# 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

Andrew D. 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

## Abstract

Lifelong object re-identification incrementally learns from a stream of re-identification tasks. The objective is to learn a representation that can be applied to all tasks and that generalizes to previously *unseen* re-identification tasks. The main challenge is that at inference time the representation must generalize to previously *unseen* identities. To address this problem, we apply continual meta metric learning to lifelong object re-identification. To prevent forgetting of previous tasks, we use knowledge distillation and explore the roles of positive and negative pairs. Based on our observation that the distillation and metric losses are *antagonistic*, we propose to remove positive pairs from distillation to robustify model updates. Our method, called Distillation without Positive Pairs (DwoPP), is evaluated on extensive intra-domain experiments on person and vehicle re-identification datasets, as well as inter-domain experiments on the LReID benchmark. Our experiments demonstrate that DwoPP significantly outperforms the state-of-the-art.

## 1 Introduction

*Object re-identification (ReID)* aims to associate the identity of a query image with those in a gallery set [13, 75]. It is applied to many applications, including person re-identification [8,

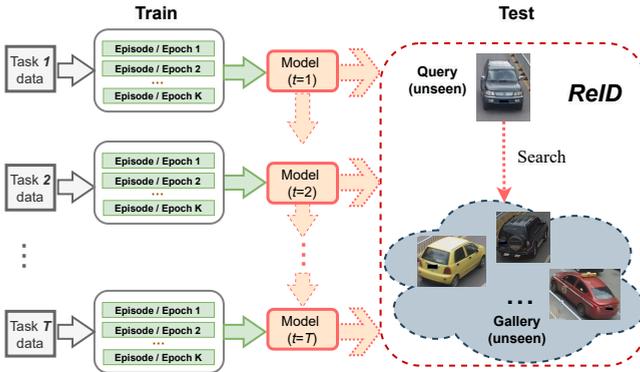*These authors contributed equally to this work.

Figure 1: Lifelong Object ReID with continual meta-metric learning. Unlike conventional object re-identification, data are presented sequentially in discrete tasks of disjoint classes. Data from previous tasks are unavailable in successive ones and the learner must incrementally update when a new task arrives. Furthermore, in object re-identification the test identities are not seen during training, which demands generalization of the learned metric.

[31, 72], vehicle re-identification [24, 39, 79], and face verification [60, 61]. Most existing approaches assume that the test and training dataset are drawn from the same distribution and that all training data is available jointly when training the network [31, 39, 41, 75, 79]. In domain generalization ReID [2, 9, 11, 47, 56] all source domain data is assumed available during training. This assumption is not realistic for many applications as all training data might not be available from the start and its distribution could vary over time. In addition, the trained system could be applied at inference time to new data never seen during training. Only recently, the problem of Lifelong ReID has been proposed [50]. This setting requires learning from a *sequence* of domains, and evaluates the algorithm on *unseen* domains.

Continual learning [12, 38, 44, 46] addresses the problem of learning from non-stationary streams of data. It has developed several techniques including regularization-based methods [1, 14, 26, 35, 74], parameter-isolation [42, 43, 45, 53], and replay-based methods [16, 36, 52, 57, 58, 70]. In this paper we consider exemplar-free continual learning where it is not allowed to save any samples (exemplars) of previous tasks for the problem of object re-identification. This requirement is out of the privacy considerations in person ReID problems.

Most continual learning methods specifically consider the incremental learning of classification problems. The considered setup for object re-identification (Fig. 1) is different in two main aspects. Firstly, they usually do not incrementally learn a *classifier*, instead they incrementally learn a feature representation. Secondly, the aim is to perform evaluation on new unseen tasks. So the real goal is to incrementally learn a metric space that generalizes to previously unseen tasks. Pu et al. [50] propose a method to address the first problem but ignore the second consideration: *the representation should generalize to unseen tasks*.

Meta-learning [3, 8, 13, 22, 48, 55, 59] focus on generalising to unseen tasks and has been applied to few-shot learning [4, 27, 53, 57, 65, 71]. Object ReID can be considered a few-shot learning problem, since the object identities at test time are not shown during the training and we only have few support images. To exploit the generalization capability of meta learning, Chen et al. [6] propose Deep Meta Metric Learning (DMML) that formulates

the deep metric learning as a meta learning problem. Since the main challenges of object re-identification are learning from a sequences and generalization to previously unseen domains, in this paper we propose *Continual Meta Metric Learning* to address this problem.

To further endow continual meta metric learning with a mechanism to mitigate forgetting knowledge from previous tasks, we introduce a temporary classifier for the support set and study the potential of directly applying knowledge distillation [19, 32]. However, we find that the distillation and metric learning losses are antagonistic. We therefore propose **D**istillation **w**ith**o**ut **P**ositive **P**airs (DwoPP). DwoPP, different from naive distillation, which distills knowledge from the previous to the current task classifier over *all* classes in the current task, distills only using *negative* examples. In this way, we avoid the antagonistic relationship between the metric and distillation losses which is from positive pairs distillation.

The main contributions of our paper are: **1)** we show that meta metric learning is superior to global metric learning for object re-identification; **2)** we explicitly explore the roles of positive and negative pairs in distillation and propose a novel distillation scheme called DwoPP for Continual Meta Metric Learning; **3)** we propose task splits for evaluation of continual metric learning methods on intra-domain object ReID for three ReID datasets and evaluate on much longer sequences than existing benchmarks; and **4)** we perform extensive experimental analysis demonstrating that, DwoPP achieves significantly better performance on person and vehicle ReID, as well as on the lifelong re-identification (LReID) benchmark [50].

