ECAI 2023
K. Gal et al. (Eds.)

doi:10.3233/FAIA230578

Online Privacy Preservation for Camera-Incremental Person Re-Identification

Sheng Wu[a], Wenhang Ge[b], Jiong Wu[c], Jingke Meng[d] and Huang Zhang[a;*]

[a]School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha, China
[b]The Hong Kong University of Science and Technology (Guangzhou)
[c]Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, USA
[d]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China

Abstract. Task-incremental person re-identification aims to train a model with consecutively available cross-camera annotated data in the current task and a small number of saved data in preceding tasks, which may lead to individual privacy disclosure due to data storage and annotation. In this work, we investigate a more realistic online privacy preservation scenario for camera-incremental person re-identification, where data storage in preceding cameras is not allowed, while data in the current camera are intra-camera annotated online by a pedestrian tracking algorithm without cross-camera annotation. In this setup, the missing data of previous cameras not only results in catastrophic forgetting as task-incremental learning, but also makes the cross-camera association infeasible, which further leads to the incapability of person matching across cameras due to the camera-wise domain gap. To solve these problems, we propose an **O**nline **P**rivacy **P**reservation (OPP) framework based on the generated exemplars of previous cameras by DeepInversion, where generated exemplars used as supplements to alleviate forgetting and enable cross-camera association to be feasible for camera-wise domain shift mitigation, meanwhile further improving the cross-camera matching capability. Specifically, we propose to mine underlying cross-camera positive pairs between samples of the current camera and exemplars of previous cameras by similarity cues. Furthermore, we introduce a mixup learning strategy to handle the domain gap with mixed samples and labels. Finally, intra-camera incremental learning and cross-camera incremental learning are aggregated into the OPP framework. Extensive experiments on Re-ID benchmarks validate the superiority of the OPP framework as compared with state-of-the-art methods.

1 Introduction

Person re-identification (Re-ID) is to match query images across a set of gallery images from non-overlapping camera views [36, 39, 9, 8]. Thanks to the large-scale cross-camera labelled datasets [42, 29], supervised person Re-ID methods have achieved excellent performance. However, this strong supervision across multiple camera views from surveillance data is costly and time-consuming, meanwhile results in privacy disclosure due to manual annotation. Though this supervision is not necessary in unsupervised person Re-ID, most unsupervised methods still have a big performance margin compared with supervised counterparts in the absence of supervision [15, 5].

* Corresponding author. Email: zhanghuang@csust.edu.cn

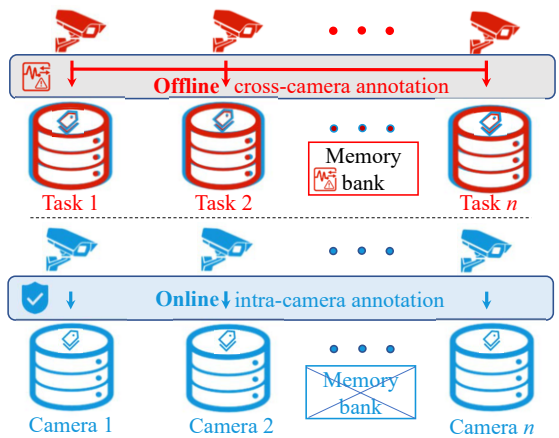


Figure 1. Comparison between traditional task-incremental learning (Top) and camera-incremental learning (Bottom) for person Re-ID.

Currently, some representative works for person Re-ID turn to exploit the offline intra-camera supervised learning strategy to avoid the costly cross-camera annotation [10, 43].

In vision surveillance cameras, large-scale data usually becomes available gradually over time [11, 17, 37]. This often results in a high re-training cost when new online data becomes available, making them poorly scalable for the above methods when they perceive data sequentially. Recently, task-based lifelong learning and continual learning in the case of person Re-ID exhibited their scalable superiority [32, 9, 23]. However, as illustrated in Figure 1, these conventional Re-ID methods focus on training with strong supervision across camera views and relies on the memory bank for the storage of representative samples [9], which may lead to the possibility of individual privacy disclosure during the offline data storage and cross-camera annotation [1, 31]. The solution of cross-camera manual annotation for continually-generated data is thus unrealistic. However, pedestrians in each single camera can be readily detected and identified by human tracking algorithm [12, 7, 6]. As shown in Figure 2, the pedestrians within the i-th camera view can be inartificially detected and identified with each other without human intervention and annotation [30, 24, 22]. Therefore, the tracking algorithm can readily determine the intra-camera annotation within a camera view without privacy disclosure. By this way, online data with the intra-camera annotation in the i-th camera view can be arranged as the i-th data set and sequentially perceived by the learning machine, which we refer

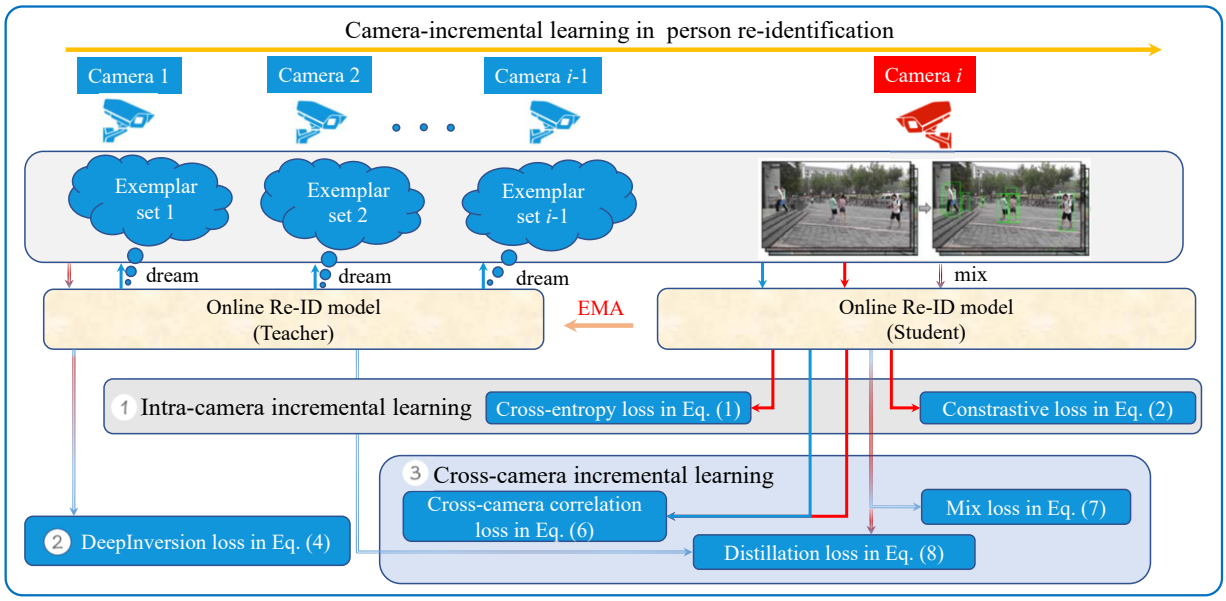


Figure 2. Diagram of the OPP framework, where the online Re-ID model continually perceive intra-camera labeled data without storing old data. Here, a flow of online camera-wise data sets are non-repetitively perceived by the student model, while the teacher model takes charge of retrospecting or dreaming those exemplars by DeepInversion in preceding camera views. The union of the previous exemplar sets and current image set is used as input to the student model for intra-camera incremental learning in Subsection 3.2 and cross-camera incremental learning in Subsection 3.3. The three-stage optimization procedure are tagged by the sequential numbers ($1 \rightarrow 2 \rightarrow 3 \rightarrow 1$).

as camera-incremental learning person re-identification.

