Lifelong Person Re-Identification via Adaptive Knowledge Accumulation

Nan Pu[1] Wei Chen[1] Yu Liu[2]* Erwin M. Bakker[1] Michael S. Lew[1]
[1]LIACS Media Lab, Leiden University, The Netherlands
[2]International School of Information Science & Engineering, Dalian University of Technology, China
{n.pu, w.chen, erwin, m.s.k.lew}@liacs.leidenuniv.nl, liuyu8824@dlut.edu.cn

Abstract

Person re-identification (ReID) methods always learn through a stationary domain that is fixed by the choice of a given dataset. In many contexts (e.g., lifelong learning), those methods are ineffective because the domain is continually changing in which case incremental learning over multiple domains is required potentially. In this work we explore a new and challenging ReID task, namely lifelong person re-identification (LReID), which enables to learn continuously across multiple domains and even generalise on new and unseen domains. Following the cognitive processes in the human brain, we design an Adaptive Knowledge Accumulation (AKA) framework that is endowed with two crucial abilities: knowledge representation and knowledge operation. Our method alleviates catastrophic forgetting on seen domains and demonstrates the ability to generalize to unseen domains. Correspondingly, we also provide a new and large-scale benchmark for LReID. Extensive experiments demonstrate our method outperforms other competitors by a margin of 5.8% mAP in generalising evaluation. The codes will be available at https://github.com/TPCD/LifelongReID.

1. Introduction

Person re-identification (ReID) seeks to linking the same pedestrian across disjoint camera views. While advanced deep learning methods [55, 49, 30, 46, 38, 31, 47] have shown powerful abilities for ReID [35, 10], their training process is limited heavily by a fixed and stationary dataset [52, 54, 40]. However, this limitation violates many practical scenarios where the data is continuously increasing from different domains. For instance, smart surveillance systems [53, 15] over multiple crossroads capture millions of new images every day, and they are required to have the ability of incremental or lifelong learning.

To overcome the above limitation, we propose a new yet practical ReID task, namely *lifelong person re-identification* (LReID), which requires the model to accumulate informative knowledge incrementally from several seen domains and then adapt the knowledge to the test sets of both seen and unseen domains (Fig. 1). Our LReID task has two challenging problems, compared to previous tasks. First, unlike conventional lifelong learning [27, 32], LReID further considers improving the generalization ability on unseen classes that never appear in the lifelong training stage. Second, LReID is a fine-grained lifelong learning task, in which inter-class appearance variations are significantly subtler than standard lifelong learning benchmarks like CIFAR-100 [13] and ImageNet [33].

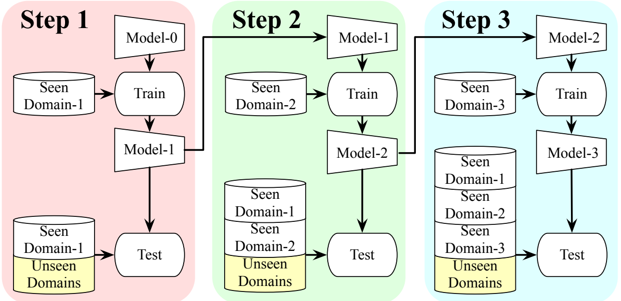

Figure 1: Pipeline of the proposed lifelong person re-identification task. The person identities among the involved domains are completely disjoint.

To tackle the challenges in LReLD, we propose a new *adaptive knowledge accumulation* (AKA) framework which can continually accumulate knowledge information from old domains, so as to have a better generalization quality on any new domain. This idea is inspired by a new perspective of human cognitive processes. Recent discoveries [4, 39] in cognitive science indicate that a cognitive process could be broadly decomposed into *"representations"* and *"operations"*. The structure of the knowledge representations (KRs) plays a key role for stabilizing memory, which shows our brain has potential relations with graph structure. Adaptive update and retrieval contained in the knowledge operations (KOs) promotes the efficient use of knowledge. Such complex yet elaborate KRs and KOs enable our brain to perform life-long learning well. Motivated by this, we endow AKA with two abilities to separately ac-

*Corresponding Author.

complish *knowledge representation* and *knowledge operation*. Specifically, we first represent transferable knowledge as a knowledge graph (KG), where each vertex represents one type of knowledge (e.g., the similar appearance between two persons). For image samples in one mini-batch, we temporally construct a similarity graph based on their relationships. Then, AKA establishes cross-graph links and executes a graph convolution. Such operation enables KG to transfer previous knowledge to each current sample. Meanwhile, KG is updated by summarizing the information underlying the relationships among current instances. Furthermore, for encouraging KG to improve learned representation while considering the forgetting problem, *plasticity loss* and *stability loss* are integrated to achieve an optimal balance for generalization on unseen domain. Our contributions are three-fold:

Task contribution. We exploit a new yet practical person ReID task, namely LReID, which considers person re-identification problem under a lifelong learning scenario.

Technical contribution. We propose a new AKA framework for LReID. AKA maintains a learnable knowledge graph to adaptively update previous knowledge, while transferring the knowledge to improve generalization on any unseen domains, with the plasticity-stability loss.

Empirical contribution. We provide a new benchmark and evaluation protocols for LReID. AKA shows promising improvements over other state-of-the-art methods.

2. Related Work

2.1. Person Re-identification Setups

As summarized in Tab. 1, previous person ReID works are performed in four different setups: 1) Fully-supervised (FS) methods investigate and exploit different network structures and loss functions [53, 31, 47, 30]; 2) Unsupervised domain adaption (UDA) is introduced to mitigate the domain gaps between source and target domain, caused by discrepancies of data distribution or image style [54, 38, 49, 55]; 3) Pure-unsupervised (PU) setting is less researched as it has to handle learning robust representation without using any label information [22]. 4) Domain generalization (DG) is an open-set problem. Lately, DG ReID task is explored by [35]. However, all the above setups do not address the lifelong learning challenge in our LReID.

The most related works [19] and [48] proposed an online-learning method for one-pass person ReID and a continual representation learning setting for bio-metric identification, respectively. However, both of them focused on intra-domain continual learning instead of our inter-domain incremental learning. Since there are relatively narrow domain gaps between the training and the testing set, their settings are less challenging for keeping learned knowledge while improving generalization.

Table 1: The comparison of fully-supervised (FS), unsupervised domain adaption (UDA), pure unsupervised (PU), domain generalization (DG), and lifelong person re-identification (LReID). "S." and "T." denote source and target domain, respectively.

