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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**82.5 (5.4)**	3	164.8 (9.8)	5	229.7 (14.2)	4	180.6 (12.3)	2	151.3 (6.8)	2	147.9 (8.0)
	2	**1**	**76.6 (5.8)**	3	217.3 (17.3)	5	339.0 (25.1)	4	245.8 (16.5)	3	208.4 (9.7)	2	197.1 (10.6)
	3	**1**	**60.4 (3.8)**	2	151.7 (14.3)	4	238.9 (16.3)	3	179.7 (12.4)	2	151.5 (6.8)	2	138.7 (8.5)
S_2	1	**1**	**62.7 (3.6)**	2	104.2 (9.8)	3	145.2 (7.7)	2	107.6 (7.8)	2	106.9 (4.5)	2	102.3 (5.2)
	2	**1**	**74.5 (4.9)**	2	145.1 (20.2)	4	215.5 (9.9)	3	162.6 (11.7)	3	172.7 (7.2)	2	147.6 (7.2)
	3	**1**	**50.8 (2.7)**	2	83.4 (7.5)	5	149.3 (7.7)	4	114.4 (9.6)	4	121.7 (5.2)	3	95.2 (4.8)
S_3	1	**1**	**50.2 (2.3)**	2	78.2 (6.2)	4	95.9 (4.3)	2	73.8 (5.2)	4	97.1 (2.7)	3	84.6 (2.6)
	2	**1**	**66.4 (5.0)**	2	101.0 (5.4)	4	157.1 (6.2)	3	118.2 (6.5)	5	175.6 (7.5)	3	116.6 (3.6)
	3	**1**	**45.4 (3.1)**	2	65.3 (8.6)	4	101.1 (4.8)	3	81.1 (5.9)	5	121.5 (5.4)	3	76.7 (2.4)
S_4	1	**1**	**46.9 (2.6)**	3	66.1 (2.6)	4	72.9 (3.0)	2	61.6 (4.9)	6	94.6 (2.3)	5	79.9 (1.9)
	2	**1**	**62.6 (5.7)**	2	84.4 (3.5)	4	116.0 (4.9)	3	100.2 (6.3)	5	181.7 (6.6)	3	102.6 (3.1)
	3	**1**	**40.1 (1.6)**	2	50.2 (1.8)	4	75.9 (4.0)	3	68.3 (5.6)	5	124.0 (4.8)	3	65.6 (2.1)
S_5	1	**1**	**42.3 (1.9)**	4	75.1 (10.7)	3	62.6 (2.6)	2	49.4 (3.5)	5	91.0 (1.6)	4	77.0 (2.0)
	2	**1**	**55.1 (3.4)**	3	96.7 (13.2)	3	101.9 (4.5)	2	84.2 (4.2)	4	173.1 (4.9)	3	99.5 (2.5)
	3	**1**	**35.9 (1.1)**	2	43.2 (1.5)	4	61.7 (2.9)	3	48.3 (2.7)	5	116.6 (3.5)	4	62.9 (1.6)
Average r		**1.0**		2.4		4		2.9		4		2.9	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	2	1353.0 (191.7)	**1**	**1074.7 (142.0)**	4	2354.1 (239.5)	3	1784.2 (289.8)	3	1719.7 (174.6)	4	2328.5 (188.7)
	2	2	105.4 (8.4)	2	110.6 (22.4)	2	99.3 (7.6)	**1**	**75.7 (10.7)**	4	409.2 (20.9)	3	194.7 (9.6)
	3	**1**	**41.9 (6.7)**	3	74.6 (10.6)	2	54.2 (8.0)	2	51.4 (12.4)	5	245.4 (12.3)	4	103.1 (8.4)
S_2	1	**1**	**754.1 (143.1)**	**1**	**723.7 (109.5)**	3	1781.2 (203.5)	2	1485.5 (230.4)	2	1373.0 (96.5)	3	1920.2 (198.1)
	2	2	59.1 (4.0)	2	63.2 (12.4)	2	57.8 (3.6)	**1**	**36.3 (4.5)**	4	288.3 (13.7)	3	196.6 (5.5)
	3	2	26.9 (2.4)	4	37.5 (3.3)	3	31.4 (2.7)	**1**	**23.9 (2.8)**	6	151.2 (5.8)	5	99.9 (4.7)
S_3	1	**1**	**438.6 (90.4)**	**1**	**495.4 (86.4)**	3	1388.4 (122.0)	2	1040.7 (136.5)	3	1422.6 (86.1)	4	1646.7 (161.4)
