Table 1: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for DeepArch.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	2	82.5 (5.4)	3	95.3 (1.6)	6	273.4 (21.1)	5	215.7 (10.3)	**1**	**68.2 (1.2)**	4	117.4 (4.5)
	2	**1**	**76.6 (5.8)**	**1**	**72.6 (0.0)**	3	196.2 (15.4)	2	147.9 (5.9)	**1**	**72.0 (3.2)**	3	189.3 (10.7)
	3	2	60.4 (3.8)	3	71.1 (0.4)	6	283.7 (22.6)	5	202.9 (8.6)	**1**	**48.5 (2.1)**	4	132.5 (6.9)
S_2	1	**1**	**62.7 (3.6)**	2	95.7 (1.5)	4	171.1 (9.7)	4	164.7 (6.3)	**1**	**60.7 (0.8)**	3	104.2 (3.6)
	2	2	74.5 (4.9)	2	71.6 (0.0)	4	129.0 (6.0)	3	117.7 (3.8)	**1**	**62.0 (1.9)**	5	198.1 (9.3)
	3	2	50.8 (2.7)	3	63.3 (0.5)	6	175.0 (9.4)	5	153.5 (5.3)	**1**	**40.1 (1.4)**	4	131.5 (6.1)
S_3	1	**1**	**50.2 (2.3)**	3	85.4 (1.0)	5	123.5 (4.9)	6	152.5 (4.7)	2	57.1 (0.7)	4	95.5 (2.5)
	2	3	66.4 (5.0)	**1**	**55.7 (0.9)**	4	97.8 (4.0)	5	108.1 (3.3)	2	61.1 (1.6)	6	189.1 (8.3)
	3	2	45.4 (3.1)	3	55.1 (0.7)	4	124.3 (5.8)	5	140.6 (4.3)	**1**	**38.3 (1.2)**	4	124.8 (5.9)
S_4	1	**1**	**46.9 (2.6)**	3	76.2 (0.9)	5	103.6 (3.3)	6	149.1 (3.3)	2	55.3 (0.5)	4	94.9 (2.2)
	2	**1**	**62.6 (5.7)**	**1**	**59.5 (1.2)**	2	81.1 (2.7)	3	107.8 (2.9)	**1**	**62.5 (1.7)**	4	190.6 (6.7)
	3	2	40.1 (1.6)	3	60.2 (2.2)	4	101.2 (3.8)	6	135.9 (3.6)	**1**	**38.2 (0.8)**	5	125.4 (4.8)
S_5	1	**1**	**42.3 (1.9)**	3	84.2 (1.0)	5	95.2 (3.5)	6	145.0 (2.7)	2	54.7 (0.5)	4	90.9 (1.3)
	2	**1**	**55.1 (3.4)**	2	58.7 (1.4)	4	75.1 (2.4)	5	113.1 (2.8)	3	62.7 (1.6)	6	184.4 (4.7)
	3	**1**	**35.9 (1.1)**	3	55.5 (0.7)	4	89.8 (3.6)	6	136.3 (3.2)	2	37.8 (0.7)	5	117.5 (3.4)
Average r		**1.5**		2.4		4.4		4.8		**1.5**		4.3	

Table 2: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for SaC.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**1353.0 (191.7)**	4	2918.7 (47.2)	3	2720.2 (234.1)	3	2615.0 (43.8)	4	3079.0 (260.1)	2	2174.6 (200.8)
	2	**1**	**105.4 (8.4)**	2	245.6 (1.4)	2	241.5 (6.5)	3	253.5 (1.7)	4	359.4 (13.5)	4	367.6 (17.5)
	3	**1**	**41.9 (6.7)**	4	177.9 (0.3)	5	185.4 (5.4)	3	166.1 (0.6)	6	352.0 (18.7)	2	122.1 (4.6)
S_2	1	**1**	**754.1 (143.1)**	5	3874.7 (264.3)	3	2579.9 (111.2)	3	2598.9 (60.4)	4	2840.7 (128.4)	2	1646.5 (112.4)
	2	**1**	**59.1 (4.0)**	5	375.8 (23.4)	2	223.1 (2.8)	3	251.6 (1.9)	4	320.5 (12.1)	3	257.0 (10.4)
	3	**1**	**26.9 (2.4)**	6	350.2 (37.7)	3	163.2 (2.6)	4	165.3 (0.6)	5	277.9 (12.4)	2	115.8 (5.5)
S_3	1	**1**	**438.6 (90.4)**	6	3621.6 (349.9)	3	2474.6 (84.8)	4	2630.2 (77.8)	5	2869.9 (129.4)	2	1464.7 (75.5)
	2	**1**	**30.8 (2.3)**	6	344.6 (22.4)	3	215.2 (2.2)	4	251.6 (2.4)	5	297.2 (10.0)	2	204.2 (7.0)
