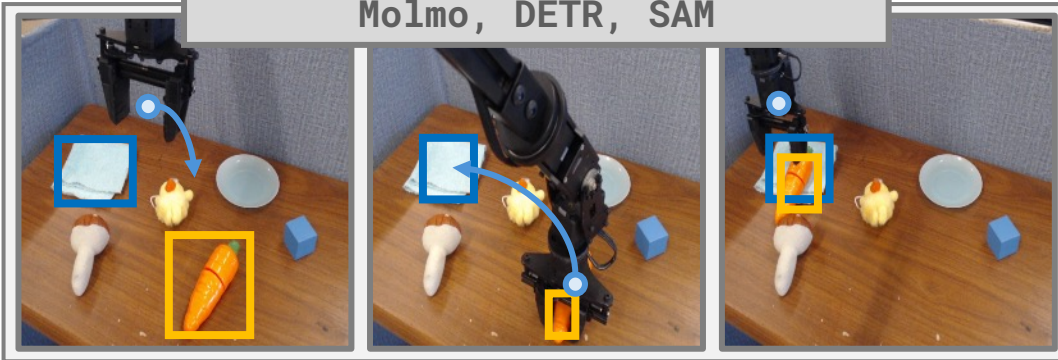


(1) Annotate robot trajectories with embodied reasoning features and subtasks

MoIms, DETR, SAM



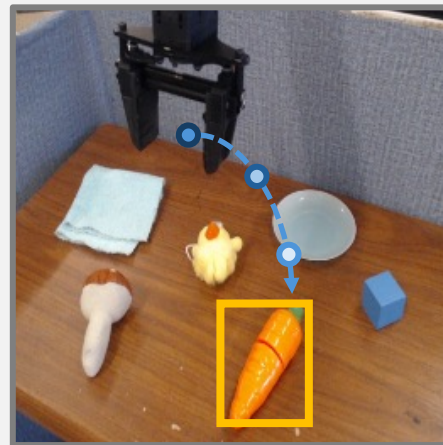
Gemini

Subtask Decomposition:

1. reach for carrot, 2. firmly grasp carrot,
3. lift above towel, 4. drop carrot on towel

(2) Generate diverse steering commands for all subtasks using VLMs

Generate steering commands using:
task | motions | gripper coords | bboxes



Reach for carrot:

- Go right to the carrot
- Reach for the carrot at `<carrot bbox>`
- Move right and down
- Go from $[x_1 \ y_1]$, $[x_2 \ y_2]$, $[x_3 \ y_3]$

Firmly grasp carrot:

...