

Semesterarbeit/Masterarbeit

Spatiotemporal Reasoning of Autonomous Driving Disengagements using Vision-Language Models

With the continuous advancement of autonomous driving (AD), modern AD stacks are steadily approaching SAE Level 4. However, even advanced systems frequently encounter complex long-tail corner cases (CCs) resulting in critical system disengagements. The root causes of these disengagements are rarely limited to simple perception failures of specific dynamic objects. Instead, they often stem from highly unpredictable, unbounded scenarios—such as ambiguous interactions between multiple traffic agents, complex or unmapped road topologies, irregular traffic rules, and severe environmental anomalies.

Traditional AD modules, which rely on closed-set perception and predefined planning rules, lack the high-level semantic understanding required to interpret these diverse and unbounded situations. This necessitates a paradigm shift towards open-vocabulary reasoning. By leveraging large Vision-Language Models (VLMs), we can holistically analyze the exact causes of disengagements. Specifically, by processing sequences (e.g., continuous multi-view camera feeds) leading up to a disengagement, VLMs can extract rich spatiotemporal context and provide human-interpretable, open-vocabulary reasoning about why the AD system failed, far beyond identifying mere object-level errors.

Work packages:

This work will develop a spatiotemporal reasoning framework that deduces the root causes of AD disengagements using Vision-Language Models. You will investigate how effectively VLMs can interpret complex driving contexts and identify open-vocabulary hazards from a sequence of multi-sensor data. The project encompasses the following tasks:

- Literature review: Existing approaches regarding safety-critical reasoning.
- Implementation:
 - Develop a framework: to integrate sequential disengagement logs (multiview camera videos, and vehicle states if needed) with existing VLMs.
 - Fine-tune & prompt-engineer: to perform trustworthy safety-critical reasoning, leading to identification of the cause of disengagement.
 - [Optional] To Incorporate point cloud for enabling long-range reasoning.
- Evaluation: To validate the reasoning capabilities and accuracy on real-world corner case datasets

Requirements:

- Very good programming skills in Python and PyTorch.
- Knowledge of Computer Vision and Deep Learning (Must), Foundation models such as VLMs (Desired).
- High personal motivation and independent working style.
- Very good language proficiency in English.

Recommended literature:

- [WaymoQA: A Multi-View Visual Question Answering Dataset for Safety-Critical Reasoning in Autonomous Driving](#)
- [CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving](#)
- [NuRisk: A Visual Question Answering Dataset for Agent-Level Risk Assessment in Autonomous Driving](#)

Possibility for publication in case of excellent work.

The thesis should clearly document the individual work steps. The candidate undertakes to complete the term paper independently and to indicate all scientific aids used. The submitted work remains the property of the chair as an examination document.

Prof. Dr.-Ing. M. Lienkamp

Betreuer: Hojun Lim, M. Sc.

Ausgabe: _____

Abgabe: _____