	3	**1**	**20.9 (1.1)**	6	370.5 (36.3)	3	153.7 (2.0)	4	165.5 (0.6)	5	240.5 (9.8)	2	108.5 (3.5)
S_4	1	**1**	**287.3 (72.9)**	5	3687.4 (227.2)	3	2492.3 (107.3)	3	2455.7 (94.6)	4	2871.8 (130.5)	2	1477.6 (85.3)
	2	**1**	**32.8 (2.2)**	6	369.6 (27.0)	3	211.8 (2.5)	4	252.9 (2.7)	5	286.0 (7.5)	2	187.8 (5.6)
	3	**1**	**17.7 (0.9)**	6	256.2 (30.4)	3	151.7 (2.0)	4	164.6 (0.7)	5	227.4 (8.7)	2	111.6 (3.9)
S_5	1	**1**	**234.4 (63.4)**	6	3621.9 (315.6)	4	2402.6 (116.0)	3	2135.1 (87.9)	5	2672.4 (139.1)	2	1345.2 (92.8)
	2	**1**	**30.4 (2.9)**	6	351.5 (26.2)	3	211.2 (2.6)	4	254.6 (2.9)	5	273.7 (6.9)	2	172.0 (4.3)
	3	**1**	**15.9 (0.7)**	6	305.1 (45.6)	3	150.6 (1.3)	4	164.0 (0.8)	5	223.1 (8.8)	2	108.9 (4.3)
Average r		**1.0**		5.3		3.1		3.5		4.7		2.2	

Table 3: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for SQLite.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**1.0 (0.0)**	5	5.5 (0.0)	3	3.7 (0.0)	4	4.5 (0.1)	6	15.5 (0.5)	2	1.6 (0.0)
	2	**1**	**1.0 (0.0)**	5	11.7 (0.0)	4	7.9 (0.0)	3	4.8 (0.2)	6	17.8 (0.6)	2	1.7 (0.0)
	3	**1**	**1.1 (0.1)**	5	3.6 (0.0)	3	2.2 (0.0)	4	3.1 (0.1)	6	15.6 (0.5)	2	1.2 (0.0)
	4	**1**	**1.2 (0.1)**	3	3.5 (0.0)	2	2.1 (0.0)	4	4.0 (0.2)	5	15.6 (0.5)	**1**	**1.2 (0.0)**
S_2	1	**1**	**0.9 (0.0)**	5	4.5 (0.0)	3	3.6 (0.0)	4	4.1 (0.1)	6	5.2 (0.3)	2	1.4 (0.0)
	2	**1**	**0.8 (0.0)**	6	10.6 (0.0)	4	7.9 (0.0)	3	4.2 (0.1)	5	8.4 (0.2)	2	1.3 (0.0)
	3	**1**	**0.9 (0.0)**	4	2.6 (0.0)	3	2.1 (0.0)	4	2.6 (0.1)	5	4.5 (0.3)	2	1.1 (0.0)
	4	**1**	**1.0 (0.0)**	4	2.5 (0.0)	3	2.0 (0.0)	5	3.4 (0.2)	6	4.5 (0.3)	2	1.1 (0.0)
S_3	1	**1**	**0.9 (0.0)**	5	4.4 (0.0)	3	3.5 (0.0)	4	3.7 (0.1)	4	3.6 (0.1)	2	1.2 (0.0)
	2	**1**	**0.7 (0.0)**	5	9.6 (0.0)	4	7.8 (0.0)	3	4.1 (0.1)	4	7.8 (0.1)	2	1.1 (0.0)
	3	**1**	**0.9 (0.0)**	6	2.6 (0.0)	3	2.1 (0.0)	4	2.2 (0.1)	5	2.5 (0.1)	2	1.0 (0.0)
	4	**1**	**1.0 (0.0)**	5	2.5 (0.0)	3	2.0 (0.0)	6	3.2 (0.2)	4	2.4 (0.1)	2	1.1 (0.0)
S_4	1	**1**	**0.8 (0.0)**	4	6.3 (0.5)	3	3.5 (0.0)	3	3.5 (0.2)	3	3.5 (0.0)	2	1.1 (0.0)
	2	**1**	**0.7 (0.0)**	4	5.5 (0.5)	6	7.9 (0.0)	3	3.9 (0.1)	5	7.8 (0.1)	2	1.0 (0.0)
	3	**1**	**0.9 (0.0)**	5	3.1 (0.3)	3	2.1 (0.0)	3	2.1 (0.1)	4	2.2 (0.0)	2	1.0 (0.0)
	4	**1**	**1.0 (0.0)**	5	2.6 (0.1)	3	2.1 (0.0)	6	3.0 (0.2)	4	2.1 (0.0)	2	1.0 (0.0)
S_5	1	**1**	**0.8 (0.0)**	5	6.2 (0.4)	3	3.5 (0.0)	3	3.5 (0.2)	4	3.6 (0.0)	2	1.0 (0.0)
	2	**1**	**0.7 (0.0)**	4	7.1 (0.5)	6	7.9 (0.0)	3	3.9 (0.1)	5	7.7 (0.1)	2	0.9 (0.0)
	3	**1**	**0.9 (0.0)**	5	2.4 (0.2)	4	2.1 (0.0)	3	1.9 (0.1)	5	2.5 (0.1)	2	0.9 (0.0)