	2	**1**	**30.8 (2.3)**	2	39.6 (4.5)	2	35.7 (2.3)	**1**	**31.9 (2.0)**	4	239.3 (7.5)	3	196.5 (4.8)
	3	**1**	**20.9 (1.1)**	3	33.3 (2.8)	2	22.8 (1.6)	**1**	**21.5 (1.4)**	5	134.0 (5.2)	4	99.7 (3.9)
S_4	1	**1**	**287.3 (72.9)**	2	617.5 (147.8)	4	1297.0 (111.9)	3	1110.1 (125.0)	5	1471.2 (84.5)	6	1618.2 (146.0)
	2	3	32.8 (2.2)	3	33.5 (3.3)	2	26.2 (2.0)	**1**	**21.0 (2.7)**	5	220.3 (7.0)	4	194.9 (4.5)
	3	**1**	**17.7 (0.9)**	2	26.9 (1.9)	**1**	**18.1 (1.2)**	**1**	**18.2 (2.0)**	4	123.2 (4.5)	3	98.9 (3.2)
S_5	1	**1**	**234.4 (63.4)**	2	323.2 (89.3)	3	1147.5 (121.8)	3	1005.0 (142.2)	4	1393.8 (94.9)	4	1382.4 (120.3)
	2	3	30.4 (2.9)	3	28.6 (1.5)	2	21.4 (1.8)	**1**	**16.1 (2.5)**	4	200.4 (6.2)	4	197.5 (4.2)
	3	**1**	**15.9 (0.7)**	3	24.2 (1.7)	2	17.3 (1.1)	2	18.0 (2.0)	5	115.2 (3.1)	4	100.8 (2.9)
Average r		**1.5**		2.3		2.5		1.7		4.2		3.9	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		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)**	3	1.3 (0.1)	**1**	**1.0 (0.0)**	2	1.3 (0.0)	5	16.0 (0.3)	4	4.0 (0.1)
	2	**1**	**1.0 (0.0)**	2	1.2 (0.1)	**1**	**1.0 (0.0)**	2	1.1 (0.1)	4	16.0 (0.3)	3	3.9 (0.1)
	3	2	1.1 (0.1)	4	1.7 (0.4)	**1**	**1.0 (0.0)**	3	1.2 (0.0)	6	16.1 (0.3)	5	4.0 (0.1)
	4	2	1.2 (0.1)	4	1.9 (0.3)	**1**	**1.1 (0.0)**	3	1.4 (0.1)	6	16.1 (0.3)	5	4.0 (0.1)
S_2	1	**1**	**0.9 (0.0)**	2	1.2 (0.1)	**1**	**0.9 (0.0)**	2	1.2 (0.0)	4	9.9 (0.4)	3	3.5 (0.0)
	2	2	0.8 (0.0)	4	1.0 (0.1)	**1**	**0.7 (0.0)**	3	1.0 (0.0)	6	9.9 (0.4)	5	3.4 (0.0)
	3	**1**	**0.9 (0.0)**	3	1.3 (0.1)	**1**	**0.9 (0.0)**	2	1.2 (0.0)	5	9.9 (0.4)	4	3.5 (0.1)
	4	**1**	**1.0 (0.0)**	2	1.3 (0.0)	**1**	**1.0 (0.0)**	3	1.4 (0.0)	5	9.9 (0.4)	4	3.5 (0.0)
S_3	1	**1**	**0.9 (0.0)**	2	1.1 (0.1)	**1**	**0.9 (0.0)**	3	1.2 (0.0)	5	5.3 (0.3)	4	3.2 (0.0)
	2	**1**	**0.7 (0.0)**	2	0.9 (0.1)	**1**	**0.7 (0.0)**	2	0.9 (0.0)	4	5.3 (0.3)	3	3.2 (0.0)
	3	**1**	**0.9 (0.0)**	3	1.1 (0.1)	2	0.9 (0.0)	4	1.2 (0.0)	6	5.3 (0.3)	5	3.2 (0.0)
	4	**1**	**1.0 (0.0)**	2	1.4 (0.1)	**1**	**1.0 (0.0)**	3	1.4 (0.0)	5	5.4 (0.3)	4	3.2 (0.0)
S_4	1	**1**	**0.8 (0.0)**	3	1.1 (0.1)	2	0.9 (0.0)	4	1.2 (0.0)	5	3.2 (0.2)	5	3.1 (0.0)
	2	**1**	**0.7 (0.0)**	2	0.8 (0.0)	**1**	**0.7 (0.0)**	3	0.9 (0.0)	4	3.1 (0.2)	4	3.1 (0.0)
	3	**1**	**0.9 (0.0)**	3	1.1 (0.1)	2	0.9 (0.0)	4	1.2 (0.0)	5	3.2 (0.2)	5	3.1 (0.0)
	4	**1**	**1.0 (0.0)**	3	1.3 (0.1)	2	1.0 (0.0)	4	1.4 (0.0)	6	3.3 (0.2)	5	3.1 (0.0)
