Analysis and Evaluation of VLMs in multimodal scene understanding

TECHNISCHE UNIVERSITÄT DARMSTADT

MaTh

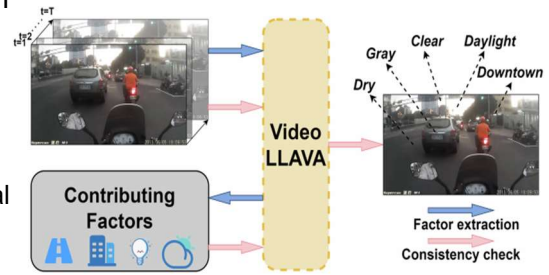
The growing complexity of automated driving demands scalable pipelines to extract relevant traffic scenes from real-world data automatically. This work, in collaboration with Porsche Engineering, explores how state-of-the-art Vision-Language Models can identify and retrieve predefined driving scenarios in large-scale datasets.

Task Details

- State-of-the-art analysis for scene and scenario extraction based on vision-language and video-language models.
- Implementation of processing strategies that balance detection performance, computational load and model size for large-scale datasets.
- Conducting ablation studies across different model families (e.g., LLaMA-Vision, Qwen-VL, Florence 2) and network configurations.

Profile

- Hands-on experience with PyTorch, proficient in Python development on Linux systems
- Practical knowledge or experience with VLMs is a plus



Guan, Y., Liao, H., Wang, C. *et al.* World model-based end-to-end scene generation for accident anticipation in autonomous driving.

NOTICE: All projects and theses at FZD can be done in English or German, as prefered.
ANMERKUNG: Alle Projekte und Arbeiten bei FZD können wahlweise in Englisch oder Deutsch durchgeführt werden.

Anton Kuznietsov| anton.kuznietsov@tu-darmstadt.de| Room 405 | 06151 – 16 24204

AI safety

KI-Absicherung

TECHNISCHE UNIVERSITÄT DARMSTADT

M.Sc.

Anton Kuznietsov

AUTOtech.agil

Room 405
Phone 06151 / 16 24204
Email anton.kuznietsov@tu-darmstadt.de

MASCHINENBAU
We engineer future

FZD