Anti-Forgetting Adaptation for Unsupervised Person Re-identification

Hao Chen, Francois Bremond, Nicu Sebe, Shiliang Zhang

***Abstract*—Regular unsupervised domain adaptive person re-identification (ReID) focuses on adapting a model from a source domain to a fixed target domain. However, an adapted ReID model can hardly retain previously-acquired knowledge and generalize to unseen data. In this paper, we propose a Dual-level Joint Adaptation and Anti-forgetting (DJAA) framework, which incrementally adapts a model to new domains without forgetting source domain and each adapted target domain. We explore the possibility of using prototype and instance-level consistency to mitigate the forgetting during the adaptation. Specifically, we store a small number of representative image samples and corresponding cluster prototypes in a memory buffer, which is updated at each adaptation step. With the buffered images and prototypes, we regularize the image-to-image similarity and image-to-prototype similarity to rehearse old knowledge. After the multi-step adaptation, the model is tested on all seen domains and several unseen domains to validate the generalization ability of our method. Extensive experiments demonstrate that our proposed method significantly improves the anti-forgetting, generalization and backward-compatible ability of an unsupervised person ReID model.**

***Index Terms*—Re-identification, incremental learning, contrastive learning, domain generalization, backward compatible representation.**

I. Introduction

PERSON re-identification (ReID) [3] targets at matching a person of interest across non-overlapping cameras. Although significant improvement has been witnessed in both supervised [4], [5] and unsupervised domain adaptive [2], [6] person ReID, traditional methods only consider the performance of a single fixed target domain. In the single target domain scenario, people usually assume that all training data are available before training a ReID model. However, a real-world video monitoring system can record new data every day and from new locations, when new cameras are installed into an existing system. When a model needs to be frequently updated with new data, regular unsupervised person ReID methods can lead to three problems: 1) Once adapted to a new domain, a model is prone to lose the acquired knowledge of previous seen domains. 2) A model can hardly learn domain-shared features and generalize to unseen domains. 3) An adapted retrieval model usually shows low backward-compatible ability.

Firstly, as the weather and season are repetitive, losing previous knowledge usually results in a less robust ReID system. To tackle the forgetting problem, lifelong person ReID [7], [8] has been recently proposed to incrementally accumulate domain knowledge from several seen datasets. Lifelong person ReID is closely related to incremental (or continuous) learning [9]. There are three fundamental scenarios for incremental learning, including class-incremental learning, domain-incremental learning, and task-incremental learning. Facing non-stationary streams of data, lifelong person ReID is supposed to simultaneously learn incrementally added new classes and new domain knowledge, which can be defined as a joint class-incremental and domain-incremental learning task. However, previous supervised lifelong person ReID [7],

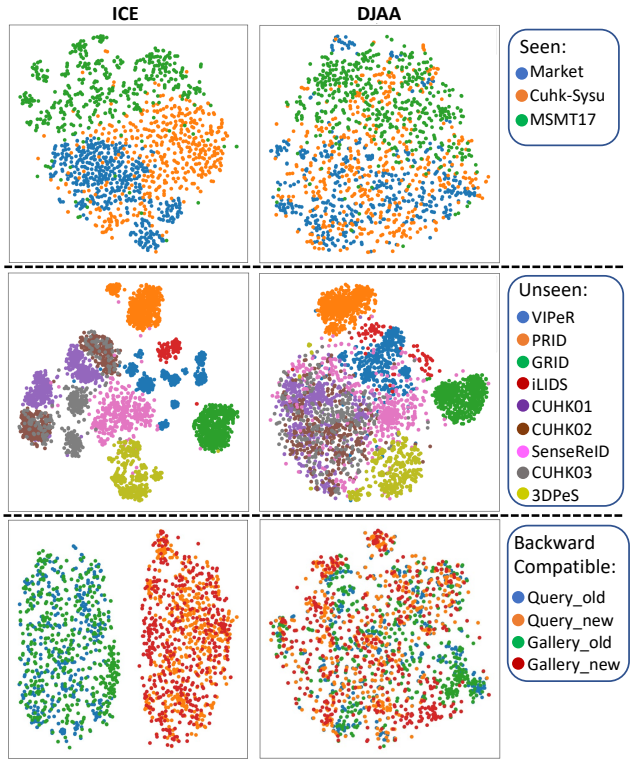


Fig. 1. ReID feature space comparison [1] of regular unsupervised ReID method ICE [2] and unsupervised lifelong ReID method DJAA on three scenarios. 1) **Seen domain non-forgetting ability**: DJAA preserves previously acquired knowledge, which reduces the gap in the feature space between seen domains. 2) **Unseen domain generalization ability**: DJAA accumulates domain-shared features that reduce the gap between unseen domains. 3) **Backward-compatible ability**: With DJAA, updated query (Query_new) and gallery (Gallery_new) representations remain in the same feature space with the previously extracted representations (Query_old, Gallery_old). Query_old and Gallery_old are Market1501 representations extracted after one adaptation step, while Query_new and Gallery_new are Market1501 representations extracted after three adaptation steps.

H. Chen and S. Zhang are with Peking University, No.5 Yiheyuan Road, Beijing 100871, China. E-mail: {hchen, slzhang.jdl}@pku.edu.cn

F. Bremond is with Inria, 2004 Route des Lucioles, 06902 Valbonne, France. E-mail: francois.bremond@inria.fr

N. Sebe is with University of Trento, Via Sommarive 9 - 38123 Povo - Trento, Italy.E-mail: niculae.sebe@unitn.it

[8] relies on human annotations for cross-domain fine-tuning. Replacing supervised cross-domain fine-tuning with unsupervised domain adaptation can substantially enhance the flexibility of a lifelong person ReID algorithm in real-world deployments.

Secondly, to enhance the model generalization ability on unseen domain, domain generalization methods [10] jointly train a model on several domains to learn domain-shared features. Regular domain generalization requires that several large-scale domain data are available before training a model. Lifelong person ReID can be an alternative for learning a generalizable model, which does not require collecting all data in advance. By regularizing the model consistency between old and new domains, we gradually accumulate domain-shared information into a unified model. With domain-shared features, an incrementally trained model [8] has proven to be generalizable on unseen domains.

Thirdly, after a domain adaptation step, an updated person ReID model also faces backward-compatiblility [11], [12] problem. To obtain optimal performance on new data, a lifelong adaptive person ReID system needs to be frequently updated. During the lifelong adaptation, feature representations of previous gallery images need to be re-extracted to maintain a consistent feature space for the pairwise distance calculation. However, re-extracting features after each adaptation step can be time-consuming. An optimal setting for lifelong person ReID system is that an updated model only extracts feature representations from queries and incoming gallery images, while keeping previous gallery image representations unchanged. How to improve the feature consistency between a model and its updated versions remains a key problem for building an efficient system, which is neglected by previous lifelong person ReID methods. We show that anti-forgetting and backward-compatible ability can be unified into the same question for lifelong person ReID, i.e., preserving the representation similarity relationship.

In this paper, we propose an unsupervised lifelong person ReID method to simultaneously address the forgetting, generalization ability, and backward-compatibility problems. Unsupervised lifelong person ReID is a challenging yet practical problem. Compared with generic lifelong domain adaptation [13], [14], unsupervised lifelong person ReID faces more technical problems. On the one hand, lifelong domain adaptation usually considers fixed classes across different styles, while unsupervised lifelong person ReID has to deal with non-overlapping identities across domains. On the other hand, unsupervised lifelong ReID aims at learning fine-grained human features, which are usually more sensitive to domain changes than generic object features. Our objective is to train a robust person ReID model that is discriminative to each seen domain while being generalizable to unseen domains.

Inter-instance Contrastive Encoding (ICE) [2] formulates image-to-prototype and image-to-image contrastive learning to enhance unsupervised representation learning for person ReID. The proposed dual-level contrastive learning has proven to be effective in enhancing intra-cluster compactness and inter-cluster separability. ICE shows remarkable performance in unsupervised person ReID. However, when adapted to a new domain, ICE rapidly loses previously acquired knowledge from the source domain. In this paper, we extend ICE to unsupervised lifelong person ReID to enhance the anti-forgetting, generalization and backward-compatible ability during the multi-step adaptation, as shown in Fig. 1. We incorporate pseudo-label based contrastive learning and rehearsal-based incremental learning into a Dual-level Joint Adaptation and Anti-forgetting (DJAA) framework, which enhances anti-forgetting, generalization and backward-compatible ability at the same time.

Our proposed DJAA consists of an adaptation module and a rehearsal module. In the adaptation module, we first use a clustering algorithm to assign pseudo labels to unlabeled data. Based on the pseudo labels, we use an image-to-prototype contrastive loss to make images with the same pseudo-label converge to a unique cluster prototype. We also use an image-to-image contrastive loss to further reduce the distance between hard positives. In the rehearsal module, a small number of old domain samples and the corresponding cluster prototypes are selected and stored in a long-term memory buffer. While adapting a model to a new domain, regularizing the image-to-image and image-to-prototype similarity relationship helps to prevent forgetting previous knowledge. Given a frozen old domain model and a trainable model, we set a consistency regularization condition: the image-to-image and image-to-prototype similarity calculated by the frozen old domain model and the current domain model should be consistent during the adaptation. Based on this condition, we regularize the representation similarity relationship with both image-level and cluster-level similarity between the frozen and the trainable models, so that the trainable current domain model can be updated in a way that suits old knowledge.

The main contributions of our work are four folds. 1) We comprehensively investigate the forgetting, generalization, and backward-compatible problems in domain adaptive person ReID. Our study shows that, those three challenges could be jointly addressed in a Dual-level Joint Adaptation and Anti-forgetting (DJAA) framework. To the best of our knowledge, this is an original person ReID work addressing those three challenges within a unified framework. 2) We propose an adaptation module that combines both cluster-level and instance-level contrastive losses to learn new domain features. 3) We propose a rehearsal module to retain previously-acquired knowledge during the domain adaptation. By introducing data augmentation and domain gap perturbations, we regularize the representation relationship at both cluster and instance levels. 4) Extensive experiments with various setups have been conducted to validate the effectiveness of our proposed method. DJAA shows remarkable non-forgetting, generalization, and backward-compatible ability on mainstream person ReID datasets.

II. Related Work

A. Person ReID

Depending on the number of training/test domains and availability of human annotation, recent person ReID research is conducted under different settings, as shown in Table I.

TABLE I
COMPARISON OF SUPERVISED (S), UNSUPERVISED DOMAIN ADAPTATION (UDA), DOMAIN GENERALIZATION (DG), SUPERVISED LIFELONG (SL) AND UNSUPERVISED LIFELONG (UL) REID SETTINGS. 'SD', 'TD' AND 'UD' RESPECTIVELY REFER TO SOURCE DOMAIN, TARGET DOMAIN AND UNSEEN DOMAINS.

Setting	Domain	Train	Label	Test
S	one	TD	TD	TD
UDA	two	SD&TD	SD	TD
DG	multi	all SD	SD	UD
SL	multi	one by one	SD	SD&UD
UL	multi	one by one	None	SD&UD

As the most studied setting, supervised person ReID [5], [15]–[17] has shown impressive performance on large-scale datasets thanks to deep neural networks and human annotation. However, as a fine-grained retrieval task, a ReID model trained on one domain can hardly generalize to other domains. Unsupervised domain adaptation methods [2], [18]–[22] are proposed to adjust a ReID model to a target domain with unlabeled target domain images. To maximize the model generalization ability on unseen data, domain generalization ReID [10], [23]–[27] is proposed to jointly train multiple labeled domains, in order to learn a generalizable model that can extract domain-invariant features. The above-mentioned settings simply assume that all training data is available before training. However, in most real-world cases, it is hard to prepare enough diversified data to directly train a generalizable model. Instead, new domain data can be recorded when time and season change or a new camera is installed. Supervised lifelong person ReID [7], [8], [28]–[30] is thus proposed to learn incrementally added data without forgetting previous knowledge. LSTKC [31] introduces a relation matrix-based erroneous knowledge filtering and rectification mechanism to distill correct knowledge for supervised lifelong person ReID. However, continuously annotating new domains can be a cumbersome task for ReID system administrators. CLUDA [32] combines meta learning and knowledge distillation to address the forgetting problem for Unsupervised lifelong person ReID. However, LSTKC [31] and CLUDA [32] have not considered the backward-compatible problem for lifelong person ReID. In addition, LSTKC and CLUDA neglect informative cluster prototypes that help to regularize the similarity relationship during adaptation.

B. Contrastive learning

The main idea of contrastive learning is to maximize the representation similarity between a positive pair composed of differently augmented views of a same image, so that a model can be invariant to view differences. While attracting a positive pair, some contrastive methods also consider other images as negatives and push away negatives stored in a memory bank [33], [34] or in a large mini-batch [35]. Contrastive methods show great performance in unsupervised representation learning, which makes it the main approach in unsupervised person ReID. Based on clustering generated pseudo-labels, state-of-the-art unsupervised person ReID methods build positive pairs with cluster centroids [21], camera-aware centroids [36] and generated positive views [37], [38]. ICE [2] combines image-to-prototype and image-to-image contrastive losses to reach the maximal agreement between positive images and cluster prototypes. However, these contrastive methods only consider single-domain view agreements, which suffer from catastrophic forgetting in multi-domain learning.

