Paper ###-2020

Lyft Jacking: Analysis of Lyft's Open Source Autonomous Driving Data

John Tomaselli, Chidananda s.a. Nayak, and Jorge Vargas Amezcua,

University of North Carolina Wilmington

ABSTRACT

Proliferation of autonomous vehicle testing is no longer a dream, but a reality. Dozens of companies actively test platforms on live city streets that navigate many of the same hazards as human drivers (Scale AI Inc. , 2020). Contributing to this ecosystem first entails understanding the basic inputs to driving, one of which is object identification. This paper seeks to lay the groundwork for future research by characterizing general patterns of object recognition observed in a dataset collected on public roads. Using the Lyft Level 5 Autonomous Vehicle dataset, two questions are asked which explore object identification from different perspectives. Question one explores whether different camera calibrations significantly influence object detection. With object annotations varying by as much as 16.3% between different calibration bins, select groups of camera calibrations do appear to be affecting object recognition. Question two seeks to document the precision with which object actions associate with object categories, or identities. Data analysis reveals strict adherence to expected classifications, albeit with minor variance in at least one relatively obscure object action.

INTRODUCTION

Autonomous vehicles already occupy American roads. According to pcmag, well over 1.7 million miles of autonomous driving occurred during 2018 in California alone (Marvin, 2019). Supporting further development of these vehicles by a wider audience, several academic and industry groups recently released datasets containing both images and supporting metadata from select sets of their autonomous vehicle fleets (Scale AI Inc. , 2020). Lyft joined these groups on July 23, 2019 by releasing their Level 5 Autonomous Vehicle dataset (Vincent, 2019). The depth and extent of Lyft's data enables detailed examination of many different aspects of autonomous driving that allow for Lyft's vehicles to safely drive.

Fundamentally, this project's objective is to leverage that depth and extent to identify optimal patterns for object recognition, a critical aspect of safe driving. Two questions achieve this objective by examining distinct, but complimentary, aspects of object recognition. The first question answers whether available camera calibrations affect the ability of the Lyft Level 5 Platform to detect objects. If a subset of camera calibrations in this dataset allows for greater object detection relative to other camera calibrations under similar circumstances, this could indicate a set of optimized calibrations across the wider population of Lyft's Fleet. The second question determines how strongly an object's characteristics influence the object's final classification. Discerning how tightly an object's size or actions match the object's final classification may indicate the significance of these parameters in larger datasets. By the conclusion of this paper, enough evidence should be presented to provide basic insight into the factors that enable the Lyft Level 5 Platform to recognize and characterize external objects.