Setup	Step	Train	Label	Test
FS [53]	one	S.	S.	S.
UDA [38]	one or two	S. & T.	S.	T.
PU [22]	one	S.	-	S.
DG [35]	one	all S.	all S.	T.
LReID	multiple	current S.	current S.	S. & T.

2.2. Lifelong Learning

Lifelong or incremental learning [29, 2, 28] dates back several decades, but now is attracting an ever-increasing attention due to impressive progresses in deep neural networks. Existing methods focus on common vision tasks like object recognition [2, 32], object detection [34] and image generation [42]. The key challenge for lifelong learning is *catastrophic forgetting*, which means that the model has performance degradation on previous tasks after training on new tasks. Existing methods can be divided into three categories, including knowledge distillation by the teacher-student structure [21], regularizing the parameter updates [45] when new tasks arrive, and storing or generating image samples of previous tasks [32, 42].

However, these methods are not suitable for LReID for various reasons. 1) The number of classes in ReID is much larger than that in conventional lifelong learning tasks, e.g., the popular benchmarks for them include MNIST [14], CORe50 [24], CIFAR-100 [13], CUB [37] and ImageNet [33]. Except ImageNet, other benchmarks are small-scale in terms of classes numbers. In contrast, one of the popular ReID benchmarks, MSMT17_V2 [40] includes 4,101 classes/identities. 2) ReID datasets are more imbalanced [23], that means the number of samples per class ranges from 2 to 30. Because model degradation typically happens when learning from tail classes, LReID also suffers from a few-shot learning challenge. 3) Similar with the fine-grained retrieval task [3]. The inter-class appearance variations for ReID are significantly subtler than generic classification tasks. It is particularly challenging in the lifelong learning scenario. 4) Previous works use the same classes for both training and testing, while ReID always need to handle with unseen classes. Fortunately, we find that remembering previously seen classes is beneficial for generalising on newly unseen classes.

3. Lifelong Person Re-Identification

3.1. Problem Definition and Formulation

In terms of LReID, one unified model needs to learn T domains in an incremental fashion. Suppose we have a stream of datasets $\mathcal{D} = \{D^{(t)}\}_{t=1}^{T}$. The dataset of the t-

th domain is represented as $D^{(t)} = \{D_{tr}^{(t)}, D_{te}^{(t)}\}$, where $D_{tr}^{(t)} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{\left|D_{tr}^{(t)}\right|}$ contains training images and their corresponding labels set $Y_{tr}^{(t)}$, and $D_{te}^{(t)}$ indicates the testing set with $Y_{te}^{(t)}$. The training and testing classes are disjoint, so that $Y_{tr}^{(t)} \cap Y_{te}^{(t)} = \emptyset$. Note that, only $D_{tr}^{(t)}$ is available at the t-th training step, and the data from previous domains are *not* available any more. For evaluation, we test retrieval performance on all encountered domains with their corresponding testing sets. In addition, the generalization ability is evaluated via new and unseen domains D_{un} with unseen identities Y_{un}. Henceforth, we will drop the subscript $\{tr, te\}$ for simplicity of notation.

3.2. Baseline Approach

We introduce a baseline solution based on knowledge distillation to address LReID. The baseline model consists of a feature extractor $h(\cdot;\theta)$ with parameters θ and a classifier $g(\cdot;\phi)$ with parameters ϕ. The whole network $f(\cdot;\theta,\phi)$ is the mapping from the input space directly to confidence scores, which is defined as: $f(\cdot;\theta,\phi) := g(h(\cdot;\theta);\phi)$. Training the parameters θ and ϕ in the network is optimized by a cross-entropy loss,

$$\mathcal{L}_c = -\sum_{(\mathbf{x},\mathbf{y})\in D} \mathbf{y} \log\left(\boldsymbol{\sigma}\left(f\left(\mathbf{x};\theta,\phi\right)\right)\right), \tag{1}$$

where $\boldsymbol{\sigma}$ is *softmax* function. In addition, we adopt the knowledge distillation (KD) [21] technique for mitigating forgetting on previous domains. Omitting the superscript (t), the loss function is defined as:

$$\mathcal{L}_d = \\ -\sum_{\mathbf{x}\in D}\sum_{j=1}^{n} \boldsymbol{\sigma}\left(f(\mathbf{x};\hat{\theta},\hat{\phi})\right)_j \log\left(\boldsymbol{\sigma}\left(f(\mathbf{x};\theta,\phi)\right)_j\right), \tag{2}$$

where $n = \sum_{i=1}^{t-1}\left|Y^{(i)}\right|$ is the number of the old classes, $\hat{\theta}$ and $\hat{\phi}$ are copied from θ and ϕ before current-step training, respectively. The total objective of baseline method is:

$$\mathcal{L}_{base} = \mathcal{L}_c + \gamma\mathcal{L}_d, \tag{3}$$

where γ is a trade-off factor for the knowledge distillation loss and the cross-entropy loss.

4. Adaptive Knowledge Accumulation

In this section, we introduce the details of the proposed AKA framework. The goal of AKA is to facilitate both learning process of new domain and generalization on unseen domains by leveraging transferable knowledge learned from previous domains. Referring to biological prior knowledge, AKA mimics the brain's cognitive process [4] to construct two sub-processes: *knowledge representation* and *knowledge operation*, illustrated by Fig. 2. In the following subsections, we elaborate both sub-processes and their optimization, respectively.

4.1. Knowledge Representation

To respectively represent the knowledge underling current samples, and the accumulated knowledge learned from already-trained domains, we parameterize the knowledge *"representations"* by constructing two different graph structures: *instance-based similarity graph* (ISG) and *accumulated knowledge graph* (AKG).

Instance-based Similarity Graph. Given a mini-batch of samples from a certain domain, the extracted features are defined as $\mathbf{V}^S = h(\mathbf{x};\phi)$. Inspired by [26], we first investigate the relationships among these samples and represent the relationships by a fully-connected graph $\mathcal{G}^S(\mathbf{A}^S, \mathbf{V}^S)$, namely ISG, where $\mathbf{A}^S$ is the edge set and the extracted features serve as vertices $\mathbf{V}^S$ in the graph. The edge weight $\mathbf{A}_{ij}^S$ between two vertices $\mathbf{V}_i^S$ and $\mathbf{V}_j^S$ is measured by a learnable L_1 distance between them:

$$\mathbf{A}_{ij}^S = \boldsymbol{\rho}\big(\mathbf{W}^S\left|\mathbf{V}_i^S - \mathbf{V}_j^S\right| + \mathbf{b}^S\big), \tag{4}$$

where $\mathbf{W}^S$ and $\mathbf{b}^S$ represent learnable parameters, and $\boldsymbol{\rho}$ is *Sigmoid* function. This is, the ISG is build by parameterized weight as shown in Fig. 2. For each mini-batch with N^b samples, AKA temporarily constructs a $\mathcal{G}^S$, in which $\mathbf{V}^S \in \mathbb{R}^{N^b\times d}$ denotes a feature set with dimensions d and $\mathbf{A}^S \in \mathbb{R}^{N^b\times N^b}$ gives the adjacency matrix. This matrix indicates the proximity between instances.

