

Model	Put spoon on towel		Put carrot on plate		Stack green cube on yellow cube		Put eggplant in basket		Average	
	Grasp/Put	Success	Grasp/Put	Success	Grasp/Put	Success	Grasp/Put	Success	Grasp/Put	Success
Octo [40] (Zero-Shot)	34% / 26%	9%	35% / 20%	7%	28% / 0%	0%	65% / 51%	33%	40.5% / <u>30.1%</u>	12.2%
Octo (Bridge V2 SFT)	52% / 41%	29%	32% / 37%	14%	47% / 4%	3%	60% / 11%	5%	47.5% / 23.0%	12.8%
Octo+RoboEngine [51]	67% / 21%	14%	43% / 37%	16%	43% / 5%	2%	0% / 0%	0%	38.2% / 20.9%	8.0%
Octo+RoboVIP (Text prompt)	59% / 7%	4%	69% / 46%	32%	55% / 9%	5%	63% / 17%	11%	61.5% / 21.1%	<u>13.0%</u>
Octo+RoboVIP (Text prompt with ID)	59% / 63%	37%	37% / 62%	23%	47% / 15%	7%	37% / 19%	7%	45.0% / 41.1%	18.5%
π_0 [7] (Zero-Shot)	52% / 63%	33%	0% / 0%	0%	28% / 7%	2%	31% / 42%	13%	27.75% / 43.2%	12%
π_0 (Bridge V2 SFT)	57% / 63%	36%	44% / 43%	19%	30% / 7%	2%	29% / 41%	12%	40% / 43.1%	17.25%
π_0 +RoboEngine [51]	61% / 70%	43%	33% / 30%	10%	61% / 11%	7%	31% / 45%	14%	46.5% / 39.8%	18.5%
π_0 +RoboVIP (Text prompt)	74% / 84%	62%	52% / 40%	21%	49% / 14%	7%	36% / 64%	23%	52.75% / 55.0%	29%
π_0 +RoboVIP (Text prompt with ID)	73% / 64%	47%	49% / 41%	20%	52% / 13%	7%	53% / 70%	37%	56.75% / <u>48.9%</u>	27.75%