C. Incremental learning

Incremental (also called continuous or lifelong) learning aims at learning new classes, domains or tasks without forgetting previously acquired knowledge. Previous methods can be roughly categorized into three directions, i.e., architecture-based, regularization-based and rehearsal-based methods. Architecture-based methods combine task-specific parameters to build a whole network. These methods progressively extend network structure when new tasks are added into an existing model [39]–[41]. Regularization-based methods consist in regularizing model updates on new data in a way that does not contradict the old knowledge. A common approach is to freeze the old model as a teacher for previous knowledge distillation [42]–[44]. Rehearsal-based (also called recall or replay) methods address the forgetting problem by storing a small number of old image samples [45]–[48] or a generative model [49], [50]. The stored old data or generated data are used to remind the model of previous knowledge during incremental training. In addition to the above-mentioned supervised incremental methods, several attempts have been made in unsupervised lifelong adaptation, such as setting gradient regularization [13] in contrastive learning and consolidating the internal distribution [14]. CoTTA [51] proposes to use weight-averaged and augmentation-averaged pseudo-labels for test-time adaptation. Differently, our method is designed for lifelong domain adaptation, where we regularize both cluster and instance-level relationship with data augmentation and domain gap perturbations in the training. Moreover, general lifelong adaptation [13], [14], [51] has identical classes across different domains, which is not suitable for lifelong ReID that has to learn fine-grained identity representations from non-overlapping classes across domains.

D. Backward Compatible Learning

Backward-compatible learning aims at enhancing the backward consistency of feature representations, so that previously extracted representations can be comparable with newly extracted representations in retrieval. BCT [11] employs a cross-entropy distillation loss between old and new classifiers to constrain new representations. Neighborhood Consensus Contrastive Learning (NCCL) [52] proposes a neighborhood consensus supervised contrastive learning method to constrain new representations at the sub-cluster level. AdvBCT [53] uses adversarial learning to minimize the distributional distance between old and new encoders. However, the above-mentioned backward-compatible learning usually focuses on single-step learning models. In a long session learning scenario, more adaptation steps make it easier to lose knowledge of beginning steps. To handle the multi-step learning scenario, Wan et al. [12] introduce a continual learner for visual search (CVS), which effectively improves the backward compatibility in the

class-incremental task. CVS only considers class-incremental setting within a single domain, which is sub-optimal for the domain-incremental scenario. Oh et al. [54] propose a part-assisted knowledge consolidation method that leverages both local and global features to enhance the backward compatibility in lifelong person ReID. Instead of using local features, our framework leverages informative cluster prototypes and instances to enhance the backward compatibility during domain-incremental learning.

E. Difference with previous methods

In this paper, we explore the possibility of jointly addressing the forgetting, generalization and backward-compatible problems existing in regular unsupervised person ReID methods. Although our method is based on contrastive learning, incremental learning and backward-compatible learning, there are major differences between our method and previous methods. Compared with previous contrastive learning that mainly considers single-domain performance, we propose to further reach maximal image-to-prototype and image-to-image similarity consistency across different domains for anti-forgetting. Compared with previous lifelong adaptation methods, our proposed unsupervised lifelong ReID method is able to accumulate fine-grained identity information from non-overlapping classes in a domain-incremental scenario. Moreover, our method uses a dual-level relationship regularization along with contrastive learning to better mitigate the forgetting problem. Previous backward-compatible learning methods, such as CVS [12], only consider class-incremental settings within a single domain, which is sub-optimal for the domain-incremental scenario. As incrementally added person ReID domains bring in both class and domain gaps, it is more difficult to retain backward consistency in lifelong ReID. Our proposed method regularizes the representation consistency at both cluster and instance levels, which enhances the backward compatibility during domain-incremental learning.

III. Methodology

A. Overview

Given a stream of N person ReID datasets, unsupervised lifelong person ReID aims at learning a generalizable model via sequential unsupervised learning on the training set of each domain $D_1 \rightarrow ... \rightarrow D_{s-1} \rightarrow D_s ... \rightarrow D_N$. After the whole pipeline, the adapted model is tested respectively on the test set of each seen domain $D_1, ..., D_N$, as well as on multiple unseen domains.

We use θ_{new} and θ_{old} to respectively represent the adapted and old models. Incremental learning aims at retaining previous knowledge while acquiring new knowledge. Given a triplet of an anchor x_i, a positive x_p and a negative x_n, we have initially $d(f(x_i|\theta_{old}), f(x_p|\theta_{old})) < d(f(x_i|\theta_{old}), f(x_n|\theta_{old}))$ with the old model. After incremental adaptation, the distance between the anchor $f(x_i|\theta_{new})$ and the positive $f(x_p|\theta_{new})$ should remain smaller than that between the anchor and the negative $f(x_n|\theta_{new})$. An incremental learning criterion can be defined as:

$$\begin{aligned} &d(f(x_i|\theta_{new}), f(x_p|\theta_{new})) < d(f(x_i|\theta_{new}), f(x_n|\theta_{new})), \\ &\forall (i,p,n) \in \{(i,p,n) : y_i = y_p \neq y_n\}, \end{aligned} \tag{1}$$

where $d(\cdot,\cdot)$ denotes the distance between two representations.

Backward-compatible learning also aims at retaining previous knowledge while acquiring new knowledge. Different to incremental learning, backward-compatible learning focuses more on the compatibility between new query representations and old gallery representations. The distance between a new query $f(x_i|\theta_{new})$ and a stored gallery positive $f(x_p|\theta_{old})$ should be smaller than that between the query $f(x_i|\theta_{new})$ and a stored gallery negative $f(x_n|\theta_{old})$. A backward-compatible criterion can be defined as:

$$\begin{aligned} &d(f(x_i|\theta_{new}), f(x_p|\theta_{old})) < d(f(x_i|\theta_{new}), f(x_n|\theta_{old})), \\ &\forall (i,p,n) \in \{(i,p,n) : y_i = y_p \neq y_n\}, \end{aligned} \tag{2}$$

In this paper, we show that the incremental learning criterion Eq. (1) and the backward-compatible criterion Eq. (2) can be both satisfied with a unified framework. We present a Dual-level Joint Adaptation and Anti-forgetting (DJAA) method for unsupervised lifelong person ReID. The general architecture of DJAA is illustrated in Fig. 2 (a). To mitigate the forgetting, we build a hybrid memory buffer that stores a small number of informative images and corresponding cluster prototypes for rehearsing old knowledge. The proposed DJAA consists of an adaptation module and a rehearsal module, which work collaboratively to achieve the unsupervised lifelong adaptation without forgetting. For step s, we jointly train new samples on domain D_s and old samples stored in the hybrid memory buffer to adapt the encoder θ_{s-1} to become the encoder θ_s.

The Adaptation Module of DJAA is shown in Fig. 2 (b). We use the pseudo label based contrastive learning method ICE [2] as a baseline, which leverages both an online encoder and a momentum encoder. In DJAA, the momentum encoder serves as a knowledge collector that gradually accumulates knowledge of each seen domain. During the step s, the momentum encoder (weights noted as θ_m) accumulates new knowledge with exponential moving averaged weights of the online encoder (weights noted as θ):

$$\theta_m^t = \alpha \theta_m^{t-1} + (1-\alpha)\theta^t, \tag{3}$$

where the hyper-parameter α controls the speed of the knowledge accumulation. t and $t-1$ refer respectively to the current and last iterations. We extract all image representations with the stable momentum encoder and generate corresponding pseudo labels with a density-based clustering algorithm DBSCAN [55]. Based on the clustered pseudo labels, we build cluster prototypes for prototype contrastive adaptation loss $\mathcal{L}_{pa}$ and image contrastive adaptation loss $\mathcal{L}_{ia}$ (described in Section III-B) on the new domain D_s.

The Rehearsal Module of DJAA is shown in Fig. 2 (c). Before adaptation, we freeze the momentum encoder from the last step θ_{s-1} as an old knowledge expert model. We set the consistency regularization condition: the image-to-image and image-to-prototype similarity encoded by the frozen old

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2024.3490777

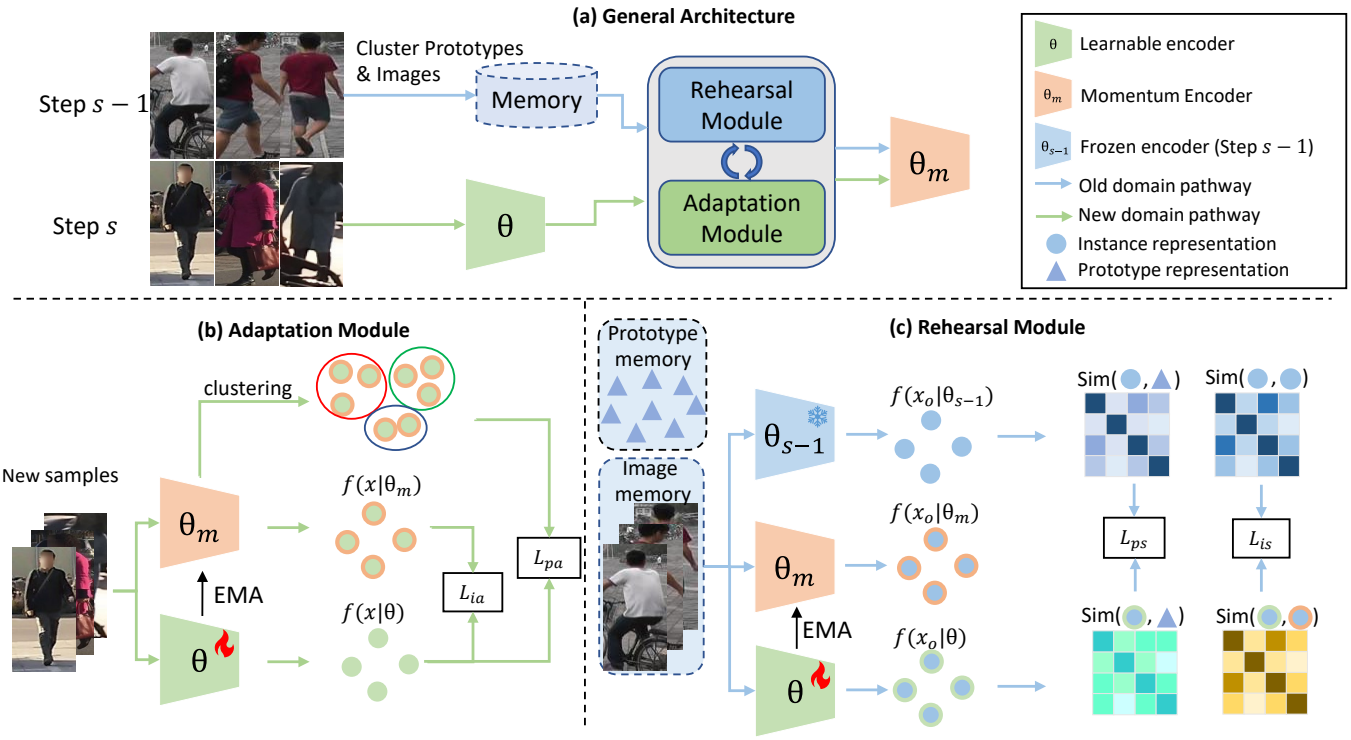


Fig. 2. **(a) General architecture of DJAA**: our proposed framework is composed of an adaptation module and a rehearsal module. A memory buffer stores a small number of images and corresponding cluster prototypes from previous step $s-1$, which are trained jointly with data from step s. The two modules work collaboratively to achieve the adaptation without forgetting. **(b) Adaptation Module**: This module follows the new domain pathway (→) to learn new domain knowledge with dual-level contrastive adaptation losses $\mathcal{L}_{pa}$ and $\mathcal{L}_{ia}$. **(c) Rehearsal Module**: This module follows the old domain rehearsal pathway(→) to rehearse old domain knowledge with dual-level similarity consistency losses $\mathcal{L}_{ps}$ and $\mathcal{L}_{is}$.

domain model and the trainable model should be consistent during the adaptation. Based on the condition, we set an image-to-prototype similarity consistency loss $\mathcal{L}_{ps}$ and an image-to-image similarity consistency loss $\mathcal{L}_{is}$ (described in Section III-C) to regularize the model updates during the lifelong adaptation.

To train our proposed framework, we combine the above-mentioned four losses into an overall unsupervised lifelong loss:

$$\mathcal{L}_{overall} = \mathcal{L}_{pa} + \lambda_{ia}\mathcal{L}_{ia} + \lambda_{ps}\mathcal{L}_{ps} + \lambda_{is}\mathcal{L}_{is}. \tag{4}$$

B. Adaptation Module

Our adaptation module contains a prototype-level contrastive adaptation loss and an instance-level contrastive adaptation loss. Combining these two adaptation losses permits adapting a model to a new target domain, while reducing intra-cluster variance for better knowledge accumulation.