To protect privacy, all perceived data are deleted online when next training data arrives. Without available data of previous cameras, training the model exhibits three challenges. First, the model trends to forget previous knowledge acquired from previous camera views; Second, the absence of cross-camera positive pairs leads to the difficulty of the cross-camera association which is the key to endow the cross-camera matching capability of the Re-ID model. Third, the camera-wise domain shift cannot be explicitly mitigated which damages the online Re-ID performance.

To address these issues, we propose an **O**nline **P**rivacy **P**reservation (OPP) framework, which online dreams exemplars of previous camera views by DeepInversion [38]. Based on these dreamed exemplars, the old knowledge can be renewed and the cross-camera association becomes feasible. We propose a cross-camera correlation loss, which can mine underlying cross-camera positive pairs by similarity cues to alleviate domain gap. To further mitigate the gap, we introduce a mixup learning loss, which mix samples of current camera and exemplars of previous camera for camera-invariant representation learning. Finally, we combine intra-camera incremental learning, cross-camera incremental learning and knowledge distillation in our framework to address all issues discussed above. The main contributions of our work are summarized as follows:

- We introduce a more practical camera-incremental learning setting for person Re-ID, where storing the data of previous cameras is not permitted and cross-camera annotation is not available due to privacy pretection. To supplement the missing, we apply an exemplar-based method to online dream identity-specific exemplars.
- To address the incapability of cross-camera matching induced by the camera-wise domain gap, we incrementally exploit the similarity cues to associate potential cross-camera positive pairs and mix the generated exemplars with current samples to guide the camera-invariance representation learning.
- By using knowledge distillation, we exploit the teacher-student framework with exponential moving average (EMA) strategy to maintain the memory consistency and mitigate catastrophic forgetting when the model continually perceives data in camera-incremental person re-identification.

2 Related Work

2.1 *Intra-Camera Person Re-identification*

Most supervised person Re-ID methods require precise annotation of accurately finding those identities matched across camera views [20, 10, 24, 9]. However, the Re-ID pedestrian data are usually generated online and captured by video footage of public surveillance cameras, which could easily disclose individual confidential information such as personal activities or daily individual whereabouts due to the artificial annotation. The unauthorized collection of online pedestrian data for the person Re-ID task exists some risk of privacy leakage due to the storage of all data [40]. Therefore, numerous supervised Re-ID methods [35, 18, 28] mostly focus on boosting their performance on public Re-ID datasets while neglecting privacy protection in the data storage and annotation. Although offline unsupervised learning methods have been proposed recently without any supervision to learn person Re-ID models [34, 14, 33], full data collection into a storage device in the unsupervised setup still incurs the possible privacy leakage and inferior performance due to the absence of clear annotation.

To overcome these fundamental limitations, some representative works for person Re-ID turn to exploit the intra-camera supervised learning strategy, which is able to associate the underlying cross-camera positive pairs to avoid the privacy disclosure and efforts during annotation [43, 10, 16, 33]. However, high re-training cost is needed for these supervised, unsupervised, and/or intra-camera supervised methods above when new online data becomes available,

disclosing their poorly scalability for the scenario where they perceive data sequentially. It is unrealistic to make the individual extension for these person Re-ID methods in a more practical and challenging scenario, where online continually-generated surveillance data lead to the data storage cost and privacy leakage.

2.2 Lifelong Person Re-identification

To tackle the scalability of a Re-ID model, lifelong learning is a learning strategy capable of continually upgrading a system with a flow of new data stream [9, 32, 4, 2]. A key challenge in lifelong learning lies in minimising catastrophic forgetting, which means that the model leverages new information for its upgrades while preserving old knowledge learned in the previous tasks. Many approaches have been developed to address knowledge forgetting for common vision tasks such as object detection [27] and segmentation [21]. Recently, there were rare works tackling the problem of camera-wise continual learning in the case of person Re-ID, which requires a Re-ID model to be incrementally generalised without forgetting knowledge already learned in previous camera views. However, lifelong learning methods recently emerge in the case of person Re-ID through a cross-domain learning way. For example, supervised augmented geometric distillation (AGD) framework proposed in [19] keeps evolving to train on a sequence of Re-ID domain tasks, where the model still used cross-camera supervision, which could damage the confidentiality of online data. Huang et al. [13] addresses an unsupervised scenario where stored images from previous domain tasks were allowed. We consider the above compromise case between supervised and unsupervised lifelong learning, where pedestrian images can be detected and identified by the object detector to form the intra-camera annotation without any privacy leakage, and stored intra-camera images from previous camera views is not permitted.

3 The Proposed Approach

In this section, we formally introduce the framework of the online privacy preservation (OPP). Section 3.1 introduces the problem formulation. Section 3.2 introduces the intra-camera incremental learning which aims to improve the intra-camera discriminative ability by identity loss with the gradually-perceived data. Section 3.3 introduces cross-camera incremental learning to further guarantee the cross-camera matching capability and alleviate forgetting. In this section, we first employ DeepInversion technology to online dream exemplars to supplement the missing samples of previous camera views in Subsection 3.3.1. Based on these exemplars and the real samples of current camera view, we propose cross-camera correlation loss in Subsection 3.3.2 to associate underlying cross-camera positive pairs by similarity cues and mix learning strategy in Subsection 3.3.3 to alleviate the camera-wise domain shift.

3.1 Problem Formulation

As illustrated in Figure 2, we assume an online Re-ID model non-repetitive inspects from n cameras, and is currently perceiving the i-th ($1 \leq i \leq n$) camera data set $\mathcal{X}_i$ in which the intra-camera annotation $\mathcal{Y}_i$ was created by a human tracking algorithm without any human-induced privacy leakage. The online Re-ID model consists of a student model and a teacher model as illustrated in Figure 2, in which both the teacher and student model contain a feature extractor ϕ_θ, contrastive module φ_ϑ and incremental module ψ_ω with parameters θ, ϑ and ω, respectively. The EMA upgrade strategy is used in the teacher-student framework to maintain the memory consistency between them. We define a set to online collect all intra-camera labeled data in $\boldsymbol{\mathcal{R}} = \cup_{i=1}^{n}\{(\mathcal{X}_i, \mathcal{Y}_i)\}$, in which the observed camera data can not be assessed twice and will be deleted online. Let C_i be the number of captured identities in the i-th camera and $S_i = \sum_{k=1}^{i} C_k$ denotes those perceived identities covered from the 1st to the i-th camera view. Note that identities in different cameras may be overlapping, i.e., the same identities could appear in multiple cameras. By using the cues of the cross-camera identity overlapping, our goal is to incrementally uncover the potential cross-camera identity similarity and use it to online train a robust feature representation network ϕ_θ. We define the descriptor of the k-th sample $\boldsymbol{x}_k^i \in \mathcal{X}_i$ by $\boldsymbol{f}_k^i = \phi_\theta(\boldsymbol{x}_k^i)$.

