

Table 1: Accuracy (%) of linear probing (LP) over CLIP on CIFAR-10.

n	CD [1]	DeepFool [52]	Forgetting [69]	Glistner [38]	GraNd [56]	Herdning [80]	Entropy [15]	LeastConfidence [15]	Margin [15]	Uniform	SkMM-LP
1000	91.98 $\pm$ 0.11	91.68 $\pm$ 0.32	91.51 $\pm$ 0.01	91.09 $\pm$ 0.08	88.48 $\pm$ 0.90	91.68 $\pm$ 0.27	88.66 $\pm$ 1.01	89.38 $\pm$ 1.73	91.22 $\pm$ 0.64	91.68 $\pm$ 0.45	92.96 $\pm$ 0.07
2000	92.53 $\pm$ 0.05	91.78 $\pm$ 0.79	92.15 $\pm$ 0.02	91.60 $\pm$ 0.23	88.89 $\pm$ 0.53	92.20 $\pm$ 0.17	89.96 $\pm$ 0.42	90.49 $\pm$ 1.47	91.60 $\pm$ 0.92	92.22 $\pm$ 0.25	93.38 $\pm$ 0.01
3000	92.81 $\pm$ 0.25	91.93 $\pm$ 0.89	92.61 $\pm$ 0.01	91.83 $\pm$ 0.08	89.04 $\pm$ 0.24	92.72 $\pm$ 0.51	90.27 $\pm$ 1.11	90.83 $\pm$ 1.43	91.94 $\pm$ 0.83	92.60 $\pm$ 0.14	93.67 $\pm$ 0.01
4000	92.99 $\pm$ 0.24	92.18 $\pm$ 0.70	92.81 $\pm$ 0.01	91.87 $\pm$ 0.09	89.54 $\pm$ 0.46	93.00 $\pm$ 0.45	91.06 $\pm$ 0.75	91.26 $\pm$ 1.30	92.09 $\pm$ 0.81	92.79 $\pm$ 0.12	93.78 $\pm$ 0.04

Table 2: Accuracy (%) of FT over (the last two layers of) ResNet18 (FT2) on CIFAR-10.

n	CD [1]	DeepFool [52]	Forgetting [69]	Glistner [38]	GraNd [56]	Herdning [80]	Entropy [15]	LeastConfidence [15]	Margin [15]	Uniform	SkMM-FT2
2500	77.24 $\pm$ 0.08	77.25 $\pm$ 0.09	77.32 $\pm$ 0.20	77.46 $\pm$ 0.13	77.22 $\pm$ 0.10	77.58 $\pm$ 0.17	77.55 $\pm$ 0.21	77.46 $\pm$ 0.23	77.24 $\pm$ 0.15	77.55 $\pm$ 0.16	77.75 $\pm$ 0.08
3000	77.65 $\pm$ 0.10	77.70 $\pm$ 0.21	77.87 $\pm$ 0.21	77.95 $\pm$ 0.15	77.74 $\pm$ 0.07	77.74 $\pm$ 0.19	77.74 $\pm$ 0.12	77.93 $\pm$ 0.14	77.81 $\pm$ 0.23	78.04 $\pm$ 0.18	78.12 $\pm$ 0.04
3500	78.17 $\pm$ 0.07	78.04 $\pm$ 0.12	78.05 $\pm$ 0.04	78.19 $\pm$ 0.10	78.31 $\pm$ 0.11	78.37 $\pm$ 0.14	78.23 $\pm$ 0.20	78.21 $\pm$ 0.17	78.32 $\pm$ 0.17	78.46 $\pm$ 0.09	78.66 $\pm$ 0.06
4000	78.22 $\pm$ 0.11	78.46 $\pm$ 0.13	78.53 $\pm$ 0.09	78.54 $\pm$ 0.08	78.49 $\pm$ 0.10	78.39 $\pm$ 0.11	78.41 $\pm$ 0.08	78.60 $\pm$ 0.10	78.52 $\pm$ 0.15	78.83 $\pm$ 0.15	79.11 $\pm$ 0.02

Table 3: Accuracy and F1 (%) of linear probing (LP) over CLIP on StanfordCars.