	4	**1**	**1.0 (0.0)**	3	2.1 (0.1)	3	2.0 (0.0)	5	2.7 (0.2)	4	2.5 (0.1)	2	1.0 (0.0)
Average r		**1.0**		4.7		3.5		3.9		4.8		1.9	

Table 4: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for NGINX.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	3	130.2 (23.3)	**1**	**24.2 (1.7)**	4	347.5 (51.2)	2	61.7 (2.8)	6	682.8 (15.3)	5	649.2 (9.0)
	2	3	186.4 (34.2)	**1**	**25.2 (1.5)**	4	347.9 (51.1)	2	62.0 (2.7)	6	683.6 (15.4)	5	649.8 (9.1)
	3	4	313.4 (48.7)	**1**	**18.8 (2.0)**	3	229.7 (33.6)	2	36.0 (2.0)	6	525.3 (15.1)	5	502.7 (9.6)
	4	3	252.7 (51.3)	**1**	**24.8 (1.0)**	2	167.5 (22.6)	**1**	**24.5 (1.5)**	4	427.4 (16.2)	5	456.8 (13.2)
S_2	1	2	26.4 (3.2)	**1**	**23.5 (1.8)**	4	79.9 (30.0)	3	33.3 (1.9)	6	668.7 (9.4)	5	645.7 (7.9)
	2	**1**	**11.7 (1.2)**	2	27.8 (1.7)	4	80.1 (30.0)	3	33.2 (1.9)	6	669.9 (9.5)	5	646.8 (8.0)
	3	**1**	**31.8 (20.7)**	**1**	**18.5 (1.4)**	2	49.4 (19.7)	**1**	**16.0 (1.4)**	4	525.3 (11.0)	3	508.2 (9.2)
	4	3	36.9 (20.7)	2	24.7 (1.1)	3	43.2 (13.2)	**1**	**15.0 (0.6)**	4	445.5 (14.7)	5	471.2 (13.4)
S_3	1	**1**	**17.0 (1.4)**	3	27.0 (2.6)	2	20.7 (1.3)	3	26.7 (1.1)	5	658.2 (7.9)	4	642.5 (7.3)
	2	**1**	**9.1 (1.2)**	3	28.4 (2.8)	2	20.8 (1.3)	3	26.9 (1.1)	5	659.3 (8.0)	4	643.3 (7.4)
	3	**1**	**5.4 (0.5)**	4	17.3 (1.0)	2	9.3 (0.9)	3	10.8 (0.7)	6	517.4 (8.4)	5	507.9 (7.7)
	4	**1**	**6.7 (0.8)**	4	23.9 (0.7)	3	16.3 (0.5)	2	13.1 (0.2)	5	436.1 (10.9)	6	469.9 (10.9)
S_4	1	**1**	**10.4 (0.7)**	4	28.7 (2.6)	2	16.9 (0.8)	3	23.3 (0.4)	6	657.3 (7.7)	5	639.5 (6.9)
	2	**1**	**7.4 (0.6)**	3	22.5 (1.1)	2	17.0 (0.8)	3	23.4 (0.4)	5	658.4 (7.8)	4	640.4 (7.0)
	3	**1**	**4.6 (0.3)**	4	19.5 (1.8)	2	6.7 (0.5)	3	8.9 (0.3)	6	517.6 (8.0)	5	505.8 (7.1)
	4	**1**	**5.2 (0.3)**	4	24.7 (1.0)	3	15.0 (0.3)	2	11.9 (0.1)	5	435.0 (10.4)	6	467.3 (10.0)
S_5	1	**1**	**6.4 (0.6)**	4	24.9 (2.1)	2	14.7 (0.4)	3	22.6 (0.5)	6	651.0 (5.1)	5	632.9 (3.9)
	2	**1**	**7.9 (1.6)**	4	24.7 (1.7)	2	14.8 (0.4)	3	22.8 (0.5)	6	652.1 (5.1)	5	633.4 (4.0)
	3	**1**	**4.2 (0.3)**	4	18.0 (1.5)	2	4.9 (0.2)	3	8.7 (0.3)	6	515.9 (5.4)	5	503.1 (4.1)
	4	**1**	**4.9 (0.3)**	4	23.9 (0.8)	3	13.8 (0.2)	2	11.5 (0.1)	5	433.5 (6.4)	6	464.6 (5.8)
Average r		**1.6**		2.8		2.6		2.4		5.4		4.9	

Table 5: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for SPEAR.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	3	130.2 (23.3)	**1**	**24.2 (1.7)**	4	347.5 (51.2)	2	61.7 (2.8)	6	682.8 (15.3)	5	649.2 (9.0)
	2	3	186.4 (34.2)	**1**	**25.2 (1.5)**	4	347.9 (51.1)	2	62.0 (2.7)	6	683.6 (15.4)	5	649.8 (9.1)
