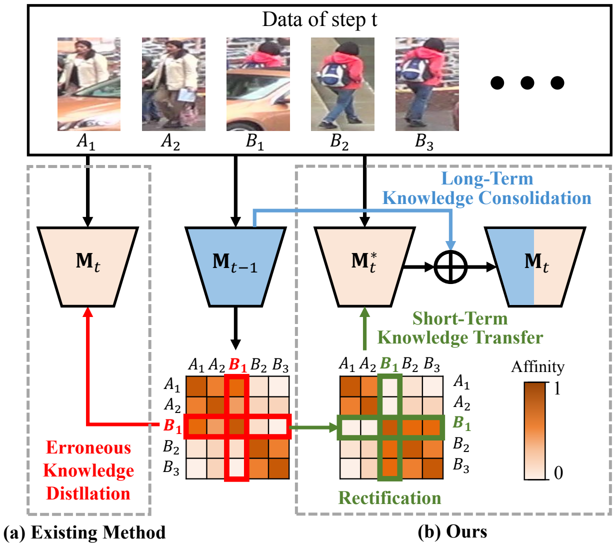
LSTKC: Long Short-Term Knowledge Consolidation for Lifelong Person Re-identification

Kunlun Xu[1], **Xu Zou**[2], **Jiahuan Zhou**[1*]
[1]Wangxuan Institute of Computer Technology, Peking University
[2]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
xkl@stu.pku.edu.cn, zoux@hust.edu.cn, jiahuanzhou@pku.edu.cn

Abstract

Lifelong person re-identification (LReID) aims to train a unified model from diverse data sources step by step. The severe domain gaps between different training steps result in catastrophic forgetting in LReID, and existing methods mainly rely on data replay and knowledge distillation techniques to handle this issue. However, the former solution needs to store historical exemplars which inevitably impedes data privacy. The existing knowledge distillation-based models usually retain all the knowledge of the learned old models without any selections, which will inevitably include erroneous and detrimental knowledge that severely impacts the learning performance of the new model. To address these issues, we propose an exemplar-free LReID method named Long-Short Term Knowledge Consolidation (LSTKC) that contains a Rectification-based Short-Term Knowledge Transfer module (R-STKT) and an Estimation-based Long-Term Knowledge Consolidation module (E-LTKC). For each learning iteration within one training step, R-STKT aims to filter and rectify the erroneous knowledge contained in the old model and transfer the rectified knowledge to facilitate the short-term learning of the new model. Meanwhile, once one training step is finished, E-LTKC proposes to further consolidate the learned long-term knowledge via adaptively fusing the parameters of models from different steps. Consequently, experimental results show that our LSTKC exceeds the state-of-the-art methods by 6.3%/9.4% and 7.9%/4.5%, 6.4%/8.0% and 9.0%/5.5% average mAP/R@1 on seen and unseen domains under two different training orders of the challenging LReID benchmark respectively.



(a) Existing Method

(b) Ours

Figure 1: To prevent erroneous knowledge distillation and facilitate old knowledge consolidation, we propose a Short-Term Knowledge Transfer mechanism to distill the correct knowledge of the old model and a Long-Term Knowledge Consolidation mechanism to accomplish knowledge-guided model fusion.

Introduction

Person re-identification (ReID) aims to retrieve the person of interest from a collection of images. However, numerous investigations (Wang et al. 2022a; Zhao et al. 2021b) have observed that ReID models trained on a specific and stationary dataset often exhibit inadequate performance when confronted with new datasets. This limitation has sparked increasing interest in lifelong person re-identification (LReID), which focuses on continuously learning informative knowledge from a stream of datasets (Pu et al. 2021). Similar to other lifelong learning tasks (Liang et al. 2022), the challenge of catastrophic forgetting emerges as a critical obstacle due to the discrepancy in knowledge across diverse datasets. To handle this issue, several LReID approaches aim to retain exemplars from the old datasets as the rehearsal of historical knowledge for the learning of new models (Wu and Gong 2021; Ge et al. 2022; Yu et al. 2023). However, this solution will indeed impede data privacy and suffer from considerable computational overheads. Therefore, we concentrate on a more practical but challenging LReID setting where no exemplars can be preserved for new model learning.

In addition to exemplar preservation, a majority of existing LReID approaches strive to alleviate catastrophic forgetting through knowledge distillation (Pu et al. 2021; Sun and Mu 2022; Yu et al. 2023), primarily aimed at ensuring out-

[*]Corresponding author

put consistency for new data across both old and new models. However, such stringent constraints significantly curtail the model's ability to acquire new knowledge (Wu and Gong 2021; Pu et al. 2021). Indiscriminately distilling the knowledge from old models to the new ones will inevitably introduce erroneous and detrimental information conflicting with the new data, leading to significant performance degradation (as is shown in Figure 1 (a)). On the other hand, without preserving any historical exemplars, it is hard to effectively and comprehensively retain all the informative knowledge from old models through knowledge distillation using only new data. This oversight leads to severe forgetting of old knowledge in the long run and subsequently results in inferior overall LReID performance.

To address the above issues, we propose a novel Long Short-Term Knowledge Consolidation model, named LSTKC, which can not only actively filter and rectify the erroneous knowledge to enhance short-term knowledge acquisition, but also adaptively balance the old and new knowledge via model fusion to mitigate the long-term catastrophic forgetting. Specifically, we first represent the transferable knowledge as a pair-wise relation matrix (RM), where each element represents the affinity between two samples within one mini-batch of new data. Both the old and new LReID models are utilized to achieve a pair of RM. Then, as is shown in Figure 1 (b), a Rectification-based Short-Term Knowledge Transfer module (R-STKT) is proposed to filter and rectify the erroneous knowledge contained in the RM based on the annotation information of new data. A short-term relation knowledge transfer loss is adopted to transfer the knowledge within rectified old model relation matrix to the new model. Furthermore, when the new model is updated on the new dataset, an Estimation-based Long-Term Knowledge Consolidation module (E-LTKC) is proposed to automatically estimate the degree of long-term knowledge forgetting by leveraging the differences between the aforementioned relation matrices of old and new models. Finally, a knowledge-guided model fusion strategy is designed to adaptively balance the new and old knowledge.

