

rc_study_table_case1

RoboCerebra | goal: The task involves organizing a white bowl, a wooden tray, tomato sauce

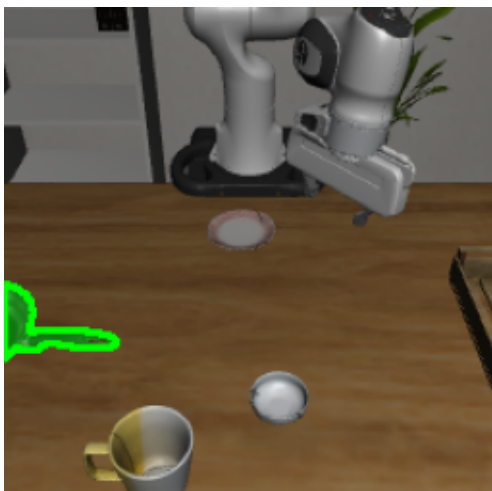
6 annotated steps + annotation inventory + methodology



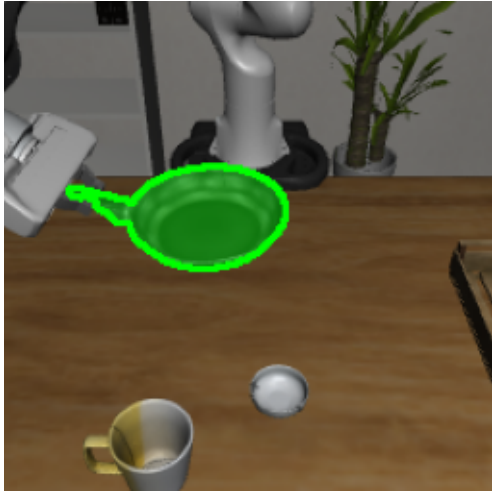
Subgoal 1 [grasp]: Pick up the tomato sauce from the study table | target='tomato sauce' (green GT mask, focus_kind=object, 440px) | related_objects: tomato sauce, study table | frames {'start': 0, 'end': 648}



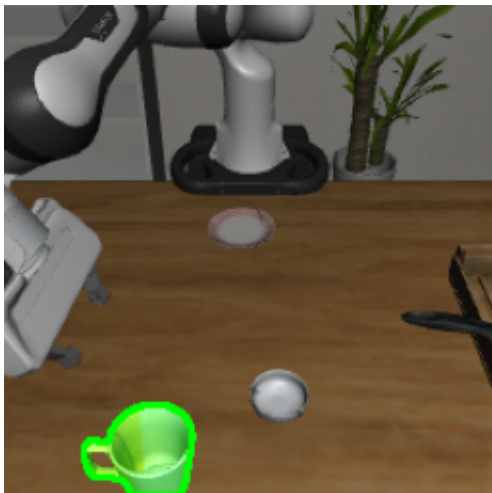
Subgoal 2 [release]: Place the tomato sauce on the wooden tray | target='wooden tray' (green GT mask, focus_kind=container, 440px) | related_objects: tomato sauce, wooden tray | frames {'start': 648, 'end': 871}



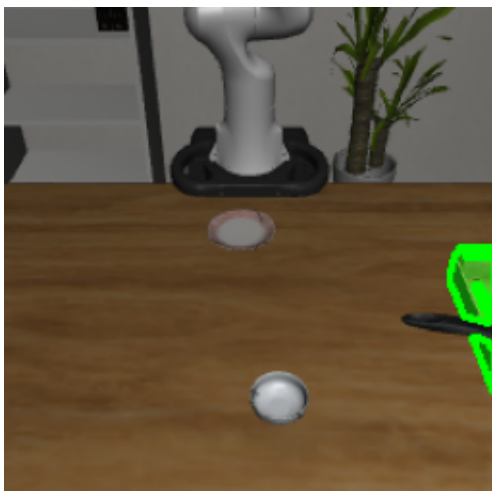
Subgoal 3 [grasp]: Pick up the frypan from the study table | target='chefmate 8 frypan' (green GT mask, focus_kind=object, 665px) | related_objects: frypan, study table | frames {'start': 871, 'end': 1451}



Subgoal 4 [release]: Place the frypan on the wooden tray
 | target='chefmate 8 frypan' (green GT mask,
 focus_kind=object, 2887px) | related_objects: frypan,
 wooden tray | frames {'start': 1451, 'end': 1988}



Subgoal 5 [grasp]: Pick up the white yellow mug from the
 study table | target='white yellow mug' (green GT mask,
 focus_kind=object, 1508px) | related_objects: white
 yellow mug, study table | frames {'start': 1988, 'end':
 2448}



Subgoal 6 [release]: Place the white yellow mug on the
 wooden tray | target='wooden tray' (green GT mask,
 focus_kind=container, 699px) | related_objects: white
 yellow mug, wooden tray | frames {'start': 2448, 'end':
 2983}

ANNOTATION INVENTORY -- rc_study_table_case1

long-horizon goal: The task involves organizing a white bowl, a wooden tray, tomato sauce, a chefmate 8" frypan, and a white-yellow mug from the study table, placing all items onto the wooden tray for meal preparation.

1. Subgoal decomposition . 6 subtasks (LLM-decomposed, human-verified) -> subtask + phase
[TEMPORAL]

2. related_objects per-subtask target object list (from task_description) [SEMANTIC]

3. timestep per-subtask frame range (start,end) [TEMPORAL]

4. Focus GT mask sim instance-seg of the target object, per subgoal -> green overlay
[SPATIAL, engine-exact]

focus_kind / mask_px .. how the target was matched to a sim instance + mask area.

5. Forced-memory markers . keyframes=6 ; onset_occluded_steps=2 -> steps whose target is
already occluded at onset (answer needs MEMORY, not current frame) -- the held-out test
condition.

6. states / actions full recorded MuJoCo trajectory (enables the re-sim that yields the
masks).

How these annotations were produced (RoboCerebra + our pipeline):

- RoboCerebra ships RGB video + (states, actions) + a task_description (high-level instruction LLM-decomposed

into a subtask sequence, each with a frame range 'timestep' and its 'related_objects'). Human-verified.

- Forced-memory structure: meta.json marks 'keyframes' and 'onset_occluded_steps' -- subgoals whose target

is already occluded at the step onset, so the answer must come from MEMORY, not the current frame.

- Focus MASK (green): we RE-SIMULATE each trajectory in the LIBERO/MuJoCo engine (set every recorded mujoco

state) and read the simulator's GROUND-TRUTH instance segmentation for the subgoal's target object

(matched by name). Engine-exact, obtained by replay -- not hand-drawn. focus_kind = how it was matched;

mask_px = mask area. RoboCerebra does NOT ship masks.

- Sim render -> exact masks but synthetic appearance (visual sim2real gap).