

Introduction

Motivation:

Traditional robot programming is rigid and expert-dependent. While Behavior Trees (BTs) offer modularity and reactivity, the framework integrates high-level cognitive planning with low-level reactive control, with BTs serving both as the structured output of planning and as the governing structure for execution.

Contributions:

- Video-to-BT Long-Horizon Planning:

A novel framework that directly translates human demonstration videos into structured, executable BTs, achieving highly accurate planning across diverse complex robotic assemblies.

- Perception-Driven Control:

A supervisory system integrating a BT executor with real-time semantic perception for precise, context-aware scheduling in dynamic environments.

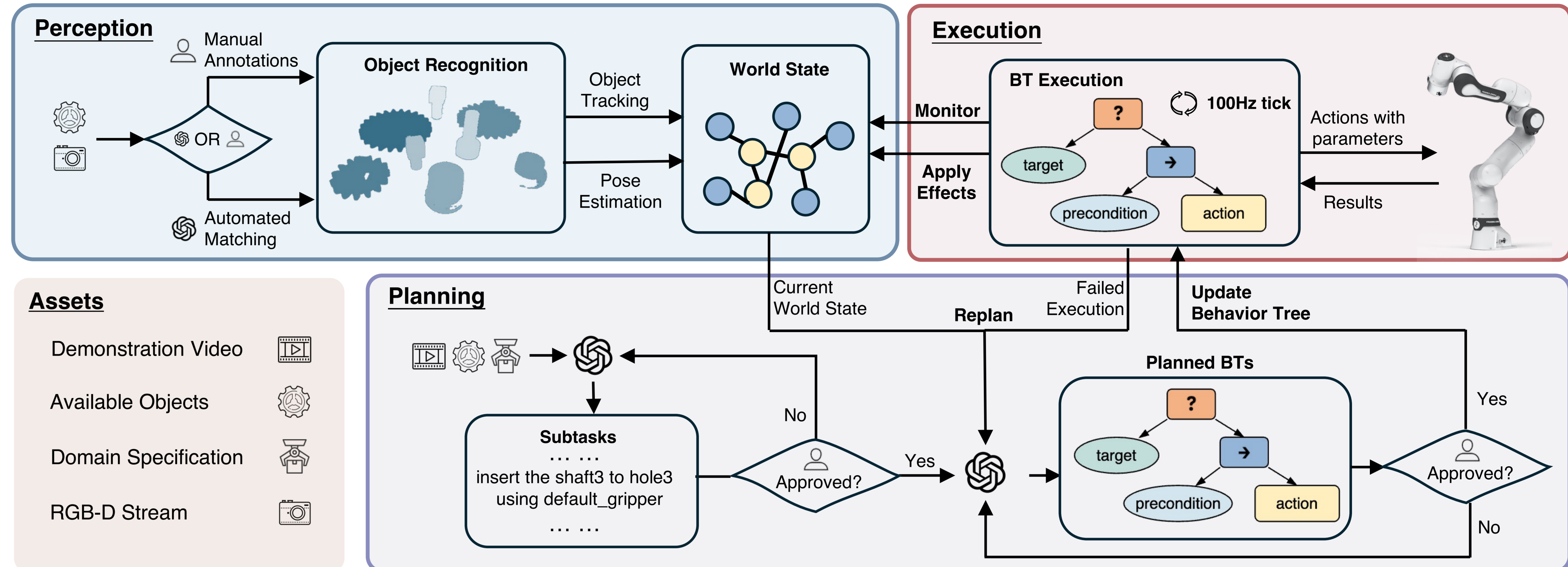
- Robust Execution Against Disturbances:

Sustains superior completion rates for long-horizon tasks, demonstrating exceptional resilience even under severe external physical impacts