Accumulated Knowledge Graph. Furthermore, to represent accumulated knowledge, we construct an AKG, whose vertices represent different types of knowledge (e.g., the representative person appearance and structure) and edges are automatically constructed to reflect the relationship between such knowledge. Specifically, Given an vertex set $\mathbf{V}^K \in \mathbb{R}^{N^k\times d}$ and an adjacent matrix $\mathbf{A}^K \in \mathbb{R}^{N^k\times N^k}$, we define the knowledge graph as $\mathcal{G}^K(\mathbf{A}^K, \mathbf{V}^K)$, where N^k is the number of the AKG's vertices. To better explain the construction of the AKG, we first discuss the vertex representation $\mathbf{V}^K$. During domain-incremental training, domains arrive sequentially and their corresponding vertices representations are expected to be updated dynamically and timely. Therefore, the vertex representations of the AKG is parameterized and learned at the training time. Moreover, to encourage the diversity of knowledge encoded in the AKG, the vertex representations are randomly initialized. Analogous to the definition of weight in the ISG, the parameterized weight of AKG is defined as:

$$\mathbf{A}_{ij}^K = \boldsymbol{\rho}\big(\mathbf{W}^K(\left|\mathbf{V}_i^K - \mathbf{V}_j^K\right|) + \mathbf{b}^K\big), \tag{5}$$

where $\mathbf{W}^K$ and $\mathbf{b}^K$ represent learnable parameters.

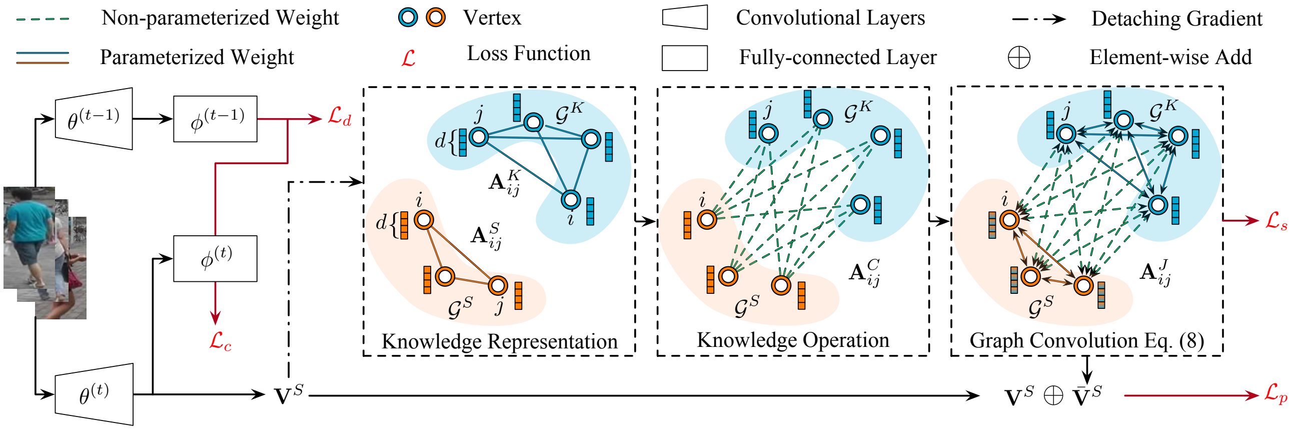

Figure 2: Overview of the proposed AKA framework. AKA maintains the AKG parameterized by ψ, to organize and memorize previous learned knowledge. Given a mini-batch images from a certain domain, similarity graph $\mathcal{G}^S$ is constructed by the extracted features $\mathbf{V}^S$. Meanwhile it taps into AKA to acquire relevant knowledge from $\mathcal{G}^K$, resulting in the vectored representations $\bar{\mathbf{V}}^S$ of acquired knowledge. Further, the required knowledge $\bar{\mathbf{V}}^S$ are summed with corresponding input features $\mathbf{V}^S$, which generates enhanced representation with better generalization capability.

Remark: The weights in $\mathcal{G}^S$ and $\mathcal{G}^K$ are calculated by independent learnable parameters, as the manners of knowledge organization in two graph have distinct differences. One focuses on the relationship among current samples. The other is required to consider both its own structure and efficient knowledge transformation. Such design is distinct different from the graph matching network [20] that shares same weights of two graphs like a Siamese network.

4.2. Knowledge Operation

Based on such knowledge representations, we further decompose the *"operations"* into *knowledge transfer* and *knowledge accumulation*, to enhance the learning of new domains with involvement of previous knowledge, and update these accumulated knowledge, correspondingly.

Knowledge Transfer. We first discuss how to organize and extract knowledge from the previous learning process and then explain how to leverage such knowledge to benefit the training of a new domain. The edges in $\mathcal{G}^S$ and $\mathcal{G}^K$ are also reserved in the joint graph $\mathcal{G}^J$. We connect $\mathcal{G}^S$ with $\mathcal{G}^K$ by creating links between the prototype-based relational graph and the knowledge graph. The cross-graph edge between a pair of vertices in $\mathcal{G}^S$ and $\mathcal{G}^K$ is weighted by the similarity between them. Specifically, for each instance pair $\mathbf{V}^S_i$ and $\mathbf{V}^K_j$, the cross-graph weight $\mathbf{A}^C_{ij}$ is calculated by applying a Softmax over Euclidean distances between $\mathbf{V}^S_i$ and $\mathbf{V}^K_j$, which is a non-parameterized similarity:

$$\mathbf{A}^C_{ij} = \frac{\exp(-\frac{1}{2}\left\|\mathbf{V}^S_i - \mathbf{V}^K_j\right\|^2_2)}{\sum_{k=1}^{N^k}\exp(-\frac{1}{2}\left\|\mathbf{V}^S_i - \mathbf{V}^K_k\right\|^2_2)}. \tag{6}$$

Taking Eq. 4, 5 and 6, the joint graph is formulated as:

$$\mathbf{A}^J = \begin{bmatrix} \mathbf{A}^S & \mathbf{A}^C \\ (\mathbf{A}^C)^T & \mathbf{A}^K \end{bmatrix}, \mathbf{V}^J = \begin{bmatrix} \mathbf{V}^S \\ \mathbf{V}^K \end{bmatrix}, \tag{7}$$

where the adjacent matrix $\mathbf{A}^J \in \mathbf{R}^{(N^b+N^k)\times(N^b+N^k)}$ and vertex matrix $\mathbf{V}^J \in \mathbf{R}^{(N^b+N^k)\times d}$ define joint graph $\mathcal{G}^J$.

