Maestro

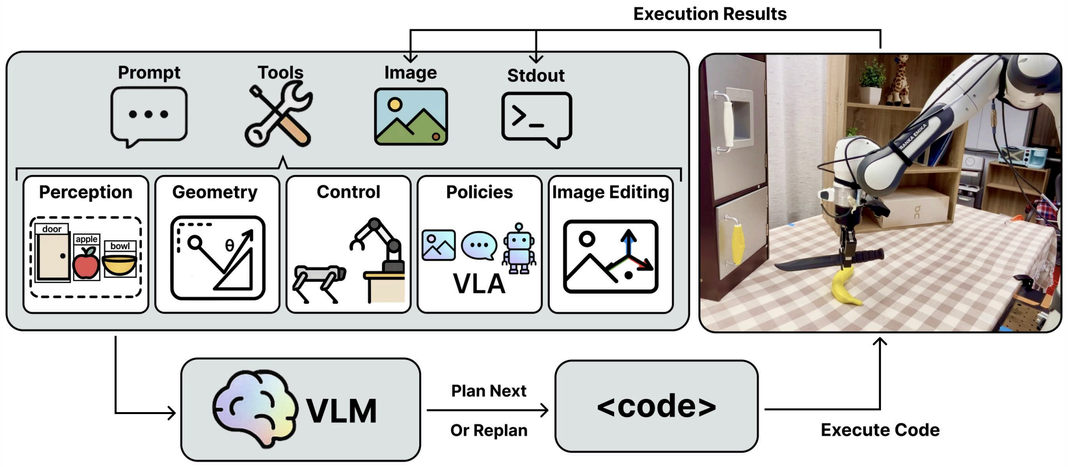
Orchestrating Robotics Modules with Vision-Language Models for Zero-Shot Generalist Robots

Junyao Shi*, Rujia Yang*, Kaitian Chao*, Bingqing Wan, Yifei Shao, Jiahui Lei, Jianing Qian, Long Le, Pratik Chaudhari, Kostas Daniilidis, Chuan Wen, Dinesh Jayaraman

Penn Engineering
GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

Overview

Method



System Tools

Tool Category	Prior Work Examples	MAESTRO Modules
Tabletop Manipulation		
Perception	Raw sensory inputs (RGB + proprioception); Segmentation/Bounding Box centers [2, 8, 13]	Raw sensory inputs (RGB + proprioception); Segmentation centers; **Active perception (zoom/look around with wrist camera)**; **FoundationStereo [25] depth**; **Gemini pointing; VLM-selected task-relevant keypoints (ReKep-inspired [26])**
Control	Cartesian control, gripper control [2, 8]; Movement primitives [13]	Cartesian control, gripper control; **cuRobo collision-free motion planning**
Learned Visuomotor Policies	m2t2 grasp model [27]	**GraspGen grasp model [28]**; **$\pi_{0.5}$ VLA with high-frequency closed-loop monitoring (Qwen-2.5-VL)**
Geometry & Linear Algebra (new)	*None*	**Distance measurement, vector construction, vector rotation, relative rotation between vectors**
Image Editing (new)	*None*	**Draw points, overlay 6D poses to improve visual grounding**
Extra Modules for Mobile Manipulation		
Perception	Build global/local map [6]	Mobile base state estimation; **Active perception tools (look left/right/ground, view carry-on basket, log object location)**
Locomotion	Navigation [6]	Navigation; **Fine-grained "nudge" tool for local adjustment**

Results

Tabletop Manipulation

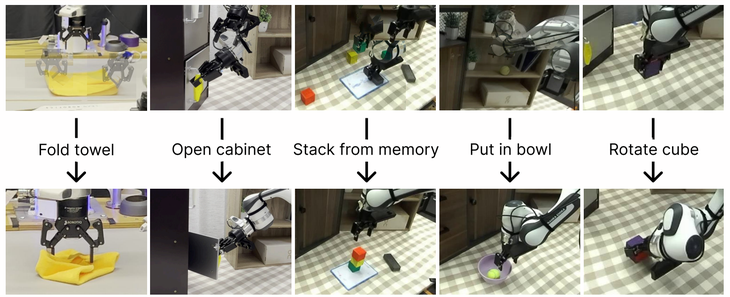
Challenge	Task Description	Gemini Robotics Agent	π_0	$\pi_{0.5}$	MAESTRO
Pick–Place	Put item in bowl	73.3 ± 46.2	74.0 ± 37.1	70.0 ± 41.1	**98.0 ± 4.5**
Deformable object	Fold the four corners of the towel into the center	40.0 ± 17.3	47.0 ± 25.1	70.0 ± 15.4	**71.3 ± 21.4**
Articulated object	Open cabinet	3.3 ± 5.8	8.3 ± 2.9	0.0 ± 0.0	**68.0 ± 31.3**
Spatial reasoning	Rotate cube purple side up	23.6 ± 3.5	29.0 ± 1.7	10.0 ± 0.0	**60.0 ± 38.1**
Memory & long-horizon & semantic	Erase instructions on whiteboard, then follow instruction to stack cups	26.7 ± 24.7	12.0 ± 12.0	22.0 ± 22.8	**63.0 ± 16.8**

Mobile Manipulation

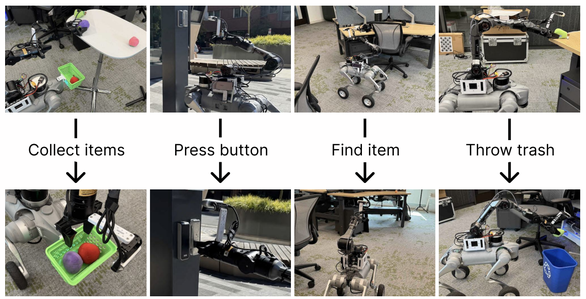
Task Category	Task Description	MAESTRO
Long-horizon manipulation	Collect all toys on table	85.0 ± 22.4
Long-horizon loco-manipulation	Throw green ball into garbage can	76.7 ± 14.9
Active exploration	Search item and put on table	96.0 ± 8.9
Object affordance	Press button to open door	93.3 ± 14.9

Experiments

Tabletop Manipulation



Mobile Manipulation



Evolution