Intention Prediction in Vehicular Environments

Mitch Mathieu
Supervisor: Sameh Sorour
CISC 499

Background

As autonomous vehicles (AVs) become commonplace on our roads, it is important that we build decision making models that are strong and robust. One of the shortcomings of current AV models is the inability to understand the intentions of other drivers on the road, a skill that human drivers possess innately.

With the use of advanced machine learning techniques, it should be possible to give AV models the ability to predict the near-future intentions of other drivers. Using these predictions, our autonomous vehicles should be able to improve both the safety and efficiency of their decision making.

Data

The data used for this project came from The KITTI Vision Benchmark Suite, a project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago. KITTI provides raw data recordings for autonomous vehicles, including colour stereo image sequences and 3D Velodyne point clouds, both of which were used in this project.

KITTI also provides a range of tools for training and evaluating 3D object detection models, including 7481 training images and 7518 test images as well as the corresponding point clouds. All together, KITTI offers a total of 80,256 labeled objects. These tools were used to train and test the underlying object detection framework of this project.

Approach

The proposed model for this project was a deep neural network with two input streams: RGB colour images and 3D LiDAR point clouds. The backbone of the model is a convolutional neural network (CNN) that processes both inputs and provides 3D bounding box predictions as a vector. The head of the network is a linear regression model trained to predict the position, size and orientation of the bounding box at certain time steps in the future.

An unofficial TensorFlow implementation of VoxelNet was chosen as the underlying object detection model. VoxelNet is an end-to-end trainable deep learning network that uses voxel-based feature extraction to perform 3D object detection straight on the raw LiDAR data. VoxelNet is also a region proposal network, making it faster than standard CNNs and therefore better equipped for vehicular environments.

Each car detected in a sequence of images was assigned a unique identifier and linear regression model. An algorithm was designed to track the cars from image to image and associate new bounding boxes with the correct car. The regression models were trained only on bounding boxes belonging to that same car. This regression model was then used to predict the bounding box position for that car in near-future times.

Results

Object Detection

The VoxelNet implementation was able to correctly detect cars and provide accurate 3D bounding boxes for each instance. The average precision (AP) of the predictions is 53.43, 48.78 and 48.06 for easy, moderate and hard detection, respectively, when tested on the KITTI 3D object detection benchmark. AP is the standard metric used to compare models on the KITTI benchmark.

Object Tracking and Intention Prediction

The object detection algorithm was able to consistently match boxes to the correct car and track each car across image frames, allowing accurate regression models to be built for each car.

The regression model for each car was able to provide appropriate bounding box predictions for near-future times (Fig 3). The average mean absolute error (MAE) for the x, y, and z coordinates of the future boxes was between 0.05 and 2.12, depending on the car.

Conclusions

This research has shown that LiDAR and image data from autonomous vehicle sensors can be used not only to detect cars in the field of view, but to track them through the scene and predict their position and orientation in the near-future as well. The ability to track objects through a scene and understand where they are heading is of great value to an autonomous vehicle and can hopefully allow AVs to make better decisions that improve overall safety and efficiency.

The research accomplished in this project provides a framework for future work in LiDAR-based, vehicular intention prediction models.