Website: https://video2bt.github.io/video2bt_page/



Framework Overview



Methods

Stage A: Video to Subtask sequence

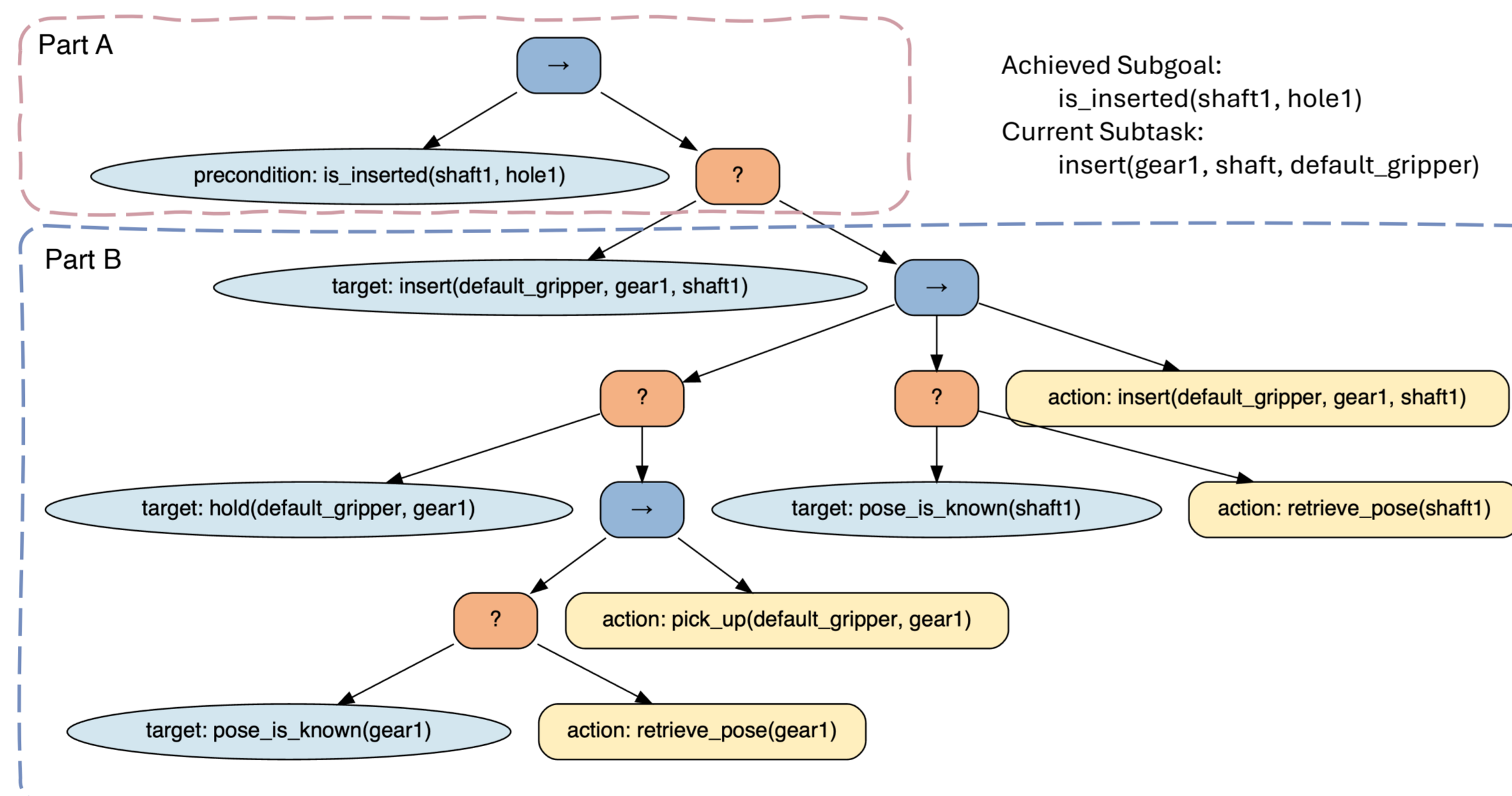
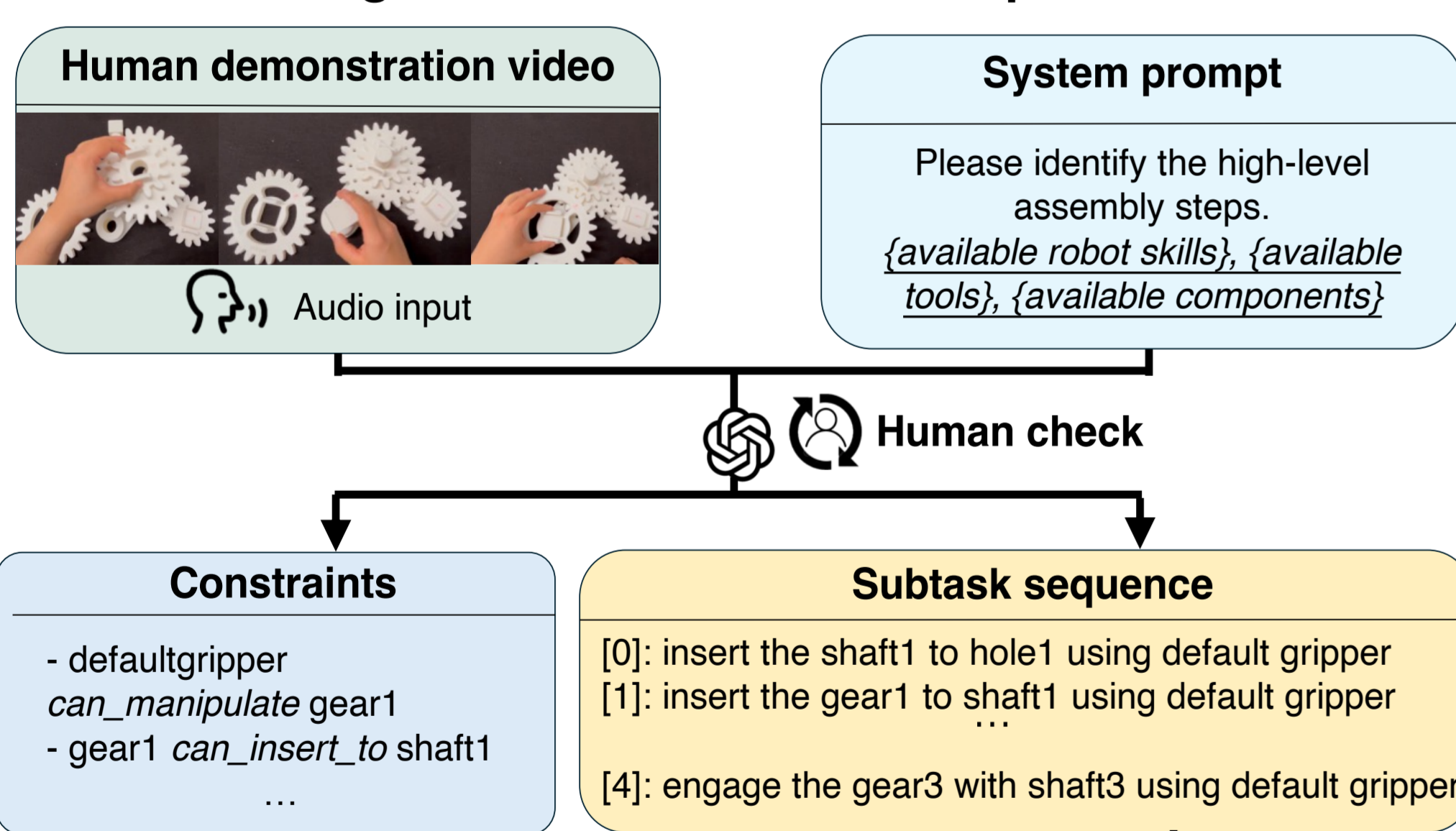