After constructing the joint graph $\mathcal{G}^J$, we propagate the most related knowledge from $\mathcal{G}^K$ to $\mathcal{G}^S$ via a Graph Convolutional Network (GCN) [11], which is formulated as:

$$\mathbf{V}^G = \boldsymbol{\delta}\big(\mathbf{A}^J(\mathbf{V}^J\mathbf{W}^J)\big), \tag{8}$$

where $\mathbf{V}^G \in \mathbf{R}^{(N^b+N^k)\times d}$ is the vertex embedding after one-layer "message-passing" [5] and $\mathbf{W}^J$ is a learnable weight matrix of the GCN layer followed by a non-linear function $\boldsymbol{\delta}$, e.g., ReLU [1]. We employ only one layer to accomplish information propagation for simplicity, while it is natural to stack more GCN layers. After passing features through GCN, we obtain the information-propagated feature representation of the $\mathbf{V}^S$ from the top-N^b rows of $\mathbf{V}^G$, which is denoted as $\bar{\mathbf{V}}^S = \{\mathbf{V}^G_i | i \in [1, N^b]\}$.

Knowledge Accumulation. Maintaining a knowledge graph within limited storage resource during lifelong learning is inevitably expected to compact memorized knowledge and selectively update the AKG. To achieve this goal, we first aggregate $\mathbf{V}^S$ and $\bar{\mathbf{V}}^S$ by summing them, which results in a set of summed representation $\mathbf{F} = \left(\mathbf{V}^S + \bar{\mathbf{V}}^S\right)/2$. Then, to guide $\bar{\mathbf{V}}^S$ that improves the generalization of $\mathbf{V}^S$, we introduce a plasticity objective:

$$\mathcal{L}_p = \frac{1}{N^b}\sum_{(a,p,n)} \ln\Big(1+\exp\left(\Delta(\mathbf{F}_a,\mathbf{F}_p) - \Delta(\mathbf{F}_a,\mathbf{F}_n)\right)\Big), \tag{9}$$

where Δ denotes a distance function, e.g., L_2 distance and cosine distance. a, p and n donate the anchor, positive and negative instances in a mini-batch while we utilize an online hard-mining sampling strategy [44] to boost generalization capability of learned representation.

Furthermore, we observed that only encouraging the knowledge graph to adapt the current domain easily results in significant over-fitting, which would further lead to catastrophic forgetting. Thus, we propose a stability loss to punish the large movements of vertices in $\mathcal{G}^K$ when they update from the ending state $\hat{\mathbf{V}}^K$ of last training step:

$$\mathcal{L}_s = \frac{1}{N^k} \sum_{i=1}^{N^k} \ln\Big(1 + \exp\left(\Delta(\mathbf{V}_i^K, \hat{\mathbf{V}}_i^K)\right)\Big). \quad (10)$$

This loss term constrains the vertices in $\mathcal{G}^K$ to approximate their initial parameters. Eq. 9 and Eq. 10 are used to co-optimize the parameters of AKG but detaching the gradient flowing into CNN, which is discussed in Sec. 4.4. Through imposing such stability-plasticity dilemma, AKG accumulates more refine and general knowledge from comparison with previous knowledge, so as to generate better representation for generalizable ReID.

4.3. Optimization

According to [4, 39], when a visual cognitive process starts, our brain retrieves relevant representational content (knowledge) from high-dimensional memories based on similarity or familiarity. Then, our brain will summarize the captured information and update relevant knowledge or allocate new memory. Motivated by this, we query the ISG in the AKG to obtain the relevant previous knowledge. The ideal query mechanism is expected to optimize both graphs simultaneously at the training time and guide the training of both graphs to be mutual promotion. At the training step t, we train the whole model $\Theta^{(t)} = \{\theta^{(t)}, \phi^{(t)}, \psi^{(t)}\}$ on $D^{(t)}$ with mini-batch SGD and detaching the gradient between $\theta^{(t)}$ and ψ^t. The overall loss function is:

$$\mathcal{L}_{total} = \mathcal{L}_{base} + \lambda_p \mathcal{L}_p + \lambda_s \mathcal{L}_s, \quad (11)$$

where λ_s and λ_p are plasticity-stability trade-off factors. Here, we discuss how our proposed AKG works. When λ_p is relatively larger than λ_s, $\mathcal{G}^K$ focuses on learning new knowledge with minimal weight on taking into account previous knowledge. On the contrary, our model can only benefits for improving generalization in first two domain-incremental steps with approximately fixed vertices of knowledge graph. Intuitively, the optimal balance of these two terms not only ensures the stability of knowledge graph, but also endows AKG with a plasticity that allows new knowledge to be incorporated and accumulated.

4.4. Discussion

(1) Why does AKA respectively use non-parameterized and parameterized weight for knowledge operation and representation? In the sight of [12], the partial parameters of top layers favor becoming domain-specific during incremental training on different domains, which leads to severe performance degradation on previous domains. In addition, according to the biological inspiration [4], the representation and operation should be independent. To this end, when performing knowledge transformation, a non-parameterized metric allows model to treat different domains with less bias. As for the knowledge representation, summarizing and updating knowledge require the power of parameters.

(2) Why does AKA detach the gradient of GCN? As shown in Fig. 5, AKA without detaching gradient tends to transfer relatively similar knowledge through all training domains, which is caused by the degradation of GCN [9]. However, detaching the gradient encourages AKA to learn independently so that AKA enables to adaptively generate different knowledge for different domains.