1) Prototype-level Contrastive Adaptation Loss: Given an encoded query q and a set of encoded samples $K = \{k_0, k_1, k_2, ...\}$, we use k_+ to denote the positive match of the query q. We define a softmax cosine similarity function:

$$S(q, K, \tau) = \frac{\exp(q \cdot k_+/\tau)}{\sum_{i=1}^{K} \exp(q \cdot k_i/\tau)}, \tag{5}$$

where $q \cdot k$ is the cosine similarity between q and k. τ is a temperature parameter that controls the similarity scale.

Inside an unsupervised lifelong ReID pipeline, our model incrementally learns new knowledge on each domain. For step s, $D_s = \{x_1, ..., x_n\}$ where n is the number of current domain unlabeled images. We first use the momentum encoder θ_m to extract image representations and calculate re-ranking [56] based similarity between each image pair. Based on the pair-wise similarity, the DBSCAN clustering algorithm is utilized to assign pseudo labels $\{y_1, ..., y_n\}$ to unlabeled images $\{x_1, ..., x_n\}$. Given a current domain image x_i, $f(x_i|\theta)$ and $f(x_i|\theta_m)$ denote respectively the online and the momentum representations. The prototype of a cluster j is defined as the averaged momentum representations of all the samples with a same pseudo-label y_j:

$$p_j = \frac{1}{n_j} \sum_{x_i \in y_j} f(x_i|\theta_m), \tag{6}$$

where n_j is the number of images in the cluster j.

We use P to denote all the cluster prototypes in the current domain. A cluster prototype adaptation loss maximizes the similarity between a sample x_i and the positive prototype in P, which can be defined as:

$$\mathcal{L}_{pa} = -\log S(f(x_i|\theta), P, \tau_{pa}) \tag{7}$$

where τ_{pa} is a temperature hyper-parameter.

2) Instance-level Contrastive Adaptation Loss: The prototype-level adaptation loss $\mathcal{L}_{pa}$ makes samples from the same cluster converge to a common prototype and push them away from other clusters. However, images belonging

TABLE II
COMPARISON OF POSITIVE VIEW SELECTION METHODS ON MARKET.

Method	Market	
	mAP	R1
Top-1 hardest positive (ours)	**82.3**	**93.8**
Average of top-2 hard positives	**82.3**	93.4
Average of top-3 hard positives	81.8	93.1

to the same class can be easily affected by noisy factors, such as illumination and view-point, leading to high intra-class variance. When the adaptation module is only trained with $\mathcal{L}_{pa}$, all the image samples converge at a same pace. Consequently, representations of hard positive samples remain far away after optimization. We use an identity-aware sampler to construct mini-batches. A mini-batch X is composed of n_p identities, where each identity has n_k positive instances. Thus, the batch-size is $n_{bs} = n_p \times n_k$. Given an anchor instance $f(x_i|\theta)$, we sample the hardest positive momentum representation $f(x_j|\theta_m)$ that has the lowest cosine similarity with $f(x_i|\theta)$. To improve the intra-class compactness, we formulate an instance contrastive adaptation loss:

$$\mathcal{L}_{ia} = -\log S(f(x_i|\theta), f(X|\theta_m), \tau_{ia}) \tag{8}$$

where $f(X|\theta_m)$ denotes the mini-batch containing one hard positive $f(x_j|\theta_m)$ and $(n_p - 1) \times n_k$ negatives, and τ_{ia} is a temperature hyper-parameter.

Remark. In our instance-level contrastive adaptation loss, we select the hardest positive to maximally increase the similarity between an anchor and the hard positive. Another possibility is to use the average of top-k hard positives as the contrastive positive view. We compare the positive selection methods in Table II. The results show that the hardest positive slightly outperforms the average of top-k hard positives. We thus choose to select the hardest positive in our method.

By reducing the intra-cluster variance between hard positive samples, the instance-level adaptation loss $\mathcal{L}_{ia}$ allows for purifying the new knowledge on D_i before being accumulated into the model θ_m. Person ReID is a cross-camera image retrieval task, where the individual camera style is the main factor that causes intra-cluster variance. As shown in ICE [2], using camera labels to reduce camera style variance can further mitigate the noise during the knowledge accumulation. A camera proxy p_{jk} is defined as the averaged momentum representations of all the instances belonging to the cluster j captured by camera c_k:

$$p_{jk} = \frac{1}{n_{jk}} \sum_{x_i \in y_j \cap x_i \in c_k} f(x_i|\theta_m), \tag{9}$$

where n_{jk} is the number of instances belonging to the cluster j captured by camera c_k. We form a set of camera prototypes P_{cam}, which combines one positive cross-camera prototype and n_{neg} nearest negative prototypes. A camera contrastive loss is the softmax log loss that minimizes the distance between an anchor $f(x_i|\theta)$ and each of positive cross-camera prototypes:

$$\mathcal{L}_{cam} = -\frac{1}{|\mathcal{P}|} \sum_{x_i \in y_j \cap x_i \notin c_k} \log S(f(x_i|\theta), P_{cam}, \tau_c), \tag{10}$$

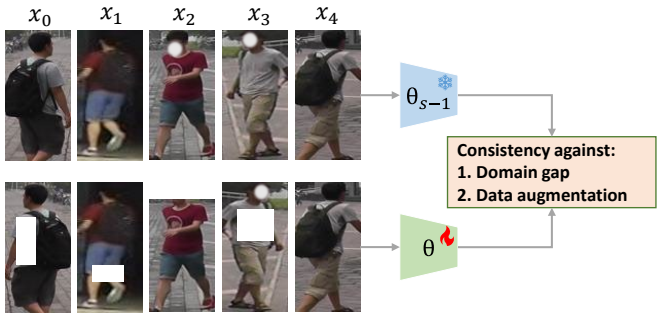


Fig. 3. We introduce two kinds of perturbations, including domain gap from different encoders and data augmentation. Our target is to make the model invariant to domain gap and data augmentation perturbations.

where τ_c is a cross-camera temperature hyper-parameter, which is set to 0.07 following [2]. $|\mathcal{P}|$ is the number of cross-camera positive prototypes. We provide extra results with the camera contrastive loss for knowledge accumulation in Section IV.

C. Rehearsal Module

In our rehearsal module, buffered samples and prototypes are utilized to prevent catastrophic forgetting during the domain adaptation. This module contains an Image-to-prototype Similarity Consistency loss and Image-to-image Similarity Consistency loss to maximally retain the old knowledge. ICE [2] introduces a consistency regularization method to enhance a model's invariance to data augmentation perturbations. In this paper, we further consider the encoder style as a kind of perturbation. As shown in Figure 3, we enhance the similarity consistency between differently augmented views encoded by current domain encoder θ and last step encoder θ_{s-1}. In this way, our model can be robust to both domain style and data augmentation changes.

1) Image-to-prototype Similarity Consistency loss: Technically, person ReID is a representation similarity ranking problem, in which the objective is to have high similarity scores between positive pairs and low similarity scores between negative pairs. However, when a model is adapted into a new domain, the similarity relationship between old domain samples could be affected by the new domain knowledge. As a cluster prototype (the averaged representation of all cluster samples) contains generic information of a cluster, the prototype memory enables the current model to have access to generic old domain cluster information without storing all the images. For the memory buffer of n_{mem} buffered images and prototypes, we first use the buffered cluster prototypes $P^o = \{p_1^o, ..., p_{n_{mem}}^o\}$ as anchors to rehearse the old domain knowledge. As the similarity relationship between images and prototype should be consistent before and after a domain adaptation step, we propose an image-to-prototype similarity loss to ensure the upcoming new knowledge would not change the image-to-prototype similarity.

As the frozen old model θ_{s-1} from the last domain can be regarded as an expert on the old domain, we calculate the cosine similarity between representations encoded by

θ_{s-1} and prototypes as a reference to regularize the online model θ. Given a mini-batch of old images $\{x_1^o, ..., x_{n_{bs}}^o\}$ randomly sampled from the memory buffer where n_{bs} is the batch size, the online image-to-prototype similarity between a buffered image x_i^o and buffered prototypes P^o is defined as $S(f(x_i^o|\theta), P^o, \tau_{ps})$. For the same mini-batch, we calculate the image-to-prototype similarity with the frozen model θ_{s-1} as the reference similarity. The reference similarity between the same image-prototype pair is defined as $S(f(x_i^o|\theta_{s-1}), P^o, \tau_{ps})$.

We formulate an image-to-prototype similarity consistency loss with a Kullback-Leibler (KL) Divergence between the two similarity distributions:

$$\mathcal{L}_{ps} = \mathcal{D}_{KL}(S(f(x_i^o|\theta), P^o, \tau_{ps})|| \\ S(f(x_i^o|\theta_{s-1}), P^o, \tau_{ps})). \tag{11}$$

The buffered prototypes can be regarded as anchors of previously-acquired knowledge. We expect that the new domain knowledge does not change the distance relationship between an image and the prototypes, so that we can retain the previously-acquired knowledge for anti-forgetting adaptation. KL divergence measures the information loss when one distribution is used to approximate another. By minimizing $\mathcal{L}_{ps}$, we encourage the image-to-prototype similarity distribution $S(f(x_i^o|\theta), P^o, \tau_{ps})$ calculated with current domain knowledge to be consistent with that calculated with old domain distribution $S(f(x_i^o|\theta_{s-1}), P^o, \tau_{ps})$.

2) Image-to-image Similarity Consistency loss: Cluster prototypes contain general cluster information, for example, the most salient feature shared by different views of a person. Differently, image instances contain more detailed information in a specific view, which is complementary to prototype-level general information. As the similarity relationship between the same images should be consistent before and after a domain adaptation step, we propose an image-to-image similarity loss that regularizes the similarity relationship updates in a way that does not contradict the old knowledge. Similar to the image-to-prototype similarity, we also use the frozen old model θ_{s-1} from the last domain as an expert to calculate the reference similarity. Given a mini-batch of old images $X^o = \{x_1^o, ..., x_{n_{bs}}^o\}$ where n_{bs} is the batch size, the image-to-image similarity distribution can be calculated with a softmax function on the cosine similarity between each image pair in the mini-batch. The image-to-image similarity between an old image x_i^o and a mini-batch X^o is calculated with both the online encoder θ and the momentum encoder θ_m, i.e., $S(f(x_i^o|\theta), f(X^o|\theta_m), \tau_{is})$. For the same mini-batch, we calculate the image-to-image similarity distribution with the frozen old model θ_{s-1} as a reference for the constraint. The reference similarity between the same image x_i^o and the mini-batch X^o is $S(f(x_i^o|\theta_{s-1}), f(X^o|\theta_{s-1})), \tau_{is})$.

We formulate an image-to-image similarity constraint loss with the KL Divergence between the two distributions:

$$\mathcal{L}_{is} = \mathcal{D}_{KL}(S(f(x_i^o|\theta), f(X^o|\theta_m), \tau_{is})|| \\ S(f(x_i^o|\theta_{s-1}), f(X^o|\theta_{s-1})), \tau_{is})). \tag{12}$$

By minimizing $\mathcal{L}_{is}$, we encourage the similarity relationship calculated with current domain knowledge θ to be consistent with that calculated with old domain knowledge θ_{s-1}.

Remark. As shown in Fig.3, we introduce both domain gap and data augmentation perturbations. The domain gap perturbations are introduced by using online and frozen encoders. The data augmentation perturbations are introduced by using two data augmentation settings on same images. Inspired by consistency regularization from semi-supervised learning [57], we use weak data augmentation on reference similarity calculation and strong data augmentation on prediction similarity calculation. As data augmentation can mimic image perturbations, augmentation consistency regularization allows for enhancing the model robustness against perturbations in real deployments.

D. Memory Buffer Update

We store a small number of informative samples and cluster prototype representations in a hybrid memory buffer, which is updated at the end of each step. The total size of the hybrid memory buffer is set to n_{mem} images and n_{mem} prototypes. In our proposed DJAA framework, the default value of n_{mem} is 512. Suppose that $|P|$ is the number of current domain clusters and $|P^o|$ is the number of buffered cluster prototypes, we update the memory buffer with n_{new} samples from the new domain and n_{old} samples from the memory buffer:

$$n_{new} = \frac{|P|}{|P^o| + |P|} \times n_{mem}, \tag{13}$$

$$n_{old} = \frac{|P^o|}{|P^o| + |P|} \times n_{mem}. \tag{14}$$

Previous methods [29], [32] usually store several samples from randomly selected classes, which is sub-optimal for our proposed image-to-prototype similarity consistency loss. We propose a simple yet effective clustering-guided selection algorithm to select informative samples. As the cluster prototype is defined as the averaged representation, the prototype can cover more sample information if a cluster contains more samples. We thus use the number of samples to rank each cluster and preferentially select clusters that contain more samples. Once a cluster is selected, we proceed to calculate the cosine similarity between the cluster prototype and cluster samples. Under unsupervised setting, we argue that the sample with the highest similarity score is the most credible sample belonging to the cluster. To mitigate the pseudo-label noise and enhance the data diversity, we select 1 sample that is closest to the prototype in each selected cluster. We provide training and memory buffer updating details in Algorithm 1. We validate the effectiveness of our proposed memory buffer updating method in Section IV-G2.

