

# Report on Distribution

## **New Datasets (Truck, van, bus, construction)**

In the past week we scraped data for vans and buses on the highways. I used the keywords “highway van traffic” and “highway bus traffic” because they seemed to yield the best results when going through Google Images.

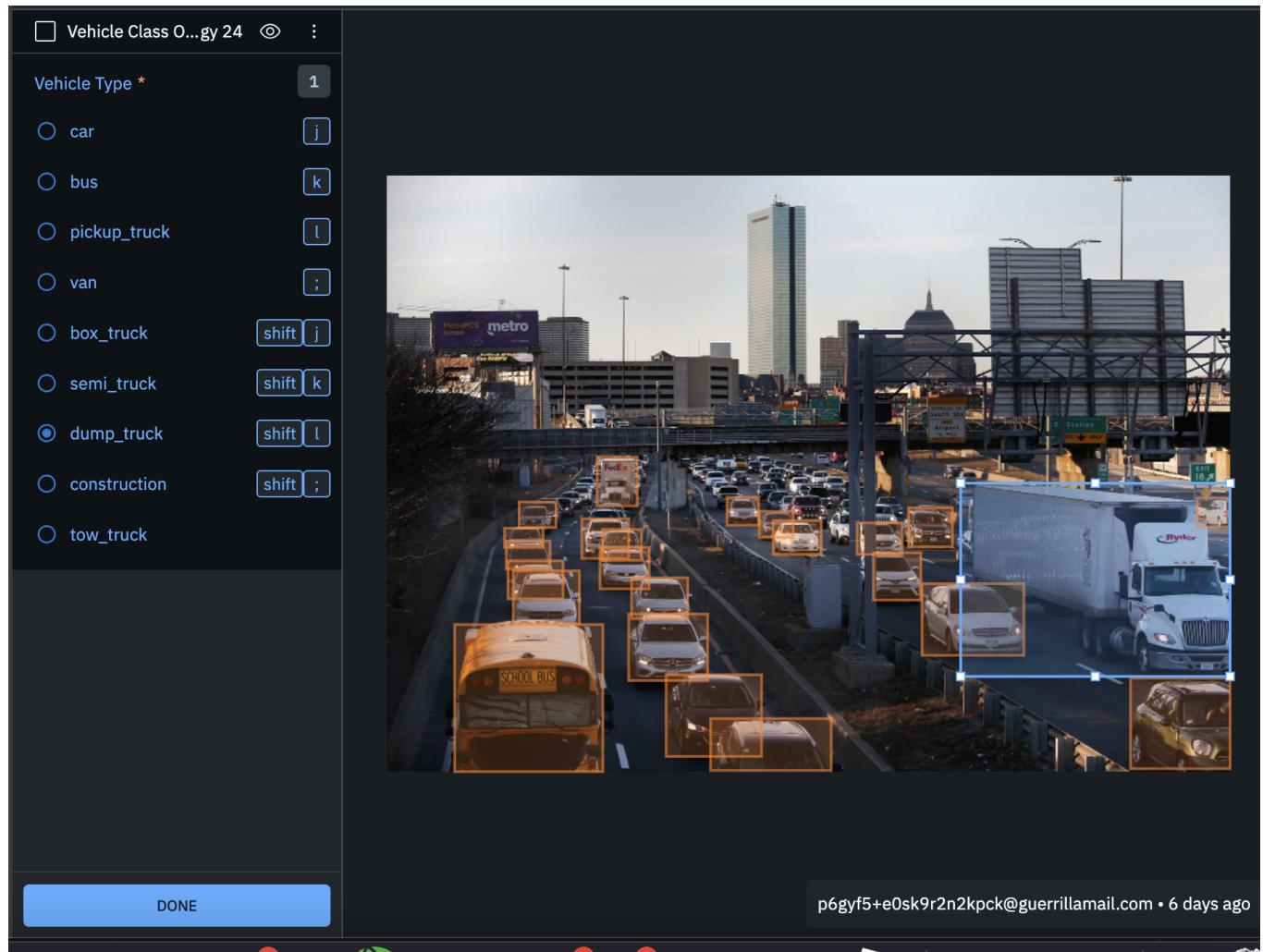
### **Trucks dataset**

It seems like there are many issues with the trucks dataset at this point with the amount of iterations we've done. The primary problems we have are ambiguous classes, which seem to mostly been flagged with issues or rejected.