Figure b. Example of an executable BT

Stage B: Subtask sequence to BTs

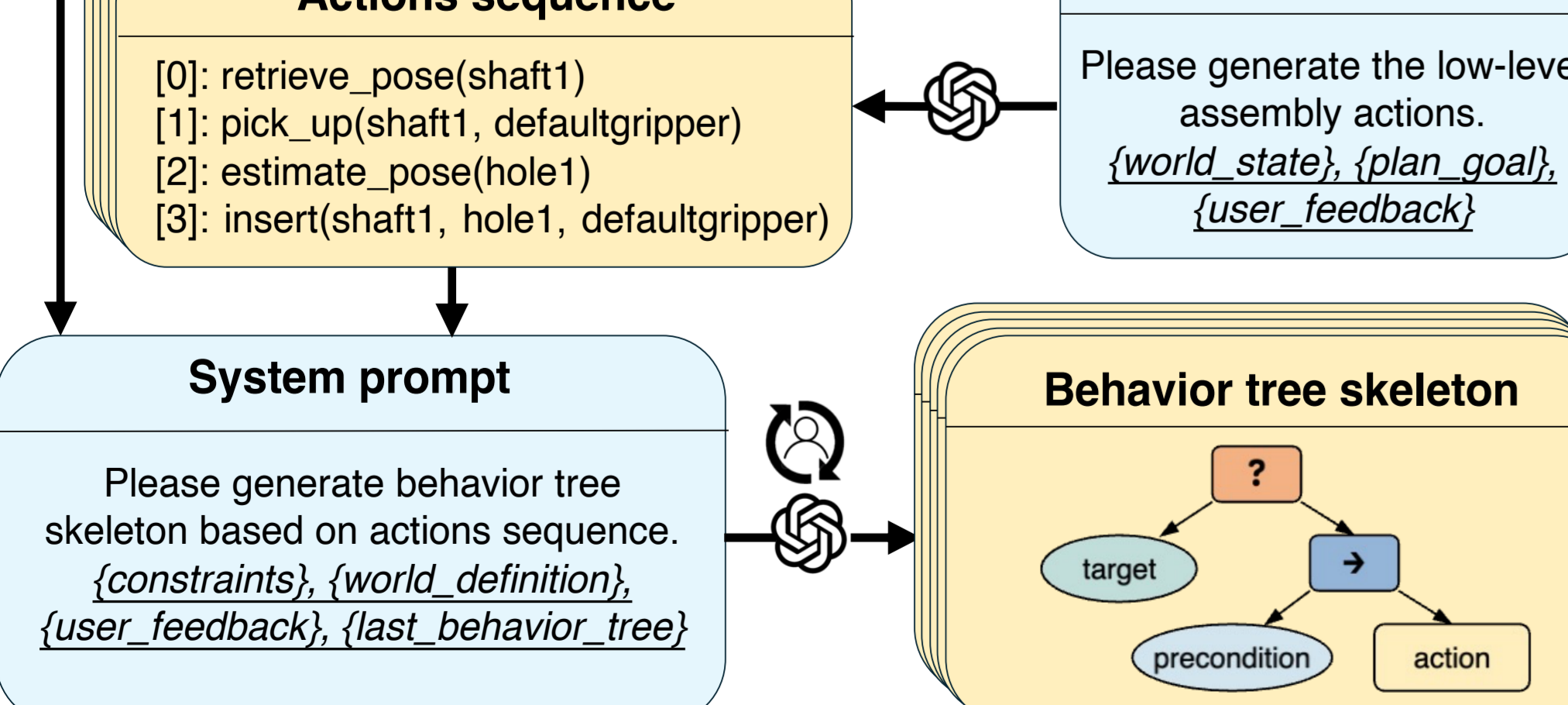


Figure a. BT generation from human demonstration video

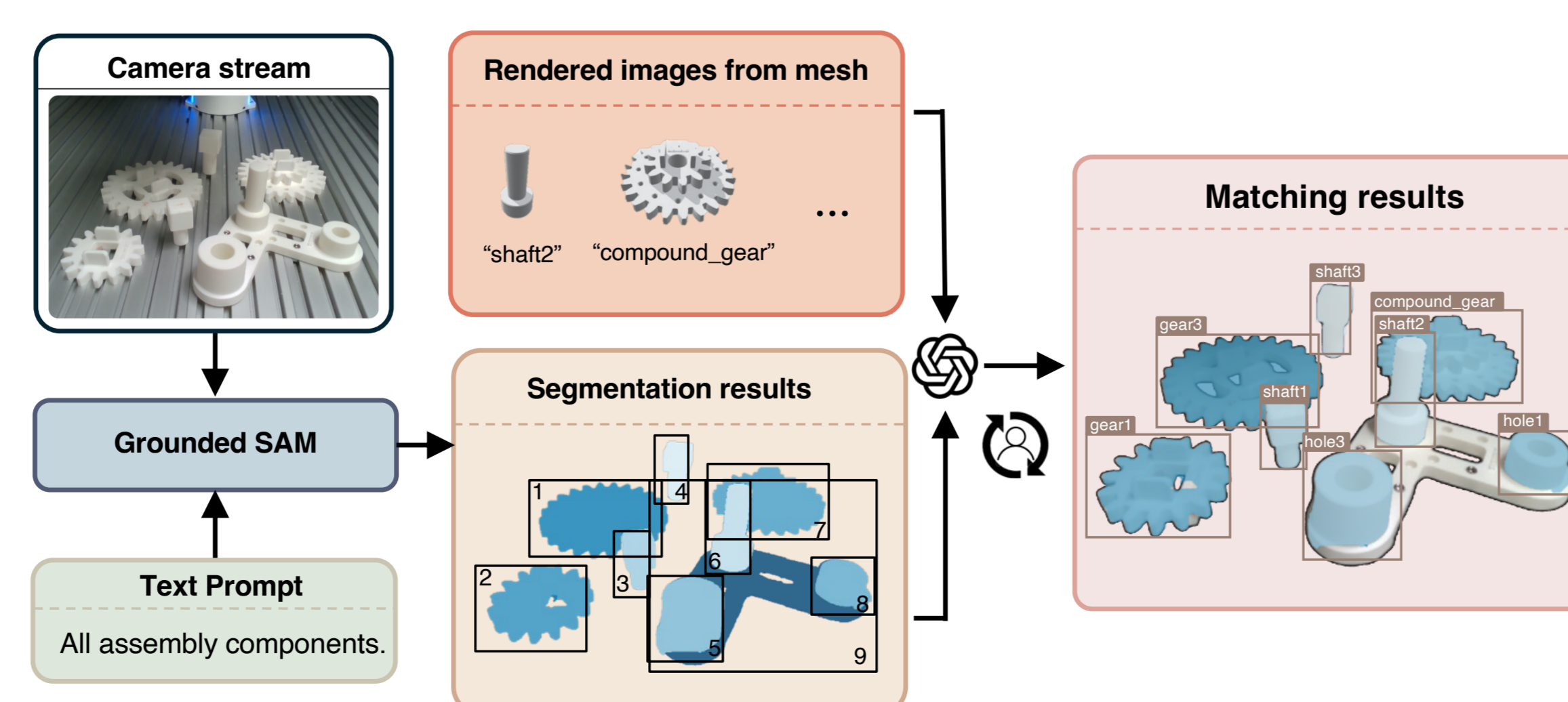
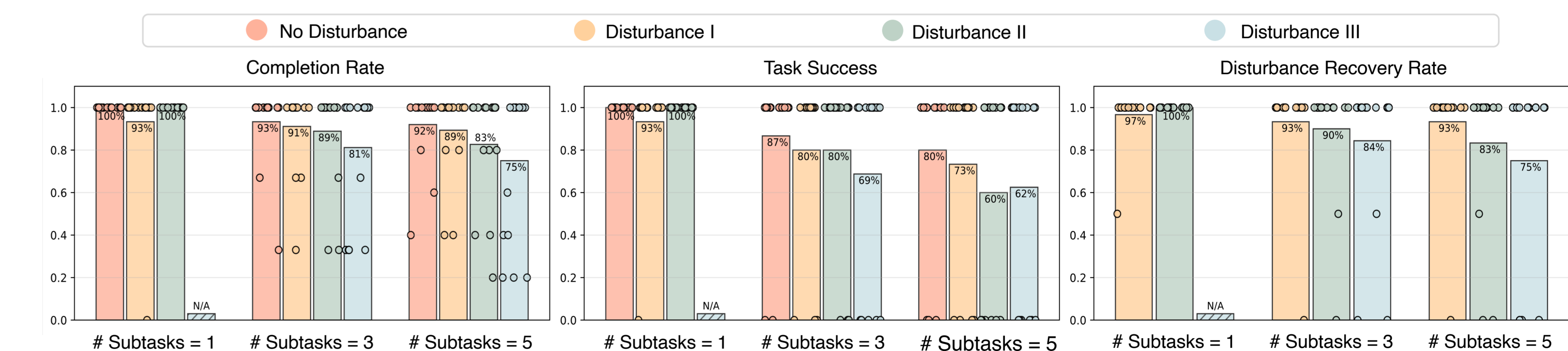
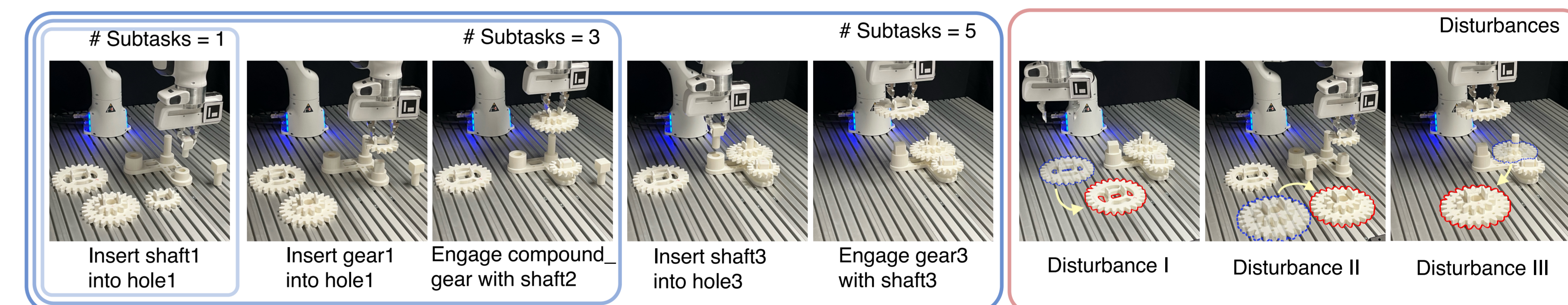


Figure c. Scene object recognition

Evaluation



Affiliation:

* - Equal contribution
 1- Munich Institute of Robotics and Machine Intelligence (MIRMI), Technical University of Munich (TUM), Munich, Germany.
 2- Aalto University, Espoo, Finland.
 3- Shanghai University, Shanghai, China.
 4- Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE.

Contacts:

M.Sc. Xiwei Zhao
zhaoxiwei118@gmail.com
 B.Sc. Yiwei Wang
yiwei.wang.03@tum.de

