Robotics and Embedded Systems

KI.FABRIK

Master/Bachelor Thesis: Assemble-as-I-do

Learning LLM-based Task Planning for Multi-stage Manipulation

Background

Large Language Models (LLMs) have gained popularity in robot task planning for long-horizon manipulation tasks [1]. To enhance the validity of LLM-generated plans, visual demonstrations and online videos have been widely employed to guide the planning process [2][3]. However, for manipulation tasks involving subtle movements but rich contact interactions (such as assembly), visual perception alone is not insufficient for the LLM to fully interpret the demonstration. Incorporating tactile and force-torque information from human demonstrations will greatly enhance LLMs' ability to generate plans for new task scenarios [4][5]. Additionally, audio guidance may also benefit LLM's understanding about the demonstration.

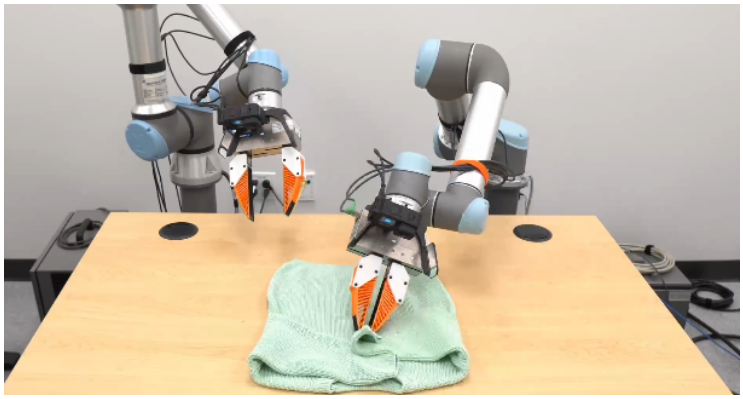
Your Tasks

In this project, you will develop a LLM-based task planning framework based on our previous work [4]. Specifically, your task will include:

1. collect multi-modal human demonstration for manipulation tasks using Universal Manipulation Interface [6][7],
2. extends the current framework in terms of accurate event segmentation, task parameter grounding and generation of executable task plans,
3. evaluate the framework on a wide-range of manipulation tasks.

Requirement

- Self-motivation and love for robots;
- Basic knowledge of robot task planning and LLMs;
- At least six-month working time;
- Python programming experiences;
- Working experience with LLMs will be a plus



Supervisor: Prof. Alois Knoll

Advisor: Kejia Chen kejia.chen@tum.de

Lehrstuhl für Robotik, Künstliche Intelligenz und Echtzeitsysteme,

Fakultät für Informatik, Technische Universität München

[1] Ao, Jicong, et al. "Llm as bt-planner: Leveraging llms for behavior tree generation in robot task planning." 2025 IEEE International Conference on Robotics and Automation (ICRA).

[2] C. Wang, et al."MimicPlay: Long-Horizon Imitation Learning by Watching Human Play," in Proceedings of The 7th Conference on Robot Learning. PMLR, Dec. 2023, pp. 201–221.

[3] N. Wake, et al. "GPT-4V(ision) for Robotics: Multimodal Task Planning From Human Demonstration," in IEEE Robotics and Automation Letters, vol. 9, no. 11, pp. 10567-10574, Nov. 2024.

[4] Chen, Kejia, et al. "Learning Task Planning from Multi-Modal Demonstration for Multi-Stage Contact-Rich Manipulation." 2025 IEEE International Conference on Robotics and Automation (ICRA).

[5] Jones, Joshua, et al. "Beyond Sight: Finetuning Generalist Robot Policies with Heterogeneous Sensors via Language Grounding." arXiv preprint arXiv:2501.04693 (2025).

[6] Chi, Cheng, et al. "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots." arXiv preprint arXiv:2402.10329 (2024).

[7] Liu, Wenhai, et al. "ForceMimic: Force-Centric Imitation Learning with Force-Motion Capture System for Contact-Rich Manipulation." arXiv preprint arXiv:2410.07554 (2024).