

Visual Language Model in Autonomous Driving

Background

As part of the research project Providentia++ funded by the federal ministry of transport and digital infrastructure under the initiative “Digital Test Beds for Autonomous Driving”, a group has come up to conduct research in the field of intelligent transportation systems, and to come up with solutions and recommendations for improving traffic safety, efficiency and comfort. Within the framework of this project, an existing infrastructure for real-time localization of traffic participants on the Highway A9 will be extended from the highway into an adjacent urban environment. Project video is available on: <https://youtu.be/4oCnQlGFuc4>.

Description

Visual language model (VLM) has shown notably impressive capabilities in many challenging tasks and is gaining more and more attention in recent years. It has been applied to various domains within computer vision, giving rise to many interesting new topics and new tasks. However, how to utilize VLM to assist autonomous driving, including improving perception performance and helping traffic scene understanding, remains an unexplored territory. Multiple research topics are available under this scope, including but not limited to:

- **Traffic Environment 3D Perception Enhancement leveraging VLM**
- **Traffic Scene Understanding with VLM**
- **Image/Video Retrieval with VLM**

Your Tasks

- Learn perception algorithms, popular visual language models and relevant machine learning topics via literature research
- Develop and implement algorithms to solve research tasks
- Algorithms evaluation and experiments summary

Requirements

- Strong interest in visual language model, computer vision, and autonomous driving
- Experience with deep learning frameworks and pipelines (**PyTorch**, **Docker**, etc.)
- Good knowledge of at least one programming language (ideally **Python**)

Supervisor: Prof. Alois Knoll
Contact: M.Sc. Xingcheng Zhou
xingcheng.zhou@tum.de

Technical University of Munich
Faculty of Informatics, Chair of Robotics,
Artificial Intelligence and Real-time Systems