3.2 Intra-Camera Incremental Learning

Given the feature extractor ϕ_θ^{s} and incremental module ψ_ω^{s} in the student network, along with the continual inspection of camera views, we optimize the camera-specific parameter ω_i for the current learning task of the i-th camera view by the cross-entropy loss as below:

$$\mathcal{L}_{\mathrm{ce}}(\omega_i) = \frac{-1}{|\mathcal{B}_i|} \sum_{(\boldsymbol{x}^i, \boldsymbol{y}^i) \in \mathcal{B}_i} 1\{\boldsymbol{y} = \boldsymbol{y}^i\} \log(\psi_{\omega_i}^{\mathrm{s}}(\phi_\theta^{\mathrm{s}}(\boldsymbol{x}^i))) \tag{1}$$

where $\mathcal{B}_i$ is a batch of current data in the i-th camera view, and $(\boldsymbol{x}^i, \boldsymbol{y}^i)$ denotes an arbitrary input-to-target pair from the batch $\mathcal{B}_i$. The number of parameters and output dimension of the incremental module will increase along with the continual inspection to fit the new added classes.

To further boost the discriminative ability of the model, we employ the contrastive loss within the i-th camera view. Specifically, Given arbitrary sample pairs $\boldsymbol{x}_p^i, \boldsymbol{x}_q^i \in \mathcal{X}_i$ with intra-camera annotation $\boldsymbol{y}_p^i, \boldsymbol{y}_q^i \in \mathcal{Y}_i$ and feature representations $\boldsymbol{f}_p^i$ and $\boldsymbol{f}_q^i$, The contrastive loss for the i-th currently inspected camera view is described as below:

$$\begin{aligned} &\mathcal{L}_{\mathrm{con}}(\varphi_\vartheta^{\mathrm{s}}, \phi_\theta^{\mathrm{s}}) \\ &= \frac{1}{2|\mathcal{X}_i|} \sum_{p,q,p\neq q} \Big(1\{\boldsymbol{y}_p^i = \boldsymbol{y}_q^i\} \|\varphi_\vartheta^{\mathrm{s}}(\boldsymbol{f}_p^i) - \varphi_\vartheta^{\mathrm{s}}(\boldsymbol{f}_q^i)\|_2^2 \\ &\quad + 1\{\boldsymbol{y}_p^i \neq \boldsymbol{y}_q^i\} \max\left(0, m - \|\varphi_\vartheta^{\mathrm{s}}(\boldsymbol{f}_p^i) - \varphi_\vartheta^{\mathrm{s}}(\boldsymbol{f}_q^i)\|_2^2\right) \Big), \end{aligned} \tag{2}$$

where $\varphi_\vartheta^{\mathrm{s}}$ is a contrastive module in the student network for intra-camera contrastive learning and m denotes the margin. Finally, the intra-camera incremental learning loss is formulated by:

$$\mathcal{L}_{\mathrm{ICI}} = \mathcal{L}_{\mathrm{ce}} + \mathcal{L}_{\mathrm{con}} \tag{3}$$

The intra-camera incremental learning above fails to consider the catastrophic forgetting and cross-camera matching capability with camera-wise domain shift. To address these problems, we present a cross-camera incremental learning scheme in the ensuing subsection.

3.3 Cross-Camera Incremental Learning

3.3.1 Dreaming Exemplars by DeepInversion

In view of the data privacy preservation in the online Re-ID setup, the storage of images from the 1st to the $(i-1)$-th camera view is not permitted. However, the samples in previous cameras are necessary for mitigating forgetting and alleviating the camera-wise domain shift. So we adopt the DeepInversion technology to online dream exemplars in the previous camera views as a supplement of missing data.

As shown in the exemplar sets of Figure 2, the randomly initialized exemplar $\hat{\boldsymbol{x}}$ in these sets as the optimizable variables are feed to the teacher network to output the identity label $\hat{\boldsymbol{y}}$ for the retrospection of those identities from previous inspected camera views. More specifically, the identity-specific exemplar $\hat{\boldsymbol{x}}$ is generated by the following DeepInversion loss:

$$\begin{aligned}&\mathcal{L}_{\text{DI}}(\hat{\boldsymbol{x}}) \\ &= -\sum_{(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})\in\mathcal{B}} 1\{\boldsymbol{y}=\hat{\boldsymbol{y}}\}\log(\psi_\omega^{\text{t}}(\phi_\theta^{\text{t}}(\hat{\boldsymbol{x}}))) + \mathcal{R}_{\text{DI}}(\hat{\boldsymbol{x}}),\end{aligned} \tag{4}$$

where $\mathcal{B}$ is a batch of optimizable inputs and $\mathcal{R}_{\text{DI}}(\hat{\boldsymbol{x}})$ is the DeepInversion loss [38] to facilitate $\hat{\boldsymbol{x}}$ to retrospect the identity-specific knowledge in previous camera views. We assume the exemplar generation per identity is p, and thus there are total pS_{i-1} exemplars generated by the teacher network, with which we aim at mitigating forgetting and alleviating domain shift.

3.3.2 Cross-Camera Correlation by Similarity Cues

With dreamed exemplars, matching the same identity across cameras becomes feasible when the online Re-ID model continually perceives the sequential union of camera data sets. The main challenge is how to associate potential cross-camera positive pairs without available cross-camera annotation. To handle this problem, we leverage the similarity cues between exemplars in preceding cameras and real samples in the current camera for cross-camera correlation.

More specifically, we correlate the feature representation $\boldsymbol{f}_p$ of a real image in the current camera and the averaged counterpart $\boldsymbol{f}_q$ of generated exemplars sharing the same identity in previous cameras according to the feature similarity computed as below:

$$s_{p,q} = \exp\big(-\|\boldsymbol{f}_p - \boldsymbol{f}_q\|_2^2/\sigma^2\big), \tag{5}$$

if $\boldsymbol{f}_p \in \mathcal{N}_k(\boldsymbol{f}_q) \ \wedge \ \boldsymbol{f}_q \in \mathcal{N}_k(\boldsymbol{f}_p)$ and 0 otherwise, where $\mathcal{N}_k(\cdot)$ indicates the set of the k nearest neighbors of the alternative $\boldsymbol{f}_p$ or $\boldsymbol{f}_q$. We randomly select a batch data from the i-th currently camera view and create a similarity matrix by Eq. (5), and normalize each row to obtain a correlation matrix denoted by $\boldsymbol{W}$, in which each row element indicates the similar possibility to correlate an identity in the current batch and the others in previous inspected camera views. We define the following cross-camera correlation (CCC) loss to incrementally correlate those similar identities across camera views:

$$\mathcal{L}_{\text{ccc}}(\phi_\theta^{\text{s}},\psi_\omega^{\text{s}}) = -\frac{1}{|\mathcal{B}|}\sum_{k=1}^{|\mathcal{B}|}\boldsymbol{W}^{(k)}\log(\psi_\omega^{\text{s}}(\phi_\theta^{\text{s}}(\boldsymbol{x}_k))), \tag{6}$$

where $\mathcal{B}$ denotes the batch size; $\boldsymbol{W}^{(k)}$ denotes the k-th row of $\boldsymbol{W}$ and $\boldsymbol{x}_k$ is the k-th sample from a batch data derived from the i-th currently camera view.

