Lifelong Person Re-identification by Pseudo Task Knowledge Preservation

Wenhang Ge[1,2,3], **Junlong Du**[2], **Ancong Wu**[1,3,*], **Yuqiao Xian**[1], **Ke Yan**[2], **Feiyue Huang**[2], **Wei-Shi Zheng**[1,3,4]

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[2] Youtu Lab, Tencent,
[3] Pazhou Lab, Guangzhou, China,
[4] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
gewh@mail2.sysu.edu.cn; jeffdu@tencent.com; wuanc@mail.sysu.edu.cn; xianyq3@mail2.sysu.edu.cn; kerwinyan@tencent.com; feiyuehuang@gmail.com; wszheng@ieee.org

Abstract

In real world, training data for person re-identification (Re-ID) is collected discretely with spatial and temporal variations, which requires a model to incrementally learn new knowledge without forgetting old knowledge. This problem is called lifelong person re-identification (LReID). Variations of illumination and background for images of each task exhibit task-specific image style and lead to task-wise domain gap. In addition to missing data from the old tasks, task-wise domain gap is a key factor for catastrophic forgetting in LReID, which is ignored in existing approaches for LReID. The model tends to learn task-specific knowledge with task-wise domain gap, which results in stability and plasticity dilemma. To overcome this problem, we cast LReID as a domain adaptation problem and propose a pseudo task knowledge preservation framework to alleviate the domain gap. Our framework is based on a pseudo task transformation module which maps the features of the new task into the feature space of the old tasks to complement the limited saved exemplars of the old tasks. With extra transformed features in the task-specific feature space, we propose a task-specific domain consistency loss to implicitly alleviate the task-wise domain gap for learning task-shared knowledge instead of task-specific one. Furthermore, to guide knowledge preservation with the feature distributions of the old tasks, we propose to preserve knowledge on extra pseudo tasks which jointly distills knowledge and discriminates identity, in order to achieve a better trade-off between stability and plasticity for lifelong learning with task-wise domain gap. Extensive experiments demonstrate the superiority of our method [1] as compared with the state-of-the-art lifelong learning and LReID methods.

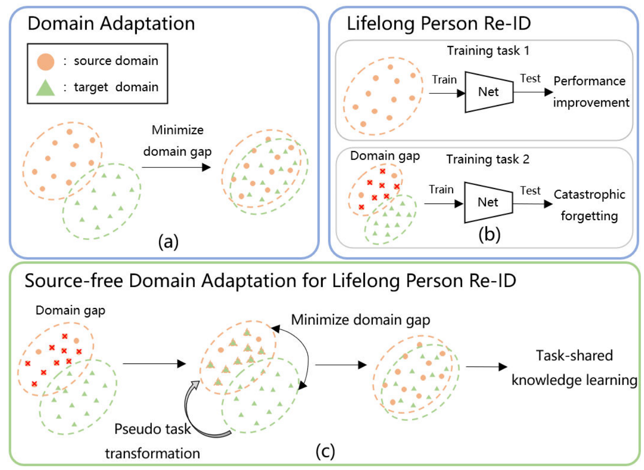

Figure 1: Comparison between domain adaptation and lifelong learning for Re-ID. (a) Domain adaptation which minimizes the domain gap with access to sufficient samples from both source and target domain. (b) Lifelong person Re-ID training with task-wise domain shift without access to sufficient source data, which results in catastrophic forgetting. (c) Our pseudo task knowledge preservation framework based on source-free domain adaptation for learning task-shared knowledge instead of task-specific one.

Introduction

Person Re-Identification (Re-ID) has been widely investigated as a person retrieval issue across non-overlapping cameras. Given a query person of interest, the purpose of person Re-ID is to retrieve the person who shares the same identity with the query in a large scale gallery. Due to the real-world application, a model is expected to learn in an incremental manner to meet a fast adaptation requirement. For example, in a large-scale and dynamic surveillance system for person Re-ID, when a new monitored area is added, extra training data will be collected along with the deployment of new cameras. Therefore, the system needs to be updated for the tasks in new domains without losing discriminative capability in learned domains. Thus, it is imperative to study Re-ID model training on sequential tasks with dynamic domain shift to accumulate knowledge incrementally and then adapt the knowledge for both seen and unseen domains.

Most of the existing Re-ID methods focus on learning knowledge with pre-prepared dataset with fixed data distribution. However, learning on sequential tasks is under-explored. To address this problem, recent works (Wu and Gong 2021; Pu et al. 2021) study lifelong person Re-ID, which has two main challenges. First, task-wise domain gap. Training data of each task collected under different background and illumination leads to task-wise domain gap, which is one of the key reasons for catastrophic forgetting

[*]Corresponding author.

[1]Code available at https://github.com/g3956/PTKP

in LReID. Second, the identities in training set and test set are non-overlapped which is different from lifelong learning of recognition problem like ImageNet (Deng et al. 2009), which is essentially formulated as an open-set problem. However, existing LReID methods (Wu and Gong 2021; Pu et al. 2021) ignore the task-wise domain gap and most lifelong learning methods (Li and Hoiem 2017; Tung and Mori 2019; Zhao et al. 2020a; Rebuffi et al. 2017; Zhao et al. 2021) focus on closed-set recognition problem. When learning new knowledge or distilling old knowledge, the model tends to learn or distill task-specific knowledge instead of task-shared knowledge due to the task-wise domain gap without access to sufficient samples from the old tasks.

To address this issue, we cast lifelong person Re-ID as a domain adaptation problem where the old tasks are treated as the source domain and the new task is treated as the target domain, we aim to alleviate the domain gap between them for learning task-shared knowledge. To this end, we propose a **P**seudo **T**asks **K**nowledge **P**reserving framework (PTKP), which consists of a pseudo task transformation module for mapping the features of the new task into the feature space of the old tasks, from which we supplement limited exemplars of the old tasks to model task-specific data distribution and mine task-shared information. To align the distinct distributions among tasks for domain adaptation, we propose a task-specific domain consistency loss. To exploit pseudo tasks beyond domain consistency learning and guide the knowledge preservation with the feature distributions of the old tasks, we propose a pseudo task knowledge distillation loss and a pseudo task identity discrimination loss. Finally, we develop a pseudo task knowledge preservation framework. The main contributions of our work are summarized as follows:

- We formulated lifelong person Re-ID as a domain adaptation problem. To supplement the missing samples of the old tasks, we design a pseudo task transformation module to map the features of the new task into the feature space of the old tasks.
- Based on the transformed pseudo tasks, we propose a task-specific domain consistency loss, pseudo task knowledge distillation loss and pseudo task identity discrimination loss which aim at alleviating task-wise domain shift and mining task-shared information.
- Extensive experiments validate the proposed framework significantly outperformed the state-of-the-art methods both in anti-forgetting capability and generalization.

