

Masterarbeit

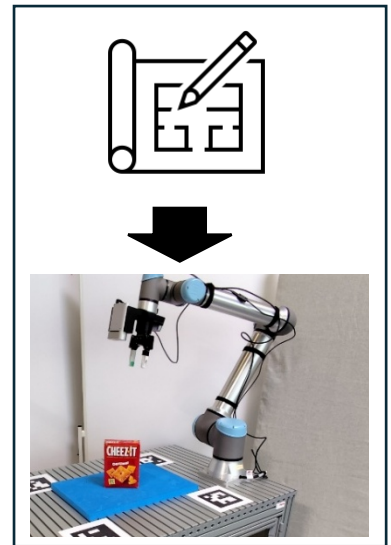
Sketch-to-Skill: Using Vision-Language Models to Interpret Hand-Drawn Manipulation Instructions

Problem formulation

Teaching robots new manipulation tasks typically requires teleoperation, programming, or collecting demonstrations, which can be time-consuming and unintuitive for non-expert users. Humans, however, can easily communicate task intentions through simple sketches or diagrams [1]. Recent vision-language foundation models have shown strong capabilities in interpreting visual instructions and contextual information. This thesis investigates whether hand-drawn sketches can serve as an intuitive teaching modality for robot manipulation, where a robot interprets human sketches in the context of a scene to infer manipulation goals and interaction strategies.

Task definition

The student will develop a system that allows users to teach robot manipulation tasks using hand-drawn sketches overlaid on images of the robot workspace. A vision-language model will interpret the sketch together with the scene image to infer task intent, relevant objects, and desired motion or interaction regions. These outputs will be translated into executable robot actions such as grasping, pushing, or placing. The approach will be implemented and evaluated on a real robot platform across selected manipulation tasks, analyzing usability, task success, and robustness.



Requirements

- Solid knowledge base and experience in deep learning and robotics.
- Good coding skills in Python.
- Experience with ROS2 is beneficial

Literature

[1] P. Yu, A. Bhaskar, A. Singh, Z. Mahammad, and P. Tokekar, "Sketch-to-Skill: Bootstrapping Robot Learning with Human Drawn Trajectory Sketches," Mar. 14, 2025, arXiv: arXiv:2503.11918. doi: 10.48550/arXiv.2503.11918.

Betreuer: M.Sc. Edgar Welte

Betr. Hochschullehrerin: Prof. Dr.-Ing. Rania Rayyes

Bearbeitungszeitraum: SS 2027

Bearbeiter: offen