International Journal of Computer Vision
https://doi.org/10.1007/s11263-024-02110-x

Exemplar-Free Lifelong Person Re-identification via Prompt-Guided Adaptive Knowledge Consolidation

Qiwei Li[1,2] · Kunlun Xu[1,2] · Yuxin Peng[1,2] · Jiahuan Zhou[1,2]

Received: 20 September 2023 / Accepted: 29 April 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Lifelong person re-identification (LReID) refers to matching people across different cameras given continuous data streams. The challenge of catastrophic forgetting of old knowledge and the effective acquisition of new knowledge form a significant dilemma for LReID. Most current LReID methods propose to retain abundant exemplars from historical data, which are further rehearsed to fully fine-tune the whole model. However, such a learning paradigm will inevitably hinder data privacy and result in substantial computation costs. In this paper, we propose a paradigm for exemplar-free LReID through model re-parameterization. Without retaining any exemplars, our designed method adopts a novel Prompt-guided Adaptive Exponential Moving Average (PAEMA) strategy to achieve dynamic knowledge consolidation. Our key idea is to leverage visual prompting as the guidance for model re-parameterization to benefit knowledge preservation. Conventional Exponential Moving Average (EMA) methods rely on fixed or time-varied constants as weighting parameters, the unpredictable correlation between new and old data streams may lead to varying levels of model parameter drifting during LReID learning. Hence, we argue that a proper weighting parameter should be conditioned on the variation of new and old models to provide an adaptive knowledge consolidation for LReID. To do so, an adaptive mechanism is proposed to utilize the visual prompt as a surrogate for model variation estimation. Consequently, without using any exemplars, the forgetting issue in LReID is greatly alleviated. Experiments on various LReID benchmarks have verified the superiority of our method against the state-of-the-art lifelong learning and LReID approaches. Code is available at https://github.com/zhoujiahuan1991/IJCV2024-PAEMA/.

Keywords Lifelong person re-identification · Prompt learning · Exemplar-free · Adaptive knowledge consolidation

1 Introduction

Person re-identification (ReID), aiming to retrieve the same people across different camera views, has played a crucial role in many computer vision tasks (Luo et al., 2019; Zhang et al., 2019; Wang et al., 2021). Though recent ReID methods have achieved promising performance, most of them assume that all the training data can be accessed at once (as shown in Fig. 1a), and their performance drops dramatically in a practical scenario where the new training data come continually (Zhang et al., 2022; Huang et al., 2022). Thus, existing ReID models need to be incrementally updated from sequential learning of intermittent new data, and the task of Lifelong Person ReID (LReID) has attracted increasing attention recently (as shown in Fig. 1b).

Similar to other lifelong learning-based tasks (Rebuffi et al., 2017; Wang et al., 2023; Kalb & Beyerer, 2023; Liu et al., 2023), the main challenge of LReID is the catastrophic forgetting of the knowledge learned from old datasets. To mitigate this issue, most recent works (Wu & Gong, 2021; Ge et al., 2022; Huang et al., 2022; Yu et al., 2023) propose to address it by retaining exemplars of learned tasks and emphasizing the feature consistency of exemplars between the current and previous models (as shown in Fig. 1c). How-

Communicated by Jingdong Wang.

✉ Jiahuan Zhou
jiahuanzhou@pku.edu.cn

Qiwei Li
lqw@pku.edu.cn

Kunlun Xu
xukunlun2023@gmail.com

Yuxin Peng
pengyuxin@pku.edu.cn

[1] Wangxuan Institute of Computer Technology, Peking University, Beijing 100871, China

[2] National Key Laboratory for Multimedia Information Processing, Peking University, Beijing 100871, China

Published online: 27 May 2024

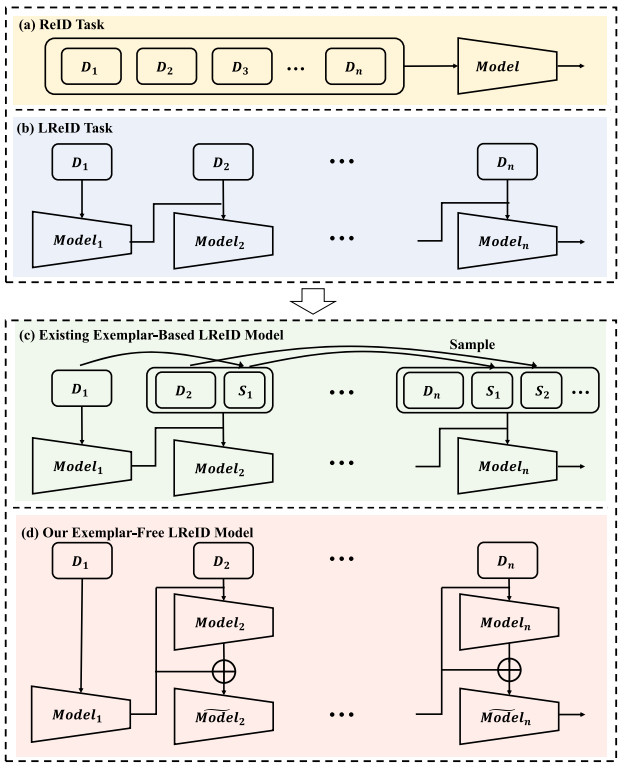

Fig. 1 The comparison between the general ReID (**a**) and LReID tasks (**b**), as well as the existing exemplar-based LReID methods (**c**) and our proposed exemplar-free model (**d**)

ever, such a learning paradigm will not only inevitably hinder data privacy, but also result in substantial storage and training consumption with the increase of exemplar amounts over time. Although a few latest works (Pu et al., 2021; Sun & Mu, 2022; Pu et al., 2022) have begun to investigate the exemplar-free LReID scenario, their performance is severely limited due to the lack of effective anti-forgetting designs.

Therefore, in this paper, we focus on such a more challenging scenario in LReID that *no exemplar is allowed to retain for learning*. In this setting, the catastrophic forgetting issue of historical data streams will be even exaggerated since there is no explicit prior knowledge provided. To tackle it, we design a novel re-parameterization-based learning paradigm that a **P**rompt-guided **A**daptive **E**xponential **M**oving **A**verage (PAEMA) strategy is adopted to achieve dynamic knowledge consolidation without exemplars (as shown in Fig. 1d). Different from the existing EMA methods that usually utilize a fixed or time-varied constant as the balancing parameter (Cai et al., 2021; Lin et al., 2022; Yu et al., 2023), we argue that a proper balancing parameter should be conditioned on the variation of new and old models. This is because the correlation between new and old data streams is always unpredictable and may lead to different levels of model parameter drifting during LReID.

To achieve this, we propose a novel adaptive mechanism to leverage the learnable prompts as a surrogate for model variation prediction. Based on a ViT (Dosovitskiy et al., 2020) backbone, multiple learnable prompts are learned for each multi-head self-attention (MSA) layer to encode the knowledge of the new data. After training on the new data, these prompts can automatically predict the variety of models with a balancing parameter for PAEMA. By adaptively fusing the models at different learning stages, our approach can readily achieve a better balance of new knowledge acquisition and old knowledge forgetting. Even if no exemplars are retained, the catastrophic forgetting issue in LReID is greatly alleviated. In summary, the main contributions of this paper are three-fold:

(1) To tackle a more challenging exemplar-free scenario in LReID, we propose a prompt-integrated ViT-based LReID model along with a novel adaptive model re-parameterization algorithm. (2) To mitigate the catastrophic forgetting issue, a novel Prompt-guided Adaptive Exponential Moving Average (PAEMA) strategy is proposed to achieve dynamic knowledge consolidation for LReID. (3) Even without retaining any old exemplars, extensive experiments on various LReID benchmarks have demonstrated our superiority against the existing exemplar-based state-of-the-art LReID approaches.

2 Related Work

2.1 Person Re-identification

Person re-identification (ReID) aims to retrieve the person of interest from the given gallery set (Ahmed et al., 2015; Li et al., 2018; Luo et al., 2019). Owing to the collection of tremendous labeled data, supervised ReID methods (Zhuang et al., 2020; He et al., 2021; Chen et al., 2017) have achieved remarkable performance on various benchmarks but suffered severely from the heavy annotation bottleneck and poor generalization ability across different datasets (Liao & Shao, 2022; Ni et al., 2022). The main reason is that the domain gap inhibits these well-trained models from well handling different datasets (Liu et al., 2019; Song et al., 2019; Jin et al., 2020). To settle this problem, recently, various unsupervised learning ReID approaches (Yu et al., 2019; Wang & Zhang, 2020; Zheng et al., 2021; Lin et al., 2019; Isobe et al., 2021; Cho et al., 2022) are proposed which generalize from the labeled source domain to the unlabeled target domain. These methods assume that all training data are available beforehand and neglect the factor that, in real scenarios, the training data may not be available at once but come sequentially.

2.2 Lifelong learning

In the area of computer vision, deep learning methods have shown remarkable capabilities, often surpassing human performance when applied to fixed datasets. However, when confronted with a continuous stream of training data instead of a static dataset, certain methods may falter due to the evolving nature of knowledge. To address this challenge, lifelong learning, also named continual learning or incremental learning, was introduced (Rebuffi et al., 2017; Rannen et al., 2017), aiming to strike a balance between the acquisition of new knowledge and the forgetting of previously learned knowledge. Existing lifelong learning methods can be grouped into three categories: exemplar-based methods, architecture-based methods, and regularization-based methods. Exemplar-based methods (Prabhu et al., 2020; Liu et al., 2021; Luo et al., 2023) maintain a small set of previous data to replay and reinforce prior knowledge. However, these methods often raise concerns about data privacy since they require storing previous data. Architecture-based methods (Wang et al., 2022a, b; Zhou et al., 2022; Hu et al., 2023) design specific parameters for each lifelong learning step and dynamically expand the models to retain previous knowledge. However, they suffer from continuous increases in parameter numbers, leading to substantial storage and training consumption. Regularization-based methods (Kirkpatrick et al., 2017; Li & Hoiem, 2017; Tung & Mori, 2019; Sun et al., 2023) adopt knowledge distillation to transfer knowledge from the old model to the new model and regularize the model change, thus mitigating knowledge forgetting. Unfortunately, most of these methods primarily focus on distilling knowledge from the logit predictions, which may not perform optimally in scenarios like person re-identification, where datasets may encompass thousands of distinct individuals.