S_5	1	**1**	**0.8 (0.0)**	3	1.0 (0.1)	2	0.9 (0.0)	4	1.1 (0.0)	5	2.0 (0.1)	6	3.0 (0.0)
	2	**1**	**0.7 (0.0)**	3	0.7 (0.0)	2	0.7 (0.0)	4	0.9 (0.0)	5	2.0 (0.1)	6	3.0 (0.0)
	3	**1**	**0.9 (0.0)**	3	1.0 (0.0)	2	0.9 (0.0)	4	1.2 (0.0)	5	2.1 (0.1)	6	3.0 (0.0)
	4	**1**	**1.0 (0.0)**	3	1.2 (0.1)	2	1.0 (0.0)	4	1.4 (0.0)	5	2.1 (0.1)	6	3.1 (0.0)
Average r		**1.1**		2.8		1.4		3.1		5		4.6	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**130.2 (23.3)**	2	262.8 (44.2)	4	503.3 (34.5)	3	369.6 (54.1)	5	619.3 (11.4)	2	275.1 (51.9)
	2	**1**	**186.4 (34.2)**	2	284.2 (62.3)	4	502.7 (34.2)	3	340.4 (52.1)	5	620.2 (11.5)	2	275.4 (51.8)
	3	3	313.4 (48.7)	5	423.8 (65.0)	4	359.2 (25.8)	2	253.1 (40.7)	6	483.2 (12.5)	**1**	**186.7 (34.1)**
	4	3	252.7 (51.3)	4	371.4 (49.3)	3	294.8 (23.2)	2	211.5 (35.5)	5	429.2 (14.6)	**1**	**148.7 (25.2)**
S_2	1	**1**	**26.4 (3.2)**	2	64.4 (25.4)	3	168.4 (28.7)	3	139.4 (42.5)	4	633.9 (7.8)	2	51.3 (27.8)
	2	**1**	**11.7 (1.2)**	2	46.5 (18.0)	4	167.9 (28.7)	3	122.1 (39.3)	5	635.1 (7.9)	2	51.0 (27.8)
	3	**1**	**31.8 (20.7)**	2	62.8 (26.1)	3	118.0 (18.9)	3	110.2 (33.8)	4	501.8 (9.1)	**1**	**38.6 (18.2)**
	4	**1**	**36.9 (20.7)**	**1**	**63.2 (27.9)**	2	98.4 (14.4)	2	97.7 (30.0)	3	462.7 (12.3)	**1**	**38.0 (13.4)**
S_3	1	3	17.0 (1.4)	2	13.4 (1.3)	4	74.2 (14.2)	4	52.0 (27.4)	5	635.6 (7.6)	**1**	**8.4 (0.4)**
	2	**1**	**9.1 (1.2)**	2	29.7 (8.8)	3	74.4 (14.1)	3	53.6 (28.7)	4	636.4 (7.7)	**1**	**8.3 (0.4)**
	3	**1**	**5.4 (0.5)**	3	12.6 (2.5)	4	52.2 (9.5)	4	52.0 (27.4)	5	504.0 (7.9)	2	9.2 (0.3)
	4	**1**	**6.7 (0.8)**	2	13.0 (1.7)	3	44.3 (7.2)	3	52.4 (27.1)	4	465.8 (10.8)	2	13.8 (0.4)
S_4	1	2	10.4 (0.7)	3	14.2 (1.9)	4	32.8 (10.1)	**1**	**8.1 (0.7)**	5	636.6 (7.6)	**1**	**7.6 (0.4)**
	2	**1**	**7.4 (0.6)**	2	11.0 (1.7)	3	32.9 (10.0)	**1**	**7.1 (0.7)**	4	637.4 (7.7)	**1**	**7.6 (0.4)**
	3	**1**	**4.6 (0.3)**	3	8.8 (0.9)	4	23.4 (6.7)	2	7.4 (0.5)	5	505.6 (7.7)	3	8.8 (0.3)
	4	**1**	**5.2 (0.3)**	3	12.1 (3.6)	4	21.5 (5.1)	2	9.0 (0.5)	5	467.5 (10.6)	3	13.5 (0.4)
S_5	1	**1**	**6.4 (0.6)**	3	9.5 (1.0)	3	9.7 (3.1)	**1**	**6.2 (0.5)**	4	629.8 (5.0)	2	6.8 (0.1)
	2	3	7.9 (1.6)	3	9.0 (0.6)	3	9.6 (3.1)	**1**	**5.5 (0.3)**	4	630.7 (5.0)	2	6.7 (0.1)
	3	**1**	**4.2 (0.3)**	2	6.0 (0.5)	3	7.8 (2.1)	3	7.6 (0.5)	5	503.8 (5.2)	4	8.2 (0.1)
	4	**1**	**4.9 (0.3)**	3	7.7 (0.6)	4	9.3 (1.7)	2	6.7 (0.5)	6	466.0 (6.9)	5	13.0 (0.3)