### **Issues**

A lot of issues are raised in the trucks dataset. When labeling, I found 20 issues in about 200 images.

In the prelabeling for the new datasets, it seems like dump truck is overrepresented. It also doesn't seem like all non-maximal suppression is working, because there are still instances of double boxes in the prelabeling



## Confusion matrix

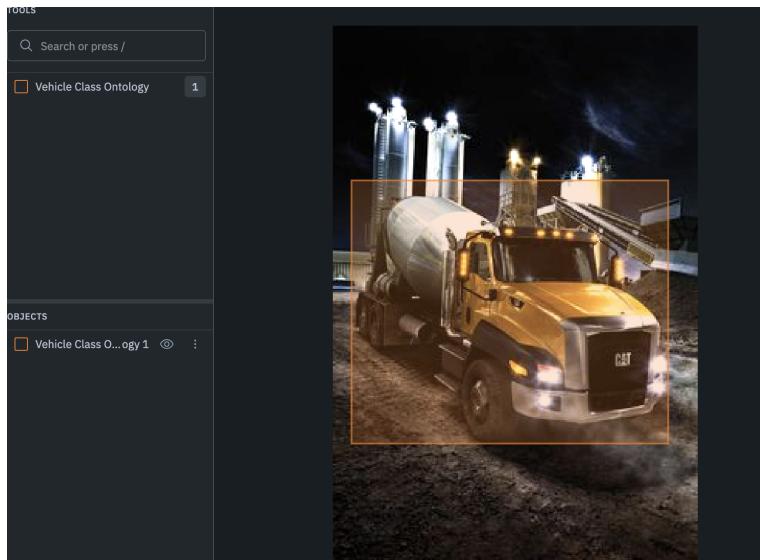
Reviewing Trucks Dataset:

<u>Label</u>	<u>Correct</u>	<u>Missed</u>			
Car	81	2			
Bus	1				
Pickup Truck	15				
Van	9				
Box Truck	66				
Semi Truck	25	1			
Dump Truck	1				

<u>Label</u>	<u>Correct</u>	<u>Missed</u>			
Construction	5	1			
Tow Truck	4				

## Questions:

What should this be labeled as?



Should we include a motorcycle class?

## Van Dataset

Several of the photos are taken from advertisements, and often feature people in, on or around the van. On occasion doors, or the trunk are open. They are also likely to be set against scenic backdrops like mountains or forest, and not often in actual traffic settings:



There was a very evident switch going into the images scraped using highway van traffic. We can see this looking at the below data.

<u>Label</u>	<u>Correct</u>	<u>Missed</u>	<u>Mislabeled</u>	<u>Extra Label</u>
Car	6		1	
Bus	2			
Pickup Truck				
Van	68		1	

<u>Label</u>	<u>Correct</u>	<u>Missed</u>	<u>Mislabeled</u>	<u>Extra Label</u>
Box Truck				
Semi Truck	1			
Dump Truck				1
Construction	1			
Tow Truck				

<u>Label</u>	<u>Correct</u>	<u>Missed</u>	<u>Mislabeled</u>	<u>Extra Label</u>
Car	179		5	5
Bus	2			
Pickup Truck	10		1	
Van	17		2	
Box Truck	8			
Semi Truck	6			
Dump Truck	2			5
Construction	4			
Tow Truck				

There are still a few accident images which are challenging to label, and especially in the case of the larger vehicles, they're often in an unnatural shape post accident which could introduce noise into the model