# 2 Related work

**Object re-identification and metric learning.** Metric learning has been widely applied to object re-identification [18, 75], mainly focusing on person ReID [5, 41, 72, 76], vehicle ReID [24, 39, 79] and face verification [60, 61, 76]). Deep metric learning methods can be divided into three categories based on the loss used: contrastive loss with pairwise inputs [10], triplet loss with triplet inputs [21], and N-pair loss with batch inputs [54]. In general, deep metric learning works well but does not take generalization of the learned metrics into account and neglects relationships between inter-class samples. DMML [6] formulates metric learning for object re-identification from a meta learning perspective. We build upon DMML for our *continual learning* view of meta metric learning.

**Continual learning.** Continual learning methods can be categorized into three groups: parameter-isolation, regularization-based and replay-based methods [12]. The most relevant to our work are regularization-based methods [1, 26, 32, 35, 35, 49, 73, 74]. Knowledge distillation is a widely used regularization method which decreases forgetting by either aligning features [36, 57] or the predicted probabilities [32]. To adapt knowledge distillation to Continual Meta Metric Learning, we propose a variant of knowledge distillation by introducing a temporary classifier for the current support set, and more importantly the distillation in the paper is without considering positive pairs. Replay-based continual learning overcomes forgetting by saving a set of exemplars from each task [16, 23, 23, 64, 57, 68, 68, 70]. We focus on exemplar-free continual learning. And continual learning applied to persons in particular has privacy considerations which makes retaining data problematic.

**(Incremental) Meta learning.** Meta learning based on metrics or optimization-based approaches are the main directions of current research [65]. ProtoNets [55] and RelationNets [58] are canonical representatives of metric-based approaches, while MAML [13] and Reptile [48] are representative optimization-based methods. Incremental meta learning

(IDA [34], ERD [53]) methods have been mainly developed for incremental few-shot learning, however, they can also be applied to lifelong object ReID and we will compare to them in the experimental section. There are a few methods on incremental metric learning which approach the problem as one of representation learning with a metric-based classification loss. Examples include CRL [76], FGIR [7], and AKA [50]. However, these works all focus on distillation over seen classes and thus neglect the need to recognize unseen identities.

# 3 Methodology

## 3.1 Preliminaries

There are two main approaches to metric learning applied to object ReID: those based on global optimization of a metric embedding over the training set, and those based on episodic meta learning. Most global optimization metric learning methods minimize a metric loss over the whole dataset $D = (\mathbf{X}, \mathbf{Y})$ of inputs $\mathbf{X}$ and corresponding labels $\mathbf{Y}$. For comparison in this paper we use the popular softmax-triplet loss as used in Bag-of-Tricks (BoT) [41].

**Deep meta metric learning (DMML).** In DMML [6], the authors instead formulate metric learning as a meta learning problem. They decompose the training data into a series of sub-tasks, called *episodes* in meta learning, and then learn a meta metric that generalizes well to all sub-tasks. Assuming the unseen test task is drawn from the same distribution of sub-tasks from the training set, this learned meta metric should generalize to this unseen test task.

Assume that we sample $K$ episodes in total for training, that each episode $E_k$ is composed of $N$ classes, and that each class contains $n_s$ images in the support set $S_k$ and $n_q$ images in the query set $Q_k$. In each episode, we learn the meta metric to correctly predict the query samples from support samples. The learning problem for DMML is:

$$\theta^* = \arg\min_\theta \mathbb{E}_{k \in [1,K]} \left[ \mathcal{L}_{\text{eps}}(\theta; S_k, Q_k) \right] \tag{1}$$

where $\mathcal{L}_{\text{eps}}$ is the episode level hard-mining metric loss proposed in DMML [6].

The episodic loss $\mathcal{L}_{\text{eps}}$ is defined in terms of positive and negative pairs. In the current episode $E_k$ with the class set $\mathbb{C}$, a query point $q_c \in Q_k$ is drawn from a specific class $c \in \mathbb{C}$. We construct the *positive pairs* $[q_c, s_c]$ from the query point and support points $s_c \in S_k$ from the class $c$, and negative pairs $[q_c, s_{c'}]$ from the query point and support points $s_{c'} \in S_k$ from *different* classes $c' \neq c$. Hard mining is performed over the positive pairs by finding largest Euclidean distance from $q_c$ to a positive support sample $d_c = \max_{s_c \in S_k} d(q_c, s_c)$, and over negative pairs by finding the smallest distance from $q_c$ to the a negative support sample $d_{c'} = \min_{s_{c'} \in S_k} d(q_c, s_{c'})$. $\mathcal{L}_{\text{eps}}$ is defined in terms of these hard-mined distances ($\tau$ is a margin):

$$\mathcal{L}_{\text{eps}}(\theta; S_k, Q_k) = \sum_{q_c \in Q_k} \log(1 + \sum_{c' \in \mathbb{C} \setminus \{c\}} \exp(d_{c'} - d_c + \tau)), \tag{2}$$

## 3.2 Continual Metric Learning

In continual metric learning, tasks $t \in [1, T]$ arrive sequentially as disjoint datasets $D_t$. The aim is to learn $\theta_t$ incrementally in a *training session* for each task $t$ and to ensure it accumulates knowledge from the previous tasks so as to generalize better to unseen test tasks:

$$\theta_t^* = \arg\min_{\theta_t} \mathcal{L}_{\text{cml}}(\theta_t; D_t). \tag{3}$$
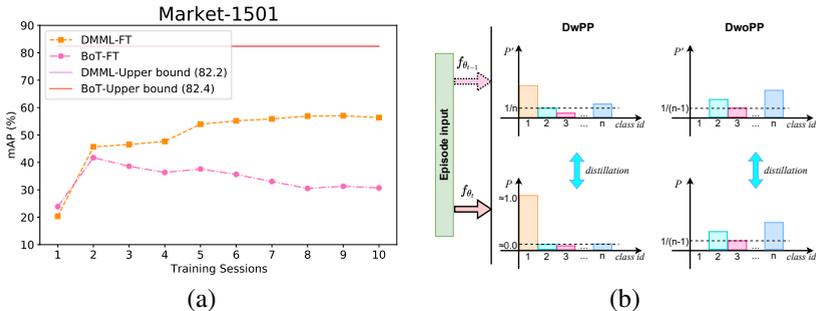
Figure 2: (a) Comparing continual meta-metric learning (DMML-FT [6]) with continual metric learning (BoT-FT [41]). We finetune on 10 equally split Market-1501 tasks. Upper bounds are joint training on all data. (b) Comparison between DwPP and DwoPP (class 1 is the positive class). The old model has never seen class 1 and so likely produces an output less than 1 although we want positive pairs to map to the exact same point in latent space. Also, the dominance of the positive class inhibits distillation of negative pair information.

