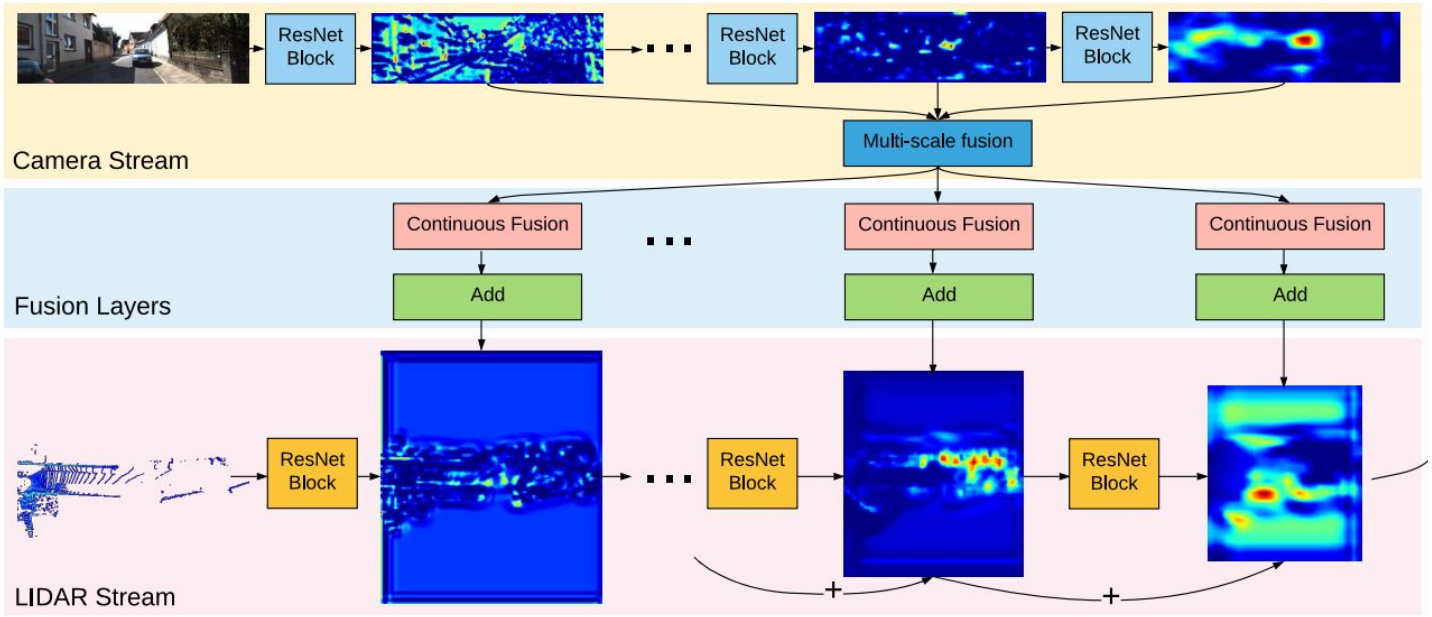
State of the art methods

3D Object Detection with low level sensor fusion:

4. Multi-Task Multi-Sensor Fusion for 3D Object Detection(CVPR,2019)

use dense point cloud to fuse





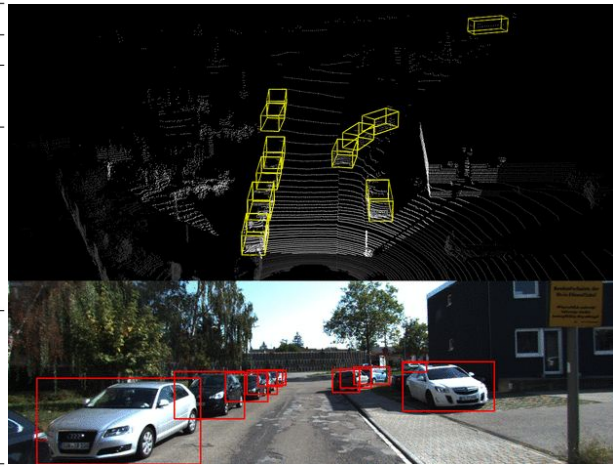
State of the art methods

3D Object Detection with low level sensor fusion:

4. Multi-Task Multi-Sensor Fusion for 3D Object Detection(CVPR,2019)

Detector	Input Data		Time	2D AP (%)			3D AP (%)			BEV AP (%)		
	LiDAR	IMG	(ms)	easy	mod.	hard	easy	mod.	hard	easy	mod.	hard
SHJU-HW [35, 7]		✓	850	90.81	90.08	79.98	-	-	-	-	-	-
RRC [20]		✓	3600	90.61	**90.23**	87.44	-	-	-	-	-	-
MV3D [5]	✓		240	89.80	79.76	78.61	66.77	52.73	51.31	85.82	77.00	68.94
VoxelNet [36]	✓		220	-	-	-	77.49	65.11	57.73	89.35	79.26	77.39
SECOND [32]	✓		50	90.40	88.40	80.21	83.13	73.66	66.20	88.07	79.37	77.95
PIXOR [34]	✓		**35**	-	-	-	-	-	-	87.25	81.92	76.01
PIXOR++ [33]	✓		**35**	-	-	-	-	-	-	89.38	83.70	77.97
HDNET [33]	✓		50	-	-	-	-	-	-	89.14	86.57	78.32
MV3D [5]	✓	✓	360	90.53	89.17	80.16	71.09	62.35	55.12	86.02	76.90	68.49
AVOD [12]	✓	✓	80	89.73	88.08	80.14	73.59	65.78	58.38	86.80	85.44	77.73
ContFuse [13]	✓	✓	60	-	-	-	82.54	66.22	64.04	88.81	85.83	77.33
F-PointNet [17]	✓	✓	170	90.78	90.00	80.80	81.20	70.39	62.19	88.70	84.00	75.33
AVOD-FPN [12]	✓	✓	100	89.99	87.44	80.05	81.94	71.88	66.38	88.53	83.79	77.90
Our MMF	✓	✓	80	**91.82**	90.17	**88.54**	**86.81**	**76.75**	**68.41**	**89.49**	**87.47**	**79.10**

Table 1. Evaluation results on the testing set of KITTI 2D, 3D and BEV object detection benchmark (car). We compare with previously published detectors on the leaderboard ranked by Average Precision (AP) in the moderate setting.



MMF (fusion)

Method	Modality	Car (IoU=0.7)			Pedestrian (IoU=0.5)			Cyclist (IoU=0.5)		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
MV3D [4]	RGB + LiDAR	71.09	62.35	55.12	-	-	-	-	-	-
UberATG-ContFuse [17]	RGB + LiDAR	82.54	66.22	64.04	-	-	-	-	-	-
AVOD-FPN [14]	RGB + LiDAR	81.94	71.88	66.38	50.80	42.81	**40.88**	64.00	52.18	46.61
F-PointNet [25]	RGB + LiDAR	81.20	70.39	62.19	**51.21**	**44.89**	40.23	71.96	56.77	50.39
VoxelNet [43]	LiDAR	77.47	65.11	57.73	39.48	33.69	31.51	61.22	48.36	44.37
SECOND [40]	LiDAR	83.13	73.66	66.20	51.07	42.56	37.29	70.51	53.85	46.90
Ours	LiDAR	**85.94**	**75.76**	**68.32**	49.43	41.78	38.63	**73.93**	**59.60**	**53.59**

Table 1. Performance comparison of 3D object detection with previous methods on KITTI *test* split by submitting to official test server. The evaluation metric is Average Precision(AP) with IoU threshold 0.7 for car and 0.5 for pedestrian/cyclist.

PointRCNN (lidar only)

Plan for next few weeks

- Implement some of the available networks like PointRCNN, VoxelNet for 3D Object Detection from point clouds in Lidar
- Implement object detection from monocular images from camera
- Implement fusion network
- Compare the performances from all the different networks.

Thank you!