Term Paper on Challenges Faced by Autonomous Vehicles

With the advent of machine learning and computer vision, more and more self-driving technologies have emerged. Tech companies such as Google, Uber, Tesla have joined the auto industry alongside big car conglomerates such as Ford, Nissan, and General Motors to develop autonomous cars. While system details vary, most self-driving AIs use a wide array of hardware sensors to create an internal mapping of their surroundings. Complex AI systems using a combination of neural networks and other learning methods use this mapping to instruct a variety of actuators which in turn control basic functioning of the car such as acceleration, braking, steering, lane assist. The initial approach which has been around for decades simply included rule-based systems which controlled driving assists such as cruise control or warning signs if the car was swerving off the roads. But now, with the advent of ML, we can foresee a near-future with completely autonomous cars with complete decision-making capabilities.

FirstComputer vision has an important role to play in the development of autonomous driving cars, now and in the future. The most evident use of computer vision that comes to mind is of object detection for collision avoidance. If the car cannot see, how will it drive. The number one priority for us, when it comes to driving cars in general is to not cause accidents, and this is where object detection and classification comes in. Not only can CV help predict possible collisions, but we can also extend this application to help avoid accidents as well. By using freespace detectors, the car can compute possible maneuvers to change lanes, keep safe distance from other vehicles and when the need presents itself, avoid possible accidents.

The next task after capturing of images is the 3D internal map creation. This can be achieved either through multiple cameras taking images from various angles and triangulating positions of objects; or with the much simpler solution using 3D cameras that use RADAR or LIDAR. Translating bounding boxes found in 2D to 3D will give us precise information about the object's position and orientation.

Lane assistive software already exists in many cars today, but these features usually require human interaction. In a perfectly autonomous system, we can imagine the car making lane changing decisions purely out of efficiency and of its own accord. Choosing to use the carpool lane, using a lane to make a future turn, or even using a low trafficked lane computed by using long range 3D sensors; these are all entities that can be provided by computer vision.

SecondComing back to object detection, the various practices we have studied in class from traditional computer vision techniques such as selective search algorithms to the cutting edge deep neural networks such as SSD all are making strides into making object detection faster, more reliable, and more efficient. Beyond detection, classification is just as important. Various classes of objects hold different weights. Object classes such as vehicles, pedestrians, traffic signals must be predicted with the highest accuracy, whereas trees, fences, graffiti and other background objects have relatively less importance.

As with any machine learning algorithm there is an immense significance on the data used to train the systems. Collection of data may seem easy with the multitude of cameras being used for dashcams, traffic rules enforcement, surveillance videos, etc., but all this data also needs to be processed, normalized, and annotated. One approach is to annotate manually through third party data labeling companies, which requires time and manpower. Alternatively, some companies use auto-labeling technique that involves a combination of neural networks, radar data, and human reviews.

Collected data also needs to be sufficiently diverse including edge cases that may never happen. Training a model using large datasets from one source could mean it performs better in one type of neighborhood than another. There are also studies that have shown that designing pedestrian detection models inadvertently includes algorithmic biases. A study from Georgia Tech concluded that on average light skinned individuals are detected with 5% higher accuracy than dark skinned persons. The solution presented to such problems is to add more images of dark-skinned persons (increase the diversity of the dataset), and to place greater weight on prediction accuracy on those images. Data security is also of growing concern, as we realize the need for protection against adversarial attacks. Training dataset needs to include certain images that we can say will cause system to give unexpected results at test time. Thus, eliminating the threat of such attacks during product release.

There are also many challenges being faced in the task of 3D mapping. Experts still debate the use of 3D cameras over 2D cameras. 2D inferencing using images uses considerable computation power and is not suited for surface texture predictions, whereas 3D LIDAR cameras are generally more expensive. Tesla has been a vocal champion for the pure vision-based approach to autonomous driving. They are of the opinion that vast amounts of diverse data from simple cameras are enough to guarantee success in the 3D inference tasks. Waymo on the other hand, who have tested their cars through millions of miles on the streets of San Francisco swear by their single integrated system of sensors consisting of Radar, Lidar, and 360-degree cameras.

