Supplementary Material for Mean Field Theory in Deep Metric Learning

Anonymous Author(s)
Affiliation
Address
email

A Additional experimental results

In this section, we present the experimental results that cannot be shown in the main paper due to the page limit.

MLRC results. Tables 1 – 3 show the complete results of Table 1 in the main paper, which are obtained the modern benchmark protocol proposed in the "Metric Learning Reality Check" (MLRC) paper [1]. In the CUB-200-2011 (CUB) dataset [2], MeanFieldContrastive (MFCont.) and MeanFieldClassWiseMultiSimilarity (MFCWMS) losses outperform the others in Mean Average Precision at R (MAP@R) and R-Precision (RP), while ProxyAnchor loss [3] is better in Precision at 1 (P@1) in the separated case. In contrast, in the Stanford Online Products (SOP) dataset [4], the MFCWMS loss shows the best performance in all the metrics.

Learning curves. Figure 1 shows learning curves obtained in the traditional evaluation protocol [3, 5] in fixed seeds, associated with Tables 2 and 3 in the main paper. Both MFCont. and MFCWMS losses show faster convergence than the ProxyAnchor loss. In the smaller datasets (CUB and Cars), accuracies of our mean field losses seem to decrease faster while we don't see such behaviors in the larger datasets (SOP and InShop [6]). This phenomenon might be caused by strong repulsive interactions with negative mean fields. For larger datasets, the embedding spaces may be sufficiently populated to balance the repulsive force, while this may not be the case for smaller datasets. It might not occur for ProxyAnchor loss since repulsive forces for ProxyAnchor loss are weighted depending on distances between proxy and negative samples.

Impact of batch size in InShop. Table 4 compares the MAP@R in ProxyAnchor and MFCWMS losses in the InShop dataset varying the batch size. As mentioned in the main paper, the accuracy of ProxyAnchor starts to decrease gradually for large batch sizes, while that of MFCWMS loss drops at batch size 150. Moreover, Table 5 shows the MAP@R in ProxyAnchor and MFCWMS for the InShop dataset without the query–gallery split of test data. In this case, accuracies of both losses start to decrease gradually around batch size 150. Thus, we conclude the accuracy drop in the MFCWMS loss probably comes from the specific query–gallery split.

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

Table 1: MLRC evaluation results in CUB-200-2011 [2]. We carry out 10 test runs and show averaged metrics with their confidence intervals.

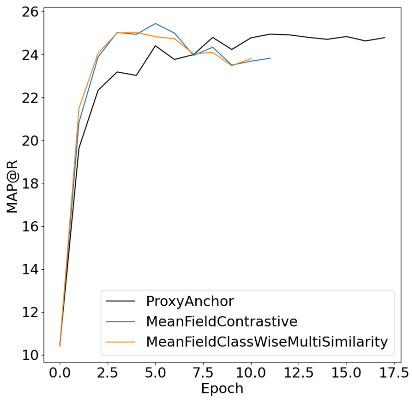
	Separated (128D)			Concatenated (512D)		
Loss	MAP@R	P@1	RP	MAP@R	P@1	RP
ArcFace	21.46 ± 0.13	59.98 ± 0.22	32.31 ± 0.14	26.39 ± 0.16	67.11 ± 0.23	37.23 ± 0.17
CosFace	21.19 ± 0.22	59.74 ± 0.28	32.00 ± 0.23	$\mathbf{26.54 \pm 0.29}$	67.14 ± 0.29	$\mathbf{37.38 \pm 0.28}$
MS	20.98 ± 0.16	59.38 ± 0.27	31.84 ± 0.15	26.20 ± 0.16	67.34 ± 0.35	36.99 ± 0.16
MS+Miner	20.78 ± 0.17	59.02 ± 0.25	31.67 ± 0.16	25.94 ± 0.18	67.08 ± 0.32	36.77 ± 0.16
ProxyNCA	18.75 ± 0.18	57.06 ± 0.27	29.64 ± 0.21	23.84 ± 0.22	65.60 ± 0.28	34.82 ± 0.25
ProxyAnch.	$\mathbf{21.67 \pm 0.22}$	$\underline{\mathbf{60.80 \pm 0.33}}$	$\mathbf{32.53 \pm 0.23}$	26.48 ± 0.23	$\mathbf{67.72 \pm 0.30}$	37.30 ± 0.23
Cont.	21.02 ± 0.14	59.35 ± 0.33	31.80 ± 0.15	26.37 ± 0.18	$\mathbf{67.67 \pm 0.25}$	37.10 ± 0.19
MFCont.	$\mathbf{22.01 \pm 0.10}$	$\mathbf{60.29 \pm 0.23}$	$\mathbf{32.85 \pm 0.10}$	$\underline{\mathbf{27.16 \pm 0.07}}$	67.64 ± 0.27	$\underline{\mathbf{37.95 \pm 0.07}}$
CWMS	21.48 ± 0.27	60.09 ± 0.27	32.32 ± 0.26	26.94 ± 0.29	$\underline{\mathbf{68.24 \pm 0.42}}$	37.69 ± 0.27
MFCWMS	$\underline{\mathbf{22.11 \pm 0.08}}$	$\mathbf{60.28 \pm 0.10}$	$\underline{\mathbf{32.96 \pm 0.08}}$	$\mathbf{27.03 \pm 0.12}$	67.63 ± 0.21	$\mathbf{37.83 \pm 0.12}$