To sum up, the contributions of this paper are as follows: (1) A long short-term knowledge consolidation (LSTKC) model is proposed for LReID, which contains a novel Rectification-based Short-Term Knowledge Transfer Module (R-STKT) and an effective Estimation-based Long-Term Knowledge Consolidation Module (E-LTKC). (2) The proposed R-STKT mechanism performs relation matrix-based erroneous knowledge filtering and rectification to facilitate the correct knowledge transfer within the short-term learning stage. (3) The proposed E-LTKC module proposes to actively balance the forgetting and acquisition of old and new knowledge via a knowledge-guided model fusion strategy to consolidate long-term knowledge. (4) Extensive experimental results demonstrate that our proposed LSTKC model exceeds state-of-the-art LReID methods by a large margin in different settings.

Related Work

Person Re-Identification

Person Re-Identification (ReID) aims to justify if given images from distinct cameras, times, and locations contain the same person (Ahmed, Jones, and Marks 2015; Li, Zhu, and Gong 2018; Luo et al. 2019). It has been extensively studied in a close setting, where the scenarios of the test data are identical to the training ones (Zhuang et al. 2020; He et al. 2021; Chen et al. 2017). However, such a training procedure though works well on seen datasets, often exhibits significant performance degradation on different datasets, inhibiting the practical usage of the existing ReID models (Liu et al. 2019; Song et al. 2019). Therefore, in this paper, we study the Lifelong Person Re-Identification (LReID) task to enable the model to consistently learn from labeled data of diverse datasets that could adapt to various conditions.

Lifelong Person Re-Identification

LReID has drawn increasing attention in recent years. Similar to other lifelong learning tasks (Wang et al. 2022b), the catastrophic forgetting problem (Li and Hoiem 2017; Shmelkov, Schmid, and Alahari 2017) that the ReID performance on historical datasets will degrade seriously when a model is trained on new ones, is also the main challenge. To overcome this problem, various methods have been proposed which can be mainly categorized into two branches: data replay-based methods and knowledge distillation-based ones. The data reply-based approaches aim to prevent knowledge forgetting via storing and replaying exemplars from historical datasets (Wu and Gong 2021; Ge et al. 2022; Yu et al. 2023; Chen, Lagadec, and Bremond 2022; Huang et al. 2022). However, such a strategy tends to hinder data privacy and incur substantial computational overheads.

The knowledge distillation technique is widely used in LReID (Pu et al. 2021; Wu and Gong 2021; Ge et al. 2022; Sun and Mu 2022) by forcing the new model to generate consistent outputs as the old model. (Pu et al. 2021) was one of the initial works that introduced Knowledge distillation into the LReID task and adopted logistic distillation which forced the new model to generate the same classification score as the old model. (Sun and Mu 2022) designed a patch distillation module that can recognize the important regions of the image and distill the patch logits within such regions. Besides, (Sun and Mu 2022) also proposed patch relation distillation that constrains the new model to generate the same relative inter-instance distances as the old model. (Pu et al. 2022) claimed the crucial role of Batch Normalization (BN) in data distribution shifting and proposed reconciliation normalization to constrain the learning of BN layers. Apart from the aforementioned exemplar-free LReID methods, various exemplar-based approaches also adopt knowledge distillation to alleviate catastrophic forgetting (Wu and Gong 2021; Ge et al. 2022; Yu et al. 2023).

However, imposing strict constraints on the output of the new model often hampers its ability to adapt to new domains. For example, PatchKD (Sun and Mu 2022) and Lwf (Li and Hoiem 2017) experienced considerable performance degradation when trained for several steps and applied to new datasets, compared to training them individually on those datasets. Besides, the knowledge of the old model may be erroneous, which could mislead the new model if treated uniformly with the correct knowledge. Finally, when employing knowledge distillation strategies, it is crucial to emphasize that only the knowledge that can be reflected by the new data is distilled, while any knowledge that cannot be manifested in the new data is inevitably forgotten.

Method

Formulation

Lifelong person re-identification (LReID) assumes that various training datasets from different domains are given for learning step by step, and the data of previous and later steps are unavailable for the current step (Pu et al. 2021; Sun and Mu 2022). Specifically, the training data consists of a stream of T datasets $\mathcal{D} = \{D_t\}_{t=1}^{T}$ where each $D_t = \{(x_i, y_i)\}_{i=1}^{n_t}$ contains n_t images x_i and their identity labels y_i. Note that the identities between different training datasets are disjoint. After the t-th training step, the leaned model is $\mathbf{M}_t^*$ and the finally obtained model after our method is $\mathbf{M}_t$. Thus, $\mathbf{M}_T$ is the final model. During testing, to evaluate the acquisition and anti-forgetting capacity of $\mathbf{M}_T$, a series of T testing datasets $\mathcal{D}^{test} = \{D_t^{test}\}_{t=1}^{T}$ collected from all the seen domains are evaluated. Besides, to verify the generalization capacity of $\mathbf{M}_T$, an additional series of U datasets $\mathcal{D}^{un} = \{D_t^{un}\}_{t=1}^{U}$ from unseen domains are tested as well.

Overview

As shown in Figure 2, our method mainly contains a backbone network, *i.e.*, (a) and (b), a Rectification-based Short-Term Knowledge Transfer module (c) and an Estimation-based Long-Term knowledge Consolidation module (d).

