

Lyft Releases Data Set on Level 5 Autonomous Vehicles and Hosting Competition on Use of Data

In what can be seen as a huge step forward in the arena of autonomous vehicle technology, [Lyft has announced](#) that it will share with the public its level 5 dataset from its autonomous vehicle data. The dataset is described as a “comprehensive, large-scale dataset featuring the raw sensor camera and LiDAR inputs as perceived by a fleet of multiple, high-end, autonomous vehicles in a bounded geographic area.” The Lyft [Level 5 Dataset](#) includes:

- Over 55,000 human-labeled 3D annotated frames;
- Data from seven cameras and up to three LiDARs;
- A drivable surface map; and
- An underlying HD spatial semantic map (including lanes, crosswalks, etc.)

In addition, Lyft is also launching a competition for individuals to train on the algorithms on the dataset, including completing testing on 3D object detection over the semantic maps. Although specific details about the competition have yet to be released, Lyft has indicated there will be \$25,000 in prizes, and they will fly the top researchers to the [NeurIPS Conference](#) in December, as well as allow the winners to interview with Lyft for a position within the company.

So, what does all this really mean for the advancement of autonomous vehicles? One of the biggest issues in the autonomous vehicle arena right now is cooperation among the manufacturers and developers. This is not to say that these entities are not creating consortium groups or teaming up on manufacturing ventures, but this cooperation remains very segmented. The sharing of data within the entire industry is not common place, so it is very exciting to see such an extensive dataset being shared so openly by Lyft. Sharing knowledge of this kind is crucial if autonomous vehicles are to become a reality on the roadways, as these vehicles are going to need to work in cooperation with not only each other but the infrastructure of the areas in which these vehicles operate.

It also seems, on the surface, that this dataset is being shared not only among manufacturers and technology developers, but with the entire public. While this is technically true, as anyone can download the dataset, unless you are capable of reading the data, there is no way for the average person to understand what it is that has been collected or what it really all means with respect to autonomous vehicles. I downloaded the available dataset (there is additional information being released within the coming weeks) and while it is very impressive, it doesn't enhance my knowledge of this technology in anyway.

It's obvious that Lyft is partially using the release of this data, in conjunction with a competition, as a way in which the company can find new talent to add to its autonomous vehicle team, and maybe even acquire some free research out of it by the submissions received in the competition. Though not a criticism of Lyft, this is an observation that this data release is really not to the public at large, but to a select group of individuals who have the capability of reading and interpreting the data.

As the public's understanding of and trust in this technology is also a crucial factor in the acceptance of autonomous vehicles on our roadways, there needs to a way to help relay this type of information to the general public so they can continue to learn, along with the manufacturers and developers, about how the technology works, its safety and improvements being made as issues arise. While this dataset may not be capable of such relatability and dissemination, this is a reminder to all involved that no amount of technology will be worth much, without the public's trust in and use of autonomous vehicles.