Prototype-guided Knowledge Propagation with Adaptive Learning for Lifelong Person Re-identification

Zhijie Lu[1] , **Wuxuan Shi**[1] , **He Li**[1*] , **Mang Ye**[1,2*]
[1]School of Computer Science, Wuhan University, Wuhan, China
[2] Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, China
{zhijielu, wuxuanshi, lihe404, yemang}@whu.edu.cn

Abstract

Lifelong Person Re-identification (LReID) is essential in dynamic camera networks, which continually adapts to new environments while preserving previously acquired knowledge. Existing LReID techniques often preserve samples from past datasets to maintain old knowledge, potentially leading to privacy risks. While prototype-based methods offer privacy advantages, current approaches primarily focus on adjusting classifiers for image classification tasks, neglecting representation biases between old and new identities in person re-identification. This study introduces a novel Prototype-guided Knowledge Propagation (PKP) method, which mitigates discrepancies in similar identity images between old and new tasks by guiding prototype construction through triplet loss constraints. Additionally, to address disparities between prototypes and the updated feature extractor, an Adaptive Parameter Evolution (APE) strategy is proposed. APE optimizes the integration of the old and new models by assessing the importance of the new tasks, dynamically selecting the most pertinent parameters for updates according to their contribution to the current task. Extensive experiments on the LReID benchmark demonstrate that our approach surpasses state-of-the-art prototype-based LReID methods in terms of mAP and rank-1 accuracy. Code is available at https://github.com/joyner-7/IJCAI2025-PKA.

1 Introduction

Person re-identification (ReID) is a fundamental task in computer vision that aims to match the same person across different locations and times [Ye *et al.*, 2021; Leng *et al.*, 2019; Ye *et al.*, 2024]. Traditional ReID methods have achieved outstanding results by leveraging deep learning models and large-scale static datasets, where all training data are available simultaneously [Li *et al.*, 2024; Dai *et al.*, 2018; Zhang *et al.*, 2016; Ye *et al.*, 2018]. However, in real-world scenarios, such as surveillance systems that generate continuous streaming data, these models face significant challenges due to the inability to handle incremental and dynamic data effectively [Ge *et al.*, 2022; Wu and Gong, 2021]. This limitation has motivated the emergence of Lifelong Person Re-identification (LReID), which aims to enable ReID models to acquire new knowledge from streaming data while retaining previously learned knowledge.

The primary challenge in LReID lies in addressing catastrophic forgetting, a phenomenon common in lifelong learning tasks. This issue is particularly pronounced in LReID due to the unique characteristics of the task. First, as a fine-grained classification problem, the intra-person variations caused by temporal, environmental, and camera view changes are often significant[Ye *et al.*, 2023]. Second, subtle inter-person differences can lead to severe distribution overlaps, making it difficult to preserve discriminative knowledge for each individual. These factors exacerbate the forgetting of previously learned knowledge when learning new data.

To tackle catastrophic forgetting, most existing LReID methods employ additional memory to store exemplar data from previous tasks, which can be reused during training with new datasets [Ge *et al.*, 2022; Wu and Gong, 2021; Yu *et al.*, 2023a]. However, such memory-based approaches raise privacy concerns and introduce additional computational overhead [Wu *et al.*, 2025]. These limitations are particularly acute given the private nature of pedestrian images. Some methods replaced sample-based approaches in Class-Incremental Learning with prototype-based methods to solve privacy and memory issues [Xu *et al.*, 2024a]. However, these methods struggle to adapt to the training process of LReID, which is designed as a retrieval task and places greater emphasis on the embedding capability of the feature extractor. This requires constructing a more discriminative feature space that is effective for both seen and unseen domains. Therefore, prototype-based methods commonly used in CIL face difficulties in achieving satisfactory performance in LReID. After completing each training task, existing LReID methods merge the old model with the new one to ensure compatibility between stored prototypes and the updated feature extractor. However, the static merging strategy ignores the unique characteristics of each task. This approach struggles to balance old knowledge retention and new knowledge learning during training.

In this paper, we propose a novel non-exemplar-based

*Corresponding authors.

method for LReID, which provides an effective solution to privacy and memory constraints commonly faced in LReID tasks. When a new task is introduced, the model compares the new features to the existing prototypes. If the new features are similar to the old prototypes, the model pushes them apart, ensuring a clear distinction between old and new knowledge. By creating a clear separation, this approach not only safeguards previously learned knowledge but also sharpens the differentiation between old and new identity features, thus enabling better propagation of knowledge within the model over the entire training period and maintaining powerful embedding representation capability. In addition, we introduce an Adaptive Parameter Evolution (APE) strategy. APE evaluates the parameters in the model to assess which ones have a greater impact on the current training task and selectively updates them. And it dynamically evaluates the impact of new tasks, and based on this evaluation, it updates the model fusion method to better align with the requirements of the current task. This approach ensures robust alignment between all prototypes and the feature extractor while maintaining a high level of compatibility between previously acquired and newly learned knowledge throughout the training process.

