CLOC network for object detection - Project proposal

Aneesh Chodisetty
Department of Computer Science
University of Maryland
aneeshch@umd.edu

Koundinya Vinnakota
Department of Computer Science
University of Maryland
vkd@umd.edu

Sai Surya Sriramoju
Department of Computer Science
University of Maryland
saisurya@umd.edu

1 Motivation

It is to nobody's surprise that the growing demand of product availability everywhere increases the need for more transportation services that take care of logistics. One of the best ways to improve the logistics efficiency is by making the driving part of the logistics as autonomous as possible. This means that the vehicle that is driving itself needs to identify the objects around it so that it can navigate around safely.
Our project focuses on suggesting a method that would help analyze the surroundings and identify 3D objects in real-time using a Camera-LiDAR Object Candidates (CLOCs) fusion network. CLOCs fusion provides a low-complexity multi-modal fusion framework that significantly improves the performance of single-modality detectors. Any 2D(camera) and any 3D(LiDAR) detector's combined output candidates are used by CLOCs, which is trained to take advantage of their geometric and semantic consistency to provide more accurate final 3D and 2D detection results.

2 Literature Survey

[4] A Survey on 3D Object Detection Methods for Autonomous Driving Applications by Eduardo Arnold et .al, this paper gives us the fundamental reasoning behind the need of 3D detection methods and categorizes different detection methods into categories of monocular , point cloud based and fusion methods. Also a limitation of KITTI dataset has been presented regarding the characteristic daylight scenes and very standard weather conditions of the data. The conclusion of the paper states that monocular methods are not reliable and fusion methods are very useful in each modality to achieve state of the art results.

[5] PointPillars: Fast Encoders for Object Detection from Point Clouds by Alex H. Lang et. al. This paper talks about a novel approach of Point pillars that can be used to get pint cloud data using pointNets. Various experiments have been conducted and the results showcase that PointPillars outperform outperform encoders in terms of speed and accuracy drastically. Though only LiDAR has been used the entire detection pipeline seems to perform well even among fusion methods with some shortcomings.

[6] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection,"- This paper talks about the existing end to end learning of point cloud data using VoxelNet, a generic 3D detection network that unifies feature extraction and bounding box prediction into a single stage.

[7] Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy by DuarteFernandes et. al in this paper various advancements in the field of perception using LIDAR data have been clearly analysed. More specifically the main focus is of the point cloud data high dimensional and sparse nature. Comparison between different performance results have been analysed. the paper concludes by stating the the superior accuracy of LiDAR and field of view compared to radar based solutions. Also it talks about various potential methods

that have not been fully explored the complete understanding of the model in its different stages is fundamental to address.

3 Planned Methods

For this project we have planned to use the KITTI data-set to train our model using the CLOCs fusion network architecture. The paper suggested that CLOCs exploits the geometric and semantic consistencies between 2D and 3D detection and automatically learns fusion parameters. Also CLOCs is ment to work on any 2D and 3D detector combination.

4 Planned Experiments

In the cited paper [1], the detectors used for 2D are: RRC, MS-CNN and Cascade R-CNN; and the 3D detectors are: SECOND, PointPillars, PointRCNN and PV-RCNN. For our project we plan to use YOLOv5 as the 2D detector and complexer YOLO as the 3D detector and compare the results with the ones in the paper. Individually YOLOv5 and complexer YOLO do not have that high of mAP but they are fast, by combining those two using the CLOCs framework, we would like to check performance of the combined models.

References

[1] S. Pang, D. Morris and H. Radha, "CLOCs: Camera-LiDAR Object Candidates Fusion for 3D Object Detection," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 10386-10393, doi: 10.1109/IROS45743.2020.9341791.

[2] W. Souidene Mseddi, M. A. Sedrine and R. Attia, "YOLOv5 Based Visual Localization For Autonomous Vehicles," 2021 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 746-750, doi: 10.23919/EUSIPCO54536.2021.9616354.

[3] M. Simon et al., "Complexer-YOLO: Real-Time 3D Object Detection and Tracking on Semantic Point Clouds," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1190-1199, doi: 10.1109/CVPRW.2019.00158.

[4] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby and A. Mouzakitis, "A Survey on 3D Object Detection Methods for Autonomous Driving Applications," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 10, pp. 3782-3795, Oct. 2019, doi: 10.1109/TITS.2019.2892405.

[5] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang and O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 12689-12697, doi: 10.1109/CVPR.2019.01298.

[6] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4490-4499, doi: 10.1109/CVPR.2018.00472.

[7] Duarte Fernandes, António Silva, Rafael Névoa, Cláudia Simões, Dibet Gonzalez, Miguel Guevara, Paulo Novais, João Monteiro, Pedro Melo-Pinto, Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy, Information Fusion, Volume 68, 2021, ISSN 1566-2535