n	Metric	CD [1]	DeepFool [52]	Forgetting [69]	Glistner [38]	GraNd [56]	Herdning [80]	Entropy [15]	LeastConfidence [15]	Margin [15]	Uniform	SkMM-LP
2000	Acc	67.64 $\pm$ 0.13	67.77 $\pm$ 0.29	67.59 $\pm$ 0.10	67.60 $\pm$ 0.24	67.27 $\pm$ 0.07	67.22 $\pm$ 0.16	67.95 $\pm$ 0.11	67.68 $\pm$ 0.11	67.53 $\pm$ 0.14	67.63 $\pm$ 0.17	68.27 $\pm$ 0.03
	F1	64.51 $\pm$ 0.17	64.16 $\pm$ 0.68	64.85 $\pm$ 0.13	64.50 $\pm$ 0.34	64.04 $\pm$ 0.09	64.07 $\pm$ 0.23	64.55 $\pm$ 0.10	64.09 $\pm$ 0.20	64.16 $\pm$ 0.15	64.54 $\pm$ 0.18	65.29 $\pm$ 0.03
2500	Acc	70.82 $\pm$ 0.23	70.73 $\pm$ 0.22	70.99 $\pm$ 0.05	70.85 $\pm$ 0.27	70.38 $\pm$ 0.07	71.02 $\pm$ 0.13	71.00 $\pm$ 0.10	70.99 $\pm$ 0.14	71.19 $\pm$ 0.09	70.59 $\pm$ 0.19	71.53 $\pm$ 0.05
	F1	68.18 $\pm$ 0.25	68.49 $\pm$ 0.53	68.53 $\pm$ 0.07	68.07 $\pm$ 0.38	67.48 $\pm$ 0.09	68.28 $\pm$ 0.15	67.95 $\pm$ 0.12	68.03 $\pm$ 0.20	68.33 $\pm$ 0.14	67.79 $\pm$ 0.23	68.75 $\pm$ 0.06
3000	Acc	72.66 $\pm$ 0.12	73.24 $\pm$ 0.22	72.54 $\pm$ 0.07	73.07 $\pm$ 0.26	72.56 $\pm$ 0.05	73.17 $\pm$ 0.22	73.28 $\pm$ 0.10	73.04 $\pm$ 0.05	73.09 $\pm$ 0.14	72.49 $\pm$ 0.19	73.61 $\pm$ 0.02
	F1	70.05 $\pm$ 0.11	70.93 $\pm$ 0.32	70.30 $\pm$ 0.05	70.47 $\pm$ 0.35	69.81 $\pm$ 0.08	70.64 $\pm$ 0.28	70.68 $\pm$ 0.12	70.30 $\pm$ 0.07	70.37 $\pm$ 0.17	70.00 $\pm$ 0.20	71.14 $\pm$ 0.03
3500	Acc	74.46 $\pm$ 0.17	74.57 $\pm$ 0.23	74.81 $\pm$ 0.05	74.63 $\pm$ 0.21	74.67 $\pm$ 0.06	74.64 $\pm$ 0.18	75.02 $\pm$ 0.08	74.65 $\pm$ 0.09	74.66 $\pm$ 0.11	74.16 $\pm$ 0.22	75.12 $\pm$ 0.01
	F1	72.13 $\pm$ 0.15	72.44 $\pm$ 0.27	72.59 $\pm$ 0.04	72.18 $\pm$ 0.25	72.13 $\pm$ 0.05	72.22 $\pm$ 0.26	72.46 $\pm$ 0.12	72.02 $\pm$ 0.10	72.03 $\pm$ 0.11	71.77 $\pm$ 0.23	72.64 $\pm$ 0.02
4000	Acc	75.77 $\pm$ 0.12	75.71 $\pm$ 0.15	75.74 $\pm$ 0.01	76.00 $\pm$ 0.20	75.77 $\pm$ 0.12	75.71 $\pm$ 0.29	75.82 $\pm$ 0.06	75.58 $\pm$ 0.08	75.57 $\pm$ 0.13	75.40 $\pm$ 0.16	76.34 $\pm$ 0.02
	F1	73.35 $\pm$ 0.07	73.79 $\pm$ 0.15	73.74 $\pm$ 0.02	73.69 $\pm$ 0.24	73.44 $\pm$ 0.13	73.26 $\pm$ 0.39	73.29 $\pm$ 0.04	73.15 $\pm$ 0.12	73.14 $\pm$ 0.20	73.14 $\pm$ 0.12	74.02 $\pm$ 0.10

