HAL open science

Source-Guided Similarity Preservation for Online Person Re-Identification

Hamza Rami, Jhony H. Giraldo, Nicolas Winckler, Stéphane Lathuilière

To cite this version:

Hamza Rami, Jhony H. Giraldo, Nicolas Winckler, Stéphane Lathuilière. Source-Guided Similarity Preservation for Online Person Re-Identification. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Jan 2024, Waikoloa, United States. pp.1700-1709, 10.1109/WACV57701.2024.00173 . hal-04541703

HAL Id: hal-04541703

https://hal.science/hal-04541703v1

Submitted on 11 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Source-Guided Similarity Preservation for Online Person Re-Identification

Hamza Rami[1,2], Jhony H. Giraldo[1], Nicolas Winckler[2], Stéphane Lathuilière[1]
[1]LTCI, Télécom Paris, Institut Polytechnique de Paris.
[2]Atos.
{hamza.rami, jhony.giraldo, stephane.lathuiliere}@telecom-paris.fr, nicolas.winckler@atos.net

Abstract

Online Unsupervised Domain Adaptation (OUDA) for person Re-Identification (Re-ID) is the task of continuously adapting a model trained on a well-annotated source-domain dataset to a target domain observed as a data stream. In OUDA, person Re-ID models face two main challenges: catastrophic forgetting and domain shift. In this work, we propose a new Source-guided Similarity Preservation (S2P) framework to alleviate these two problems. Our framework is based on the extraction of a support set composed of source images that maximizes the similarity with the target data. This support set is used to identify feature similarities that must be preserved during the learning process. S2P can incorporate multiple existing UDA methods to mitigate catastrophic forgetting. Our experiments show that S2P outperforms previous state-of-the-art methods on multiple real-to-real and synthetic-to-real challenging OUDA benchmarks.

1. Introduction

Person Re-Identification (*Re-ID*) is the task of recognizing a person of interest (*i.e.*, query) across a set of images taken by non-overlapping cameras (gallery) [43]. Person Re-ID has attracted a lot of interest because of the rising need for public safety and intelligent surveillance systems. Recently, the accuracy of Re-ID models has significantly improved when using supervised deep learning [39]. However, the performance of these approaches drastically decreases when they are deployed in data that visually differ from the training dataset [30]. Since collecting data for every new environment is not practical, previous studies have introduced Unsupervised Domain Adaptation (*UDA*) for person Re-ID [13, 14, 38].

UDA methods combine a well-annotated dataset (*source domain*) and an unlabeled dataset corresponding to the *target domain*, aiming to train a model that can perform well in the new environment. Despite progress in recent years [13, 14], UDA for person Re-ID suffers from three main issues that prevent its practical use. First, when collecting the target data required to adapt the model, images are generally gathered as a stream that continually sends photos from various cameras/locations. Consequently, collecting a large target dataset may take time and delay deployment. In addition, in UDA, the model is frozen after deployment and does not benefit from the new data, which are continuously captured. Finally, numerous countries have adopted privacy regulations that forbid technology providers to store images of individuals. Thus, collecting a large target dataset is not possible.

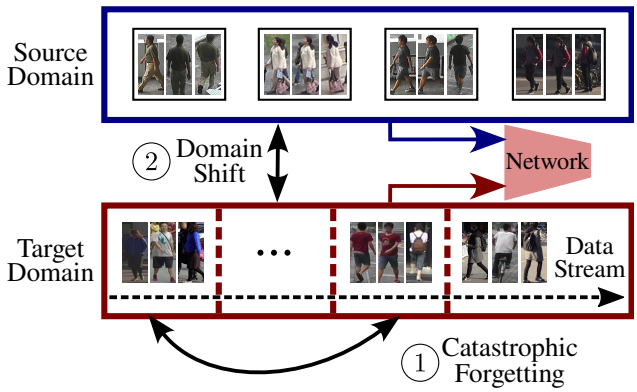


Figure 1. In OUDA for person Re-ID, the images of the target domain are available as a stream of data, and past images cannot be stored. Two main challenges should be addressed: 1) catastrophic forgetting and 2) domain shift.

Since deploying algorithms that conform with policies of data privacy protection has become a legal obligation in a growing number of countries, the Online Unsupervised Domain Adaptation for person Re-Identification (*OUDA-Rid*) setting was introduced in [32] to address the limitations of traditional UDA techniques. In the OUDA-Rid framework, we operate under the assumption that we have access to annotated source data as well as unlabeled target data. However, in contrast to traditional UDA settings, the target dataset is treated as an online stream of data, aligning with the constraint that camera-captured images

cannot be stored. In addition to complying with privacy-protection regulations, this setting also enables the person Re-ID model to be continuously updated as new target data becomes available, thereby improving its adaptability to changes in the target domain.

The performance of existing UDA methods for OUDA-Rid shows significant drops in performance regarding the offline setting [32]. This drop can be explained by the two main difficulties of OUDA-Rid illustrated in Fig. 1: catastrophic forgetting and domain shift. Catastrophic forgetting appears when the model only observes a few target identities, and consequently, the model tends to forget previously learned identities. Domain shift is a change in the data distribution between the source and target domains. Addressing the domain shift is especially challenging in the online setting since, at every training step, we observe only a small and possibly biased subset of the target domain.

In this work, we consider that these two difficulties must be addressed jointly since mitigating catastrophic forgetting can lead to target representations that better capture the full target distribution, and consequently facilitate source-target distribution alignment. We introduce a unified Source-guided Similarity Preservation (*S2P*) framework for OUDA-Rid that addresses these two challenges jointly. We take inspiration from *replay-based* strategies [1, 3] to introduce a Knowledge Distillation (*KD*) mechanism. By transferring the knowledge acquired with a teacher model to a student model, the KD [19] method enables the learning of more robust and generalizable features. However, unlike existing replay-based approaches, we do not store any target image to conform to the *privacy protection* requirement. To this end, we extract a support set composed of source images that are similar to the previously seen images of the target. This support is thus used to regularize the learning process and alleviate catastrophic forgetting. Our framework combines both explicit source-target distribution alignment and pseudo-labeling to address domain shift. S2P can easily integrate almost any existing UDA approaches [13, 14] and readily outperforms all state-of-the-art methods for OUDA-Rid in several challenging conditions in real-to-real and synthetic-to-real tasks. Our main contributions can be summarized as follows:

- We introduce a novel S2P algorithm that uses source-guided similarity preservation to jointly alleviate the *catastrophic forgetting* and *domain shift* while respecting the *privacy protection* requirements.
- S2P can easily incorporate almost any existing UDA approach. In particular, we present the integration of the *MMT* [13], *SpCL* [14] and *IDM* [5] methods into our framework, which achieve remarkable results in the UDA setting.
- We perform extensive experiments[1] in real-to-real and synthetic-to-real OUDA tasks with four datasets. S2P readily improves previous state-of-the-art UDA methods for OUDA-Rid. A set of ablation studies validate each component of our algorithm.