	3	4	313.4 (48.7)	**1**	**18.8 (2.0)**	3	229.7 (33.6)	2	36.0 (2.0)	6	525.3 (15.1)	5	502.7 (9.6)
	4	3	252.7 (51.3)	**1**	**24.8 (1.0)**	2	167.5 (22.6)	**1**	**24.5 (1.5)**	4	427.4 (16.2)	5	456.8 (13.2)
S_2	1	2	26.4 (3.2)	**1**	**23.5 (1.8)**	4	79.9 (30.0)	3	33.3 (1.9)	6	668.7 (9.4)	5	645.7 (7.9)
	2	**1**	**11.7 (1.2)**	2	27.8 (1.7)	4	80.1 (30.0)	3	33.2 (1.9)	6	669.9 (9.5)	5	646.8 (8.0)
	3	**1**	**31.8 (20.7)**	**1**	**18.5 (1.4)**	2	49.4 (19.7)	**1**	**16.0 (1.4)**	4	525.3 (11.0)	3	508.2 (9.2)
	4	3	36.9 (20.7)	2	24.7 (1.1)	3	43.2 (13.2)	**1**	**15.0 (0.6)**	4	445.5 (14.7)	5	471.2 (13.4)
S_3	1	**1**	**17.0 (1.4)**	3	27.0 (2.6)	2	20.7 (1.3)	3	26.7 (1.1)	5	658.2 (7.9)	4	642.5 (7.3)
	2	**1**	**9.1 (1.2)**	3	28.4 (2.8)	2	20.8 (1.3)	3	26.9 (1.1)	5	659.3 (8.0)	4	643.3 (7.4)
	3	**1**	**5.4 (0.5)**	4	17.3 (1.0)	2	9.3 (0.9)	3	10.8 (0.7)	6	517.4 (8.4)	5	507.9 (7.7)
	4	**1**	**6.7 (0.8)**	4	23.9 (0.7)	3	16.3 (0.5)	2	13.1 (0.2)	5	436.1 (10.9)	6	469.9 (10.9)
S_4	1	**1**	**10.4 (0.7)**	4	28.7 (2.6)	2	16.9 (0.8)	3	23.3 (0.4)	6	657.3 (7.7)	5	639.5 (6.9)
	2	**1**	**7.4 (0.6)**	3	22.5 (1.1)	2	17.0 (0.8)	3	23.4 (0.4)	5	658.4 (7.8)	4	640.4 (7.0)
	3	**1**	**4.6 (0.3)**	4	19.5 (1.8)	2	6.7 (0.5)	3	8.9 (0.3)	6	517.6 (8.0)	5	505.8 (7.1)
	4	**1**	**5.2 (0.3)**	4	24.7 (1.0)	3	15.0 (0.3)	2	11.9 (0.1)	5	435.0 (10.4)	6	467.3 (10.0)
S_5	1	**1**	**6.4 (0.6)**	4	24.9 (2.1)	2	14.7 (0.4)	3	22.6 (0.5)	6	651.0 (5.1)	5	632.9 (3.9)
	2	**1**	**7.9 (1.6)**	4	24.7 (1.7)	2	14.8 (0.4)	3	22.8 (0.5)	6	652.1 (5.1)	5	633.4 (4.0)
	3	**1**	**4.2 (0.3)**	4	18.0 (1.5)	2	4.9 (0.2)	3	8.7 (0.3)	6	515.9 (5.4)	5	503.1 (4.1)
	4	**1**	**4.9 (0.3)**	4	23.9 (0.8)	3	13.8 (0.2)	2	11.5 (0.1)	5	433.5 (6.4)	6	464.6 (5.8)
Average r		**1.6**		2.8		2.6		2.4		5.4		4.9	

Table 6: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for Storm.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**31.2 (1.7)**	3	225.7 (5.5)	2	157.4 (39.3)	4	934.4 (21.1)	5	21531.3 (558.9)	6	62365.6 (1192.9)
	2	2	144.5 (29.1)	4	529.8 (8.1)	3	373.6 (54.5)	**1**	**63.9 (0.6)**	6	43178.7 (853.9)	5	21690.5 (474.8)
	3	**1**	**251.3 (18.3)**	4	1578.8 (10.4)	3	1113.8 (16.8)	2	500.8 (34.4)	6	36052.8 (728.3)	5	6480.9 (356.7)
S_2	1	**1**	**31.4 (2.6)**	2	228.2 (6.7)	2	255.1 (62.3)	3	963.0 (23.5)	4	23140.7 (483.9)	5	63627.1 (992.0)
	2	2	119.0 (21.9)	3	530.1 (8.4)	3	506.6 (90.3)	**1**	**64.0 (0.8)**	5	44703.3 (631.2)	4	21993.3 (374.0)
	3	**1**	**214.9 (15.2)**	4	1560.1 (11.6)	3	1121.7 (23.6)	2	532.1 (32.8)	6	37365.1 (484.6)	5	6101.7 (264.1)
S_3	1	**1**	**27.4 (1.2)**	2	241.6 (5.7)	2	234.3 (34.3)	3	999.4 (26.7)	4	23757.2 (417.3)	5	64316.7 (833.9)
