

A Metric Learning Reality Check: Supplementary Material

Kevin Musgrave¹, Serge Belongie¹, Ser-Nam Lim²

¹Cornell Tech ²Facebook AI

1 Additional results

Table 1 shows results on CUB200 with a batch size of 256 instead of 32. The increase in batch size gives FastAP a significant boost in accuracy, and as a result, it performs on par with the rest of the methods, rather than underperforming. See github.com/KevinMusgrave/powerful-benchmark for more supplementary material, including the source code, configuration files, log files, and interactive bayesian optimization plots.

Table 1. Accuracy on CUB200, batch size 256

	Concatenated (512-dim)			Separated (128-dim)		
	P@1	RP	MAP@R	P@1	RP	MAP@R
Pretrained	51.05	24.85	14.21	50.54	25.12	14.53
Contrastive	67.60 ± 0.40	37.08 ± 0.14	26.25 ± 0.16	59.61 ± 0.18	31.85 ± 0.10	21.03 ± 0.10
Triplet	63.92 ± 0.32	34.32 ± 0.39	23.51 ± 0.38	56.08 ± 0.23	29.82 ± 0.25	19.02 ± 0.23
NTXent	66.88 ± 0.36	37.04 ± 0.16	26.13 ± 0.18	59.38 ± 0.14	32.14 ± 0.12	21.18 ± 0.12
ProxyNCA	66.21 ± 0.30	36.38 ± 0.12	25.53 ± 0.13	58.83 ± 0.24	31.60 ± 0.15	20.76 ± 0.16
Margin	64.98 ± 0.43	35.15 ± 0.29	24.13 ± 0.28	56.38 ± 0.42	29.76 ± 0.22	18.81 ± 0.20
Margin / class	66.51 ± 0.42	36.11 ± 0.19	25.18 ± 0.19	58.29 ± 0.32	31.03 ± 0.18	20.04 ± 0.17
N. Softmax	66.18 ± 0.41	36.15 ± 0.22	25.33 ± 0.23	59.08 ± 0.30	31.78 ± 0.18	20.95 ± 0.18
CosFace	66.73 ± 0.31	37.07 ± 0.16	26.23 ± 0.18	59.82 ± 0.22	32.06 ± 0.11	21.24 ± 0.11
ArcFace	66.61 ± 0.44	36.94 ± 0.24	26.08 ± 0.25	60.08 ± 0.30	32.46 ± 0.16	21.60 ± 0.15
FastAP	66.54 ± 0.47	37.07 ± 0.23	26.18 ± 0.24	59.26 ± 0.25	32.23 ± 0.18	21.32 ± 0.17
SNR	67.20 ± 0.30	37.38 ± 0.13	26.59 ± 0.13	59.71 ± 0.25	32.24 ± 0.17	21.43 ± 0.15
MS	68.00 ± 0.18	37.66 ± 0.13	26.86 ± 0.14	60.48 ± 0.13	32.83 ± 0.08	21.94 ± 0.09
MS+Miner	68.52 ± 0.37	37.95 ± 0.17	27.08 ± 0.19	60.41 ± 0.37	32.61 ± 0.15	21.66 ± 0.15
Soft Triplet	66.74 ± 0.38	37.06 ± 0.22	26.26 ± 0.23	59.94 ± 0.26	32.21 ± 0.11	21.45 ± 0.11