Cluster A

Sensing and perception

A1 - Self-Supervised Learning for Camera–Radar Perception in Autonomous Driving

Autonomous driving relies on robust perception from camera and radar data. Supervised training is limited by labeled data, while self-supervised learning (SSL) enables learning from large unlabeled datasets.

Pretraining camera and radar backbones with SSL can improve feature quality, robustness, and generalization. Combining these modalities is especially promising due to their complementary strengths.

In this project, you will

- develop self-supervised pretraining methods for camera and radar backbones
- Explore multi-modal learning approaches (e.g., contrastive, masked, cross-modal)
- Train on unlabeled driving data
- Evaluate on downstream tasks
- Pretrained multi-modal backbone
- Evaluation on perception tasks
- Insights into SSL for radar - camera backbones

A2 -Self-supervised learning via a joint embedding predictive architecture on autonomous driving images

I-JEPA is trained on and achieves impressive results on ImageNet which consists of single prominent object images. Applying this learning strategy directly on AD images however yields poor results as the objective function is completely misaligned with the multi-object driving scenes in our data.

In this project we want to investigate how the I-JEPA training strategy and model architecture can be adapted for the complexity of AD scenes.

In this master thesis project, you will focus on:

- Investigating a hierarchical masking approach suitable for the differences in object scales.
- How the supervision signal should be applied to supervise objects at different scales.
- Look at how to gain local features with the JEPA objective (known to produce noisy feature maps)
- Investigate the learned features consistency between frames.

A3 - Enhance dynamic object state prediction using synthetic data

Ground truth data for rare events is often hard to find, such as for example emergency breaking scenarios. Furthermore, pseudo-annotations and manual annotations may suffer from position jitter throughout a sequence making it hard to get reliable GT for motion state predictions.

In this project, we aim to make these predictions more accurate by enhancing the training dataset with synthetic data.

In this master thesis project, you will focus on:

- Learning about IDSplat for rendering camera and lidar
- Setting up a recurrent model for predicting object states
- Rendering edge of distribution cases for example high acceleration, high yaw rate, velocity changes.
- Compare performance with post processing filter based model.

A4 - Towards Better Occupancy World Models for Autonomous Driving

Driving world models have recently emerged as a promising research direction for autonomous driving. Unlike conventional perception systems that only estimate the current state of the environment, driving world models aim to predict or generate future driving scenes conditioned on past observations, ego motion, maps, or other contextual information. By explicitly modeling the temporal evolution of driving environments, they provide a unified framework for future scene prediction and have broad applications in simulation, scenario generation, safety validation, and potentially closed-loop autonomous driving.

In this project, you will investigate better occupancy world models for autonomous driving by exploring different components of the world modeling pipeline. Possible research directions include

- improved occupancy tokenizers
- more effective latent scene representations
- enhanced spatial-temporal attention mechanisms
- geometry-aware training objectives
- multi-scale modeling strategies

which better capture both global scene evolution and fine-grained local structures.

A5 3D-Grounded World Models for Autonomous Driving

Accurate and geometrically consistent world models are increasingly recognized as a key enabling technology for end-to-end autonomous driving. Autonomous systems must be able to predict the future evolution of complex dynamic 3D/4D environments under different actions, requiring scene representations that are both physically plausible and geometrically accurate. Despite recent progress, current world models still struggle to produce accurate and consistent long-horizon predictions.

In this project, you will investigate whether explicit depth supervision can improve the geometric consistency and downstream vehicle performance of Gaussian World Models (GWMs). Improving the geometric fidelity of GWMs through explicit depth supervision could enhance the reliability of future-state prediction, supporting more effective policy learning, simulation, and decision-making in safety-critical driving scenarios.

A6 - Semantic Superquadric Splatting for Structured 3D Scenes

Recent Gaussian Splatting (GS) advances have demonstrated its effectiveness in high-fidelity rendering and 3D reconstruction. In parallel to this research direction is structured 3D object reconstruction, where compact geometric primitives are used to describe scenes in an interpretable and semantically meaningful manner. While 3D planes and ellipsoids are commonly used due to their simplicity, they are limited in representing the diversity of real-world geometry, particularly for intricate structures (imagine a teapot or a tree). Instead, superquadrics strike a balance between a compact and yet expressive parametric representation, capable of modelling a wide range of shapes while maintaining geometric interpretability with a handful of parameters.

In this project, you will develop a method, Semantic Superquadric Splatting, that simultaneously optimizes image semantics and geometry for scene-level reconstruction. Ensure that reconstructed models include semantics suitable for downstream applications such as autonomous car navigation. You will also extend reconstruction beyond indoor scenes by conducting comprehensive tests on challenging autonomous driving outdoor environments.