WASP Project Course 2025

Performance Evaluation of using Multi-Modal LLMs for Out-of-Distribution Detection

Background

Modern advanced driver assistance systems (ADAS) like advanced emergency braking systems (AEBS) and automated lane keeping systems (ALKS) aim to reduce the number of fatally injured traffic participants. Besides directly "driving" the vehicle, AI/ML-enabled systems play a key role in an automated driving system (ADS) to perceive their surroundings. However, ADS and ADAS inevitably face complex real-world situations that can lead to fatal traffic accidents, as has already been reported in both the newspapers and the academic literature. To better cope with these dangerous situations, Out-of-Distribution (OOD) detection approaches can help to evaluate the confidence of AI/ML-enabled systems and to guide the overall ML development process [1].

At Volvo Cars, we work to build a OOD-based run-time monitoring system to provide indications about the reliability of the AI model's predictions (Figure 1). However, systematic benchmarking concepts for OOD covering real-world settings with multi-modal, time-series data are scarce. This hinders the curation of OOD datasets that are free from in-distribution samples.

In this project, we invite WASP PhD students to systematically explore and analyze the performance of Large Language Models (LLMs) and other generative AI (GenAI) models to act as OOD detectors on the example of automotive datasets. For this, we plan to use the Zenseact Open Dataset (ZOD) and state-of-the-art multi-modal LLMs like MiniCPM-V, Gemma3, and LLaVA to setup an experimental OOD pipeline that we systematically feed with known in-distribution samples and out-of-distribution samples to investigate modern GenAI's performance to serve as OOD guard.

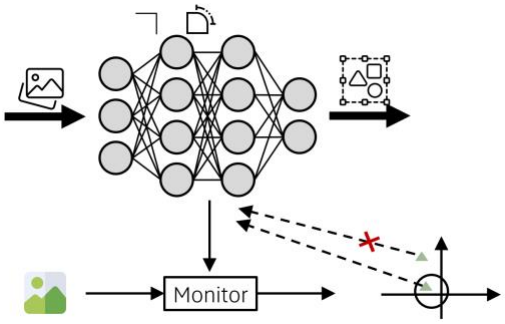

Figure 1: Monitor real world inputs with run-time safety monitors.

Constraints: Gothenburg

Participants

Industrial partner: Volvo Cars

Industrial supervisor: Qinglei Ji, qinglei.ji@volvocars.com

Academic supervisors:

- Christian Berger, christian.berger@chalmers.se, Chalmers University of Technology and University of Gothenburg, Sweden
- Beatriz Cabrero-Daniel, beacab@chalmers.se, Chalmers University of Technology and University of Gothenburg, Sweden

Challenges to investigate

- How to acquire OOD confidence score on AI models' inferences using Multi-Modal LLMs?
- How does the Multi-Modal LLMs approach compare with the currently used distance-based feature comparison methods?

Resources

Automotive datasets (Zenseact Open Dataset), complementary datasets like Waymo Open Dataset, and nuScenes dataset

Deliverables

- Proposal for a methodology how LLMs/VLMs/FMs can be used for OOD.
- Validation of the methodology in an industrially relevant context.
- Illustration video for project course.
- Manuscript for submission to relevant conference/journal.

References

[1] *Yang, Jingkang, et al. "Generalized out-of-distribution detection: A survey." International Journal of Computer Vision (2024): 1-28.*

Keywords

ML-Ops, LLMs, VLMs, Foundation Models, Python