	2	2	146.3 (25.5)	3	482.2 (7.3)	3	469.7 (47.3)	**1**	**64.4 (1.0)**	5	45230.8 (449.2)	4	22230.5 (270.5)
	3	**1**	**193.0 (12.6)**	4	1561.8 (16.7)	3	1104.5 (19.1)	2	613.4 (35.1)	6	37711.7 (325.0)	5	6126.7 (242.3)
S_4	1	**1**	**27.5 (1.2)**	3	286.2 (32.2)	2	244.0 (26.8)	4	1025.4 (32.4)	5	24266.3 (396.3)	6	64923.8 (763.5)
	2	2	149.5 (13.0)	4	600.2 (78.7)	3	484.2 (36.0)	**1**	**65.0 (1.3)**	6	46878.5 (564.7)	5	22415.4 (268.5)
	3	**1**	**173.4 (9.7)**	4	1694.5 (92.2)	3	1103.5 (18.2)	2	720.8 (34.9)	6	38262.6 (292.4)	5	5925.8 (151.5)
S_5	1	**1**	**25.7 (0.9)**	3	260.8 (22.6)	2	219.3 (18.8)	4	1047.5 (43.4)	5	24695.6 (435.9)	6	65453.6 (836.2)
	2	2	153.5 (14.8)	4	771.1 (116.8)	3	445.9 (25.8)	**1**	**65.4 (1.8)**	6	47293.9 (654.2)	5	22805.4 (200.4)
	3	**1**	**161.2 (10.1)**	4	1505.2 (94.3)	3	1085.9 (25.7)	2	771.7 (32.3)	6	38455.1 (325.3)	5	5867.1 (125.6)
Average r		**1.3**		3.4		2.7		2.2		5.4		5.1	

Table 7: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for ImageMagick.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL_{RF}		RF+		SeMPL_{SPL}		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	2	5.3 (0.3)	**1**	**3.7 (0.0)**	4	9.9 (1.3)	3	6.0 (0.2)	6	27.3 (0.8)	5	25.7 (0.7)
	2	**1**	**6.0 (0.2)**	3	13.8 (0.0)	4	15.5 (1.2)	2	8.3 (0.2)	5	27.0 (0.9)	5	26.1 (0.7)
	3	2	6.9 (0.7)	**1**	**4.7 (0.0)**	3	10.1 (1.2)	2	6.8 (0.3)	5	27.6 (0.8)	4	25.6 (0.7)
	4	**1**	**7.3 (0.6)**	3	14.0 (0.0)	4	15.1 (1.3)	2	8.5 (0.4)	6	32.0 (0.8)	5	25.3 (0.6)
S_2	1	2	4.5 (0.1)	**1**	**3.7 (0.0)**	4	5.6 (0.2)	3	5.0 (0.2)	5	25.1 (0.7)	5	25.1 (0.6)
	2	**1**	**5.6 (0.2)**	4	13.7 (0.1)	3	11.9 (0.3)	2	7.1 (0.2)	5	24.8 (0.8)	6	26.1 (0.7)
	3	2	5.4 (0.3)	**1**	**4.6 (0.0)**	4	6.4 (0.1)	3	5.8 (0.2)	5	25.5 (0.7)	5	25.0 (0.6)
	4	**1**	**5.7 (0.3)**	4	14.0 (0.1)	3	11.7 (0.2)	2	7.5 (0.3)	6	30.2 (0.8)	5	24.9 (0.6)
S_3	1	2	4.2 (0.1)	**1**	**3.8 (0.1)**	3	4.5 (0.1)	4	5.3 (0.2)	5	24.6 (0.6)	5	24.9 (0.6)
	2	**1**	**5.0 (0.2)**	4	13.4 (0.1)	3	11.3 (0.1)	2	7.3 (0.2)	5	24.1 (0.6)	6	26.2 (0.7)
	3	**1**	**4.2 (0.1)**	2	4.6 (0.1)	3	5.5 (0.1)	4	6.1 (0.2)	5	25.0 (0.6)	5	24.8 (0.6)
	4	**1**	**4.4 (0.2)**	4	13.7 (0.1)	3	11.2 (0.2)	2	6.7 (0.2)	6	30.0 (0.7)	5	24.9 (0.5)
S_4	1	2	3.7 (0.1)	**1**	**3.6 (0.1)**	3	4.0 (0.1)	4	5.4 (0.2)	5	24.2 (0.7)	5	24.3 (0.7)
	2	**1**	**4.4 (0.1)**	4	13.2 (0.1)	3	11.0 (0.1)	2	7.5 (0.2)	5	24.5 (0.8)	6	25.4 (0.8)
	3	**1**	**3.8 (0.1)**	2	4.4 (0.1)	3	4.9 (0.1)	4	6.1 (0.2)	5	24.4 (0.7)	5	24.4 (0.7)
	4	**1**	**3.9 (0.2)**	4	13.6 (0.1)	3	11.4 (0.1)	2	6.3 (0.2)	6	29.3 (0.8)	5	24.1 (0.6)