Average r		**1.4**		2.5		3.5		2.4		4.7		1.9	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**44.5 (8.8)**	3	89.5 (12.6)	2	69.7 (8.0)	**1**	**49.2 (3.1)**	4	244.6 (12.6)	**1**	**44.8 (2.1)**
	2	**1**	**59.3 (15.3)**	4	800.3 (242.1)	4	771.6 (134.3)	2	73.5 (3.0)	5	2609.3 (152.5)	3	192.8 (43.7)
	3	**1**	**34.1 (2.1)**	4	202.0 (96.8)	4	238.7 (40.1)	2	48.1 (2.8)	5	1013.3 (52.6)	3	71.9 (12.3)
S_2	1	**1**	**34.2 (1.3)**	3	47.2 (2.3)	2	42.4 (2.0)	3	45.3 (2.7)	4	86.1 (6.0)	2	42.3 (1.6)
	2	**1**	**30.1 (2.6)**	3	117.5 (38.0)	2	69.5 (3.8)	2	68.1 (2.4)	5	1234.0 (105.4)	4	155.4 (23.0)
	3	2	52.4 (33.0)	2	59.4 (9.6)	**1**	**43.2 (2.2)**	**1**	**43.6 (2.1)**	3	405.8 (43.8)	2	55.7 (5.6)
S_3	1	**1**	**40.9 (1.9)**	2	42.6 (2.0)	**1**	**38.9 (1.8)**	**1**	**40.5 (2.2)**	3	55.8 (3.8)	2	42.1 (1.5)
	2	**1**	**28.4 (3.4)**	3	130.1 (55.9)	2	64.8 (3.1)	2	62.6 (1.6)	4	868.2 (81.5)	3	131.4 (19.1)
	3	**1**	**19.0 (2.2)**	4	147.7 (96.1)	2	39.2 (1.7)	2	40.6 (1.4)	5	253.9 (32.3)	3	55.1 (5.3)
S_4	1	**1**	**37.7 (1.6)**	3	39.8 (1.9)	**1**	**37.2 (1.4)**	2	38.8 (1.4)	4	49.1 (2.8)	3	41.2 (1.3)
	2	**1**	**26.2 (2.7)**	3	91.1 (20.7)	2	60.2 (1.0)	2	59.5 (1.9)	5	751.2 (60.8)	4	131.3 (17.7)
	3	**1**	**21.1 (2.2)**	4	47.3 (3.5)	3	36.8 (1.0)	2	35.2 (1.1)	5	217.2 (25.3)	4	51.8 (4.7)
S_5	1	**1**	**32.6 (1.0)**	3	38.1 (1.2)	2	36.7 (1.2)	2	36.4 (1.7)	5	47.1 (2.2)	4	40.1 (1.2)
	2	**1**	**29.0 (3.1)**	3	69.2 (6.1)	2	58.2 (1.0)	2	57.9 (1.6)	5	673.2 (49.8)	4	102.8 (12.0)
	3	**1**	**14.6 (1.0)**	3	44.1 (3.2)	2	35.7 (0.8)	2	34.6 (1.3)	4	200.0 (21.5)	3	46.3 (3.5)
Average r		**1.1**		3.1		2.1		1.9		4.4		3	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	3	31.2 (1.7)	4	278.8 (70.2)	**1**	**11.4 (0.4)**	2	14.1 (0.5)	6	56472.1 (1238.4)	5	4119.9 (275.2)
	2	4	144.5 (29.1)	3	54.5 (13.6)	**1**	**8.4 (0.4)**	2	9.9 (0.5)	6	21531.1 (490.2)	5	4072.3 (157.6)
	3	3	251.3 (18.3)	4	618.8 (297.9)	2	197.8 (18.8)	**1**	**142.0 (21.3)**	6	4592.9 (271.7)	5	2295.2 (89.0)
S_2	1	3	31.4 (2.6)	4	2101.6 (1796.5)	**1**	**11.0 (0.3)**	2	13.9 (0.3)	6	58133.5 (989.4)	5	4152.4 (255.7)
	2	4	119.0 (21.9)	3	72.4 (15.2)	**1**	**8.0 (0.2)**	2	9.7 (0.3)	6	21900.5 (393.9)	5	4339.3 (113.1)
	3	2	214.9 (15.2)	3	888.1 (420.8)	**1**	**142.2 (19.0)**	**1**	**132.1 (31.1)**	5	4296.8 (170.0)	4	2108.8 (88.8)
S_3	1	3	27.4 (1.2)	4	323.8 (97.0)	**1**	**10.5 (0.1)**	2	13.4 (0.2)	6	59208.8 (861.9)	5	4518.0 (171.3)