Our contributions are summarized as follows:

- We propose a non-exemplar-based LReID method that constructs prototypes to mitigate catastrophic forgetting while addressing privacy and memory concerns.
- We introduce an Adaptive Parameter Evolution (APE) strategy that dynamically integrates old and new knowledge by assessing task variations and selectively updating parameters, enhancing the adaptability of the model.
- Our method achieves superior performance on benchmark datasets, demonstrating its effectiveness and setting a new standard for LReID tasks.

2 Related Work

2.1 Lifelong Person Re-identification

Lifelong person re-identification (LReID) is an emerging area that seeks to enable ReID models to learn continuously from non-stationary data streams, a more realistic scenario than the traditional static batch learning setup [Pu *et al.*, 2021]. The core challenge in LReID is mitigating catastrophic forgetting, which refers to the tendency of neural networks to rapidly forget previously learned knowledge when trained on new tasks. This phenomenon is particularly pronounced in LReID due to the fine-grained nature of the task and the variability of person appearances over time and across different environments. Current research in LReID can be broadly categorized into two main branches: rehearsal-based mthods and knowledge distillation-based methods. Rehearsal-based approaches mitigate forgetting by storing exemplar images from previous tasks and replaying these during the training process of new tasks [Ge *et al.*, 2022; Huang *et al.*, 2022; Yu *et al.*, 2023b; Wu and Gong, 2021]. While effective, this approach raises practical concerns, such as privacy issues related to storing human images, and scalability problems due to the growing memory requirements. Knowledge distillation-based methods preserve past knowledge by enforcing consistency between the outputs of old and new models [Huang *et al.*, 2023; Pu *et al.*, 2022; Pu *et al.*, 2023]. While these methods have shown promise in terms of anti-forgetting capability, the strict consistency constraints may hinder the plasticity of the model, limiting its ability to effectively learn new and potentially different data distributions [Xu *et al.*, 2024a]. Our work aims to address the challenges of both branches by exploring a non-exemplar based approach that utilizes prototypes to represent previously learned knowledge, thereby mitigating forgetting while avoiding data storage issues.

2.2 Prototype-based Class Incremental Learning

In the field of Class Incremental Learning (CIL), many prototype-based methods have been proposed, which are distinguished by their ability to avoid storing historical samples, significantly reducing storage overhead and mitigating potential privacy concerns. Prototypes are typically derived by averaging or aggregating the features of instances belonging to the same class, and have demonstrated their effectiveness in tasks such as few-shot learning and clustering. In CIL, prototypes are widely employed to represent past knowledge, where prototypes are mostly used for classifier calibration [Zhu *et al.*, 2021; Goswami *et al.*, 2024] or knowledge distillation at the output level [Zhu *et al.*, 2022; Shi and Ye, 2024]. While some attempts have been made to adapt prototype-based methods for LReID [Xu *et al.*, 2024a], these approaches remain inadequate as they often fail to fully address the unique requirements of LReID, such as the need for fine-grained feature discrimination and robust feature transfer across incremental tasks. Directly applying CIL methods often leads to suboptimal performance, as they overlook the critical need for effective and robust feature-based knowledge transfer that preserves the fine-grained discriminative capabilities essential for accurate and reliable performance in complex and dynamic retrieval scenarios.

3 Method

3.1 Problem Formulation

We address the challenge of Non-Exemplar Lifelong Person Re-identification. Formally, we are given a stream of sequential training datasets $\mathcal{D} = \{D_1, D_2, \ldots, D_T\}$, where T represents the total number of tasks. Each dataset $D_t = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$ contains N_t samples with corresponding identity labels. During the training phase for the t-th task, access to previous datasets $\{D_1, \ldots, D_{t-1}\}$ is restricted due to privacy considerations. To mitigate the problem of catastrophic forgetting, we construct a prototype set $\mathcal{P}_t = \{p_i\}_{i=1}^{N_t}$ for the t-th task. Each prototype in $\mathcal{P}_t$ represents a unique identity, and it is constructed by averaging the features of the corresponding identity.