Related Work

Person Re-identification

Previous person Re-ID methods mainly focus on the following five categories: (1) Fully supervised methods which are commonly based on learning distance metric (Chen et al. 2017; Xiong et al. 2014; Zheng, Gong, and Xiang 2012; Liu et al. 2019) and deep learning (Li, Zhu, and Gong 2018b; Sun et al. 2018; Chen et al. 2019; Li, Wu, and Zheng 2021). (2) Intra-camera supervised methods which either use intra-camera supervision and cross-camera association for training (Li, Zhu, and Gong 2018a; Qi et al. 2020, 2019; Wu, Zhu, and Gong 2020) or only use cross-camera unpaired data for training (Zhang et al. 2020; Ge et al. 2021). (3) Unsupervised domain adaptation methods which assume a labelled source domain and an unlabelled target domain with different data distribution or image style. Recent works (Zheng et al. 2021; Wu, Zheng, and Lai 2019; Zhao et al. 2020b; Zhang et al. 2021; Bai et al. 2021) try to mitigate the domain gap between them for domain adaptation. (4) Pure-unsupervised methods which suppose labelled source data is not available. These methods (Wang et al. 2020; Zeng et al. 2020; Lin et al. 2019) aim at learning robust discriminative representation without any annotation. (5) Domain generalization methods (Song et al. 2019; Jin et al. 2020; Chen et al. 2021) which aim to learn from source data and then generalize to target data without access to target data. Nevertheless, most of them do not address the challenge in LReID.

In this work, we formulate lifelong person Re-ID as a supervised domain adaptation problem, where we regard the old tasks as the source domain and the new task as the target domain. The goal is to alleviate the domain gap for learning task-shared knowledge instead of task-specific one.

Lifelong Re-ID

In lifelong person Re-ID, training data of each task is gradually collected over time or from different locations with task-specific image style, which requires a model to learn in a task-incremental manner with a dynamic domain shift. Recent works focus on distilling old knowledge when learning new knowledge. GwFReID (Wu and Gong 2021) exploits classification coherence, distribution coherence and representation coherence to avoid forgetting. AKA (Pu et al. 2021) designs an adaptive knowledge accumulation framework which maintains a learnable knowledge graph to adaptively update previous knowledge and transfer the knowledge to improve generalization. Both of them ignore the dynamic domain shift among tasks, with which task-specific knowledge is learned which results in forgetting.

Lifelong Learning

Most of the existing methods for lifelong learning can be categorized into two general types. First, class-incremental lifelong learning supposes that the task ID is not available at inference time, the model is expected to distinguish between all classes of all tasks. Second, task-incremental lifelong learning assumes that task ID is known at inference time, the model is expected to distinguish between classes for each single task. We focus on task-incremental lifelong learning in this work. The methods can be divided into three categories, including knowledge distillation based methods (Wu and Gong 2021; Li and Hoiem 2017; Zhao et al. 2020a), parameter isolation based methods (Aljundi, Chakravarty, and Tuytelaars 2017; Fernando et al. 2017; Serra et al. 2018) and replay based methods (Wu et al. 2018; Rebuffi et al. 2017). However, all of these methods ignore the task-wise domain gap which is a special character of LReID. With several task-wise domain gap, the model tends to learn task-specific knowledge which exacerbates the catastrophic forgetting. In our work, we treat LReID as a domain adaptation problem and mitigate the discrepancy of data distributions.

Approach

Formulation

In lifelong person Re-ID, a continual stream of person Re-ID datasets denoted by $\mathcal{D} = \{\mathbf{D^{(s)}}\}_{s=1}^{S}$ are collected under different background and illumination, which leads to task-wise domain gap. These datasets are used for training the model in sequence. $\mathbf{D^{(s)}} = \{\mathbf{D}_{train}^{(s)}, \mathbf{D}_{test}^{(s)}\}$ represents the s-th step training and corresponding test set. $\mathbf{D}_{train}^{(s)} = \{(\boldsymbol{x}_i^{(s)}, \boldsymbol{y}_i^{(s)})\}_{i=1}^{N^{(s)}}$, where $\boldsymbol{x}_i^{(s)}$ is a person image, $\boldsymbol{y}_i^{(s)}$ is corresponding identity label and $N^{(s)}$ is the number of samples. We aim at training a feature extractor $H(\cdot;\boldsymbol{\Theta})$ parameterized by $\boldsymbol{\Theta}$ and a classifier $g(\cdot;\boldsymbol{\phi})$ parameterized by $\boldsymbol{\phi}$ with datasets $\mathcal{D}$ input in sequence. The trained feature extractor is expected to perform well on all test sets $\{\mathbf{D}_{test}^{(s)}\}_{s=1}^{S}$ to validate its anti-forgetting capability. We follow a common setting where a small amount of samples from the old tasks can be saved in an exemplar memory bank. Specifically, at the s-th training step, besides new samples from $\mathbf{D}_{train}^{(s)}$, a small number of old exemplars from the old datasets $\{\mathbf{D}_{train}^{(1)}, \mathbf{D}_{train}^{(2)}, \ldots, \mathbf{D}_{train}^{(s-1)}\}$ are also available. Since there is several task-wise domain gap, we treat lifelong person Re-ID as a domain adaptation problem, how to bridge the gap is the key to handle this issue.

Task-specific Domain Consistency Learning

Each Re-ID dataset is collected discretely with spatial and temporal variations, which results in task-wise domain gap. Training on the new task will result in dramatic performance degradation on the old tasks. On the contrary, if the distributions of two tasks are very similar, training on new task will further improve the performance on the old task (Wu and Gong 2021). Hence, we regard LReID as a domain adaptation problem where the old tasks are viewed as the source domain and the new task is viewed as the target domain. We propose that the distributions of all tasks should be aligned for learning task-shared knowledge instead of task-specific one. However, in LReID, only few samples from source domains are available, which results in obstacle for bridging the domain shift. Thus, LReID is inherently formulated as a source-free domain adaptation problem. To overcome this issue, we introduce a pseudo task transformation module, then we propose a task-specific domain consistency learning objective for source-free domain adaptation.

Pseudo Task Transformation Module. Since most samples of the old tasks are missing, it is hard to model task-specific data distribution of the old tasks which further leads to obstacle for aligning distributions. To supplement the missing samples, we propose to map the features of the new task to the feature space of the old tasks by a pseudo task transformation module for mining task-shared information.