3.3.3 Mix Learning by Exemplars and Images

Inspired by the Mixup approach in [3], where labelled/unlabelled data and manual/guessing labels are interleaved or mixed to calculate batch normalization for the distribution consistency of labelled and unlabelled data. We feed old-camera exemplars mixed with the images in the current camera to obtain camera invariant features. More specifically, we mix a generated exemplar $\hat{\boldsymbol{x}}$ and real sample $\boldsymbol{x}$ randomly by the vanilla Mixup approach, i.e., $\boldsymbol{x}_{\text{m}} = (\hat{\boldsymbol{x}} + \boldsymbol{x})/2$ with corresponding mixed label $\boldsymbol{y}_{\text{m}} = (\hat{\boldsymbol{y}} + \boldsymbol{y})/2$, where $\hat{\boldsymbol{y}}$ and $\boldsymbol{y}$ are the guessed labels of $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}$ generated by a soften function $s(\boldsymbol{p},T)_v := \boldsymbol{p}_v^{1/T}/\sum_{j=1}^{S_{i-1}}\boldsymbol{p}_j^{1/T}$ with temperature T, which is used to softly enhance those smaller entries corresponding to the potential cues among similar identities across camera views, while compromise those confident cues corresponding to the incredible similar identities across camera views in a reversible way. We define the mix learning loss as below:

$$\mathcal{L}_{\text{mix}}(\phi_\theta^{\text{s}},\psi_\omega^{\text{s}}) = \sum_{(\boldsymbol{y}_{\text{m}},\boldsymbol{x}_{\text{m}})\in\mathcal{B}}\|\boldsymbol{y}_{\text{m}} - \psi_\omega^{\text{s}}(\phi_\theta^{\text{s}}(\boldsymbol{x}_{\text{m}}))\|_2^2. \tag{7}$$

Besides intra-camera incremental learning and inter-camera incremental learning, we adopt the knowledge distillation to further alleviate forgetting. Specifically, the teacher network is updated in a slower manner compared to student work which is updated in a faster manner to learn new knowledge, and hence contains more knowledge of previous data. We enforce the student network to mimic the teacher network to alleviate forgetting and define a knowledge distillation loss formulated by:

$$\begin{aligned}&\mathcal{L}_{\text{kd}}(\phi_\theta^{\text{s}},\psi_\omega^{\text{s}}) \\ &= \frac{-1}{|\mathcal{B}|S_i}\sum_{\boldsymbol{x}\in\mathcal{B}}\sum_{k=1}^{S_i}\overline{\sigma}_k(\psi_\omega^{\text{t}}(\phi_\theta^{\text{t}}(\boldsymbol{x})))\log(\sigma_k(\psi_\omega^{\text{s}}(\phi_\theta^{\text{s}}(\boldsymbol{x})))),\end{aligned} \tag{8}$$

where $\mathcal{B}$ is a batch which contains both generated exemplars and current samples; σ_k denotes the k-th output entry of the softmax function and $\overline{\sigma}_k$ denotes the detach operation from teacher network. Finally, the cross-camera incremental learning loss is formulated by:

$$\mathcal{L}_{\text{CCI}} = \mathcal{L}_{\text{kd}} + \alpha_i\mathcal{L}_{\text{ccc}} + \mathcal{L}_{\text{mix}}, \tag{9}$$

where $\alpha_i = i/n$ and $1 < i \leq n$, linearly increases during the training stage since the model becomes increasingly confident in the cross-camera correlation when more identities appear in camera views, which enhances the possibility of matching similar identities across camera views.

3.3.4 Online Privacy Preservation Framework

Our framework is learned with joint guidance of intra-camera incremental learning, DeepInversion, cross-camera incremental learning by:

$$\mathcal{L}_{\text{OPP}} = \{\mathcal{L}_{\text{ICI}},\mathcal{L}_{\text{DI}},\mathcal{L}_{\text{CCI}}\}, \tag{10}$$

where the three-stage learning procedure is an alternative and cycled optimization by $\mathcal{L}_{\text{ICI}} \rightarrow \mathcal{L}_{\text{DI}} \rightarrow \mathcal{L}_{\text{CCI}} \rightarrow \mathcal{L}_{\text{ICI}}$.

4 Experiments

4.1 Implementation Details

To implement our proposed OPP method, we adopt Resnet-50 as our backbone for feature extraction. The output dimension of the last layer in Resnet-50 is 2048. Following the last layer, we append a contrastive module for contrastive intra-camera learning and an incremental module for camera-wise incremental learning. The learning rates of the SGD optimizers for the base ResNet-50 backbone, contrastive module and incremental module are $1e-5$, $1e-3$ and $1e-3$. In the exemplar generation in DeepInversion, we optimize random exemplars with 1000 iterations and 64 batch size in two rounds of exemplar random rolling. All input images and exemplars

Market-1501	r=1	r=5	r=10	r=20	mAP
Full OPP model	**88.21**	**94.92**	**96.38**	**97.68**	**73.36**
w/o CCI loss	76.70	88.26	90.24	91.80	66.45
w/o ICI loss	72.28	83.16	83.52	86.49	63.47
baseline	71.44	80.06	82.15	84.21	59.43
DukeMTMC	r=1	r=5	r=10	r=20	mAP
Full OPP model	**74.78**	**86.64**	**90.00**	**92.37**	**58.32**
w/o CCI loss	68.75	85.31	86.22	88.17	55.25
w/o ICI loss	62.14	82.27	83.87	85.10	53.53
baseline	63.02	82.63	83.25	86.61	52.43
MARS	r=1	r=5	r=10	r=20	mAP
Full OPP model	**63.94**	**70.01**	**72.76**	**76.69**	**46.30**
w/o CCI loss	62.20	67.73	68.50	74.54	44.28
w/o ICI loss	61.75	65.96	66.48	70.48	42.30
baseline	59.24	64.38	66.12	72.47	41.55
MSMT17	r=1	r=5	r=10	r=20	mAP
Full OPP model	**55.73**	**67.82**	**72.30**	**76.29**	**37.89**
w/o CCI loss	52.43	61.56	69.77	73.30	35.73
w/o ICI loss	50.32	58.60	67.84	74.63	33.25
baseline	51.89	58.24	64.22	72.83	32.18

Table 1. Ablation study of the proposed OPP method. Here, the CCI loss in Eq. (9) and ICI loss in Eq. (3) the OPP method was integrally removed to investigate its removal impact on the online Re-ID model. The **w/o** CCI and ICI loss functions denote that the OPP model was carried out without the CCI and ICI loss functions. The baseline denotes that the OPP model was carried out with the ICI loss.