And the data from previous tasks (i.e. $D_{t'}$ for $t' < t$) are *not* available to the learner at task $t$. Eq. 3 defines the *continual* learning setup where one only has access to data of a single task at a time. It is a general equation applicable to continual learning setups. In Eq. 3 $\mathcal{L}_{cml}$ could be replaced with a metric learning loss (yielding continual metric learning) or with a meta metric learning loss, like Eq. 1, to obtain continual meta metric learning. The challenge of continual learning is preventing forgetting of previous knowledge.

Here we consider two different losses for $\mathcal{L}_{cml}$ in Eq. 3, either based on meta-learning (like DMML [6]) or on the softmax (like BoT [41]). The majority of Person Re-Identification approaches (including the LReID benchmark [50]) are based on the softmax-triplet loss. We compare these methods in the Continual Metric Learning setting on Market-1501 in Fig. 2(a) by simply applying fine-tuning (FT) without any mitigation of forgetting. We clearly see that continual metric learning is quickly surpassed by continual meta-metric learning. The underlying reason for this marked improvement is that re-identification aims to recognize *unseen* objects (each object identity is represented by only one query image at test time). This is the central characteristic of few-shot recognition. Instead, the conventional softmax-triplet loss optimizes recognition on seen classes. It does not explicitly aim for generalization to unseen classes. Moreover, the focus on current task classes can also lead to increased forgetting of previous classes. The meta learning DMML loss, however, tends to learn a better representation space that generalizes to future unseen tasks and thus suffers less from forgetting. In brief, DMML is a more principled approach for continual metric learning than the softmax-triplet loss and we propose to use DMML as the basis.

## 3.3 Distillation without Positive Pairs (DwoPP)

To adapt the DMML loss defined in Eq. 2 to continual metric learning, we compute it for task $t$ over episodes $E_k^t$ drawn only from the current task data $D_t$. We denote the support set and query set of each episode during task $t$ as $S_k^t$ and $Q_k^t$. Then the DMML loss is defined with the current model $f_{\theta_t}$ as $\mathcal{L}_{eps}(\theta_t; S_k^t, Q_k^t)$ (see Eq. 2).

Episodic meta learning with the DMML loss will not mitigate forgetting in a continual metric learning. Knowledge Distillation [19, 32] is a common technique for alleviating

catastrophic forgetting when learning over a sequence of tasks. Note, however, distillation assumes that a classifier over classes from the previous tasks is available on which to perform knowledge distillation – something that for continual meta metric learning we do not have. However, based on the sampled episodes we can construct two temporary classifiers, one based on the previous and one based on the current tasks' feature extractor. We can then define a new distillation loss in terms of these temporary classifiers.

**Class Prototypes.** To construct the temporary classifier, we compute prototypes as the centroid of embedded samples of each class $\mathbf{u}_c$ ($c$ is class label):

$$\mathbf{u}_c = \frac{1}{n_s} \sum_{(x_i, y_i) \in S_k^t} f_\theta(x_i) \delta_c(y_i), \tag{4}$$

where $\delta_c(y) = 1 \Leftrightarrow y = c$ is an indicator function.

**DwPP: Distillation with Positive Pairs.** With the class prototypes $\mathbf{u}_c$, the prediction for class $c \in \mathbb{C}$ of query image $\hat{x} \in Q_k^t$ with the model $f_{\theta_t}$ is given by:

$$g_c(S_k^t, \hat{x}; \theta) = \frac{[\exp(-d(f_\theta(\hat{x}), \mathbf{u}_c))]^{1/T}}{\sum_{c' \in \mathbb{C}} [\exp(-d(f_\theta(\hat{x}), \mathbf{u}_{c'}))]^{1/T}}, \tag{5}$$

where $T$ is the temperature and $d$ is the Euclidean distance. These predictions are used to distill knowledge from task $t$-1 into task $t$ by constructing two temporary classifiers, one using $\theta_t$ and another using $\theta_{t-1}$, and considering all negative and positive pairs:

$$\mathcal{L}_{\text{DwPP}}(\theta_t; \theta_{t-1}, S_k^t, Q_k^t) = \sum_{\hat{x} \in Q_k^t} KL\left[\mathbf{g}(S_k^t, \hat{x}; \theta_{t-1}) \,||\, \mathbf{g}(S_k^t, \hat{x}; \theta_t)\right]. \tag{6}$$

Here $\mathbf{g}$ is a classifier constructed by concatenating the predictions $g_c$ defined in Eq. 5 for all classes in the episode.

Knowledge distillation for continual meta metric learning requires careful attention to which pairs are included in the distillation loss. Consider the hypothetical case illustrated in Fig. 2(b) where we show the predictions of the two temporary classifiers (class 1 is the query class). In task $t$, the new classes from $D_t$ are not well-discriminated from each other – that is, the margin between positive and negative pairs in $D_t$ is not guaranteed by the model from task $t - 1$ and the predicted probabilities are distributed as in the upper left column of Fig. 2(b). After learning task $t$ we would like it to be a peaked distribution around the correct class, and simultaneously we also wish to maintain the relative probabilities of all classes (via knowledge distillation). Although this distillation will maintain model stability and mitigate forgetting, the estimate of the old model for the correct label is likely to be unreliable and will prevent the metric loss from pushing similar labels to the same position in the embedding space. Furthermore, the dominance of the positive class prevents distillation of the relevant negative pair information (also known as dark knowledge [19]), which weakens the alignment of classes in the feature space.