Table 2: MLRC evaluation results in Cars-196 [7]. We carry out 10 test runs and show averaged metrics with their confidence intervals.

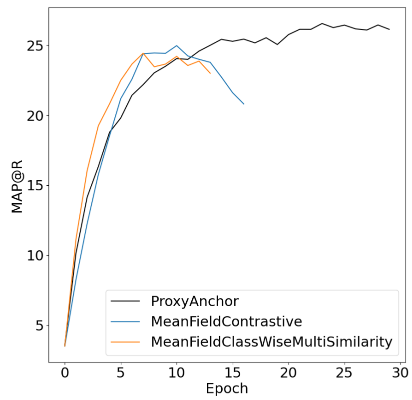
	Separated (128D)			Concatenated (512D)		
Loss	MAP@R	P@1	RP	MAP@R	P@1	RP
ArcFace	18.25 ± 0.12	71.12 ± 0.36	28.63 ± 0.13	$\mathbf{27.63 \pm 0.15}$	84.39 ± 0.15	$\mathbf{37.45 \pm 0.15}$
CosFace	18.49 ± 0.13	74.66 ± 0.21	28.75 ± 0.12	26.96 ± 0.25	85.29 ± 0.26	36.80 ± 0.24
MS	18.66 ± 0.30	71.89 ± 0.33	29.42 ± 0.29	27.19 ± 0.41	84.03 ± 0.30	37.39 ± 0.36
MS+Miner	18.49 ± 0.23	71.99 ± 0.28	29.20 ± 0.23	26.89 ± 0.38	83.89 ± 0.36	37.09 ± 0.33
ProxyNCA	17.43 ± 0.11	70.96 ± 0.26	27.85 ± 0.10	26.78 ± 0.18	84.31 ± 0.24	36.83 ± 0.17
ProxyAnch.	$\underline{\mathbf{19.44 \pm 0.17}}$	$\underline{\mathbf{76.15 \pm 0.25}}$	$\mathbf{29.89 \pm 0.18}$	26.81 ± 0.27	$\underline{\mathbf{85.53 \pm 0.30}}$	36.76 ± 0.26
Cont.	17.04 ± 0.26	69.77 ± 0.40	27.48 ± 0.26	24.93 ± 0.46	81.87 ± 0.35	35.12 ± 0.42
MFCont.	$\mathbf{18.12 \pm 0.13}$	$\mathbf{71.77 \pm 0.28}$	$\mathbf{28.54 \pm 0.14}$	$\mathbf{27.37 \pm 0.18}$	$\mathbf{84.56 \pm 0.21}$	$\mathbf{37.19 \pm 0.18}$
CWMS	$\mathbf{19.27 \pm 0.26}$	$\mathbf{74.19 \pm 0.30}$	$\underline{\mathbf{29.95 \pm 0.25}}$	$\underline{\mathbf{27.80 \pm 0.33}}$	$\mathbf{85.18 \pm 0.28}$	$\underline{\mathbf{37.89 \pm 0.29}}$
MFCWMS	18.85 ± 0.16	73.02 ± 0.20	29.55 ± 0.15	26.98 ± 0.31	84.00 ± 0.22	37.11 ± 0.27

Table 3: MLRC evaluation results in Stanford Online Products [4]. We carry out 10 test runs and show averaged metrics with their confidence intervals. We remove ProxyAnchor because it fails to converge in our settings.