2.3 Lifelong Person Re-identification

Recently, lifelong person re-identification (LReID) (Wu & Gong, 2021; Ge et al., 2022; Pu et al., 2022; Yu et al., 2023) has become more notable owing to its importance in handling realistic data streams. Simultaneously, the classical challenge in lifelong learning, catastrophic forgetting (Li & Hoiem, 2017; Shmelkov et al., 2017) also occurs in LReID. To alleviate this issue, Wu and Gong (2021) designed a scheme to simultaneously maintain the classification, representation, and distribution coherence between the old and new models. Pu et al. (2021) focused on old knowledge accumulation in order to propagate the previously learned knowledge to new domains. Thus, a similarity graph and a knowledge graph were designed for new knowledge acquisition and old knowledge accumulation. Sun and Mu (2022) proposed a differentiable patch sampler to adaptively select discriminative patches for coherence learning, trying to alleviate the influence of large data distribution discrepancy between old and new data. Ge et al. (2022) formulated LReID as a domain adaption problem and proposed a pseudo-task transformation to minimize the domain gap between different datasets. Yu et al. (2023) proposed a rehearsal and refreshing method to achieve both positive forward and backward knowledge transfer.

However, most of the aforementioned LReID methods have to retain sufficient exemplars to alleviate forgetting. Thus, they always suffer from severe data privacy disclosure and heavy storage costs. Even worse, when tackling privacy-sensitive tasks, no exemplars are allowed to be retained which causes them to fail completely. Although few works (Pu et al., 2021; Sun & Mu, 2022) do not need to keep exemplars, their performance is largely deteriorated compared with the exemplar-based ones. In this paper, without retaining any old data, we propose a novel exemplar-free LReID method that can consistently outperform those exemplar-based methods.

2.4 Prompting in Lifelong Learning

Prompts are firstly designed in natural language processing (NLP) (Houlsby et al., 2019) which pretend instructions to the input so that the pre-trained model can obtain information on downstream tasks. Although some recent works (Petroni et al., 2019; Cui et al., 2021) demonstrated that manually designed prompts can improve the generalization ability of models, designing prompts needs specific domain knowledge which is indeed difficult. Recently, prompt tuning (PT) (Lester et al., 2021) has been studied by regarding prompts as learnable parameters. In the field of computer vision, VPT (Jia et al., 2022) adapted prompt tuning to learn task-specific tokens for the encoder layers in a vision transformer (ViT) (Dosovitskiy et al., 2020). This is because most visual prompting methods assume that the pretrained ViT model is strong enough to extract discriminative features and introducing minor learnable parameters can adapt the model to the new tasks (Wang et al., 2022a). Recently, Wang et al. (2022b); Douillard et al. (2022); Wang et al. (2022); Smith et al. (2023) combined prompt learning with lifelong learning and exhibited promising results in the task of classification. Among them, Wang et al. (2022b) optimized a pool of prompts among tasks and selected the most similar prompts for inference, and (Wang et al., 2022; Douillard et al., 2022; Wang et al., 2022a; Smith et al., 2023) proposed task-independent prompts to mitigate the influence of different tasks.

In our method, we further leverage prompts from a different but important perspective. Besides utilizing prompts to benefit new knowledge learning, we also treat them as a surrogate for modeling knowledge variation during lifelong learning. A prompt-guided balancing parameter can be read-

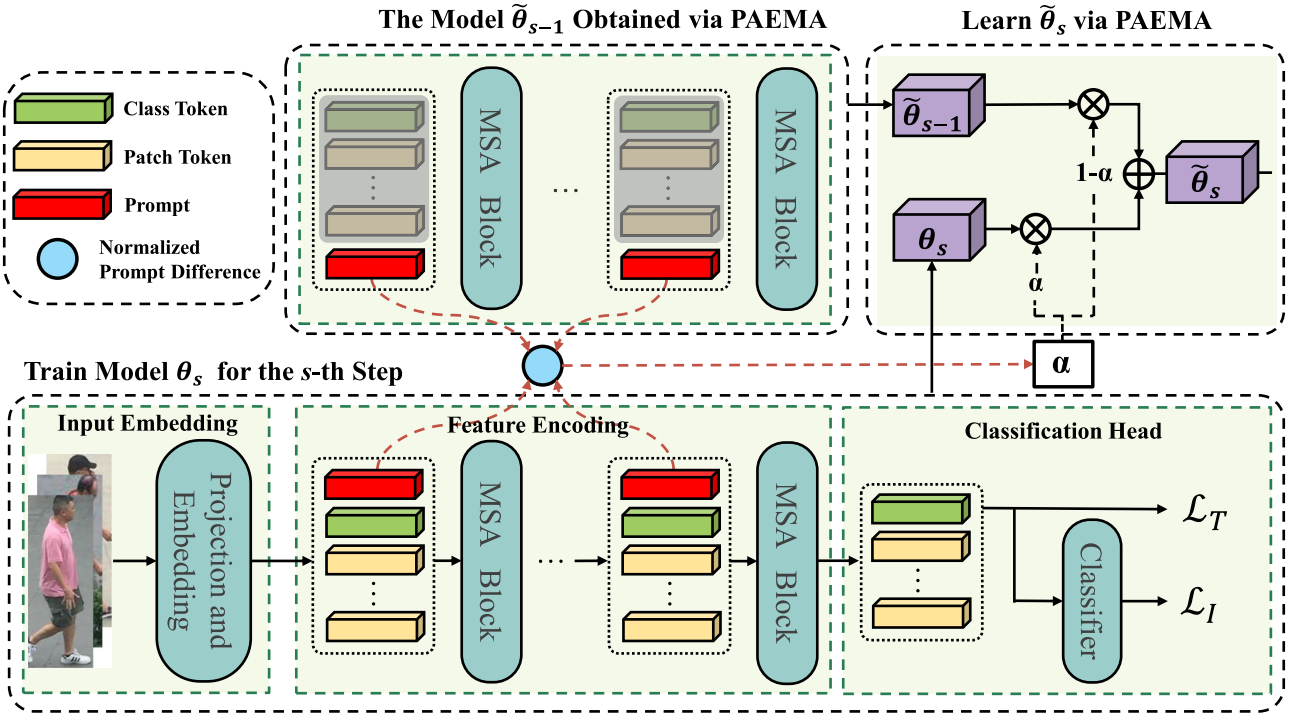

Fig. 2 The overall pipeline of our proposed PAEMA LReID model. For the s-th training dataset, the obtained model θ_s is further fused with the PAEMA model $\widetilde{\theta}_{s-1}$ at the $s-1$ training step via a prompt-guided adaptive balancing parameter α. The gray mask means these tokens are not generated

ily obtained for PAEMA to mitigate the forgetting issue in LReID.

3 The Proposed Method

3.1 Problem Formulation

In the task of LReID, a stream of S datasets denoted as $\mathcal{D} = \{D^s\}_{s=1}^{S}$ are collected in sequence to incrementally train the model. Each dataset D^s has a training set D^s_{train} and a testing set D^s_{test} which have no overlapping identities. In our method, we assume the model can not access the data from previous training steps. At step s, the model can only utilize the training set D^s_{train} for learning. For testing, all the test sets $\{D^i_{test}\}_{i=1}^{S}$ are utilized to evaluate the anti-forgetting and acquisition ability of the LReID models. Besides, to further verify the generalization ability of different LReID models, a set of U unseen datasets $\mathcal{D}_U = \{D^i_u\}_{i=1}^{U}$ is directly evaluated using the obtained LReID models trained on $\mathcal{D}$.

3.2 Overview of PAEMA

As illustrated in Fig. 2, we propose a prompt-guide anti-forgetting model for exemplar-free LReID. The whole model consists of a ViT-based LReID backbone for feature extraction (Sect. 3.3) and a Prompt-guided Adaptive Exponential Moving Average (PAEMA) algorithm to dynamically consolidate learned knowledge (Sect. 3.5). The model trained at step s is denoted as $\theta_s = \theta^c_s \circ \theta^a_s \circ \theta^e_s$, where θ_s, θ^e_s, θ^a_s, and θ^c_s represent the whole model, the input embedding module, the feature encoding module, and the classification head module respectively.

During training, our model is initially trained on the first training set and the learned model parameter is θ_1 which also serves as the re-parameterized model parameter of $\widetilde{\theta}_1$ directly. At the following training step $s \geq 2$, the re-parameterized model parameter $\widetilde{\theta}_{s-1}$ at the (s-1)-th step is adopted as the initial parameter for step s and fine-tuned on the new training set D^s_{train} to learn θ_s. Once θ_s is obtained, the proposed PAEMA algorithm is explored to adapt it to $\widetilde{\theta}_s$ via adaptively fusing $\widetilde{\theta}_{s-1}$ and θ_s for a better balance between new knowledge acquisition and old knowledge forgetting. The core is to leverage a set of globally-shared prompts to depict the model knowledge variation which can act as guidance for automatically balancing $\widetilde{\theta}_{s-1}$ and θ_s. To be noticed, throughout the whole LReID learning procedure in our work, no data from the previous training datasets in any form are preserved.

3.3 A ViT-based LReID Backbone

Currently, Vision Transformer (ViT) (Dosovitskiy et al., 2020) has shown overwhelming performance in many computer vision areas. A recent work (He et al., 2021) illustrated the effectiveness of pure-transformer models in tackling the ReID task. Inspired by He et al. (2021), we first build a ViT-based backbone for discriminative representation learning of LReID. For simplicity, we omit the dataset index s in this section.