In essence, the metric and distillation losses are *antagonistic* due to the inclusion of positive pairs in knowledge distillation. Thus we propose to remove positive pairs from distillation. As shown in the right column of Fig. 2(b), since the other classes are negatives for class 1, they can be easily aligned with the previous probabilities to overcome forgetting. At the same time, the peaked distribution in the bottom left of Fig. 2(b) can also be achieved by the metric loss. To further analyze the role of positive and negative pairs, we decouple the
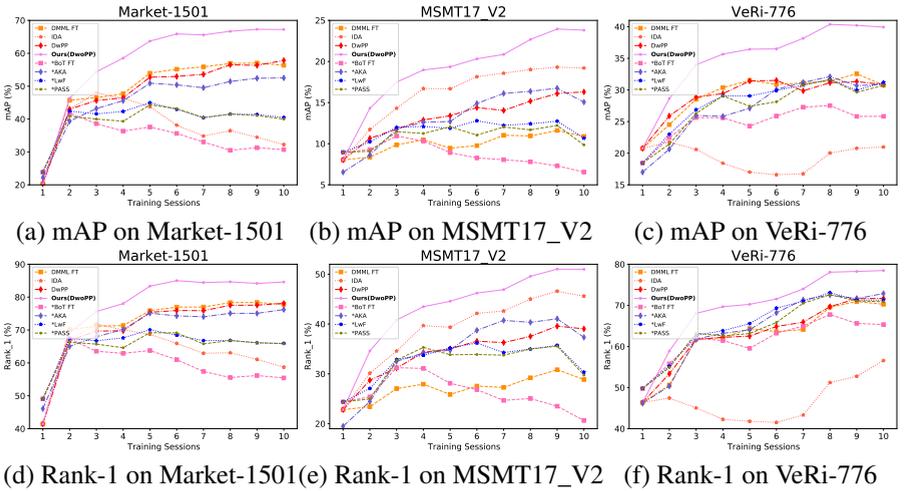
(a) mAP on Market-1501 (b) mAP on MSMT17_V2 (c) mAP on VeRi-776

(d) Rank-1 on Market-1501 (e) Rank-1 on MSMT17_V2 (f) Rank-1 on VeRi-776

Figure 3: mAP and Rank-1 performance. Methods with "*" use the softmax-triplet loss.

KL divergence into positive and negative pair distillation as proposed by DKD [7], showing that positive pair distillation leads to performance degradation (see Table 3).

**DwoPP: Distillation without Positive Pairs.** To remove positive pairs from DwPP distillation, we exclude class $\hat{y}$ which is the class label of the query image $\hat{x} \in Q_k^t$ from the temporary classifier and rewrite the Eq. 5 as:

$$g_c'(S_k^t, \hat{x}, \hat{y}; \theta) = \frac{[\exp(-d(f_\theta(\hat{x}), \mathbf{u}_c))]^{1/T}}{\sum_{c' \in \mathbb{C} \setminus \{\hat{y}\}} [\exp(-d(f_\theta(\hat{x}), \mathbf{u}_{c'}))]^{1/T}} \tag{7}$$

Then the DwoPP distillation can be rewritten as:

$$\mathcal{L}_{\text{DwoPP}}(\theta_t; \theta_{t-1}, S_k^t, Q_k^t) = \sum_{(\hat{x}, \hat{y}) \in Q_k^t} KL[\mathbf{g}'(S_k^t, \hat{x}, \hat{y}; \theta_{t-1}) \| \mathbf{g}'(S_k^t, \hat{x}, \hat{y}; \theta_t)]. \tag{8}$$

With the above defined DwoPP distillation loss and episode DMML loss, the continual metric learning loss function for each episode is defined as:

$$\mathcal{L}_{\text{cml}}(\theta_t; \theta_{t-1}, S_k^t, Q_k^t) = \mathcal{L}_{\text{eps}}(\theta_t; S_k^t, Q_k^t) + \lambda \mathcal{L}_{\text{DwoPP}}(\theta_t; \theta_{t-1}, S_k^t, Q_k^t). \tag{9}$$

To demonstrate the necessity of removing positive pairs from the distillation, we compare DwPP and DwoPP in Sec. 4 and perform an ablation on $T$ in both.

# 4 Experimental Results

## 4.1 Experimental setup

**Datasets for Intra-domain Object ReID.** We propose continual metric learning splits for two Person ReID datasets and one vehicle ReID dataset. **(1) Market-1501 [80]** consists of 32,668 images of 1,501 identities captured by 6 cameras. The dataset is divided into a training set with 12,968 images of 751 identities and a test set containing 3,368 query images

and 19,732 gallery images of 750 identities. For continual metric learning setup, we split the 751 training identities into 10 disjoint tasks, each with 75 identities (the first with 76). **(2) MSMT17_V2 [56]** consists of 126,441 images of 4101 persons captured by 15 cameras. Its training set includes 30,248 images of 1041 persons, and its test set covers the remaining 3060 persons with 11,659 query images and 82,161 gallery images. For MSMT17_V2, we split the training persons into 10 tasks also, each task with 104 persons (the first task with 105 persons). **(3) VeRi-776 [37]** contains 49,357 images of 776 vehicles, which are captured by 20 cameras. Among them, 576 vehicles are used for training and the remaining 200 are used for testing. In total, VeRi-776 consists of 37,778 training images, 1,678 query images, and 11,579 gallery images. For continual metric learning, we split the training 576 vehicles into 10 tasks, each task with 57 vehicles (the first task with 63 vehicles).

