Supplementary Material: Positive Pair Distillation Considered Harmful: Continual Meta Metric Learning for Lifelong Object Re-Identification

Kai Wang[1*]
kwang@cvc.uab.es
Chenshen Wu[1*]
chenshen@cvc.uab.es
Andy Bagdanov[3]
andrew.bagdanov@unifi.it
Xialei Liu (*corresponding author*)[2]
xialei@nankai.edu.cn
Shiqi Yang[1]
syang@cvc.uab.es
Shangling Jui[4]
jui.shangling@huawei.com
Joost van de Weijer[1]
joost@cvc.uab.es

[1] Computer Vision Center
Universitat Autònoma de Barcelona
Barcelona, Spain

[2] College of Computer Science
Nankai University
Tianjin, China

[3] MICC
University of Florence
Florence, Italy

[4] Huawei Kirin Solution
Shanghai, China

1 Implementation details.

We follow the same network structure and training strategy as DMML [1] for methods based on the DMML loss, and use the network and training protocol of BoT [10] for methods based on the softmax-triplet. For our person ReID experiments, we use ResNet-50 [3] pretrained on ImageNet [12] as our feature extractor. The last spatial downsampling operation in the network is removed to maintain high resolution. We resize input images to 256×128 for all methods. For vehicle ReID, we also use a ResNet-50 backbone pretrained on ImageNet as the embedding architecture, and use input images of size 224×224 augmented with random horizontal flips. We use the Adam optimizer [5] with a base learning rate of $LR = 0.0002$ and weight decay of 0.0001. All models are trained for 600 epochs with fixed learning rate of 0.0002 for the first 300 epochs, after which the learning rate is reduced by a factor of $0.005^{1/300}$ each epoch until the end. We set the trade-off coefficient to $\lambda = 1.0$, the margin as $\tau = 0.4$ as in DMML [1], and the temperature to $T = 1.0$ for DwoPP and $T = 10.0$ for DwPP. The number of classes, support images, and query images are in each episode are $N = 32, n_s = 5, n_q = 1$, respectively.

Continual learning metrics on Market 1501							
	BoT FT	LwF	AKA	IDA	DMML FT	DwPP	DwoPP
Plasticity	9.7	9.5	9.7	7.2	10.8	10.7	8.1
Forgetting	-8.9	-7.7	-6.5	-5.9	-6.8	-6.5	-2.9
Overall	0.8	1.8	3.2	1.3	4.0	4.2	5.2

Table 4: Average forgetting and plasticity in mAP (%) on Market-1501 together with the overall mAP change (defined as plasticity plus forgetting).

		continual metric learning training tasks				Unseen-test task
Task id:		1	2	...	10	
Identities per task:	Market-1501	76	75	...	75	750
	MSMT17_V2	105	104	...	104	3060
	VeRi-776	63	57	...	57	200

Table 5: Our proposed 10-task split of two Person ReID datasets and one Vehicle ReID dataset for Continual Metric Learning.

2 Evaluating forgetting.

To further analyze the results we measure forgetting and plasticity on the Market 1501 dataset. Continual learning aims to counter forgetting (stability) while optimally learning new tasks (plasticity). To measure these, we track the change in mAP for each identity in the unseen test set after each task: a drop is added to forgetting, an increase to plasticity. In Table 4 we report the plasticity and forgetting averaged over tasks. We see that DwoPP has greatly reduced forgetting at the price of only a small decrease in plasticity.

3 Continual Metric Learning splitting protocols

Our proposed Continual Metric Learning splits for two Person ReID datasets and one Vehicle ReID dataset are shown in Table. 5. For all three datasets, we try to uniformly distribute the training identities into 10 continual metric learning tasks. The query and gallery set are fixed and serve as the unseen-test task.

4 DMML loss illustration

An illustration of the DMML loss [1] is shown in Fig. 5. The hard-mining DMML loss finds the largest distance to a positive example and the smallest distance to a negative sample to compute the metric loss with a margin.

