NVIDIA: Building a compute and data platform for self-driving cars

Summary

To build safe and reliable AI models for Autonomous Vehicles (AVs), enormous amounts of compute power and training data are needed, along with the skills, resources and expertise required at that scale. Under such conditions, we are likely to see an emergence of larger platforms that can pool data from multiple participants, aggregate sufficient demand to justify the large investments required, and ultimately enable a new business model where AV software can be offered as a service to carmakers and fleet operators.

Opportunity: Solving complex problems to drive mass adoption of AV

Driver support systems are becoming more widespread and can now perform parking, emergency braking, lane-changing and other functions. Once vehicles are fully self-driving, the applications will be numerous, for example freight (e.g., trucks), mass transit (e.g., buses) and on-demand transportation (e.g., robo-taxis). With the market for Autonomous Vehicles (AVs) projected to grow to over $2,000 billion in 2030 (up from $54 billion in 2019, representing a compound annual growth rate of 39%)[1], this represents an enormous opportunity.

Challenge: Enormous data and compute requirements

The data requirement for fully autonomous vehicles is enormous because of the range of planning and control tasks that need to be performed (e.g., finding pedestrians, detecting road markings and traffic lights). These functions must be robust to varying environmental conditions (e.g., weather, locale) and must be able to handle transient, rare events (e.g., being cut off unexpectedly). Current systems are improving on the number and effectiveness of these functions, but there is a high bar to demonstrate their reliability in terms of reduced fatalities and injuries; the RAND corporation has estimated that to match a human-level error rate could take 11 billion miles of testing, equivalent to 100 vehicles being test-driven continuously for over 500 years[2]. NVIDIA themselves estimate that good performance on certain AV tasks requires training examples in the order of 1,000,000 scenes.

Considering that each of these scenes will entail data from numerous sensors, the compute challenge for autonomous vehicles is enormous. To create the perception models needed for a full AV stack, NVIDIA estimates that a productive development team could require in the order of 5000 dedicated GPUs (Graphical Processing Units)[3]. A single model can take 3-6 days to run on 32 GPUs, and there can be 25-50 Deep Learning experiments for each task. Individual car companies typically don't have the resources in terms of skills, experience, hardware, and data to develop these systems on their own.

Solution: Shared data and compute platform across multiple customers

NVIDIA is addressing these challenges through the following:

- **Extending a common data platform across multiple customers**: Pooling data between several companies[4] will increase the data available for training and enable greater model performance, particularly with edge cases. The quality of the data will be enforced through a reference architecture[5] that specifies the standards for sensor specifications and placement.

- **Simulation for training and testing**: Hundreds of millions of driving scenarios can be simulated to supplement real-world data and help bootstrap models for silent on-street testing and iteration e.g., running AI in the vehicles to compare what it would have done relative to the driver's actual behavior.

- **Common processing of visual tasks:** NVIDIA was able to minimize the compute required by jointly training multiple tasks on a single ResNet-based model architecture. Once the full model is trained, the heads (later layers) of the model can be optimized for each given task, without the need to re-train the trunk (earlier layers) of the model. Noting that the compute was not much greater than that required for a single task alone, suggests that there is a lot of common processing possible, which makes intuitive sense for the domain of computer vision.

Outcomes: New business model that enables competition at different level of the stack

Centralizing the management of data in this way enables new possibilities for AV technology. Depending on their needs and on their existing capabilities, participating carmakers can either lease AV hardware to train their own models based on a larger dataset or use pre-trained AV models from NVIDIA. In either case, instead of making significant capital investments in hardware and development capability, carmakers can book the AV technology as operating expenses, and benefit from improvements as the hardware and software improves.

It also represents the start of a new market dynamic. On one side are vertically integrated carmakers (e.g., Tesla), that can co-design their software and hardware for more seamless experiences. On the other are increasingly modularized carmakers who compete on the quality of their hardware and buy their software from centralized players like NVIDIA (which greatly reduces cost of entry to the AV market and is likely to stimulate greater competition as a result). The success of either of these two paradigms depends on how important the quality of AV software is in terms of the overall experience.

[1] https://www.marketwatch.com/press-release/autonomous-vehicle-market-share-2021-global-trend-segmentation-size-business-growth-top-key-players-analysis-industry-opportunities-and-forecast-to-2030-2021-07-21-5197440

[2] https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf

[3] GPUs are arranged into purpose-built DL systems (e.g., the NVIDIA DGX, which comprises 8 GPUs per server)

[4] NVIDIA recently announced a partnership with Mercedes-Benz, and has plans to extend this to other carmakers

[5] Hyperion