IV. Experiments

A. Datasets and Evaluation Protocols

We use 4 seen datasets for domain-incremental training and 10 unseen datasets for generalization ability evaluation, as shown in Table III.

Algorithm 1 DJAA for unsupervised lifelong person ReID

Input: Unlabeled domains $D_1 \rightarrow ... \rightarrow D_s \rightarrow ... \rightarrow D_N$, a hybrid memory buffer of size n_{mem}, an online encoder θ and a momentum encoder θ_m.
Output: Encoder θ_m after N-step adaptation.

for $D_s = D_1$ to D_N **do**
  Get θ_m from the last step D_{s-1}. Freeze a copy of momentum encoder $\theta_{s-1} \leftarrow \theta_m$. Initialize the online encoder with the momentum encoder $\theta \leftarrow \theta_m$;
  for $epoch = 1$ to E_{max} **do**
    Generate pseudo labels on D_s;
    Calculate cluster prototypes P in Eq. (6) on D_s;
    for $iter = 1$ to I_{max} **do**
      Sample a mini-batch from the current domain D_s for $\mathcal{L}_{pa}$ in Eq. (7) and $\mathcal{L}_{ia}$ in Eq. (8);
      Sample a mini-batch from the memory buffer for $\mathcal{L}_{ps}$ in Eq. (11) and $\mathcal{L}_{is}$ in Eq. (12);
      Train θ with $\mathcal{L}_{overall}$ in Eq. (4);
      Update θ_m with Eq. (3);
    end for
  end for
  Update the memory buffer with n_{new} new samples and n_{old} previous samples, following Eq. (13) and Eq. (14);
  Store θ_m for next step D_{s+1};
end for

TABLE III
DATASET STATISTICS. UNSEEN DOMAINS ARE ONLY USED FOR TESTING.

Type	Dataset	#train img	#train id	#test img	#test id
Seen	PersonX [58]	9840	410	35952	856
	Market [59]	12936	751	19281	750
	Cuhk-Sysu [60]	15088	5532	8347	2900
	MSMT17 [61]	32621	1041	93820	3060
Unseen	VIPeR [62]	-	-	632	316
	PRID [63]	-	-	749	649
	GRID [64]	-	-	1025	125
	iLIDS [65]	-	-	120	60
	CUHK01 [66]	-	-	1944	486
	CUHK02 [67]	-	-	956	239
	SenseReID [68]	-	-	4428	1718
	CUHK03 [69]	-	-	1930	100
	3DPeS [70]	-	-	512	96
	MMP-Retrieval [32]	-	-	28907	7

Incremental training on seen datasets: We set 2 incremental training pipelines on seen datasets, i.e., one-cycle full set benchmark and two-cycle subset benchmark. The one-cycle full-set benchmark targets at evaluating the effectiveness of handling imbalanced class numbers between different adaptation steps, while the two-cycle subset benchmark aims to mimic the season and weather cycle, which may re-appear after several adaptation steps. In the one-cycle full set benchmark, we do not use any supervised pre-training. The model is directly adapted to Market1501, CUHK-SYSU, and MSMT17. We further define two training orders, i.e., Market→Cuhk-Sysu→MSMT17 and MSMT17→Market→Cuhk-Sysu. In the two-cycle subset benchmark, following CLUDA [32], we first pre-train our model on PersonX in a supervised manner. Then, we adapt our pre-trained model to Market1501, CUHK-SYSU, and MSMT17 in an unsupervised manner. The adaptation steps are repeated twice, with each stage involving a subset of 350 identities. The two-cycle training order is defined as PersonX→Market →Cuhk-Sysu→MSMT17→Market→Cuhk-Sysu→MSMT17. Cumulative Matching Characteristics (CMC) at Rank1 accuracy and mean Average Precision (mAP) are used in our experiments. We also report the averaged $\bar{s}Rank1$ and $\bar{s}mAP$ on the seen domains and unseen domains.

Generalization ability on unseen datasets: We use 10 person ReID datasets to maximally evaluate the model generalization ability on different unseen domains, including VIPeR, PRID, GRID, iLIDS, CUHK01, CUHK02, SenseReID, CUHK03, 3DPeS and MMP-Retrieval. These 10 datasets cover all the unseen domains that are considered in previous supervised lifelong ReID methods [7], [8] and domain generalizable ReID methods [10]. We use the traditional training/test split on CUHK03 dataset. Rank1 accuracy and mAP results are respectively reported on the test set of each unseen domain after the final step.

Backward-compatible evaluation: To evaluate the backward compatibility of feature representations, we report retrieval performance between current query images and current gallery images (termed as self-test), and that between current query images and stored gallery images (termed as cross-test).

B. Implementation details

1) Training: Our method is implemented under Pytorch [71] framework. The total training time with 4 Nvidia 1080Ti GPUs is around 6 hours. We use an ImageNet [72] pre-trained ResNet50 [73] as our backbone network. For the strong data augmentation, we resize all images to 256×128 and augment images with random horizontal flipping, cropping, Gaussian blurring and erasing [74]. For the weak data augmentation, we only resize images to 256×128.

At each step in the one-cycle full set training, we train our framework 30 epochs with 400 iterations per epoch using a Adam [75] optimizer with a weight decay rate of 0.0005. For the two-cycle subset evaluation, we follow CLUDA [32] to train our model for 60 epochs without specified iterations. The learning rate is set to 0.00035 with a warm-up scheme in the first 10 epochs. No learning rate decay is used in the training. Pseudo labels on the current domain are updated on re-ranked Jaccard distance [56] at the beginning of each epoch with a DBSCAN [55], in which the minimum cluster sample number is set to 4 and the distance threshold is set to 0.55. The momentum encoder is updated with a momentum hyper-parameter $\alpha = 0.999$. To balance the model ability on old domains and the new domain, we separately take a mini-batch of current domain images and a mini-batch of buffered images of the same batch size n_{bs}, which is set to $n_{bs} = 32$ in our experiments. Furthermore, we use a random identity sampler to construct mini-batches to handle the imbalanced images of different identities. Following the clustering setting on the current domain, the 32 current domain images are composed of 8 identities and 4 images per identity. Inside the mini-batch of buffered images, the 32 buffered images are composed of 32 identities and 1 image per identity.

TABLE IV
SEEN-DOMAIN RESULTS (%) OF UNSUPERVISED DOMAIN ADAPTATION (U) AND UNSUPERVISED LIFELONG (UL) METHODS ON ONE-CYCLE FULL SET BENCHMARK. THE TRAINING ORDER IS MARKET→CUHK-SYSU→MSMT17. * REFERS TO USING THE CAMERA LOSS EQ.(10) TO REDUCE INTRA-CLUSTER VARIANCE. THE BEST PERFORMANCE IS MARKED IN BOLD.

Method	Memory size	Type	Market		Cuhk-Sysu		MSMT17		Average	
			mAP	Rank1	mAP	Rank1	mAP	Rank1	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
ICE [2]	0	U	29.0	60.4	72.5	76.3	21.8	49.0	41.1	61.9
CC [76]	0	U	31.0	58.9	74.6	77.3	25.7	51.8	43.8	62.7
PPRL [22]	0	U	29.5	58.6	75.6	79.5	**32.9**	**63.2**	46.0	67.1
LwF [42]	0	UL	27.5	59.0	70.5	74.7	20.3	48.6	39.5	60.8
iCaRL [45]	512	UL	37.4	67.6	79.5	81.9	19.9	45.4	45.5	65.0
Co^{2}L [47]	512	UL	35.3	62.0	78.3	80.8	24.2	50.7	46.0	64.5
CVS [12]	512	UL	56.8	78.7	74.6	77.4	15.7	38.6	49.0	64.9
LSTKC [31]	512	UL	48.8	74.9	77.2	79.7	20.6	47.6	48.8	67.4
DJAA (ours)	512	UL	60.2	82.5	**83.9**	**85.6**	20.7	46.7	54.9	71.6
DJAA* (ours)	512	UL	**65.2**	**86.3**	81.8	84.1	23.7	51.6	**56.9**	**74.0**

TABLE V
UNSEEN-DOMAIN RESULTS (%) OF UNSUPERVISED (U), UNSUPERVISED LIFELONG (UL) AND DOMAIN GENERALIZATION (DG) METHODS ON ONE-CYCLE FULL SET BENCHMARK. THE TRAINING ORDER IS MARKET→CUHK-SYSU→MSMT17. * REFERS TO USING THE CAMERA LOSS EQ.(10) TO REDUCE INTRA-CLUSTER VARIANCE. THE BEST UNSUPERVISED PERFORMANCE IS MARKED IN BOLD.

Method	Memory Size	Type	VIPeR		PRID		GRID		iLIDS		CUHK01		CUHK02		SenseReID		CUHK03		3DPeS		Average	
			mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	$\bar{s}_{mAP}$	$\bar{s}_{R1}$
ICE [2]	0	U	35.7	25.9	39.0	29.0	20.6	14.4	71.4	61.7	60.6	60.0	48.2	45.8	33.6	27.9	17.3	29.7	48.5	55.4	41.7	38.9
CC [76]	0	U	43.3	32.6	41.6	28.0	21.1	15.2	79.5	73.3	66.3	66.0	57.2	57.3	37.9	30.6	25.6	24.0	54.4	59.6	47.4	43.0
PPRL [22]	0	U	40.5	30.1	43.8	33.0	15.0	9.6	74.3	63.3	66.3	65.9	54.5	53.3	38.4	31.2	22.2	20.3	50.0	62.4	45.0	41.0
LwF [42]	0	UL	41.0	29.7	40.1	30.0	19.7	13.6	74.9	66.7	62.4	61.9	52.8	51.9	33.9	27.3	20.0	34.1	52.2	58.6	44.1	41.5
iCaRL [45]	512	UL	48.1	38.0	44.5	34.0	29.4	20.0	82.1	76.7	66.9	66.0	58.3	55.2	42.1	35.5	26.2	41.3	57.1	63.6	50.5	47.8
Co^{2}L [47]	512	UL	43.8	32.0	50.7	41.0	30.4	21.6	**83.6**	**76.7**	69.5	69.3	58.1	54.2	39.6	31.6	25.3	41.6	55.4	62.7	50.7	47.8
CVS [12]	512	UL	44.3	33.2	31.6	23.0	31.0	23.2	78.5	71.7	62.8	60.8	58.2	55.2	42.6	34.7	32.3	41.0	54.5	60.9	48.4	44.9
LSTKC [31]	512	UL	43.6	30.7	45.0	35.0	31.7	24.0	74.4	68.3	66.5	66.2	61.8	59.8	43.4	36.4	29.3	40.3	57.8	66.4	50.4	47.5
DJAA (ours)	512	UL	49.0	38.3	**56.0**	**44.0**	46.0	36.0	82.9	**76.7**	69.3	**69.3**	**67.3**	**65.9**	47.5	39.5	31.4	**48.8**	64.5	**70.5**	57.1	**54.3**
DJAA* (ours)	512	UL	**50.9**	**39.6**	53.4	**44.0**	**47.2**	**37.6**	79.8	71.7	**70.3**	**69.7**	66.9	63.8	**50.1**	**42.2**	**36.1**	46.7	**64.8**	70.5	**57.7**	54.0
ACL [25]	All	DG	69.2	60.8	60.0	50.0	58.8	48.8	89.9	86.7	78.4	77.8	77.1	76.4	62.4	53.5	65.1	67.3	71.1	78.8	70.2	66.7

We use grid search to set the optimal temperature and balancing hyper-parameters in our proposed losses. Based on grid search results, we set the temperature hyper-parameters $\tau_{pa} = 0.5$, $\tau_{ia} = 0.1$, $\tau_{ps} = 0.1$ and $\tau_{is} = 0.2$. To make rehearsal and adaptation losses on the same scale, we set the balancing hyper-parameters $\lambda_{ia} = 1$, $\lambda_{ps} = 10$ and $\lambda_{is} = 20$. For the memory buffer, we set $n_{mem} = 512$, where the total identity number equals 512 and 1 image per cluster. After the whole training, only the momentum encoder is saved for inference.

2) Compared methods: We re-implement 3 types of unsupervised methods, including unsupervised domain adaptation, unsupervised lifelong and supervised domain generalization methods to compare with our method.