	2	4	146.3 (25.5)	3	71.9 (17.2)	**1**	**7.7 (0.2)**	2	9.3 (0.2)	6	22043.2 (248.3)	5	4394.0 (91.2)
	3	3	193.0 (12.6)	4	460.1 (171.3)	2	109.4 (10.0)	**1**	**97.6 (9.1)**	6	4246.2 (158.3)	5	2065.3 (89.4)
S_4	1	3	27.5 (1.2)	4	286.0 (77.0)	**1**	**10.3 (0.1)**	2	13.4 (0.2)	6	59990.2 (761.8)	5	4378.3 (149.7)
	2	4	149.5 (13.0)	3	108.8 (27.8)	**1**	**7.6 (0.1)**	2	9.1 (0.2)	6	22264.3 (267.0)	5	4486.6 (83.0)
	3	3	173.4 (9.7)	4	1377.7 (485.8)	**1**	**99.9 (6.4)**	2	112.0 (10.5)	6	4238.4 (132.7)	5	1891.8 (51.2)
S_5	1	3	25.7 (0.9)	4	269.5 (88.0)	**1**	**10.2 (0.1)**	2	13.4 (0.2)	6	60627.2 (795.3)	5	4449.5 (132.6)
	2	4	153.5 (14.8)	3	83.1 (19.1)	**1**	**7.8 (0.1)**	2	9.5 (0.2)	6	22719.5 (204.0)	5	4449.6 (69.6)
	3	2	161.2 (10.1)	3	605.3 (304.5)	**1**	**89.7 (6.0)**	**1**	**94.5 (7.6)**	5	4268.0 (99.9)	4	1886.2 (61.9)
Average r		3.2		3.5		**1.1**		1.7		5.9		4.9	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)	r	MRE (SEM)
S_1	1	**1**	**5.3 (0.3)**	3	17.5 (5.0)	3	17.8 (1.2)	2	13.3 (2.3)	4	27.4 (1.1)	2	12.2 (1.7)
	2	**1**	**6.0 (0.2)**	2	11.3 (2.3)	3	20.0 (1.3)	2	11.6 (2.0)	4	29.5 (1.3)	2	12.3 (1.7)
	3	**1**	**6.9 (0.7)**	2	10.7 (0.8)	4	18.2 (1.1)	3	12.6 (1.7)	5	27.3 (1.1)	3	12.1 (1.7)
	4	**1**	**7.3 (0.6)**	2	10.2 (1.0)	4	16.6 (1.1)	3	13.1 (2.0)	5	26.8 (1.1)	3	12.2 (1.6)
S_2	1	**1**	**4.5 (0.1)**	2	6.1 (0.3)	2	6.1 (0.3)	3	6.7 (0.3)	5	22.5 (0.5)	4	7.1 (0.2)
	2	**1**	**5.6 (0.2)**	2	6.0 (0.2)	2	6.3 (0.3)	3	6.7 (0.3)	5	23.3 (0.5)	4	7.5 (0.2)
	3	**1**	**5.4 (0.3)**	2	6.2 (0.2)	3	6.9 (0.3)	4	7.4 (0.4)	5	22.6 (0.5)	4	7.5 (0.2)
	4	2	5.7 (0.3)	**1**	**5.3 (0.2)**	3	6.1 (0.3)	4	6.5 (0.3)	6	22.6 (0.4)	5	7.6 (0.2)
S_3	1	**1**	**4.2 (0.1)**	2	4.7 (0.2)	**1**	**4.3 (0.1)**	2	4.8 (0.1)	4	24.3 (0.5)	3	6.0 (0.2)
	2	2	5.0 (0.2)	2	5.0 (0.2)	**1**	**4.7 (0.1)**	3	5.3 (0.1)	5	25.7 (0.6)	4	6.5 (0.2)
	3	**1**	**4.2 (0.1)**	2	4.5 (0.1)	3	5.1 (0.2)	4	5.8 (0.2)	6	24.2 (0.5)	5	6.5 (0.2)
	4	**1**	**4.4 (0.2)**	**1**	**4.5 (0.1)**	**1**	**4.6 (0.1)**	2	5.2 (0.2)	4	24.4 (0.5)	3	6.6 (0.2)
S_4	1	**1**	**3.7 (0.1)**	2	4.0 (0.2)	**1**	**3.7 (0.1)**	3	4.4 (0.1)	5	23.9 (0.7)	4	5.6 (0.2)
	2	2	4.4 (0.1)	2	4.6 (0.1)	**1**	**3.9 (0.1)**	3	4.7 (0.1)	5	25.1 (0.8)	4	5.9 (0.2)
	3	**1**	**3.8 (0.1)**	**1**	**3.8 (0.1)**	2	4.4 (0.1)	3	4.7 (0.1)	5	23.9 (0.7)	4	6.1 (0.2)