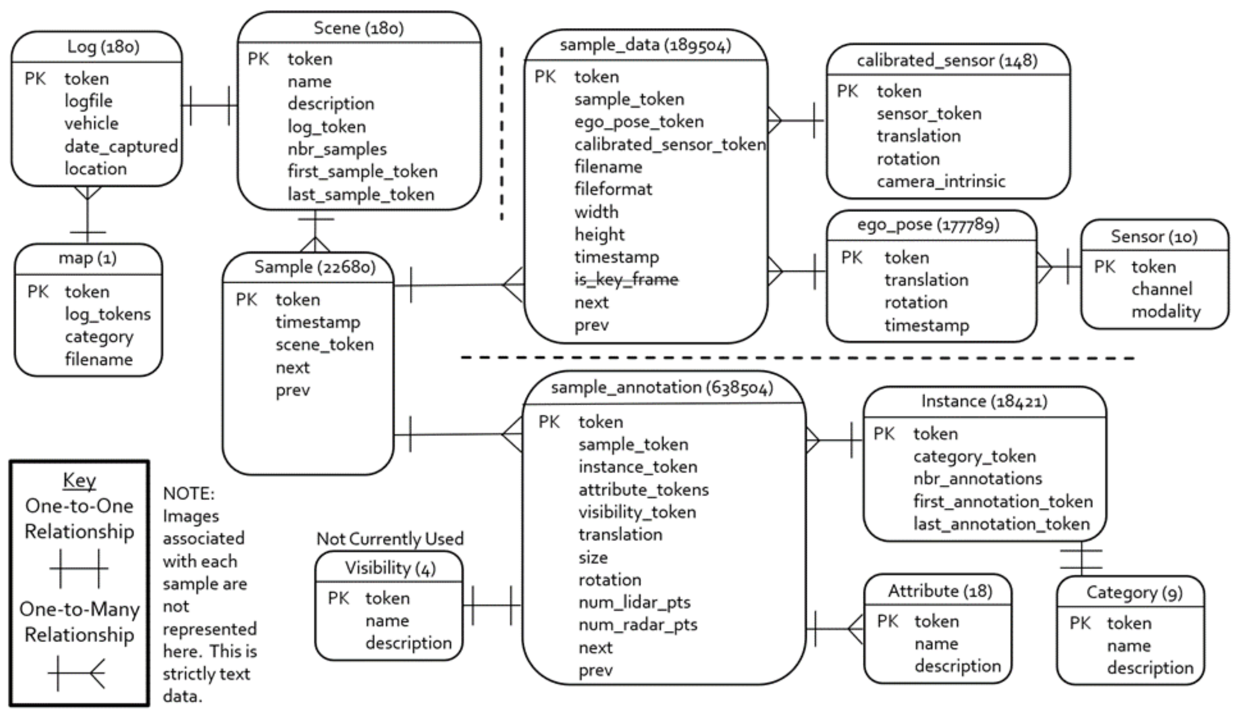
Note that all figures in this paper, unless explicitly included, can be found individually in the hyperlinks associated with each figure number or collectively in Attachment 1 (Lyft

Paper Figures). All SAS code to support the below analysis can be found in the Code folder on this project's Github page.

DATA SET AND DEFINITIONS

Lyft's Level 5 Dataset organizes data according to the nuScenes data format (Lyft Inc. , 2019). 13 individual schema in JSON format allow access to different types of metadata. The entry relationship diagram shown in figure 1 below displays how these schemas logically divide data to allow both one-to-one and many-to-one relationships. Forming these relationships permits relatively flexible analysis of patterns across data from different short videos of vehicles driving called scenes.

Figure 1: Lyft Level 5 Entry Relationship Diagram (ERD)



Scenes form the basis for analysis. In the Lyft Level 5 Dataset, these short driving videos, showing approximately 25 seconds of a Lyft vehicle's journey, consist of a set of snapshots at specific timestamps known as samples. From each sample, the Lyft Level 5 Platform catalogues recognized objects using the *sample_annotation* and associated schema. Conversely, the components of the Lyft Level 5 Platform used to collect a sample, such as the vehicle's location and the sensor being used, are aggregated by the *sample_data* schema. Figure 2 goes into more detail concerning the definition of each schema. Script 1 contains a Python script that manipulates the Lyft Software Development Kit to produce videos of different scenes for a given sensor. Examples of these videos are in the *CAM_FRONT* and *LIDAR* folders of this project's Github page. Reviewing the above videos can help you better understand and visualize the parameters and patterns discussed below.

QUESTION 1: SENSOR CALIBRATION INFLUENCE

Examining how different camera calibrations affect object detection first requires exploring response and predictor variables. The only viable response variable for this project proves to be the *nbr_annotations* variable in the *Instance* schema. This variable fluctuates based on the total time an object is detected by the Lyft Level 5 Platform over the course of a scene. Predictor variables for camera calibrations are aggregated in the *camera_intrinsic* variable of the *calibrated_sensor* schema. As shown in figure 3, this 3x3 matrix contains the *focal length* and *optical center* variables, the values of which sufficiently vary to allow analysis (MathWorks Inc. , n.d.). Figure 4 displays how the optical center of an image corresponds to a camera's intrinsic focus and is broken up into *cx* and *cy* variables. Figure 4 further illustrates that focal length is the distance between a camera's lens and the sensor used to capture an image.

