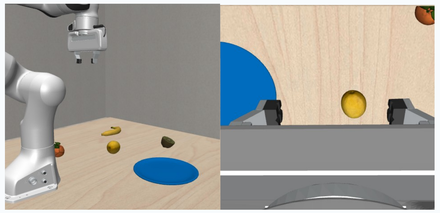


"Put the orange on the plate"

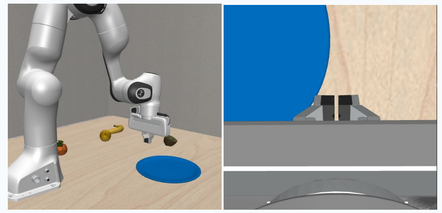
Obs. & Instr.

Vision-Language-Action Model

Pre-trained VLM | Action Expert

Joint or Pose

RLFT



Obs. & Instr.

Correction Instr.

Failure Detection and Recovery

in Robotic Manipulation

Action or Execution Results

Correction Action

Foundation Model

Large Language Model

The orange slipped from the gripper.

Locate the position of the orange, then grasp it and put it on the plate.

Multimodal Large Language Model

Audio Force Tactile

MLLM

1. Multimodal Cross-checking

2. Replanning Correction

World Model