



b)

01 Scene-Grounding ETC

Sequence in 3D space → Camera Calibration →

02 Task-Oriented ETC

Grasp here to pick up the mango
[295, 397] Affordance

Put the mouse into the box
Planning

Sim

01 Coffee Bag

Real-world setup

Predict the trajectory of: Pick up the COL beans.

[[<loc_128>, <loc_612>], [<loc_230>, <loc_590>], [<loc_333>, <loc_569>], [<loc_410>, <loc_547>], [<loc_486>, <loc_533>]]

[[128, 612], [230, 590], [333, 569], [410, 547], [486, 533]]

Real

03 BBQ

Real-world setup

Given the bounding box, is this region suitable for a stable grasp?

Yes.

Yes.

03 Action-Aligned Planning ETC

Gripper position to Book 3:
- Front/Back: in front
- Up/Down: slightly above
- Left/Right: centered around it

OpenAI

Spatial Understanding

Which is the 3-pin tile?
The tile with three dots.

Locate the 3-pin.

[250, 218, 288, 255].

Object description

02 Flower

Real-world setup

Generated

Replace the monstera with lily.

Locate the lily with bounding box.

[551, 11, 17, 931]

ETC Data