**The Lifelong ReID (LReID) benchmark.** We adapt the train set of the inter-domain LReID benchmark by building it from four datasets: Market-1501 [80], CUHK-SYSU ReID [69], MSMT17_V2 [56], and CUHK03 [30].* After training, the model is evaluated on the test query and gallery sets LReID-Seen of these four datasets (i.e. over *seen* domains). We also test on LReID-Unseen test set which combines seven person ReID datasets: VIPeR [15], PRID [20], GRID [40], i-LIDS [81], CUHK01 [29], CUHK02 [28], and SenseReID [78].

**Implementation details.** We follow the same network structure and training strategy as DMML [6] for our method. The ResNet-50 [17] pretrained on ImageNet [52] works as our feature extractor for all methods. The feature extractor is further trained during continual training. We use the Adam optimizer [25] with a base learning rate of $LR = 0.0002$ and weight decay of 0.0001. We set the trade-off coefficient to $\lambda = 1.0$, the margin as $\tau = 0.4$, and the temperature to $T = 1.0$ for DwoPP and $T = 10.0$ for DwPP. The number of classes, support and query images in each episode are $N = 32, n_s = 5, n_q = 1$.

**Compared methods and metrics.** Our evaluation is divided into two parts: **(1)** To compare with conventional continual learning methods, we train models with the softmax-triplet loss of BoT [41]. For methods using this loss without exemplars, we selected AKA [60], PASS [82], and LwF [32]. For methods using exemplars, we selected FT+, iCaRL [51] and LwF+ [32]. **(2)** For comparison with incremental meta learning methods, we build upon the DMML loss [6]. For methods without exemplars we selected IDA [34]. For methods with exemplars, we selected ERD [63]. Note that AKA is the state-of-the-art in LifelongReID and IDA is the state-of-the-art in incremental meta learning. For all exemplar-based methods we store 500 exemplars for all experiments. We use mean Average Precision (mAP) and Accuracy at Rank-1 as metrics [77]. We compute the mAP and Rank-1 Accuracy of the model on the unseen test set after each task. All results are averages over three runs.

## 4.2   Comparative performance evaluation

**Intra-domain Lifelong Object ReID.** Fig. 3 gives the mAP and Rank-1 curves on Market-1501, MSMT17_V2, and VeRI-776. We report the performance of all methods after task $t = 10$ and the average metrics over all training sessions in Table 1. On all three datasets, finetuning with softmax-triplet loss is always sub-optimal to finetuning with the meta metric loss. The performance gap between the mAP for the DMML-FT and BoT-FT after the last task is 25.8, 4.3, and 5.0 on three datasets, respectively. Note that the two losses result in a similar joint training performance. This demonstrates that meta metric learning is more suitable to the Continual Metric Learning problem, as we discussed in Sec. 3.2. For contin-

---

*We removed DukeMTMC-reID from the LReID benchmark due to its retraction on account of privacy issues.

| Metric: | mAP | | | | | | Rank-1 Accuracy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset: | Market | | MSMT17 | | VeRi-776 | | Market | | MSMT17 | | VeRi-776 | |
| Based on episodic optimization with DMML loss [6] | | | | | | | | | | | | |
| Joint training: | 82.2 | | 44.7 | | 73.6 | | 92.6 | | 68.9 | | 92.1 | |
| Sessions: | last | avg | last | avg | last | avg | last | avg | last | avg | last | avg |
| without exemplars | | | | | | | | | | | | |
| DMML-FT (ICCV'19) | 56.3 | 49.1 | 10.9 | 10.0 | 30.8 | 29.3 | 77.8 | 71.5 | 28.9 | 27.4 | 70.3 | 62.4 |
| IDA (ECCV'20) | 32.2 | 37.8 | 19.2 | 16.8 | 21.0 | 18.4 | 58.7 | 63.1 | 45.6 | 38.2 | 56.6 | 45.4 |
| DwPP | 57.8 | 48.4 | 16.3 | 13.3 | 30.9 | 28.9 | 78.1 | 70.7 | 39.0 | 33.9 | 71.7 | 63.3 |
| *Ours (DwoPP)* | **67.2** | **57.6** | **23.8** | **19.1** | **39.9** | **35.3** | **84.6** | **77.1** | **51.0** | **42.6** | **78.5** | **69.3** |
| with 500 exemplars in total | | | | | | | | | | | | |
| ERD (CVPRW'22) | 63.5 | 53.9 | 21.7 | 17.2 | 38.2 | 33.8 | 81.8 | 74.5 | 46.6 | 39.4 | 72.9 | 65.5 |
| Based on global optimization with softmax-triplet loss from BoT [41] | | | | | | | | | | | | |
| Joint training: | 82.4 | | 43.2 | | 69.2 | | 93.0 | | 71.1 | | 92.7 | |
| Sessions: | last | avg | last | avg | last | avg | last | avg | last | avg | last | avg |
| without exemplars | | | | | | | | | | | | |
| BoT-FT (CVPR'19) | 30.7 | 33.5 | 6.6 | 8.4 | 25.8 | 24.9 | 55.4 | 58.8 | 20.6 | 25.4 | 65.3 | 62.1 |
| LwF (ECCV'16) | 40.5 | 40.2 | 10.7 | 11.8 | 31.2 | 28.1 | 65.9 | 65.4 | 30.3 | 32.3 | 71.3 | 65.8 |
| PASS (CVPR'21) | 40.0 | 40.1 | 9.9 | 11.6 | 30.7 | 27.3 | 65.8 | 64.2 | 29.7 | 31.9 | 70.9 | 64.4 |
| AKA (CVPR'21) | 52.5 | 45.6 | 15.1 | 13.3 | 30.9 | 27.1 | 76.2 | 69.9 | 37.3 | 34.6 | 72.9 | 64.4 |
| with 500 exemplars in total | | | | | | | | | | | | |
| BoT-FT+ (CVPR'19) | 61.5 | 52.4 | 21.5 | 17.5 | 36.7 | 32.3 | 81.0 | 74.4 | 47.7 | 41.3 | 76.2 | 69.6 |
| iCaRL (CVPR'17) | 58.0 | 52.2 | 21.6 | 18.3 | 38.0 | 33.3 | 78.7 | 74.5 | 47.5 | 42.2 | 78.1 | 70.9 |
| LwF+ (ECCV'16) | 60.7 | 54.0 | 20.8 | 17.5 | 38.3 | 33.3 | 80.3 | 75.4 | 46.6 | 40.8 | 77.9 | 70.1 |