3.2 Overview

Our proposed method, Prototype-guided Knowledge Propagation with Adaptive Learning (PKA), addresses Lifelong Person Re-identification (LReID) through two key components: Prototype-guided Knowledge Propagation (PKP) and Adaptive Parameter Evolution (APE). At each training stage t with dataset D_t, The sampled prototypes, enhanced with

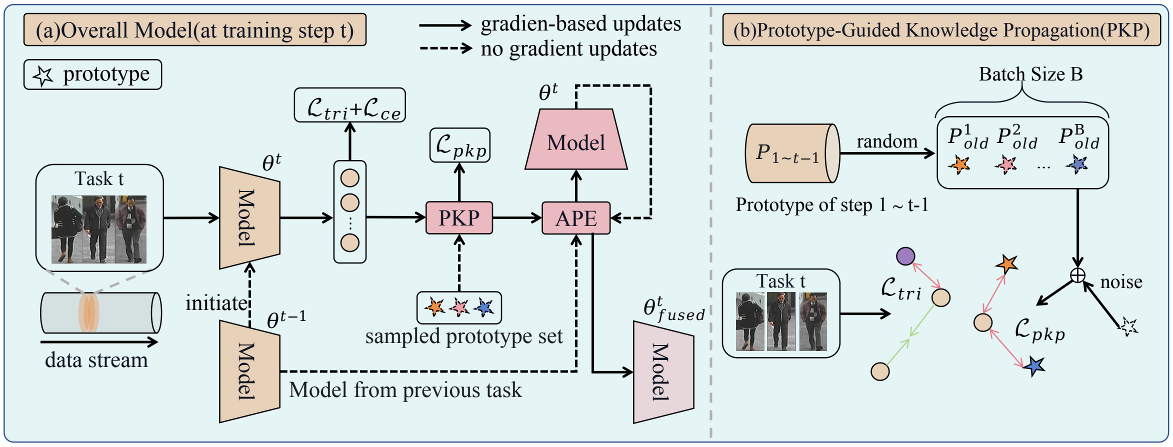


Figure 1: Overview of the proposed model Architecture. Solid arrows indicate the gradient-based updates, and dashed arrows represent no gradient updates. Prototype-guided Knowledge Propagation (PKP) module uses a modified triplet loss $\mathcal{L}_{pkp}$ and a standard triplet loss $\mathcal{L}_{tri}$ to propagate knowledge while ensuring discriminability. Different colors in the figure represent features or prototypes associated with different identities.

added noise, are passed through the PKP module. This process encourages them to diverge from the current task features, enabling the extraction of more discriminative feature embeddings while effectively leveraging prior knowledge, as illustrated in Fig. 1.

The APE module dynamically manages model parameters, assessing the relevance of D_t and selecting parameters based on their impact on the current task. It then fuses the new model's parameters (θ_t) with those of the previous model (θ_{t-1}), as shown by the dashed arrows in Fig. 1. This ensures adaptive evolution of model parameters while retaining past knowledge, resulting in a new model θ_{fused} for the next training stage $t+1$.

3.3 Prototype-guided Knowledge Propagation

To mitigate catastrophic forgetting in lifelong person reidentification (LReID), we propose a novel prototype-based non-exemplar learning paradigm. our approach introduces a novel perspective by leveraging prototypes to guide both knowledge propagation and feature learning for new tasks, as illustrated in Fig. 2. we generate more discriminative embeddings, which in turn improves retrieval performance for the LReID model.

Existing methods that often employ triplet loss directly on input features of new tasks for feature discrimination. Our method aims to leverage prototypes to push apart identities from previous tasks and new tasks within the embedding space, creating a clear distinction. This facilitates the generation of more refined embeddings. To achieve this, we define a prototype set $P = \{p_1, \ldots, p_M\}$. During training for a new task, we randomly sample a subset P_s from it, where P_s contains prototypes of size half the batch size. To enhance the generalization and robustness of these prototypes, we add Gaussian noise, resulting in augmented prototypes $\tilde{P}_s$. Specifically, for each prototype $p_l \in P_s$, we add Gaussian noise ϵ:

$$\tilde{p}_l = p_l + \beta\epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I). \tag{1}$$

Here $\tilde{p}_l$ represents the augmented prototype, ϵ is drawn from a Gaussian distribution with zero mean and covariance matrix $\sigma^2 I$, and β is a hyperparameter controlling the magnitude of the noise. The prototypes enhanced with noise can cover a broader feature space during training, preventing the prototypes from becoming too concentrated. This enables a more comprehensive separation of the features distributed around them in the new task, thereby improving clustering among distinct classes and optimizing the representation capability of the embedding space.

