

# Labeling Standards Report

## Primary Overarching Problems

1. Not enough data in important areas of focus
  - a. Highways
2. Feature issue
  - a. Seems like one of the primary problems of our model is it picks up cars and trucks as “box shaped things.” But not all box-shaped things are trucks/cars. This could be most problematic with stationary box-shaped things in camera view

## Trucks Dataset

I don't know how helpful this is going to be. I did most of the labeling for the trucks dataset and all of the images were zoomed in, single truck images. There was some slight augmentation, but it still wasn't real life focused. It could help our model get off the ground, but it seems like most of our problems were focused on highways. But one thing I noticed online was if we use our Selenium web-scraper to write “trucks on the highway” instead of just “trucks”, we come up with much better searches.

## Pickup Trucks

Our model is not very good with distinguishing pickup trucks from regular cars. In our more recent pre-labeling tasks, it seems to be overshooting on pickup trucks, but also missing them. The primary problem with this is on highways.

## Vans

Similarly to pickup trucks, our model is pretty bad at labeling vans when zoomed out. There doesn't seem to be a ton of vans data in our dataset (only a few hundred based on Luke's distribution data from last week). On the highway, they just get mixed in with regular cars. We could try to fix this similarly to trucks (Web scraper with “vans on the highway”), but when I was on Google Images, the images weren't as useful.

## Billboards

One thing that our model consistently does is pick up billboards as trucks and cars. This could be problematic because if we have some cameras consistently in front of billboards, they could really mess up our numbers. In the end, this will be solved by getting more billboard data on highways in our dataset and making sure everything is labeled correctly.

## Residential Areas

Similarly to billboards, our model isn't great at distinguishing buildings and square-like objects in our images from cars and trucks. The solution is the same with more data and good labeling.

## Double Box Problem

When going through some of the images in “Review” I saw that quite a few images had “double boxes”, or were labeled by the same bounding box twice. I’m not sure if this is a problem or not.

### Questions:

1. How should we label this?



2. How should we label this?