Controlling for unintentional variations with the three predictor variables, *focal_length*, *cx*, and *cy*, is challenging. Figure 5 shows how vehicle use fluctuated considerably over the 30 days for which data is available. When aggregated, the locations in which the *nbr_annotations* variable were captured, as displayed in figure 6, appears to match the complete track used by Lyft. However, examination of the *nbr_annotations* associated with each vehicle in figure 7 reveals that vehicles with 5 or more recorded days of driving cover a much larger portion of the track than vehicles with fewer than 5 days of driving. As a final consideration, each Lyft Platform contains 7 camera settings with 6 camera settings providing a 360 degree perspective and the front-facing camera possessing an extra zoomed mode. These location, date, and orientation variances must be considered when determining the optimal camera calibrations.

Initial examination of the data shows why controls are necessary. Figure 8 contains multiple slides charting our journey through the data. Our first attempts to find an optimal set of variables to produce the maximum *nbr_annotations* returned low adjusted R-squared values, indicating a poor fit to the data. However, high significance is associated with each identified variable as demonstrated by p values below 0.05. Subsequent HPBIN results show that two different clusters of *focal length*, *cx*, and *cy* values exist due to the *CAM_FRONT_ZOOMED* camera setting. Removing the *CAM_FRONT_ZOOMED* setting still produces a low adjusted R-squared value, but shows significantly different numbers of responses for distinct ranges of focal length. Figure 9 further develops this observation with the average *nbr_annotations* for each *focal_length* range appearing to vary by as much as 14.86% of the lowest observation. Figure 10 and figure 11 respectively reinforce this observation with the cy and cx variables observing variances of 15.39% and 16.30%, respectively.

Thankfully, even with relatively poor fitting data, reasons exist to lend confidence to variables with higher *nbr_annotations*. Controlling for the date, vehicle, and location using the REPEATED statement in PROC GENMOD produces the results observed in slide 1 of figure 12. These results attribute significance to the *focal_length, cx,* and *cy* variables due to the observed low p values. Examining the two-factor interactions amongst the predictor variables using PROC LOGISTIC and PROC GLMSELECT, found in slides 2 and 3, reinforces the above evidence. Logistical analysis reveals generally synergistic effects between the *focal_length* and *cx* variables as well as between the *cx* and *cy* variables. Moreover, slide 3 shows how select bins of *focal_length, cx,* and *cy* variables continually produce higher t values, indicating greater overall responses. These observations all appear to lend credibility to select sets of *focal_length*, *cx*, and possibly *cy* values producing higher average values of *nbr_annotations* and, therefore, optimizing object detection.

QUESTION 2: OBJECT CLASSIFICATION

Moving on, the second question asks how strongly an object's characteristics influence the object's final classification. Said otherwise, once an object is recognized, what available variables most profoundly influence how the object is identified as a car, truck, pedestrian, or some other entity. The response variable for this endeavor is the *name* variable in the *Category* schema. Each object, known as an instance, forms a one-to-one relationship with 1 of 9 categories. 7 of those categories are vehicles, being recognized as a *bicycle*, *bus*, *car*, *emergency_vehicle*, *motorcycle*, *truck*, or *other_vehicle*. *Animals* and *pedestrians* complete the list as not being vehicles. Predictor variables likewise fall into two types. The *size* variable in the *sample_annotation* schema contains the *height*, *width*, and *depth* of an object. Conversely, *attributes*, which describe an object's actions, form one-to-one or many-to-one relationships with objects, depending if an object is recognized as performing one action or many actions at the same timestamp.

First examining the *size* variable, we see clear correlations between an object's classification and its dimensions. Figure 13 shows the results of a proportional odds assessment attempting to determine whether the dimensions contained in the *size* variable contribute to whether an object is classified as a vehicle or not a vehicle. With low p values associated with each dimension and pairing of dimensions, the answer is yes. Greater associations occur with an object's *height* and *width* while *depth*, with a weaker p value and lower estimate, might function as a derived value. No matter the dimension, each value appears to have a positive individual influence on the object's final classification with pairs of values forming weaker antagonistic effects.