[1]Code available: https://github.com/ramiMMhamza/S2P

2. Related Work

UDA for person Re-ID. Existing methods can be divided into *domain translation-based* and *pseudo-labeling*.

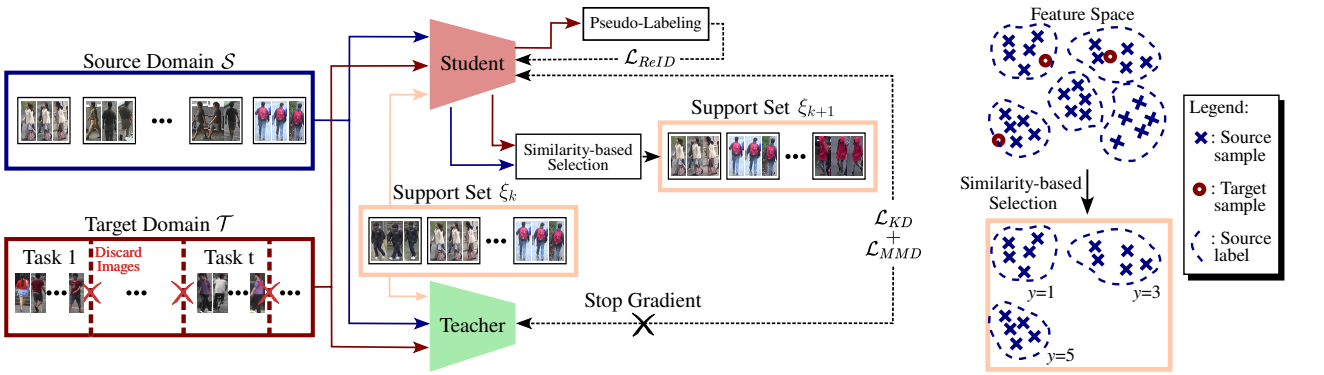
Domain translation-based methods [4, 15, 29] modify the source domain images to resemble the appearance of the target set with style transfer approaches [51]. Recent research focuses on enhancing translation by preserving self-similarity [8] or performing camera-specific translation [49].

Pseudo-labeling methods employ an iterative process alternating between clustering and fine-tuning [6, 10, 28, 33, 42]. Fan *et al.* [9] proposed a simple and effective baseline where the Re-ID model is fine-tuned using cluster indices as labels. Several studies have expanded on this framework, such as self-similarity grouping [12], Mutual-Mean Teaching (*MMT*) [13], and Self-paced Contrastive Learning (*SpCL*) [14]. MMT adopts a teacher-student framework where two student networks are jointly trained using pseudo-labels generated by themselves and soft pseudo-labels generated by their mean-teacher networks. On the other hand, SpCL takes a different approach by gradually constructing more reliable clusters to refine a hybrid memory containing both source and target images. More recently, the use of an Intermediate Domain Module (IDM) [5] has also been explored as means to bridge the gap between source and target domains.

We adopt the pseudo-labeling framework as it has outperformed previous techniques in almost all datasets [13, 14] and avoids the computational overhead of transfer-based methods. Our S2P overall framework can incorporate existing pseudo-labeling methods toward better performance in the OUDA setting.

Lifelong learning for Re-ID. Lifelong learning, also called Continual Learning (*CL*) or incremental learning [21, 22, 35], is a field that aims at developing adaptive agents, like the way humans learn throughout their lifetime. The main problem of CL is *catastrophic forgetting*, meaning that the model tends to forget the previously acquired knowledge. Recently, several methods have been developed to solve this issue in typical vision tasks [2, 46, 50]. We can categorize existing lifelong learning approaches into three main categories: 1) teacher-student [27, 41], 2) regularization [44], and 3) replay methods [40].

Few studies have tackled the problem of lifelong learning in person Re-ID. For instance, Pu *et al.* [31] proposed an Adaptive Knowledge Accumulation (*AKA*) framework, which is fully supervised. AKA addresses the domain-incremental setting where each task corresponds to a dif-



(a) Overall Pipeline of S2P.

(b) Construction of the support set via similarly-based selection.

Figure 2. The pipeline of S2P. a) S2P incorporates knowledge distillation $\mathcal{L}_{KD}$, discrepancy $\mathcal{L}_{MMD}$ loss functions, and a teacher model to mitigate the catastrophic forgetting and domain-shift problems. b) Our algorithm employs a similarity-based selection to construct the support set ξ_k from the source domain that maximizes the similarity with the target images.

ferent target domain. Huang *et al.* [20] also adopt an incremental scenario, although storing images from the previous task is permitted.

While previous methods in continual learning for person Re-ID, such as [20, 31], have adopted a less restrictive setting that allows keeping images from previous tasks, we follow the more challenging and privacy-preserving OUDA-Rid setting proposed in [32]. To address domain shift and catastrophic forgetting in OUDA-Rid, we introduce two key technical contributions: a source-guided knowledge distillation strategy and an explicit domain alignment. Gong *et al.* [16] introduced a technique based on landmarks that is similar to our support set selection. However, these landmarks were proposed to solve the domain gap in the context of UDA with classical machine learning techniques, while we have to also consider the catastrophic forgetting problem in OUDA-Rid using end-to-end deep learning models.

3. Source-Guided Similarity Preservation

OUDA-Rid problem definition. In OUDA-Rid, we assume having access to a well-annotated source domain dataset $\mathcal{S} = \{(\boldsymbol{x}_i^S, \boldsymbol{y}_i^S)\}_{i=1}^{Ns}$, and an unlabeled target domain dataset $\mathcal{T} = \{\boldsymbol{x}_i^t\}_{i=1}^{Nt}$. Here, both domain images are not necessarily drawn from the same distribution. We consider that we have access to the target domain dataset in the form of an ongoing stream of data. Following the common batch-based approximation of the online learning setting [11], we consider that we observe a sequence of N_T tasks $\{\mathcal{T}^1 \cup \mathcal{T}^2 \cup ... \cup \mathcal{T}^{N_T}\}$. Each task $\mathcal{T}^k, 1 \leq k \leq N_T$ is a set of images captured by several cameras and depicting an unknown number of identities. To align with practical scenarios, we consider that each identity can be observed by different cameras simultaneously. However, it is unlikely for an identity to appear at widely separated time intervals (*e.g.* different days). Therefore we can assume that identities do not overlap across different tasks, although this assumption is not strictly required in our approach.