We utilize a modified triplet loss to encourage separation between features from the new task and augmented prototypes. The standard triplet loss, as defined in [Sun and Mu, 2022; Yu *et al.*, 2023a; Schroff *et al.*, 2015], serves as our foundation:

$$L_{tri} = \max(0, \|a-p\|_2^2 + \alpha - \|a-n\|_2^2). \tag{2}$$

Here a represents the anchor feature, p represents the positive feature, and n represents the negative feature, all drawn from the features of the new task. Triplet loss aims to reduce the distance between same-identity embeddings and increase the distance between embeddings of different identities, thereby enhancing the model's ability to distinguish between them and improving retrieval performance.

Our method incorporates two loss terms. The first term focuses on pushing the features of the new task away from the augmented prototypes to guarantee the discrimination between old and new tasks. To achieve this, the triplet loss is modified by removing the positive sample. The following loss term is used, which aims to maximize the distance between the anchor a (new task feature) and the negative sample n (augmented prototype):

$$L_{pkp} = \frac{1}{N_1}\sum_{i=1}^{N_1} \max(0, \gamma - \|a-n\|_2^2). \tag{3}$$

Here γ is a margin to enforces a minimum distance between the new task features and the augmented prototypes, and N_1

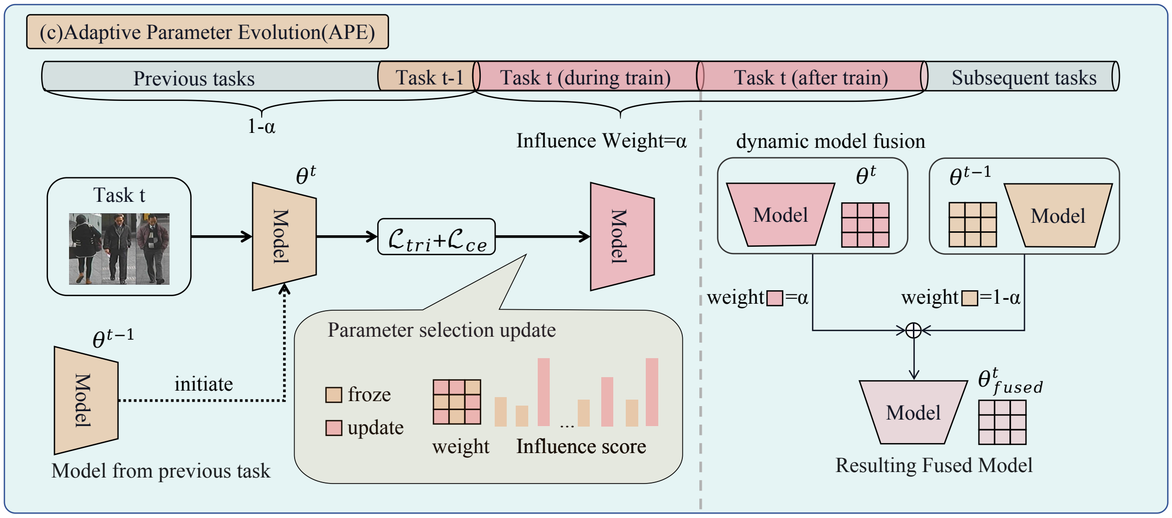


Figure 2: The Adaptive Parameter Evolution (APE) strategy employs gradient-informed parameter selection and dynamic model parameter fusion. We employ gradient-informed selection (left) to update high-influence parameters and performs dynamic model fusion (right) with weight α. The resulting model is then used for the next task.

represents the number of triplets sampled from the augmented prototypes. This loss function is denoted as L_{pkp}. The second term ensures the discrimination between different identities within the new task. We use the following standard triplet loss function:

$$L_{tri} = \frac{1}{N_2} \sum_{j=1}^{N_2} \max \left(0, \|z_j^t - z_p^t\|_2^2 + \alpha - \|z_j^t - z_n^t\|_2^2\right). \tag{4}$$

Here z_k^t represents the feature of the k-th sample in the new task, and z_p^t and z_n^t are the positive and negative samples, respectively, within the new task. N_2 is the number of triplets sampled from the new task.

3.4 Adaptive Parameter Evolution Strategy

To further facilitate effective knowledge propagation in lifelong person re-identification, we introduce the Adaptive Parameter Evolution (APE) strategy. a dynamic mechanism that orchestrates the evolution of model parameters in response to the ongoing learning process. APE is characterized by two synergistic components: gradient-informed parameter selection and dynamic model parameter fusion. These are designed to maintain the alignment between previously learned prototypes and the retrained feature extractor, while balancing the retention of old knowledge and the acquisition of new knowledge throughout the training process.

