Investigation of LiDAR-based 3D Object Detection

Ashley Mojica
Dept. of Computer Science
University of Central Florida
Ashley.Mojica@knights.ucf.edu

Molika So
Dept. of Computing
University of North Florida
Molika.So.23@gmail.com

Muhammad Shahbaz
Dept. of Civil, Environmental & Construction Engineering
University of Central Florida
Mshahbaz@knights.ucf.edu

Shaurya Agarwal
Dept. of Civil, Environmental & Construction Engineering
University of Central Florida
Shaurya.Agarwal@ucf.edu

Damla Turgut
Dept. of Computer Science
University of Central Florida
Damla.Turgut@ucf.edu

Samiul Hasan
Dept. of Civil, Environmental & Construction Engineering
University of Central Florida
Samiul.Hasan@ucf.edu

Abstract

Light Detection and Ranging (LiDAR) sensors have recently become a product of interest for autonomous driving, in contrast to cameras, because of their ability to collect in-depth 3D data (point clouds) that are not impacted by lighting/weather changes. Various detection methods [2] have been developed to process this 3D data using machine learning to detect objects. However, obtaining and processing the large amount of labeled data required for training can be costly. In our research, we analyzed the relationship between mean average precision (mAP), the amount of data utilized in training, and the time required to perform training for car, pedestrian, and cyclist detection.

Methods

- Utilized the open-source KITTI dataset [3]
- Collected sample sets with sizes incrementing by 500 up to approximately 7,500 frames (15 sets in total)
- Each set is used to train a model with the Sparsely Embedded Convolutional Detection (SECOND) method (Figure 1) [4]
- The model was tested on a controlled test set of 7,518 frames
- The mAP for the detected cars, pedestrians, and cyclists were calculated from each test
  - Mean Average Precision (mAP) is calculated by averaging the Average Precision (AP) of each class [2]
- The entirety of the experimentation utilized two NVIDIA GeForce RTX 2080 Ti GPUs

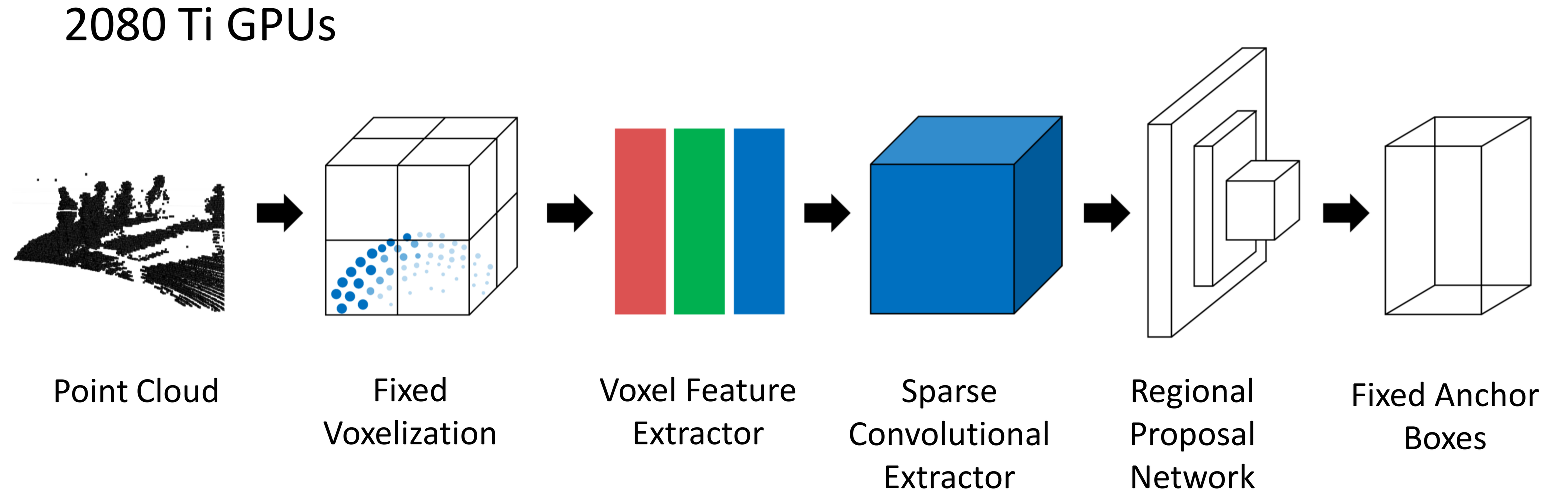

Figure 1. The SECOND (Sparsely Embedded Convolutional Detection) method pipeline used in experimentation as proposed by [1].