Table 1: Results in mAP and Rank-1 Accuracy (in %) after last task and average over all tasks. The top half reports results for meta metric learning, and the lower half for global optimization methods using the softmax-triplet loss (BoT [41]). Results are further split into methods with and without exemplars. The best exemplar-free results are highlighted in **bold**.

| | mAP | | | | | | Rank-1 Accuracy | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | market | sysu | msmt17 | cuhk03 | *seen avg.* | *unseen* | market | sysu | msmt17 | cuhk03 | *seen avg.* | *unseen* |
| BoT-FT | 11.6 | 54.6 | 0.8 | 31.2 | 24.6 | 32.4 | 31.6 | 61.6 | 2.8 | 35.1 | 32.8 | 32.8 |
| LwF | 21.0 | 58.0 | 1.7 | 48.0 | 32.2 | 43.3 | 46.5 | 64.7 | 5.8 | 53.8 | 42.7 | 42.9 |
| AKA | 18.7 | 56.3 | 1.6 | 48.6 | 31.3 | 43.6 | 42.3 | 63.1 | 5.8 | 53.9 | 41.3 | 43.6 |
| DMML-FT | 22.5 | 56.8 | 2.3 | 67.0 | 37.2 | 42.8 | 47.3 | 62.6 | 8.4 | **73.8** | 48.0 | 42.6 |
| DwPP | 23.2 | 56.7 | 2.2 | **67.9** | 37.5 | 44.7 | 49.1 | 63.2 | 7.5 | 72.4 | 48.0 | 44.2 |
| Ours (DwoPP) | **34.4** | **67.3** | **4.1** | 53.5 | **39.8** | **48.5** | **58.6** | **73.0** | **12.3** | 59.6 | **50.9** | **47.8** |

Table 2: Results after learning the last task. BoT [41] (above) and DMML [6] (below).

ual learning methods without exemplars, our method DwoPP performs best on all datasets. Compared to the DMML-FT metrics after task 10, DwoPP improves by between 9.1 to 12.9 in mAP. Note that on Market-1501 and VeRi-776 DMML-FT outperforms most of the methods that actively counter fogetting. Furthermore, we also include a comparison with rehearsal methods in Table 1. The methods iCaRL, LwF+, FT+ and ERD obtain similar results, and improved performance compared to DMML-FT. Our exemplar-free method DwoPP performs better than exemplar-based methods on these three datasets (only marginally worse in average Rank-1 Accuracy on VeRi-776). We also ablate our distillation and report results for distillation with all pairs (DwPP). The results of DwPP show that naive application of knowledge distillation to continual meta metric learning does hardly improve results. The removal of positive pairs (DwoPP) results in large performance gains after the last task: gains between 7.5 to 9.4 in mAP.

**Inter-Domain Lifelong Person ReID (LReID).** In Table 2, we compare DwoPP with other methods on the LReID [50] benchmark. Similar to the results for the intra-domain ReID setting, the DMML-FT baseline outperforms BoT-FT by a large margin for both seen and unseen tasks. Our method performs best, outperforming AKA by 8.5/9.6 (mAP/Rank-1 Accuracy) on seen tasks and 4.9/4.2 (mAP/Rank-1 Accuracy) on unseen tasks. The difference between DwPP and DwoPP on LReID further highlights the importance of removing positive pairs from knowledge distillation. See the Supplementary Material for more analysis.

(a) $\lambda$ for DwoPP      (b) $T$ for DwoPP      (c) $T$ for DwPP
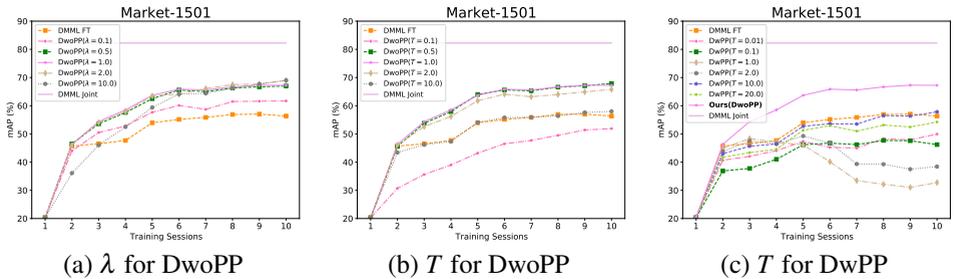
Figure 4: Ablation study on hyperparameters $\lambda$ and $T$.

|  |  | DwoPP | DKD [ ] |  |  |  | DwPP |
|---|---|---|---|---|---|---|---|
|  | $\alpha$ | 0.0 | 0.1 | 0.3 | 0.5 | 1.0 | 1.0 |
|  | $\beta$ | 1.0 | 0.9 | 0.7 | 0.5 | 0.0 | 1-$\rho$ |
| mAP | last | **67.2** | 62.9 | 48.2 | 36.0 | 25.9 | 32.8 |
|  | avg | **57.6** | 53.7 | 46.8 | 39.1 | 32.1 | 37.8 |

Table 3: Decoupling Eq. 6 into PPKD and NPKD with coefficients $\alpha$ and $\beta$ on Market-1501 with temperature $T = 1.0$. $\rho$ is the positive probabilities as in DKD [ ].