Martket-1501	r=1	r=5	r=10	r=20	mAP
WA-iCaRL [25]	76.46	90.55	92.61	95.20	63.22
RM [2]	70.42	86.33	89.52	92.84	65.36
Co^{2}L [4]	82.36	89.40	91.47	96.76	68.20
AGD [19]	80.43	87.65	90.22	94.37	66.24
PTKP [9]	84.80	90.26	92.58	94.72	69.54
OPP (Ours)	**88.21**	**94.92**	**96.38**	**97.68**	**73.36**
DukeMTMC	r=1	r=5	r=10	r=20	mAP
WA-iCaRL [25]	60.83	78.37	83.64	90.25	45.70
RM [2]	58.25	68.61	76.35	84.86	49.38
Co^{2}L [4]	63.67	76.56	81.64	87.27	47.20
AGD [19]	68.50	82.04	86.12	87.35	49.76
PTKP [9]	66.27	77.36	82.40	88.58	49.29
OPP (Ours)	**74.78**	**86.27**	**90.00**	**92.37**	**54.32**
MARS	r=1	r=5	r=10	r=20	mAP
WA-iCaRL [25]	46.44	61.13	67.56	75.03	35.78
RM [2]	46.27	60.32	66.94	73.53	36.81
Co^{2}L [4]	48.35	59.84	67.73	73.06	37.78
AGD [19]	52.53	75.68	72.25	75.33	36.41
PTKP [9]	54.16	62.75	68.40	73.81	39.16
OPP (Ours)	**62.94**	**70.01**	**72.76**	**76.69**	**44.30**
MSMT17	r=1	r=5	r=10	r=20	mAP
WA-iCaRL [25]	36.36	65.24	69.38	74.06	29.34
RM [2]	32.72	60.46	68.28	73.53	28.12
Co^{2}L [4]	30.50	61.37	68.86	72.26	28.54
AGD [19]	35.38	62.50	64.39	75.87	32.18
PTKP [9]	41.78	63.29	69.63	73.51	30.44
OPP (Ours)	**51.73**	**67.82**	**72.30**	**76.29**	**35.89**

Table 2. Performance comparison with continual learning methods on four Re-ID benchmark datasets within a single domain. The best results are in black boldface font.

are shaped to 256×128. In the intra-camera training stage, the online Re-ID model was trained with 250 epoches for intra-camera contrastive learning. In the cross-camera incremental learning stage, the online Re-ID model was trained with 50 epoches for cross-camera incremental learning. In the label guessing stage, we adopt two random augmentations for the soften label guessing, i.e., $K = 2$ for the augmentation of exemplars and images. The order of the camera sequence was shuffled randomly with the fixed random seed in the experiments.

4.2 Datasets and Settings

The performance of the OPP method in the experiments were evaluated on four benchmark datasets: Market-1501 [42], DukeMTMC [26], MSMT17 [29] and MARS [41]. In the online training stage, we first create continually-perceived sets in advance for sequentially accessing training data in each of these datasets. The element in the continually-perceived data sets contains either video tracklets or identity frames in the individual camera. In the testing stage, we use all images in the query set to match possible images in the gallery set, where the performance of our OPP method was evaluated by the mean average precision (mAP) and we also report the Rank-1 ($r = 1$), Rank-5 ($r = 2$), Rank-10 ($r = 10$) and Rank-20 ($r = 20$) scores. More specifically, the continually-perceived data sets divided from each of the following datasets were presented as follows.

The **Market-1501** dataset contained a large number of identity images from 6 disjoint cameras and used the DPM as pedestrian detector. The naming rule of this dataset is `0001_c1s1_001051_00.jpg`, where `0001` denotes the person identity and `c1s1` is the `0001` identity appearing in the sequence 1 of camera 1. Therefore, we can first form an intra-camera `c1s1` data set containing different intra-camera identities in our OPP setup. Thereafter, total 25 continually-perceived data sets were created by the union of all cxsy data sets and gradually observed by the online Re-ID model. The continually-perceived data sets for **DukeMTMC** and **MSMT17** were created by the similar way. In the **MARS** dataset, all bounding boxes and tracklets are generated automatically by DPM and GMMCP. The naming rule of this dataset is `0000C6T3036F006.jpg`, where `0000` denotes the person identity; `C6` and `T3036` are the `0000` identity appearing in the tracklet 3036 of camera 6. Therefore, we can use the tracklet `T3036` as a data set if it has the number of images larger than 20000; Otherwise, we merge the tracklet with others into the data set until the number of images of the data set is larger than 20000. Finally, we obtain total 22 continually-perceived data sets.

Figure 3. Selective exemplar images generated by the DeepInversion for the replay with new arriving camera data.

4.3 Evaluation of the OPP Method

In our OPP method, we introduce the intra-camera incremental (ICI) loss in Eq. (3) and the cross-camera incremental (CCI) loss in Eq.

Methods	Reference	Train: Market1501 → DukeMTMC → MARS → MSMT17									
		Market-1501		DukeMTMC		MARS		MSMT17		Average	
		$r=1$	mAP	$r=1$	mAP	$r=1$	mAP	$r=1$	mAP	$\overline{s}_{r=1}$	$\overline{s}_{\text{mAP}}$
JTA (Upper bound)*	AAAI21	94.60	86.30	89.10	76.30	67.36	49.42	78.50	53.10	82.39	66.28
WA-iCaRL†	CVPR20	76.46	63.22	61.52	46.43	48.04	37.27	37.29	30.75	55.83	44.42
RM†	CVPR21	70.42	65.36	58.34	48.11	47.40	38.87	33.53	28.27	52.42	45.15
Co^{2}L†	ICCV21	82.36	68.23	64.20	48.53	48.57	38.69	31.36	29.25	56.62	46.18
AGD†	CVPR22	81.27	67.54	63.49	48.21	50.45	41.36	36.22	30.73	57.86	46.96
PTKP†	AAAI22	84.80	69.54	67.35	50.47	54.42	40.76	43.55	31.26	62.53	48.01
OPP	Ours	**88.21**	**73.36**	**76.29**	**55.20**	**63.85**	**45.67**	**55.14**	**36.58**	**70.87**	**52.70**

Table 3. Comparison with state-of-the-art methods across domains. The notation "*" means the results of models using the cross-camera annotation of sequential camera-wise data. "†" means we implement the released code on our baseline, where no cross-camera annotation was used.