3.3.1 Feature Extractor

In the input embedding module θ^e, an input training image $x \in \mathbb{R}^{H\times W\times C}$ is first split into L non-overlapped patches and flattened into vectors $\{x_p^i\}_{i=1}^L \in \mathbb{R}^{M^2C}$, where H, W, C represent image height, width, and the number of channels respectively, L is the number of patches and M is the patch size. Then the input embedding layer θ^e maps these vectors into a set of d-dimension patch embedding, each of which serves as a patch token. A learnable class token x_{cls} is concatenated to the patch tokens and its corresponding output token serves as the global feature. The position encoding $P \in \mathbb{R}^{(L+1)d}$ is also added to tokens to encode the spatial configuration. Thus, the input x is re-formed as:

$$x^* = [x_{cls}, \theta^e(x_p^1), \theta^e(x_p^2), ..., \theta^e(x_p^L)] + P. \tag{1}$$

Then x^* is fed into the feature encoding model θ^a which consists of N consecutive encoder blocks (Vaswani et al., 2017). The key part of each block is the multi-head self-attention (MSA) layer:

$$\begin{aligned} \mathrm{MSA}(Q, K, V) &= \mathrm{Concat}(h_1, ..., h_m)W^O \\ h_i &= \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{2}$$

where Q, K, and V are input query, key, and value to the MSA layer, W^O, W_i^Q, W_i^K, and W_i^V are the projection matrices, h_i is the i-th attention head, and m is the number of heads. In our model, following (Dosovitskiy et al., 2020), we set Q=K=V=x^* as the inputs for the first MSA layer. Therefore, the output feature embedding of the feature extractor can be formulated as:

$$f_x = \theta^a(x^*) \tag{3}$$

3.3.2 Classifier and Optimization

Though ReID models usually regard the ReID task as a person ranking problem and take feature vectors as outputs instead of person identities, Luo et al. (2019) has found that an identity classification loss can separate the embedding spaces into different subspaces to facilitate the ReID training. Therefore, in our LReID model, we train a dataset-specific classification head θ^c to predict probabilities from the corresponding output feature $f_{cls,x} \in f_x$ of the class token x_{cls}. Then, the predicted probabilities are supervised by:

$$\mathcal{L}_{ID} = \sum_{x\in\mathcal{B}} L_{CE}\left(y, \theta^c\left(f_{cls,x}\right)\right) \tag{4}$$

where $\mathcal{B}$ is a training batch with B samples. $\left(f_{cls,x}, y\right)$ are the input feature and its class label. L_{CE} is the cross-entropy loss without label smoothing. Besides, we also adopt the triplet loss with a soft-margin to enhance intra-class compactness and inter-class separability, which can be formulated as:

$$\mathcal{L}_T = \sum_{x\in\mathcal{B}} \log\Big[1 + \exp\big(\|f_x - f_p\|_2^2 - \|f_x - f_n\|_2^2\big)\Big], \tag{5}$$

where (f_p, f_n) are features of positive and negative samples for x respectively (He et al., 2021).

Thus, the above ViT-based LReID backbone is optimized via a combined loss consisting of $\mathcal{L}_{ID}$ and $\mathcal{L}_T$:

$$\mathcal{L} = \mathcal{L}_I + \mathcal{L}_T \tag{6}$$

3.4 EMA: Exponential Moving Average

In LReID, when fine-tuning the model parameter θ_s using the s-th training dataset, the model might be forced to adapt to the new data distribution of D_{train}^s. However, without retaining the exemplars from the previous s-1 datasets, the fine-tuned model θ_s will result in serious forgetting of historical knowledge. A straightforward way to bring back the lost knowledge is by directly fusing the fine-tuned model θ_s and the old model $\widetilde{\theta}_{s-1}$, which is known as Exponential Moving Average (EMA) (Cai et al., 2021; Xu et al., 2021):

$$\widetilde{\theta}_s = (1-\alpha)\widetilde{\theta}_{s-1} + \alpha\theta_s, \tag{7}$$

where α is the balancing parameter that controls the proportion of θ_s and $\widetilde{\theta}_{s-1}$ in $\widetilde{\theta}_s$.

Here we would like to investigate why EMA could benefit lifelong learning. Intuitively, the adopted optimization loss Eq. 6 aims to guide the model to distinguish different people of new distribution perfectly, even to an extent, generate excess distance between different people. Motivated by Sankaranarayanan et al. (2017); Wang et al. (2018), a certain degree of disturbance of the fine-tuned model might not cause serious performance degradation on the new data domain, while the recovered old knowledge by the model addition could contribute to the old data domains substantially.

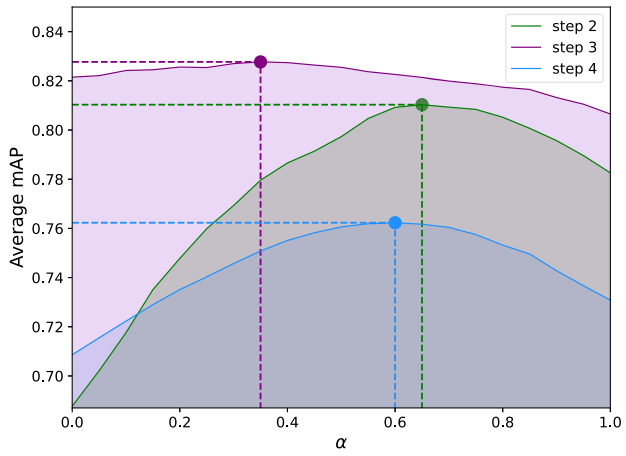

Fig. 3 A verification experiment to demonstrate the relationship between α and mAP at different learning steps in LReID

3.5 PAEMA: Prompt-Guided Adaptive EMA

Generally, conventional EMA (Lin et al., 2022) usually adopt a fixed constant α or a time-varied value $\alpha = 1/s$ in Eq. 7. However, we empirically observe that the proper choices of α vary significantly at different LReID steps, distinct from fixed constant or $1/s$. As illustrated in Fig. 3, we perform lifelong learning on four ReID datasets, where after each training step, different α from 0 to 1 are sampled to obtain the best value of α. The optimal values of α are 0.64, 0.35, and 0.60 at 2∼4-th training step. Therefore, how to adaptively determine the α for different training datasets is necessary and important to maximize the benefit brought by EMA. However, due to the nonlinearity and complexity of large-scale deep models, there is no direct correlation between the changes in output features and model parameters, forming a significant challenge to the adaptive determination of α.

Thus we propose a novel prompt-guided knowledge shifting estimation scheme, PAEMA, which not only utilizes prompts to capture the information of certain tasks but also exploits prompts as a surrogate for knowledge variation estimation between different tasks. As a set of parameters in the form of tokens, prompts have an intrinsic relationship with the extracted features of input samples that share the same token form. Therefore, besides instructing the model to learn task-specific knowledge as commonly used by previous work (Wang et al., 2022a, b), we build globally-shared prompts optimized for every task to investigate the relationships between the change of prompt parameters and knowledge shifting.

To do so, learnable prompts are added to MSA layers in θ^a instead of only to the embedding feature of input image (Wang et al., 2022b; Douillard et al., 2022). Based on the function of the MSA layer in Eq. 2, our method learns two prompt $p_K, p_V \in \mathbb{R}^{L_p \times D}$ and concatenates them to K and V respectively:

$$\mathrm{MSA}(Q, [p_K; K], [p_V; V]). \tag{8}$$

To estimate a proper α in Eq. 7, we denote N as the number of total prompts and p_s^n as the n-th prompt for the model of step s. N is calculated by $N = 2 \times L_p \times L_e$, where L_e represents the number of MSA layers and for each MSA layer, L_p Key prompts and L_p Value prompts are contained. The difference of the prompts learned in step s and $s-1$ could be simply represented as $p_s^n - p_{s-1}^n \in \mathbb{R}^D$. Then, we propose a Normalized Prompt Difference formula which directly take prompt parameters as input to map all prompt differences in step s to a finite scalar Δp_s:

$$\Delta p_s = \frac{1}{N} \sum_{n=1}^{N} \frac{2\|p_s^n - p_{s-1}^n\|_1}{\|p_s^n\|_1 + \|p_{s-1}^n\|_1}. \tag{9}$$

Since the supervised training of the deep network always leads to massively over-parameterization on the current dataset D^s, there is a boundary of tolerance ΔB_s for new model disturbance, within which the performance on the domain of D^s will change slightly. According to Eq. 7, let $\Delta B_s = \widetilde{\theta}_s - \theta_s$ and we obtain $(1-\alpha) \cdot (\widetilde{\theta}_{s-1} - \theta_s) = \Delta B_s$ which means that α and $\widetilde{\theta}_{s-1} - \theta_s$ have positive correlation. This illustrates that the larger the knowledge shifting is, the closer the boundary should be to the new model. Therefore, the balancing parameter α can be calculated as:

$$\alpha = \Delta p_s. \tag{10}$$

An overview of our proposed PAEMA is presented in Algorithm 1.

Algorithm 1 PAEMA training procedure.

Input: Data stream $\mathcal{D} = \{D^s\}_{s=1}^S$, initial model $\widetilde{\theta}_0$.
Output: Final model $\widetilde{\theta}_S$.

```
for s ← 1 to S do
    Initialise θ_s as θ̃_{s−1};
    Train θ_s with L_I + L_T (Eq. 6);
    if s = 1 then
        θ̃_s ← θ_s
    else
        Calculate α with (Eq. 9 and Eq. 10);
        Calculate θ̃_s ← (1 − α)θ̃_{s−1} + αθ_s (Eq. 7).
    end if
end for
```

In summary, our proposed PAEMA is designed to estimate the knowledge shifting by the variations of prompt parameters to instruct the EMA for a better balance of forgetting and acquisition. For the use of prompts, we add learnable prompts to all MSA layers for the following three reasons.