Since training set $\{\mathbf{D}_{train}^{(s)}\}_{i=1}^{S}$ of each task is collected discretely and the style of each person image is task-specific, we suppose that the training set of each task follows a task-specific Gaussian distribution denoted as $\{N(\boldsymbol{\mu}^{(s)}, {\boldsymbol{\sigma}^{(s)}}^2)\}, s \in \{1, 2, \ldots, S\}$. Meanwhile, we note that Batch Normalization (BN) layer can estimate moment statistics of feature distribution (e.g., mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\theta}$), we choose S task-specific batch normalization (TSBN) layers denoted by $\{\mathrm{TSBN}^{(s)}\}_{s=1}^{S}$ as our pseudo task transformation module, each of which estimates the task-specific moment statistics $\{\boldsymbol{\mu}^{(s)}, \boldsymbol{\sigma}^{(s)}\}$ and maps the output of the shared structure to task-specific feature space.

Given a set of features $\{\boldsymbol{f}_i^{(s)}\}_{i=1}^{N^{(s)}}$ from task $\mathbf{D}_{train}^{(s)}$, The mean and variance of $\mathrm{TSBN}^{(s)}$ are estimated by

$$\boldsymbol{\mu}^{(s)} = \frac{1}{N^{(s)}} \sum_{i=1}^{N^{(s)}} \boldsymbol{f}_i^{(s)}, \quad {\boldsymbol{\sigma}^{(s)}}^2 = \frac{1}{N^{(s)}} \sum_{i=1}^{N^{(s)}} (\boldsymbol{f}_i^{(s)} - \boldsymbol{\mu}^{(s)})^2, \tag{1}$$

where $\boldsymbol{\mu}^{(s)}$ and ${\boldsymbol{\sigma}^{(s)}}^2$ are task-specific mean and variance.

At the s-th training step, given a sample $\boldsymbol{x}_i^{(s)}$, we encode it into a feature vector through feature extractor $H(\cdot, \boldsymbol{\Theta})$ denoted by $\boldsymbol{f}_i^{(s)} = H(\boldsymbol{x}_i^{(s)}, \boldsymbol{\Theta})$. Then we input the feature $\boldsymbol{f}_i^{(s)}$ into $\{\mathrm{TSBN}^{(j)}\}_{j=1}^{s}$ and project the feature into task-specific feature space formulated as:

$$\hat{\boldsymbol{f}}_{i,j}^{(s)} = \mathrm{Transform}(\boldsymbol{f}_i^{(s)}, j) = \boldsymbol{\gamma} \frac{\boldsymbol{f}_i^{(s)} - \boldsymbol{\mu}^{(j)}}{\sqrt{{\boldsymbol{\sigma}^{(j)}}^2 + \epsilon}} - \boldsymbol{\beta}, \tag{2}$$

where j is the subscript which indicates j-th $\mathrm{TSBN}^{(j)}$, ϵ is a small value to keep stability. Note that the statistics of old TSBN keep unchanged for avoiding bias, only the $\mathrm{TSBN}^{(s)}$ with respect to the new task is dynamic changed during training. After the transformation, the feature $\boldsymbol{f}_i^{(s)}$ is mapped to $\hat{\boldsymbol{f}}_{i,j}^{(s)}$ under the feature distribution of task j to supplement the missing samples of j-th task, which contains cross-task consistent information to be mined.

We assume that the transformed features of the pseudo tasks follow the same relationship of similarity and identity with that of the original features.

Domain Consistency Learning on Pseudo Tasks. In this section, we explore a task-specific **D**omain **C**onsistency **L**earning objective (DCL), which focuses on mining consistency information from pseudo tasks and mitigating task-wise domain shift. We assume that the features of all tasks follow a global Gaussian distribution. To learn task-shared knowledge, we apply a BN layer called shared BN on $\boldsymbol{f}_i^{(s)}$ to obtain a normalized feature $\boldsymbol{f}_{i,\mathrm{shared}}^{(s)}$, which is also the feature for inference. We regard the pseudo task features $\left\{\hat{\boldsymbol{f}}_{i,j}^{(s)}\right\}_{j=1}^{s}$ as different views of $\boldsymbol{f}_{i,\mathrm{shared}}^{(s)}$ and pull normalized feature $\hat{\boldsymbol{f}}_{i,\mathrm{shared}}^{(s)}$ by shared BN close to corresponding pseudo task features $\left\{\hat{\boldsymbol{f}}_{i,j}^{(s)}\right\}_{j=1}^{s}$ to implicitly mitigate the task-wise domain gap. Specifically, to pull these features together, we minimize the cosine distance between $\hat{\boldsymbol{f}}_{i,\mathrm{shared}}^{(s)}$ and pseudo task features $\left\{\hat{\boldsymbol{f}}_{i,j}^{(s)}\right\}_{j=1}^{s}$ formulated by

$$\mathcal{L} = \frac{1}{N} \frac{1}{s} \sum_{i=1}^{N} \sum_{j=1}^{s} \mathcal{D}\left(\hat{\boldsymbol{f}}_{i,\mathrm{shared}}^{(s)}, \mathrm{SG}(\hat{\boldsymbol{f}}_{i,j}^{(s)})\right), \tag{3}$$

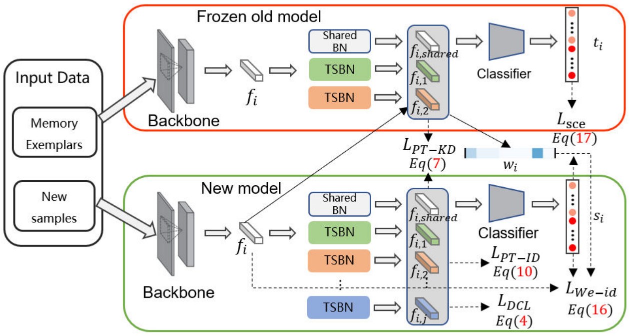

Figure 2: An overall of our framework.

where N is the batch-size and SG indicates stop gradient operation. However, identity supervision is ignored in Eq. (3), to further leverage the supervision, we slightly modify our DCL loss as follow:

$$\mathcal{L}_{\text{DCL}} = \frac{1}{N}\frac{1}{s}\frac{1}{|\mathcal{P}_i|}\sum_{i=1}^{N}\sum_{j=1}^{s}\sum_{k\in\mathcal{P}_i}\mathcal{D}\left(\hat{\boldsymbol{f}}^{(s)}_{i,\text{shared}}, \text{SG}(\hat{\boldsymbol{f}}^{(s)}_{k,j})\right), \tag{4}$$

where $\mathcal{P}_i$ denotes the set of indices of samples that share the same identity with $\mathbf{x}_i$ and $|\mathcal{P}|$ is the cardinality of $\mathcal{P}$.

Discussion. When minimizing the $\mathcal{L}_{\text{DCL}}$, the discrepancy between global statistics and task-specific statistics is mitigated, so the task-wise domain gap is alleviated. Commonly used distribution alignment losses (e.g., MMD (Long et al. 2015) and CORAL (Sun and Saenko 2016)) model domain-specific distribution and then minimize distribution distance. However, with few samples from the old tasks, it is hard to estimate unbiased distribution measurement which is harmful for domain adaptation. While we keep the task-specific statistics in the TSBN without access to source samples, which is a *lifelong source-free domain adaptation* method.