Limitations

- Quality/quantity of the GPUs used for experimentation were not ideal
- Newer datasets are incompatible with the SECOND implementation
- Dataset is small compared to newer datasets

Results

- **Easy:** Min. bounding box height: 40 Px, Max. occlusion level: Fully visible, Max. truncation: 15 % [3]
- **Moderate:** Min. bounding box height: 25 Px, Max. occlusion level: Partly occluded, Max. truncation: 30 % [3]
- **Hard**: Min. bounding box height: 25 Px, Max. occlusion level: Difficult to see, Max. truncation: 50 % [3]

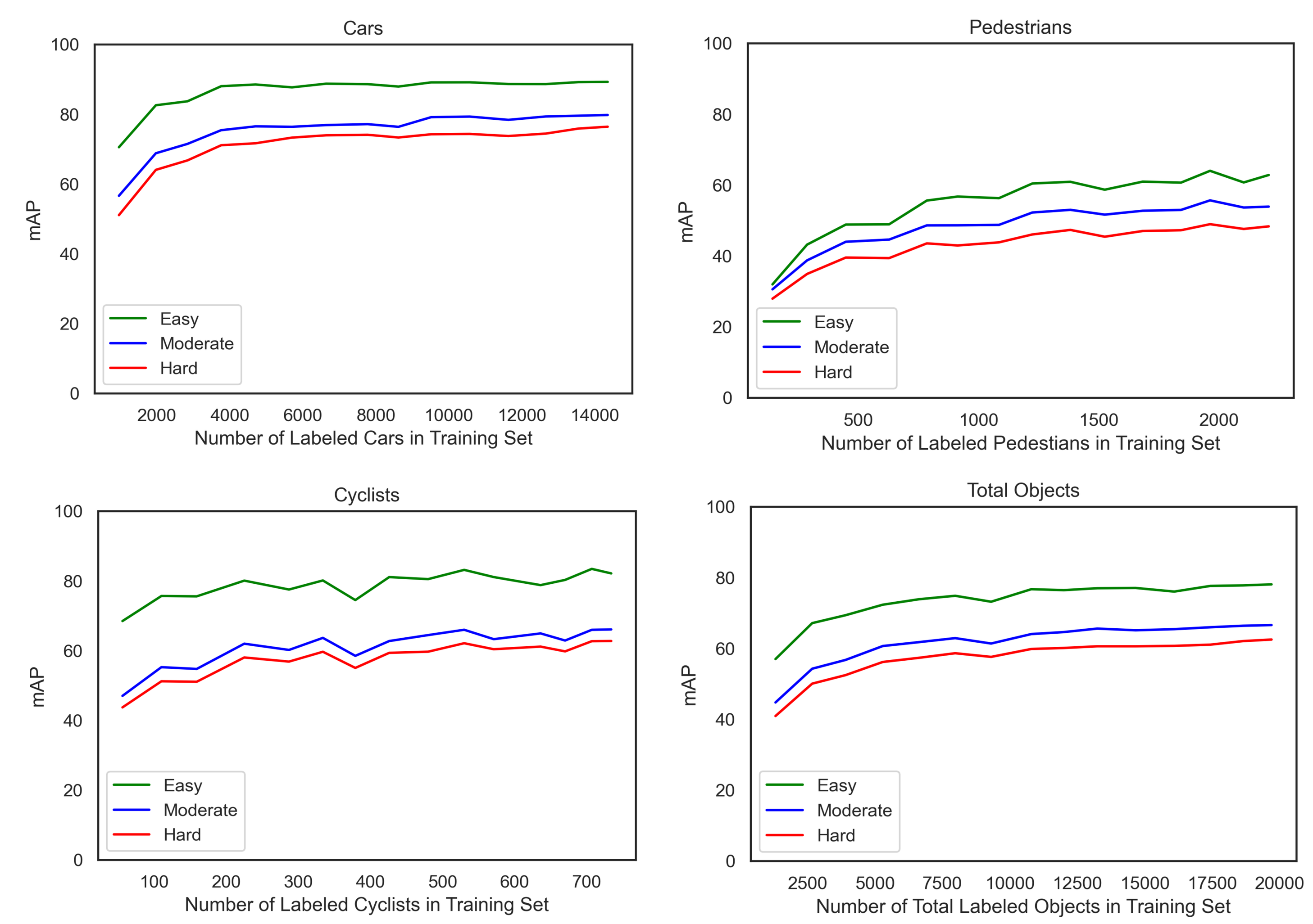