The unsupervised domain adaptive methods include ICE [2], CC [76] and PPRL [22], which are trained sequentially on each seen domain. The lifelong methods include three general-purpose lifelong methods (LwF [42], iCaRL [45] and Co^{2}L [47]), one backward-compatible class-incremental method CVS [12] and one lifelong ReID method LSTKC [31]. LwF is a regularization-based method, which does not store old samples for rehearsal. LwF uses a prediction-level cross-entropy distillation [77] between old and new domain models. iCaRL and Co^{2}L are rehearsal-based methods. iCaRL conducts prediction-level distillation on new and stored old images for rehearsal. Co^{2}L proposes an asymmetric supervised contrastive loss and a relation distillation for supervised continual learning. CVS formulates an inter-session data coherence loss and a neighbor-session triplet loss to enhance model coherence during the incremental learning. For general-purpose methods LwF, iCaRL, Co^{2}L and CVS, we **combine our unsupervised adaptation module ($\mathcal{L}_{pa} + \mathcal{L}_{ia}$) and the lifelong learning techniques of each paper** to convert these methods to person ReID and conduct a fair comparison with our method. For example, in LwF, we combine our contrastive baseline for learning current domain knowledge and the prediction-level distillation for mitigating the forgetting. To convert the supervised lifelong method LSTKC [31] into an unsupervised lifelong method, we combine our unsupervised adaptation module ($\mathcal{L}_{pa} + \mathcal{L}_{ia}$) and the rectification-based knowledge distillation. We also re-implement a supervised domain generalization method ACL [25] to show the upper bound of the generalization ability. The compared lifelong methods, such as LwF [42], iCaRL [45], Co^{2}L [47], CVS [12] and LSTKC [31], follow the same data augmentation. For ICE [2], CC [76], PPRL [22] and ACL [25], we directly use their original augmentation setting, including random cropping, flipping and erasing.

C. Seen-domain adaptation performance evaluation

For the one-cycle full set benchmark, we report seen-domain results after the final step in Table IV. Designed for maximally learning domain-specific features inside a single domain, regular unsupervised adaptation methods ICE, ClusterContrast (CC) and PPRL cannot learn domain-agnostic generalized features for lifelong ReID. We convert several incremental learning methods, such as LwF, iCaRL, Co^{2}L and CVS, into unsupervised lifelong ReID methods. Among these lifelong methods, the rehearsal-based methods iCaRL and Co^{2}L yield better averaged performance than the pure regularization-based

TABLE VI
ADAPTATION PERFORMANCE (%) EVALUATION ON THE TWO-CYCLE SUBSET BENCHMARK. THE TRAINING ORDER IS PERSONX→MARKET→CUHK-SYSU→MSMT17→MARKET→CUHK-SYSU→MSMT17.

Method	Market(t=1)		Cuhk-Sysu(t=2)		MSMT17(t=3)		Market(t=4)		Cuhk-Sysu(t=5)		MSMT17(t=6)	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
CLUDA [32]	52.96	75.21	71.35	75.38	11.04	28.67	64.87	82.79	78.39	81.86	14.64	33.91
DJAA (ours)	**56.5**	**79.4**	**81.1**	**83.4**	**16.3**	**38.5**	**65.3**	**84.7**	**82.9**	**84.8**	**16.8**	**39.5**

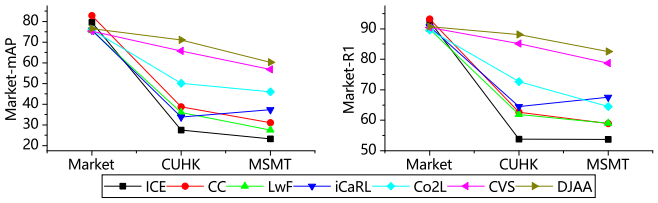


Fig. 4. Non-forgetting evaluation with mAP and Rank1 on the first seen domain Market-1501. The training order is Market→Cuhk-Sysu→MSMT17.

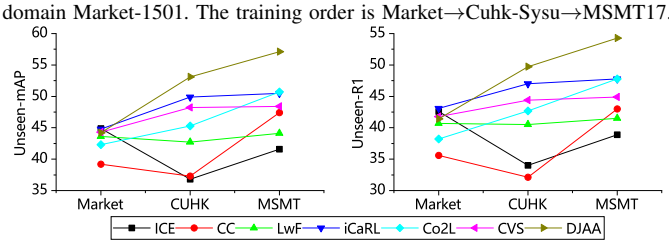


Fig. 5. Generalization ability evaluation with averaged mAP and averaged Rank1 on all the unseen domains. The training order is Market→Cuhk-Sysu→MSMT17.

method LwF. We also report the performance of state-of-the-art unsupervised lifelong ReID method CLUDA. Under the unsupervised lifelong setting, our proposed DJAA outperforms the state-of-the-art method CVS by 5.9% on averaged mAP and 6.7% on averaged Rank1. We further add camera-aware contrastive loss to reduce the camera style variance. With less camera noise being accumulated into our model, DJAA* achieves higher averaged performance on seen domains. DJAA* does not performs well on CUHK-SYSU, because no camera labels are available on this dataset. We further draw mAP/Rank1 variation curves on the first seen domain Market1501 in Fig. 4, which confirms that our proposed DJAA has a better anti-forgetting capacity. After 3 adaptation steps, our proposed DJAA forgets the least knowledge among all the compared methods.

For the two-cycle subset benchmark, we compare the adaptation performance after each step in Table VI and the anti-forgetting performance in Table VII. As shown in Table VI, our method DJAA shows superior adaptation performance on the two domain adaptation cycles. These results demonstrate the effectiveness of our proposed adaptation module in adapting a model to new environments. As shown in Table VII, our method DJAA shows slightly inferior anti-forgetting performance on PersonX, while significantly outperforming CLUDA on Market1501 and CUHK-SYSU. The anti-forgetting performance of DJAA surpasses that of CLUDA on the averaged results over PersonX, Market and Cuhk-Sysu. These results demonstrate the effectiveness of our proposed rehearsal module in alleviating catastrophic forgetting.

D. Unseen-domain generalization ability evaluation

For the one-cycle full set benchmark, we report unseen-domain generalization ability results in Table V. Similar to seen-domain results, ICE, ClusterContrast (CC) and PPRL can hardly learn domain-agnostic generalized features, which leads to low performance on unseen domains. On the contrary, lifelong methods accumulate knowledge from each adapted domain and eventually learn domain-agnostic generalized features. With the same baseline, the rehearsal-based methods iCaRL, Co^{2}L and CVS outperform the pure regularization-based method LwF. Under the unsupervised lifelong setting, our proposed DJAA outperforms the second best method Co^{2}L by 6.4% on averaged mAP and 6.5% on averaged Rank1. We also add camera-aware contrastive loss to reduce the camera style variance. With camera information, the performance of DJAA* is on par with DJAA on unseen domains, showing that camera information does not bring in any further improvement in generalization ability. We further re-implement a domain generalization method ACL, to show the generalization ability of the supervised multi-domain generalization method on unseen domains. ACL is jointly trained on three datasets with human-annotated labels. With more expensive training setup than the unsupervised lifelong learning, ACL shows a strong domain generalization ability. The generalization ability of DG methods is related to the data distribution. If the training datasets and unseen test datasets have significantly different styles, DG methods could fail to generalize to unseen test datasets. Compared to DG, UL methods have better flexibility in data preparation and label annotation.

For the two-cycle subset benchmark, we compare the generalization performance after the final adaptation step in Table VIII. Following CLUDA [32], we uniformly downsample the original video sequences of MMPTRACK [78] with a ratio of 128, and divide each downsampled sequence into two halves. On MMP-Retrieval dataset, we report the Rank-1 and mAP scores averaged over all the five scenarios as the final results. Our proposed method DJAA significantly outperforms CLUDA in the Rank-1 score. The results validate the strong domain generalization ability of our method in the repetitive domain adaptation scenario.

E. Backward-compatible ability evaluation

To validate the effectiveness of our proposed method, we compare backward-compatible ability between state-of-the-art methods and our proposed method DJAA in the domain-incremental scenario. As shown in Table IX, each adaptation step helps to acquire new domain knowledge, while losing previous domain knowledge. ICE is designed for traditional one-step unsupervised domain adaptation, which has the most

TABLE VII
ANTI-FORGETTING PERFORMANCE (%) EVALUATION ON THE TWO-CYCLE SUBSET BENCHMARK. THE TRAINING ORDER IS PERSONX→MARKET→CUHK-SYSU→MSMT17→MARKET→CUHK-SYSU→MSMT17.

Method	t=3						t=6					
	PersonX		Market		Cuhk-Sysu		PersonX		Market		Cuhk-Sysu	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
CLUDA [32]	**69.48**	**80.66**	45.60	69.97	69.44	73.24	**58.23**	**76.47**	49.12	76.25	75.43	78.28
DJAA (ours)	64.4	80.6	**48.2**	**73.6**	**81.4**	**84.1**	56.0	75.6	**55.6**	**78.2**	**80.4**	**82.7**

TABLE VIII
GENERALIZATION ABILITY PERFORMANCE (%) EVALUATION ON MMP-RETRIEVAL DATASET. THE TRAINING ORDER IS PERSONX→MARKET→CUHK-SYSU→MSMT17→MARKET→CUHK-SYSU→MSMT17.

Method	MMP-Retrieval	
	mAP	R1
CLUDA [32]	41.0	67.8
DJAA (ours)	**42.7**	**78.1**

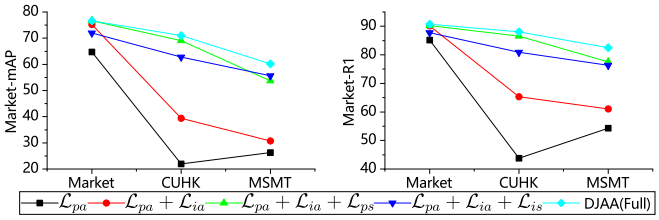


Fig. 6. Ablation study on non-forgetting evaluation with mAP and Rank1 on the first seen domain Market-1501.

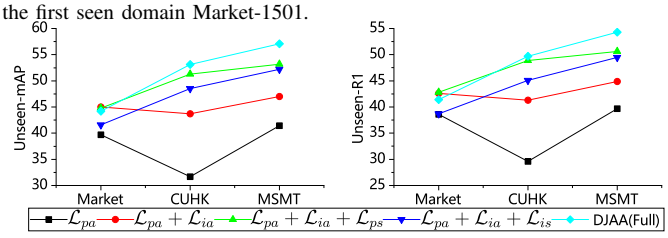


Fig. 7. Ablation study on generalization ability evaluation with averaged mAP and averaged Rank1 on all the unseen domains.

evident performance drop in the cross-test. Co^{2}L is a class-incremental learning method, which shows better backward-compatible ability than ICE. CVS aims at simultaneously addressing class-incremental learning and backward-compatible learning. Our proposed DJAA uses both image-to-prototype and image-to-image similarity to maximally retain old domain knowledge, which shows optimal backward-compatible ability. The stored gallery features are extracted with the first-domain expert model, which has more discrimination ability than the updated model on the first domain. With compatible gallery features extracted with our DJAA, the cross-test even outperforms the self-test for lifelong person ReID.

F. Ablation study

1) Adaptation and rehearsal losses: To tackle the forgetting problem during the unsupervised domain adaptation, we propose a dual-level joint adaptation and anti-forgetting method. The performance improvement of DJAA over the baseline mainly comes from the combination of the adaptation losses and the rehearsal losses '$\mathcal{L}_{pa}$', '$\mathcal{L}_{ia}$', '$\mathcal{L}_{ps}$' and '$\mathcal{L}_{is}$'. To validate the effectiveness of each loss, we conduct ablation experiments by gradually adding one of them onto the baseline. In Table X, when we only use a prototype adaptation loss '$\mathcal{L}_{pa}$', our model tends to lose most of the previous knowledge. The instance-level adaptation loss '$\mathcal{L}_{ia}$' reduces intra-cluster variance, which prevents the noise accumulation in the multi-step adaptation. The prototype-level similarity consistency loss '$\mathcal{L}_{ps}$' and the instance-level similarity consistency loss '$\mathcal{L}_{is}$' respectively make the image-to-prototype and image-to-image similarity relationships consistent to domain knowledge changes. A cluster prototype contains general information of a cluster, while an instance contains specific information from a single view. As the two consistency losses work on different levels, '$\mathcal{L}_{ps}$' and '$\mathcal{L}_{is}$' are complementary to each other. By combining all the above-mentioned losses, our full DJAA framework yields the highest performance on both seen and unseen domains. We draw the forgetting curve in Fig. 6 and the generalization ability curve in Fig. 7, which further validate the effectiveness of our proposed losses.

The backward compatible learning aims at maintaining the consistency of representations after training more data, so that previously extracted representations can be directly compared with newly extracted representations. The backward compatibility is strongly correlated with the anti-forgetting ability of a method. We perform an ablation study to validate the effectiveness of our Rehearsal Module in addressing the backward compatibility problem. The difference between the self-test and the cross-test reflects the backward compatibility. As shown in Table XI, the cross-test performance obviously degrades when we only use the adaptation losses $\mathcal{L}_{pa}$ and $\mathcal{L}_{ia}$. The improvement in the backward compatibility comes mainly from our rehearsal module ($\mathcal{L}_{ps}$ and $\mathcal{L}_{is}$). In particular, the image-to-image similarity consistency loss $\mathcal{L}_{is}$ provides the most significant performance boost.