Knowledge Preservation on Pseudo Tasks

Task-specific domain consistency loss only considers the intra-class relationship in the new model while ignores the inter-class relationship and old model. Under the guidance of the feature distributions of the old tasks, we propose to preserve knowledge on pseudo tasks with a more comprehensive learning objective including a pseudo task knowledge distillation loss and a pseudo task identity discrimination loss which further considers the intra-inter-class relationship of new task under the constraint of keeping the discriminative capability on old tasks, which achieves a better trade-off with imposed stability and plasticity dilemma in lifelong person Re-ID.

Pseudo Task Knowledge Distillation. In LReID, catastrophic forgetting is inevitable. To mitigate the issue, we should also distill old knowledge to new model when learning new knowledge. To this end, knowledge distillation is commonly employed for preserving old knowledge.

Since Re-ID task is essentially an open-set problem, we aim at learning good representation instead of classification scores, we propose that the pair-wise relationship between new model and old model for saved exemplars from the old tasks should be consistency to preserve already learned knowledge, which formulates a representation-level relation based knowledge distillation loss which is ignored in previous methods for LReID. Specifically, given a mini-batch samples $\{\boldsymbol{x}_i\}_{i=1}^{N}$, we extract N d-dimensional features $\mathcal{X}^{(s-1)} = [\boldsymbol{f}_1^{(s-1)}, \boldsymbol{f}_2^{(s-1)}, \ldots, \boldsymbol{f}_N^{(s-1)}] \in \mathcal{R}^{d\times N}$ and $\mathcal{X}^{(s)} = [\boldsymbol{f}_1^{(s)}, \boldsymbol{f}_2^{(s)}, \ldots, \boldsymbol{f}_N^{(s)}] \in \mathcal{R}^{d\times N}$ via the old feature extractor $H(\cdot;\boldsymbol{\Theta}^{(s-1)})$ and the new feature extractor $H(\cdot;\boldsymbol{\Theta}^{(s)})$, respectively. Then we compute the pairwise similarity matrix with cosine similarity as follows:

$$\mathcal{M}^{(s-1)} = {\mathcal{X}^{(s-1)}}^{\mathrm{T}}\mathcal{X}^{(s-1)}, \quad \mathcal{M}^{(s)} = {\mathcal{X}^{(s)}}^{\mathrm{T}}\mathcal{X}^{(s)}. \tag{5}$$

We aim to preserve the old knowledge by minimizing the distance between the two similarity matrices as follow:

$$\mathcal{L}_{\text{KD}} = \left\|\mathcal{M}^{(s-1)} - \mathcal{M}^{(s)}\right\|_F, \tag{6}$$

where F is set to 1 as default and $\|\cdot\|_F$ is the L1 norm.

Instead of using the representation output by the feature extractor which ignores the guidance of distributions of the old tasks, we utilize the pseudo task representation for distillation. In particular, we propose a pseudo tasks knowledge distillation loss (PT-KD) formulated by

$$\mathcal{L}_{\text{PT-KD}} = \frac{1}{s^2}\sum_{j=1}^{s}\sum_{k=1}^{s}\left\|\mathcal{M}_j^{(s-1)} - \mathcal{M}_k^{(s)}\right\|_F, \tag{7}$$

where $\mathcal{M}_j$ indicates the similarity matrix computed by $\{\hat{\boldsymbol{f}}^{(s)}_{i,j}\}_{i=1}^{N}$ with omitting the subscript. Note that Eq. (7) not only minimizes intra-pseudo task similarity matrix, but also minimizes inter-pseudo task similarity matrix.

Pseudo Task Identity Discrimination. The pseudo task knowledge distillation takes the pair-wise relationship between new and old model into consideration while ignores the intra-inter-class relationship of the new task to keep the discriminative capability. We assume that transformed features in the feature space of the old tasks follow the same intra-inter-class relationship and guaranteeing the discriminative capability on them encourages both the new task and the old tasks to learn better. Thus, we propose to pull the features close to their hard positive pair and push away from their hard negative pair within each pseudo task $\left\{\hat{\boldsymbol{f}}^{(s)}_{i,j}\right\}_{i=1}^{N}$, which is called intra-pseudo-task identity discrimination loss (Intra-PT-ID) formulated by

$$\begin{aligned}\mathcal{L}_{\text{Intra-PT-ID}} = &\frac{1}{N}\frac{1}{s}\sum_{i=1}^{N}\sum_{j=1}^{s}\max(d(\hat{a}_{i,j},\hat{p}_{i,j}) \\ &- d(\hat{a}_{i,j},\hat{n}_{i,j}) + \text{margin}, 0),\end{aligned} \tag{8}$$

where $d(\hat{a}_{i,j},\hat{p}_{i,j})$, $d(\hat{a}_{i,j},\hat{n}_{i,j})$ is the distance between anchor feature $\hat{\boldsymbol{f}}_{i,j}$ and corresponding positive pair $\hat{\boldsymbol{f}}_{p,j}$ and negative pair $\hat{\boldsymbol{f}}_{n,j}$, respectively. Besides, we also propose an inter-pseudo-task identity discrimination loss (Inter-PT-ID) to mine cross-pseudo-task information. Specifically, given an anchor features $\hat{\boldsymbol{f}}_{i,j}$, we mine its hard positive and hard negative counterpart in another pseudo task features $\{\hat{\boldsymbol{f}}_{i,k}\}_{i=1}^{N}, (k \neq j)$ formulated by

$$\begin{aligned}\mathcal{L}_{\text{Inter-PT-ID}} =& \frac{1}{N}\frac{1}{s}\frac{1}{s-1}\sum_{i=1}^{N}\sum_{j=1}^{s}\sum_{k\neq j}^{s}\max(d(\hat{a}_{i,j},\hat{p}_{i,k}) \\ & - d(\hat{a}_{i,j},\hat{n}_{i,k}) + \text{margin}, 0),\end{aligned} \tag{9}$$

where $d(\hat{a}_{i,j},\hat{p}_{i,k})$, $d(\hat{a}_{i,j},\hat{n}_{i,k})$ is the distance between anchor feature $\hat{\boldsymbol{f}}_{i,j}$ and inter-pseudo-task positive pair $\hat{\boldsymbol{f}}_{p,k}$ and negative pair $\hat{\boldsymbol{f}}_{n,k}$, respectively.

