Università di Parma

3D Perception, Learning-Based Data Fusion Exam

3D Object Detection On NuScenes

Francesco Marotta
Simone Maravigna

November 13, 2023

Project Report

Goal of the Project

The goal of this project is to create an object detection system tailored for autonomous vehicles navigating urban environments by integrating LiDAR and camera inputs. Our approach involves leveraging the Faster R-CNN[1] model initialized with pretrained weights from the COCO dataset and fine-tuned specifically for 2D detection on camera images.
Additionally, for 3D detection on the LiDAR point cloud, we utilized SSN[2] (Shape Signature Networks) pretrained on the NuScenes[3] dataset.

Dataset

The nuScenes dataset is a comprehensive and diverse dataset designed for autonomous vehicle research and development. It provides a rich collection of sensor data, including lidar, radar, and camera information, captured from a variety of urban driving scenarios. Notably, for the purposes of our project, we utilized the mini version of the nuScenes dataset due to memory constraints, allowing us to efficiently manage and process the dataset while retaining its richness in content.
The dataset is structured as a relational database, where each row in every table can be uniquely identified by its primary key token.
Therefore, the data loader is able to navigate in such database to retrieve the required information and format them appropriately.

Sensors

- Spinning LiDAR:
  - Capture Frequency: 20Hz
  - Field of View (FOV): 360° Horizontal, +10° to -30° Vertical (Uniform Azimuth Angles)
  - Range: 80m-100m, Usable Returns up to 70 meters
  - Accuracy: ±2 cm
  - Points per Second: Up to 1.39 Million
- Camera:
  - Capture Frequency: 12Hz
  - Lens: Evetar Lens F1.8 f5.5mm 1/1.8”
  - Sensor: 1/1.8” CMOS sensor of 1600x1200 resolution
  - Region of Interest (ROI): 1600x900 ROI is cropped from the original resolution to reduce processing and transmission bandwidth

In the pursuit of generating a high-quality multi-sensor dataset, the calibration of both extrinsic and intrinsic parameters for each sensor is imperative. Extrinsic coordinates are defined in relation to the ego frame, specifically the midpoint of the rear vehicle axle.
To ensure robust cross-modality data alignment between LiDAR and cameras, the camera's exposure is synchronized with the top LiDAR sweep, precisely when it traverses the center of the camera's FOV. This method, owing to the nearly instantaneous exposure time of the camera, consistently results in effective data alignment.
It is noteworthy that the cameras operate at a frequency of 12Hz, while the LiDAR operates at 20Hz. To optimize data distribution, the 12 camera exposures are evenly distributed across the 20 LiDAR scans. Consequently, not every LiDAR scan corresponds to a camera frame due to the frequency disparity.

Inference and Fusion

We utilized the identical model and weights from the prior project for performing 2D object detection on the images.
Additionally, for 3D object detection on LiDAR point clouds, we employed the mmdetection[4] suite, with a specific focus on integrating SSN (Shape Signature Networks).
Following the acquisition of both sets of predictions, we applied translation matrices inherent in the dataset to project them onto the image plane and we were able to extract some information and some demo.

Results

The results indicate that the model excels in recognizing nearby objects when utilizing LiDAR, whereas with the camera, it demonstrates the capability to retrieve information even from distant, small objects.
It is crucial to note that we did not train the LiDAR model, and due to computational resource constraints, we selected the lightest model available from the mmdetection model portfolio. Consequently, with a more sophisticated model, such as PointPillars, we anticipate achieving superior and more accurate 3D predictions.

Figure 1: Example of an input raw image

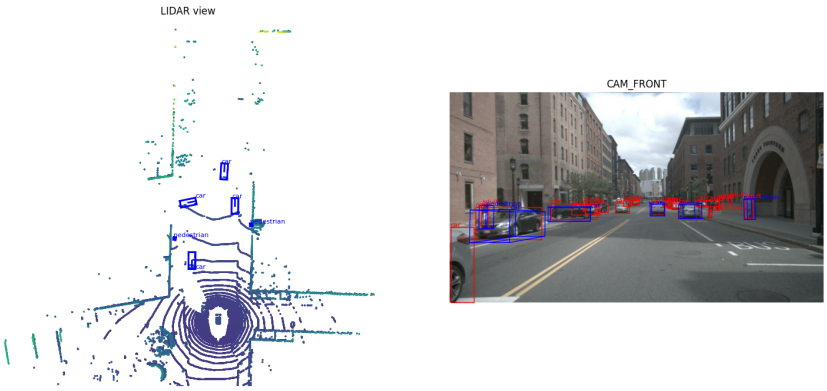

Figure 2: Example the LiDAR output (left) and the fused detection output (right)

References

[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[2] Xinge Zhu, Yuexin Ma, Tai Wang, Yan Xu, Jianping Shi, and Dahua Lin. *SSN: Shape Signature Networks for Multi-class Object Detection from Point Clouds*, pages 581–597. 11 2020.

[3] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[4] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.