(3) Why is the proposed straightforward $\mathcal{L}_s$ efficient? Intuitively, the unity of $\mathcal{L}_s$ and $\mathcal{L}_p$ forms a bottleneck mechanism, which forces $\mathcal{G}^K$ to learn sparse knowledge from each domain. In this work, we utilize a simple yet effective method, restricting the vertices only, to preserve knowledge. Even though the vertices are almost fixed, the weight of transferable knowledge is learnable. Ideally, $\mathcal{G}^K$ could adaptively modify the transformation weight so as to reorganize old knowledge for representing new knowledge. That means we maintain the topology of vertices and leverage flexible non-parameter transformation to adapt feature representations in a new environment.

5. Experiments

5.1. Implementation Details

We remove the last classification layer of ResNet-50 and use the retained layers as the feature extractor to yield 2048-dimensional features. The AKA network consists of one GCN layer. In each training batch, we randomly select 32 identities and sample 4 images for each identity. All images are resized to 256 × 128. Adam optimizer with learning rate 3.5×10^{-4} is used. The model is trained for 50 epochs, and decrease the learning rate by × 0.1 at the 25^{th} and 35^{th} epoch. We follow [48] to set the balance weight γ as 1, and explore the effect of other hyper-parameters. The N^K, λ_p, and λ_s are set as 64, 1, and 10, respectively. The hyper-parameter analysis is given in Sec. 5.5. The retrieval of testing data is based on Euclidean distance of feature embeddings. For all experiments, we repeat five times and report means and standard deviations.

5.2. New Benchmark for LReID

We present a new and large-scale benchmark including LReID-Seen and LReID-Unseen subsets. The presented benchmarks are different from existing ReID benchmarks in three main aspects: 1) The proposed LReID benchmarks are specifically designed for person re-identification that is

Table 2: The statistics of ReID datasets involved in our experiments. '*' denotes that we modified the original dataset by using the ground-truth person bounding box annotation for our lifelong ReID experiments rather than using the original images which were originally used for person search evaluation. '-' denotes these data are not used for lifelong training.

Benchmark	Datasets Name	Scale	Original Identities			Selected Identities		
			Train	Query	Gallery	Train	Query	Gallery
LReID-Seen	CUHK03[18]	mid	767	700	700	500	700	700
	Market-1501[52]	large	751	750	751	500	751	751
	MSMT17_V2 [40]	large	1041	3060	3060	500	3060	3060
	DukeMTMC-ReID[54]	large	702	702	1110	500	702	1110
	CUHK-SYSU ReID*[43]	mid	942	2900	2900	500	2900	2900
LReID-Unseen	VIPeR[6]	small	316	316	316	-	316	316
	PRID[8]	small	100	100	649	-	100	649
	GRID[25]	small	125	125	126	-	125	126
	i-LIDS[41]	small	243	60	60	-	60	60
	CUHK01[17]	small	485	486	486	-	486	486
	CUHK02[16]	mid	1677	239	239	-	239	239
	SenseReID[50]	mid	1718	521	1718	-	521	1718

the fine-grained retrieval task, while existing lifelong learning benchmarks focus on general image classification; 2) The total number of classes in our benchmark ($|Y| \approx$14K) is much larger than existing benchmarks ($\leq$ 1K); 3) We test the model on novel identities that have never appeared in the training set even on unseen domains, while existing benchmarks test on new images of learned (seen) classes.

LReID-Seen. In total, 40,459 training images of the 2,500 identities are employed for the lifelong training set. The training identities are uniformly split into 5 subsets in accordance with their domains, for 5-step domain-incremental training. Their original testing sets are kept to evaluate the model's domain forgetting and performance of the current domain. Specifically, we selected five relatively large-scale datasets, CUHK03 (CU) [18], Market-1501 (MA) [52], MSMT17_V2 (MS) [40], DukeMTMC-ReID (DU) [54] and CUHK-SYSU ReID (SY) [43], and sampled 500 identities from each of their training sets to construct five training domains so that each domain has an equal number of classes. Note that for the SY [43] dataset, we modified the original dataset by using the ground-truth person bounding box annotation and selected a subset in which each identity includes at least 4 bounding boxes, rather than using the original images which were originally used for person search evaluation. For testing on this dataset, we fixed both query and gallery sets instead of using variable gallery sets. We used 2,900 query persons, with each query containing at least one image in the gallery, which resulted in 942 training identities, called CUHK-SYSU ReID in Tab. 2.

LReID-Unseen. To verify raising the model's abilities resulting from progressively accumulated knowledge from previous domains, we reorganize 7 popular person ReID datasets as shown in Tab. 2. Specifically, we first merge VIPeR [6], PRID [8], GRID [25], i-LIDS [41], CUHK01 [17], CUHK02 [16], SenseReID [50] in accordance with their original train/test splits as a new benchmark. Then, the merged test set, including 3,594 different identities with total 9,854 images, is adopted to evaluate the generalization ability of learned features on unseen domain, called LReID-Unseen in Tab. 2.

Evaluation metrics. We use $\bar{u}$ (average performance on unseen domains) to measure the capacity of generalising on unseen domains and $\bar{s}$ (average performance on seen domains) to measure the capacity of retrieving incremental seen domains. Note that the performance gap of $\bar{s}$ between joint training and a certain method indicates the method's ability to prevent forgetting. $\bar{u}$ and $\bar{s}$ are measured with mean average precision (mAP) and rank-1 (R-1) accuracy. These metrics are calculated after the last training step.

5.3. Seen-domain Non-forgetting Evaluation

Less forgetting performance refers to the effectiveness of one method which mitigates the accuracy degradation on previous domains. We evaluated AKA on LReID task against the state-of-the-art. The methods for comparison include 1) sequential fine-tuning (SFT): Fine-tuning model with new datasets without distilling old knowledge; 2) learning without forgetting (LwF): The baseline method [21] introduced in Sec. 3.2; 3) similarity-preserving distillation (SPD): A competitor with advanced feature distillation [36]; 4) Continual representation learning (CRL) [48]: We first reproduce their method and achieve the reported results on their published benchmark. Then, we apply their methods to our domain-incremental person ReID benchmark and report these new results in Table. 3; 5) Joint-CE serves as an upper-bound by training model on all data of the seen domains with $\mathcal{L}_c$. For a fair comparison, SFT-T, CRL-T and Joint-CE denote directly adding the widely-used triplet loss [7] for co-optimizing learned features.

In practice, the order of input domains is agnostic. Thus, we investigate the influence caused by different training orders and analyze two representative results. *Order-1* and *Order-2* are denoted by **MA→SY→DU→MS→CU** and **DU→MS→MA→SY→CU**, respectively. As shown in

Table 3: Seen-domain non-forgetting evaluation. We test model after sequentially training on all seen domains (t=5). Each experiment is repeated by 5 times to report mean and std of all seen domains. The training order is MA→SY→DU→MS→CU.

