Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving

CVPR 2019

Summary

In this paper a novel framework for 3D object detection from image data, called Pseudo-LiDAR, is put forward. The authors propose to extract point cloud-like data from the output of a depth estimation model and use it as input to a LiDAR-based 3D object detector. It is argued that the performance gap between image-based methods and LiDAR-based method can be attributed to the representation of the data rather than the inaccurate estimation of depth. The Pseudo-LiDAR framework is independent of the algorithms used for depth estimation an object detection. Experiments show that the accuracy of image-based 3D object detection can be increased substantially by leveraging Pseudo-LiDAR representation without modifications to the underlying depth estimation algorithm.

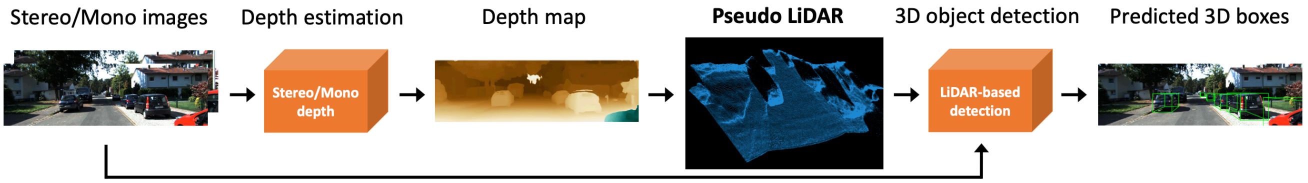


Figure 1: Illustration of a detection pipeline under the Pseudo-LiDAR framework. The projection from depth map to Pseudo-LiDAR can be computed efficiently and object detectors can make use of either the Psuedo-LiDAR alone or combine it with the original input images used for depth estimation. Thus, a variety of algorithms can be applied without changing the input pipeline.

Main Contributions

- **Generic Object Detection Framework** The Pseudo-LiDAR framework presents an object detection pipeline that can be used with any depth estimation algorithm for mono or stereo images as well as a any LiDAR-based object detector. Besides the improvements of detection accuracy, Pseudo-LiDAR allows for more unified comparisons of object detections algorithms using different sensor modalities.

- **Analysis of Data Representation for Depth Estimation** The authors present an analysis on why 3D object detectors using conventional depth maps struggle to achieve good performance despite accurate depth information. It is argued that scale variability and image proximity of objects that are far apart in real-world space violate fundamental assumptions of convolutional operators and therefore degrade the performance of image-based methods.

- **Extensive Evaluation on KITTI 3D Object Detection Benchmark** A large number of experiments is conducted to evaluate the performance of the Pseudo-LiDAR framework under different combinations of well-known depth estimators and object detectors. Results are reported for depth estimation from both, mono and stereo images, using variants of DispNet and PSM-Net. For object detection, AVOD and F-PointNet are used.

Implementation Details

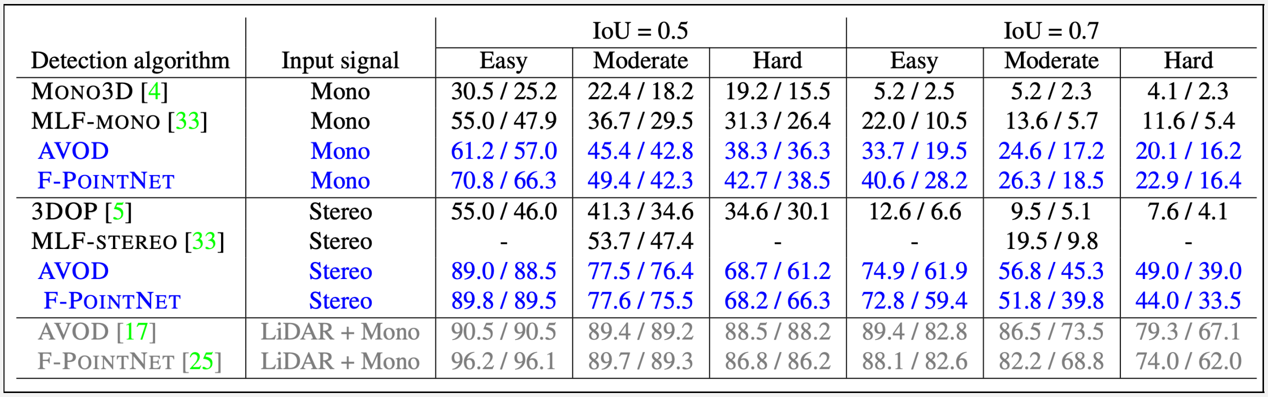
- **Pseudo-LiDAR Generation** Given a depth map $z = D(u, v)$ for each pixel in the input image, the corresponding x and y coordinates in real-world metric space can be computed according to the following formulas:

$$x = \frac{((u - c_U) \times z)}{f_U} \qquad y = \frac{((v - c_V) \times z)}{f_V}$$

with (c_U, c_V) being the pixel corresponding to the camera center. f_U and f_V denote horizontal and vertical focal lengths. In order to ensure consistency with real LiDAR data, projected points above a certain height are removed. The number of points is determined by the number of pixels in the input image and roughly corresponds to the number of points in a high-resolution LiDAR scan. Due to missing information, the reflectance value of all Pseudo-LiDAR points is set to 1.

Evaluation

- **KITTI 3D and BEV Object Detection** The main evaluation is conducted on the KITTI validation set. Results are reported on all three classes (Easy, Medium, Hard) for both, IoU 0.5 and IoU 0.7. The authors evaluate two baseline methods for both, mono and stereo images and compare them to the performance of Pseudo-LiDAR methods with AVOD and F-PointNet. Furthermore, the performance of AVOD and F-PointNet using real LiDAR data is reported. The results show a substantial increase in performance for image-based detectors using Pseudo-LiDAR compared to conventional data representations. Although real LiDAR data still outperforms Pseudo-LiDAR by large margins, especially on difficult examples, the performance increase compared to previous methods underlines the huge potential of the Pseudo-LiDAR framework.



		IoU = 0.5			IoU = 0.7		
Detection algorithm	Input signal	Easy	Moderate	Hard	Easy	Moderate	Hard
Mono3D [4]	Mono	30.5 / 25.2	22.4 / 18.2	19.2 / 15.5	5.2 / 2.5	5.2 / 2.3	4.1 / 2.3
MLF-mono [33]	Mono	55.0 / 47.9	36.7 / 29.5	31.3 / 26.4	22.0 / 10.5	13.6 / 5.7	11.6 / 5.4
AVOD	Mono	61.2 / 57.0	45.4 / 42.8	38.3 / 36.3	33.7 / 19.5	24.6 / 17.2	20.1 / 16.2
F-PointNet	Mono	70.8 / 66.3	49.4 / 42.3	42.7 / 38.5	40.6 / 28.2	26.3 / 18.5	22.9 / 16.4
3DOP [5]	Stereo	55.0 / 46.0	41.3 / 34.6	34.6 / 30.1	12.6 / 6.6	9.5 / 5.1	7.6 / 4.1
MLF-stereo [33]	Stereo	-	53.7 / 47.4	-	-	19.5 / 9.8	-
AVOD	Stereo	89.0 / 88.5	77.5 / 76.4	68.7 / 61.2	74.9 / 61.9	56.8 / 45.3	49.0 / 39.0
F-PointNet	Stereo	89.8 / 89.5	77.6 / 75.5	68.2 / 66.3	72.8 / 59.4	51.8 / 39.8	44.0 / 33.5
AVOD [17]	LiDAR + Mono	90.5 / 90.5	89.4 / 89.2	88.5 / 88.2	89.4 / 82.8	86.5 / 73.5	79.3 / 67.1
F-PointNet [25]	LiDAR + Mono	96.2 / 96.1	89.7 / 89.3	86.8 / 86.2	88.1 / 82.6	82.2 / 68.8	74.0 / 62.0

Figure 2: Results of the main evaluation of KITTI validation set. Reported scores are AP_{BEV} and AP_{3D}. The results give a comprehensive insight into the capabilities of the Pseudo-LiDAR representation. Especially the results for Pseudo-LiDAR based on stereo depth estimation indicate high potential to further close the gap between image-based and LiDAR-based object detection.

References

This summary is solely based on my understanding of the original paper. All images used here are taken from the original paper as well. The paper can be found under the following link:
`https://arxiv.org/pdf/1812.07179.pdf`