Finally, pseudo task identity discrimination loss (PT-ID) is formulated by

$$\mathcal{L}_{\text{PT-ID}} = \mathcal{L}_{\text{Intra-PT-ID}} + \mathcal{L}_{\text{Inter-PT-ID}}. \tag{10}$$

Pseudo Task Knowledge Preservation Framework

Besides preserving knowledge for the old tasks, for learning new knowledge, we adopt the commonly used cross-entropy loss and triplet loss on new samples. For saved exemplars, we only employ triplet loss. To select samples with more task-shared information for new knowledge learning, we introduce an ambiguity-aware instance weighting strategy. Finally, We aggregate task-specific domain consistency loss, knowledge preservation on pseudo tasks and new knowledge learning to form an unified framework.

Ambiguity-aware Instance Weighting Strategy. To learn new knowledge, cross-entropy loss $\mathcal{L}_{\text{ce}}$ is usually used for classification formulated by

$$\mathcal{L}_{\text{ce}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C_n+C_o} p_{i,c}\log(q_{i,c}), \tag{11}$$

where $q_{i,c}$ is the softmax output score belonging to class c for sample $\boldsymbol{x}_i$, $p_{i,c}$ is the ground truth, N is the batch-size, C_n and C_o is the class number of new task and all old tasks.

For a more compact representation learning, triplet loss $\mathcal{L}_{\text{tr}}$ is also employed for metric learning formulated by

$$\mathcal{L}_{\text{tr}} = \frac{1}{N}\sum_{i=1}^{N}\max(d(a_i,p_i) - d(a_i,n_i) + \text{margin}, 0), \tag{12}$$

where $d(a_i,p_i)$ is the distance between anchor $\boldsymbol{x}_i$ and its positive pair, $d(a_i,n_i)$ is the distance between anchor $\boldsymbol{x}_i$ and its negative pair, margin is set to 0.0 as default. Then, basic identity loss for learning new knowledge is formulated by

$$\mathcal{L}_{\text{ID}} = \mathcal{L}_{\text{ce}} + \mathcal{L}_{\text{tr}}. \tag{13}$$

Since the difference of unseen task data distribution leads to low matching confidence, intuitively, if a feature $\boldsymbol{f}_i^{(s)}$ of the new task has higher similarity with the task-specific distribution of the old tasks, this indicates the high matching confidence and less ambiguous. Based on the above discussion, we propose an ambiguity-aware instance weighting strategy to reweight samples in a mini-batch. Given a mini-batch features $\{\boldsymbol{f}_i^{(s)}\}_{i=1}^{N}$ in the s-th training step, we compute the mean similarity vector between $\{\boldsymbol{f}_i^{(s)}\}_{i=1}^{N}$ and frozen task-specific mean $\{\boldsymbol{\mu}^{(k)}\}, k \in \{1, 2, \ldots, (s-1)\}$ estimated by $\{\text{TSBN}^{(k)}\}_{k=1}^{(s-1)}$ as follow:

$$\hat{w}_i = \frac{1}{s-1}\sum_{k=1}^{s-1}{\boldsymbol{f}_i^{(s)}}^T\boldsymbol{\mu}^{(k)} \tag{14}$$

where $\hat{w}_i$ is the mean cosine similarity between sample $\boldsymbol{f}_i^{(s)}$ and task-specific mean. To obtain the weight for $\boldsymbol{f}_i^{(s)}$, we employ a softmax function to normalize the similarity as

$$w_i = \frac{\exp(\hat{w}_i/\tau)}{\sum_{j=1}^{N}\exp(\hat{w}_j/\tau)}, \tag{15}$$

where τ is the temperature factor to control the distribution. Finally, the weighted identity loss (We-ID) is formulated as

$$\mathcal{L}_{\text{We-ID}} = \sum_{i=1}^{N} w_i\mathcal{L}_{\text{ID}}. \tag{16}$$

Logits-level Knowledge Distillation. Since pseudo task knowledge distillation only considers pair-wise relationship among saved exemplars while ignores the relationship between new classes and old classes, we further adopt a logits-level knowledge distillation to mitigate forgetting on old tasks as a complement for pseudo task knowledge distillation. In particular, we use soft cross-entropy loss to minimize the distribution discrepancy of logits between new and old model for saved exemplars as follow:

$$\begin{aligned}&\boldsymbol{t}_i = \boldsymbol{\sigma}(g(H(\boldsymbol{x}_i^{(j)};\boldsymbol{\Theta}^{(s-1)});\boldsymbol{\phi}^{(s-1)})), \\ &\boldsymbol{s}_i = \boldsymbol{\sigma}(g(H(\boldsymbol{x}_i^{(j)};\boldsymbol{\Theta}^{(s)});\boldsymbol{\phi}^{(s)})), \\ &\mathcal{L}_{\text{sce}} = -\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{t}_i\log\boldsymbol{s}_i,\end{aligned} \tag{17}$$

where $\boldsymbol{\sigma}(\cdot)$ is softmax operation, $\boldsymbol{t}_i$ works as supervision signal from the old model and $\boldsymbol{s}_i$ is classification scores output by the next-step new model.

Finally, the total loss function can be formulated as

$$\mathcal{L} = \mathcal{L}_{\text{We-ID}} + \mathcal{L}_{\text{sce}} + \mathcal{L}_{\text{DCL}} + \mathcal{L}_{\text{PT-KD}} + \mathcal{L}_{\text{PT-ID}}. \tag{18}$$

Experiments

Experiments Setting

Datasets. To evaluate the effectiveness of our method, we conducted experiments in the LReID setting as GwFReID (Wu and Gong 2021) on benchmark person Re-ID datasets, of which four were used as sequential input datasets (i.e., Market-1501 (Zheng et al. 2015), DukeMTMC[2] (Zheng, Zheng, and Yang 2017), CUHK-SYSU (Xiao et al. 2017) and MSMT17 (Wei et al. 2018)) to imitate the task-incremental lifelong learning. CUHK-SYSU is modified from original dataset prepared for person search, we followed the process method of GwFReID (Wu and Gong 2021) to obtain training and test data. Besides, to validate the generalization of the model, we also tested on four unseen Re-ID benchmarks (i.e., CUHK01 (Li, Zhao, and Wang 2012), CUHK03 (Li et al. 2014), SenseReID (Zhao et al. 2017) and GRID (Loy, Xiang, and Gong 2010)).

[2]We used DukeMTMC for fair comparison and only for academic use without identifying or showing the person images.