**Influence of positive pairs on distillation.** To better understand the role of positive pairs (PP) and negative pairs (NP) in knowledge distillation, we decouple the knowledge distillation (following DKD [ ]) from Eq. 6 into PPKD and NPKD by $\mathcal{L}_{DwPP*} = \alpha * PPKD + \beta * NPKD, \alpha + \beta = 1.0$ (here we use $T = 1.0$). Note that $\mathcal{L}_{DwPP} = PPKD + \rho * NPKD$ (see Supplementary Material for further explanations). In Table 3, we observe that the performance drastically decreases with higher participation of positive pairs.

**Ablation on $\lambda$ in DwoPP and temperature $T$ in both DwoPP and DwPP.** In Fig. 4(a) we vary $\lambda$ which controls the tradeoff between metric and distillation losses. Except for $\lambda = 10.0$ and $\lambda = 0.1$, DwoPP performance is stable to changing $\lambda$. We set $\lambda = 1.0$ for DwoPP in all experiments. In Fig. 4(b), we vary the temperature hyperparameter $T$ in DwoPP. A high temperature smooths the distribution and decreases the influence of the dominant class. For DwoPP $T = 10.0$ performs similarly to finetuning, and $T = 0.1$ causes the model to focus only on the highest probability. Thus we set $T = 1.0$ for DwoPP. In Fig. 4(c) we vary the temperature $T$ in DwPP to determine if larger temperatures benefit it. However, even with the best $T = 10.0$, DwPP performs similarly to DMML-FT and much worse than DwoPP. Again showing that naive knowledge distillation does not improve results for continual meta metric learning. We use $T = 10$ for DwPP in all experiments.

# 5   Conclusions

We demonstrate that meta learning approaches perform better than those based on global metric loss optimization for Object ReID. We therefore proposed an approach based on Continual Meta Metric Learning. To overcome forgetting, we propose Distillation without Positive Pairs (DwoPP) as an approach that eliminates positive samples from distillation. This distillation makes the metric learning model accumulate knowledge from the previous and current tasks and generalize better to unseen tasks. Extensive experiments on newly proposed intra-task object re-identification datasets and the existing LReID benchmark demonstrate the effectiveness of our approach. Furthermore, experiments confirm that naive knowledge distillation does not improve results for continual meta metric learning, and only after the removal of positive pairs is forgetting of previous tasks effectively countered.

# Acknowledgement

# References

[1] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *European Conference on Computer Vision*, pages 139–154, 2018.

[2] Yan Bai, Jile Jiao, Wang Ce, Jun Liu, Yihang Lou, Xuetao Feng, and Ling-Yu Duan. Person30k: A dual-meta generalization network for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2123–2132, 2021.

[3] Sungyong Baik, Janghoon Choi, Heewon Kim, Dohee Cho, Jaesik Min, and Kyoung Mu Lee. Meta-learning with task-adaptive loss function for few-shot learning. In *Proceedings of the International Conference on Computer Vision*, pages 9465–9474, 2021.

[4] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved few-shot visual classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14493–14502, 2020.

[5] Guangyi Chen, Jiwen Lu, Ming Yang, and Jie Zhou. Spatial-temporal attention-aware learning for video-based person re-identification. *IEEE Transactions on Image Processing*, 28(9):4192–4205, 2019.

[6] 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.

[7] Wei Chen, Yu Liu, Weiping Wang, Tinne Tuytelaars, Erwin M Bakker, and Michael Lew. On the exploration of incremental learning for fine-grained image retrieval. *Proceedings of the British Machine Vision Conference*, 2020.

[8] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang. Meta-baseline: Exploring simple meta-learning for few-shot learning. In *Proceedings of the International Conference on Computer Vision*, pages 9062–9071, 2021.

[9] Seokeon Choi, Taekyung Kim, Minki Jeong, Hyoungseob Park, and Changick Kim. Meta batch-instance normalization for generalizable person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3425–3435, 2021.

[10] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 539–546. IEEE, 2005.

[11] Yongxing Dai, Xiaotong Li, Jun Liu, Zekun Tong, and Ling-Yu Duan. Generalizable person re-identification with relevance-aware mixture of experts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16145–16154, 2021.

[12] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017.

[14] Alex Gomez-Villa, Bartlomiej Twardowski, Lu Yu, Andrew D Bagdanov, and Joost van de Weijer. Continually learning self-supervised representations with projected functional regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3867–3877, 2022.

[15] 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.

[16] Tyler L Hayes, Kushal Kafle, Robik Shrestha, Manoj Acharya, and Christopher Kanan. Remind your neural network to prevent catastrophic forgetting. In *European Conference on Computer Vision*, 2020.

[17] 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.

[18] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. *Proceedings of the International Conference on Computer Vision*, 2021.

[19] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *Advances in Neural Information Processing Systems*, 2014.

[20] 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.

[21] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In *International workshop on similarity-based pattern recognition*, pages 84–92. Springer, 2015.

[22] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[23] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 831–839, 2019.

[24] Pirazh Khorramshahi, Amit Kumar, Neehar Peri, Sai Saketh Rambhatla, Jun-Cheng Chen, and Rama Chellappa. A dual-path model with adaptive attention for vehicle re-identification. In *Proceedings of the International Conference on Computer Vision*, 2019.

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

[26] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[27] Kai Li, Yulun Zhang, Kunpeng Li, and Yun Fu. Adversarial feature hallucination networks for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 13470–13479, 2020.