Base Model

The overall architecture of our model follows the typical setting of existing LReID approaches (Pu et al. 2021; Sun and Mu 2022) that adopt a CNN backbone to extract input image features and utilize a classifier to predict person identities. Specifically, the backbone and classifier of $\mathbf{M}_t$ are denoted as Θ_t and Φ_t respectively. Given an input image $x \in \mathbb{R}^{H\times W\times C}$, Θ_t converts x into a feature vector $v \in \mathbb{R}^d$, where H, W and C are the image height, width and channel respectively, and d is the feature dimension. Then Φ_t takes v as the input and generates logistic predictions. Therefore, the overall model could be represented as $\mathbf{M}_t(x; \Theta_t, \Phi_t) = \Phi_t(\Theta_t(x))$. The learnable parameters in Θ_t and Φ_t are optimized by a cross-entropy loss together:

$$\mathcal{L}_{ID} = -y \log\left(\sigma(\mathbf{M}_t(x; \Theta_t, \Phi_t))\right), \tag{1}$$

where σ is the softmax function and y is the identity label of x.

Furthermore, to enhance the discriminative capacity of the model, a normalization-guided Triplet loss (Liu et al. 2017) is adopted:

$$\mathcal{L}_{Tri} = \log(1 + \exp(\parallel \tilde{v}_a - \tilde{v}_p \parallel_2^2 - \parallel \tilde{v}_a - \tilde{v}_n \parallel_2^2)), \tag{2}$$

where $\tilde{v}$ is the L_2-normalized version of v (corresponding to the "Norm" module in Figure 2 (a) and (b)), and $\langle a, p, n\rangle$ is a triplet set. Therefore, the overall optimization loss of the base model is:

$$\mathcal{L}_{Base} = \mathcal{L}_{ID} + \gamma \mathcal{L}_{Tri}, \tag{3}$$

where γ is a hyperparameter to balance two components.

Following the existing works (Ge et al. 2022), for the first training step, the parameters of Θ_1 are initialized with the ImageNet pre-trained model, and the parameters of Φ_1 are randomly initialized. For the later steps, the parameters of Θ_t are initialized with Θ_{t-1} and the parameters of Φ_{t-1} which is a linear layer are initialized with the mean feature of the identities in D_t.

Rectification-based Short-Term Knowledge Transfer (R-STKT)

A core function of LReID models is to evaluate the similarity between different person images. In light of this, we delve into knowledge transfer through the lens of relation distillation. Initially, we introduce a pairwise relation matrix to effectively capture and encapsulate the knowledge embedded within the features extracted from LReID models.

Pair-wise Relation Matrix: During training step t ($t > 1$), given a batch of data $\mathcal{I} = \{(x_{t,i}, y_{t,i})\}_{i=1}^{B}$ where B is the batch size. The feature extractor of $\mathbf{M}_t^*$ covert $\mathcal{I}$ into a feature matrix denoted as $V_t \in \mathbb{R}^{B\times d}$. Then we use L_2 normalization to process V_t at the channel dimension and obtain $\tilde{V}_t$. A pair-wise similarity matrix $S_t \in \mathbb{R}^{B\times B}$ is calculated by $S_t = \tilde{V}_t \cdot (\tilde{V}_t)^\top$ which can be utilized to further calculate the pair-wise relation matrix $R_t \in \mathbb{R}^{B\times B}$:

$$R_t[i,j] = \frac{\exp(S_t[i,j]/\tau)}{\sum_{k=1}^{B} \exp(S_t[i,k]/\tau)}, \tag{4}$$

where $[i, j]$ denotes the i-th row, j-th column of the corresponding matrix, and τ is a temperature hyper-parameter. Note that each row of R_t is a softmax-like function that maps each row of similarity matrix S_t into a relation score vector which reflects the overall affinity score between $x_{t,i}$ and other images in $\mathcal{I}$. When $1/\tau$ is large, the high similarity pairs will dominate the relation scores in R_t which can assure that similar pairs are assigned more attention compared to dissimilar ones during the following knowledge transfer.

Besides, to represent the pair-wise relation knowledge of the old model $\mathbf{M}_{t-1}$, we feed the current batch $\mathcal{I}$ into $\mathbf{M}_{t-1}$ to obtain the pair-wise relation matrix R_{t-1} following the above manner.

Erroneous Knowledge Filtering: Due to the data variations between different lifelong learning steps, directly distilling the knowledge in the relation matrix R_{t-1} of the old model to learn R_t will inevitably introduce erroneous knowledge that the matching relationships in R_{t-1} are incorrect. The potential presence of such incorrect knowledge from the old model might misguide the learning process of the new model. To address this issue, we propose to filter out and rectify the erroneous relation knowledge from R_{t-1}, ensuring the transferred knowledge is correct and informative.