S_5	1	**1**	**3.7 (0.1)**	**1**	**3.6 (0.1)**	2	3.8 (0.1)	3	5.4 (0.2)	4	24.2 (0.8)	4	24.2 (0.7)
	2	**1**	**4.3 (0.1)**	4	13.1 (0.2)	3	11.0 (0.1)	2	7.6 (0.3)	5	24.8 (0.8)	5	25.4 (0.8)
	3	**1**	**3.8 (0.1)**	2	4.4 (0.1)	3	4.8 (0.1)	4	6.1 (0.2)	5	24.4 (0.8)	5	24.3 (0.8)
	4	**1**	**3.9 (0.2)**	4	13.6 (0.2)	3	11.4 (0.1)	2	6.2 (0.2)	6	29.2 (0.8)	5	24.2 (0.6)
Average r		**1.3**		2.5		3.2		2.7		5.2		5	

Table 8: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for ExaStencils.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL_{RF}		RF+		SeMPL_{SPL}		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**3.1 (0.0)**	3	16.8 (0.3)	2	9.0 (0.3)	2	8.7 (0.9)	4	18.2 (0.2)	5	18.9 (0.2)
	2	**1**	**4.3 (0.5)**	3	18.7 (0.7)	2	8.9 (0.3)	2	9.4 (1.1)	3	18.3 (0.2)	3	18.3 (0.2)
	3	**1**	**4.4 (0.4)**	3	16.5 (0.6)	2	8.8 (0.3)	2	9.5 (1.0)	4	17.4 (0.2)	3	16.6 (0.2)
	4	**1**	**3.4 (0.3)**	5	16.7 (0.6)	2	8.1 (0.3)	3	11.2 (0.8)	5	17.1 (0.2)	4	16.2 (0.2)
S_2	1	**1**	**2.7 (0.0)**	4	17.5 (0.5)	2	6.8 (0.2)	3	7.6 (0.9)	4	17.9 (0.1)	5	18.7 (0.1)
	2	**1**	**2.9 (0.2)**	4	18.4 (0.6)	2	6.7 (0.2)	3	7.5 (1.0)	4	18.0 (0.1)	4	18.1 (0.1)
	3	**1**	**3.0 (0.1)**	4	15.8 (0.4)	2	6.7 (0.1)	3	7.8 (0.9)	6	17.0 (0.1)	5	16.3 (0.1)
	4	**1**	**2.5 (0.1)**	5	16.9 (0.7)	2	6.2 (0.1)	3	9.8 (0.7)	5	16.7 (0.2)	4	15.9 (0.1)
S_3	1	**1**	**2.3 (0.0)**	4	16.8 (0.3)	3	5.0 (0.1)	2	4.7 (0.3)	5	17.7 (0.1)	6	18.6 (0.1)
	2	**1**	**2.3 (0.0)**	5	18.8 (0.9)	2	4.9 (0.1)	2	4.8 (0.5)	3	17.9 (0.1)	4	18.0 (0.1)
	3	**1**	**2.3 (0.0)**	3	16.2 (0.6)	2	5.0 (0.1)	2	5.0 (0.5)	4	16.8 (0.1)	3	16.2 (0.1)
	4	**1**	**1.9 (0.0)**	5	16.9 (0.6)	2	4.7 (0.1)	3	7.4 (0.5)	5	16.6 (0.1)	4	15.8 (0.1)
S_4	1	**1**	**2.0 (0.0)**	4	17.9 (0.5)	3	4.5 (0.0)	2	4.3 (0.3)	4	17.7 (0.1)	5	18.5 (0.1)
	2	**1**	**2.0 (0.0)**	3	17.7 (0.4)	2	4.4 (0.1)	2	4.2 (0.3)	3	17.8 (0.1)	3	17.9 (0.1)
	3	**1**	**2.0 (0.0)**	3	15.8 (0.3)	2	4.5 (0.0)	2	4.5 (0.4)	5	16.7 (0.1)	4	16.2 (0.1)
	4	**1**	**1.4 (0.0)**	5	16.7 (0.6)	2	4.2 (0.0)	3	7.1 (0.4)	5	16.5 (0.1)	4	15.8 (0.1)
S_5	1	**1**	**1.8 (0.0)**	4	17.6 (0.5)	3	4.0 (0.0)	2	3.8 (0.2)	4	17.6 (0.1)	5	18.4 (0.1)
	2	**1**	**1.4 (0.0)**	4	17.1 (0.6)	2	3.8 (0.0)	3	3.9 (0.1)	5	17.8 (0.1)	5	17.8 (0.1)
	3	**1**	**1.1 (0.1)**	4	16.8 (0.5)	2	3.9 (0.0)	2	4.0 (0.4)	4	16.7 (0.1)	3	16.1 (0.1)
	4	**1**	**1.4 (0.0)**	4	15.5 (0.5)	2	3.7 (0.0)	3	6.7 (0.4)	5	16.5 (0.1)	4	15.7 (0.1)
Average r		**1.0**		4		2.1		2.5		4.3		4.2	

Table 9: The Scott-Knott rank (r), mean MRE, and standard error (SEM) of all target-tasks and training sizes for x264.