Methods	Reference	Train: Market1501 → DukeMTMC → CUHK-SYSU → MSMT17									
		Market-1501		DukeMTMC		CUHK-SYSU		MSMT17		Average	
		R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	$\bar{s}_{\text{R-1}}$	$\bar{s}_{\text{mAP}}$
Joint-Train	-	94.6	86.3	89.1	76.3	94.1	92.9	78.5	53.1	89.1	77.2
FineTune	-	76.8	50.2	66.4	46.9	84.5	81.6	71.1	43.3	74.7	55.5
MMD†	JMLR15	88.0	69.6	73.4	55.7	87.9	85.9	65.1	37.2	78.6	62.1
CORAL†	ECCV16	87.4	70.7	75.3	58.2	88.4	86.5	64.9	37.7	79.0	63.3
LwF†	TPAMI17	88.3	72.4	74.6	59.3	87.2	86.1	63.0	36.1	78.3	63.5
SPD†	ICCV19	87.8	71.4	75.1	58.5	88.0	86.5	64.3	37.7	78.9	63.4
CRL†	WACV21	83.7	60.5	71.8	51.9	86.4	83.7	67.1	41.2	77.3	59.3
WA*	CVPR20	73.2	52.1	54.6	36.0	83.7	81.1	39.6	17.2	62.8	46.6
BiC*	CVPR19	75.8	53.4	55.4	37.7	84.2	81.4	33.0	13.1	62.1	46.4
GwFReID*	AAAI21	81.6	60.9	66.5	46.7	83.9	81.4	52.4	25.9	71.1	53.7
Ours	Ours	**90.1**	**77.0**	**78.0**	**63.8**	**90.1**	**88.4**	**67.5**	**41.9**	**81.4**	**67.6**

Table 1: Comparison with state-of-the-art methods in lifelong person Re-ID setting for seen tasks. The notation "*" means we implement the reported results. "†" means we implement the released code on our baseline.

Method	CUHK01	CUHK03	GRID	SenseReID
Joint-Train	78.2	38.8	29.6	42.7
FineTune	69.3	37.8	14.4	38.6
MMD	75.6	45.0	23.8	40.8
CORAL	74.3	44.0	24.8	42.0
LwF	73.0	42.8	20.1	39.1
SPD	73.9	43.4	20.8	42.1
CRL	71.4	41.6	18.7	40.1
WA	-	30.5	-	-
BiC	-	31.2	-	-
GwFReID	-	40.2	-	-
Ours	**79.1**	**54.1**	**31.5**	**44.8**

Table 2: Comparison with state-of-the-art methods in lifelong person Re-ID setting for unseen tasks. The results (Rank-1) are reported at the end of the last training step.

Evaluation Protocol. We used mean average precision (mAP) and Rank-1 accuracy (R-1) (Zheng et al. 2015) on each task. We also computed average incremental accuracy (Rebuffi et al. 2017) $\bar{s}_{\text{R-1}}$ and $\bar{s}_{\text{mAP}}$ for seen tasks, where $\bar{s}_{\text{R-1}} = \frac{1}{s}\sum_{i=1}^{s} \text{mAP}_i$. Specifically, after each training phase, we tested on all seen tasks observed so far and computed the average mAP ($\bar{s}_{\text{mAP}}$), average Rank-1 ($\bar{s}_{\text{R-1}}$) and computed Rank-1 on all unseen tasks.

Implementation Details

ResNet-50 (He et al. 2016) pre-trained on ImageNet (Deng et al. 2009) was adopted as our backbone. The details of its structure followed the design in BoT (Luo et al. 2019). We replaced the global average pooling (GAP) with generalized mean pooling. The mini-batch size for both new task and replay task sampled from the exemplar memory bank was 128 per task. For constructing exemplar memory bank with less memory cost, we randomly chose 250 classes. For each class, we saved furthest two samples away from the class center. For sampling strategy, we randomly selected 64 identities for each task and 2 images for each identity. We used Adam (Kingma and Ba 2014) as our optimization algorithm. The learning rate was set to 0.00035 initially and decayed by 0.1 after 40^{th}, 70^{th} epochs in 80 epochs totally for training the first dataset. For subsequent tasks, the learning rate was set to 0.000035 initially and decayed by 0.1 in 30^{th} epochs in 60 epochs totally. The weights for all loss functions were set to 1 for simplicity. τ was set to 0.5.

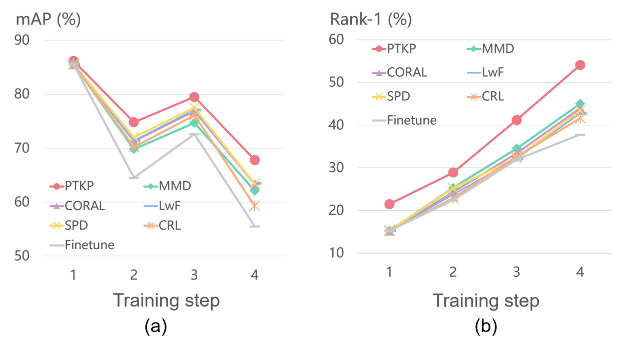

Figure 3: (a) Incremental learning performance ($\bar{s}_{\text{mAP}}$) evaluation on seen task. (b) Incremental learning performance (Rank-1) evaluation on CUHK03.

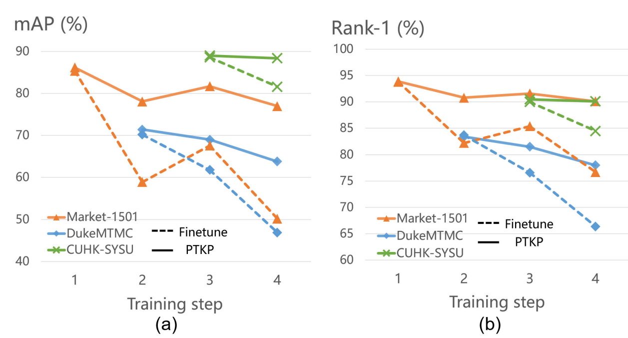

Figure 4: (a) Results (mAP) on each seen task. (b) Results (Rank-1) on each seen task. Finetune method is compared.

Comparison with State-of-the-Art Methods

We evaluated several methods including the state-of-the-art lifelong learning method (i.e., GwFReID (Wu and Gong 2021)) for LReID, lifelong learning methods for recognition issue, including SPD (Tung and Mori 2019), CRL (Zhao et al. 2021), WA (Zhao et al. 2020a), BiC (Wu et al. 2019) and LwF (Li and Hoiem 2017). Besides, since we cast LReID as a domain adaptation problem, we also compared distribution alignment methods including Maximum Mean Discrepancy (MMD) loss (Long et al. 2015) and CORAL loss (Sun and Saenko 2016). In addition, we also compared with Finetune method which incrementally finetunes new datasets with identity loss $\mathcal{L}_{\text{ID}}$ and Joint-Train method which