In the rest of this section, we present our S2P framework to alleviate the two major challenges of the OUDA setting: catastrophic forgetting and domain shift. First, our framework integrates a teacher model that distills previously acquired knowledge. The KD strategy of S2P is based on feature space similarity preservation and only requires images from the source domain, hence respecting the *privacy protection* norms. Second, we minimize the discrepancy between the source domain and the target domain to reduce the domain shift and further enhance the stability of the S2P.

3.1. Overview of the Approach

Fig. 2 shows the pipeline of our S2P framework. In every task of the OUDA-Rid problem, the target labels are not available and we assume that the identities are different even if our S2P does not strictly require this assumption. Furthermore, we construct a *support set* that plays the role of a memory bank for *source-guided knowledge distillation*. We could keep a few samples from previous tasks if there were no privacy constraints. However, in S2P the support set only includes images from the source domain. We choose those images based on their similarities to previously seen images, ensuring a good approximation of the previously learned feature spaces during continual learning.

In this work, we follow an overall training scheme that was adopted by multiple UDA methods for Re-ID [13, 14]. More concretely, we use a student model that consists of a feature extractor $\mathcal{F}(\cdot)$. First, the student model is pretrained on source data $\mathcal{S}$, and then fine-tuned on the unlabeled target data $\mathcal{T}$ with three different loss functions:

- $\mathcal{L}_{KD}$: the knowledge distillation loss in the feature similarity space is proposed to preserve the previously ac-

quired knowledge. To this end, a *similarity-based selection* strategy is applied to the source domain to construct the support set, and a teacher model $\bar{\mathcal{F}}(\cdot)$ is added to the main pipeline (Sec. 3.2).

- $\mathcal{L}_{MMD}$: the Maximum Mean Discrepancy (*MMD*) loss is minimized to reduce explicitly the domain shift. In other words, we want to construct a feature space that is domain invariant and can regroup features from both the source and the target domains (Sec. 3.3).
- $\mathcal{L}_{ReID}$: this loss corresponds to the loss of the UDA method that is integrated into our framework. This loss is jointly minimized on the source domain $\mathcal{S}$ and the target domain images $\mathcal{T}$ together with their pseudo-labels. The pseudo labels are estimated by a clustering algorithm assigning each image to a cluster label (Sec. 3.4).

3.2. Source-Guided Knowledge Distillation

When learning a new task $\mathcal{T}^k$, the model must be updated to better discriminate the appearance of the new individuals. However, the model should also preserve the knowledge acquired on previous tasks $\mathcal{T}^i \; \forall \, 1 \leq i \leq k-1$. Therefore, we employ a teacher model that progressively distills the knowledge to the student model. Distillation is performed in the feature space over a set of source-based support images. Since target images cannot be stored, we propose to use images from the source domain as the support set. More precisely, we select images that are similar to the images from the target domain seen in previous tasks. This solution encourages the student model to project the images into a common feature space, resulting in more discriminant and task-invariant representations.

Support set collection. Fig. 2 (b) depicts the construction of the support set in S2P. We construct the support set based on the cosine similarity in the feature space between the current target images and the source domain images. For each image $\boldsymbol{x}^t$ in the target task $\mathcal{T}^k$, we identify the image $\xi_x(\boldsymbol{x}^t)$ and its corresponding identity label $\xi_y(\boldsymbol{x}^t)$ from the source domain that maximizes the cosine similarity in the feature space:

$$\big(\xi_x(\boldsymbol{x}^t), \xi_y(\boldsymbol{x}^t)\big) = \underset{(\boldsymbol{x}^s, \boldsymbol{y}) \in \mathcal{S}}{\operatorname{argmax}} \frac{\mathcal{F}(\boldsymbol{x}^s) \cdot \mathcal{F}(\boldsymbol{x}^t)}{\|\mathcal{F}(\boldsymbol{x}^s)\| \|\mathcal{F}(\boldsymbol{x}^t)\|}. \tag{1}$$

Then, we add to the support set all the images from the source that correspond to the selected identity $\xi_y(\boldsymbol{x}_t)$:

$$\xi_k = \bigcup_{\boldsymbol{x}^t \in \mathcal{T}^k} \{(\boldsymbol{x}^s, \boldsymbol{y}) \in \mathcal{S}, \boldsymbol{y} = \xi_y(\boldsymbol{x}^t)\}. \tag{2}$$

While learning a new task $\mathcal{T}^{k+1}$, ξ_k is used as a memory that best approximates previously seen images.

Teacher-student framework. As a teacher, we need a model that has accumulated knowledge from previous tasks and can effectively guide the student's learning on a new task. We use the Exponential Moving Average (*EMA*) parameters update [24, 34] of the current model. At every iteration i, the parameters $\bar{\boldsymbol{\theta}}_i$ of the teacher model are given by:

$$\bar{\boldsymbol{\theta}}_i = \alpha \bar{\boldsymbol{\theta}}_{i-1} + (1-\alpha)\boldsymbol{\theta}, \tag{3}$$

where $\boldsymbol{\theta}$ denotes the current parameters of the student model and $\alpha \in [0, 1)$ is the weighting factor. At the first iteration of our framework, $\bar{\boldsymbol{\theta}}_0$ is initialized using a model pretrained on the source dataset. Once the adaptation process is performed on a specific task, only the teacher is used for inference on the test set.

KD loss. Knowledge distillation commonly uses softened softmax labels from the teacher in training the student network [19, 27]. However, we argue that this formulation is not suitable for Re-ID. In classification problems, the absolute position of the samples in the feature space must be preserved to remain compatible with the learned classifiers. On the contrary, in Re-ID, we are interested only in preserving the relative distance between samples. Therefore, we employ a distillation loss that acts on similarity matrices to offer the model more freedom to adjust the position of the features in the learned space.