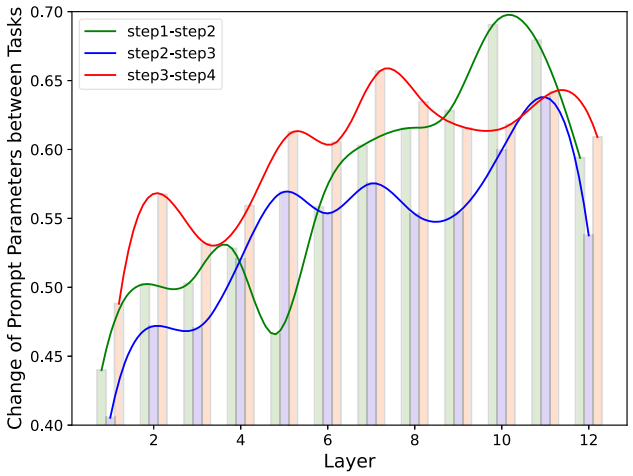

Fig. 4 Visualization of model changes between adjacent steps at different layers. A verification experiment is conducted to demonstrate that the prompts added to different layers exhibit significant variations of parameter changes according to Eq. 9

Firstly, the gradient vanishing phenomenon is still inevitable in current advanced deep networks (Glorot & Bengio, 2010; He et al., 2016), measuring the change of models by only a few layers of prompts, especially the early layers, is not accurate enough. Secondly, as illustrated in Fig. 4, prompts from different layers have different changing scales. Solely considering partial layers may result in biased knowledge-shifting estimation (Table 2). At last, different layers propose to learn different knowledge, e.g., in the lower layers, local structural information is always learned, but in the higher layers, the high-level semantic information can be better modeled. Thus, to better estimate the model variation, all layers should be considered comprehensively which are verified by the ablation experiments in Table 8.

4 Experiments

4.1 Experimental Settings

4.1.1 Benchmarks

There are two streams of benchmarks, i.e., GwFReID benchmark and AKA benchmark, in LReID proposed by GwFReID (Wu & Gong, 2021) and AKA (Pu et al., 2021) respectively. **GwFReID benchmark** adopts four widely-used datasets as the training set with the showing order of Market-1501 (Zheng et al., 2015), DukeMTMC (Ristani et al., 2016), CUHK-SYSU (Xiao et al., 2016), and MSMT17_v1 (Wei et al., 2018), and evaluate the generalization of methods on another four test sets, i.e., CUHK01 (Li et al., 2012), CUHK03 (Li et al., 2014), GRID (Loy et al., 2010), and SenseReID (Zhao et al., 2017). **AKA benchmark** consists of twelve datasets, five of which are chosen as the training sets, i.e., Market-1501, DukeMTMC, CUHK-SYSU, MSMT17_V2 (Wei et al., 2018), and CUHK03 (Li et al., 2014), Seven of which are used as test-only sets to evaluate the generalization on unseen domains, i.e., CUHK01, CUHK02 (Li & Wang, 2013), VIPeR (Gray & Tao, 2008), PRID (Hirzer et al., 2011), i-LIDS (Zheng et al., 2009), GRID, and SenseReID. Note that the GwFReID benchmark provides all the person identities and images of the training datasets, whereas the AKA benchmark selects 500 identities and corresponding images from each training dataset. Besides, AKA benchmark has two classical training orders, i.e. Market-1501 → CUHK-SYSU → DukeMTMC → MSMT17_v2 → CUHK03, named **AKA-Order-1**, and DukeMTMC → MSMT17_v2 → Market-1501 → CUHK-SYSU → CUHK03, named **AKA-Order-2**. The detailed statistics for these datasets in our experiments are provided in Table 1.

Table 1 The statistics of person re-identification datasets in our experiments

Dataset	Scale	Person identities		
		Train	Query	Gallery
Market-1501	Large	751	750	751
DukeMTMC	Large	702	702	1110
CUHK-SYSU	Mid	942	2900	2900
MSMT_v1	Large	1041	3060	3060
MSMT_v2	Large	1041	3060	3060
CUHK03	Mid	767	700	700
CUHK02	Mid	1577	239	239
CUHK01	Small	485	486	486
GRID	Small	125	125	126
SenseReID	Mid	1718	521	1718
VIPeR	Small	316	316	316
PRID	Small	100	100	649
i-LIDS	Small	59	60	60

4.1.2 Evaluation Metrics

The widely adopted mean Average Precision (mAP) and Rank-1 (R@1) accuracy are explored in our experiments to evaluate each individual dataset. To further verify the lifelong learning capacity of models, we calculate the average of mAP and R@1 results on all datasets as an overall evaluation metric. We also evaluate the Average Forgetting (Chaudhry et al., 2018) which represents the average performance degradation compared to the trained step on each seen domain. Specifically, the forgetting of j-th stage after training on t-th stage can be computed as:

Table 2 Comparison with SOTA lifelong learning and LReID methods on seen datasets of GwFReID benchmark

		Train: Market-1501 → DukeMTMC → CUHK-SYSU → MSMT17_v1									
		Market-1501		DukeMTMC		CUHK-SYSU		MSMT17_v1		Average	
Methods	E	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
Joint-Train	×	92.7	83.2	86.4	76.3	94.6	93.7	73.5	53.0	86.8	76.5
GwFReID	✓	81.6	60.9	66.5	46.7	83.9	81.4	52.4	25.9	71.1	53.7
PTKP	✓	90.1	77.0	78.0	63.8	90.1	88.4	67.5	41.9	81.4	67.6
PTKP-ViT‡	✓	**90.3**	**78.8**	81.7	68.7	92.3	91.1	72.4	48.9	84.2	71.9
KRKC‡	✓	81.6	61.0	73.3	59.3	90.3	88.7	68.3	43.9	78.4	63.2
KRKC-ViT‡	✓	84.9	68.0	75.1	61.5	92.1	91.0	65.1	40.5	79.3	65.3
PAEMA	✓	90.0	78.4	**82.1**	**71.0**	**93.6**	**92.5**	**76.8**	**55.8**	**85.6**	**74.4**
LwF‡	×	82.0	63.6	72.4	60.5	92.2	90.7	61.2	38.9	77.0	63.4
SPD‡	×	82.0	63.5	72.8	60.5	92.0	90.7	71.8	49.6	79.6	66.1
CRL‡	×	83.0	66.5	73.0	60.9	92.2	90.7	61.0	39.6	77.3	64.4
L2P‡	×	84.9	67.9	46.2	30.0	84.3	82.0	27.1	11.1	60.6	47.7
Dualprompt‡	×	84.3	63.1	66.8	48.9	72.2	66.6	28.1	9.8	62.8	47.1
AKA‡	×	74.8	49.4	55.0	38.1	83.8	81.1	52.4	29.9	66.3	49.6
PatchKD‡	×	**90.2**	**75.9**	61.2	44.8	83.9	82.2	33.1	16.0	67.1	54.7
PTKP‡	×	73.5	47.4	67.1	47.4	77.9	75.2	65.1	37.4	70.9	51.9
PTKP-ViT‡	×	78.4	54.4	68.9	50.1	90.0	88.1	63.9	37.8	75.3	57.6
NAPA-VQ‡	×	82.0	63.2	73.2	60.9	91.3	89.7	69.7	48.3	79.0	65.5
PRAKA‡	×	84.3	67.8	74.7	63.4	91.6	90.2	71.1	49.3	80.4	67.7
PAEMA	×	87.3	71.4	**78.6**	**67.7**	**93.6**	**92.6**	**73.5**	**51.0**	**83.2**	**70.7**

The bold numbers represent the best results
E means that sampled images of historical datasets are retained as exemplars and are replayed when new data comes
‡ represents that we reproduce the results based on their code
PAEMA with exemplars means adopting the same sampling strategy as PTKP (Ge et al., 2022)

$$f_j^t = \max_{l\in\{1,\ldots,t-1\}} a_{l,j} - a_{t,j} \tag{11}$$

where $a_{t,j}$ represents the results of task j, after training the model from 1 to t. Then the Average Forgetting at stage t can be computed as:

$$F_t = \frac{1}{t-1}\sum_{j=1}^{t-1} f_j^t \tag{12}$$

4.2 Implementation Details

Following (He et al., 2021), ViT pre-trained on ImageNet-21k (Deng et al., 2009) and then fine-tuned on ImageNet-1k (Deng et al., 2009) is utilized to initialize the ViT-based LReID backbone. The parameters of the ViT backbone are optimized by an SGD (Amari, 1993) optimizer with a momentum of 0.9, a weight decay of 1e-4, and a learning rate of 8e-3. The number of MSA layers is $N = 12$. The prompts in our model are initialized by a uniform distribution between 0 and 1, then optimized by an Adam (Kingma & Ba, 2014) optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 5e-3. Note that the optimizer configurations of the ViT backbone and prompts are kept consistent with the recent works (He et al., 2021; Zhu et al., 2022). For training data, all images are resized to 256×128 and augmented with random horizontal flipping, padding, random cropping, normalization, and random erasing. The batch size B is set to 64. For the first dataset, we train our model for 80 epochs in total, and the learning rate decays by 0.1 after the 40th and 70th epochs. For the later datasets, we train the network for 60 epochs in total, and the learning rate decays by 0.1 after the 30th epoch. The proposed method is implemented with Pytorch and trained on a single NVIDIA 4090 GPU.

4.3 Comparison with the State-of-the-art (SOTA)

4.3.1 Comparison Methods

In our experiments, five lifelong learning methods (LwF (Li & Hoiem, 2017), SPD (Tung & Mori, 2019), CRL (Zhao et al., 2021)), NAPA-VQ (Malepathirana et al., 2023), PRAKA (Shi & Ye, 2023), latest prompt based lifelong learning method (L2P (Wang et al., 2022b), Dualprompt (Wang

et al., 2022a)) and the LReID approaches (GwFReID (Wu & Gong, 2021), AKA (Pu et al., 2021), PatchKD (Sun & Mu, 2022), PTKP (Ge et al., 2022), KRKC (Yu et al., 2023)) are compared with our PAEMA.

Considering that our prompt parameters are specifically designed for ViT, we readily choose ViT as our backbone. Besides, we also replace the ResNet-50 backbone in several important methods (e.g., LwF, SPD, CRL, NAPA-VQ, PRAKA, PTKP, KRKC) with our adopted ViT-based backbone for fair comparisons with the official configuration in the original paper. Moreover, a special experimental setting Joint-Train is also conducted on our baseline, which denotes collecting the data from all the training datasets together to train the ReID model at once. Thus, Joint-Train is commonly regarded as the upper bound of LReID models.

4.3.2 Results on Seen Datasets of GwFReID Benchmark

The comparison results with the traditional lifelong learning methods and state-of-the-art (SOTA) LReID methods are reported in Table 6 where the symbol E denotes exemplars are used during LReID training.