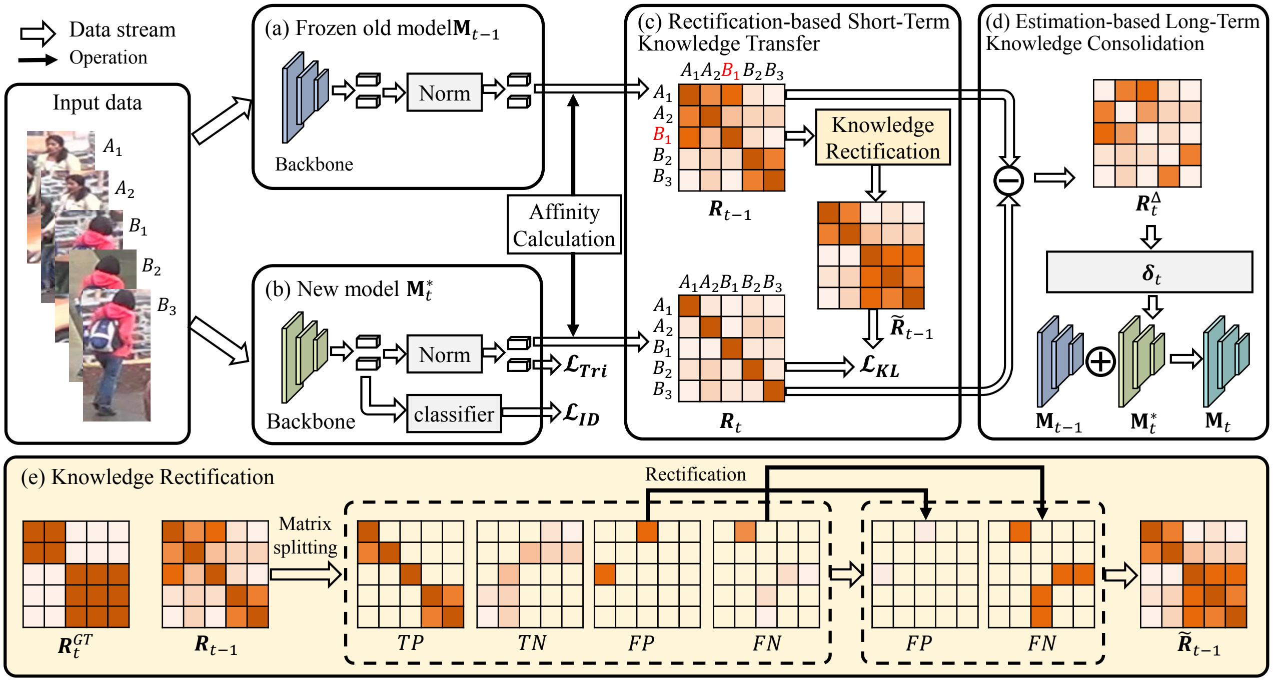


Figure 2: The overall pipeline of our proposed LSTKC. (a) and (b) are the old and new models respectively. (c) is the Rectification-based Short-Term Knowledge Transfer module (R-STKT) that rectifies the erroneous old knowledge in a relation matrix, then distills the rectified relation matrix to the new model to achieve correct knowledge transfer. (d) is the Estimation-based Long-Term Knowledge Consolidation module (E-LTKC) that evaluates the knowledge difference between the new and old models so as to accomplish knowledge-guided model fusion. (e) is the illustration of the knowledge rectification procedure.

As is illustrated in Figure 2 (e), we split R_{t-1} into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) groups where FP and FN are the erroneous relations, where true/false denotes a specific element in the relation matrix is correct/erroneous, and positive/negative denotes the corresponding pair of the element is/not the same person according to annotation. To achieve this, we design two thresholds to separate TP/FP and TN/FN respectively. Specifically, given the i-th row of R_{t-1}, the positive and negative pairs could be separated beforehand according to the identity annotations. Then the maximum negative relation score $s_{i,n}$ and the minimum positive relation score $s_{i,p}$ of i-th row of R_{t-1} could be obtained. Finally, $s_{i,n}$ and $s_{i,p}$ are set as the thresholds for positive and negative pairs respectively, where for the positive pairs, the ones whose relation score are higher than $s_{i,n}$ are set as TP, and for the negative pairs, the ones whose relation score are higher than $s_{i,n}$ are set as FP.

Erroneous Knowledge Rectification: For a given row i-th of R_{t-1}, we replace the FP elements with $s_{i,p}$ and the FN elements with $s_{i,n}$. The resulting rectified matrix is denoted as R_{t-1}^{re}. Note that when false predictions occur, $s_{i,p}$ is always smaller than $s_{i,n}$, so it could be assured that all scores of the true pairs are always higher than the scores of the false pairs in rectified matrix R_{t-1}^*. Then, in order to ensure that the rows of the rectified relation matrix adhere to the probability law, we apply L_1 normalization to each row of R_{t-1}^*, and the resulting matrix is denoted $\tilde{R}_{t-1}$.

To transfer the relation knowledge from the old model $\mathbf{M}_{t-1}$ to the current model $\mathbf{M}_t$, we adopt Kullback Leibler divergence (KL) which is calculated by:

$$\mathcal{L}_{KL} = \frac{1}{B}\sum_{i=1}^{B} KL(R_{t-1}[i,:]||R_t[i,:]), \tag{5}$$

where $[i,:]$ denotes the i-th row of corresponding matrix. The normalized rectified relation matrix $\tilde{R}_{t-1}$ is set as the target distribution and the relation matrix R_t generated by the current model is set as the source distribution.

Estimation-based Long-Term Knowledge Consolidation (E-LTKC)

Although the R-STKT accomplishes short-term correct knowledge transfer between adjacent training steps, due to the domain gap between different datasets, the knowledge of the old model is hard to fully reflect in new data. In this section, we aim to achieve long-term knowledge consolidation by fusing the models of different stages. Instead of using fixed fusion parameters, We propose adaptively balancing the new and old knowledge based on the knowledge difference estimation of the new and old models.

Knowledge-Guided Model Fusion With a little abuse of R_{t-1} and R_t which denotes the extracted relation matrix from all training data D_t by $\mathbf{M}_{t-1}$ and $\mathbf{M}_t^*$ respectively here. As is shown in Figure 2(d), firstly, we obtain

Method	Market		CUHK-SYSU		DukeMTMC		MSMT17		CUHK03		Seen-Avg		Unseen-Avg	
	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
Finetune	32.7	58.3	58.0	60.6	25.2	43.8	4.5	13.1	41.3	43.4	32.3	43.9	38.4	34.4
JointTrain	68.1	85.2	81.4	83.8	60.4	75.7	24.6	48.9	42.7	43.6	55.4	67.5	49.8	46.3
SPD (Tung and Mori 2019)	35.6	61.2	61.7	64.0	27.5	47.1	5.2	15.5	42.2	44.3	34.4	46.4	40.4	36.6
LwF (Li and Hoiem 2017)	56.3	77.1	72.9	75.1	29.6	46.5	6.0	16.6	36.1	37.5	40.2	50.6	47.2	42.6
CRL (Zhao et al. 2021a)	58.0	78.2	72.5	75.1	28.3	45.2	6.0	15.8	37.4	39.8	40.5	50.8	47.8	43.5
AKA (Pu et al. 2021)	51.2	72.0	47.5	45.1	18.7	33.1	16.4	37.6	27.7	27.6	32.3	43.1	44.3	40.4
AKA* (Pu et al. 2021)	47.2	69.8	72.8	76.5	26.6	44.3	6.2	17.2	42.2	43.6	39.0	50.3	47.7	41.6
PatchKD (Sun and Mu 2022)	**68.5**	**85.7**	75.6	78.6	33.8	50.4	6.5	17.0	34.1	36.8	43.7	53.7	49.1	45.4
Ours	54.7	76.0	**81.1**	**83.4**	**49.4**	**66.2**	**20.0**	**43.2**	**44.7**	**46.5**	**50.0**	**63.1**	**57.0**	**49.9**

Table 1: Training order-1: Market-1501→ CUHK-SYSU→ DukeMTMC-reID→ MSMT17-V2→ CUHK03. '*' denotes our re-implementation with the same batch size as ours based on AKA official code.