Assuming an input tensor $\mathbf{X}$ corresponding to a mini-batch of n images from the support set $\{\boldsymbol{x}_i\}_{i=1}^n$, we use the student network $\mathcal{F}$ to compute the feature representations $\mathbf{F} = \mathcal{F}(\mathbf{X}) \in \mathbb{R}^{n \times c}$, where c is the dimension of the feature space. Similarly, we compute the features with the teacher network $\bar{\mathbf{F}} = \bar{\mathcal{F}}(\mathbf{X}) \in \mathbb{R}^{n \times c}$. Then, we calculate the similarity matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$ and $\bar{\mathbf{S}} \in \mathbb{R}^{n \times n}$ containing the pairwise scalar product between the current features of all images in the current batch of the support set:

$$\mathbf{S} = \mathbf{F}\mathbf{F}^\mathsf{T}, \text{ and } \bar{\mathbf{S}} = \bar{\mathbf{F}}\bar{\mathbf{F}}^\mathsf{T}. \tag{4}$$

Moreover, we minimize the Frobenius norm $\|\cdot\|_F$ between the similarity matrices of the teacher and the student. The source-guided knowledge distillation loss can thus be formulated as follows:

$$\mathcal{L}_{KD}(\bar{\mathbf{S}}, \mathbf{S}) = \left\| \frac{\bar{\mathbf{S}}}{\|\bar{\mathbf{S}}\|} - \frac{\mathbf{S}}{\|\mathbf{S}\|} \right\|_F^2. \tag{5}$$

3.3. Source-Target Distribution Alignment

To achieve successful knowledge distillation over the support set, it is crucial to ensure that the selected images from the source domain are visually similar to the previously seen target images. To this end, we introduce an additional training loss that explicitly aligns the source and the target feature distribution. We use the Maximum Mean Discrepancy (MMD) loss [17] to reduce the domain shift by minimizing the discrepancy between the source and target domains. Formally, given an input batch of images $\{\boldsymbol{x}_i^s\}_{i=1}^n, \{\boldsymbol{x}_j^t\}_{j=1}^n$ coming from both $\mathcal{S}$ and $\mathcal{T}^k$, we compute the feature representations from both the teacher and

the student models: $\bar{\mathbf{B}} = (\bar{\mathbf{b}}_i)_{i=1}^n, \mathbf{B} = (\mathbf{b}_j)_{j=1}^n \in \mathbb{R}^{n\times c}$, where:

$$\bar{\mathbf{b}}_i = \bar{\mathcal{F}}(\boldsymbol{x}_i^s), \text{and } \mathbf{b}_j = \mathcal{F}(\boldsymbol{x}_j^t). \tag{6}$$

As shown in [17], assuming a positive semi-definite kernel K, the MMD loss can be empirically estimated as follows:

$$\mathcal{L}_{MMD}(\bar{\mathbf{B}}, \mathbf{B}) = \frac{1}{n^2} \sum_{i,j=1}^{n} [K(\bar{\mathbf{b}}_i, \bar{\mathbf{b}}_j) + K(\mathbf{b}_i, \mathbf{b}_j) - 2\mathrm{K}(\bar{\mathbf{b}}_i, \mathbf{b}_j)]. \tag{7}$$

We follow the common practice and employ the Gaussian kernel [26] with bandwidth parameter σ:

$$K(\bar{\mathbf{b}}_i, \mathbf{b}_j) = \exp\left(-\frac{\|\bar{\mathbf{b}}_i - \mathbf{b}_j\|^2}{2\sigma^2}\right), \tag{8}$$

where we set the bandwidth σ to the estimated variance of each minibatch as in [26].

3.4. Incorporating Pseudo-Labeling into S2P.

We now detail how we integrate three state-of-the-art pseudo-labeling-based frameworks into S2P: MMT [13], SpCL [14] and IDM [5].

MMT employs two networks $\mathcal{F}_1$ and $\mathcal{F}_2$ instead of a single feature extractor $\mathcal{F}$ as discussed above. The classifier C_1 for the feature extractor $\mathcal{F}_1$ is trained to predict the clustering labels obtained from $\mathcal{F}_2$ and vice-versa. Mean teacher networks $\bar{\mathcal{F}}_1$ and $\bar{\mathcal{F}}_2$ are introduced. In addition to the cross-entropy loss $\mathcal{L}_{ce}$, and the triplet loss $\mathcal{L}_{tri}$ introduced in the *strong baseline* [9], the two networks $\mathcal{F}_1$ and $\mathcal{F}_2$ are also optimized using a soft classification loss $\mathcal{L}_{sce}$ and a soft triplet loss $\mathcal{L}_{stri}$ with their mean networks [27]. Finally, $\mathcal{L}_{ReID}$ is a weighted sum of the four aforementioned losses. To integrate MMT into S2P, the two similarity matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are estimated using respectively $\mathcal{F}_1$ and $\mathcal{F}_2$ as student networks from a support set mini-batch. Similarly, two teacher similarity matrices $\bar{\mathbf{S}}_1$ and $\bar{\mathbf{S}}_2$ are estimated from the two mean teachers. The total knowledge-distillation loss is defined as the sum of $\mathcal{L}_{KD}(\bar{\mathbf{S}}_1, \mathbf{S}_1)$ and $\mathcal{L}_{KD}(\bar{\mathbf{S}}_2, \mathbf{S}_2)$. In the same way, $\mathcal{L}_{MMD}$ is jointly optimized on the source and the target domains in the feature spaces of both student-teacher couples $(\mathcal{F}_1, \bar{\mathcal{F}}_1)$ and $(\mathcal{F}_2, \bar{\mathcal{F}}_2)$.

SpCL adopts a contrastive training scheme in the feature space over a hybrid memory that is continually updated by the estimated pseudo-labels. The hybrid memory stores three types of feature representations: 1) the centroids for every class of the source domain, 2) the centroids for every cluster from the target domain, and 3) the feature representations of the outliers. Finally, $\mathcal{L}_{ReID}$ is a contrastive loss that jointly distinguishes classes, clusters, and unclustered instances in the feature space of the hybrid memory. For more details, the readers are referred to [14]. The integration of SpCL into our S2P is straightforward. We first add the teacher model, which is the EMA of the fine-tuned model. Then, for each new task, the support set is constructed to add $\mathcal{L}_{KD}$ and $\mathcal{L}_{MMD}$ to the S2P pipeline.

IDM is based on a module designed to generate intermediate domain representations by mixing the hidden representations of the source and target domains. Network training is regularized with additional losses, which promote diversity among the domain variables and ensure that the intermediate domain lies between the source and target domains. To integrate IDM into our S2P framework, we first add a teacher model which is obtained through EMA over the model's weights, including the IDM module. Then, during the optimization, we sum the two losses of S2P, $\mathcal{L}_{KD}$ and $\mathcal{L}_{MMD}$, to the IDM losses.