Method	Market		SYSU		Duke		MSMT17		CUHK03		$\bar{s}$	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
SFT	24.1±0.2	48.5±0.4	30.5±0.3	32.7±0.3	14.4±0.2	27.0±0.2	12.0±0.3	30.1±0.3	45.6±0.2	48.5±0.3	25.3	37.4
SFT-T	25.8±0.3	48.4±0.5	32.0±0.4	34.8±0.5	15.1±0.4	27.7±0.5	13.8±0.3	32.4±0.4	**48.0**±0.4	**50.1**±0.5	26.9	38.7
SPD	30.5±0.3	50.7±0.4	37.6±0.4	39.9±0.4	14.6±0.3	27.2±0.3	12.2±0.3	30.3±0.2	40.7±0.3	42.5±0.3	27.1	38.1
LwF	47.1±0.3	65.1±0.4	43.9±0.2	44.3±0.2	16.0±0.3	28.0±0.2	14.8±0.2	33.6±0.3	26.1±0.2	26.2±0.2	29.6	39.4
CRL	48.5±0.5	66.6±0.3	45.2±0.2	43.3±0.3	16.2±0.2	27.9±0.3	16.1±0.3	34.3±0.2	28.1±0.3	29.8±0.4	30.8	40.4
CRL-T	49.2±0.5	67.0±0.3	45.6±0.2	43.9±0.4	15.9±0.3	27.5±0.4	15.8±0.3	33.9±0.4	26.5±0.3	26.7 ±0.4	30.6	39.8
AKA	**51.2**±0.2	**72.0**±0.3	**47.5**±0.5	**45.1**±0.6	**18.7**±0.3	**33.1**±0.4	**16.4**±0.2	**37.6**±0.3	27.7±0.4	27.6±0.5	**32.3**	**43.1**
Joint-CE	71.9±0.2	83.2±0.2	61.2±0.3	62.5±0.3	65.1±0.2	76.8±0.3	25.3±0.3	50.7±0.5	48.7±0.1	50.3±0.2	54.4	64.7
Joint-CE-T	74.8±0.2	87.0±0.3	63.3±0.3	65.5±0.4	68.3±0.2	80.1±0.3	27.9±0.3	54.1±0.5	50.8±0.2	56.6±0.2	57.0	67.7

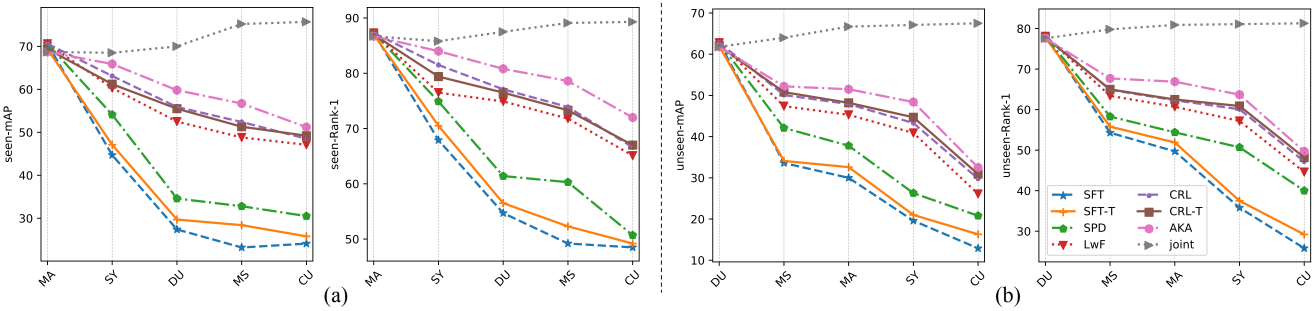

Figure 3: Illustration of seen-domain non-forgetting evaluation. (a) depicts the trend of mAP and Rank-1 score on the first training domain during training process following *Order-1*. Likewise, (b) shows the results of *Order-2*.

Fig. 3, training order significantly impacts the model's ability to prevent forgetting. Specifically, for *Order-1*, AKA ranks the first with accuracy degradation of 17.5%/14.7% in mAP/R-1, which demonstrates that AKA is able to preserve old knowledge while mitigating catastrophic forgetting. In comparison, AKA outperforms SFT by around 30% in R-1 and is superior to most competitive CRL by 6% in mAP. Note that SFT-T and CRL-T (with additional triplet loss) is not beneficial for the first three training steps, because when the number of training identities is large enough, triplet loss contributes less on performance and even leads to conflict with cross-entropy loss [51]. On the other hand, KD-based methods are obviously superior to feature distillation or SFT methods. For *Order-2*, AKA ranks the first with performance degradation of 29.3%/27.9% in mAP/R-1 as well.

5.4. Unseen-domain Generalising Evaluation

To demonstrate that our LRe-ID is more challenging than the latest CRL-ReID [48] task, we re-implement their method and evaluate on both their CRL-ReID dataset [48] and our LReID-Unseen benchmarks. Despite our setting needs to overcome larger domain gaps, our AKA can automatically transfer and update knowledge based on different input. Thus, the results shown in the first two rows of Tab. 4 indicate that LRe-ID setting is more difficult and our method outperforms the compared methods significantly.

For the experiments on LReID-Unseen, we assumed that a model was sequentially trained with the *Order-1*. Then, we report all results in the final step when all domains are trained. As shown in Tab. 4, AKA achieves best performance compared with other competitive methods. Specifically, AKA achieves averaged 31.8% mAP on seen domains and averaged 44.3% mAP on unseen domains, which are significantly better than the baseline methods. Interestingly, as shown in Fig. 4, the methods without KD reach a better performance on 2^{nd} step, but they fail to accumulate previous knowledge to further improve generalization ability. The similar phenomenon appears in *order-2* as well. However, our results are still obviously lower than the upper-bound. The gap indicates the challenges of LReID on the proposed benchmark.