5 More random orders on Market-1501 dataset.

In the main paper, we split tasks according to the *object IDs*. Thus, for the purpose of verifying the robustness of our proposed method to various orderings of the tasks, we randomly generate three different orderings of person IDs from Market-1501 to split the tasks (see Fig. 6). The results show that the trends are similar as those reported in Table 1. Results are averages and standard deviations in mAP and Rank-1 Accuracy over these three runs.

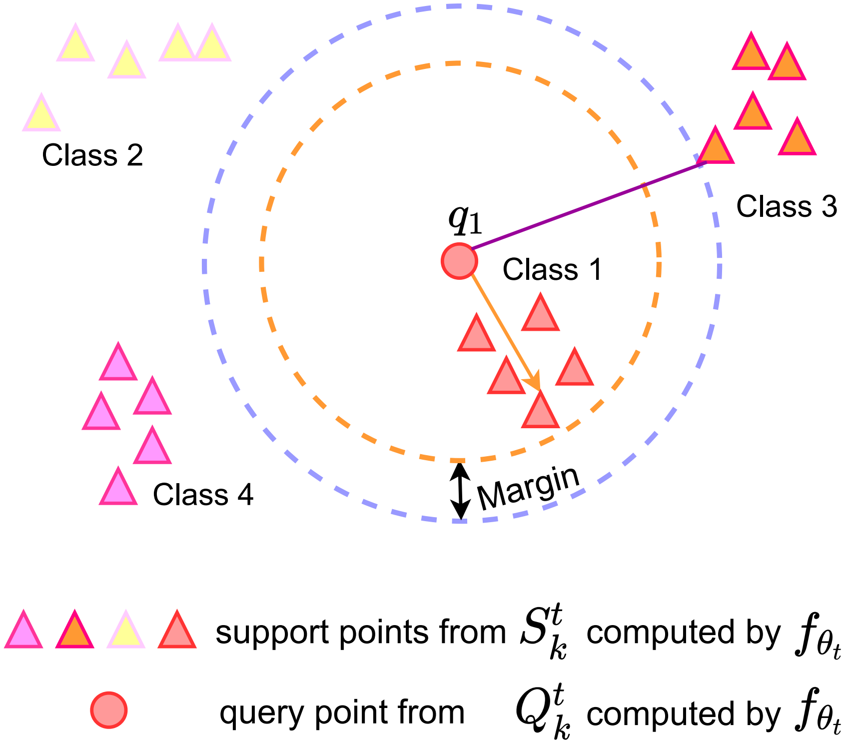
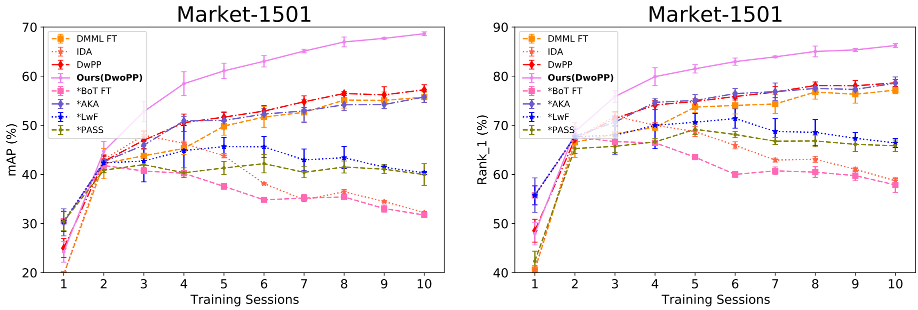


Figure 5: Supposing the query point is from class 1, the hard-mining DMML loss selects the farthest positive point and nearest negative point to compute the distance. It forces a margin between negative and positive distances.

