MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

ENGINEERING HONORS PROGRAM
UNIVERSITY OF MICHIGAN

Gatik

M city
UNIVERSITY OF MICHIGAN

Improving Long-Range 3D Object Detection Methods for Autonomous Box Trucks using Sensor Fusion

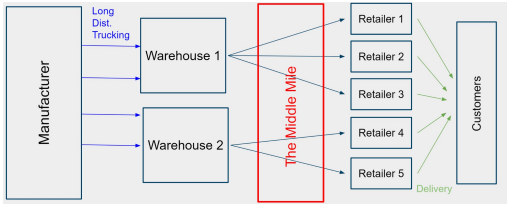
Ashutosh Bhown (Honors Capstone), Rahul Srinivasan, Ruohua Li, Callie Hastie, Nick Moroz (Advisor)

Background

Gatik AI is the first autonomous trucking startup to focus on automating business-to-business delivery with driverless box trucks. They are currently operating with multiple partners such as Walmart in Arkansas and Sam's Club in Dallas.

As Gatik is the first company to bring autonomous trucks to mainly urban and suburban environments, their vehicles face the unique challenge of having to detect smaller agents such as pedestrians and cyclists at long range to have more time to make downstream decisions, which to date has not been heavily explored, as only smaller autonomous cars that do not need as much time to react have had to manage driving in these areas.

One of Gatik's autonomous Class 6 box trucks



Schematic of the logistics supply chain; Gatik's target segment is boxed in red

Experiment

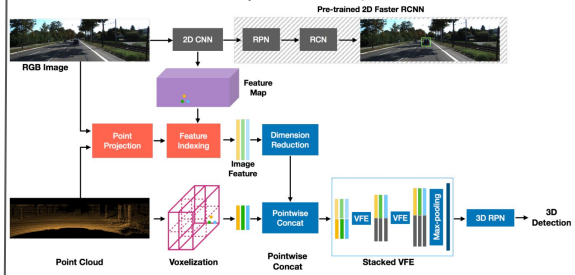
The goal of this project was to improve detection of pedestrians and cyclists at long range in a 3D environment using existing sensor fusion techniques between LiDAR and monocular cameras.

Using the MVXNet architecture as a baseline for performance and evaluating on the KITTI dataset, we sought to improve the architecture such that it would marginally improve detection of pedestrians and cyclists in the "3D/hard" subsections of the KITTI dataset.

Augmented MVXNet architectures* were trained on the KITTI dataset (7,424 images) for 40 epochs and then tested on KITTI's test dataset (3,769 images) to obtain results.

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

KITTI Dataset Easy/Medium/Hard partitions



Overview of MVXNet Sensor Fusion Architecture

*Augmentation details may not be discussed due to Non-Disclosure Agreement

Results

When examining the 3D/Hard sections of the KITTI dataset with 50% overlap with (1) ground truth bounding boxes counting as correct classification and (2) using 40 points to calculate bounding box overlap, we noticed improvement in mean average precision (mAP) of bounding boxes across all classes, especially pedestrian detection.

Performance Comparison of Stock/Augmented MVXNet Architectures

Class	Stock MVXNet mAP	Augmented MVXNet mAP
Ped.	61.69	73.28
Cyclist	51.60	53.94
Car	87.60	90.09

KITTI scenes with predictions made by stock and augmented MVXNets