4. Experiments and Results

This section introduces the datasets used in the current work, the evaluation protocol, the implementation details, as well as the results and discussions of S2P. We compare our algorithm against four state-of-the-art approaches for UDA for person Re-ID: the *strong baseline* [9], MMT [13], SpCL [14], and IDM [5]. Finally, we perform a set of ablation studies to analyze each component of S2P, including the construction of the support set, the choice of the teacher, and the loss functions. In particular, we compare our KD loss $\mathcal{L}_{KD}$ with alternatives [36, 45] previously introduced in the literature for similar tasks.

Datasets. We evaluate S2P on four widely used person Re-ID datasets in domain adaptation:

- *Market 1501* (M) [47] has $1,501$ identities captured by six cameras. It includes $32,668$ images, with $12,936$ training images from 751 identities and $19,732$ test images from the remaining 750 identities. The official protocol matches $3,368$ query images to the test images.
- *MSMT17* (MS) [38] includes videos from 15 cameras. The training set has $32,621$ images of $1,042$ identities, while the test set comprises $11,659$ query images and $82,161$ gallery images from $3,060$ identities.
- *CUHK03* (C) [25] comprises $14,097$ photos of $1,467$ individual identities from six cameras, each identity is recorded by two cameras. It includes both manual and automatic bounding boxes. We utilize manually-annotated bounding boxes for training and testing.
- *RandPerson* (RP) [37] is a synthetic dataset containing $8,000$ identities and $1,801,816$ images. We use a subset of $132,145$ images from the original $8,000$ identities.

Evaluation protocol. We follow the experimental protocol introduced in [32]. We evaluate the performance of all methods using the standard training/testing splits proposed by the original authors for *Market 1501* and *MSMT17*. In *CUHK03*, we use a more challenging testing protocol proposed in [48], which consists of splitting the dataset into 767 and 700 identities for training and testing, respectively.

Method	MS → M		MS → C		M → MS		RP → M	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Strong Baseline [9]	$51.4_{\pm 1.8}$	$72.3_{\pm 0.5}$	$5.3_{\pm 1.2}$	$4.3_{\pm 1.9}$	$6.1_{\pm 0.1}$	$18.1_{\pm 0.3}$	$43.1_{\pm 1.3}$	$67.6_{\pm 1.6}$
MMT [13]	$65.8_{\pm 0.1}$	$83.7_{\pm 0.1}$	$32.2_{\pm 1.6}$	$32.2_{\pm 2.4}$	$15.1_{\pm 1.9}$	$36.9_{\pm 0.1}$	$58.7_{\pm 0.7}$	$77.5_{\pm 0.1}$
SpCL [14]	$53.5_{\pm 0.4}$	$76.0_{\pm 0.3}$	$15.6_{\pm 3.1}$	$15.7_{\pm 1.7}$	$14.7_{\pm 0.2}$	$36.7_{\pm 2.3}$	$50.5_{\pm 2.8}$	$72.1_{\pm 3.5}$
IDM [5]	$57.5_{\pm 0.2}$	$78.6_{\pm 0.2}$	$8.3_{\pm 0.2}$	$7_{\pm 0.3}$	$7.9_{\pm 0.5}$	$21.5_{\pm 0.1}$	$60.8_{\pm 0.2}$	$80.4_{\pm 0.1}$
S2P-MMT (ours)	$\underline{70}_{\pm 0.4}$	$\underline{87.1}_{\pm 0.4}$	$\mathbf{40.4}_{\pm 0.8}$	$\mathbf{42.4}_{\pm 0.9}$	$\underline{19.5}_{\pm 0.1}$	$\underline{43.3}_{\pm 0.7}$	$\underline{61.4}_{\pm 0.1}$	$\underline{81}_{\pm 0.2}$
S2P-SpCL (ours)	$69.1_{\pm 0.1}$	$87.1_{\pm 0.1}$	$\underline{34.3}_{\pm 0.3}$	$\underline{35.1}_{\pm 0.5}$	$\mathbf{20.2}_{\pm 0.1}$	$\mathbf{46.1}_{\pm 0.2}$	$59_{\pm 0.1}$	$80.5_{\pm 0.2}$
S2P-IDM (ours)	$\mathbf{71.3}_{\pm 0.1}$	$\mathbf{88.0}_{\pm 0.1}$	$17.5_{\pm 0.5}$	$16.6_{\pm 0.5}$	$14.2_{\pm 0.3}$	$33.9_{\pm 0.2}$	$\mathbf{70.2}_{\pm 0.2}$	$\mathbf{86.1}_{\pm 0.4}$

Table 1. Performance of S2P and four state-of-the-art methods in the last task in three real-to-real and one synthetic-to-real OUDA-Rid tasks. The best and second-best methods on each dataset are highlighted in **bold** and underlined, respectively.

RP is always used as a source dataset in this work.

We evaluate S2P for OUDA-Rid in several real-to-real and synthetic-to-real configurations: MS→M, MS→C, M→MS, and RP→M. These configurations are widely used in the literature [13, 14, 37] and illustrate domain shifts of diverse difficulties. For each configuration, we randomly and uniformly split the training identities into five subsets, corresponding to five tasks for OUDA-Rid, each having a distinct set of identities. We also perform additional experiments where we increase the number of tasks in the target domain, which are detailed in the supplementary material due to space limitations.

We adopt the commonly used metrics for evaluation in Re-ID [13, 14]: mean Average Precision (mAP) and CMC Rank-1 [47] accuracies. These metrics are computed on the entire test set of the target domain after each task during the online adaptation process. We report the average mAP and Rank-1 over three repetitions with different seeds.

Implementation details. We follow the common practices in the UDA person Re-ID field by adopting ResNet50 [18] pre-trained on ImageNet [7] as a backbone. We employ the features computed after the global average pooling layer. We use DBSCAN for clustering, which is commonly employed in pseudo-labeling methods because it requires no prior assumption on the number of clusters. For each new task, Adam [23] optimizer is adopted with an initial learning rate (LR) equal to $3.5e{-}4$, a linear LR scheduler, and weight decay of $5e{-}4$ [13, 14]. Same as [32], the number of epochs per task is set to 20. For the EMA, we follow [13] and set α to 0.999 to update the teacher model parameters. Finally, all the images are resized to 256×128 before being fed into the backbone (or backbones for MMT), and the batch size was set to 64 corresponding to 16 different identities with 4 images per ID.