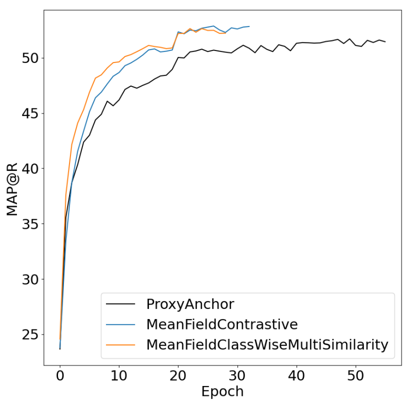
	Separated (128D)			Concatenated (512D)		
Loss	MAP@R	P@1	RP	MAP@R	P@1	RP
ArcFace	41.47 ± 0.24	71.39 ± 0.20	44.35 ± 0.23	$\mathbf{47.37 \pm 0.23}$	$\mathbf{76.13 \pm 0.16}$	$\mathbf{50.22 \pm 0.22}$
CosFace	41.01 ± 0.24	71.03 ± 0.22	43.89 ± 0.24	46.77 ± 0.20	75.69 ± 0.13	49.63 ± 0.20
MS	41.87 ± 0.21	71.10 ± 0.18	45.00 ± 0.20	46.70 ± 0.18	75.21 ± 0.15	49.70 ± 0.17
MS+Miner	41.90 ± 0.30	71.08 ± 0.25	45.05 ± 0.30	46.57 ± 0.28	75.09 ± 0.19	49.57 ± 0.28
ProxyNCA	$\mathbf{42.73 \pm 0.11}$	$\mathbf{71.77 \pm 0.08}$	$\mathbf{45.72 \pm 0.11}$	46.73 ± 0.13	75.24 ± 0.10	49.61 ± 0.13
Cont.	41.09 ± 0.18	70.04 ± 0.16	44.18 ± 0.19	45.35 ± 0.19	73.88 ± 0.15	48.28 ± 0.19
MFCont.	$\mathbf{43.62 \pm 0.36}$	$\mathbf{72.74 \pm 0.29}$	$\mathbf{46.55 \pm 0.35}$	$\mathbf{47.01 \pm 0.21}$	$\mathbf{75.57 \pm 0.16}$	$\mathbf{49.85 \pm 0.20}$
CWMS	41.53 ± 0.20	70.76 ± 0.16	44.50 ± 0.21	45.13 ± 0.16	73.99 ± 0.11	47.99 ± 0.16
MFCWMS	$\underline{\mathbf{44.57 \pm 0.16}}$	$\underline{\mathbf{73.32 \pm 0.11}}$	$\underline{\mathbf{47.53 \pm 0.16}}$	$\underline{\mathbf{48.33 \pm 0.18}}$	$\underline{\mathbf{76.38 \pm 0.14}}$	$\underline{\mathbf{51.17 \pm 0.18}}$



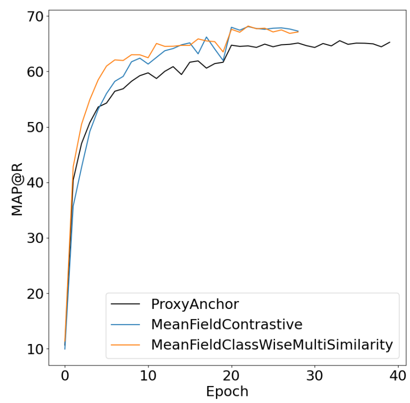
(a) CUB



(b) Cars



(c) SOP



(d) InShop

Figure 1: The test accuracy (MAP@R) plotted against the number of epochs for the (a) CUB, (b) Cars, (c) SOP, and (d) InShop datasets, comparing ProxyAnchor, MFCont., and MFCWMS.

Table 4: Test accuracies on InShop with the test dataset split into queries and galleries.

Batch size	ProxyAnchor	MFCWMS
30	63.6 ± 1.4	67.4 ± 0.2
60	$\mathbf{65.7 \pm 0.2}$	67.6 ± 0.2
90	65.5 ± 0.3	67.6 ± 0.2
120	65.6 ± 0.3	$\mathbf{67.8 \pm 0.2}$
150	65.5 ± 0.2	67.0 ± 0.6
300	64.5 ± 0.2	67.0 ± 0.4
500	63.3 ± 0.2	67.1 ± 0.1

Table 5: Test accuracies on InShop with the test dataset *not* split into queries and galleries.

Batch size	ProxyAnchor	MFCWMS
30	61.7 ± 0.6	64.7 ± 0.1
60	62.8 ± 0.4	$\mathbf{65.1 \pm 0.2}$
90	$\mathbf{62.9 \pm 0.3}$	65.0 ± 0.5
120	$\mathbf{62.9 \pm 0.3}$	64.9 ± 0.5
150	62.7 ± 0.1	65.0 ± 0.4
300	61.9 ± 0.2	64.7 ± 0.4
500	60.6 ± 0.2	64.6 ± 0.1

References

[1] Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 681–699. Springer, 2020.

[2] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[3] Sungyeon Kim, Dongwon Kim, Minsu Cho, and Suha Kwak. Proxy anchor loss for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3238–3247, 2020.

[4] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4004–4012, 2016.

[5] Yair Movshovitz-Attias, Alexander Toshev, Thomas K Leung, Sergey Ioffe, and Saurabh Singh. No fuss distance metric learning using proxies. In *Proceedings of the IEEE international conference on computer vision*, pages 360–368, 2017.

[6] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1096–1104, 2016.

[7] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 554–561, 2013.