Market-1501	r=1	mAP	**MSMT17**	r=1	mAP
MATE [43]	**88.70**	71.10	MATE [43]	46.00	19.10
PTKP [9]	84.80	69.54	PTKP [9]	41.78	30.44
OPP (Ours)	88.21	**73.36**	OPP (Ours)	**51.73**	**35.89**

Table 4. Performance comparison with intra-camera supervised methods.

DukeMTMC	r=1	r=10	mAP
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{mix}}$ + $\mathcal{L}_{\text{kd}}$	**74.78**	**90.00**	**54.32**
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{mix}}$	64.84	84.24	53.69
baseline + $\mathcal{L}_{\text{kd}}$ + $\mathcal{L}_{\text{mix}}$	68.82	87.04	53.71
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{kd}}$	68.76	87.17	53.80
baseline	63.02	83.25	52.43

Market>DukeMTMC>MARS>MSMT17	r=1	mAP
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{mix}}$ + $\mathcal{L}_{\text{kd}}$	**90.00**	**54.32**
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{mix}}$	84.24	53.69
baseline + $\mathcal{L}_{\text{kd}}$ + $\mathcal{L}_{\text{mix}}$	87.04	53.71
baseline + $\mathcal{L}_{\text{ccc}}$ + $\mathcal{L}_{\text{kd}}$	87.17	53.80
baseline	83.25	52.43

Table 5. Each loss is added one by one in the OPP method.

(9). To investigate the efficiency of the two loss functions, we removed each of them from our OPP method and explored the magnitude of performance degradation individually. We reported the performance comparison to the baseline method, which only used the cross-entropy loss in Eq. (1) and basic contrastive loss in Eq. (2) to learn the online Re-ID model. The results are reported in Table 1, where the notations "w/o CCI" and "w/o ICI" represented results without the CCI and ICI loss, respectively.

Comparing the full OPP method to the baseline in Table 1, it exhibited the efficiency of our OPP method in the case of the person Re-ID with online privacy preservation. More specifically, the reported results represent notable improvements in the rank-1 matching accuracy, e.g., 16.77%, 11.76%, 4.70% and 3.84% improvements were approximately observed on the Market-1501, DukeMTMC, MARS and MSMT17 datasets, respectively. Note that the gains on the Market-1501 and DukeMTMC datasets were more remarkable than the gains on the other datasets. Considering mAP, we also obtain 4∼14% improvement on these continually intra-camera labeled datasets.

Moreover, as reported in Table 1, the ablation study indicates that the removal of the ICI loss will lead to the notable performance degradation. For example, on the Market-1501 and DukeMTMC datasets, we observed approximate 15.93% and 12.64% degradation, respectively, which indicates that the ICI loss was quietly useful to improve the feature discriminative confidence in the current task. More detailed impact of each individual loss was presented in Table 5.

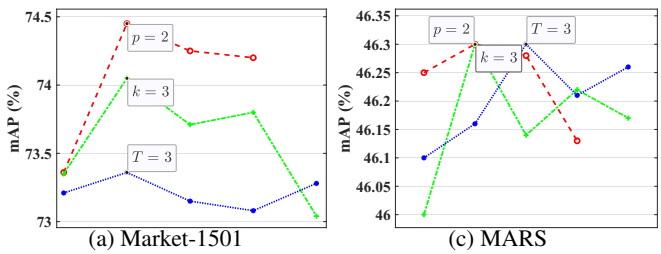


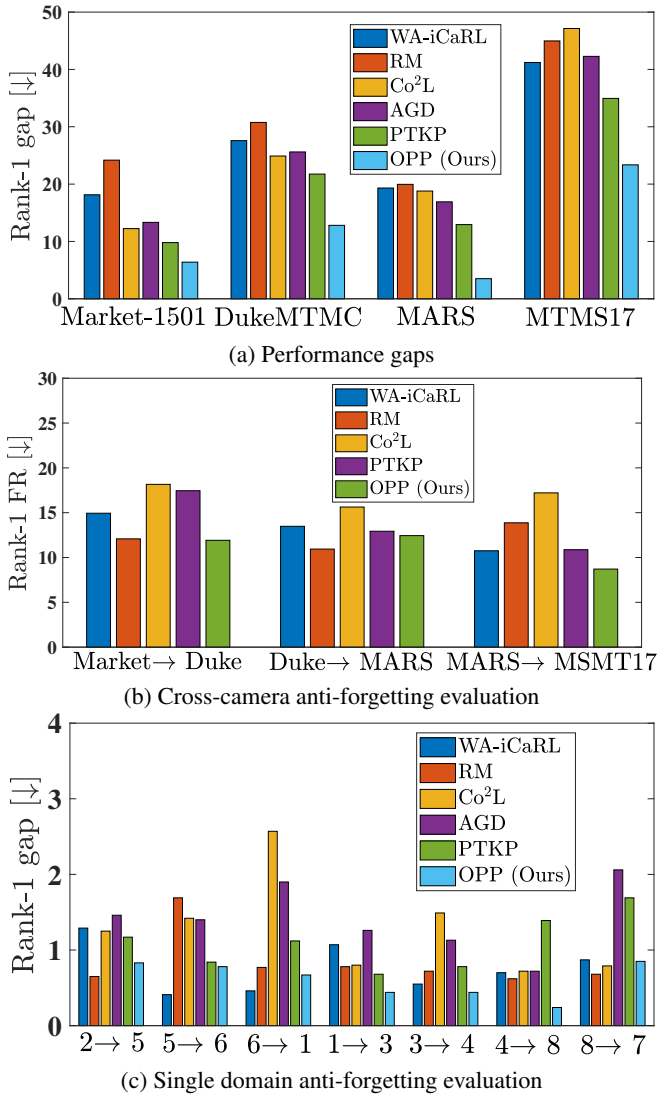
Figure 4. Accurate grid search for parameter-sensitive exploration on the Market-1501 and DukeMTMC datasets, where the OPP performance changed with respect to the values of the three parameters p, k and T.

4.4 Evaluation on Camera-wise Learning

This work mainly focuses on handling the privacy disclosure issue by the online camera-wise perception within a single domain. Beyond the performance evaluation within a single domain, we still evaluate cross-domain continual learning performance in this section for the proposed OPP method. The compared methods include related state-of-the-art alternative methods: WA-iCaRL [25], JTA [32], Co^{2}L [4], RM [2], AGD [19] and PTKP [9]. Since only intra-camera annotation was available in our online Re-ID setup, the contrastive learning method Co^{2}L can be directly used for comparison in this scenario; We only retain the components using intra-camera annotation in the PTKP method for the fair comparison. The others were also trained by the intra-camera supervision. Note that the JTA method, a baseline assembling all datasets in advance for joint training, was used to show the upper bound for reference.