6 Lifelong Person ReID (LReID) benchmark

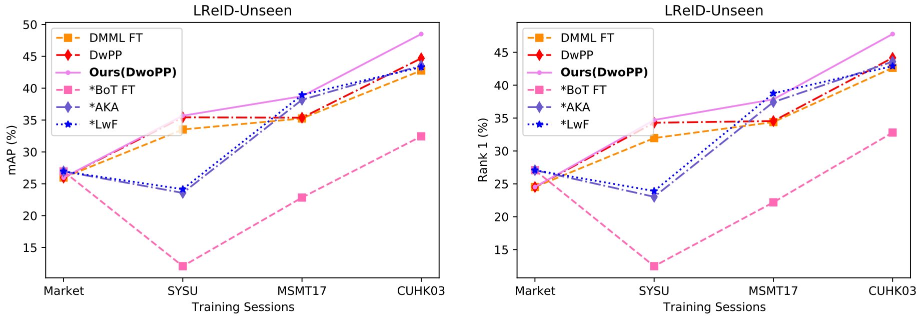
For the LReID benchmark[11] we removed the DukeMTMC-ReID dataset due to its retraction on account of privacy issues. Except for this change, we keep the same training order as LReID Order-1: Market-1501 [16]→ CUHK-SYSU [14] → MSMT17_V2 [13] → CUHK03 [8]. After training each task, we evaluate the model over the test query and gallery sets in LReID-Seen for these four datasets, and also on the LReID-Unseen test set consisting of seven person ReID datasets: VIPeR [2], PRID [4], GRID [9], i-LIDS [17], CUHK01 [7], CUHK02 [6], and SenseReID [15]. All the performance curves on LReID-Seen are shown in Fig. 8 and the curves on LReID-Unseen are shown in Fig. 7.

An interesting phenomenon we observed in the main paper is that DwPP is always better in the current task evaluation. We assume this is because DwPP forces the predictions to be aligned with the probability distributions of the old model, which contain some information about the relative distances of these identities. This extra information further enhances representation learning in the current task, thus leading to better performance on the current task even compared to the finetuning baseline (which is usually better on the current task).



(a) mAP on Market-1501 (b) Rank-1 on Market-1501

Figure 6: Performance on Market-1501 averaged over three random ID orders with standard deviation.

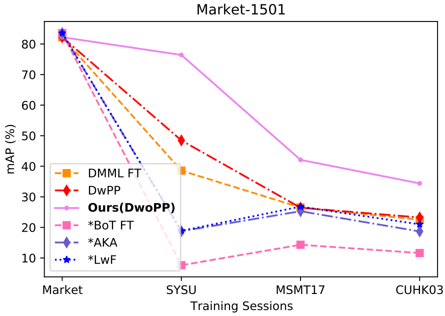


(a) mAP on LReID-Unseen (b) Rank-1 Accuracy on LReID-Unseen

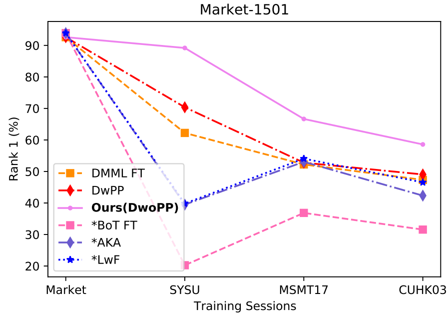
Figure 7: Results in mAP and Rank-1 Accuracy the LReID-Unseen test set of the LReID benchmark. The training order is (Market-1501→ CUHK-SYSU → MSMT17_V2 → CUHK03).

7 Limitations and ethical considerations

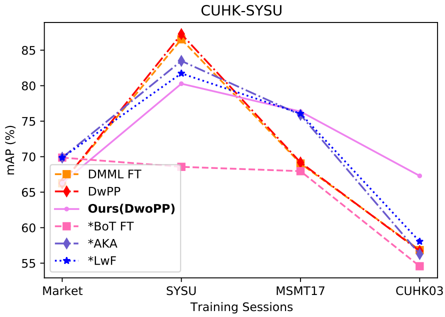
Person ReID is fraught with ethical concerns over its potential to violate the privacy of observed subjects. Although continual learning for Person ReID offers the possibility of learning and updating models without the need for long-term retention of sensitive data, it also runs the risk of "baking" biases into the model that, due to mitigation of forgetting, become difficult to remove. For real applications there is still a large gap between joint and continual training for object ReID, and a limitation of the experiments in this work is the relatively short task sequences we consider.



