Robust 3D Object Detection for Autonomous Vehicles using Sensor Fusion

Mohan Rao Divate Kodandarama
University of Wisconsin-Madison
Madison, WI
divatekodand@wisc.edu

Aditya Rungta
University of Wisconsin-Madison
Madison, WI
arungta@wisc.edu

Abstract

Accurate detection of objects in 3D is a central problem in autonomous navigation. Although, there are some works in this area, most of them suffer from one of these problems - very high latency, poor accuracy at far off range and poor accuracy on minority classes. In this work, we try to address some of these by fusing multi-sensor (Image and LiDAR) information and temporal information.

1. Problem Definition and Motivation

It is becoming clear that the future transportation would be dominated by autonomous vehicles. Autonomous vehicle technology is already being developed by the likes of Lexus, BMW and Mercedes, Waymo and Apple. A central problem in autonomous navigation is the accurate detection of 3D objects around the autonomous vehicle in real-time. More formally, the problem is to use data form multiple sensors (cameras, LiDAR's, and RADAR) embedded on the autonomous vehicles and output robust detection of objects around the autonomous vehicle. Although, there have been a plethora of works in this field, all of them have at least one of the following limitations -

1. Very high inference latency.
2. Low detection accuracy for far off objects and minority class.
3. Fails to fuse data from multiple sensors (e.g. Image and LiDAR).

In our work, we intend to address these limitations of current state of the art methods by utilizing both image and LiDAR data. If time permits, we would also like to utilize temporal information.

2. Related Work

[6] is one of the first papers that proposed an architecture for processing point cloud data. [7] improves [6] by utilizing hierarchical features and weight sharing. [9] further improves [6] by modelling the relationship between the points in a point cloud using graph convolutions.

[11] is one of the first papers that proposed an end-to-end learnable system for 3D object detections using point cloud data. It divides the 3D world into voxels, computes voxel-wise features and employs 3D convolutional network as a detection head .However, expensive 3D convolution operations resulted in very high inference latency. [4] refines the architecture of [11] by eliminating 3D convolution operations and utilizing better loss function.

[5] uses LiDAR data as range view images. It uses a fully convolutional network to predict a multimodal distribution over 3D boxes for each point and then it efficiently fuses these distributions to generate a prediction for each object.

3. Approaches and Evaluation

3.1. Approaches

To keep the inference time low, we would plan to use the LiDAR data as range view images. We would upscale the range view image using some of the recent advances in field of super resolution and use it with camera images in a fully convolutional architecture.

3.2. Evaluation

We would use mean Average Precision (mAP) as described in [2] at different ranges to evaluate the performance of detection algorithms. More formally, we would report mAP scores considering the objects within the range of 30 meters, 50 meters and 70 meters. This would help evaluate the variation in performance at different ranges from the autonomous vehicle.

3.3. Datasets

We would use the following datasets for all our experiments.

1. lyft
2. argo
3. waymo

4. nuScenes
5. kitti

References

[1] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[2] Mark Everingham, Luc Gool, Christopher K. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision*, 88(2):303–338, June 2010.

[3] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven Lake Waslander. Joint 3d proposal generation and object detection from view aggregation. *CoRR*, abs/1712.02294, 2017.

[4] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. *CoRR*, abs/1812.05784, 2018.

[5] Gregory P. Meyer, Ankit Laddha, Eric Kee, Carlos Vallespi-Gonzalez, and Carl K. Wellington. Lasernet: An efficient probabilistic 3d object detector for autonomous driving. *CoRR*, abs/1903.08701, 2019.

[6] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *CoRR*, abs/1612.00593, 2016.

[7] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *CoRR*, abs/1706.02413, 2017.

[8] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. *CoRR*, abs/1812.04244, 2018.

[9] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph CNN for learning on point clouds. *CoRR*, abs/1801.07829, 2018.

[10] Hengshuang Zhao, Li Jiang, Chi-Wing Fu, and Jiaya Jia. Pointweb: Enhancing local neighborhood features for point cloud processing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[11] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection, 2017.