

rc_coffee_table_case1

RoboCerebra | goal: Arrange a plate, red coffee mug, Chefmate 8 frypan, and wine bottle fr

8 annotated steps + annotation inventory + methodology



Subgoal 1 [grasp]: Pick up the plate from coffee table |
 target='plate' (green GT mask, focus_kind=object, 133px)
 | related_objects: plate, coffee table | frames
 {'start': 0, 'end': 642}



Subgoal 2 [release]: Place the plate onto the wooden two
 layer shelf top region | target='plate' (green GT mask,
 focus_kind=object, 134px) | related_objects: plate,
 wooden two layer shelf | frames {'start': 642, 'end':
 853}



Subgoal 3 [grasp]: Pick up the red coffee mug from coffee
 table | target='red coffee mug' (green GT mask,
 focus_kind=object, 339px) | related_objects: red coffee
 mug, coffee table | frames {'start': 853, 'end': 1545}



Subgoal 4 [release]: Place the red coffee mug onto the wine rack top region | target='red coffee mug' (green GT mask, focus_kind=object, 499px) | related_objects: red coffee mug, wine rack | frames {'start': 1545, 'end': 2034}



Subgoal 5 [grasp]: Pick up the frypan from coffee table | target='chefmate 8 frypan' (green GT mask, focus_kind=object, 159px) | related_objects: frypan, coffee table | frames {'start': 2034, 'end': 2969}



Subgoal 6 [release]: Place the frypan onto the bowl drainer left region | target='chefmate 8 frypan' (green GT mask, focus_kind=object, 299px) | related_objects: frypan, bowl drainer | frames {'start': 2969, 'end': 3794}



Subgoal 7 [grasp]: Pick up the wine bottle from coffee table | target='wine bottle' (green GT mask, focus_kind=object, 121px) | related_objects: wine bottle, coffee table | frames {'start': 3794, 'end': 4572}



Subgoal 8 [release]: Place the wine bottle onto the wooden two layer shelf bottom region | target='wine bottle' (green GT mask, focus_kind=object, 202px) | related_objects: wine bottle, wooden two layer shelf | frames {'start': 4572, 'end': 5540}

ANNOTATION INVENTORY -- rc_coffee_table_case1

long-horizon goal: Arrange a plate, red coffee mug, Chefmate 8 frypan, and wine bottle from the coffee table onto designated storage surfaces including a wooden two-layer shelf, wine rack, and bowl drainer for organized display.

1. Subgoal decomposition . 8 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=8 ; onset_occluded_steps=0 -> 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).