We posit that not all parameters are equally relevant to a given learning task, and that a more judicious parameter update strategy is needed [Zhang *et al.*, 2024]. To this end, we compute the gradient of the loss function $\mathcal{L}$ with respect to each model parameter θ_i, denoted by $\nabla_{\theta_i}\mathcal{L}$. These gradients quantify the sensitivity of the loss to each parameter, providing a measure of the parameter's relevance to the task. The absolute value of the gradient is then computed as $|\nabla_{\theta_i}\mathcal{L}|$. Instead of naively updating all parameters, we introduce a threshold τ. We selectively update a parameter only if its gradient magnitude is greater than the threshold τ, otherwise, the parameter is preserved. This process can be expressed as:

$$\theta_i \leftarrow \begin{cases} \theta_i - \eta\nabla_{\theta_i}\mathcal{L}, & \text{if } |\nabla_{\theta_i}\mathcal{L}| > \tau \\ \theta_i, & \text{if } |\nabla_{\theta_i}\mathcal{L}| \leq \tau \end{cases}, \tag{5}$$

where η is the learning rate and τ is a predefined gradient magnitude threshold. This approach not only enhances the efficiency of the training process by focusing on the parameters that are most influential for the current task, but also mitigates the risk of overfitting by preserving the parameters that are less sensitive to the current task.

After gradient descent training, dynamic model parameter fusion is performed, which emphasizes maintaining the compatibility between new and old knowledge. Rather than applying a static fusion strategy [Xu *et al.*, 2024a], APE dynamically adjusts the fusion weight according to the relative influence of the current task [Xiao *et al.*, 2023], which is estimated by the size of its training dataset. This allows us to overcome the shortcomings of previous methods which are less sensitive to variation between different tasks. After training a new task t, θ_{new}^t and θ_{old}^{t-1} represent the parameters of the new and old models, respectively. The fused model parameters θ_{fused}^t are determined as:

$$\theta_{fused}^t = \alpha_t\theta_{new}^t + (1-\alpha_t)\theta_{old}^{t-1}. \tag{6}$$

The fusion weight α_t is dynamically determined based on the relative impact of the current task's dataset. We recognize that datasets with larger sizes typically offer a more comprehensive representation of the task, hence they should exert a greater influence on the model parameters. This influence is quantified using the dataset size, therefore, α_t is defined as:

$$\alpha_t = \frac{N_t}{\sum_{i=1}^{t} N_i}, \tag{7}$$

where N_t denotes the number of samples in the training dataset for task t.

	Method	Non-Exemplar	Market1501		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
CIL	Joint-Train [Xu *et al.*, 2024a]		75.3	90.1	84.5	86.0	66.9	81.6	31.6	57.1	58.5	61.4	63.4	75.2	55.2	48.2
	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
	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
	PRAKA* [Shi and Ye, 2023]	✓	37.4	61.3	69.3	71.8	35.4	55.0	10.7	27.2	**54.0**	**55.6**	41.3	54.2	47.7	41.6
	PRD* [Asadi *et al.*, 2023]		7.3	18.0	33.5	35.6	3.7	7.6	0.8	2.4	33.8	33.8	15.8	19.5	23.0	17.7
LReID	CRL [Xu *et al.*, 2024b]		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]		58.1	77.4	72.5	74.8	28.7	45.2	6.1	16.2	38.7	40.4	40.8	50.8	47.6	42.6
	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
	MEGE [Pu *et al.*, 2023]		39.0	61.6	73.3	76.6	16.9	30.3	4.6	13.4	36.4	37.1	34.0	43.8	47.7	44.0
	DKP [Xu *et al.*, 2024a]	✓	**60.3**	80.6	83.6	85.4	51.6	68.4	19.7	41.8	43.6	44.2	51.8	64.1	59.2	51.6
	PKA(Ours)	✓	57.7	80.2	**85.0**	**86.9**	**59.4**	**75.6**	**31.0**	**56.1**	44.1	44.6	**55.4**	**68.7**	**62.3**	**55.3**

Table 1: Training Order-1: Market-1501 → CUHK-SYSU → DukeMTMC-ReID → MSMT17-V2 → CUHK03. * denotes the results are reproduced by the released official code. † denotes the results reported by [Sun and Mu, 2022]. The best and second-best results are marked in **bold** and underlined.