4.1. Quantitative Results

Comparison with the state of the art. Table 1 reports the mAP accuracy and CMC Rank-1 score obtained at the end of training with all methods in three *real-to-real* configurations: MS→M, MS→C, M→MS, and one *synthetic-to-real* RP→M. The reported metrics are computed at the end of the adaptation process in each case. The low scores of the *strong baseline* are due to the presence of the domain shift, which cannot be appropriately addressed with this method. The state-of-the-art UDA methods MMT, SpCL and IDM struggle when deployed in the OUDA-Rid setting. The drop in performances of MMT, SpCL and IDM is partially explained by the presence of catastrophic forgetting. Furthermore, MMT outperforms SpCL and IDM in almost all configurations, showing that their student-teacher framework is well suited to OUDA-Rid.

Table 1 shows that S2P-MMT, S2P-SpCL or S2P-IDM outperforms all previous state-of-the-art UDA methods in OUDA-Rid over all configurations. For example, S2P improves the mAP of SpCL from 15.6 to 34.3 and from 14.7 to 20.2 in MS→C and M→MS, respectively.

As for IDM, our S2P significantly improves its performances, from 8.3 to 17.5 and from 7.9 to 14.2, in the same configurations: MS→C and M→MS. Finally, for MMT, S2P improves the mAP, from 32.2 to 40.4 and from 15.1 to 19.5, in MS→C and M→MS, respectively. The gain for SpCL and IDM is greater than for MMT because MMT already integrates a teacher in its knowledge distillation loss function (soft cross entropy and soft triplet loss), whereas SpCL and IDM are only optimized using hard pseudo labels without any refinement.

Similarly, we can see that in the *synthetic-to-real* scenario RP→M, S2P noticeably improves the performance of the three state-of-the-art methods. S2P improves: from 58.7 to 61.4, from 50.5 to 59, and from 60.8 to 70.2 the performances of MMT, SpCL, and IDM, respectively. These results demonstrate that S2P can be successfully deployed in OUDA-Rid applications where we cannot have access to a real and well-annotated dataset for the source domain[2].

Continual behavior. To delve deeper into the analysis on the continual behavior of the different methods, we com-

[2]Additional experiments in different configurations can be found in the supplementary materials.

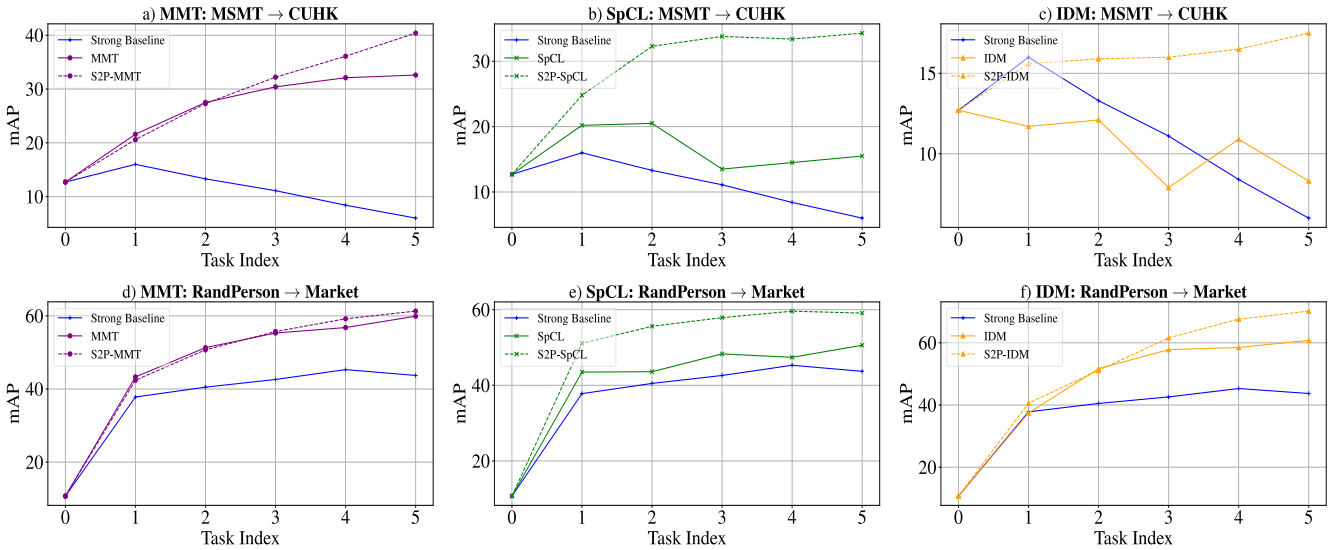


Figure 3. Comparison of S2P with four state-of-the-art methods in terms of mAP vs. task index in two different OUDA-Rid tasks, MSMT→CUHK and RandPerson→Market.

pare in Fig. 3 the mAP at the end of each task before and after incorporating the three state-of-the-art methods MMT, SpCL and IDM into our S2P framework. For this analysis, we choose two different configurations: MS→C (Fig. 3-a, -b, and -c) and RP→M (Fig. 3-d, -e and -f). In general, the low performances of the direct inference (*i.e.* the mAP at task 0) and the *strong baseline* show that the chosen configurations are of varying degrees of difficulty.

Fig. 3 also shows the effect of catastrophic forgetting as a drop in performance in new tasks in several situations. For example, the *strong baseline* presents degradation of performance in both configurations in new tasks. Similarly, SpCL and IDM both lose accuracy when confronted with new incoming data due to catastrophic forgetting and domain shift in the later tasks. For MS→C configuration: in b) the mAP of SpCL goes from 20.5 in the second task to 13.5 in the third task, while in c) the performance of IDM drops from 10.9 to 8.3 in the last task. Finally, for MMT we can notice in a) that the performance reaches an undesirable plateau after the third task in the same configuration. This shows that the knowledge acquired during the first stages of OUDA-Rid is lost during the adaptation process. Furthermore, the fluctuations of the mAP of SpCL and IDM in b), c), e) and f) in Fig. 3 illustrate the inability of the models to maintain a general structure of the feature space that captures the whole target domain distribution.

On the contrary, S2P-MMT, S2P-SpCL and S2P-IDM show a steady improvement in performance on the two configurations. Specifically, all the three methods achieve better performance when learning later tasks when incorporated into our S2P framework and deliver consistent results across the different configurations.

Moreover, it is clear from the learning curves across all the different tasks that S2P successfully adapts UDA methods to the continual setting OUDA-Rid, resulting in a superior learning process evolution and a solid accumulation of prior knowledge.

4.2. Ablation Studies

We perform three ablation studies about: 1) the loss functions, 2) the knowledge distillation design, and 3) the choice of the teacher model. We run those experiments with S2P-SpCL as the pseudo-labeling method in OUDA-Rid configurations, namely, MS→C and RP→M.