Method	DukeMTMC		MSMT17		Market		CUHK-SYSU		CUHK03		Seen-Avg		Unseen-Avg	
	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
Finetune	26.1	45.7	3.3	10.3	29.1	54.1	57.2	60.0	40.3	40.9	31.2	42.2	36.1	32.0
JointTrain	60.4	75.7	24.6	48.9	68.1	85.2	81.4	83.8	42.7	43.6	55.4	67.5	49.8	46.3
SPD (Tung and Mori 2019)	28.5	48.5	3.7	11.5	32.3	57.4	62.1	65.0	43.0	45.2	33.9	45.5	39.8	36.3
LwF (Li and Hoiem 2017)	42.7	61.7	5.1	14.3	34.4	58.6	69.9	73.0	34.1	34.1	37.2	48.4	44.0	40.1
CRL (Zhao et al. 2021a)	43.5	63.1	4.8	13.7	35.0	59.8	70.0	72.8	34.5	36.8	37.6	49.2	45.3	41.4
AKA (Pu et al. 2021)	32.5	49.7	-	-	-	-	-	-	-	-	-	-	40.8	37.2
AKA* (Pu et al. 2021)	37.9	55.9	5.2	14.4	36.6	59.0	72.9	76.0	41.6	41.9	38.8	49.4	44.9	38.5
PatchKD (Sun and Mu 2022)	**58.3**	**74.1**	6.4	17.4	43.2	67.4	74.5	76.9	33.7	34.8	43.2	54.1	48.6	44.1
Ours	49.9	67.6	**14.6**	**34.0**	**55.1**	**76.7**	**82.3**	**83.8**	**46.3**	**48.1**	**49.6**	**62.1**	**57.6**	**49.6**

Table 2: Training order-2: DukeMTMC-reID→ MSMT17-V2→Market-1501→ CUHK-SYSU→ CUHK03. '*' denotes our re-implementation with the same batch size as ours based on AKA official code.

the element-wise absolute difference of R_{t-1} and R_t named $R_t^{\Delta} \in \mathbb{R}^{n_t \times n_t}$. Then we convert R_t^{Δ} into a scalar δ_t by:

$$\delta_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \left(\sum_{j=1}^{n_t} R_t^{\Delta}[i,j] \right), \tag{6}$$

where $\sum_{j=1}^{n_t} R_t^{\Delta}[i,j]$ denotes row-wise addition that evaluates the knowledge difference reflected on image $x_{t,i}$, and $\frac{1}{n_t} \sum_{i=1}^{n_t}(\cdot)$ calculates average knowledge difference of D_t, which represents the between $\mathbf{M}_{t-1}$ and $\mathbf{M}_t^*$. Then we obtain the final model $\mathbf{M}_t$ in step t by:

$$\mathbf{M}_t = (1-\delta_t)\mathbf{M}_t^* + \delta_t \mathbf{M}_{t-1}, \tag{7}$$

where δ_t serves as a weight balancing the new and old knowledge.

After the lifelong learning procedure with T steps, the final model $\mathbf{M}_T$ could be represented as

$$\begin{aligned} \mathbf{M}_T &= (1-\delta_T)\mathbf{M}_T^* + \delta_T((1-\delta_{T-1})\mathbf{M}_{T-1}^* + \delta_{T-1}...) \\ &= \beta_T \mathbf{M}_T^* + \beta_{T-1}\mathbf{M}_{T-1}^* + ... + \beta_1 \mathbf{M}_1^* \end{aligned}, \tag{8}$$

where $\beta_t = (1-\delta_t) \prod_{i=t+1}^{T} \delta_i$, and $\delta_1 = 0$ because the model of the first step does not need fusion. It is obvious that the final model is equivalent to a weighted fusion of all previous models that assures long-term knowledge consolidation.

Training and Inference

During the t-th training step, only the old model $\mathbf{M}_{t-1}$ and the new data D_t are utilized to learn $\mathbf{M}_t^*$. The proposed R-STKT is adopted along the model learning and the total training loss function is formulated as

$$\mathcal{L} = \mathcal{L}_{Base} + \mathcal{L}_{KL}, \tag{9}$$

At the end of the t-th training step, we adopt E-LTKC to fuse $\mathbf{M}_{t-1}$ and $\mathbf{M}_t^*$ and obtain the final model $\mathbf{M}_t$.

During inference, the model $\mathbf{M}_T$ obtained after training step T, followed by L_2 normalization, is used to extract image features for image ranking.

Experiments

Datasets

We conducted all our experiments on the widely-used LReID benchmark (Pu et al. 2021), which consists of 12 datasets. Among them, five datasets (Market-1501 (Zheng et al. 2015), DukeMTMC-reID (Ristani et al. 2016), CUHK-SYSU (Xiao et al. 2016), MSMT17-V2 (Wei et al. 2018), and CUHK03 (Li et al. 2014)) are seen datasets used for LReID training and anti-forgetting testing. The remaining seven datasets (CUHK01 (Li, Zhao, and Wang 2012), CUHK02 (Li and Wang 2013), VIPeR (Gray and Tao 2008), PRID (Hirzer et al. 2011), i-LIDS (Branch 2006), GRID (Loy, Xiang, and Gong 2010), and SenseReID (Zhao et al. 2017)) are unseen datasets used solely for testing purposes. Note that the LReID benchmark selects 500 identities from each seen dataset for training. Typically, two training orders are adopted: Market-1501→CUHK-SYSU→DukeMTMC-reID→MSMT17→CUHK03 and DukeMTMC-reID→MSMT17→Market-1501→CUHK-SYSU→CUHK03. For more detailed information about these datasets, please refer to the Supplementary.