ThirdAs we have studied in class; various new technologies are emerging every day in the field of computer vision that will help us reach Level 5 (Full Driving Automation) self-driving cars. The question is not about if but when we will achieve this goal. Current state of the art systems we studied such as Mask-RCNN have provided good qualitative results but are far from perfect. Newer technologies such as vision transformers and more robust vision models will shape the future of CV and autonomous vehicles.

Beyond these challenges faced by computer vision, there is also a problem faced by all self-driving AI systems in in general: an immense importance on the speed of operations. If there are immediate obstacles in the way and collision is imminent, the AI in the car has to: take multiple images all around the car every second, preprocess the data, detect the obstacle in the way, classify it as a solid object/collision threat, calculate whether there is a possible way to avoid accidents, and then perform evasive maneuvers. As we can see, speed is of the essence. The progress of self-driving cars is mostly held up, not by software algorithms but by hardware shortcomings in today's vehicles. Processing power and sensor limitations are the biggest bottlenecks in the development pipeline. For all their technical prowess Tesla and Cadillac are both dependent on sensor technology, causing their cars to lose functionality in adverse conditions such as bad weather, heavy traffic, or sensor failure. This has known to cause accidents and system breakdowns and is one of the reasons we are still in the Level 2 (Partial Automation) of autonomous vehicles.

As of now, self-driving cars are an emerging technology. Both established auto companies and startups alike are trying to get in on the market, but it is hard to gain a foothold as the general public is skeptical about the use cases and safety of these systems. Companies have to work extra hard, not just on the capabilities of the systems, but also to improve the public perception of these systems.

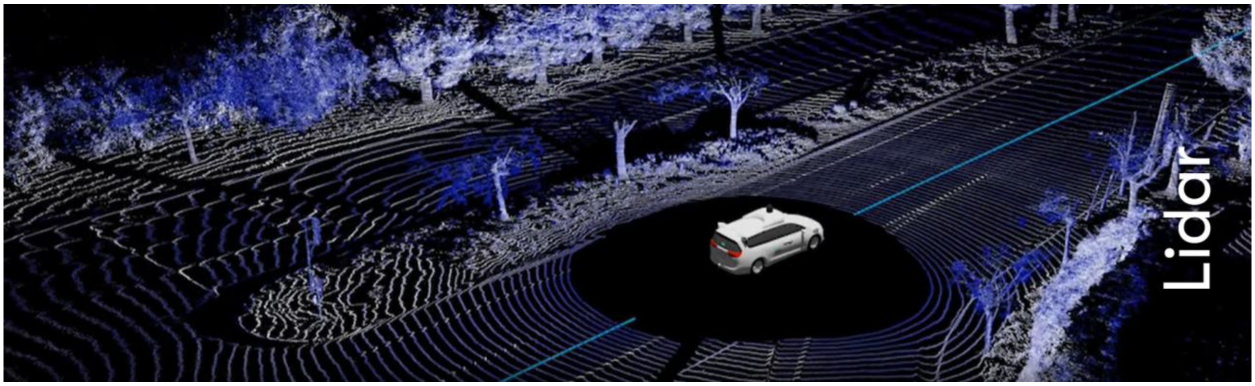

Figure 1: Lidar Scan Example from Waymo.com

https://bdtechtalks.com/2021/06/28/tesla-computer-vision-autonomous-driving/

https://www.vox.com/future-perfect/2019/3/5/18251924/self-driving-car-racial-bias-study-autonomous-vehicle-dark-skin

https://arxiv.org/pdf/1902.11097.pdf

https://www.blickfeld.com/blog/levels-of-autonomous-driving/

https://www.cnn.com/style/article/ces-toyota-big-smart-city/index.html