The impact of the two main losses of S2P. The two main loss functions (KD and MMD) of S2P were introduced in Sec. 3.2 and 3.3. In this ablation, we study the influence of different configurations of the losses $\mathcal{L}_{MMD}$ and $\mathcal{L}_{KD}$ in the performance of S2P as shown in Table 2. The performance of the baseline significantly improves in almost all the configurations by only integrating either the $\mathcal{L}_{MMD}$ or $\mathcal{L}_{KD}$. For example, the configuration MS→ M shows a gain in performance. The mAP goes from 53.5 to 62.4 with $\mathcal{L}_{MMD}$ and from 53.5 to 65.1 with $\mathcal{L}_{KD}$ for S2P-SpCL. Furthermore, combining both losses leads to an additional overall improvement in performance in all cases.

Knowledge Distillation Design. We delve into our knowledge distillation mechanism focusing on two key factors: the loss function and the selection of the support set.

Regarding the support set construction, our similarity-based selection relies on a cosine similarity function ξ given in Eq. (2). We explore two different approaches to compute the support set as shown in Table 3. The first strategy employs all the images of the source domain $\mathcal{S}$ to construct

$\mathcal{L}_{MMD}$	$\mathcal{L}_{KD}$	S2P-SpCL								S2P-MMT							
		MS → M		MS → C		M → MS		RP → M		MS → M		MS → C		M → MS		RP → M	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
✗	✗	53.5	76.0	15.6	15.7	14.7	36.7	50.5	72.1	65.8	83.7	32.2	32.2	15.1	36.9	58.7	77.5
✓	✗	62.4	82.9	24.1	23.6	15.2	38.5	55.4	77.5	62.6	81.4	27.4	26.4	15.3	37	60.8	80.2
✗	✓	65.1	85.1	28.2	26.7	16	40	55.5	78.9	67	85.5	35.2	35.1	17.8	41.1	60.4	80.1
✓	✓	**69.1**	**87.1**	**34.3**	**35.1**	**20.2**	**46.1**	**59**	**80.5**	**70**	**87.1**	**40.4**	**42.4**	**19.5**	**43.3**	**61.4**	**81**

Table 2. Ablation study on the effectiveness of the $\mathcal{L}_{MMD}$ and $\mathcal{L}_{KD}$ loss functions using S2P-SpCL and S2P-MMT.

Dist. Loss	Support Set	MS → C		RP → M	
		mAP	Rank-1	mAP	Rank-1
$\mathcal{L}_{KD}$	Source Domain $\mathcal{S}$	29.3	28.1	56.3	78.3
$\mathcal{L}_{KD}$	Rank-1 NN	29.8	29.6	56.4	78.9
$\mathcal{L}_{KD}$	Similarity-based ξ	**34.3**	**35.1**	**59**	**80.5**
$\mathcal{L}_{SP}$ [36]	Similarity-based ξ	26.5	25	55.4	78.8
$\mathcal{L}_{AT}$ [45]	Similarity-based ξ	26.4	25.6	55.5	78.8

Table 3. Ablation study on the design of our knowledge distillation mechanism using S2P-SpCL. We assess the impact of two key factors: the loss function and the selection function of the support set. See text for details.

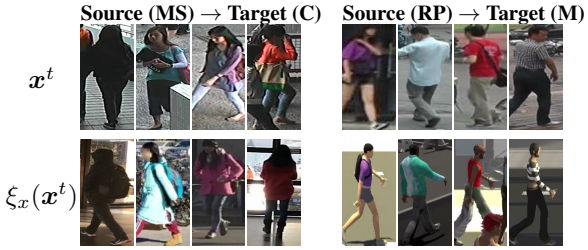


Figure 4. The support set construction based on the similarities between the source domain MS (RP respectively) and the target domain C (M respectively).

the support set. The second (Rank-1 NN) selects only the most similar image from the source domain to each previously seen image, without considering its identity's other images. The similarity-based selection strategy ξ shows the best results in almost all cases as shown in Table 3. Furthermore, we compare our $\mathcal{L}_{KD}$ with two different losses that are widely used in the literature: $\mathcal{L}_{SP}$ [36] which uses pairwise activation similarities to supervise the training of the student model, and $\mathcal{L}_{AT}$ [45] where only the activations are used to compute a mean squared error between the student and the teacher models. The results of Table 3 allow us to draw the conclusion that our knowledge distillation design better suits the setting of OUDA-Rid and outperforms both the other knowledge distillation losses and support set selection strategies.

To qualitatively illustrate the construction of our support set, in Fig. 4, we show some random samples of the support set for MS→C and RP→M, where $\boldsymbol{x}^t$ is the image in the target domain and $\xi_x(\boldsymbol{x}^t)$ is its most similar image in the source domain.

The choice of the teacher. As described in Sec. 3.2 for S2P, knowledge distillation is performed with a teacher network obtained via EMA updates. In this ablation study, we investigate alternative solutions for the choice of the teacher model as shown in Table 4. We analyze three teacher models: 1) at the start of each task t, the teacher is frozen and initialized by the weights of the fine-tuned model on the previous task $\mathcal{F}_{t-1}$; 2) the teacher is an EMA of the student model, being updated only at the end of the previously seen tasks $\bar{\mathcal{F}}_{t-1}$; and 3) the mean teacher $\bar{\mathcal{F}}$ obtained via EMA after each iteration (*i.e.*, one mini-batch pass) as in Sec 3.2. The results in Table 4 suggest that the choice of the teacher model is highly critical to alleviating the problem of catastrophic forgetting and that the proposed solution outperforms other alternatives.

Teacher Model	MS → C		RP → M	
	mAP	Rank-1	mAP	Rank-1
Task-specific $\mathcal{F}_{t-1}$	14.3	14.9	28.7	57
EMA of task-specific $\bar{\mathcal{F}}_{t-1}$	14.8	15.1	28.3	55.7
EMA of the student $\bar{\mathcal{F}}$	**34.3**	**35.1**	**59**	**80.5**

Table 4. Ablation study on the choice of the teacher model for Knowledge Distillation using S2P-SpCL.

5. Conclusions

In this paper, we introduced a new Source-guided Similarity Preservation (S2P) algorithm for the problem of Online Unsupervised Domain Adaptation for person Re-identification (OUDA-Rid). S2P jointly addresses catastrophic forgetting and domain shift with a knowledge distillation mechanism that respects data privacy regulations. This mechanism is based on a support set composed of source images similar to previously seen identities in the target dataset. We also introduced an explicit source-target distribution alignment and a pseudo-labeling strategy to alleviate the domain shift. We performed extensive experiments where S2P straightforwardly incorporates existing state-of-the-art UDA methods and consistently outperformed them by significant margins.