Figure 2. The mAP vs the number of cars, pedestrians, and cyclists used in training.

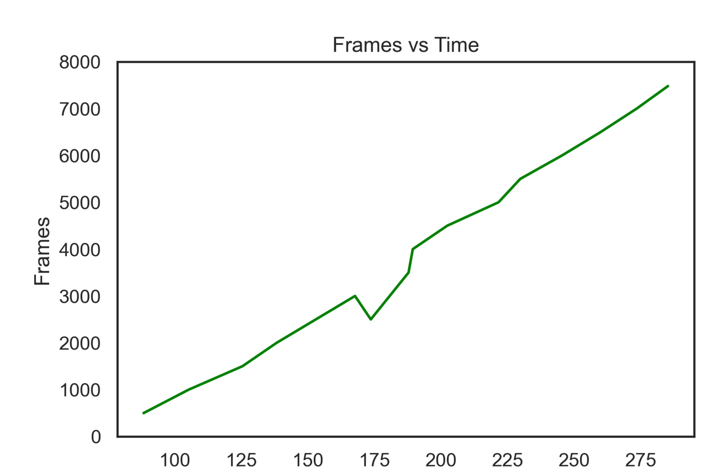

Figure 3. The relationship between frames and training time.

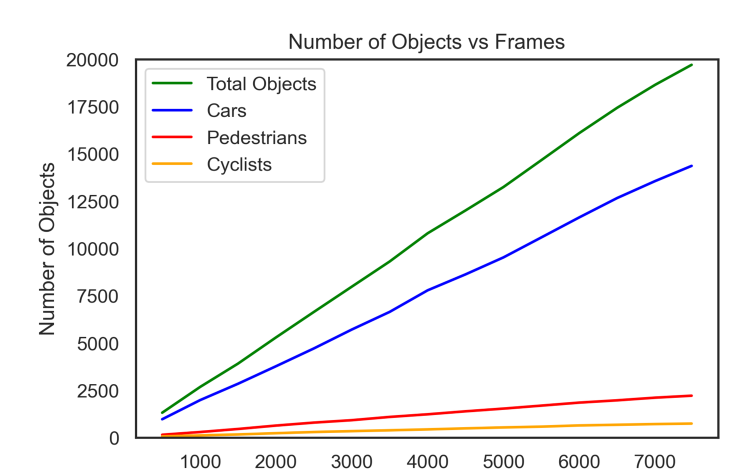

Figure 4. The relationship between the number of objects and frames used in training.

Analysis

- As the number of objects used for training increases, the mAP increases quickly and levels out
- There were fewer cyclists in the dataset than cars and pedestrians, resulting in less improvement in mAP
- The number of each type of object varied in each frame used for training