4.3.2.1 PAEMA w/o Exemplar vs. SOTA w/o Exemplar Compared with the LReID approach without using exemplar, AKA and PatchKD, our proposed PAEMA outperforms the best player by 17.4%/22.9%, 9.7%/10.4%, and 21.1%/21.1% of R@1/mAP on the last three datasets. Compared with the latest PTKP approach, without using any exemplars, our proposed PAEMA can significantly outperform them (PTKP, PTKP-ViT) by at least 8.9%/17.0%, 9.7%/17.6%, 3.6%/4.5%, and 8.4%/13.2% of R@1/mAP on all four datasets respectively. It can be attributed to the anti-forgetting ability of our proposed prompt-guided adaptive knowledge consolidation model. Compared to lifelong learning methods (LwF, SPD, CRL, NAPA-VQ, PRAKA, L2P, Dualprompt), our proposed PAEMA can also outperform the best player by 2.4%/3.5%, 3.9%/4.3%, 1.4%/1.9%, and 1.7%/1.4% of R@1/mAP on four datasets respectively. This observed superiority can be attributed to the inherent differences in the tasks addressed by these methods compared to the Person Re-identification task tackled by PAEMA. These lifelong learning methods are primarily designed for image classification, with a focus on alleviating catastrophic forgetting in the classification head. However, person re-identification is fundamentally an image retrieval task, where the labels of test images differ from those of the training set. Consequently, these methods achieve inferior results in comparison to PAEMA.

4.3.2.2 PAEMA w/o exemplar versus SOTA w/ exemplar Even compared with exemplar-based methods, our PAEMA still greatly exceeds GwFReID, PTKP, and KRKC. Overall, our PAEMA achieves 1.8%/3.1% improvement on Average R@1/mAP over the original PTKP with ResNet-50 backbone, and comparable results with our implemented PTKP-ViT, verifying our effectiveness in balancing knowledge acquisition and forgetting. It is noticed that compared to PTKP, PTKP-ViT, and PatchKD, our PAEMA achieves a little inferior performance on Market-1501. This is because their methods reserve exemplars of old datasets or keep the old models during training for knowledge distillation. These techniques retain old knowledge but also impair the ability of knowledge acquisition, resulting in inferior average performance.

4.3.2.3 PAEMA w/ exemplar Versus Others w/ exemplar Furthermore, we conduct an extra experiment to incrementally train our model using exemplars in the same way as PTKP. The results show that our method can consistently and greatly beat the second-best player PTKP-ViT on DukeMTMC, CUHK-SYSU, and MSMT17 by 0.4%/2.3%, 1.3%/1.4%, and 4.4%/6.9% of R@1/mAP. As for the Market-1501 dataset which is learned at the initial stage, the R@1/mAP performance difference is just 0.3%/0.4%, which is comparable with PTKP-ViT. Thus, the overall 1.4%/2.5% improvement on Average R@1/mAP demonstrates that our PAEMA is indeed complementary to exemplar-based strategies. Even if the exemplars can be retained for rehearsal, our PAEMA still presents promising performance to facilitate LReID.

4.3.2.4 Average Forgetting Results To verify the anti-forgetting ability of different methods, we also provide the average forgetting results in Table 6. Without using any exemplar, our PAEMA achieves the average forgetting of 3.37%/6.44% on R@1/mAP, which is significantly lower than the existing best model PRAKA (4.58%/7.52%). Besides, the average forgetting of PAEMA can be further reduced when having access to exemplars. These results show the superior anti-forgetting capability of PAEMA.

4.3.2.5 Performance Curve To present the R@1 and mAP performance across LReID training steps in detail, we show the results of methods on different training steps in Fig. 5. It can be observed that, with similar R@1 and mAP accuracy for the initial stage, PAEMA achieves the best results across subsequent training steps. These results underscore that PAEMA could better consolidate the learned knowledge along the training steps, owing to its capacity for adaptively balancing knowledge acquisition and forgetting.

4.3.3 Results on Unseen Datasets of GwFReID Benchmark

To further verify the generalization ability of LReID models, experiments are conducted on four unseen datasets and

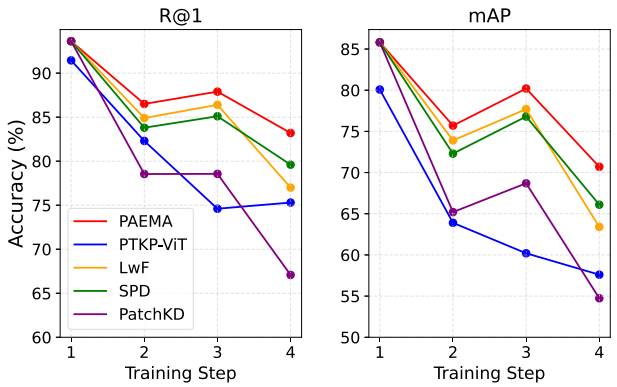

Fig. 5 The mAP and R@1 results on seen datasets of each training step

the results are shown in Table 3. Our PAEMA outperforms all compared methods and achieves SOTA results of 54.8%/58.5% on Average R@1/mAP. Specifically, compared to the exemplar-based methods, our PAEMA achieves better R@1/mAP performance on CUHK01 and SenseReID, and comparable performance on CUHK03 and GRID. Since PTKP uses the exemplars from the historical datasets for training, some knowledge bias may be brought to promote the performance on CUHK03 and GRID. Overall, our PAEMA surpasses the exemplar-based approach PTKP-ViT by 3.1%/3.5% on Average R@1/mAP of all unseen domains. Compared with other exemplar-free lifelong learning and LReID methods, our approach achieves at least 1.5%/1.5%, 6.2%/3.8%, 4.0%/5.1%, and 1.6%/1.3% of R@1/mAP improvement on four unseen datasets. Besides, our method also outperforms Joint-Train on all unseen datasets. The above results show the superiority of our PAEMA in consolidating generalizable knowledge.

4.3.4 Results on AKA-Order-1 and AKA-Order-2 Benchmark

We follow previous works such as AKA and PatchKD to compare exemplar-free methods on the AKA-Order-1 and AKA-Order-2 benchmarks. The comparison results are reported in Tables 4 and 5. For the results of the seen datasets, it can be observed that PAEMA outperforms other methods on the first four datasets and improves the average R@1/mAP of the SOTA methods by 2.2%/3.2% and 1.8%/1.8% under AKA-Order-1 and AKA-Order-2 benchmarks. For the results of the unseen datasets, PAEMA improves the average performance of the SOTA methods by 3.9%/3.9% and 2.6%/2.8% and achieves comparable results with the Joint-Train. These results of different benchmarks show the consistent superiority of PAEMA in anti-forgetting and generalization ability.

Table 3 Comparison with SOTA lifelong learning and LReID methods on unseen datasets of GwFReID benchmark

Methods	CUHK01		CUHK03		GRID		SenseReID		Average	
	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
Joint-Train	81.2	80.3	40.7	37.9	32.8	44.0	42.3	50.4	49.3	53.2
GwFReID*	–	–	40.2	–	–	–	–	–	–	–
PTKP*	79.1	–	**54.1**	–	31.5	–	44.8	–	52.4	–
PTKP-ViT*	79.1	77.9	43.7	40.9	**37.6**	46.3	46.3	54.9	51.7	55.0
KRKC*	79.3	78.1	36.4	34.6	32.8	41.0	40.6	49.5	47.3	50.8
KRKC-ViT*	79.1	78.1	40.1	35.9	15.2	23.1	37.5	45.8	43.0	45.7
LwF	75.7	75.5	40.6	37.1	32.8	42.8	44.4	53.1	48.4	52.2
SPD	80.1	79.9	42.5	40.6	30.4	40.8	46.3	55.1	49.9	54.1
CRL	77.0	76.2	39.8	37.0	32.0	42.3	44.6	53.4	48.3	52.2
L2P	55.1	57.8	26.3	24.1	17.6	26.9	33.0	40.8	33.0	37.4
Dualprompt	36.2	34.2	23.6	16.5	11.2	20.1	24.3	29.1	23.8	25.0
PatchKD	62.0	63.0	15.9	16.2	21.6	30.9	35.2	43.5	33.7	38.4
PTKP	59.6	59.2	42.1	35.9	13.6	19.3	36.8	44.7	38.0	39.8
PTKP-ViT	72.2	72.2	34.8	33.0	28.0	37.3	40.4	48.4	43.9	47.7
NAPA-VQ	77.4	77.1	41.1	39.5	32.8	43.3	45.5	54.5	49.2	53.6
PRAKA	77.1	77.8	44.9	42.3	32.0	42.9	48.3	57.2	50.6	54.8
PAEMA	**81.6**	**81.4**	51.1	**46.1**	36.8	**48.0**	**49.9**	**58.5**	**54.8**	**58.5**

The bold numbers represent the best results
* denotes that historical exemplars are used
– denotes the original paper doesn't report this result

Table 4 Results on AKA-Order-1 benchmark

	Train: Market-1501 → CUHK-SYSU → DukeMTMC → MSMT17_v2 → CUHK03													
	Market-1501		CUHK-SYSU		DukeMTMC		MSMT17_v2		CUHK03		Seen Average		Unseen Average	
Method	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
Joint-Train	92.3	82.3	92.9	92.3	86.0	75.9	67.6	44.8	64.4	65.1	80.8	71.9	64.3	70.9
LwF	78.1	59.2	88.0	86.2	70.3	56.6	47.1	24.4	59.0	57.8	68.5	56.9	58.0	66.3
SPD	82.7	67.7	89.3	87.5	69.8	55.6	38.3	17.4	52.4	51.8	66.5	56.0	59.3	66.4
CRL	78.9	60.4	87.8	86.2	71.5	57.3	47.0	24.7	55.9	54.9	68.2	56.7	57.8	65.7
L2P	83.6	65.1	79.6	76.7	45.3	28.9	20.8	7.61	17.9	18.4	49.4	39.3	40.2	47.8
Dualprompt	82.5	62.3	75.0	70.0	61.0	40.5	28.1	8.15	35.9	35.6	56.5	43.3	36.8	41.4
AKA	72.0	51.2	45.1	47.5	33.1	18.7	37.6	16.4	27.6	27.7	43.1	32.3	40.4	44.3
PatchKD	85.7	68.5	78.6	75.6	50.4	33.8	17.0	6.49	36.8	34.1	53.7	43.7	45.4	49.1
NAPA-VQ	80.6	62.6	88.0	86.1	72.4	58.7	48.7	26.4	58.9	57.5	69.7	58.3	57.9	65.2
PRAKA	78.2	59.2	88.3	86.6	71.5	58.0	49.1	25.9	**65.3**	**63.5**	70.5	58.6	57.0	65.0
PAEMA	**86.9**	**71.7**	**91.3**	**90.0**	**79.0**	**66.0**	**55.3**	**31.4**	50.8	50.0	**72.7**	**61.8**	**63.2**	**70.3**