	4	**1**	**3.9 (0.2)**	**1**	**4.0 (0.1)**	**1**	**4.0 (0.1)**	2	4.8 (0.1)	4	23.6 (0.6)	3	6.3 (0.2)
S_5	1	2	3.7 (0.1)	3	3.9 (0.2)	**1**	**3.5 (0.1)**	4	4.3 (0.1)	6	23.8 (0.7)	5	5.5 (0.2)
	2	2	4.3 (0.1)	3	4.5 (0.2)	**1**	**3.9 (0.1)**	4	4.7 (0.1)	6	25.2 (0.8)	5	5.8 (0.2)
	3	2	3.8 (0.1)	**1**	**3.6 (0.1)**	3	4.3 (0.1)	4	4.7 (0.2)	6	23.8 (0.8)	5	6.3 (0.2)
	4	**1**	**3.9 (0.2)**	2	4.1 (0.2)	**1**	**3.9 (0.1)**	3	4.5 (0.1)	5	23.7 (0.6)	4	6.3 (0.2)
Average r		**1.3**		1.9		2		3		5		3.8	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		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)**	5	17.7 (0.7)	2	9.6 (0.3)	3	11.0 (0.3)	6	19.1 (0.2)	4	11.8 (0.2)
	2	**1**	**4.3 (0.5)**	4	16.6 (0.6)	2	9.4 (0.3)	3	11.2 (0.4)	5	18.4 (0.2)	3	11.4 (0.1)
	3	**1**	**4.4 (0.4)**	5	15.3 (0.6)	2	9.3 (0.3)	3	10.7 (0.4)	6	16.8 (0.2)	4	11.5 (0.1)
	4	**1**	**3.4 (0.3)**	5	14.9 (0.5)	2	8.5 (0.3)	3	9.4 (0.3)	6	16.4 (0.2)	4	10.7 (0.1)
S_2	1	**1**	**2.7 (0.0)**	4	10.7 (0.7)	2	6.7 (0.1)	3	7.8 (0.2)	6	18.7 (0.1)	5	11.3 (0.1)
	2	**1**	**2.9 (0.2)**	4	10.2 (0.7)	2	6.8 (0.1)	3	8.0 (0.2)	6	18.0 (0.1)	5	10.9 (0.1)
	3	**1**	**3.0 (0.1)**	4	9.1 (0.5)	2	6.8 (0.1)	3	8.4 (0.2)	6	16.4 (0.1)	5	10.8 (0.1)
	4	**1**	**2.5 (0.1)**	4	9.0 (0.6)	2	6.0 (0.1)	3	7.6 (0.2)	6	16.0 (0.1)	5	10.3 (0.1)
S_3	1	**1**	**2.3 (0.0)**	3	4.6 (0.3)	2	4.2 (0.1)	4	5.2 (0.1)	6	18.5 (0.1)	5	11.0 (0.1)
	2	**1**	**2.3 (0.0)**	2	4.7 (0.3)	2	4.4 (0.1)	3	5.2 (0.1)	5	17.9 (0.1)	4	10.5 (0.1)
	3	**1**	**2.3 (0.0)**	3	5.2 (0.3)	2	4.5 (0.1)	4	5.5 (0.1)	6	16.3 (0.1)	5	10.5 (0.1)
	4	**1**	**1.9 (0.0)**	2	4.0 (0.1)	2	4.1 (0.1)	3	5.0 (0.1)	5	15.9 (0.1)	4	9.9 (0.1)
S_4	1	**1**	**2.0 (0.0)**	3	3.6 (0.1)	2	3.5 (0.0)	4	4.2 (0.1)	6	18.4 (0.1)	5	10.9 (0.1)
	2	**1**	**2.0 (0.0)**	3	3.8 (0.2)	2	3.6 (0.0)	4	4.2 (0.1)	6	17.8 (0.1)	5	10.4 (0.1)
	3	**1**	**2.0 (0.0)**	2	3.9 (0.3)	2	3.8 (0.1)	3	4.6 (0.1)	5	16.2 (0.1)	4	10.5 (0.1)
	4	**1**	**1.4 (0.0)**	2	3.5 (0.2)	2	3.4 (0.0)	3	4.3 (0.1)	5	15.8 (0.1)	4	9.9 (0.1)
S_5	1	**1**	**1.8 (0.0)**	2	2.7 (0.1)	3	2.8 (0.0)	4	3.4 (0.1)	6	18.3 (0.1)	5	10.7 (0.1)
	2	**1**	**1.4 (0.0)**	3	3.0 (0.1)	2	2.8 (0.0)	4	3.3 (0.1)	6	17.7 (0.1)	5	10.3 (0.1)
	3	**1**	**1.1 (0.1)**	2	3.0 (0.1)	2	3.0 (0.0)	3	3.7 (0.1)	5	16.2 (0.1)	4	10.3 (0.1)
	4	**1**	**1.4 (0.0)**	2	2.5 (0.0)	3	2.8 (0.0)	4	3.5 (0.1)	6	15.8 (0.1)	5	9.8 (0.0)
Average r		**1.0**		3.2		2.1		3.4		5.7		4.5	

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