	Method	Non-Exemplar	DukeMTMC		MSMT17		Market1501		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
CIL	Joint-Train[Xu *et al.*, 2024a]		66.9	81.6	31.6	57.1	75.3	90.1	84.5	86.0	58.5	61.4	63.4	75.2	55.2	48.2
	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
	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
	PRAKA* [Shi and Ye, 2023]	✓	31.2	48.7	6.6	19.1	47.8	69.8	70.4	73.0	54.9	56.6	42.2	53.4	48.4	41.1
	PRD*[Asadi *et al.*, 2023]		3.6	8.2	0.6	1.8	8.9	22.3	34.6	36.1	35.4	35.3	16.6	20.7	20.7	15.0
LReID	CRL[Xu *et al.*, 2024b]		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]		42.2	60.1	5.4	15.1	37.2	59.8	71.2	73.9	36.9	37.9	38.6	49.4	46.0	41.7
	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
	MEGE[Pu *et al.*, 2023]		21.6	35.5	3.0	9.3	25.0	49.8	69.9	73.1	34.7	35.1	30.8	40.6	44.3	41.1
	DKP[Xu *et al.*, 2024a]	✓	53.4	70.5	14.3	33.3	60.6	81.0	83.0	84.9	**45.0**	46.1	51.3	63.2	59.0	51.6
	PKA(Ours)	✓	54.2	70.6	**24.3**	**48.0**	**68.6**	**85.3**	**85.0**	**86.8**	43.8	**46.9**	**55.2**	**67.5**	**60.3**	**52.8**

Table 2: Training Order-2: DukeMTMC-reID → MSMT17-V2→ Market-1501 → CUHK-SYSU → CUHK03. *denotes the results are reproduced by the released official code. † denotes the results reported by [Sun and Mu, 2022]. The best and second-best results are marked in **bold** and underlined.

By coupling gradient-informed parameter selection with dynamic parameter fusion, APE provides an adaptive and robust mechanism for lifelong learning, enabling the previous prototype set to maintain strong embedding representation capability in the new feature space, resulting in improved retrieval performance across all datasets.

3.5 Overall Loss Function

To optimize our model, we employ a composite loss function with three distinct components: a cross-entropy loss (L_{ce}), a triplet loss on the combined features of new samples and augmented prototypes (L_{pkp}), and a standard triplet loss on new task features (L_{tri}). The cross-entropy loss is defined as:

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}), \tag{8}$$

where N is the number of samples, C is the number of classes, y_{ic} is the true label (0 or 1), and $\hat{y}_{ic}$ is the predicted probability for class c of sample i. The overall loss function is defined as:

$$L = L_{ce} + \alpha(L_{pkp} + L_{tri}). \tag{9}$$

Here α is a variable weight controlling the influence of the two triplet loss components, ensuring a balance between preserving old knowledge and acquiring new information. This weight α is shared between the two different triplet losses. This combined loss function allows our model to learn discriminative feature representations that are robust to the challenges of lifelong learning in person re-identification tasks.

4 Experiments

4.1 Experimental Settings

Datasets. To evaluate the effectiveness of our proposed method, we conduct extensive experiments on five benchmark lifelong person ReID datasets:Market-1501 [Zheng *et al.*, 2015], CUHKSYSU [Xiao *et al.*, 2017], DukeMTMC-ReID [Ristani *et al.*, 2016], MSMT17-V2 [Wei *et al.*, 2018], and CUHK03 [Li *et al.*, 2014]. To simulate a lifelong person ReID scenario in real-world settings, we evaluate our method using two training orders as specified in, namely: Order-1: Market-1501 → CUHK-SYSU → DukeMTMC-ReID → MSMT17-V2 → CUHK03, and Order-2: DukeMTMC-ReID → MSMT17-V2 → Market-1501 → CUHK-SYSU → CUHK03. To further evaluate the generalization capacity of our models, we tested them on seven additional datasets (CUHK01 [Li *et al.*, 2013], CUHK02 [Li and Wang, 2013], VIPeR [Gray and Tao, 2008], PRID [Hirzer *et al.*, 2011], i-LIDS [Branch, 2006], GRID [Loy *et al.*, 2010], and SenseReID [Zhao *et al.*, 2017]) as unseen domains.

Evaluation Metrics. The mean Average Precision (mAP) and rank@1 accuracy (R@1) are used to evaluate the model performance on individual datasets. In addition, the average mAP and the average R@1 in all seen and unseen domains are calculated to assess and compare the overall lifelong learning and generalization capabilities of the models, respectively.

4.2 Comparison with State-of-the-arts Methods

To comprehensively evaluate our method, we compare it against various state-of-the-art(SOTA) non-exemplar LReID approaches, including DKP [Xu *et al.*, 2024b], AKA [Pu *et al.*, 2021], PatchKD [Sun and Mu, 2022], and MEGE [Pu *et al.*, 2023]. Additionally, several class-incremental learning (CIL) methods, such as LwF [Li and Hoiem, 2017], SPD [Tung and Mori, 2019], PRAKA [Shi and Ye, 2023], and PRD [Asadi *et al.*, 2023], are also tested. To ensure a fair comparison, all models are implemented using the same backbone and training configurations. Furthermore, we report the results of Joint-Train, which represents the upper bound for LReID models by assuming access to all datasets simultaneously during training.