We argue that the realistic changes such as cross-camera visual ambiguity and appearance variation caused by illumination, camera viewpoint, background and occlusions are challenging enough for continually learning a model within a single domain, especially with the consideration of the privacy disclosure in this online scenario. For comprehensive evaluation, we focus on handling the privacy disclosure issue within a single domain by the proposed OPP method. Furthermore, we still study cross-domain continual learning and explore the cross-domain performance of the proposed OPP method. The performance evaluation was divided into three parts as follows:

- Evaluation across domains. Table 3 reported the comparison results with state-of-the-art methods across domains, where the evaluation results were recorded after sequentially training on each dataset. As reported in Table 3, our OPP method outperforms the other com-



(a) Performance gaps

(b) Cross-camera anti-forgetting evaluation

(c) Single domain anti-forgetting evaluation

Figure 5. (a) Performance gaps with reference to the JTA method that uses the cross-camera annotation. (b) Rank-1 forgetting ratio for the cross-domain anti-forgetting evaluation. (c) Rank-1 forgetting ratio for the single domain anti-forgetting evaluation. Lower is better.

petitive methods in the online Re-ID setup, where only intra-camera annotation was provided without the possiblity of privacy disclosure. In contrast, the JTA method was trained by the cross-camera supervision, and thus the performance of our OPP method can not outperform to the method. More specifically, compared to the SOTA method PTKP, we achieved improvements of approximately 3.41%, 8.94%, 9.43% and 11.59% rank-1 matching accuracy on the Market-1501, DukeMTMC, MARS and MSMT17 datasets. Compared to the upper bound JTA, the performance gaps between the usage and non-usage of the cross-camera annotation were relatively small on the four datasets as shown in Figure 5(a). These results exhibited the effectiveness of the OPP method. Furthermore, the OPP method was compared with intra-camera methods and reported in Table 4.

- Evaluation within single domain. Table 2 reported the comparison results with state-of-the-art methods within a single domain, where the evaluation results were recorded after training on each dataset. Different from previous cross-domain training, there were no dataset domain knowledge that can be transfered sequentially. In contrast, there are only camera-wise knowledge that can be transferred in a single domain. As reported in Table 2, our OPP method still outperforms the compared methods in the online Re-ID setup, where only intra-camera annotation was used, but the performance can not compare with the sequential learning across domains in Table 3, since there cross-domain knowledge can be transferred from previous datasets. More specifically, compared to the SOTA method PTKP, we achieved improvements of approximately 3.41%, 8.51%, 8.78% and 9.95% on the Market-1501, DukeMTMC, MARS and MSMT17 datasets in terms of the rank-1 matching accuracy.

- Evaluation on forgetting. After sequentially training on each dataset or training with camera data, we computed each rank-1 performance difference of compared methods to measure the anti-forgetting capability. Lower difference indicates better anti-forgetting capability. More specifically, the cross-camera and single domain performance differential results were shown in Figure 5(b) and Figure5(c), respectively. Note that all intra-camera learning methods were used for the comparison except for the JTA that uses the cross-camera annotation. We observed that the proposed OPP method suffers from less forgetting than the others. This is because OPP utilizes the DeepInversion regularization strategy to regularize the model parameters by the generated exemplars as shown in Figure 3 when learning from new data, and then we feed old-camera exemplars or images along with the current task images to learn the network jointly for mitigating catastrophic forgetting.

4.5 Parameter-Sensitive Analysis

There are three hyperparameters involved in our OPP method: 1) The number of generated exemplars per identity denoted as p built by the base teacher network to preserve preceding knowledge from previous inspected camera views to address catastrophic forgetting in Subsection 3.3.1. 2) The number of k nearest neighbors in the formulation of Eq. (5). 3) The distribution temperature T in mix learning. Figure 4 shows the OPP's performance sensitive to the three parameters on the Market-1501 and DukeMTMC datasets. We recorded the mAP w.r.t. different values of the three parameters to approximately figure out their best values. As shown in Figure 4, the model approximately performs best when $p = 2$, $T = 3$, $k = 3$ or $k = 4$ and on the above datasets in a range of parameter-selection possibilities.

5 Conclusion

In view of the possibility of privacy disclosure in the online person Re-ID scenario, this work removes the needs for cross-camera annotation and data storage for online person Re-ID, where the online Re-ID model perceived data in a camera-wise manner and no cross-camera annotated and stored data incurred privacy disclosure. The proposed online Re-ID model in this setup was designed to sequentially learn across non-overlapping camera views only using the online intra-camera annotation, and meanwhile incrementally perceives the camera-wise data without forgetting previous knowledge already learned. We used generated exemplars as supplements of the missing data in previous inspected cameras, making cross-camera association, knowledge preservation and camera-wise domain shift mitigation to be feasible in the online person Re-ID scenario. More specifically, we used cross-camera correlation and mix learning strategies to discover the cross-camera similarity cues and incrementally correlate cross-camera identities potentially belonging to the same person. The generated exemplars in previous inspected cameras along with real images in the current camera were mixed to guide the camera-invariance representation learning. The experimental results have verified the effectiveness of the proposed OPP method in the online person Re-ID scenario.

Acknowledgements

We would like to thank the reviewers and meta-reviewers for their valuable comments, which helped improve this paper considerably. This work was supported in part by the National Natural Science Foundation of China under Grant 61876194 and Grant 62206093; in part by the Natural Science Foundation of Hunan Province under Grants 2023JJ30437, 2023JJ40054, 2022JJ40290 and 2021JJ30477; in part by the Youth Foundation of Hunan Province Department of Education under Grant 21B0619.

References

[1] Shafiq Ahmad, Gianluca Scarpellini, Pietro Morerio, and Alessio Del Bue, 'Event-driven Re-ID: A new benchmark and method towards privacy-preserving person re-identification', in *WACV*, pp. 459–468, (2022).

[2] Jihwan Bang, Heesu Kim, YoungJoon Yoo, Jung-Woo Ha, and Jonghyun Choi, 'Rainbow memory: Continual learning with a memory of diverse samples', in *CVPR*, pp. 8218–8227, (2021).

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel, 'Mixmatch: A holistic approach to semi-supervised learning', in *NeurIPS*, (2019).

[4] Hyuntak Cha, Jaeho Lee, and Jinwoo Shin, 'Co^{2}L: Contrastive continual learning', in *ICCV*, pp. 9516–9525, (2021).

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, 'A simple framework for contrastive learning of visual representations', in *ICML*, pp. 1597–1607, (2020).

[6] Afshin Dehghan, Shayan Modiri Assari, and Mubarak Shah, 'GMMCP tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking', in *CVPR*, pp. 4091–4099, (2015).

[7] Pedro Felzenszwalb, David McAllester, and Deva Ramanan, 'A discriminatively trained, multiscale, deformable part model', in *CVPR*, pp. 1–8, (2008).

[8] Jiawei Feng, Ancong Wu, and Wei-Shi Zheng, 'Shape-erased feature learning for visible-infrared person re-identification', in *CVPR*, pp. 22752–22761, (2023).

[9] Wenhang Ge, Junlong Du, Ancong Wu, Yuqiao Xian, Ke Yan, Feiyue Huang, and Wei-Shi Zheng, 'Lifelong person re-identification by pseudo task knowledge preservation', in *AAAI*, (2022).