Exp	Loss						Market-1501		DukeMTMC		CUHK-SYSU		MSMT17	
	$\mathcal{L}_{\text{base}}$	$\mathcal{L}_{\text{DCL}}$	$\mathcal{L}_{\text{PT-KD}}$	$\mathcal{L}_{\text{PT-ID}}$	$\mathcal{L}_{\text{KD}}$	$\mathcal{L}_{\text{We-ID}}$	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
1	✓						88.3	72.4	74.6	59.3	87.2	86.1	63.0	36.1
2	✓	✓					89.4	74.2	77.1	62.5	87.7	86.2	63.2	37.1
3	✓	✓	✓				89.7	76.6	78.5	64.2	88.2	86.6	64.3	37.9
4	✓	✓	✓	✓			89.3	76.5	77.5	63.5	89.4	88.2	67.7	42.1
5	✓	✓		✓	✓		87.6	73.7	75.6	62.1	88.1	86.4	66.0	41.2
6	✓		✓	✓		✓	88.4	74.6	74.8	60.8	89.8	88.3	65.8	40.1
7	✓	✓	✓	✓		✓	90.1	77.0	78.0	63.8	90.1	88.4	67.5	41.9

Table 3: Ablation study on individual components of PTKP. $\mathcal{L}_{\text{base}}$ is composed of $\mathcal{L}_{\text{ID}}$ and $\mathcal{L}_{\text{sce}}$.

combines all datasets for jointly training. For WA and BiC, we implemented the reported results from GwFReID (Wu and Gong 2021). For others, we implemented the released code on our baseline.

Comparison on seen tasks for validating the anti-forgetting capability. We report the results in Table 1 and Figure 3 (a). Our method surpasses lifelong learning method for Re-ID, for recognition issue and distribution alignment methods in each training step. Besides, we depict the trend for each seen task so far after each training step in Figure 4 where the Finetune method was compared.

Comparison on unseen tasks for validating the generalization. The results are shown in Table 2 and Figure 3 (b). The generalization of our method outperforms that of all other methods including Joint-Train on all datasets, which indicates that alleviating the task-wise domain gap can effectively learn task-shared knowledge which is helpful for improving the generalization of the model.

Ablation Study

We carried out ablation study to evaluate the effectiveness of each component in our framework. The experimental results are reported in Table 3.

The effectiveness of $\mathcal{L}_{\text{DCL}}$ in Eq (4). Compared with "Exp 7", the performance of "Exp 6" without $\mathcal{L}_{\text{DCL}}$ clearly dropped on four datasets and compared with "Exp 1", the performance of "Exp 2" with $\mathcal{L}_{\text{DCL}}$ clearly improved. This indicates that $\mathcal{L}_{\text{DCL}}$ can alleviate the task-wise domain gap and force $\mathcal{L}_{\text{PT-KD}}$ and $\mathcal{L}_{\text{PT-ID}}$ to learn task-shared knowledge.

The effectiveness of $\mathcal{L}_{\text{PT-KD}}$ in Eq (7). Compared with "Exp 2", the performance of "Exp 3" with $\mathcal{L}_{\text{PT-KD}}$ obviously improved on four datasets. Besides, compared with "Exp 4", the performance of "Exp 5" which used feature output by the feature extractor for distillation was obviously dropped, which indicates that the pseudo task knowledge distillation loss $\mathcal{L}_{\text{PT-KD}}$ can effectively preserve task-shared old knowledge without losing discriminative capability on new tasks.

The effectiveness of $\mathcal{L}_{\text{PT-ID}}$ in Eq (10). Compared with "Exp 3", the performance of 'Exp 4" with $\mathcal{L}_{\text{PT-ID}}$ obviously improved on new tasks without losing discriminative capability on old tasks. This indicates that $\mathcal{L}_{\text{PT-ID}}$ which further considers the intra-inter-class relationship of new task to learn task-shared identity discriminative information under the guidance of feature distributions of the old tasks can keep discriminative capability on both new and old tasks.

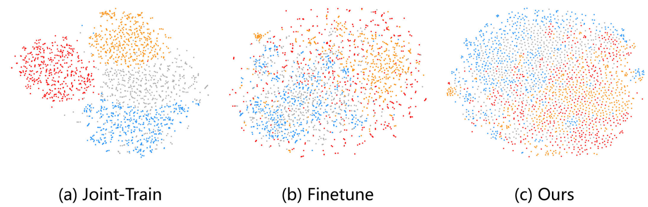

Figure 5: t-SNE visualization (Van der Maaten and Hinton 2008) of feature distribution from four seen datasets. Different color indicates samples of different tasks.

The effectiveness of $\mathcal{L}_{\text{We-ID}}$ in Eq (16). Compared with "Exp 4", the performance of 'Exp 7" with $\mathcal{L}_{\text{We-ID}}$ slightly improved on the old tasks without losing performance on the new task, which indicates that $\mathcal{L}_{\text{We-ID}}$ helps select less ambiguous samples and learn task-shared knowledge.

Visualization

To better understand the effect of our method as for domain adaptation and task-shared knowledge learning, we visualized the feature distribution learned by Joint-Train method, Finetune method and our method when the sequential training ended as shown in Figure 5. We can observe that the task-specific distribution of our method is more identical, which indicates that the task-wise domain gap is mitigated and the model learns task-shared knowledge.

Conclusion

In this paper, we study lifelong person re-identification problem, which is significant for real-world application but remains under-explored. Catastrophic forgetting leads to a great challenge which is further enhanced by the task-wise domain gap. We cast LReID as a source-free domain adaptation problem and aim to alleviate the task-wise domain shift for learning task-shared knowledge, which is ignored by existing LReID methods. To this end, we propose a pseudo task knowledge preservation framework. The framework consists of a task-wise domain consistency loss which implicitly mitigates the task-wise domain gap via intra-class relationship of transformed pseudo tasks to learn task-shared knowledge instead of task-specific one, a pseudo task knowledge distillation and identity discrimination loss which preserve knowledge under the guidance of the feature distributions of the old tasks to keep the discriminative capability on both new and old tasks. Finally, an ambiguity-aware weighting strategy is introduced for new knowledge learning. Extensive experiments validate the superiority of our method.

Acknowledgements

This work was partially done while interning at Tencent Youtu Lab. This work was supported partially by the National Science Foundation for Young Scientists of China (62106288), the NSFC (U1911401, U1811461, U21A20471), China National Postdoctoral Program for Innovative Talents (BX20200395), China Postdoctoral Science Foundation (2021M693616), Guangdong NSF Project (2020B1515120085, 2018B030312002), the Key-Area Research and Development Program of Guangzhou (202007030004).

References

Aljundi, R.; Chakravarty, P.; and Tuytelaars, T. 2017. Expert Gate: Lifelong Learning With a Network of Experts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bai, Z.; Wang, Z.; Wang, J.; Hu, D.; and Ding, E. 2021. Unsupervised Multi-Source Domain Adaptation for Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Chen, P.; Dai, P.; Liu, J.; Zheng, F.; Xu, M.; Tian, Q.; and Ji, R. 2021. Dual Distribution Alignment Network for Generalizable Person Re-Identification. *Proceedings of the AAAI Conference on Artificial (AAAI)*.