5.5. Ablation Study

We conduct two groups of ablation experiments to study the effectiveness of our method. One is to verify the improvement of adding the AKG module. Our full method AKA is composed of LwF and AKG. Comparing the performances of LwF and AKA in Tab. 3, our AKA achieves 6% improvement on both mAP and less forgetting score. The other group is to demonstrate the importance of our proposed stability and plasticity loss. In Tab. 5, "Baseline" setting is the same as the LwF method. "Baseline + $\mathcal{L}_p$" denotes LwF method added our AKG with only plasticity loss. The "Baseline + $\mathcal{L}_p$ + $\mathcal{L}_s$ " setting indicates our full method. As shown in Tab. 5, $\mathcal{L}_p$ is beneficial for only unseen domains, and $\mathcal{L}_p$ and $\mathcal{L}_s$ are complementary. The improvement of adding $\mathcal{L}_s$ indicates that greater stability of knowledge can preserve the knowledge of previous

Table 4: Unseen-domain generalising evaluation. We refer to corresponding literature and reproduce experimental results on our setting.For LReID-Unseen, the training order is MA→SY→DU→MS→CU.

Banchmark	$\bar{u}$	SFT	SFT-T	SPD	LwF	CRL	CRL-T	AKA	Joint-CE	Joint-CE-T
CRL-ReID (5-step)	mAP	44.2 ± 0.2	44.7 ± 0.3	47.1 ± 0.2	48.7 ± 0.2	51.2 ± 0.1	51.5 ± 0.2	**64.2** ± 0.1	64.8 ± 0.1	66.7 ± 0.1
	R-1	53.4 ± 0.3	53.9 ± 0.4	54.1 ± 0.4	59.6 ± 0.2	62.8 ± 0.3	63.1 ± 0.3	**74.9** ± 0.3	75.3 ± 0.1	78.6 ± 0.2
CRL-ReID (10-step)	mAP	31.7 ± 0.2	31.7 ± 0.3	40.3 ± 0.3	42.8 ± 0.2	43.8 ± 0.3	44.1 ± 0.1	**49.7** ± 0.2	64.8 ± 0.1	66.7 ± 0.1
	Rank-1	40.3 ± 0.4	40.5 ± 0.5	47.5 ± 0.4	51.7 ± 0.1	54.7 ± 0.4	54.8 ± 0.3	**58.8** ± 0.2	75.3 ± 0.1	78.6 ± 0.2
LReID-Unseen	mAP	35.2± 0.2	37.1± 0.4	36.3± 0.2	38.3± 0.2	38.5± 0.2	39.6± 0.4	**44.3**± 0.2	50.6± 0.1	53.5± 0.2
	R-1	31.1± 0.3	34.3± 0.4	32.9± 0.2	36.9± 0.3	36.7± 0.2	38.1± 0.4	**40.4**± 0.3	48.1 ± 0.1	50.0 ± 0.3

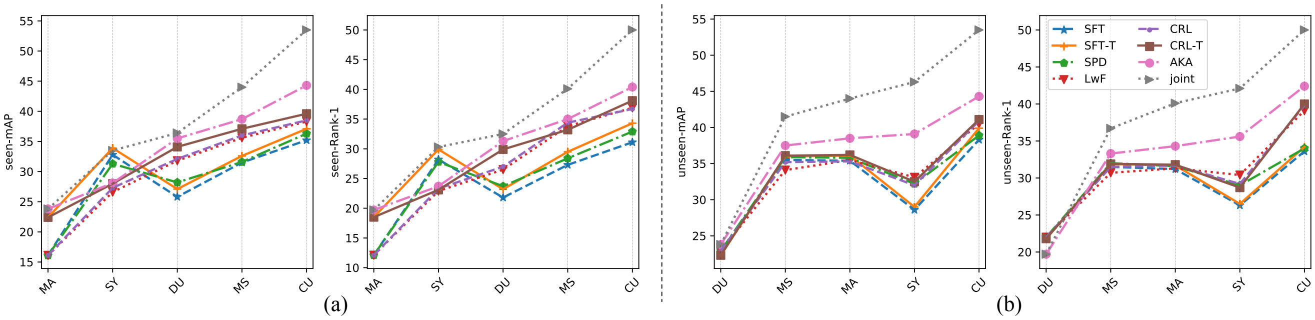

Figure 4: Illustration of unseen-domain generalising evaluation. (a) depicts the trend of mAP and Rank-1 score on unseen domains during training process following *Order-1*. Likewise, (b) shows the results of *Order-2*.

Table 5: Effectiveness of the proposed loss functions.

Setting	$\bar{s}$		$\bar{u}$	
	mAP	R-1	mAP	R-1
Baseline	29.6	39.4	38.3	36.9
Baseline + $\mathcal{L}_p$	29.5	39.6	41.6	38.3
Baseline + $\mathcal{L}_p$ + $\mathcal{L}_s$ (Full)	**32.3**	**43.1**	**44.3**	**40.4**
Full w/o $\mathcal{L}_d$	28.5	39.1	42.1	38.9

domains, which remits the unfavourable influence of catastrophic forgetting to some extent. Moreover, the improvement of adding $\mathcal{L}_p$ indicates AKG is encouraged to learn how to transfer positive knowledge to improve generalization. When λ_p becomes large enough, the model overfits on generating the same representation with the output of CNN.

Hyper-parameter analysis. The hold-off validation data are used to determine two hyper-parameters λ_p and λ_s. We first select the optimal λ_p to achieve best $\bar{u}$, then we choose the optimal λ_s based on the selected λ_p. Finally, when $\lambda_p = 1$ and $\lambda_s = 5 \times 10^{-4}$, our model achieves best balance between seen and unseen domains. Afterwards, we keep other hyper-parameters and explore the influence of $N^K \in \{32, 64, 128, 256, 512\}$ for $\bar{u}$ and $\bar{s}$ metrics calculated by mAP. The results shown in Fig. 5 indicate that N^K is not sensitive and $\bar{u}$ increases with the growth of N^K. Thus, we balance memory consumption and generalization performance, and set $N^K = 64$ in all of our experiments.

6. Conclusion

We focus on an unsolved, challenging, yet practical domain-incremental scenario, namely lifelong person re-identification, where models are required to improve generalization capability on both seen and unseen domains by leveraging previous knowledge. Hence, we propose a new AKA framework to preserve the knowledge learned from previous domains while adaptively propagating the previously learned knowledge for improving learning on new domains. Extensive experiments show that our method outperforms other competitors in terms of both mitigating forgetting on seen domains and generalising on unseen domains.