Size	$\mathcal{T}_{target}$	SeMPL		DeepPerf+		SeMPL$_{RF}$		RF+		SeMPL$_{SPL}$		SPLConqueror+	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**26.5 (0.8)**	4	200.5 (0.4)	5	254.9 (9.0)	2	95.6 (6.8)	6	317.7 (12.8)	3	122.6 (7.9)
	2	**1**	**16.3 (0.8)**	2	24.2 (0.1)	4	36.8 (2.0)	3	31.4 (1.9)	5	74.0 (3.1)	6	108.8 (5.7)
	3	**1**	**19.3 (0.9)**	2	24.3 (0.1)	3	37.7 (2.1)	2	23.9 (2.2)	4	74.6 (3.1)	5	107.4 (5.4)
	4	2	25.1 (1.4)	**1**	**12.8 (0.1)**	4	38.4 (2.7)	3	35.3 (1.9)	5	81.2 (3.8)	6	112.6 (6.5)
	5	**1**	**17.6 (0.9)**	3	31.3 (0.1)	4	67.4 (3.9)	2	28.6 (2.6)	6	115.4 (5.1)	5	80.5 (4.6)
	6	2	15.3 (1.2)	**1**	**11.6 (0.0)**	4	37.9 (2.6)	3	18.7 (2.5)	5	77.7 (3.4)	5	75.9 (3.2)
	7	2	19.6 (1.3)	**1**	**10.9 (0.1)**	3	40.4 (3.0)	2	18.3 (3.0)	5	86.5 (4.1)	4	67.2 (3.4)
	8	**1**	**11.2 (0.5)**	4	85.5 (0.5)	5	108.0 (5.9)	3	65.8 (0.7)	6	152.7 (6.7)	2	22.9 (0.7)
	9	**1**	**19.3 (1.1)**	2	27.2 (0.0)	3	39.1 (1.9)	**1**	**20.7 (2.3)**	4	70.3 (2.7)	5	100.1 (4.9)
	10	**1**	**14.4 (1.1)**	2	15.7 (0.1)	4	51.7 (3.5)	3	20.2 (2.6)	6	102.7 (4.7)	5	92.5 (4.8)
S_2	1	**1**	**21.5 (0.4)**	4	206.5 (5.0)	5	229.9 (5.8)	3	156.5 (8.1)	6	300.8 (7.5)	2	114.4 (5.3)
	2	**1**	**14.2 (0.5)**	2	21.3 (0.5)	3	29.8 (1.7)	4	40.6 (2.6)	5	71.5 (2.1)	6	94.5 (3.3)
	3	**1**	**15.7 (0.6)**	2	22.0 (0.8)	4	31.0 (1.8)	3	25.9 (1.9)	5	72.6 (2.0)	6	93.2 (3.0)
	4	2	19.0 (0.7)	**1**	**14.9 (0.5)**	3	27.4 (2.0)	4	35.7 (0.9)	5	78.9 (2.4)	6	99.1 (2.7)
	5	**1**	**13.4 (0.4)**	3	38.1 (1.8)	4	55.3 (3.2)	2	24.4 (1.2)	6	113.4 (2.8)	5	72.9 (2.8)
	6	**1**	**10.5 (0.2)**	3	14.3 (1.1)	4	28.6 (2.2)	2	12.0 (0.3)	6	75.7 (2.1)	5	66.6 (1.9)
	7	2	13.2 (0.3)	3	14.5 (0.7)	4	28.0 (2.0)	**1**	**9.4 (0.3)**	6	83.3 (2.5)	5	59.1 (1.5)
	8	**1**	**8.8 (0.3)**	4	89.4 (3.3)	5	96.8 (3.6)	3	64.3 (0.2)	6	152.8 (3.3)	2	21.0 (0.3)
	9	**1**	**13.4 (0.4)**	3	26.3 (0.7)	4	32.4 (1.6)	2	17.4 (0.7)	5	68.4 (2.1)	6	87.2 (2.3)
	10	**1**	**11.4 (0.3)**	3	22.8 (1.3)	4	39.3 (2.9)	2	17.4 (0.8)	6	99.4 (2.5)	5	81.3 (2.3)
S_3	1	**1**	**21.1 (0.5)**	4	209.1 (5.6)	4	211.7 (4.7)	3	180.3 (8.2)	5	284.3 (7.6)	2	109.5 (3.7)
	2	**1**	**13.2 (0.4)**	2	20.4 (0.7)	3	26.1 (1.2)	4	46.2 (2.7)	5	70.3 (1.8)	6	90.1 (2.1)
	3	**1**	**13.7 (0.2)**	2	21.5 (0.7)	3	27.0 (1.3)	4	29.9 (2.0)	5	71.1 (1.8)	6	88.1 (2.1)
	4	2	17.9 (0.6)	**1**	**16.1 (0.7)**	3	21.9 (1.5)	4	38.7 (0.6)	5	76.4 (1.9)	6	95.0 (2.3)
