Data Challenges for Autonomous Vehicles

Coordinators	Nr. Students
Prof. dr. ing. Adina Magda Florea **Andrei Nica (andreic.nica@gmail.com)**	**2**

1. Introduction/Context

*This research topic is part of the **NEMODRIVE** research group. Read the group profile first at: url.*

At each progressive level of vehicle autonomy the challenges become more difficult. But, among the most *mountainous* of challenges is data. Without efficient data management, the sheer resources the process will consume can dramatically slow innovation. There are rough estimates that the data generated by a single car would be around 1 TB+ /h [4], considering Nemodrive's sensor setup. It is of great importance to explore the four data considerations in the autonomous vehicle framework: Acquisition, Storage, Management and Labelling. We will be implementing a state of the art solution for data management using the Apollo framework.

Neural networks dominate the modern machine learning landscape, but their training and success still suffer from sensitivity to empirical choices of hyperparameters such as model architecture, loss function, and optimisation algorithm ([1], [2]). Given the importance that deep learning plays in achieving a fully autonomous vehicle we will invest research effort in implementing and using state of the art methods for efficient search of the best solution for our problem.

Managing experimental results in machine learning can be a daunting task. Researchers and practitioners often try a variety of algorithms, hyper-parameters, and pre-processing techniques, each resulting in different outcomes. Tracking and analyzing each of these outcomes is a burden further amplified when dealing with multiple collaborators and several computer nodes.

Consequently, the success of Nemodrive's research objectives depends on the development and deployment of a distributed Machine Learning Experiment Management Framework (e.g. [3]).

2. Objective

- Develop and deploy the full *data pipeline* system for Nemodrive`s self-driving car (from acquisition to training).
- Manage the acquisition of the first UPB self-driving dataset.
- Explore and deploy a state of the art machine learning hyperparameter optimization algorithm.
- Develop a distributed machine learning experiment management framework for the AI-MAS lab and validate the infrastructure.
- Publicly disseminate and open-source the framework solution and the UPB self-driving dataset.

3. Required and Learned Skills

- Requirements
  - Good knowledge of Python and/or C++
  - Prior Machine Learning knowledge is preferred
  - Fast learner, proactive mindset
  - Comfortable working in a team
  - Love for software architecture design is a plus :-)

- Learned skills
  - Working with a complex autonomous driving framework (Apollo)
  - Practical experience with many of the technologies sustaining autonomous-driving
  - Experience working with ML algorithms and training of deep neural networks
  - Experience working with frameworks such as: ROS, Pytorch, Scikit-learn, OpenCV, Pandas
  - Experience of contributing to the common goal of a large research team
  - Support and guidance for writing academic research papers

4. References

[1] Jaderberg, Max, et al. "Population based training of neural networks." arXiv preprint arXiv:1711.09846 (2017).

[2] Li, Lisha, et al. "Hyperband: A novel bandit-based approach to hyperparameter optimization." The Journal of Machine Learning Research 18.1 (2017): 6765-6816.

[3] Studio.ML https://github.com/studioml/studio

[4] “Training AI for Self-Driving Vehicles: the Challenge of Scale” https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/