The **bold** numbers represent the best results

Table 5 Results on AKA-Order-2 benchmark

	Train: DukeMTMC → MSMT17_v2 → Market-1501 → CUHK-SYSU → CUHK03													
	DukeMTMC		MSMT17_v2		Market-1501		CUHK-SYSU		CUHK03		Seen Average		Unseen Average	
Method	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
JointTrain	86.0	75.9	67.6	44.8	92.3	82.3	92.9	92.3	64.4	65.1	80.8	71.9	64.3	70.9
LwF	71.1	56.8	38.1	18.3	82.7	66.7	88.5	86.9	58.4	56.4	67.8	57.0	59.9	66.6
SPD	77.6	66.1	32.0	13.7	73.8	53.6	88.8	87.2	48.2	48.0	64.1	53.7	57.3	64.2
CRL	72.9	58.8	36.9	17.2	82.4	66.7	88.9	86.8	53.2	52.3	66.9	56.4	57.6	65.2
L2P	**81.6**	**69.0**	27.5	10.5	60.5	35.9	80.8	78.3	25.1	25.6	55.1	43.9	46.2	54.0
Dualprompt	81.6	68.8	35.0	13.3	77.8	55.6	80.1	77.6	41.1	40.3	63.1	51.1	44.9	50.3
AKA	60.1	42.2	15.1	5.40	59.8	37.2	73.9	71.2	37.9	36.9	49.4	38.6	41.7	46.0
PatchKD	74.1	58.3	17.4	6.39	67.4	43.2	76.9	74.5	34.8	33.7	54.1	43.2	44.1	48.6
NAPA-VQ	74.7	60.9	42.7	21.3	82.9	66.9	88.6	87.1	57.6	56.9	69.3	58.6	57.6	64.9
PRAKA	71.4	57.1	40.0	19.0	81.5	64.7	88.2	86.4	**63.9**	**62.5**	69.0	58.0	58.8	66.1
PAEMA	79.8	67.2	**49.4**	**26.0**	**85.8**	**69.8**	**91.0**	**89.9**	49.7	49.3	**71.1**	**60.4**	**62.5**	**69.4**

The bold numbers represent the best results

4.4 Ablation Study

4.4.1 The Influence of Different α Choices

We conduct an ablation study of different α choices in Table 7 to demonstrate the effectiveness of our proposed PAEMA scheme. We first train our model on Market-1501 and then finetune it on subsequent datasets. After each finetuning, the EMA with α is adopted to update the model. $\alpha = 1.0$ means only keeping the new model without retaining any knowledge from old models. Descend means adopting the time-varied strategy $\alpha = 1/s$ (Yu et al., 2023), s is the training step. The results in Table 7 show that the value of α significantly impacts final performance and the optimal α varies significantly at different LReID steps. When using a fixed α for EMA, a higher α could result in higher performance on new datasets and lower performance on old datasets. When using the adaptive parameter α obtained by PAEMA, the overall performance can outperform all the fixed EMA strategy and time-varied strategy $\alpha = 1/s$, indicating that our proposed method achieves a better balance between knowledge acquisition and catastrophic forgetting. In this case, the adaptive α calculated by our model is 0.61, 0.57, and 0.61 for three datasets respectively.

Table 6 The average forgetting results of GwFReID benchmark

Methods	E	Average Forgetting(↓) R@1	mAP
PTKP ‡	✓	3.50	5.53
PTKP-ViT‡	✓	3.00	4.58
KRKC‡	✓	8.57	14.1
KRKC-ViT‡	✓	4.47	9.97
PAEMA	✓	**2.83**	**4.36**
LwF‡	×	6.98	11.3
SPD‡	×	4.58	7.52
CRL‡	×	5.68	9.14
L2P‡	×	10.8	14.1
Dualprompt‡	×	8.48	14.5
PTKP‡	×	14.0	21.0
PTKP-ViT‡	×	8.19	13.6
NAPA-VQ‡	×	5.24	8.60
PRAKA‡	×	4.57	7.78
PAEMA	×	**3.37**	**6.44**

The bold numbers represent the best results
E means that sampled images of historical datasets are retained as exemplars and are replayed when new data comes
‡ represents that we reproduce the results based on their codes

4.4.2 The Influence of Prompt Hyperparameters

There are two important hyperparameters in PAEMA to control the adopted prompts: the prompt length L_p and which layers the prompts are embedded. Therefore, we conduct extensive experiments to verify the influence of these hyperparameters. As shown in Table 8, implementing prompts to all layers achieves the best performance, which demonstrates the analysis that all layers must be considered comprehensively to get a better estimation of model variation in Sect. 3.5. For L_p, the length of prompts directly influences the balance between the stability and plasticity of the model. A too-small L_p is not powerful enough for knowledge learning while a too-large L_p will result in a heavy computation burden and a higher rate of knowledge forgetting. It can be observed that as the prompt length increases, the average forgetting rate increases from 3.2%/5.7% to 3.7%/6.4%, indicating that the model experiences more forgetting with longer prompt lengths. Consequently, to strike a balance between knowledge forgetting and acquisition, we have chosen a prompt length hyperparameter of 5 based on the experimental results.

4.5 Training Time Comparison

In this section, we investigate the training costs of different LReID methods. Given the training epochs and batch size of all methods are set the same, we use the *mean second per batch* to reflect the training time. As shown in Fig. 7, we compare the training time with the exemplar-based method PTKP (Ge et al., 2022), exemplar-free methods AKA (Pu et al., 2021) and PatchKD (Sun & Mu, 2022), as well as the baselines (ViT-base, Res-base) with only backbones optimized. Note that only our PAEMA is built upon ViT, the other existing methods including PTKP, AKA, and PatchKD are all built upon ResNet-50.

Compared with the ViT-base methods, our PAEMA brings very little training time growth (6%). However, both exemplar-free methods AKA and PatchKD bring a significant training time growth after step 1 (60%) compared to Res-base owing to the additional designs to maintain the historical knowledge. Moreover, the training time of the exemplar-based method PTKP grows largely along the training steps mainly because of the replaying of historical data.

Table 7 Ablation study between using fixed α for EMA, and using adaptive α obtained by our proposed PAEMA

	Train: Market-1501 → DukeMTMC → CUHK-SYSU→ MSMT17_v1									
	Market-1501		DukeMTMC		CUHK-SYSU		MSMT17_v1		Average	
α	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
0.0	93.4	84.7	52.8	36.3	84.1	82.4	25.8	9.6	64.0	53.2
0.3	92.3	82.8	75.3	63.3	92.5	91.7	55.6	30.9	78.9	67.2
0.4	90.9	79.7	77.8	66.4	93.6	92.6	62.7	38.0	81.3	69.2
0.5	88.9	75.7	78.5	67.7	93.8	92.9	68.4	44.6	82.4	70.2
0.6	86.9	71.7	78.8	67.3	93.7	92.8	72.8	50.1	83.0	70.5
0.7	84.1	67.1	77.5	66.1	93.3	92.4	75.9	54.3	82.7	70.0
0.8	82.4	63.3	76.4	64.1	92.9	91.8	77.7	56.9	82.3	69.0
1.0	76.8	53.7	71.5	57.3	90.5	88.8	78.5	58.3	79.3	64.5
Descend	91.4	80.5	80.3	69.7	92.9	92.1	53.3	28.8	79.5	67.8
PAEMA	87.3	71.4	78.6	67.7	93.6	92.6	73.5	51.0	**83.2**	**70.7**

The bold numbers represent the best results
Descend means adopting the time-varied strategy $\alpha = 1/s$. s is the training step

Table 8 Ablation study on the hyperparameters including the **Length** of each prompt and MSA **Layers** with prompt integrated

		Train: Market-1501 → DukeMTMC → CUHK-SYSU → MSMT17_v1											
		Market-1501		DukeMTMC		CUHK-SYSU		MSMT17_v1		Average		Forgetting	
Length	Layers	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP
5	1–6	87.1	72.6	78.6	67.5	93.7	92.8	72.3	49.2	82.9	70.5	3.1	5.0
5	4–9	85.8	70.9	78.2	67.3	93.6	92.6	74.2	50.8	83.0	70.4	4.2	6.3
5	7–12	84.9	68.3	77.8	66.5	93.8	92.7	74.8	52.6	82.8	70.0	4.3	7.6
5	1–12	87.3	71.4	78.6	67.7	93.6	92.6	73.5	51.0	**83.2**	**70.7**	3.4	6.4
1	1–12	87.4	72.7	78.0	66.8	94.0	92.8	72.6	49.8	83.0	70.5	3.2	5.7
2	1–12	86.3	70.7	78.5	67.2	93.9	92.9	73.3	50.8	83.0	70.4	3.4	6.2
5	1–12	87.3	71.4	78.6	67.7	93.6	92.6	73.5	51.0	**83.2**	**70.7**	3.4	6.4
10	1–12	86.9	71.2	78.8	68.0	93.7	92.7	73.2	50.3	83.1	70.6	3.7	6.4

The **bold** numbers represent the best results

Therefore, the design of our PAEMA is indeed efficient in the training procedure meanwhile can achieve better LReID performance.