References

[1] Andreea Bobu, Eric Tzeng, Judy Hoffman, and Trevor Darrell. Adapting to continuously shifting domains. In *ICLR*, 2018. 2

[2] Fabio Cermelli, Dario Fontanel, Antonio Tavera, Marco Ciccone, and Barbara Caputo. Incremental learning in semantic segmentation from image labels. In *CVPR*, 2022. 2

[3] Hao Chen, Benoit Lagadec, and Francois Bremond. Unsupervised lifelong person re-identification via contrastive rehearsal. *arXiv preprint arXiv:2203.06468*, 2022. 2

[4] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Instance-guided context rendering for cross-domain person re-identification. In *ICCV*, 2019. 2

[5] Yongxing Dai, Jun Liu, Yifan Sun, Zekun Tong, Chi Zhang, and Ling-Yu Duan. IDM: an intermediate domain module for domain adaptive person re-id. In *ICCV*, 2021. 2, 5, 6

[6] Guillaume Delorme, Yihong Xu, Stéphane Lathuilière, Radu Horaud, and Xavier Alameda-Pineda. CANU-ReID: a conditional adversarial network for unsupervised person re-identification. In *ICPR*, 2021. 2

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009. 6

[8] Weijian Deng, Liang Zheng, Guoliang Kang, Yi Yang, Qixiang Ye, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, 2018. 2

[9] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM TOMM*, 2018. 2, 5, 6

[10] Hao Feng, Minghao Chen, Jinming Hu, Dong Shen, Haifeng Liu, and Deng Cai. Complementary pseudo labels for unsupervised domain adaptation on person re-identification. *IEEE TIP*, 2021. 2

[11] Enrico Fini, Stéphane Lathuiliere, Enver Sangineto, Moin Nabi, and Elisa Ricci. Online continual learning under extreme memory constraints. In *ECCV*, 2020. 3

[12] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *ICCV*, 2019. 2

[13] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *ICLR*, 2020. 1, 2, 3, 5, 6

[14] Yixiao Ge, Dapeng Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In *NeurIPS*, 2020. 1, 2, 3, 5, 6

[15] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, Xiaogang Wang, and Hongsheng Li. Structured domain adaptation with online relation regularization for unsupervised person re-id. *IEEE TNNLS*, 2022. 2

[16] Boqing Gong, Kristen Grauman, and Fei Sha. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *ICML*, 2013. 3

[17] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. *NeurIPS*, 2006. 4, 5

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 6

[19] Geoffrey Hinton, Jeff Dean, and Oriol Vinyals. Distilling the knowledge in a neural network. In *NeurIPSW*, 2014. 2, 4

[20] Zhipeng Huang, Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, and Zheng-jun Zha. Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation. In *CVPR*, 2022. 3

[21] Steven C. Y. Hung, Cheng-Hao Tu, Cheng-En Wu, Chien-Hung Chen, Yi-Ming Chan, and Chu-Song Chen. Compacting, picking and growing for unforgetting continual learning. In *NeurIPS*, 2019. 2

[22] Prakhar Kaushik, Adam Kortylewski, Alex Gain, and Alan Yuille. Understanding catastrophic forgetting and remembering in continual learning with optimal relevance mapping. In *NeurIPSW*, 2021. 2

[23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 6

[24] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. In *ICLR*, 2017. 4

[25] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 5

[26] Yujia Li, Kevin Swersky, and Rich Zemel. Generative moment matching networks. In *ICML*, 2015. 5

[27] Zhizhong Li and Derek Hoiem. Learning without forgetting. In *ECCV*, 2016. 2, 4, 5

[28] Yutian Lin, Xuanyi Dong, Liang Zheng, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, 2019. 2

[29] Jiawei Liu, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. Adaptive transfer network for cross-domain person re-identification. In *CVPR*, 2019. 2

[30] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *CVPRW*, 2019. 1

[31] Nan Pu, Wei Chen, Yu Liu, Erwin M. Bakker, and Michael S. Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *CVPR*, 2021. 2, 3

[32] Hamza Rami, Matthieu Ospici, and Stéphane Lathuilière. Online unsupervised domain adaptation for person re-identification. In *CVPRW*, 2022. 1, 2, 3, 5, 6

[33] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, 2020. 2

[34] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017. 4

[35] Cheng-Hao Tu, Cheng-En Wu, and Chu-Song Chen. Extending conditional convolution structures for enhancing multi-tasking continual learning. In *APSIPA ASC*, 2020. 2

[36] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *ICCV*, 2019. 5, 8

[37] Yanan Wang, Shengcai Liao, and Ling Shao. Surpassing real-world source training data: Random 3D characters for generalizable person re-identification. In *ACM MM*, 2020. 5, 6

[38] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer GAN to bridge domain gap for person re-identification. In *CVPR*, 2018. 1, 5

[39] Chao Wu, Wenhang Ge, Ancong Wu, and Xiaobin Chang. Camera-conditioned stable feature generation for isolated camera supervised person re-identification. In *CVPR*, 2022. 1

[40] Chenshen Wu, Luis Herranz, Xialei Liu, Yaxing Wang, Joost van de Weijer, and Bogdan C. Raducanu. Memory replay GANs: learning to generate images from new categories without forgetting. In *NeurIPS*, 2018. 2

[41] Fei Ye and Adrian G Bors. Lifelong teacher-student network learning. *IEEE TPAMI*, 2021. 2

[42] Mang Ye, Jiawei Li, Andy J Ma, Liang Zheng, and Pong C Yuen. Dynamic graph co-matching for unsupervised video-based person re-identification. *IEEE TIP*, 2019. 2

[43] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE TPAMI*, 2022. 1

[44] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. In *ICLR*, 2018. 2

[45] Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *ICLR*, 2016. 5, 8

[46] Mengyao Zhai, Lei Chen, Frederick Tung, Jiawei He, Megha Nawhal, and Greg Mori. Lifelong GAN: continual learning for conditional image generation. In *ICCV*, 2019. 2

[47] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 5, 6

[48] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, 2017. 5

[49] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero- and homogeneously. In *ECCV*, 2018. 2

[50] Wang Zhou, Shiyu Chang, Norma E. Sosa, Hendrik F. Hamann, and David D. Cox. Lifelong object detection. *arXiv preprint arXiv:2009.01129*, 2020. 2

[51] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 2