Evaluation Metrics Following the evaluation settings of the existing methods (Pu et al. 2021; Sun and Mu 2022), we calculate mean Average Precision (mAP) and Rank@1 (R@1) accuracy on each dataset and use the average of mAP and R@1 results on the seen and unseen datasets to evaluate the overall performance of our method.

Implementation Details

Following previous works (Pu et al. 2021; Sun and Mu 2022), we adopt ResNet-50 (He et al. 2016) pre-trained on ImageNet (Deng et al. 2009) as our backbone. For both training orders, the first dataset is trained for 80 epochs and the subsequent datasets are trained for 60 epochs using an SGD optimizer with a momentum of 0.9. The learning rate is set to 8×10^{-3} initially with 0.1 decay at the 30th epoch. The input images are resized to 256×128 with random cropping, erasing, and horizontal flipping augmentation. The batch size is set to 128 with 32 identities and 4 images for each identity. The hyperparameter γ and τ are set to 1 and 0.1 respectively. Our implementation is based on PyTorch. All experiments are conducted on a single NVIDIA 4090 GPU.

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

Comparison Methods: In the following experiments, we compare our LSTKC method with classical lifelong learning techniques, namely LwF (Li and Hoiem 2017), SPD (Tung and Mori 2019), as well as the most recent exemplar-free LReID approaches, AKA (Pu et al. 2021), PatchKD (Sun and Mu 2022), CRL (Zhao et al. 2021a). Besides, the term "Finetune" refers to training the datasets step by step without any anti-forgetting design. On the other hand, "JointTrain" represents aggregating all available training data to jointly train the model, which is commonly regarded as the upper bound performance of LReID. All compared experimental results follow the report from PatchKD (Sun and Mu 2022) or the official publications if not explicitly stated.

Results on Seen Datasets: Table 1 and Table 2 show the results on the LReID benchmark under training order-1 and training order-2 respectively. It can be observed that on the seen datasets, we achieve 6.3%/9.4% and 6.4%/8.0% mAP/R@1 improvement over SOTA PatchKD under training order-1 and training order-2 separately. In particular, our model achieves the highest mAP/R@1 on four of five datasets under both training orders with significant improvement. On the first dataset of both training orders, we achieve inferior performance compared to SOTA approaches, this may result from the strict knowledge distillation loss used in SOTA approaches forcing the network to maintain the learned architecture on the first dataset. However, the results in the subsequent steps show that such a strong constraint to the model output limits its new knowledge acquisition capacity. In contrast, our model, thanks to the pair-wise relation transfer design and the adaptively long-term knowledge balancing mechanism, better achieves progressive new knowledge acquisition and consolidation, obtaining state-of-the-art results on most datasets and significantly improving the average performance.

Baseline	R-STKT	E-LTKC	Seen-Avg		Unseen-Avg	
			mAP	R@1	mAP	R@1
✓			40.4	53.3	48.2	41.4
✓	✓		46.1	59.7	54.2	47.9
✓		✓	46.3	59.5	53.6	46.2
✓	✓	✓	**50.0**	**63.1**	**57.0**	**49.9**

Table 3: Ablation study on individual components of R-STKT and E-LTKC.

Strategy	Seen-Avg		Unseen-Avg	
	mAP	R@1	mAP	R@1
Baseline	40.4	53.3	48.2	41.1
Max-Min	43.4	56.8	51.9	45.3
1-Minuscule	25.3	33.8	36.0	28.7
Ours	**46.1**	**59.7**	**54.2**	**47.9**

Table 4: Ablation on knowledge rectification strategy of R-STKT. The experiments are conducted without E-LTKC.

Results on Unseen Datasets

The results on unseen datasets are shown in the "Unseen-Avg" items in Table 1 and Table 2. Our method outperforms SOTA PathchKD by 7.9%/4.5% and 9.0%/5.5% on mAP/R@1 under training order-1 and training order-2 respectively, showing the overwhelming generalization capacity of our method compared existing LReID approaches. Furthermore, our method achieves 7.2%/3.6% and 7.8%/3.3% higher mAP/R@1 than JointTrain respectively. Note that our LSTKC outperforms JointTrain on unseen domains, showing that our short-term knowledge rectification and long-term knowledge consolidation mechanisms could preserve abundant generalizable knowledge.

Ablation Studies

In this section, we conduct experiments on the LReID benchmark under training order-1 to evaluate the effectiveness of each component in our model.

The effectiveness of R-STKT and E-LTKC modules: In Table 3, we present the ablation studies of R-STKT and E-LTKC. We start from a baseline model that merely uses $\mathcal{L}_{base}$ as the loss function and adds R-STKT/E-LTKC gradually. From the second row and the third row, we can observe that R-STKT and E-LTKC could improve mAP/R@1 by about 4.8-6.2% over baseline when used alone. When they are used together, the improvement is up to 8.5-9.8%. The results show that both R-STKT and E-LTKC can work individually and cooperatively.

Ablation on knowledge rectification strategy: Knowledge rectification plays a vital role in R-STKT. We introduce different knowledge rectification strategies including replacing the false negative and false positive affinity

Strategy	Seen-Avg		Unseen-Avg	
	mAP	R@1	mAP	R@1
Baseline	40.4	53.3	48.2	41.1
MSE	40.9	53.6	49.0	42.3
MAE	41.5	54.3	50.4	43.3
JS	45.3	58.6	53.7	47.1
KL (Ours)	**46.1**	**59.7**	**54.2**	**47.9**

Table 5: Ablation on Affinity Loss of R-STKT. The experiments are conducted without E-LTKC.

