

## HCP

To GPT-4o: You are controlling a robot ..., Here is one example... Can you do it for the following input:  $[I_0]$  “Put the blocks into the corresponding shaped slots.”

## Primitive Actions

1. reach yellow cylinder
2. grasp yellow cylinder
3. ...

4. Place yellow cylinder

Timestep  $t$

To GPT-4o: Preliminaries: IOR is ...  $[\hat{I}_t]$  First, can you identify the gripper, the manipulated object, the destination in the image to complete the task “Place yellow cylinder”? And give me the IOR in the form of “Action: ...”

SAM

GroundingDINO

Initialize

Track-Anything

## IOR

Action: *place*  
Object: *yellow cylinder*  
Target: *round slot*  
Constraint: *axis (0,0,-1)*

3rd -view mask

1st -view masked depth

GEA



High freq

Update masks

Low freq

Wrong,  
Renew IOR

To GPT-4o:  $[I_{t+1}]$  Current task is “Place yellow cylinder”, which needs ... The robot’s gripper is closed. Can you give me the status of the task process ? Success, Wrong or Normal?

Success,  
Next task