Master Thesis Proposal: Integration of New Computer Vision Tech for Robotic Manipulation

Jiacheng Xu & Dr. Sichao Liu

Background

Robotic manipulation in unstructured environments remains a significant challenge in robotics research. This project aims to integrate techniques Dino-X [2] for open-vocabulary object detection, Gaussian Splatting [3] for 3D scene reconstruction, and ReKep [1] for task decomposition and spatial-temporal reasoning—to enable robots to perform complex manipulation tasks. Specifically, the system will address tasks such as identifying and grasping objects like a cup, even in cluttered or dynamic environments. ReKep provides a novel approach to representing manipulation constraints using Relational Keypoint Constraints (RKC), which are optimized through a hierarchical procedure to compute robot actions.

Application

The proposed framework targets applications in robotic manipulation for tasks such as pick-and-place, grasping, and task generalization in real-world environments like homes and warehouses. Using visual and language inputs (e.g., "pick up a cup"), the system will identify objects in 2D scenes, reconstruct their 3D representations, and compute optimal actions to achieve the task, even in scenarios involving task failure or long-horizon planning.

Problems

The project involves solving several key mathematical and computational challenges:

- **Open-Vocabulary Object Detection and Segmentation**: Efficiently identifying objects and generating precise masks using DinoX.

- **3D Scene Reconstruction**: Constructing rough but accurate object representations using Gaussian Splatting for spatial understanding.

- **Task Decomposition and Optimization**: Formulating and solving tasks as sequences of Relational Keypoint Constraints (RKC) using numerical optimization techniques.

- **Generalization and Error Handling**: Developing algorithms for task failure detection and generalization across diverse tasks and environments.

Time Schedule

- **Month 1-2**: Literature review and setup of development environment.

- **Month 3-4**: Integration of DinoX for object detection and Gaussian Splatting for 3D reconstruction.

- **Month 5**: Development of ReKep-based task decomposition and optimization framework.

- **Month 6**: Final analysis, report writing, and thesis submission.

References

[1] Wenlong Huang, Chen Wang, Yunzhu Li, Ruohan Zhang, and Li Fei-Fei. Rekep: Spatio-temporal reasoning of relational keypoint constraints for robotic manipulation. *arXiv preprint arXiv:2409.01652*, 2024.

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

[3] Tianhe Ren, Yihao Chen, Qing Jiang, Zhaoyang Zeng, Yuda Xiong, Wenlong Liu, Zhengyu Ma, Junyi Shen, Yuan Gao, Xiaoke Jiang, et al. Dino-x: A unified vision model for open-world object detection and understanding. *arXiv preprint arXiv:2411.14347*, 2024.