2) Data augmentation consistency: In our method, the data augmentation includes random horizontal flipping, cropping, Gaussian blurring and erasing. These augmentation techniques are chosen to mimic real-world perturbations, such as viewpoint variance, imperfect detection and occlusion. We use these data augmentation techniques to generate augmented views for contrastive learning. By maximizing the similarity between positive views, our model can learn robust representations. As shown in Table XII, all the chosen augmentation techniques contribute to the adaptation performance. Among these augmentation techniques, the random erasing provides the most significant performance improvement, while the random blurring provides the least. Suitable data augmentation should be diverse but realistic, which means augmentation should not bring in distortions absent in the real dataset. The

TABLE IX
BACKWARD-COMPATIBLE PERFORMANCE ON MARKET1501. FULL SET TRAINING ORDER #1: MARKET→CUHK-SYSU→MSMT17. $\mathcal{Q}_1$,$\mathcal{Q}_2$ AND $\mathcal{Q}_3$ RESPECTIVELY DENOTE MARKET1501 QUERY FEATURES EXTRACTED AFTER STEP 1, 2 AND 3. $\mathcal{G}_1$,$\mathcal{G}_2$ AND $\mathcal{G}_3$ RESPECTIVELY DENOTE MARKET1501 GALLERY FEATURES EXTRACTED AFTER STEP 1, 2 AND 3. THE COLORED NUMBER IS THE DIFFERENCE BETWEEN SELF-TEST AND CROSS-TEST.

Method	Step 1		Step 2				Step 3			
	Self-Test ($\mathcal{Q}_1,\mathcal{G}_1$)		Self-Test ($\mathcal{Q}_2,\mathcal{G}_2$)		Cross-Test ($\mathcal{Q}_2,\mathcal{G}_1$)		Self-Test ($\mathcal{Q}_3,\mathcal{G}_3$)		Cross-Test ($\mathcal{Q}_3,\mathcal{G}_1$)	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
ICE [2]	80.1	92.8	41.4	67.1	30.3(↓ 11.1)	49.3(↓ 17.8)	33.4	65.0	3.4(↓ 30.0)	6.3(↓ 58.7)
Co^{2}L [47]	73.8	88.1	47.0	70.2	41.2(↓ 5.8)	58.0(↓ 12.2)	33.7	61.3	9.6(↓ 24.1)	15.9(↓ 45.4)
CVS [12]	75.3	90.4	65.7	85.1	61.4(↓ 4.3)	78.3(↓ 6.8)	56.8	78.7	49.6(↓ 7.2)	67.1(↓ 11.6)
LSTKC [31]	74.2	89.4	63.1	83.9	64.6(↑ 1.5)	83.6(↓ 0.3)	48.8	74.9	38.5(↓ 10.3)	61.3(↓ 13.6)
DJAA (ours)	74.6	90.1	70.2	88.3	**71.7**(↑ 1.5)	**89.0**(↑ 0.7)	59.0	81.0	**65.2**(↑ 6.2)	**85.6**(↑ 4.6)

TABLE X
ABLATION STUDY ON THE ADAPTATION LOSSES ($\mathcal{L}_{pa}$ AND $\mathcal{L}_{ia}$) AND THE SIMILARITY CONSISTENCY LOSSES ($\mathcal{L}_{ps}$ AND $\mathcal{L}_{is}$). WE REPORT THE AVERAGED RESULTS ON SEEN AND UNSEEN DOMAINS.

$\mathcal{L}_{pa}$	$\mathcal{L}_{ia}$	$\mathcal{L}_{ps}$	$\mathcal{L}_{is}$	Seen		Unseen	
				$\bar{s}_{mAP}$	$\bar{s}_{R1}$	$\bar{s}_{mAP}$	$\bar{s}_{R1}$
✓				40.2	59.4	41.4	39.7
✓	✓			41.9	62.7	47.0	44.9
✓	✓	✓		51.9	66.5	52.2	49.5
✓	✓		✓	53.0	70.0	53.3	50.8
✓	✓	✓	✓	**54.9**	**71.6**	**57.1**	**54.3**

TABLE XI
ABLATION STUDY ON THE BACKWARD COMPATIBILITY. THE SELF-TEST IS REPORTED BETWEEN $\mathcal{Q}_2$ AND $\mathcal{G}_2$. $\mathcal{Q}_2$ IS THE MARKET1501 QUERY FEATURES EXTRACTED AFTER STEP 2 (CUHK-SYSU). $\mathcal{G}_2$ IS THE MARKET1501 GALLERY FEATURES EXTRACTED AFTER STEP 2. THE CROSS-TEST IS REPORTED BETWEEN $\mathcal{Q}_2$ AND $\mathcal{G}_1$. $\mathcal{G}_1$ IS THE MARKET1501 QUERY FEATURES STORED AFTER STEP 1.

$\mathcal{L}_{pa}$	$\mathcal{L}_{ia}$	$\mathcal{L}_{ps}$	$\mathcal{L}_{is}$	Self-Test($\mathcal{Q}_2,\mathcal{G}_2$)		Cross-Test($\mathcal{Q}_2,\mathcal{G}_1$)	
				mAP	$R1$	mAP	$R1$
✓				19.0	40.0	9.6(↓ 9.4)	24.1(↓ 15.9)
✓	✓			49.5	72.9	40.1(↓ 9.4)	56.1(↓ 16.8)
✓	✓	✓		68.2	86.4	64.5(↓ 3.7)	82.6(↓ 3.8)
✓	✓		✓	69.0	87.6	69.9(↑ 0.9)	89.0(↑ 1.4)
✓	✓	✓	✓	70.2	88.3	**71.7**(↑ 1.5)	**89.0**(↑ 0.7)

TABLE XII
ABLATION STUDY ON DATA AUGMENTATION, INCLUDING RANDOM HORIZONTAL FLIPPING (FLIP), RANDOM CROPPING (CROP), RANDOM BLURRING (BLUR) AND RANDOM ERASING (ERASE). WE REPORT THE AVERAGED RESULTS ON SEEN AND UNSEEN DOMAINS.

Flip	Crop	Blur	Erase	ColorJitter	Seen		Unseen	
					$\bar{s}_{mAP}$	$\bar{s}_{R1}$	$\bar{s}_{mAP}$	$\bar{s}_{R1}$
✓	✓	✓	✓		54.9	**71.6**	**57.1**	**54.3**
	✓	✓	✓		49.8	65.6	51.3	47.8
✓		✓	✓		52.4	68.5	53.4	49.7
✓	✓		✓		53.3	69.1	54.3	51.2
✓	✓	✓			48.1	65.5	49.2	45.5
✓	✓	✓	✓	✓	**55.1**	**71.6**	**57.1**	53.2

TABLE XIII
COMPARISON OF DIFFERENT DATA AUGMENTATION SETTINGS FOR AUGMENTATION CONSISTENCY REGULARIZATION.

Loss	Pred	Ref	Seen		Unseen	
			$\bar{s}_{mAP}$	$\bar{s}_{R1}$	$\bar{s}_{mAP}$	$\bar{s}_{R1}$
$\mathcal{L}_{ps}$	Weak	Weak	55.2	**72.3**	56.7	54.1
	Strong	Strong	**55.3**	71.4	56.4	53.8
	Strong	Weak	54.9	71.6	**57.1**	**54.3**
$\mathcal{L}_{is}$	Weak	Weak	48.5	67.9	50.6	48.5
	Strong	Strong	54.2	70.7	56.2	53.6
	Strong	Weak	**54.9**	**71.6**	**57.1**	**54.3**

four augmentations (random flipping, cropping, blurring and erasing) are empirically selected. The results in Table XII show that adding color jitter augmentation achieves similar results. Thus, we do not bother adding color jitter.

In addition to regularizing the consistency between two encoders of different steps, we also introduce data augmentation perturbations to further enhance the model robustness. We report the performance of different data augmentation settings in Table XIII. Different data augmentation settings on the prototype-level similarity consistency loss '$\mathcal{L}_{ps}$' have only a slight influence on the final performance. For the instance-level similarity consistency loss '$\mathcal{L}_{is}$', the influence of the data augmentation setting is more evident. We can observe that data augmentation brings in meaningful perturbations for consistency regularization. Using weakly augmented similarity as reference to regularize the strongly augmented prediction similarity with perturbations is the optimal setting for augmentation consistency regularization.

3) Hyperparameter analysis: : We use grid search to set the optimal temperature and weight hyper-parameters to balance our proposed losses. The temperature hyperparameters τ_{ps} and τ_{is} control the scale of image-to-prototype and image-to-image similarity. Based on the results in Table XIV, we set the temperature hyperparameters $\tau_{ps} = 0.1$ and $\tau_{is} = 0.2$ in image-to-prototype and image-to-image similarity rehearsal losses, respectively. The weight hyperparameters λ_{ps} and λ_{is} balance the importance of contrastive adaptation, image-to-prototype and image-to-image similarity losses. Based on the results in Table XIV, we set the balancing weight hyperparameters $\lambda_{ps} = 10$ and $\lambda_{is} = 20$. The exponential moving average hyperparameter α control the the speed of the knowledge accumulation from the online encoder to the momentum encoder. Table XIV shows that $\alpha = 0.999$ is the optimal setting for our framework. The hyperparameters are tuned in the training order Market→Cuhk-Sysu→MSMT17. The tuned hyperparameters are kept the same for the second training order MSMT17→Market→Cuhk-Sysu, which validates the effectiveness of these hyperparameters.

G. Discussion

1) Training order: As datasets are of different scales and diversity, it is meaningful to know whether the training order has a significant influence on the final results. Our primary training order starts from a medium domain Market and ends with the largest domain MSMT17. However, it is hard to control the order of upcoming domains in the real world. We

TABLE XIV
COMPARISON OF DIFFERENT VALUES FOR TRAINING HYPERPARAMETERS. τ_{ps} AND τ_{is} ARE TEMPERATURE HYPERPARAMETERS. λ_{ps} AND λ_{is} ARE WEIGHT BALANCING HYPERPARAMETERS. α IS THE EXPONENTIAL MOVING AVERAGE (EMA) HYPERPARAMETER.

τ_{ps}	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
0.05	52.5	68.5	53.9	51.3
0.1	54.9	**71.6**	**57.1**	**54.3**
0.2	**55.4**	**71.6**	56.1	52.8

τ_{is}	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
0.1	53.5	69.6	52.5	49.5
0.2	**54.9**	**71.6**	**57.1**	**54.3**
0.3	47.5	65.6	48.3	44.3

λ_{ps}	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
5	54.6	70.9	53.9	50.3
10	**54.9**	**71.6**	**57.1**	**54.3**
15	47.9	64.4	49.0	45.4

λ_{is}	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
10	54.2	69.6	54.0	50.7
20	**54.9**	**71.6**	**57.1**	**54.3**
30	52.9	71.1	55.2	52.4

α	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
0.99	53.4	70.1	53.2	50.8
0.999	**54.9**	**71.6**	**57.1**	**54.3**
0.9999	38.4	51.2	42.4	40.3

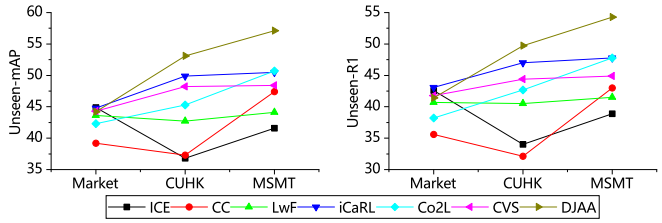


Fig. 8. Non-forgetting evaluation with mAP and Rank1 on the first seen domain MSMT17. The training order is MSMT17→Market501→Cuhk-Sysu.

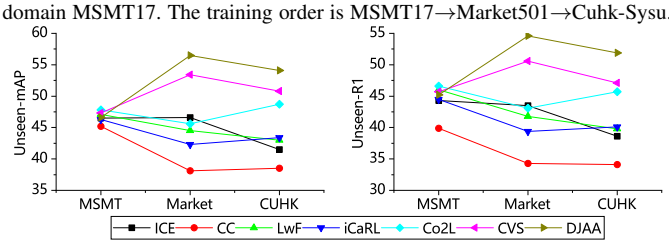


Fig. 9. Generalization ability evaluation with averaged mAP and averaged Rank1 on all the unseen domains. The training order is MSMT17→Market501→Cuhk-Sysu.

test a second order MSMT17→Market→Cuhk-Sysu that starts from the largest domain MSMT17. In the second training order, it is easier to forget more knowledge on the largest domain MSMT17. Consequently, the performance under the training order #2 is slightly inferior to that of the training order #1. As shown in Table XV, DJAA still significantly outperforms the state-of-the-art methods CLUDA and CVS on seen domains. In the meantime, as shown in Table XVI, DJAA also outperforms state-of-the-art methods on unseen domains. We provide a forgetting curve on the seen domains in Fig. 8 and a generalization ability on unseen domains in Fig. 9. Compared to previous methods, our proposed DJAA shows both strong anti-forgetting and generalization abilities. We further compare backward-compatible ability between state-of-the-art methods and our proposed method DJAA in the second training order. As shown in Table XV, DJAA consistently shows superior backward-compatible ability over previous methods in the second training order.