Point Cloud Visualization

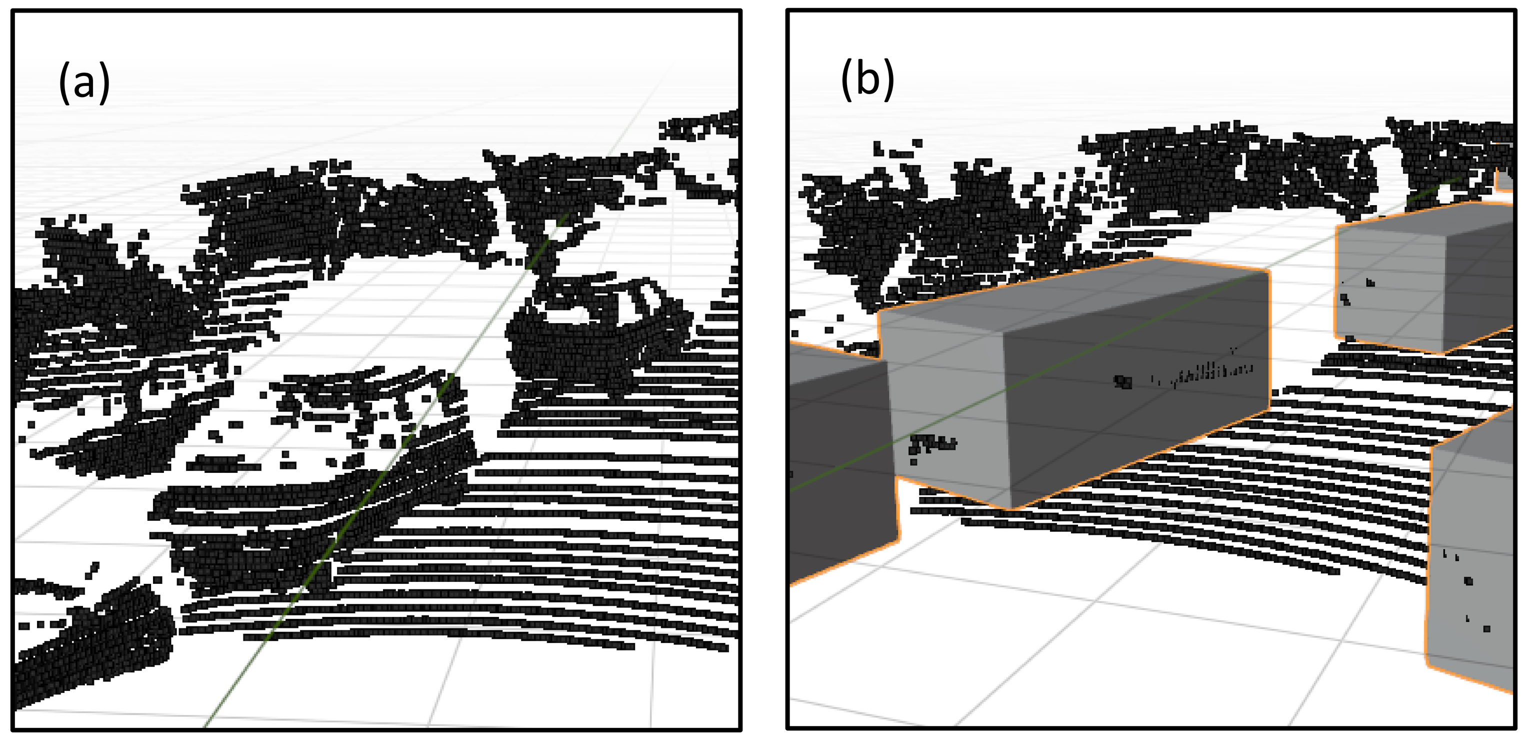

Figure 5. A sample from the KITTI dataset before (a) and after detection (b).

Conclusion & Future Work

- Sample size becomes insignificant at a certain threshold once mAP improvement stabilizes (the architecture needs improvements)
- Based on the car data, the optimal number of samples needed to reach ideal saturation is approximately 4000
- Future work may include:
  - Further investigation of transferability issues using datasets with different camera calibration and physical environments
  - Increase mAP with decreasing training set size

References

1. Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," Sensors, vol. 18, no. 10, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/10/33377
2. G. Zamanakos, L. Tsochatzidis, A. Amanatiadis, and I. Pratikakis, "A comprehensive survey of lidar-based 3d object detection methods with deep learning for autonomous driving," Computers & Graphics, vol. 99, pp. 153–181, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0097849321001321
3. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
4. M. Contributors, "MMDetection3D: OpenMMLab next-generation platform for general 3D object detection," https://github.com/openmmlab/mmdetection3d, 2020.

Acknowledgment

The support for this work was provided by the National Science Foundation REU program under Award No. 1852002. Any opinions, findings, and conclusions and recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.