[28] 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.

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

[30] 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.

[31] Yulin Li, Jianfeng He, Tianzhu Zhang, Xiang Liu, Yongdong Zhang, and Feng Wu. Diverse part discovery: Occluded person re-identification with part-aware transformer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2898–2907, 2021.

[32] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2017.

[33] Jinlu Liu, Liang Song, and Yongqiang Qin. Prototype rectification for few-shot learning. *European Conference on Computer Vision*, 2020.

[34] Qing Liu, Orchid Majumder, Alessandro Achille, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Incremental few-shot meta-learning via indirect discriminant alignment. In *European Conference on Computer Vision*, pages 685–701. Springer, 2020.

[35] Xialei Liu, Marc Masana, Luis Herranz, Joost Van de Weijer, Antonio M Lopez, and Andrew D Bagdanov. Rotate your networks: Better weight consolidation and less catastrophic forgetting. In *ICPR*, 2018.

[36] Xialei Liu, Chenshen Wu, Mikel Menta, Luis Herranz, Bogdan Raducanu, Andrew D Bagdanov, Shangling Jui, and Joost van de Weijer. Generative feature replay for class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 226–227, 2020.

[37] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2016.

[38] Vincenzo Lomonaco, Lorenzo Pellegrini, Andrea Cossu, Antonio Carta, Gabriele Graffieti, Tyler L Hayes, Matthias De Lange, Marc Masana, Jary Pomponi, Gido M Van de Ven, et al. Avalanche: an end-to-end library for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3600–3610, 2021.

[39] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle re-identification. *IEEE Transactions on Image Processing*, 28(8): 3794–3807, 2019.

[40] 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.

[41] 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.

[42] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018.

[43] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *European Conference on Computer Vision*, pages 67–82, 2018.

[44] Marc Masana, Xialei Liu, Bartlomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost van de Weijer. Class-incremental learning: survey and performance evaluation. *arXiv preprint arXiv:2010.15277*, 2020.

[45] Marc Masana, Tinne Tuytelaars, and Joost Van de Weijer. Ternary feature masks: zero-forgetting for task-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3570–3579, 2021.

[46] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24. Elsevier, 1989.

[47] Hao Ni, Jingkuan Song, Xiaopeng Luo, Feng Zheng, Wen Li, and Heng Tao Shen. Meta distribution alignment for generalizable person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2487–2496, 2022.

[48] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[49] Francesco Pelosin, Saurav Jha, Andrea Torsello, Bogdan Raducanu, and Joost van de Weijer. Towards exemplar-free continual learning in vision transformers: an account of attention, functional and weight regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3820–3829, 2022.

[50] 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.

[51] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

[52] 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.

[53] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. *International Conference on Machine Learning*, 2018.

[54] Edgar Simo-Serra, Eduard Trulls, Luis Ferraz, Iasonas Kokkinos, Pascal Fua, and Francesc Moreno-Noguer. Discriminative learning of deep convolutional feature point descriptors. In *Proceedings of the IEEE international conference on computer vision*, pages 118–126, 2015.

[55] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 2017.

[56] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 719–728, 2019.

[57] Jong-Chyi Su, Subhransu Maji, and Bharath Hariharan. When does self-supervision improve few-shot learning? In *European Conference on Computer Vision*, pages 645–666. Springer, 2020.

[58] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

[59] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial intelligence review*, 18(2):77–95, 2002.

[60] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049, 2017.

[61] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7):926–930, 2018.

[62] Kai Wang, Joost van de Weijer, and Luis Herranz. Acae-remind for online continual learning with compressed feature replay. *Pattern Recognition Letters*, 150:122–129, 2021.

[63] Kai Wang, Xialei Liu, Andrew D. Bagdanov, Luis Herranz, Shangling Jui, and Joost van de Weijer. Incremental meta-learning via episodic replay distillation for few-shot image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3729–3739, June 2022.

[64] Kai Wang, Xialei Liu, Andrew D Bagdanov, Luis Herranz, Shangling Jui, and Joost van de Weijer. Incremental meta-learning via episodic replay distillation for few-shot image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3729–3739, 2022.

[65] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53 (3):1–34, 2020.

[66] 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.

[67] Chenshen Wu, Luis Herranz, Xialei Liu, Joost van de Weijer, Bogdan Raducanu, et al. Memory replay gans: Learning to generate new categories without forgetting. *Advances in Neural Information Processing Systems*, 31:5962–5972, 2018.

[68] Yue Wu, Yinpeng Chen, Lijuan Wang, Yuancheng Ye, Zicheng Liu, Yandong Guo, and Yun Fu. Large scale incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 374–382, 2019.

[69] 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.

[70] Shipeng Yan, Jiangwei Xie, and Xuming He. Der: Dynamically expandable representation for class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3014–3023, 2021.

[71] Shuo Yang, Lu Liu, and Min Xu. Free lunch for few-shot learning: Distribution calibration. *International Conference on Learning Representations*, 2021.

[72] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[73] Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6982–6991, 2020.

[74] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International Conference on Machine Learning*, pages 3987–3995. JMLR. org, 2017.

[75] Xiao Zhang, Yixiao Ge, Yu Qiao, and Hongsheng Li. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3436–3445, 2021.

[76] Bo Zhao, Shixiang Tang, Dapeng Chen, Hakan Bilen, and Rui Zhao. Continual representation learning for biometric identification. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 1198–1208, 2021.

[77] Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, and Jiajun Liang. Decoupled knowledge distillation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11953–11962, 2022.

[78] 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.

[79] Jianan Zhao, Fengliang Qi, Guangyu Ren, and Lin Xu. Phd learning: Learning with pompeiu-hausdorff distances for video-based vehicle re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2225–2235, 2021.

[80] 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.

[81] 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.

[82] Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5871–5880, 2021.