Chen, T.; Ding, S.; Xie, J.; Yuan, Y.; Chen, W.; Yang, Y.; Ren, Z.; and Wang, Z. 2019. Abd-net: Attentive but diverse person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (CVPR)*.

Chen, Y.-C.; Zhu, X.; Zheng, W.-S.; and Lai, J.-H. 2017. Person re-identification by camera correlation aware feature augmentation. *IEEE transactions on pattern analysis and machine intelligence (TMAPI)*.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *IEEE conference on computer vision and pattern recognition (CVPR)*.

Fernando, C.; Banarse, D.; Blundell, C.; Zwols, Y.; Ha, D.; Rusu, A. A.; Pritzel, A.; and Wierstra, D. 2017. Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*.

Ge, W.; Pan, C.; Wu, A.; Zheng, H.; and Zheng, W.-S. 2021. Cross-Camera Feature Prediction for Intra-Camera Supervised Person Re-identification across Distant Scenes. In *Proceedings of the 29th ACM International Conference on Multimedia (ACMMM)*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; and Zhang, L. 2020. Style normalization and restitution for generalizable person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, H.; Wu, G.; and Zheng, W.-S. 2021. Combined Depth Space based Architecture Search For Person Re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, M.; Zhu, X.; and Gong, S. 2018a. Unsupervised person re-identification by deep learning tracklet association. In *Proceedings of the European conference on computer vision (ECCV)*.

Li, W.; Zhao, R.; and Wang, X. 2012. Human reidentification with transferred metric learning. In *Asian conference on computer vision (ACCV)*.

Li, W.; Zhao, R.; Xiao, T.; and Wang, X. 2014. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, W.; Zhu, X.; and Gong, S. 2018b. Harmonious attention network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, Z.; and Hoiem, D. 2017. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*.

Lin, Y.; Dong, X.; Zheng, L.; Yan, Y.; and Yang, Y. 2019. A bottom-up clustering approach to unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

Liu, Z.; Wang, J.; Gong, S.; Lu, H.; and Tao, D. 2019. Deep reinforcement active learning for human-in-the-loop person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Long, M.; Cao, Y.; Wang, J.; and Jordan, M. 2015. Learning transferable features with deep adaptation networks. In *International conference on machine learning (ICML)*.

Loy, C. C.; Xiang, T.; and Gong, S. 2010. Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision (IJCV)*.

Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; and Gu, J. 2019. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Transactions on Multimedia (TMM)*.

Pu, N.; Chen, W.; Liu, Y.; Bakker, E. M.; and Lew, M. S. 2021. Lifelong Person Re-Identification via Adaptive Knowledge Accumulation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qi, L.; Wang, L.; Huo, J.; Shi, Y.; and Gao, Y. 2019. Adversarial camera alignment network for unsupervised cross-camera person re-identification. *arXiv preprint arXiv:1908.00862*.

Qi, L.; Wang, L.; Huo, J.; Shi, Y.; and Gao, Y. 2020. Progressive cross-camera soft-label learning for semi-supervised person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*.

Serra, J.; Suris, D.; Miron, M.; and Karatzoglou, A. 2018. Overcoming catastrophic forgetting with hard attention to the task. In *International Conference on Machine Learning (ICML)*.

Song, J.; Yang, Y.; Song, Y.-Z.; Xiang, T.; and Hospedales, T. M. 2019. Generalizable person re-identification by domain-invariant mapping network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sun, B.; and Saenko, K. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision (ECCV)*.

Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European conference on computer vision (ECCV)*.

Tung, F.; and Mori, G. 2019. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research (JMLR)*.

Wang, M.; Lai, B.; Huang, J.; Gong, X.; and Hua, X.-S. 2020. Camera-aware Proxies for Unsupervised Person Re-Identification. *arXiv preprint arXiv:2012.10674*.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Wu, A.; Zheng, W.-S.; and Lai, J.-H. 2019. Unsupervised person re-identification by camera-aware similarity consistency learning. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Wu, C.; Herranz, L.; Liu, X.; van de Weijer, J.; Raducanu, B.; et al. 2018. Memory replay gans: Learning to generate new categories without forgetting. *Advances in Neural Information Processing Systems (NeurIPS)*.

Wu, G.; and Gong, S. 2021. Generalising without Forgetting for Lifelong Person Re-Identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

Wu, G.; Zhu, X.; and Gong, S. 2020. Tracklet self-supervised learning for unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

Wu, Y.; Chen, Y.; Wang, L.; Ye, Y.; Liu, Z.; Guo, Y.; and Fu, Y. 2019. Large scale incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xiao, T.; Li, S.; Wang, B.; Lin, L.; and Wang, X. 2017. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xiong, F.; Gou, M.; Camps, O.; and Sznaier, M. 2014. Person re-identification using kernel-based metric learning methods. In *European conference on computer vision (ECCV)*.

Zeng, K.; Ning, M.; Wang, Y.; and Guo, Y. 2020. Hierarchical clustering with hard-batch triplet loss for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, M.; Liu, K.; Li, Y.; Guo, S.; Duan, H.; Long, Y.; and Jin, Y. 2021. Unsupervised Domain Adaptation for Person Re-Identification via Heterogeneous Graph Alignment. *Proceedings of the AAAI Conference on Artificial (AAAI)*.

Zhang, T.; Xie, L.; Wei, L.; Zhang, Y.; Li, B.; and Tian, Q. 2020. Single camera training for person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

Zhao, B.; Tang, S.; Chen, D.; Bilen, H.; and Zhao, R. 2021. Continual representation learning for biometric identification. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*.

Zhao, B.; Xiao, X.; Gan, G.; Zhang, B.; and Xia, S.-T. 2020a. Maintaining discrimination and fairness in class incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhao, F.; Liao, S.; Xie, G.-S.; Zhao, J.; Zhang, K.; and Shao, L. 2020b. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *European Conference on Computer Vision (ECCV)*.

Zhao, H.; Tian, M.; Sun, S.; Shao, J.; Yan, J.; Yi, S.; Wang, X.; and Tang, X. 2017. 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 (CVPR)*.

Zheng, K.; Lan, C.; Zeng, W.; Zhan, Z.; and Zha, Z.-J. 2021. Exploiting Sample Uncertainty for Domain Adaptive Person Re-Identification. *Proceedings of the AAAI Conference on Artificial (AAAI)*.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Zheng, W.-S.; Gong, S.; and Xiang, T. 2012. Reidentification by relative distance comparison. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*.

Zheng, Z.; Zheng, L.; and Yang, Y. 2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.