[10] Wenhang Ge, Chunyan Pan, Ancong Wu, Hongwei Zheng, and Wei-Shi Zheng, 'Cross-camera feature prediction for intra-camera supervised person re-identification across distant scenes', in *ACM MM*, pp. 3644–3653, (2021).

[11] Heitor Murilo Gomes, Jesse Read, Albert Bifet, Jean Paul Barddal, and João Gama, 'Machine learning for streaming data: state of the art, challenges, and opportunities', *ACM SIGKDD*, **21**(2), 6–22, (2019).

[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 'Mask R-CNN', in *ICCV*, pp. 2961–2969, (2017).

[13] Zhipeng Huang, Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, and Zheng-jun Zha, 'Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation', in *CVPR*, pp. 14288–14297, (2022).

[14] Takashi Isobe, Dong Li, Lu Tian, Weihua Chen, Yi Shan, and Shengjin Wang, 'Towards discriminative representation learning for unsupervised person re-identification', in *ICCV*, pp. 8526–8536, (2021).

[15] Ziyu Jiang, Tianlong Chen, Ting Chen, and Zhangyang Wang, 'Improving contrastive learning on imbalanced data via open-world sampling', in *NeurIPS*, pp. 5997–6009, (2021).

[16] Minxian Li, Xiatian Zhu, and Shaogang Gong, 'Unsupervised person re-identification by deep learning tracklet association', in *ECCV*, pp. 737–753, (2018).

[17] Wei-Hong Li, Zhuowei Zhong, and Wei-Shi Zheng, 'One-pass person re-identification by sketch online discriminant analysis', *PR*, **93**, 237–250, (2019).

[18] Xinyu Lin, Jinxing Li, Zeyu Ma, Huafeng Li, Shuang Li, Kaixiong Xu, Guangming Lu, and David Zhang, 'Learning modal-invariant and temporal-memory for video-based visible-infrared person re-identification', in *CVPR*, pp. 20973–20982, (2022).

[19] Yichen Lu, Mei Wang, and Weihong Deng, 'Augmented geometric distillation for data-free incremental person ReID', in *CVPR*, pp. 7329–7338, (2022).

[20] Hao Luo, Wei Jiang, Youzhi Gu, Fuxu Liu, Xingyu Liao, Shenqi Lai, and Jianyang Gu, 'A strong baseline and batch normalization neck for deep person re-identification', *IEEE TMM*, **22**(10), 2597–2609, (2020).

[21] Andrea Maracani, Umberto Michieli, Marco Toldo, and Pietro Zanuttigh, 'Recall: Replay-based continual learning in semantic segmentation', in *ICCV*, pp. 7026–7035, (2021).

[22] Jingke Meng, Sheng Wu, and Wei-Shi Zheng, 'Weakly supervised person re-identification', in *CVPR*, pp. 760–769, (2019).

[23] Nan Pu, Wei Chen, Yu Liu, Erwin M. Bakker, and Michael S. Lew, 'Lifelong person re-identification via adaptive knowledge accumulation', in *CVPR*, (2021).

[24] Lei Qi, Lei Wang, Jing Huo, Yinghuan Shi, and Yang Gao, 'Progressive cross-camera soft-label learning for semi-supervised person re-identification', *IEEE TCSVT*, **30**(9), 2815–2829, (2020).

[25] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert, 'iCaRL: Incremental classifier and representation learning', in *CVPR*, pp. 2001–2010, (2017).

[26] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi, 'Performance measures and a data set for multi-target, multi-camera tracking', in *ECCV*, pp. 17–35, (2016).

[27] Jianren Wang, Xin Wang, Yue Shang-Guan, and Abhinav Gupta, 'Wanderlust: Online continual object detection in the real world', in *ICCV*, pp. 10829–10838, (2021).

[28] Tao Wang, Hong Liu, Pinhao Song, Tianyu Guo, and Wei Shi, 'Pose-guided feature disentangling for occluded person re-dentification based on transformer', in *AAAI*, pp. 2540–2549, (2022).

[29] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian, 'Person transfer GAN to bridge domain gap for person re-identification', in *CVPR*, pp. 79–88, (2018).

[30] Ancong Wu, Wenhang Ge, and Wei-Shi Zheng, 'Rewarded semi-supervised re-identification on identities rarely crossing camera views', *IEEE TPAMI*, (2023).

[31] Guile Wu and Shaogang Gong, 'Decentralised learning from independent multi-domain labels for person re-identification', in *AAAI*, pp. 2898–2906, (2021).

[32] Guile Wu and Shaogang Gong, 'Generalising without forgetting for lifelong person re-dentification', in *AAAI*, pp. 2889–2897, (2021).

[33] Jinlin Wu, Yang Yang, Hao Liu, Shengcai Liao, Zhen Lei, and Stan Z Li, 'Unsupervised graph association for person re-identification', in *ICCV*, pp. 8321–8330, (2019).

[34] Shiyu Xuan and Shiliang Zhang, 'Intra-inter camera similarity for unsupervised person re-identification', in *CVPR*, pp. 11926–11935, (2021).

[35] Jinrui Yang, Jiawei Zhang, Fufu Yu, Xinyang Jiang, Mengdan Zhang, Xing Sun, Yingcong Chen, and Wei-Shi Zheng, 'Learning to know where to see: A visibility-aware approach for occluded person re-identification', in *ICCV*, pp. 11865–11874, (2021).

[36] Qize Yang, Ancong Wu, and Wei-Shi Zheng, 'Person re-identification by contour sketch under moderate clothing change', *IEEE TPAMI*, **43**(6), 2029–2046, (2021).

[37] Yang Yang, Zhiying Cui, Junjie Xu, Changhong Zhong, Wei-Shi Zheng, and Ruixuan Wang, 'Continual learning with bayesian model based on a fixed pre-trained feature extractor', *Visual Intelligence*, **1**(1), 5, (2023).

[38] Hongxu Yin, Pavlo Molchanov, Jose M Alvarez, Zhizhong Li, Arun Mallya, Derek Hoiem, Niraj K Jha, and Jan Kautz, 'Dreaming to distill: Data-free knowledge transfer via deepinversion', in *CVPR*, pp. 8715–8724, (2020).

[39] Jiahang Yin, Ancong Wu, and Wei-Shi Zheng, 'Fine-grained person re-identification', *IJCV*, **128**(6), 1654–1672, (2020).

[40] Bowen Zhao, Yingjiu Li, Ximeng Liu, Hwee Hua Pang, and Robert H Deng, 'FREED: An efficient privacy-preserving solution for person re-identification', in *CDSC*, pp. 1–8, (2022).

[41] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian, 'MARS: A video benchmark for large-scale person re-identification', in *ECCV*, pp. 868–884, (2016).

[42] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian, 'Scalable person re-identification: A benchmark', in *ICCV*, pp. 1116–1124, (2015).

[43] Xiangping Zhu, Xiatian Zhu, Minxian Li, Pietro Morerio, Vittorio Murino, and Shaogang Gong, 'Intra-camera supervised person re-identification', *IJCV*, **129**, 1580–1595, (2021).