2) Data selection for memory buffer update: To update the image memory at the end of each step, our strategy is to first select n_{new} clusters with largest cluster sample numbers. The prototypes of selected clusters are used to update the memory buffer. Then, we select one image that is closest to its cluster prototype. In this way, we select the representative clusters for our image-to-prototype similarity consistency loss. As shown in Table XVIII, we compare three possible memory buffer updating strategies. 'Random' refers to randomly sample n_{new} images from the current domain, which neglects the clustering information. 'ID-wise' refers to randomly sample n_{new} clusters and one image per cluster, which has better data diversity than 'Random'. 'ours' refers to our proposed strategy in Section III-D that selects representative clusters and credible images, which brings in the best performance on both seen and unseen domains.

3) Memory buffer size: We build a hybrid memory buffer to store cluster prototypes and image samples for the rehearsal module of our method DJAA. With a default memory size $n_{mem} = 512$, our image memory stores approximately 512 images $\approx 0.8\%$ of all the training images (Market, Cuhk-Sysu and MSMT17). Meanwhile, our prototype memory stores approximately 512 prototype vectors (dimension $1 \times 2048 \times 1 \times 1$)≈4.2 MB, which is negligible compared with storing dataset images (for example, MSMT17≈2.5 GB).

To further evaluate the dependency on the memory size, we vary the value of n_{mem} and report the results in Table XIX. If the buffer size is limited to an extremely small number, such as 32 or 64, the results are even lower than our adaptation module baseline. When the buffer size is greater than or equals 128, the results are higher than the baseline. We can observe that the more data we store, the higher performance DJAA can achieve. As shown in Table IV, CLUDA [32] achieves 46.4% mAP and 62.8% mAP with a buffer size of 512. It is worth mentioning that our method achieves higher performance (49.8% mAP and 67.3% mAP) than CLUDA [32], when the buffer size equals 128.

4) Efficiency of backward compatibility: Backward compatible learning aims to maintain the consistency of representations after training more data, so that we do not need to re-extract gallery features after each adaptation step. As shown in Table IX, with Market gallery features extracted at step 1, ICE has a cross-test performance 30.3% mAP / 49.3% Rank1 at step 2 and 3.4% mAP / 6.3% Rank1 at step 3. In such case, Market gallery features need to be extracted at steps 1, 2 and 3, and Cuhk-Sysu gallery features needs to be extracted at steps 2 and 3. In contrast, reinforced by the feature back-compatibility, Market and Cuhk-Sysu gallery features need to be extracted only once in our method, at the 1^{st} and 2^{nd} step respectively. For example, with stored Market gallery features, DJAA has a cross-test performance 71.7% mAP / 89.0% Rank1 at step 2 and 65.2% mAP / 85.6% Rank1 at step 3. A real-world lifelong ReID system may involve more adaptation steps, where a backward compatible method can avoid repetitive gallery feature extraction and fasten the inference.

TABLE XV

SEEN-DOMAIN RESULTS (%) OF UNSUPERVISED (U) AND UNSUPERVISED LIFELONG (UL) METHODS ON ONE-CYCLE FULL SET BENCHMARK UNDER THE TRAINING ORDER MSMT17→MARKET→CUHK-SYSU. * REFERS TO USING THE CAMERA LOSS EQ.(10) TO REDUCE INTRA-CLUSTER VARIANCE. THE BEST PERFORMANCE IS MARKED IN BOLD.

Method	Memory size	Type	MSMT17		Market		Cuhk-Sysu		Average	
			mAP	Rank1	mAP	Rank1	mAP	Rank1	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
ICE [2]	0	U	5.4	14.0	42.1	63.4	**82.5**	**84.2**	43.3	53.9
CC [76]	0	U	4.8	14.0	40.9	64.6	74.1	76.9	39.9	51.8
PPRL [22]	0	U	4.4	12.9	45.4	70.5	80.2	83.0	43.3	55.5
LwF [42]	0	UL	6.4	20.8	37.8	66.0	74.2	77.3	39.5	54.7
iCaRL [45]	512	UL	6.5	20.3	40.0	68.7	75.7	78.7	40.7	55.9
Co^{2}L [47]	512	UL	7.9	24.4	42.8	69.5	79.8	82.3	43.5	58.8
CLUDA [32]	512	UL	13.9	31.5	50.1	77.3	82.1	84.2	48.7	64.3
CVS [12]	512	UL	14.6	39.6	53.3	79.1	75.2	77.6	47.7	65.4
LSTKC [31]	512	UL	11.7	32.1	47.8	75.4	77.6	80.4	45.7	63.6
DJAA (ours)	512	UL	18.1	43.8	53.7	79.0	82.0	84.0	51.3	68.9
DJAA* (ours)	512	UL	**20.2**	**48.5**	**59.3**	**84.1**	79.2	81.9	**52.9**	**71.5**

TABLE XVI

UNSEEN-DOMAIN RESULTS (%) OF UNSUPERVISED (U), UNSUPERVISED LIFELONG (UL) AND DOMAIN GENERALIZATION (DG) METHODS ON ONE-CYCLE FULL SET BENCHMARK. THE TRAINING ORDER IS MSMT17→MARKET→CUHK-SYSU. * REFERS TO USING THE CAMERA LOSS EQ.(10) TO REDUCE INTRA-CLUSTER VARIANCE. THE BEST UNSUPERVISED PERFORMANCE IS MARKED IN BOLD.

Method	Memory Size	Type	VIPeR		PRID		GRID		iLIDS		CUHK01		CUHK02		SenseReID		CUHK03		3DPeS		Average	
			mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	$\bar{s}_{mAP}$	$\bar{s}_{R1}$
ICE [2]	0	U	41.3	31.6	32.8	22.0	31.1	20.0	72.1	61.7	47.6	47.1	51.3	47.3	36.7	29.1	13.6	32.7	47.1	55.5	41.5	38.6
CC [76]	0	U	36.7	25.9	35.0	26.0	29.2	20.8	64.1	55.0	40.7	38.5	45.1	42.1	31.7	24.4	18.0	17.1	46.4	56.8	38.5	34.1
PPRL [22]	0	U	38.7	30.4	25.4	16.0	28.3	20.8	69.6	60.0	47.4	47.5	47.5	43.1	29.6	23.5	14.4	13.5	41.8	53.9	38.1	34.3
LwF [42]	0	UL	40.5	30.1	36.6	28.0	36.6	29.6	73.0	61.7	45.9	45.5	51.0	47.3	33.7	26.7	19.2	30.8	50.6	58.6	43.0	39.8
iCaRL [45]	512	UL	40.3	29.7	36.0	25.0	36.0	26.4	74.4	66.7	48.2	46.6	49.6	45.8	35.0	28.0	18.6	32.2	52.6	60.5	43.4	40.1
Co^{2}L [47]	512	UL	41.8	31.6	50.2	38.0	38.4	27.2	79.0	70.0	56.6	55.6	57.0	55.0	38.2	31.8	20.9	36.7	56.4	65.5	48.7	45.7
CVS [12]	512	UL	41.0	32.3	43.5	33.0	**41.3**	31.2	**85.0**	**80.0**	61.0	57.8	59.1	56.3	40.4	32.7	29.4	36.3	56.4	64.1	50.8	47.1
LSTKC [31]	512	UL	39.1	27.8	45.3	35.0	37.8	28.8	77.4	68.3	61.6	61.7	60.1	59.6	38.2	29.9	23.0	36.5	51.6	61.4	48.2	45.5
DJAA (ours)	512	UL	**47.6**	**36.4**	**51.8**	**41.0**	40.6	**32.8**	82.5	75.0	69.1	68.9	**67.4**	**65.7**	44.9	**37.6**	25.5	**45.2**	57.2	64.1	54.1	**51.9**
DJAA* (ours)	512	UL	46.1	33.9	48.5	37.0	40.5	30.4	80.5	71.7	**70.7**	**70.2**	**67.4**	65.3	**45.0**	37.1	**31.0**	43.6	**59.3**	**66.4**	**54.3**	50.6
ACL [25]	All	DG	69.2	60.8	60.0	50.0	58.8	48.8	89.9	86.7	78.4	77.8	77.1	76.4	62.4	53.5	65.1	67.3	71.1	78.8	70.2	66.7

TABLE XVII

BACKWARD-COMPATIBLE PERFORMANCE ON MSMT17. FULL SET TRAINING ORDER #2: MSMT17→MARKET→CUHK-SYSU. $\mathcal{Q}_1$,$\mathcal{Q}_2$ AND $\mathcal{Q}_3$ RESPECTIVELY DENOTE MSMT17 QUERY FEATURES EXTRACTED AFTER STEP 1, 2 AND 3. $\mathcal{G}_1$,$\mathcal{G}_2$ AND $\mathcal{G}_3$ RESPECTIVELY DENOTE MSMT17 GALLERY FEATURES EXTRACTED AFTER STEP 1, 2 AND 3. THE COLORED NUMBER IS THE DIFFERENCE BETWEEN SELF-TEST AND CROSS-TEST.

Method	Step 1		Step 2				Step 3			
	Self-Test ($\mathcal{Q}_1,\mathcal{G}_1$)		Self-Test ($\mathcal{Q}_2,\mathcal{G}_2$)		Cross-Test ($\mathcal{Q}_2,\mathcal{G}_1$)		Self-Test ($\mathcal{Q}_3,\mathcal{G}_3$)		Cross-Test ($\mathcal{Q}_3,\mathcal{G}_1$)	
	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
ICE [2]	32.7	65.7	6.6	20.1	3.8(↓ 2.8)	9.3(↓ 10.8)	7.6	22.6	1.0(↓ 6.6)	2.7(↓ 19.9)
Co^{2}L [47]	25.8	57.6	7.0	20.9	2.2(↓ 4.8)	6.1(↓ 14.8)	7.7	23.8	0.2(↓ 7.5)	0.3(↓ 23.5)
CVS [12]	27.3	59.2	18.3	44.7	19.5(↑ 1.2)	45.4(↑ 0.7)	14.6	39.6	15.3(↑ 0.7)	36.6(↓ 3.0)
LSTKC [31]	28.0	59.8	17.1	41.2	19.7(↑ 2.6)	44.7(↑ 3.5)	11.7	32.1	12.2(↑ 0.5)	30.4(↓ 1.7)
DJAA (ours)	27.1	59.3	21.6	49.4	**24.4**(↑ 2.8)	**53.9**(↑ 4.5)	18.1	43.8	**20.6**(↑ 2.5)	**48.3**(↑ 4.5)

TABLE XVIII

COMPARISON OF MEMORY BUFFER UPDATING STRATEGIES. 'RANDOM' REFERS TO RANDOMLY SAMPLE n_{new} IMAGES FROM OLD AND CURRENT DOMAINS. 'ID-WISE' REFERS TO RANDOMLY SAMPLE n_{new} CLUSTERS AND ONE RANDOM IMAGE PER CLUSTER. 'OURS' REFERS TO SAMPLE n_{new} CLUSTERS WITH THE LARGEST CLUSTER SAMPLE NUMBERS AND ONE IMAGE NEAREST TO THE PROTOTYPE.

Method	Seen		Unseen	
	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
Random	53.8	69.8	53.9	50.6
ID-wise	54.6	71.1	55.8	53.1
ours	**54.9**	**71.6**	**57.1**	**54.3**

TABLE XIX

NUMBER OF IMAGES PER PSEUDO IDENTITY IN THE MEMORY. 'BASELINE' REFERS TO ONLY USE THE ADAPTATION MODULE, I.E., $\mathcal{L}_{pa}+\mathcal{L}_{ia}$, WHICH DOES NOT STORE ANY SAMPLE FOR REHEARSAL.

Buffer Size	Seen		Unseen	
n_{mem}	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$	$\bar{s}_{mAP}$	$\bar{s}_{Rank1}$
Baseline	41.9	62.7	47.0	44.9
32	35.8	52.9	38.8	35.3
64	40.2	56.4	41.5	38.8
128	49.8	67.3	51.8	49.3
256	52.5	69.6	54.2	51.5
512	**54.9**	**71.6**	**57.1**	**54.3**

V. CONCLUSION

In this paper, we propose an anti-forgetting adaptation method for unsupervised person ReID. The traditional unsupervised domain adaptation methods aim at obtaining the optimal performance on a fixed target domain, which shows poor anti-forgetting, generalization and backward-compatible ability. To tackle the three problems at once, we propose a Dual-level Joint Adaptation and Anti-forgetting (DJAA) method, which mainly consists of an adaptation module and a rehearsal module. In the adaptation module, we leverage a prototype-level contrastive loss and an instance-level contrastive loss to maximize the positive view similarity for learning new domain features. In the rehearsal module, our method regularizes the image-to-prototype and image-to-image similarity across

domains to mitigate the knowledge forgetting. In comparison with previous lifelong methods, our proposed DJAA significantly improves the non-forgetting ability on seen domains and better generalization ability on unseen domains.