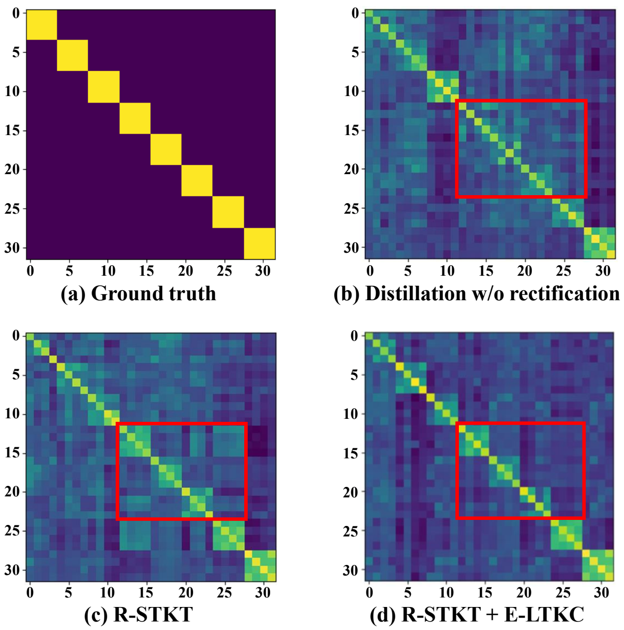
Strategy	Seen-Avg		Unseen-Avg	
	mAP	R@1	mAP	R@1
R-STKT-only	46.1	59.7	54.2	47.9
Fixed (0.5)	49.5	61.3	55.7	48.2
Time-increasing	46.3	57.0	51.7	44.9
Time-descending	48.9	62.0	56.1	49.0
Knowledge-guided (Ours)	**50.0**	**63.1**	**57.0**	**49.9**

Table 6: Ablation on choices of model fusion.

with (a) the maximum and minimum affinities of each person (noted as Max-Min), (b) 1 and minuscule value (noted as 1-Minuscule) respectively. The results are shown in Table 4. It can be observed that our method performs the best and both our method and Max-Min strategy outperform the baseline, while 1-Minuscule only obtains much inferior performance. This is because compared to our method, Max-Min modifies the original affinity more and 1-Minuscule modifies even more. Because the change of specific affinity involves the entire affinity distribution alteration within each row of the relation Matrix, too much modification would destroy the correct knowledge and hinder the correct knowledge transfer. Compared to the other two methods, our rectification strategy not only corrects the erroneous knowledge but also largely retrains the original relation distribution. Therefore, our strategy is a better selection.

Ablation on the choices of knowledge transfer loss: We also conduct ablation experiments on the choices of knowledge transfer loss by replacing our KL divergence with MSE (Park et al. 2019), MAE, and Jensen-Shannon divergence (Yu et al. 2023) that are frequently adopted in relevant works. The results are shown in Table 5, showing that KL is the best choice to accomplish knowledge transferring. This is because KL is designed for knowledge transfer from target distribution to source distribution, which corresponds with the objective of R-STKT.

Ablation on the choices model fusion: To show the effectiveness of our knowledge-guided model fusion strategy in the E-LTKC module, we compared some popular choices including (a) a fixed parameter 0.5, (b) a time-increasing parameter $1 - 1/t$, (c) a time-descending parameter $1/t$ (Pu et al. 2021; Lin, Chu, and Lai 2022). The results are shown in Table 6. It can be observed that our knowledge-guided model fusion strategy outperforms all compared choices on both seen and unseen datasets, showing our active balance strategy could better consolidate the valid knowledge.



(a) Ground truth

(b) Distillation w/o rectification

(c) R-STKT

(d) R-STKT + E-LTKC

Figure 3: We visualize the relation matrix of test images to show the effectiveness of our proposed knowledge rectification (R-STKT) and consolidation (E-LTKC) strategy.

Visualization Studies

To intuitively show the effectiveness of our knowledge rectification (R-STKT) and consolidation (E-LTKC) mechanisms, we visualize the image relation matrix shown in Figure 3 based on test samples from the DukeMTMC-reID dataset under training order-1. The Ground Truth matrix (a) is generated according to the person identity annotations. Compared to (b) which uses the distillation loss without rectification, as highlighted by the red rectangle, our R-STKT (c) significantly improves the number of true positive pairs. On the other hand, when R-STKT collaborates with E-LTKC (d), not only are the true positive pairs improved, but the false positive pairs are reduced, further interpreting how R-STKT and E-LTKC mutually reinforce each other.

Conclusion

In this paper, we propose an exemplar-free LReID method named long short-term knowledge consolidation (LSTKC). Specifically, it contains a rectification-based short-term knowledge transfer module (R-STKT) and an estimation-based long-term knowledge consolidation module (E-LTKC). R-STKT aims to filter and rectify the erroneous knowledge of the old model and transfer the rectified knowledge to the new model. R-STKT aims to automatically estimate the knowledge difference between the new and old models, and accomplish a long-term balance of acquired knowledge. Extensive experimental results show that both modules could perform effectively and mutually reinforce each other, making our performance exceed SOTA PatchKD by at least 6.3%/8.0% and 7.9%/4.5% Average mAP/R@1 on the seen and unseen domain respectively.

Acknowledgments

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

References

Ahmed, E.; Jones, M.; and Marks, T. K. 2015. An Improved Deep Learning Architecture for Person Re-identification. In *CVPR*, 3908–3916. IEEE.

Branch, H. O. S. D. 2006. Imagery library for intelligent detection systems (i-lids). In *2006 IET conference on crime and security*, 445–448. IET.

Chen, H.; Lagadec, B.; and Bremond, F. 2022. Unsupervised Lifelong Person Re-identification via Contrastive Rehearsal. arXiv:2203.06468.

Chen, Y.-C.; Zhu, X.; Zheng, W.-S.; and Lai, J.-H. 2017. Person Re-identification by Camera Correlation Aware Feature Augmentation. *PAMI*, 40(2): 392–408.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A Large-scale Hierarchical Image Database. In *CVPR*, 248–255. IEEE.

Ge, W.; Du, J.; Wu, A.; Xian, Y.; Yan, K.; Huang, F.; and Zheng, W.-S. 2022. Lifelong Person Re-identification by Pseudo Task Knowledge Preservation. In *AAAI*, 688–696.