Size	$\mathcal{E}_{target}$	SeMPL		DeepPerf		RF		DECART		SPLConqueror		XGBoost	
		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)**	3	55.9 (7.4)	3	52.2 (2.5)	3	54.7 (4.2)	4	128.1 (8.0)	2	45.1 (2.7)
	2	**1**	**16.3 (0.8)**	2	39.8 (2.7)	3	50.2 (2.6)	3	53.6 (4.1)	4	114.6 (6.0)	2	41.5 (2.5)
	3	**1**	**19.3 (0.9)**	2	42.1 (3.2)	3	51.6 (2.5)	3	52.4 (4.4)	4	112.9 (5.8)	2	41.4 (2.4)
	4	**1**	**25.1 (1.4)**	2	44.2 (3.3)	3	58.8 (3.7)	3	58.7 (4.9)	4	116.5 (6.7)	2	46.3 (3.1)
	5	**1**	**17.6 (0.9)**	2	37.3 (3.2)	3	43.6 (2.2)	3	43.6 (3.2)	4	84.1 (4.9)	2	34.9 (2.0)
	6	**1**	**15.3 (1.2)**	2	31.2 (2.1)	3	40.6 (2.0)	3	41.0 (2.8)	4	80.3 (3.5)	2	33.3 (1.9)
	7	**1**	**19.6 (1.3)**	2	29.2 (2.0)	4	39.3 (2.1)	4	42.0 (3.1)	5	71.1 (3.5)	3	32.1 (1.6)
	8	**1**	**11.2 (0.5)**	2	14.1 (0.8)	3	16.7 (1.0)	4	18.1 (1.6)	5	23.5 (0.8)	2	15.0 (0.9)
	9	**1**	**19.3 (1.1)**	2	37.8 (2.7)	3	53.7 (3.1)	3	56.2 (5.4)	4	104.3 (5.0)	2	40.5 (2.3)
	10	**1**	**14.4 (1.1)**	2	39.1 (2.9)	4	50.1 (2.6)	3	43.9 (3.1)	5	97.3 (4.8)	2	37.4 (2.1)
S_2	1	**1**	**21.5 (0.4)**	2	26.3 (1.2)	4	39.6 (1.9)	4	40.6 (2.9)	5	122.1 (5.4)	3	34.1 (1.7)
	2	**1**	**14.2 (0.5)**	2	24.3 (1.9)	4	36.5 (1.9)	4	36.7 (2.7)	5	104.5 (3.6)	3	32.6 (1.8)
	3	**1**	**15.7 (0.6)**	2	28.3 (4.0)	4	37.2 (1.8)	4	36.1 (2.6)	5	102.0 (3.4)	3	34.4 (1.9)
	4	**1**	**19.0 (0.7)**	2	25.7 (1.6)	4	40.0 (2.4)	4	39.6 (3.5)	5	105.6 (2.6)	3	34.2 (2.7)
	5	**1**	**13.4 (0.4)**	2	25.0 (3.2)	4	34.6 (2.1)	4	34.8 (2.9)	5	78.3 (3.1)	3	30.4 (1.8)
	6	**1**	**10.5 (0.2)**	2	23.2 (2.7)	3	29.2 (1.6)	3	29.5 (2.2)	4	72.6 (1.9)	2	25.1 (1.6)
	7	**1**	**13.2 (0.3)**	2	20.2 (2.1)	4	26.0 (1.3)	5	28.7 (2.4)	6	64.3 (1.4)	3	23.4 (1.2)
	8	**1**	**8.8 (0.3)**	2	9.7 (0.4)	4	11.3 (0.6)	5	12.5 (0.7)	6	21.2 (0.4)	3	10.4 (0.4)
	9	**1**	**13.4 (0.4)**	2	22.7 (1.1)	4	35.3 (1.9)	4	36.2 (3.0)	5	92.9 (2.1)	3	30.4 (2.1)
	10	**1**	**11.4 (0.3)**	2	23.2 (0.8)	4	35.6 (2.1)	4	35.9 (3.4)	5	87.7 (2.2)	3	29.7 (2.2)
S_3	1	**1**	**21.1 (0.5)**	2	23.2 (0.8)	3	29.8 (1.5)	3	28.8 (2.1)	5	113.3 (3.6)	4	31.5 (1.6)
	2	**1**	**13.2 (0.4)**	2	26.6 (5.0)	2	26.8 (1.4)	2	29.8 (2.3)	3	97.6 (2.3)	2	28.5 (1.3)
	3	**1**	**13.7 (0.2)**	2	25.4 (3.2)	2	26.9 (1.3)	2	25.5 (1.8)	4	94.7 (2.3)	3	29.1 (1.4)
	4	**1**	**17.9 (0.6)**	2	23.8 (2.3)	3	27.1 (2.1)	3	30.2 (3.5)	4	99.7 (2.2)	3	28.1 (2.3)
	5	**1**	**12.4 (0.4)**	2	22.3 (3.3)	3	25.6 (1.8)	2	23.9 (3.0)	4	72.7 (1.9)	3	27.1 (1.5)