4.6 Visualization Comparison

For a comprehensive comparison, in Fig. 6, we visualize the ReID results of our PAEMA and PTKP on the CUHK-SYSU dataset and the unseen SenseReID dataset respectively. The results on CUHK-SYSU show that our proposed PAEMA could extract more discriminative information for accurate retrieval. From the visualized results on SenseReID, we can observe that PTKP tends to retrieve images with similar styles, while our PAEMA could retrieve the correct people even if their styles are apparently different. This may be because exemplar-based methods often collect images of different datasets with various styles, and when these data are trained together, image styles could be a misleading clue,

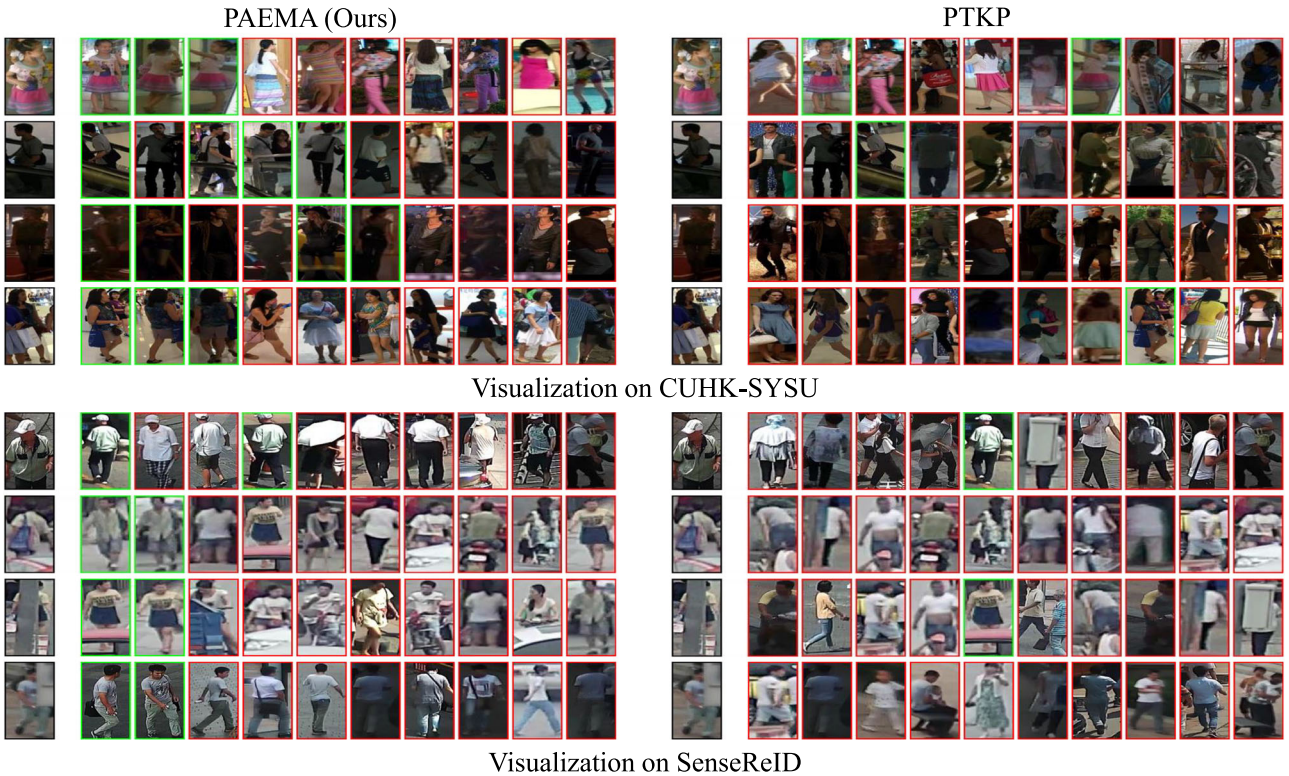

Fig. 6 Top-5 retrieval results of our PAEMA and PTKP on CUHK-SYSU (seen) and SenseReID (unseen). The images in black represent query images, and the ones in the green and red boxes represent the correct and false retrievals respectively (Color figure online)

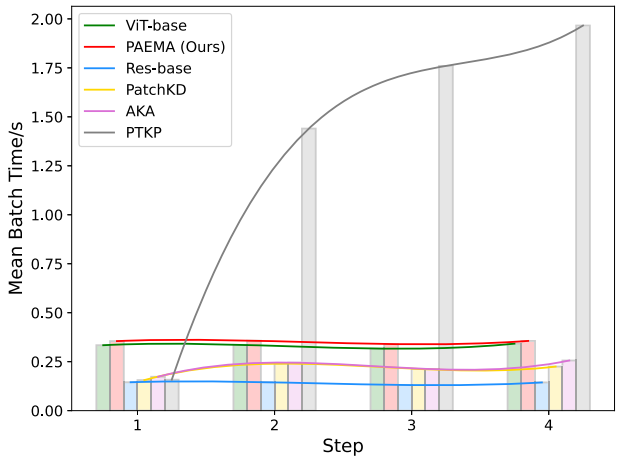

Fig. 7 Illustration of the mean second per batch over different training steps. The exemplar-based method PTKP (Ge et al., 2022), exemplar-free methods (our PAEMA, AKA (Pu et al., 2021), PatchKD (Sun & Mu, 2022)), and baselines (Res-base, ViT-base) built on backbones without any anti-forgetting designs are evaluated. The batch size of all methods and the replay batch size of PKTP are set to 64

and the resulting exemplar-based model trends to distinguish people according to the image styles. This also explains the tremendous improvement of PAEMA on SenseReID which contains people in varied styles.

5 Conclusion

We target a practical yet challenging ReID scenario named Lifelong Person Re-Identification (LReID). Training data from different datasets are given in sequence so that the ReID model needs to be incrementally updated to handle both seen and unseen domains. To mitigate the catastrophic forgetting issue in LReID, a popular solution is to retain sufficient exemplars from the old domains for rehearsal. Differently, in this paper, we propose a novel exemplar-free LReID model based on the designed Prompt-guided Adaptive Exponential Moving Average (PAEMA) algorithm to achieve dynamic knowledge preservation. By innovatively leveraging prompts as a surrogate for knowledge shifting estimation, without using any exemplars, the forgetting issue is greatly alleviated by our method.

Acknowledgements This work was supported by the National Natural Science Foundation of China (62376011, 61925201, 62132001).

Data Availability No datasets are generated and the used datasets that support the findings are Market-1501 http://www.liangzheng.org/Project/project_reid.html, DukeMTMC http://vision.cs.duke.edu/DukeMTMC/, CUHK-SYSU https://drive.google.com/file/d/1XmiNVrfK2ZmI0ZZ2HHT80HHbDrnE4l3W/view?usp=sharing, MSMT_v1, MSMT_v2 http://www.pkuvmc.com/publications/msmt17.html, CUHK01, CUHK02, CUHK03 http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html, GRID http://personal.ie.cuhk.edu.hk/~ccloy/downloads_qmul_underground_reid.html, SenseReID https://drive.google.com/file/d/0B56OfSrVI8hubVJLTzkwV2VaOWM/view?resourcekey=0-PKtdd5m_Jatmi2n9Kb_gFQ, VIPeR http://users.soe.ucsc.edu/~manduchi/VIPeR.v1.0.zip, PRID https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/PRID11/, i-LIDS http://www.eecs.qmul.ac.uk/~jason/data/i-LIDS_Pedestrian.tgz. Code is available at https://github.com/zhoujiahuan1991/IJCV2024-PAEMA/.

References

Ahmed, E., Jones, M., & Marks, T.K. (2015). An improved deep learning architecture for person re-identification. In: CVPR, IEEE, pp. 3908–3916.

Amari, S.-i. (1993). Backpropagation and stochastic gradient descent method. Neurocomputing 5(4-5), 185–196

Cai, Z., Ravichandran, A., Maji, S., Fowlkes, C., Tu, Z., & Soatto, S. (2021). Exponential moving average normalization for self-supervised and semi-supervised learning. In: CVPR, IEEE, pp. 194–203.

Chaudhry, A., Dokania, P.K., Ajanthan, T., & Torr, P.H.S. (2018). Riemannian walk for incremental learning: Understanding forgetting and intransigence. In: Proceedings of the European Conference on Computer Vision (ECCV)

Chen, Y.-C., Zhu, X., Zheng, W.-S., & Lai, J.-H. (2017). Person re-identification by camera correlation aware feature augmentation. *PAMI, 40*(2), 392–408.

Cho, Y., Kim, W.J., Hong, S., & Yoon, S.-E. (2022). Part-based pseudo label refinement for unsupervised person re-identification. In: CVPR, pp. 7308–7318.

Cui, L., Wu, Y., Liu, J., Yang, S., & Zhang, Y. (2021). Template-based named entity recognition using bart. arXiv:2106.01760

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In: CVPR, pp. 248–255. IEEE

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929

Douillard, A., Ramé, A., Couairon, G., & Cord, M. (2022). Dytox: Transformers for continual learning with dynamic token expansion. In: CVPR, IEEE, pp. 9285–9295.

Ge, W., Du, J., Wu, A., Xian, Y., Yan, K., Huang, F., & Zheng, W.-S. (2022). Lifelong person re-identification by pseudo task knowledge preservation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 688–696.

Glorot, X., Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In: ICAIS, JMLR Workshop and Conference Proceedings, pp. 249–256.

Gray, D., & Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10, pp. 262–275. Springer

He, S., Luo, H., Wang, P., Wang, F., Li, H., & Jiang, W. (2021). Transreid: Transformer-based object re-identification. In: ICCV, IEEE, pp. 14993–15002.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In: CVPR, IEEE, pp. 770–778.