Gray, D.; and Tao, H. 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 262–275. Springer.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *CVPR*, 770–778. IEEE.

He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; and Jiang, W. 2021. TransReID: Transformer-based Object Re-Identification. In *ICCV*, 14993–15002. IEEE.

Hirzer, M.; Beleznai, C.; Roth, P. M.; and Bischof, H. 2011. Person re-identification by descriptive and discriminative classification. In *Image Analysis*, 91–102. Springer.

Huang, Z.; Zhang, Z.; Lan, C.; Zeng, W.; Chu, P.; You, Q.; Wang, J.; Liu, Z.; and Zha, Z.-j. 2022. Lifelong Unsupervised Domain Adaptive Person Re-identification with Coordinated Anti-forgetting and Adaptation. In *CVPR*, 14288–14297. IEEE.

Li, W.; and Wang, X. 2013. Locally aligned feature transforms across views. In *CVPR*, 3594–3601.

Li, W.; Zhao, R.; and Wang, X. 2012. Human Reidentification with Transferred Metric Learning. In *ACCV*, 31–44. Springer.

Li, W.; Zhao, R.; Xiao, T.; and Wang, X. 2014. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In *CVPR*, 152–159. IEEE.

Li, W.; Zhu, X.; and Gong, S. 2018. Harmonious Attention Network for Person Re-identification. In *CVPR*, 2285–2294. IEEE.

Li, Z.; and Hoiem, D. 2017. Learning without Forgetting. *PAMI*, 40(12): 2935–2947.

Liang, M.; Zhou, J.; Wei, W.; and Wu, Y. 2022. Balancing between forgetting and acquisition in incremental subpopulation learning. In *ECCV*, 364–380. Springer.

Lin, G.; Chu, H.; and Lai, H. 2022. Towards Better Plasticity-Stability Trade-off in Incremental Learning: A simple Linear Connector. In *CVPR*, 89–98.

Liu, H.; Feng, J.; Qi, M.; Jiang, J.; and Yan, S. 2017. End-to-end Comparative Attention Networks for Person Re-identification. *TIP*, 26(7): 3492–3506.

Liu, J.; Zha, Z.-J.; Chen, D.; Hong, R.; and Wang, M. 2019. Adaptive Transfer Network for Cross-Domain Person Re-Identification. In *CVPR*, 7195–7204. IEEE.

Loy, C. C.; Xiang, T.; and 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.; and Jiang, W. 2019. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. In *CVPRW*, 1487–1495. IEEE.

Park, W.; Kim, D.; Lu, Y.; and Cho, M. 2019. Relational knowledge distillation. In *CVPR*, 3967–3976.

Pu, N.; Chen, W.; Liu, Y.; Bakker, E. M.; and Lew, M. S. 2021. Lifelong Person Re-Identification via Adaptive Knowledge Accumulation. In *CVPR*, 7897–7906. IEEE.

Pu, N.; Liu, Y.; Chen, W.; Bakker, E. M.; and Lew, M. S. 2022. Meta reconciliation normalization for lifelong person re-identification. In *ACMM*, 541–549.

Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; and Tomasi, C. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV*, 17–35. Springer.

Shmelkov, K.; Schmid, C.; and Alahari, K. 2017. Incremental Learning of Object Detectors without Catastrophic Forgetting. In *ICCV*, 3420–3429. IEEE.

Song, J.; Yang, Y.; Song, Y.-Z.; Xiang, T.; and Hospedales, T. M. 2019. Generalizable Person Re-Identification by Domain-Invariant Mapping Network. In *CVPR*, 719–728. IEEE.

Sun, Z.; and Mu, Y. 2022. Patch-based Knowledge Distillation for Lifelong Person Re-Identification. In *ACM MM*, 696–707.

Tung, F.; and Mori, G. 2019. Similarity-Preserving Knowledge Distillation. In *ICCV*, 1365–1374. IEEE.

Wang, W.; Yang, F.; Luo, Z.; and Li, S. 2022a. Generalized Person Re-identification by Locating and Eliminating Domain-Sensitive Features. In *ACCV*, 3258–3273.

Wang, Z.; Zhang, Z.; Ebrahimi, S.; Sun, R.; Zhang, H.; Lee, C.-Y.; Ren, X.; Su, G.; Perot, V.; Dy, J.; et al. 2022b. Dual-Prompt: Complementary Prompting for Rehearsal-free Continual Learning. arXiv:2204.04799.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person Transfer GAN to Bridge Domain Gap for Person Re-identification. In *CVPR*, 79–88. IEEE.

Wu, G.; and Gong, S. 2021. Generalising without Forgetting for Lifelong Person Re-identification. In *AAAI*, 2889–2897.

Xiao, T.; Li, S.; Wang, B.; Lin, L.; and Wang, X. 2016. End-to-end deep learning for person search. arXiv:1604.01850.

Yu, C.; Shi, Y.; Liu, Z.; Gao, S.; and Wang, J. 2023. Lifelong Person Re-Identification via Knowledge Refreshing and Consolidation. In *AAAI*, 3295–3303.

Zhao, B.; Tang, S.; Chen, D.; Bilen, H.; and Zhao, R. 2021a. Continual Representation Learning for Biometric Identification. In *WACV*, 1197–1207. IEEE.

Zhao, H.; Tian, M.; Sun, S.; Shao, J.; Yan, J.; Yi, S.; Wang, X.; and Tang, X. 2017. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion. In *CVPR*, 907–915. IEEE.

Zhao, Y.; Zhong, Z.; Yang, F.; Luo, Z.; Lin, Y.; Li, S.; and Sebe, N. 2021b. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In *CVPR*, 6277–6286.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *ICCV*, 1116–1124. IEEE.

Zhuang, Z.; Wei, L.; Xie, L.; Zhang, T.; Zhang, H.; Wu, H.; Ai, H.; and Tian, Q. 2020. Rethinking the Distribution Gap of Person Re-identification with Camera-based Batch Normalization. In *ECCV*, 140–157. Springer.