

## Inputs

### RGB-D Image



### Instruction

*Place the capybara on top of the paper labeled "Open6DOR", with the capybara facing frontwards.*

## Task Preprocessing Module

### Grounded SAM



Object Masks

[ paper ]  
[ red bowl ]  
[ **capybara** ]



Target Object  
RGB Image

## 3D-aware Position Module

GPT-4V Prediction

3D Object Point Clouds  
Object Centroids & Bboxes

GPT-4V Selection

Sim.-assisted Filtering

Large-scale Sampling

Mesh Reconstruction

## Simulation-assisted Rotation Module

## Sim.-assisted Planning Module

Goal Position

Goal Rotation

Scene Point Cloud

cuRobo  
Motion Planner



Generated Traj.