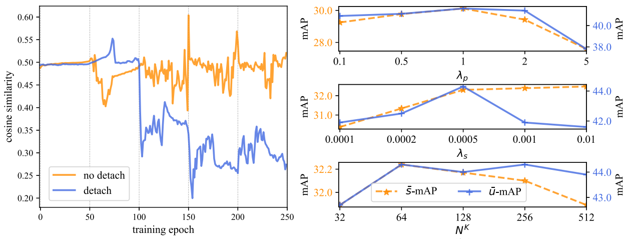

Figure 5: To investigate the effectiveness of detaching gradient, we visualize the normalized cosine similarity between $\mathbf{V}^S$ and $\bar{\mathbf{V}}^S$ during training processing in (a). The three rows in (b) study the effects of hyper-parameters λ_p, λ_s and N^K, respectively.

Acknowledgements

This work was supported mainly by the LIACS Media Lab at Leiden University and in part by the China Scholarship Council and the Fundamental Research Funds for the Central Universities. Finally, I would like to thank my wife Dr. Yuzhi Lai who gave me the invaluable love, care and encouragement in the past years.

References

[1] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018. 4

[2] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *ECCV*, pages 233–248, 2018. 2

[3] Wei Chen, Yu Liu, Weiping Wang, Tinne Tuytelaars, Erwin M Bakker, and Michael Lew. On the exploration of incremental learning for fine-grained image retrieval. In *BMVC*, 2020. 2

[4] Rosemary A Cowell, Morgan D Barense, and Patricnow S Sadil. A roadmap for understanding memory: Decomposing cognitive processes into operations and representations. *Eneuro*, 6(4), 2019. 1, 3, 5

[5] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017. 4

[6] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275, 2008. 6

[7] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 6

[8] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *scandinavian conference on image analysis*, pages 91–102. Springer, 2011. 6

[9] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*, 2020. 5

[10] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *CVPR*, pages 3143–3152, 2020. 1

[11] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017. 4

[12] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 5

[13] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. 1, 2

[14] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2

[15] Qingming Leng, Mang Ye, and Qi Tian. A survey of open-world person re-identification. *IEEE Trans. Circuit Syst. Video Technol.*, 30(4):1092–1108, 2019. 1

[16] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, pages 3594–3601, 2013. 6

[17] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*, 2012. 6

[18] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, pages 152–159, 2014. 6

[19] Wei-Hong Li, Zhuowei Zhong, and Wei-Shi Zheng. One-pass person re-identification by sketch online discriminant analysis. *Pattern Recognition*, 93:237–250, 2019. 2

[20] Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. Graph matching networks for learning the similarity of graph structured objects. In *ICML*, pages 3835–3845. PMLR, 2019. 4

[21] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(12):2935–2947, 2017. 2, 3, 6

[22] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, volume 33, pages 8738–8745, 2019. 2

[23] Jialun Liu, Yifan Sun, Chuchu Han, Zhaopeng Dou, and Wenhui Li. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In *CVPR*, pages 2970–2979, 2020. 2

[24] Vincenzo Lomonaco and Davide Maltoni. Core50: a new dataset and benchmark for continuous object recognition. *arXiv preprint arXiv:1705.03550*, 2017. 2

[25] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *Int. J. Comput. Vis.*, 90(1):106–129, 2010. 6

[26] Chuanchen Luo, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Spectral feature transformation for person re-identification. In *ICCV*, pages 4976–4985, 2019. 3

[27] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989. 1

[28] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019. 2

[29] Anastasia Pentina and Christoph Lampert. A pac-bayesian bound for lifelong learning. In *ICML*, pages 991–999, 2014. 2

[30] Angelo Porrello, Luca Bergamini, and Simone Calderara. Robust re-identification by multiple views knowledge distillation. In *ECCV*, pages 93–110, 2020. 1, 2

[31] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Dual gaussian-based variational subspace disentanglement for visible-infrared person re-identification. In *ACM MM*, pages 2149–2158, 2020. 1, 2

[32] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, pages 2001–2010, 2017. 1, 2

[33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large

scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015. 1, 2

[34] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. Incremental learning of object detectors without catastrophic forgetting. In *ICCV*, pages 3400–3409, 2017. 2

[35] Jifei Song, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Generalizable person re-identification by domain-invariant mapping network. In *CVPR*, pages 719–728, 2019. 1, 2

[36] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *ICCV*, pages 1365–1374, 2019. 6

[37] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[38] Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *CVPR*, pages 10981–10990, 2020. 1, 2

[39] Wei-Chun Wang, Nadia M Brashier, Erik A Wing, Elizabeth J Marsh, and Roberto Cabeza. Knowledge supports memory retrieval through familiarity, not recollection. *Neuropsychologia*, 113:14–21, 2018. 1, 5

[40] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88, 2018. 1, 2, 6

[41] Zheng Wei-Shi, Gong Shaogang, and Xiang Tao. Associating groups of people. In *BMVC*, pages 23–1, 2009. 6

[42] Chenshen Wu, Luis Herranz, Xialei Liu, Joost van de Weijer, Bogdan Raducanu, et al. Memory replay gans: Learning to generate new categories without forgetting. In *NeurIPS*, pages 5962–5972, 2018. 2

[43] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850*, 2(2), 2016. 6

[44] Mang Ye, Jianbing Shen, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. Augmentation invariant and instance spreading feature for softmax embedding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. 4

[45] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. *arXiv preprint arXiv:1708.01547*, 2017. 2

[46] Hong-Xing Yu and Wei-Shi Zheng. Weakly supervised discriminative feature learning with state information for person identification. In *CVPR*, pages 5527–5537, 2020. 1

[47] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *CVPR*, pages 3186–3195, 2020. 1, 2

[48] Bo Zhao, Shixiang Tang, Dapeng Chen, Hakan Bilen, and Rui Zhao. Continual representation learning for biometric identification. *arXiv preprint arXiv:2006.04455*, 2020. 2, 5, 6, 7

[49] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *ECCV*, pages 526–544, 2020. 1, 2

[50] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *CVPR*, pages 1077–1085, 2017. 6

[51] Feng Zheng, Cheng Deng, Xing Sun, Xinyang Jiang, Xiaowei Guo, Zongqiao Yu, Feiyue Huang, and Rongrong Ji. Pyramidal person re-identification via multi-loss dynamic training. In *CVPR*, pages 8514–8522, 2019. 7

[52] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015. 1, 6

[53] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016. 1, 2

[54] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *ICCV*, pages 3754–3762, 2017. 1, 2, 6

[55] Yang Zou, Xiaodong Yang, Zhiding Yu, BVK Kumar, and Jan Kautz. Joint disentangling and adaptation for cross-domain person re-identification. In *ECCV*, pages 87–104, 2020. 1, 2