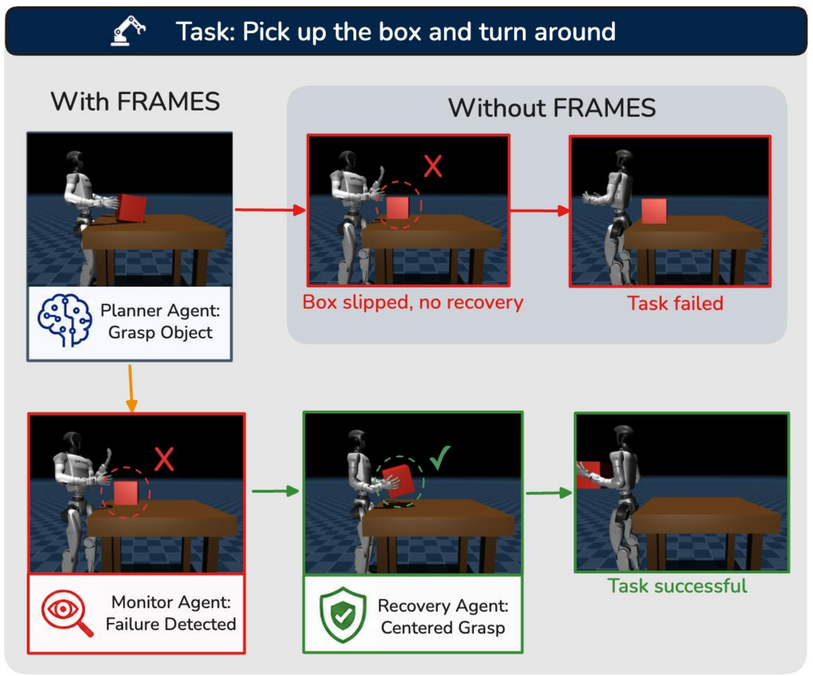
DexLab
General Robotics Lab
Duke | THOMAS LORD DEPARTMENT of MECHANICAL ENGINEERING & MATERIALS SCIENCE

FRAMES: Failure Recovery And Monitoring of Embodied Skills for Humanoid Loco-Manipulation

Ajay Vikram Periasami, Xinyuan Luo, Haoyu Li, Xianyi Cheng

IROS 2026 Workshop
Full-Shift Robot Co-Workers:
Active Perception and Interaction
for Human Environments

Project Overview



FRAMES detects failures and selects corrective actions to limit error propagation and support long-horizon task completion.

Motivation

- Humanoid loco-manipulation tightly couples locomotion and manipulation: a grasp can fail during walking or turning.
- Undetected failures leave the planner with an incorrect world state, propagating errors through subsequent actions.
- Reliable execution requires monitoring the skills, interrupting failed actions, and grounding recovery in the observed states.
- A modular supervisory layer provides these capabilities without modifying or retraining the pretrained whole-body controller.

Monitor Design

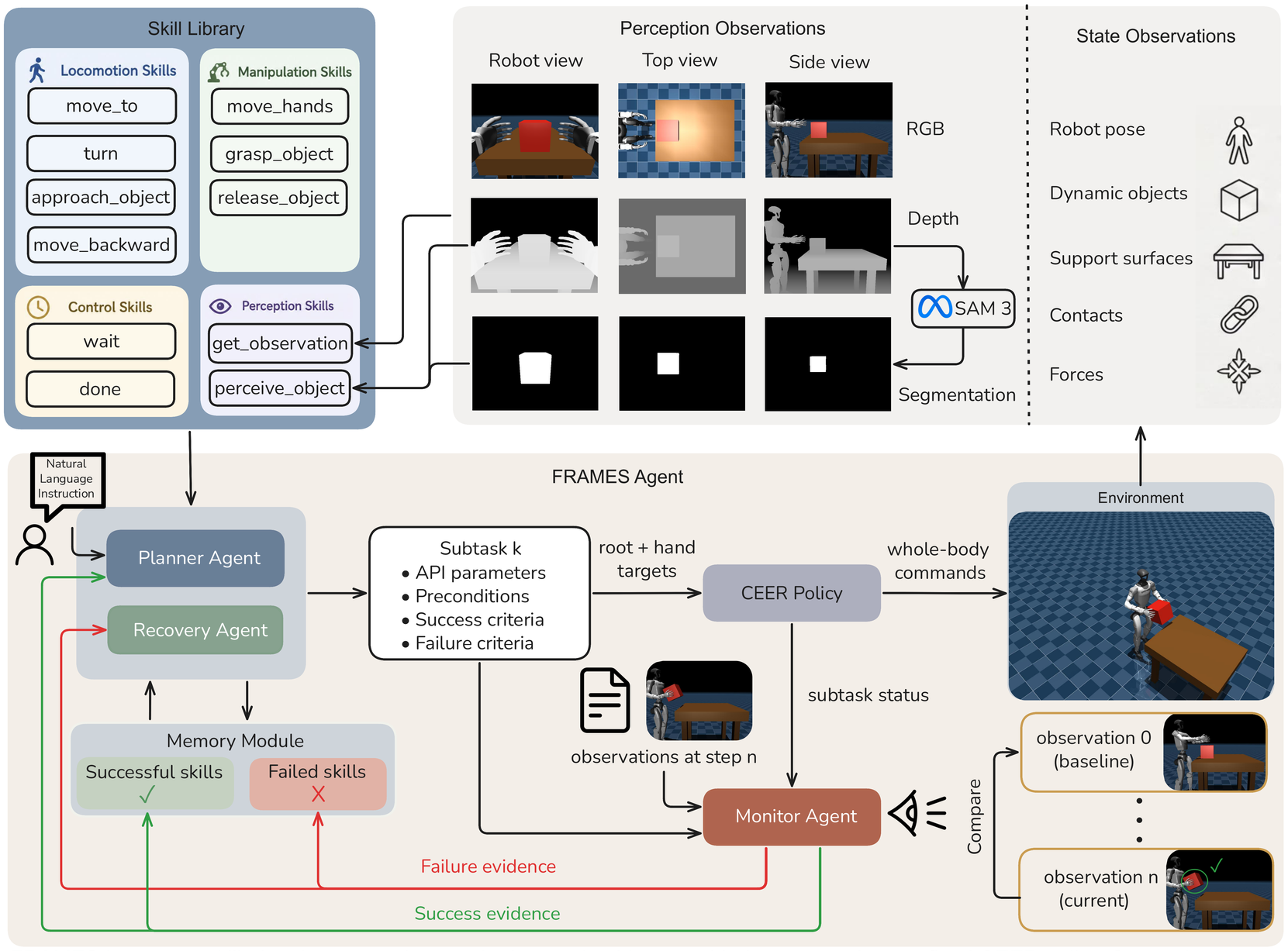
- The Monitor Agent compares skill-start and current images from three camera views, together with robot state and contact changes, against skill-specific criteria and preconditions.
- It tracks progress, confirms success after execution ends, and interrupts failed skills with grounded feedback for recovery.