We present the results of different methods on each seen domain, as well as the average performance across all seen domains (Seen-Avg) in Tab. 1 and 2, corresponding to Training Order-1 and Training Order-2, respectively.

Seen-Domain Performance Evaluation. As shown in Tab.1 and Tab.2, our PKA significantly outperforms all existing LReID models. Compared to the second-best method, DKP, our model achieves an improvement of 3.6%/4.6% and 3.9%/4.3% on the average mAP/R@1 performance for seen domains. The performance of DKP deteriorates in later training stages, likely because it adopts a static fusion strategy between the old and new models, failing to properly balance the storage of old knowledge and the acquisition of new knowledge. And our PKA demonstrates an average mAP/R@1 improvement of 14.1%/13.5% and 13%/14.1% over the best CIL method, PRAKA, across seen domains in both training orders. This indicates that by utilizing prototypes in LReID, we have successfully enhanced the propagation of old knowledge within the model, effectively mitigating the catastrophic forgetting problem, while also improving the model's generalization ability.

It should be noted that the MSMT17 dataset presents inherent challenges due to substantial variations in weather and lighting conditions. Nevertheless, our approach significantly outperforms existing methods on this dataset, achieving a notably higher mAP and Rank-1 accuracy, which underscores the superior effectiveness of our method.

To visually understand the performance trends of different models in the seen domain, we present the performance curves over different training steps in Fig 3. As can be seen in the figure, our model PKA consistently achieves the highest mAP and R@1 compared with the other methods. Although the initial performance of PKA may not be the top among these methods, it emphasizes the propagation and application of prior knowledge, granting it superior resistance to forgetting. Consequently, PKA outperforms others in later tests and exhibits more prominent performance.

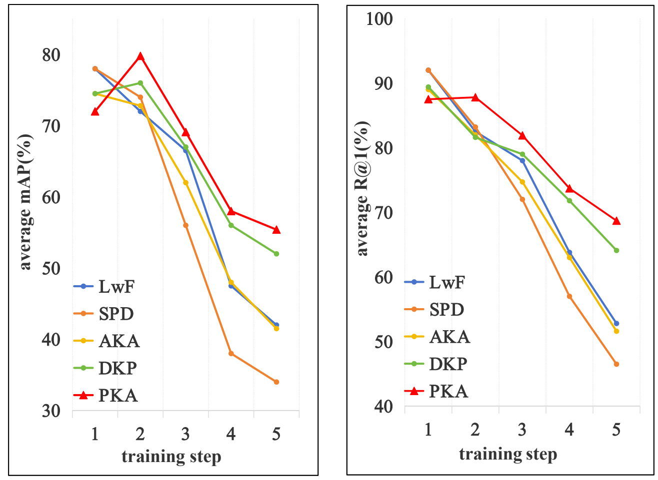


Figure 3: Illustration of performance trend on previously seen domains. After each training phase, the model is evaluated on domains it has encountered before.

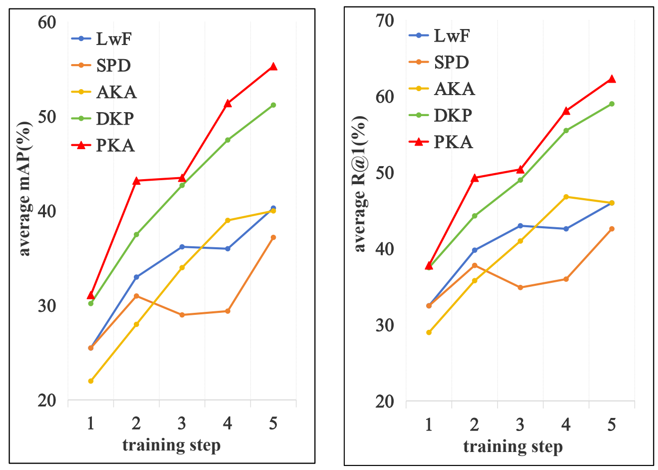


Figure 4: Illustration of performance trend on unseen domains. After each training phase, the model is evaluated on domains it has not encountered before.

