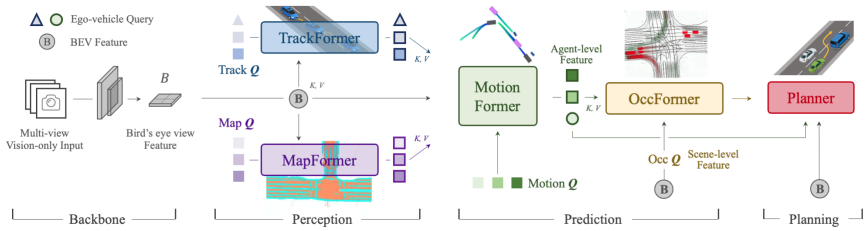
Karlsruhe Institute of Technology

Master Thesis



An end-to-end autonomous driving framework[1]

Explainable End-to-End Autonomous Driving

Challenge:

While state-of-the-art end-to-end (E2E) autonomous driving models have demonstrated significant progress, incorporating intermediate outputs such as object detection, tracking, and motion prediction, the explainability of the planning process remains a critical challenge[2][3].

A notable example of this challenge occurs when a model successfully detects an obstacle ahead, yet generates a trajectory that would result in a collision. The fundamental issue lies in how the model processes its internal representations[4]. The features utilized for motion planning exist in a high-dimensional latent space learned by the neural network. These abstract features lack direct interpretability for human understanding.

This research challenge manifests in several key aspects:

+ Limited understanding of the motion planning process, even with access to intermediate perception results.
+ Challenges in interacting with latent planning features to guide model behavior or verify reasoning in critical scenarios.
+ Difficulties in diagnosing planning-related failure modes.

Thesis Objectives:

+ Analysis of existing E2E driving models and relevant explainable AI methods.
+ Training and evaluation of an existing state-of-the-art E2E driving model.
+ Design and implementation of techniques for generating explanations of the model's planning module.
+ Assessment of the explanation quality
+ (optional) Improve the driving performance by incorporating the explanations

I am happy to answer any questions you might have on the topic, just write me an E-mail or apply directly.

[1] Planning-oriented Autonomous Driving (https://arxiv.org/abs/2212.10156)
[2] Divide and Merge: Motion and Semantic Learning in End-to-End Autonomous Driving (https://arxiv.org/abs/2502.07631)
[3] DriveTransformer: Unified Transformer for Scalable End-to-End Autonomous Driving (https://arxiv.org/abs/2503.07656)
[4] Words in Motion: Extracting Interpretable Control Vectors for Motion Transformers (https://arxiv.org/abs/2406.11624)

Institute of Measurement and Control Systems (MRT)
Prof. Dr.-Ing. Christoph Stiller

Advisor:
M.Sc. Yinzhe Shen
Dr.-Ing. Ömer Şahin Taş

Programming language(s):
Python

System, Framework(s):
Linux, PyTorch, MMCV

Profile:
- Solid theoretical foundation in deep learning
- Practical experience in implementing deep learning models, ideally in motion prediction and planning
- Strong interest in autonomous driving and AI explainability
- Highly motivated and independent working style
- English in writing and speaking

What we offer:
- Supervision by experienced researchers in deep learning and autonomous driving
- Adequate GPU resources
- Opportunity to co-author a publication **targeting a top conference** (ECCV, NeurIPS, ICLR, etc.)

Apply:

Send me an E-mail with your
- CV
- transcript of grades
- (if available) code samples of your previous programming projects with github links or a zip file

Contact:

Yinzhe Shen

Room: 033
Phone: +49 721 608-42343
Email: yinzhe.shen@kit.edu