Table 4: Accuracy and F1 score (%) of FT over (the last two layers of) ResNet18 (FT2) on StanfordCars

n	Metric	CD [1]	DeepFool [52]	Forgetting [69]	Glistner [38]	GraNd [56]	Herdning [80]	Entropy [15]	LeastConfidence [15]	Margin [15]	Uniform	SkMM-FT2
2000	Acc	28.50 $\pm$ 0.34	24.97 $\pm$ 0.20	28.61 $\pm$ 0.31	29.16 $\pm$ 0.26	28.59 $\pm$ 0.17	29.19 $\pm$ 0.21	28.87 $\pm$ 0.13	29.05 $\pm$ 0.07	29.18 $\pm$ 0.12	29.19 $\pm$ 0.37	29.44 $\pm$ 0.09
	F1	25.65 $\pm$ 0.40	22.11 $\pm$ 0.11	25.64 $\pm$ 0.25	26.33 $\pm$ 0.19	25.66 $\pm$ 0.15	25.90 $\pm$ 0.24	25.95 $\pm$ 0.17	26.18 $\pm$ 0.04	26.15 $\pm$ 0.12	26.14 $\pm$ 0.39	26.71 $\pm$ 0.10
2500	Acc	32.66 $\pm$ 0.27	29.02 $\pm$ 0.17	32.48 $\pm$ 0.28	32.91 $\pm$ 0.19	32.67 $\pm$ 0.20	32.42 $\pm$ 0.16	32.84 $\pm$ 0.20	32.88 $\pm$ 0.13	32.73 $\pm$ 0.15	32.83 $\pm$ 0.19	33.48 $\pm$ 0.04
	F1	29.79 $\pm$ 0.29	26.08 $\pm$ 0.29	29.58 $\pm$ 0.30	30.05 $\pm$ 0.28	29.70 $\pm$ 0.22	29.48 $\pm$ 0.23	30.03 $\pm$ 0.17	30.03 $\pm$ 0.14	29.66 $\pm$ 0.05	29.91 $\pm$ 0.16	30.75 $\pm$ 0.05
3000	Acc	35.67 $\pm$ 0.32	32.60 $\pm$ 0.18	35.18 $\pm$ 0.24	36.03 $\pm$ 0.20	35.83 $\pm$ 0.16	35.83 $\pm$ 0.24	35.64 $\pm$ 0.20	35.66 $\pm$ 0.18	35.67 $\pm$ 0.30	35.69 $\pm$ 0.35	36.11 $\pm$ 0.12
	F1	32.86 $\pm$ 0.31	29.83 $\pm$ 0.27	32.38 $\pm$ 0.20	33.26 $\pm$ 0.18	32.76 $\pm$ 0.16	32.89 $\pm$ 0.27	32.85 $\pm$ 0.23	32.79 $\pm$ 0.15	32.86 $\pm$ 0.30	32.80 $\pm$ 0.37	33.24 $\pm$ 0.05
3500	Acc	38.31 $\pm$ 0.15	35.59 $\pm$ 0.24	37.78 $\pm$ 0.22	38.16 $\pm$ 0.12	38.58 $\pm$ 0.15	38.30 $\pm$ 0.19	37.96 $\pm$ 0.11	38.25 $\pm$ 0.20	38.27 $\pm$ 0.20	38.31 $\pm$ 0.16	39.18 $\pm$ 0.03
	F1	35.55 $\pm$ 0.14	32.92 $\pm$ 0.33	35.16 $\pm$ 0.18	35.41 $\pm$ 0.14	35.72 $\pm$ 0.15	35.50 $\pm$ 0.22	35.19 $\pm$ 0.12	35.42 $\pm$ 0.16	35.61 $\pm$ 0.17	35.38 $\pm$ 0.19	36.38 $\pm$ 0.05
4000	Acc	40.53 $\pm$ 0.18	38.20 $\pm$ 0.22	40.24 $\pm$ 0.13	40.47 $\pm$ 0.16	40.70 $\pm$ 0.11	40.51 $\pm$ 0.19	40.29 $\pm$ 0.27	39.91 $\pm$ 0.09	40.58 $\pm$ 0.06	40.35 $\pm$ 0.26	41.77 $\pm$ 0.07
	F1	37.81 $\pm$ 0.23	35.47 $\pm$ 0.22	37.41 $\pm$ 0.14	37.63 $\pm$ 0.17	37.83 $\pm$ 0.11	37.56 $\pm$ 0.21	37.33 $\pm$ 0.34	37.14 $\pm$ 0.12	37.77 $\pm$ 0.07	37.51 $\pm$ 0.23	39.07 $\pm$ 0.10

Table 5: Summary of configurations of LP over CLIP and FT2 over ResNet18 on CIFAR-10 and StanfordCars.

Dataset	Number of classes	Fine-tuning setting	Number of parameters
CIFAR-10	10	LP over CLIP	5,130
		FT2 over ResNet18	2,364,426
Stanford Cars	196	LP over CLIP	100,548
		FT2 over ResNet18	2,459,844