Hirzer, M., Beleznai, C., Roth, P.M., & Bischof, H. (2011). Person re-identification by descriptive and discriminative classification. In: Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17, pp. 91–102. Springer

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. (2019). Parameter-

efficient transfer learning for nlp. In: ICML, PMLR, pp. 2790–2799.
Hu, Z., Li, Y., Lyu, J., Gao, D., & Vasconcelos, N. (2023). Dense network expansion for class incremental learning. In: CVPR, pp. 11858–11867.
Huang, Z., Zhang, Z., Lan, C., Zeng, W., Chu, P., You, Q., Wang, J., Liu, Z., & Zha, Z.-j. (2022). Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation. In: CVPR, IEEE, pp. 14288–14297.
Isobe, T., Li, D., Tian, L., Chen, W., Shan, Y., & Wang, S. (2021). Towards discriminative representation learning for unsupervised person re-identification. In: ICCV, IEEE, pp. 8506–8516.
Jia, M., Tang, L., Chen, B.-C., Cardie, C., Belongie, S., Hariharan, B., & Lim, S.-N. (2022). Visual prompt tuning. arXiv:2203.12119
Jin, X., Lan, C., Zeng, W., Chen, Z., & Zhang, L. (2020). Style normalization and restitution for generalizable person re-identification. In: CVPR, pp. 3143–3152.
Kalb, T., & Beyerer, J. (2023). Principles of forgetting in domain-incremental semantic segmentation in adverse weather conditions. In: CVPR, pp. 19508–19518.
Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980
Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences, 114*(13), 3521–3526.
Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. arXiv:2104.08691
Li, W., & Wang, X. (2013). Locally aligned feature transforms across views. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3594–3601
Li, W., Zhao, R., & Wang, X. (2012). Human reidentification with transferred metric learning. In: ACCV, Springer, pp. 31–44.
Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR, IEEE, pp. 152–159.
Li, W., Zhu, X., & Gong, S. (2018). Harmonious attention network for person re-identification. In: CVPR, IEEE, pp. 2285–2294.
Liao, S., & Shao, L. (2022). Graph sampling based deep metric learning for generalizable person re-identification. In: CVPR, pp. 7359–7368.
Li, Z., & Hoiem, D. (2017). Learning without forgetting. *PAMI, 40*(12), 2935–2947.
Lin, G., Chu, H., & Lai, H. (2022). Towards better plasticity-stability trade-off in incremental learning: A simple linear connector. In: CVPR, pp. 89–98.
Lin, Y., Dong, X., Zheng, L., Yan, Y., & Yang, Y. (2019). A bottom-up clustering approach to unsupervised person re-identification. In: AAAI, vol. 33, pp. 8738–8745.
Liu, Y., Schiele, B., Vedaldi, A., & Rupprecht, C. (2023). Continual detection transformer for incremental object detection. In: CVPR, pp. 23799–23808.
Liu, J., Zha, Z.-J., Chen, D., Hong, R., & Wang, M. (2019). Adaptive transfer network for cross-domain person re-identification. In: CVPR, IEEE, pp. 7195–7204.
Liu, Y., Schiele, B., & Sun, Q. (2021). Rmm: Reinforced memory management for class-incremental learning. *Advances in Neural Information Processing Systems, 34*, 3478–3490.
Loy, C. C., Xiang, T., & Gong, S. (2010). Time-delayed correlation analysis for multi-camera activity understanding. *IJCV, 90*(1), 106–129.
Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019). Bag of tricks and a strong baseline for deep person re-identification. In: CVPRW, pp. 1487–1495. IEEE
Luo, Z., Liu, Y., Schiele, B., & Sun, Q. (2023). Class-incremental exemplar compression for class-incremental learning. In: CVPR, pp. 11371–11380.
Malepathirana, T., Senanayake, D., & Halgamuge, S. (2023). Napa-vq: Neighborhood-aware prototype augmentation with vector quantization for continual learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11674–11684
Ni, H., Song, J., Luo, X., Zheng, F., Li, W., & Shen, H.T. (2022). Meta distribution alignment for generalizable person re-identification. In: CVPR, pp. 2487–2496.
Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A.H., & Riedel, S. (2019). Language models as knowledge bases? arXiv:1909.01066
Prabhu, A., Torr, P.H., & Dokania, P.K. (2020). Gdumb: A simple approach that questions our progress in continual learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, pp. 524–540. Springer.
Pu, N., Chen, W., Liu, Y., Bakker, E.M., & Lew, M.S. (2021). Lifelong person re-identification via adaptive knowledge accumulation. In: CVPR, IEEE, pp. 7897–7906.
Pu, N., Liu, Y., Chen, W., Bakker, E.M., & Lew, M.S. (2022). Meta reconciliation normalization for lifelong person re-identification. In: ACMM, pp. 541–549.
Rannen, A., Aljundi, R., Blaschko, M.B., & Tuytelaars, T. (2017). Encoder based lifelong learning. In: ICCV, pp. 1320–1328.
Rebuffi, S.-A., Kolesnikov, A., Sperl, G., & Lampert, C.H. (2017). icarl: Incremental classifier and representation learning. In: CVPR, IEEE, pp. 5533–5542.
Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In: ECCV, Springer, pp. 17–35.
Sankaranarayanan, S., Jain, A., & Lim, S.N. (2017). Guided perturbations: Self-corrective behavior in convolutional neural networks. In: ICCV, IEEE, pp. 3582–3590.
Shi, W., & Ye, M. (2023). Prototype reminiscence and augmented asymmetric knowledge aggregation for non-exemplar class-incremental learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1772–1781
Shmelkov, K., Schmid, C., & Alahari, K. (2017). Incremental learning of object detectors without catastrophic forgetting. In: ICCV, IEEE, pp. 3420–3429.
Smith, J.S., Karlinsky, L., Gutta, V., Cascante-Bonilla, P., Kim, D., Arbelle, A., Panda, R., Feris, R., & Kira, Z. (2023). Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In: CVPR, pp. 11909–11919
Song, J., Yang, Y., Song, Y.-Z., Xiang, T., & Hospedales, T.M. (2019). Generalizable person re-identification by domain-invariant mapping network. In: CVPR, IEEE, pp. 719–728.
Sun, Z., & Mu, Y. (2022). Patch-based knowledge distillation for lifelong person re-identification.
Sun, Z., Mu, Y., & Hua, G. (2023). Regularizing second-order influences for continual learning. In: CVPR, pp. 20166–20175
Tung, F., & Mori, G. (2019). Similarity-preserving knowledge distillation. In: ICCV, IEEE, pp. 1365–1374.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. NIPS 30
Wang F-Y, Zhou D-W, Liu L, Ye H-J, Bian Y, Zhan D-C, Zhao P. (2022). Beef: Bi-compatible class-incremental learning via energy-based expansion and fusion. In: *The Eleventh International Conference on Learning Representations*.
Wang, D., & Zhang, S. (2020). Unsupervised person re-identification via multi-label classification. In: CVPR, IEEE, pp. 10978–10987.

Wang, Z., He, L., Tu, X., Zhao, J., Gao, X., Shen, S., & Feng, J. (2021). Robust video-based person re-identification by ierarchical mining. CSVT

Wang, W., Hu, Y., Chen, Q., & Zhang, Y. (2023). Task difficulty aware parameter allocation & regularization for lifelong learning. In: CVPR, pp. 7776–7785.

Wang, Y., Huang, Z., & Hong, X. (2022). S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. arXiv:2207.12819

Wang, T., Yamaguchi, K., & Ordonez, V. (2018). Feedback-prop: Convolutional neural network inference under partial evidence. In: CVPR, IEEE, pp. 898–907.

Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.-Y., Ren, X., Su, G., Perot, V., & Dy, J., et al. (2022). Dualprompt: Complementary prompting for rehearsal-free continual learning. arXiv:2204.04799

Wang, Z., Zhang, Z., Lee, C.-Y., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., & Pfister, T. (2022). Learning to prompt for continual learning. In: CVPR, IEEE, pp. 139–149.

Wang, F.-Y., Zhou, D.-W., Ye, H.-J., & Zhan, D.-C. (2022). Foster: Feature boosting and compression for class-incremental learning. In: European conference on computer vision, pp. 398–414. Springer

Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person transfer gan to bridge domain gap for person re-identification. In: CVPR, IEEE, pp. 79–88.

Wu, G., & Gong, S. (2021). Generalising without forgetting for lifelong person re-identification. In: AAAI, vol. 35, pp. 2889–2897.

Xiao, T., Li, S., Wang, B., Lin, L., & Wang, X. (2016). End-to-end deep learning for person search. 2(2), 4 arXiv:1604.01850

Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., & Liu, Z. (2021). End-to-end semi-supervised object detection with soft teacher. In: ICCV, IEEE, pp. 3040–3049.

Yu, C., Shi, Y., Liu, Z., Gao, S., & Wang, J. (2023). Lifelong person re-identification via knowledge refreshing and consolidation. In: AAAI, vol. 37, pp. 3295–3303.

Yu, H.-X., Zheng, W.-S., Wu, A., Guo, X., Gong, S., & Lai, J.-H. (2019). Unsupervised person re-identification by soft multilabel learning. In: CVPR, IEEE, pp. 2143–2152.

Zhang, L., Gao, G., & Zhang, H. (2022). Spatial-temporal federated learning for lifelong person re-identification on distributed edges. arXiv:2207.11759

Zhang, W., He, X., Yu, X., Lu, W., Zha, Z., & Tian, Q. (2019). A multi-scale spatial-temporal attention model for person re-identification in videos. *TIP, 29*, 3365-3373.

Zhao, B., Tang, S., Chen, D., Bilen, H., & Zhao, R. (2021). Continual representation learning for biometric identification. In: WACV, IEEE, pp. 1197–1207.

Zhao, H., Tian, M., Sun, S., Shao, J., Yan, J., Yi, S., Wang, X., & Tang, X. (2017). Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In: CVPR, IEEE, pp. 907–915.

Zheng, W.-S., Gong, S., & Xiang, T. (2009). *Associating groups of people.*https://doi.org/10.5244/C.23.23

Zheng, K., Lan, C., Zeng, W., Zhang, Z., & Zha, Z.-J. (2021). Exploiting sample uncertainty for domain adaptive person re-identification. In: AAAI, vol. 35, pp. 3538–3546.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. In: ICCV, IEEE, pp. 1116–1124.

Zhou, D.-W., Wang, Q.-W., Ye, H.-J., & Zhan, D.-C. (2022). A model or 603 exemplars: Towards memory-efficient class-incremental learning. arXiv preprint arXiv:2205.13218

Zhu, H., Ke, W., Li, D., Liu, J., Tian, L., & Shan, Y. (2022). Dual cross-attention learning for fine-grained visual categorization and object re-identification. In: CVPR, pp. 4692–4702.

Zhuang, Z., Wei, L., Xie, L., Zhang, T., Zhang, H., Wu, H., Ai, H., & Tian, Q. (2020). Rethinking the distribution gap of person re-identification with camera-based batch normalization. In: ECCV, Springer, pp. 140–157.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.