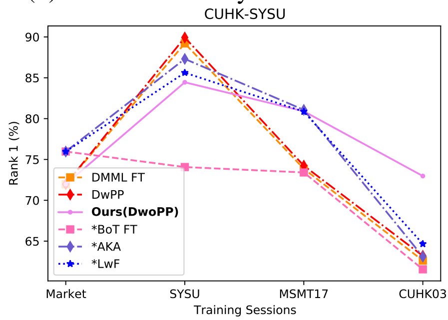
(a) mAP on Market-1501



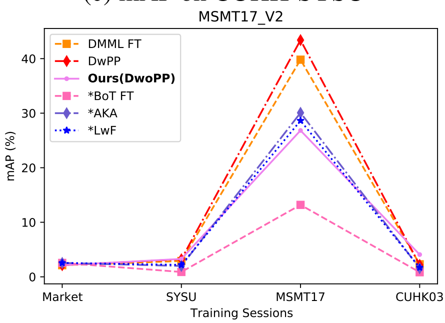
(b) Rank-1 Accuracy on Market-1501



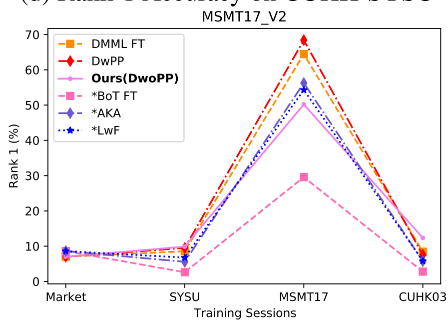
(c) mAP on CUHK-SYSU



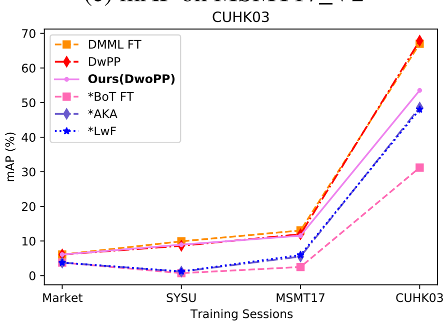
(d) Rank-1 Accuracy on CUHK-SYSU



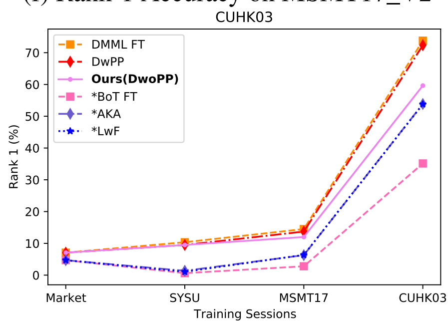
(e) mAP on MSMT17_V2



(f) Rank-1 Accuracy on MSMT17_V2



(g) mAP on CUHK03



(h) Rank-1 Accuracy on CUHK03

Figure 8: Results in mAP and Rank-1 Accuracy on the LReID benchmark. The training order is (Market-1501→ CUHK-SYSU → MSMT17_V2 → CUHK03). The first four rows show the evaluation on these four tasks respectively.

References

[1] Guangyi Chen, Tianren Zhang, Jiwen Lu, and Jie Zhou. Deep meta metric learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9547–9556, 2019.

[2] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European conference on computer vision*, pages 262–275. Springer, 2008.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[4] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, pages 91–102. Springer, 2011.

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015.

[6] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3594–3601, 2013.

[7] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Asian conference on computer vision*, pages 31–44. Springer, 2012.

[8] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 152–159, 2014.

[9] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision*, 90 (1):106–129, 2010.

[10] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

[11] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7901–7910, 2021.

[12] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115 (3):211–252, 2015.

[13] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79–88, 2018.

[14] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2(2):4, 2016.

[15] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1077–1085, 2017.

[16] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the International Conference on Computer Vision*, pages 1116–1124, 2015.

[17] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *Proceedings of the British Machine Vision Conference*, volume 2, pages 1–11, 2009.