	5	**1**	**12.4 (0.4)**	3	39.6 (1.0)	4	45.4 (2.8)	2	26.2 (1.3)	6	107.1 (1.9)	5	70.2 (1.9)
	6	**1**	**10.2 (0.3)**	3	14.7 (0.9)	4	22.8 (1.7)	2	11.3 (0.2)	6	72.9 (1.5)	5	63.4 (1.2)
	7	2	12.4 (0.3)	3	16.9 (0.7)	4	21.7 (1.5)	**1**	**9.0 (0.2)**	6	79.4 (1.7)	5	56.0 (1.1)
	8	**1**	**8.0 (0.2)**	5	92.7 (2.9)	4	86.9 (3.1)	3	64.2 (0.2)	6	148.2 (2.6)	2	20.2 (0.2)
	9	**1**	**12.9 (0.4)**	3	24.4 (1.0)	4	29.1 (1.1)	2	18.8 (0.7)	5	66.0 (1.6)	6	83.6 (1.7)
	10	**1**	**10.2 (0.3)**	3	22.9 (0.8)	4	30.5 (2.3)	2	18.6 (0.8)	6	94.7 (1.9)	5	76.1 (1.4)
S_4	1	**1**	**20.8 (0.5)**	4	204.0 (4.3)	3	198.8 (3.2)	4	203.8 (7.4)	5	272.1 (7.2)	2	106.5 (2.9)
	2	**1**	**13.0 (0.3)**	2	20.3 (0.8)	3	23.3 (0.8)	4	50.9 (2.3)	5	65.0 (1.2)	6	86.2 (1.7)
	3	**1**	**12.6 (0.3)**	2	19.6 (0.5)	3	24.0 (0.8)	4	33.2 (1.9)	5	65.8 (1.1)	6	84.1 (1.8)
	4	**1**	**16.0 (0.4)**	**1**	**16.1 (0.9)**	2	18.0 (1.1)	3	39.6 (0.5)	4	71.9 (1.5)	5	90.6 (1.7)
	5	**1**	**11.7 (0.3)**	3	39.9 (1.5)	3	38.9 (2.6)	2	27.5 (1.4)	5	103.8 (1.8)	4	68.7 (1.7)
	6	**1**	**9.8 (0.3)**	3	13.3 (0.6)	4	18.5 (1.2)	2	10.5 (0.2)	6	69.3 (1.3)	5	61.1 (1.1)
	7	2	11.6 (0.3)	3	18.6 (0.8)	3	18.2 (1.4)	**1**	**8.8 (0.2)**	5	75.4 (1.5)	4	54.4 (1.0)
	8	**1**	**7.3 (0.1)**	5	88.6 (3.2)	4	83.5 (3.6)	3	64.0 (0.3)	6	145.9 (2.2)	2	19.5 (0.2)
	9	**1**	**12.1 (0.3)**	3	22.8 (1.1)	4	26.6 (0.6)	2	20.0 (0.5)	5	63.2 (1.1)	6	80.0 (1.5)
	10	**1**	**9.5 (0.2)**	3	25.3 (1.4)	3	24.0 (1.9)	2	19.8 (0.7)	5	91.0 (1.8)	4	73.0 (1.3)
S_5	1	**1**	**21.3 (0.7)**	4	206.0 (4.6)	3	193.0 (2.7)	4	207.0 (6.3)	5	266.1 (7.9)	2	105.4 (2.4)
	2	**1**	**12.6 (0.4)**	2	20.6 (0.9)	3	23.2 (0.6)	4	52.3 (2.2)	5	64.7 (1.2)	6	85.4 (1.6)
	3	**1**	**12.2 (0.3)**	2	20.8 (0.7)	3	23.7 (0.6)	4	34.9 (1.8)	5	65.4 (1.2)	6	83.7 (1.7)
	4	**1**	**15.8 (0.4)**	**1**	**16.1 (0.9)**	2	17.0 (0.6)	3	39.7 (0.5)	4	70.4 (1.5)	5	87.3 (1.5)
	5	**1**	**11.1 (0.3)**	4	38.8 (1.2)	3	36.7 (1.7)	2	28.2 (1.2)	6	101.6 (1.9)	5	68.0 (1.4)
	6	**1**	**9.8 (0.3)**	3	14.7 (0.9)	4	17.4 (0.8)	2	10.5 (0.2)	6	67.2 (1.2)	5	60.4 (1.0)
	7	2	10.9 (0.2)	3	17.2 (0.6)	3	17.2 (0.7)	**1**	**8.8 (0.2)**	5	72.9 (1.5)	4	53.2 (0.9)
	8	**1**	**7.1 (0.2)**	5	94.6 (3.7)	4	83.3 (2.1)	3	63.7 (0.4)	6	146.9 (2.4)	2	19.3 (0.3)
	9	**1**	**12.1 (0.4)**	3	23.1 (1.3)	4	26.3 (0.5)	2	20.3 (0.5)	5	62.1 (1.1)	6	77.8 (1.3)
	10	**1**	**9.2 (0.3)**	3	23.6 (1.3)	3	22.3 (1.2)	2	20.2 (0.6)	5	88.6 (1.8)	4	71.6 (1.1)
Average r		**1.2**		2.8		3.6		2.6		5.3		4.6	