

Table 1: OGBench Evaluation

Task	Methods					
	TMD	CRL	QRL	GCBC	GCIQL	GCIVL
humanoidmaze_medium_stitch	60.5 ^(± 1.6)	36.2 ^(± 0.9)	18.0 ^(± 0.7)	29.0 ^(± 1.7)	12.1 ^(± 1.1)	12.3 ^(± 0.6)
pointmaze_teleport_stitch	29.3 ^(± 2.2)	4.1 ^(± 1.1)	8.6 ^(± 1.9)	31.5 ^(± 3.2)	25.2 ^(± 1.0)	44.4 ^(± 0.7)
humanoidmaze_large_stitch	23.0 ^(± 1.5)	4.0 ^(± 0.2)	3.5 ^(± 0.5)	5.6 ^(± 1.0)	0.5 ^(± 0.1)	1.2 ^(± 0.2)
antmaze_teleport_explore	49.6 ^(± 1.5)	19.5 ^(± 0.8)	2.3 ^(± 0.7)	2.4 ^(± 0.4)	7.3 ^(± 1.2)	32.0 ^(± 0.6)
antmaze_large_stitch	20.9 ^(± 1.7)	10.8 ^(± 0.6)	18.4 ^(± 0.7)	3.4 ^(± 1.0)	7.5 ^(± 0.7)	18.5 ^(± 0.8)
scene_noisy	19.6 ^(± 1.7)	1.2 ^(± 0.3)	9.1 ^(± 0.7)	1.2 ^(± 0.2)	25.9 ^(± 0.8)	26.4 ^(± 1.7)
visual_antmaze_teleport_stitch	38.5 ^(± 1.5)	31.7 ^(± 3.2)	1.4 ^(± 0.8)	31.8 ^(± 1.5)	1.0 ^(± 0.2)	1.4 ^(± 0.4)
visual_antmaze_large_stitch	26.6 ^(± 2.8)	11.1 ^(± 1.3)	0.6 ^(± 0.3)	23.6 ^(± 1.4)	0.1 ^(± 0.0)	0.8 ^(± 0.3)

We **bold** the best performance. Success rate (%) is presented with the standard deviation across six seeds.