Similarly stark results are seen when comparing an object's classification and associated attributes. Slide 1 of figure 14 shows a heat map demonstrating that 5 of the 6 attributes associated with *animals* and *pedestrians* are exclusively found with those two categories. Running a PROC LOGISTIC examination confirms this observation with slide 3 showing highly significant p values for 2 attributes associated with *animals* and *pedestrians* while other categories show no association. Slide 4 of figure 14 shows the only known overlap of attribute between vehicles and non-vehicles. Both sets of objects appear to be able to lose control.

CONCLUSION

This examination of the Lyft Level 5 Dataset found patterns that strongly correlate with object recognition and classification. Select focal lengths and optical centers appear to produce significant differences in the number of object annotations recognized. Both individually and, when combined, synergistically, select ranges of these variables contribute to the identification of greater numbers of objects. In a similar manner, the dimensions and actions of an object strongly correlate with how an object is classified. Taken together, these predictors, forming a small fraction of the total number of variables in the nuScenes format, form the core of how objects are fundamentally recognized and treated.

Future research on this dataset might proceed with aspects of objects that were observed, but not fully scrutinized. Figure 7, when separated into distinct days and times, might be a good starting point for analyzing traffic along the provided route. Moreover, we largely ignored the CAM_FRONT_ZOOMED camera setting after slide 6 from figure 9 in order to focus on the other 6 camera settings. Examination of this setting and comparison to data derived from the other 6 settings may produce interesting results. Finally, investigating the inflection points either directly after an object is detected or before an object is no longer detected proved beyond the scope of this project. With blank *num_lidar_pts* and *num_radar_pts* variables, insufficient data was available to give confident answers as to the limits of when objects are recognized or fail to be recognized. If future versions of the Lyft Level 5 Dataset include meaningful values with these variables, investigation of these inflection points may be warranted.

REFERENCES

Lyft Inc. . (2019, July 23). *Lyft Level 5 AV dataset and nuScenes devkit tutorial*. Retrieved from Github: https://github.com/lyft/nuscenes-devkit/blob/master/notebooks/tutorial_lyft.ipynb

Lyft Level 5. (2019). *Lyft Level 5 Dataset*. Retrieved from Lyft: https://level5.lyft.com/dataset/

Marvin, R. (2019, March 1). *Which Self-Driving Cars Put in the Most Fully Autonomous Miles?* Retrieved from PC Magazine: https://www.pcmag.com/article/366797/which-self-driving-cars-put-in-the-most-fully-autonomous-mil

MathWorks Inc. . (n.d.). *What Is Camera Calibration?* Retrieved from MathWorks: https://www.mathworks.com/help/vision/ug/camera-calibration.html

Scale AI Inc. . (2020, January 6). *Open Datasets*. Retrieved from Scale: https://scale.com/open-datasets

Vincent, L. (2019, July 23). *Unlocking Access to Self-Driving Research: The Lyft Level 5 Dataset and Competition*. Retrieved from Medium: https://medium.com/lyftlevel5/unlocking-access-to-self-driving-research-the-lyft-level-5-dataset-and-competition-d487c27b1b6c

ACKNOWLEDGMENTS

We would like to thank our mentor, Dr. James Blum, for his guidance in moving through this project.

We also thank Lyft Incorporated (Inc.) and the Lyft Level 5 Team for releasing the Level 5 dataset as well as the corresponding tutorial and SDK. The Lyft Level 5 Autonomous Vehicle dataset and associated materials remain the property of Lyft, Inc. and are licensed under version 4.0 of the Creative Commons Attribution-NonCommercial-ShareAlike license (CC-BY-NC-SA-4.0). The v1.02-train dataset used was not altered before project commencement, but was significantly altered through merging, sorting, and renaming both schema as well as individual variables in preparation for answering our questions.

RECOMMENDED READING

- *Lyft Dataset SDK (https://github.com/lyft/nuscenes-devkit)*
- *Lyft Level 5 AV dataset and nuScenes devkit tutorial (https://github.com/lyft/nuscenes-devkit/blob/master/notebooks/tutorial_lyft.ipynb)*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Tomaselli
john.tom.abc@gmail.com