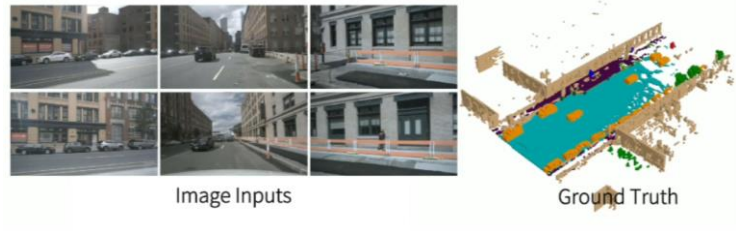


Master Thesis

Vision-based Open-Vocabulary Semantic Occupancy Forecasting for Autonomous Driving

Background:

Autonomous driving systems must perceive the current 3D scene while anticipating its future evolution. Compared with bounding boxes, semantic occupancy provides a dense representation of scene geometry and semantics, making it better suited for modeling occlusions, road structures, irregular objects, and complex dynamic environments. However, most vision-based occupancy prediction and forecasting methods remain limited to closed-set categories and struggle with rare or unseen objects. Recent works such as Cam4DOcc [1] and ForecastOcc [2] have advanced camera-only 4D occupancy forecasting and vision-based semantic occupancy forecasting from image sequences, while PG-Occ [3] explores open-vocabulary occupancy prediction with text-aligned representations and Gaussian-based scene modeling. Yet these directions remain largely separate. This thesis will study vision-based open-vocabulary semantic occupancy forecasting, aiming to predict future 3D semantic occupancy from camera observations while supporting flexible text-driven categories for more generalizable, future-aware autonomous driving scene understanding.



However, most vision-based occupancy prediction and forecasting methods remain limited to closed-set categories and struggle with rare or unseen objects. Recent works such as Cam4DOcc [1] and ForecastOcc [2] have advanced camera-only 4D occupancy forecasting and vision-based semantic occupancy forecasting from image sequences, while PG-Occ [3] explores open-vocabulary occupancy prediction with text-aligned representations and Gaussian-based scene modeling. Yet these directions remain largely separate. This thesis will study vision-based open-vocabulary semantic occupancy forecasting, aiming to predict future 3D semantic occupancy from camera observations while supporting flexible text-driven categories for more generalizable, future-aware autonomous driving scene understanding.

Your Tasks:

- Conducting a literature review on camera-only occupancy prediction, 4D occupancy forecasting, and open-vocabulary 3D scene understanding.
- Designing a model that predicts future 3D semantic occupancy from multi-view camera sequences.
- Evaluating the proposed method on autonomous driving datasets and comparing it with relevant closed-set and open-vocabulary baselines.

Your Profile:

- Strong background in computer vision, deep learning, robotics, or related fields.
- Motivation to work on open-ended research problems and conduct independent experiments.
- Solid programming skills in Python and experience with PyTorch.
- Good analytical skills and ability to communicate research ideas clearly in English.

Start date: Immediately

Duration: As per the applicable examination regulations.

If you are interested or have any questions regarding this thesis position, feel free to contact:

M.Sc. **Lei Wan**

Tel: +49 016 270-91287

E-Mail: lei.wan@partner.kit.edu

[1] Ma, Junyi, et al. "Cam4docc: Benchmark for camera-only 4d occupancy forecasting in autonomous driving applications." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[2] Mohan, Riya, et al. "ForecastOcc: Vision-based Semantic Occupancy Forecasting." arXiv preprint arXiv:2602.08006 (2026).

[3] Yan, Chi, and Dan Xu. "Progressive Gaussian Transformer with Anisotropy-aware Sampling for Open Vocabulary Occupancy Prediction." arXiv preprint arXiv:2510.04759 (2025).