Student Thesis

Master Thesis

Cloud-Assisted Neural Network Inference Acceleration

Department

Vehicle Intelligence & Automated Driving

Contact

Lennart Reiher

📞 +49 241 80 25614

✉ lennart.reiher@ ika.rwth-aachen.de

Background and Goal of the Thesis

Cloud- and edge-based computing in the context of automated driving promises to support, enhance, as well as completely replace intelligent driving functions of automated vehicles. A prerequisite for many cloud-connected driving use cases is the availability of highly optimized low runtime algorithms in order to counteract the introduced cloud communication latency.

Many driving functions today, especially perception algorithms, are deep learning-based. Here, multiple levers exist to speed-up runtime: model complexity, inference optimization, deployment strategy, etc.

The goal of this thesis is to assess and optimize runtime of neural network-based inference algorithms on vehicle and cloud computers. Additionally, further methods for reducing communication latencies as well as combining local and cloud algorithms should be explored.

You will have access to ika's extensive hardware and software infrastructure, including research vehicles, high-performance computers, simulation environments, and deep learning-frameworks. Additionally, you can expect close supervision and collaboration with other highly motivated researchers.

Working Points

- Familiarization with ika's existing software components for deep learning (training and deployment) and cloud communication

- Literature research on optimized deployment of deep learning-models as well as cloud communication strategies

- Development of techniques to speed-up runtime of neural network-based algorithms on vehicle and cloud computers

- Conceptualization and implementation of methods to reduce communication latencies and/or combine local and cloud algorithms

- Evaluation of the developed optimizations and concepts, potentially using ika's research vehicle and high-performance cloud servers

Language

German or English

Starting Date

As soon as possible

Requirements

- Reliability, commitment, and enjoyment of working independently

- Must-have experience
  - C++
  - Git
  - Unix-Shell

- Nice-to-have experience
  - Deep learning with TensorFlow
  - ROS
  - Python

Application

Informal application including a short CV and an academic transcript record