	6	**1**	**10.2 (0.3)**	2	17.5 (1.0)	3	21.9 (1.3)	3	23.4 (2.5)	4	68.3 (1.3)	3	22.3 (1.1)
	7	**1**	**12.4 (0.3)**	2	16.2 (1.0)	3	18.9 (1.1)	4	21.0 (1.6)	5	60.3 (1.1)	3	19.6 (1.1)
	8	**1**	**8.0 (0.2)**	2	8.5 (0.3)	3	8.9 (0.5)	4	9.8 (0.6)	5	20.0 (0.3)	3	8.9 (0.4)
	9	**1**	**12.9 (0.4)**	2	20.3 (1.1)	3	25.7 (1.6)	3	25.7 (1.8)	4	87.6 (1.7)	3	26.4 (1.6)
	10	**1**	**10.2 (0.3)**	2	20.2 (1.1)	3	24.8 (1.6)	3	25.1 (2.6)	4	81.7 (1.5)	3	26.2 (1.7)
S_4	1	**1**	**20.8 (0.5)**	**1**	**21.3 (0.5)**	2	23.7 (1.6)	2	23.7 (2.1)	4	108.6 (3.4)	3	28.1 (1.1)
	2	**1**	**13.0 (0.3)**	2	17.4 (0.8)	3	20.8 (1.5)	3	22.0 (2.0)	5	91.7 (2.1)	4	26.0 (1.1)
	3	**1**	**12.6 (0.3)**	2	18.0 (0.6)	3	20.6 (1.4)	3	22.2 (2.0)	5	89.7 (2.2)	4	26.3 (1.0)
	4	**1**	**16.0 (0.4)**	3	21.4 (0.9)	2	19.2 (1.1)	2	19.3 (1.0)	4	93.9 (1.8)	3	22.4 (1.3)
	5	**1**	**11.7 (0.3)**	2	15.8 (0.6)	3	21.2 (1.7)	3	20.3 (2.3)	5	69.3 (1.7)	4	24.1 (0.8)
	6	**1**	**9.8 (0.3)**	2	15.7 (0.8)	3	17.1 (1.2)	4	18.2 (1.6)	6	65.1 (1.2)	5	20.0 (1.0)
	7	**1**	**11.6 (0.3)**	2	15.1 (0.6)	2	14.6 (0.8)	3	17.0 (1.0)	4	57.3 (1.1)	3	17.5 (0.6)
	8	**1**	**7.3 (0.1)**	2	8.0 (0.2)	**1**	**7.1 (0.3)**	3	9.0 (0.5)	4	19.4 (0.3)	2	8.1 (0.3)
	9	**1**	**12.1 (0.3)**	2	19.3 (0.8)	2	19.4 (1.3)	3	23.5 (2.7)	4	82.7 (1.6)	3	22.2 (1.2)
	10	**1**	**9.5 (0.2)**	2	18.4 (0.9)	2	18.8 (1.3)	2	18.6 (1.7)	4	77.7 (1.5)	3	22.0 (1.0)
S_5	1	**1**	**21.3 (0.7)**	3	33.7 (8.7)	**1**	**21.1 (1.7)**	2	24.3 (3.0)	4	107.3 (2.5)	2	26.5 (1.0)
	2	**1**	**12.6 (0.4)**	4	23.7 (4.7)	2	17.7 (0.9)	3	21.4 (2.2)	5	90.3 (1.9)	4	24.1 (1.0)
	3	**1**	**12.2 (0.3)**	3	19.4 (0.6)	2	17.9 (0.9)	3	20.1 (2.2)	5	88.6 (1.8)	4	26.1 (0.8)
	4	**1**	**15.8 (0.4)**	3	20.4 (1.0)	**1**	**16.0 (0.8)**	2	18.5 (1.3)	4	90.9 (1.8)	3	21.0 (1.4)
	5	**1**	**11.1 (0.3)**	2	18.4 (0.8)	2	17.4 (0.9)	3	21.0 (2.5)	5	68.4 (1.3)	4	23.2 (1.0)
	6	**1**	**9.8 (0.3)**	3	15.9 (0.8)	2	14.3 (0.8)	3	15.8 (1.1)	5	63.8 (1.2)	4	17.9 (0.8)
	7	**1**	**10.9 (0.2)**	3	14.4 (0.6)	2	12.2 (0.6)	4	15.6 (0.8)	6	55.9 (1.1)	5	16.6 (0.8)
	8	2	7.1 (0.2)	4	8.5 (0.8)	**1**	**6.4 (0.2)**	4	8.1 (0.3)	5	19.2 (0.3)	3	7.7 (0.1)
	9	**1**	**12.1 (0.4)**	4	19.1 (0.8)	2	15.6 (0.8)	3	17.7 (0.9)	6	80.4 (1.6)	5	20.3 (1.1)
	10	**1**	**9.2 (0.3)**	3	19.1 (0.9)	2	15.6 (0.7)	2	16.0 (0.9)	5	75.9 (1.4)	4	20.4 (1.0)
Average r		**1.0**		2.3		2.8		3.2		4.6		3	