Results

Benchmark	Failures Detected	Successful Runs Accepted	Overall Accuracy
Failed grasp	8/10	10/10	90%
Dropped object	10/10	10/10	100%
Collision during approach	10/10	7/10	85%
Wrong-object grasp	10/10	10/10	100%
Wrong placement	10/10	9/10	95%
Overall	**48/50**	**46/50**	**94%**

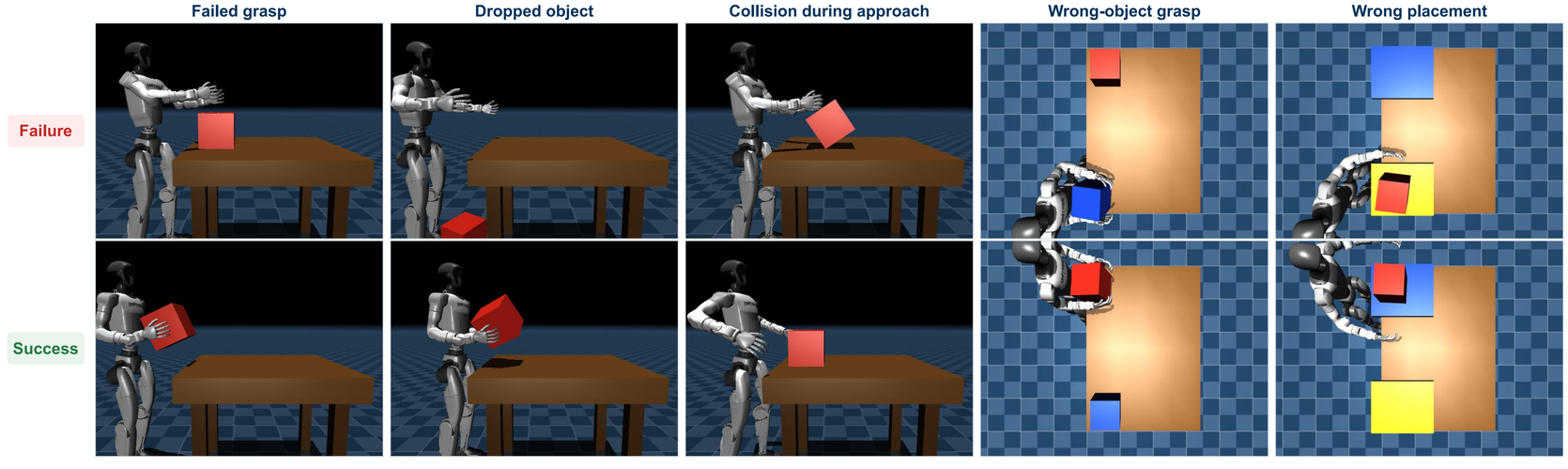
- Two failed grasps remained labelled “in progress” after execution ended.
- Safe approaches triggered false collision alarms when the robot’s hands appeared close to the box.
- A correct placement was rejected because the box appeared to remain held after release.

System Framework



FRAMES integrates skill planning, execution monitoring, and evidence-guided recovery above a fixed CEER policy. Semantic memory guides skill parameters using previous experience. Geometric grounding uses depth and segmentation to locate target objects in 3D.

Failure and Success Examples



Representative failures (top) and matched successful executions (bottom) across five MuJoCo tasks: grasping, maintaining a hold, collision-free approach, object selection, and placement. Each task includes 10 failed and 10 successful trials with varied parameters.