REFERENCES

[1] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *JMLR*, 2008.

[2] H. Chen, B. Lagadec, and F. Bremond, "Ice: Inter-instance contrastive encoding for unsupervised person re-identification," in *ICCV*, 2021.

[3] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE TPAMI*, 2021.

[4] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, "Joint discriminative and generative learning for person re-identification," in *CVPR*, 2019.

[5] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," in *ICCV*, 2021.

[6] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Learning to adapt invariance in memory for person re-identification," *IEEE TPAMI*, 2020.

[7] N. Pu, W. Chen, Y. Liu, E. M. Bakker, and M. S. Lew, "Lifelong person re-identification via adaptive knowledge accumulation," in *CVPR*, 2021.

[8] G. Wu and S. Gong, "Generalising without forgetting for lifelong person re-identification," in *AAAI*, 2021.

[9] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nature Machine Intelligence*, pp. 1–13, 2022.

[10] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," *CVPR*, 2019.

[11] Y. Shen, Y. Xiong, W. Xia, and S. Soatto, "Towards backward-compatible representation learning," in *CVPR*, 2020.

[12] T. S. Wan, J.-C. Chen, T.-Y. Wu, and C.-S. Chen, "Continual learning for visual search with backward consistent feature embedding," in *CVPR*, 2022.

[13] S. Tang, P. Su, D. Chen, and W. Ouyang, "Gradient regularized contrastive learning for continual domain adaptation," in *AAAI*, 2021.

[14] M. Rostami, "Lifelong domain adaptation via consolidated internal distribution," *NeurIPS*, 2021.

[15] H. Lee, S. Eum, and H. Kwon, "Negative samples are at large: Leveraging hard-distance elastic loss for re-identification," in *ECCV*, 2022.

[16] X. Zhou, Y. Zhong, Z. Cheng, F. Liang, and L. Ma, "Adaptive sparse pairwise loss for object re-identification," in *CVPR*, 2023.

[17] G. Zhang, Y. Zhang, T. Zhang, B. Li, and S. Pu, "Pha: Patch-wise high-frequency augmentation for transformer-based person re-identification," in *CVPR*, 2023.

[18] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *CVPR*, 2019.

[19] K. Zheng, C. Lan, W. Zeng, Z. Zhang, and Z.-J. Zha, "Exploiting sample uncertainty for domain adaptive person re-identification," in *AAAI*, 2021.

[20] K. Zheng, W. Liu, L. He, T. Mei, J. Luo, and Z.-J. Zha, "Group-aware label transfer for domain adaptive person re-identification," in *CVPR*, 2021.

[21] Y. Ge, F. Zhu, D. Chen, R. Zhao, and H. Li, "Self-paced contrastive learning with hybrid memory for domain adaptive object re-id," in *NeurIPS*, 2020.

[22] Y. Cho, W. J. Kim, S. Hong, and S.-E. Yoon, "Part-based pseudo label refinement for unsupervised person re-identification," in *CVPR*, 2022.

[23] X. Jin, C. Lan, W. Zeng, Z. Chen, and L. Zhang, "Style normalization and restitution for generalizable person re-identification," in *CVPR*, 2020.

[24] Y. Dai, X. Li, J. Liu, Z. Tong, and L.-Y. Duan, "Generalizable person re-identification with relevance-aware mixture of experts," in *CVPR*, 2021.

[25] P. Zhang, H. Dou, Y. Yu, and X. Li, "Adaptive cross-domain learning for generalizable person re-identification," in *ECCV*, 2022.

[26] B. Jiao, L. Liu, L. Gao, G. Lin, L. Yang, S. Zhang, P. Wang, and Y. Zhang, "Dynamically transformed instance normalization network for generalizable person re-identification," in *ECCV*, 2022.

[27] B. Xu, J. Liang, L. He, and Z. Sun, "Mimic embedding via adaptive aggregation: learning generalizable person re-identification," in *ECCV*, 2022.

[28] Y. Lu, M. Wang, and W. Deng, "Augmented geometric distillation for data-free incremental person reid," in *CVPR*, 2022.

[29] W. Ge, J. Du, A. Wu, Y. Xian, K. Yan, F. Huang, and W.-S. Zheng, "Lifelong person re-identification by pseudo task knowledge preservation," in *AAAI*, 2022.

[30] C. Yu, Y. Shi, Z. Liu, S. Gao, and J. Wang, "Lifelong person re-identification via knowledge refreshing and consolidation," in *AAAI*, 2023.

[31] K. Xu, X. Zou, and J. Zhou, "Lstkc: Long short-term knowledge consolidation for lifelong person re-identification," in *AAAI*, 2024.

[32] Z. Huang, Z. Zhang, C. Lan, W. Zeng, P. Chu, Q. You, J. Wang, Z. Liu, and Z.-j. Zha, "Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation," in *CVPR*, 2022.

[33] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *CVPR*, 2018.

[34] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020.

[35] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *ICML*, 2020.

[36] M. Wang, B. Lai, J. Huang, X. Gong, and X.-S. Hua, "Camera-aware proxies for unsupervised person re-identification," in *AAAI*, 2021.

[37] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, and F. Bremond, "Joint generative and contrastive learning for unsupervised person re-identification," in *CVPR*, 2021.

[38] ——, "Learning invariance from generated variance for unsupervised person re-identification," *IEEE TPAMI*, 2023.

[39] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive neural networks," *arXiv preprint arXiv:1606.04671*, 2016.

[40] J. Yoon, E. Yang, J. Lee, and S. J. Hwang, "Lifelong learning with dynamically expandable networks," in *ICLR*, 2018.

[41] A. Mallya and S. Lazebnik, "Packnet: Adding multiple tasks to a single network by iterative pruning," in *CVPR*, 2018.

[42] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE TPAMI*, 2018.

[43] A. Douillard, Y. Chen, A. Dapogny, and M. Cord, "Plop: Learning without forgetting for continual semantic segmentation," in *CVPR*, 2021.

[44] C. Shang, H. Li, F. Meng, Q. Wu, H. Qiu, and L. Wang, "Incrementer: Transformer for class-incremental semantic segmentation with knowledge distillation focusing on old class," in *CVPR*, 2023.

[45] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "icarl: Incremental classifier and representation learning," in *CVPR*, 2017.

[46] F. M. Castro, M. J. Marín-Jiménez, N. Guil, C. Schmid, and K. Alahari, "End-to-end incremental learning," in *ECCV*, 2018.

[47] H. Cha, J. Lee, and J. Shin, "Co2l: Contrastive continual learning," in *ICCV*, 2021.

[48] Z. Luo, Y. Liu, B. Schiele, and Q. Sun, "Class-incremental exemplar compression for class-incremental learning," in *CVPR*, 2023.

[49] Y. Choi, M. El-Khamy, and J. Lee, "Dual-teacher class-incremental learning with data-free generative replay," in *CVPR*, 2021.

[50] G. M. Van de Ven, H. T. Siegelmann, and A. S. Tolias, "Brain-inspired replay for continual learning with artificial neural networks," *Nature communications*, vol. 11, no. 1, p. 4069, 2020.

[51] Q. Wang, O. Fink, L. Van Gool, and D. Dai, "Continual test-time domain adaptation," in *CVPR*, 2022.

[52] S. Wu, L. Chen, Y. Lou, Y. Bai, T. Bai, M. Deng, and L.-Y. Duan, "Neighborhood consensus contrastive learning for backward-compatible representation," in *AAAI*, 2022.

[53] T. Pan, F. Xu, X. Yang, S. He, C. Jiang, Q. Guo, F. Qian, X. Zhang, Y. Cheng, L. Yang *et al.*, "Boundary-aware backward-compatible representation via adversarial learning in image retrieval," in *CVPR*, 2023.

[54] M. Oh and J.-Y. Sim, "Lifelong person re-identification with backward-compatibility," *arXiv preprint arXiv:2403.10022*, 2024.

[55] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, 1996.

[56] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *CVPR*, 2017.

[57] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *NeurIPS*, 2020.

[58] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *CVPR*, 2019.

[59] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," *ICCV*, 2015.

[60] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," *CVPR*, 2017.

[61] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *CVPR*, 2018.

[62] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *ECCV*, 2008.

[63] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person Re-Identification by Descriptive and Discriminative Classification," in *Proc. Scandinavian Conference on Image Analysis (SCIA)*, 2011.

[64] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *CVPR*, 2009.

[65] W.-S. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *BMVC*, 2009.

[66] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *ACCV*, 2012.

[67] W. Li and X. Wang, "Locally aligned feature transforms across views," *CVPR*, 2013.

[68] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," *CVPR*, 2017.

[69] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," *CVPR*, 2014.

[70] D. Baltieri, R. Vezzani, and R. Cucchiara, "3dpes: 3d people dataset for surveillance and forensics," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, 2011.

[71] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *NeurIPS*, 2019.

[72] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *IJCV*, 2015.

[73] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

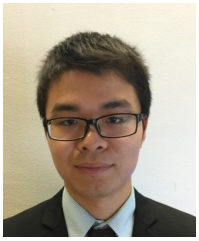
[74] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *AAAI*, 2020.

[75] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.

[76] Z. Dai, G. Wang, W. Yuan, S. Zhu, and P. Tan, "Cluster contrast for unsupervised person re-identification," in *ACCV*, 2022.

[77] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

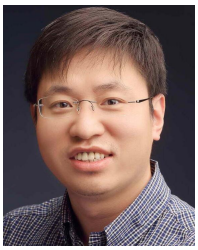
[78] X. Han, Q. You, C. Wang, Z. Zhang, P. Chu, H. Hu, J. Wang, and Z. Liu, "Mmptrack: Large-scale densely annotated multi-camera multiple people tracking benchmark," in *WACV*, 2023.



Hao Chen received the Ph.D. degree in computer science from INRIA in 2022. He is currently a Post-Doctoral Researcher at Peking University. His research interests include fine-grained image understanding, unsupervised learning and incremental learning.

Francois Bremond received the Ph.D. degree from INRIA in video understanding in 1997, and he pursued his research work as a post doctorate at the University of Southern California (USC) on the interpretation of videos taken from Unmanned Airborne Vehicle (UAV). In 2007, he received the HDR degree (Habilitation a Diriger des Recherches) from Nice University on Scene Understanding. He created the STARS team on the 1st of January 2012. He is the research director at INRIA Sophia Antipolis, France. He has conducted research work in video understanding since 1993 at Sophia- Antipolis. He is author or co-author of more than 140 scientific papers published in international journals or conferences in video understanding. He is a handling editor for MVA and a reviewer for several international journals (CVIU, IJPRAI, IJHCS, PAMI, AIJ, Eurasip, JASP) and conferences (CVPR, ICCV, AVSS, VS, ICVS). He has (co-)supervised 26 PhD theses. He is an EC INFSO and French ANR Expert for reviewing projects.

Nicu Sebe is Professor in the University of Trento, Italy, where he is leading the research in the areas of multimedia analysis and human behavior understanding. He was the General Co-Chair of the IEEE FG 2008 and ACM Multimedia 2013. He was a program chair of ACM Multimedia 2011 and 2007, ECCV 2016, ICCV 2017 and ICPR 2020. He is a general chair of ACM Multimedia 2022 and a program chair of ECCV 2024. He is a fellow of IAPR.



Shiliang Zhang received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences. He was a Post-Doctoral Scientist with NEC Laboratories America and a Post-Doctoral Research Fellow with The University of Texas at San Antonio. He is currently an Associate Professor with Tenure with the Department of Computer Science, School of Electronic Engineering and Computer Science, Peking University.

His research interests include large-scale image retrieval and computer vision. He has authored or co-authored over 100 papers in journals and conferences, including IJCV, IEEE Trans. on PAMI, IEEE Trans. on Image Processing, IEEE Trans. on NNLS, IEEE Trans. on Multimedia, ACM Multimedia, ICCV, CVPR, ECCV, NeurIPS, AAAI, IJCAI, etc. He was a recipient of the Outstanding Doctoral Dissertation Awards from the Chinese Academy of Sciences and Chinese Computer Federation, the President Scholarship from the Chinese Academy of Sciences, the NEC Laboratories America Spot Recognition Award, the NVidia Pioneering Research Award, and the Microsoft Research Fellowship. He served as the Associate Editor (AE) of Computer Vision and Image Understanding (CVIU) and IET Computer Vision, Guest Editor of ACM TOMM, and Area Chair of CVPR, AAAI, ICPR, and VCIP. His research is supported by the The National Key Research and Development Program of China, Natural Science Foundation of China, Beijing Natural Science Foundation, and Microsoft Research, etc.