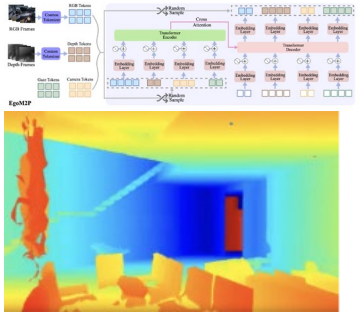
Technische Hochschule Ingolstadt

Write your Master Thesis with us!

Analysis of Semantic Scene Understanding Methods for their Deployment in Unmanned Aerial Vehicles

Recent advances in visual-language models have created many opportunities to enhance semantic understanding for UAV task applications. Computer vision research has established methods for semantic segmentation and contextual scene interpretation in ground-based scenarios which makes the system understand the objects and their semantic relationship, spatial understanding and functional roles within complex environments. However, it is challenging to deploy these semantic understanding models on UAV platforms. We focus on investigating how scene understanding models trained on ground-based data generalize to aerial perspectives while maintaining the semantic reasoning abilities.



Your Tasks

- Literature review on the state-of-the-art of real-time scene understanding
- Develop different methods to achieve real-time understanding
- Evaluate the application of the methods from real-world UAVs platform (e.g. ModalAI, Crazyflie 2.1)

Your Profile

- Studying Computer Science/Engineering (AI, HCI)
- Interest in computer vision, scene understanding and UAVs
- Experience in Python/C++/C# and ROS

Contact:

Prof. Dr. Simon Schwerd
simon.schwerd@thi.de

Diah A.I.
Diahayu.irawati@thi.de

Technologietransferzentrum Unbemannte Flugsysteme

References:

[1] Huang, C., Mees, O., Zeng, A., & Burgard, W. (2023). Visual language maps for robot navigation. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). https://doi.org/10.1109/ICRA48891.2023.10160758
[2] Huang, C., Mees, O., Zeng, A., & Burgard, W. (2025). Multimodal spatial language maps for robot navigation and manipulation. ArXiv. https://doi.org/10.48550/arXiv.2506.06862
[3] Li, G., Chen, Y., Wu, Y., Zhao, K., Pollefeys, M., & Tang, S. (2025). EgoM2P: Egocentric multimodal multitask pretraining. ArXiv. https://doi.org/10.48550/arXiv.2506.07886
[4] Shinya, Y., & S, A. (2025). CLIP-ReFine: Refining CLIP zero-shot learners with a very small amount of data. ArXiv. https://arxiv.org/abs/2504.12717
[5] You, P., Liu, Z., Jie, Z., & Lin, C. (2022). OpenScene: 3D Scene Understanding with Open Vocabularies. ArXiv. https://arxiv.org/abs/2211.15654