Unseen-Domain Generalization Evaluation. The average performance in the unseen domains is shown in the last two columns of Tab.1 and Tab.2. Our method demonstrates superior generalization capabilities compared to SOTA CIL models, exhibiting an average mAP/R@1 improvement of 14.6%/13.7% and 11.9%/11.7% across both training orders. Furthermore, our model also significantly outperforms the SOTA LReID DKP models by a margin of 3.1%/3.7% and 1.3%/1.2% in average mAP/R@1 improvement. These results show that our model effectively consolidates more generalizable knowledge. To further evaluate the generalization ability of our model on unseen domains, we provide detailed performance curves over different training steps in Fig 4. The analysis of the curves indicates that our model not only achieves high performance but also maintains stable and consistent performance improvement even in unseen environments. This demonstrates that our proposed Prototype-guided Knowledge Propagation approach, compared to LReID train-

Baseline	PKP	APE	Seen-Avg		Unseen-Avg	
			mAP	R@1	mAP	R@1
✓			42.8	57.4	50.7	45.0
✓	✓		53.4	66.0	58.7	51.4
✓		✓	54.5	67.8	60.9	54.1
✓	✓	✓	**55.4**	**68.7**	**62.3**	**55.3**

Table 3: Ablation study of different components.

ing methods that focus solely on classifier calibration,is more effective in enhancing the model's capability to extract features and shape a better embedding distribution. As a result, it achieves more discriminative embedding representations on unseen datasets, with improved generalization and flexibility.

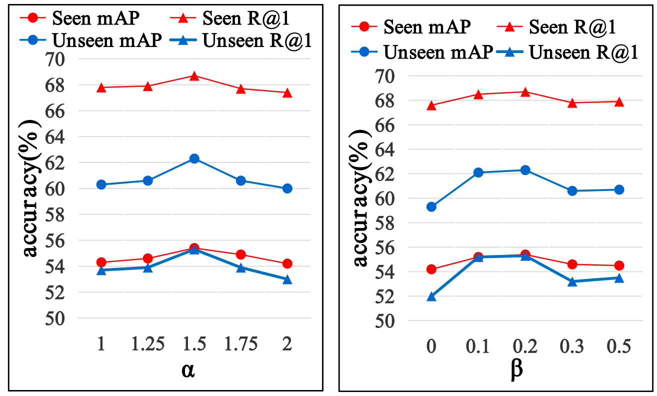
4.3 Ablation Studies

Influence of Different Components. In this section, we conduct several ablation studies on the proposed method. To better analyze the impact of the core designs, our approach is divided into two components:Prototype-guided Knowledge Propagation (PKP) and Adaptive Parameter Evolution (APE). The baseline model refers to a framework that does not incorporate the PKP and APE modules, meaning it does not utilize prototypes for knowledge propagation. Instead, after completing each training task, the old model and the new model are directly integrated using a static fusion strategy. Both the PKP and APE modules excel at propagating prior knowledge, significantly improving performance on both seen and unseen domains. As shown in Tab. 3, furthermore, the APE module facilitates the transfer of prototype knowledge constructed by the PKP module. Consequently, performance is further improved when both modules are used together.

Influence of Hyperparameters. To analyze the impact of hyperparameters on our method, we conducted experiments to evaluate the effects of different hyperparameter settings. We analyzed the effects of the weights of L_{tri} and L_{pkp}. Specifically, L_{tri} focuses on distinguishing the distributions among new data, while L_{pkp} emphasizes the separation between the distributions of new input data and the prototypes of old data. We set the weights of these two losses to the same value, and based on the results shown in Fig. 5(a), we choose $\alpha = 1.5$ as the default setting. In addition, we augmented the prototypes by adding random noise to enhance the model's ability to transfer old knowledge through prototypes. Appropriate noise augmentation can significantly improve performance on unseen domains, whereas excessive noise can degrade performance on both seen and unseen domains. As shown in Fig. 5(b) we choose $\beta = 0.2$ as the default setting.

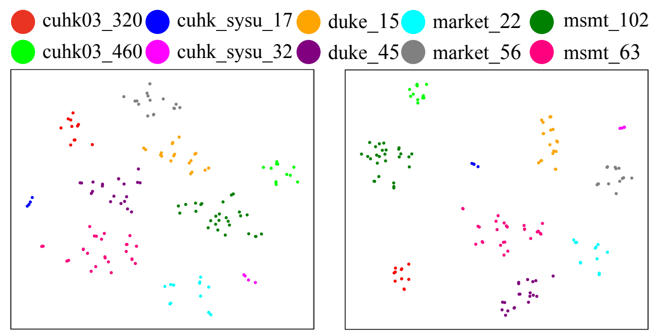
4.4 Visualization Results

To further analyze the impact of our PKA method on the feature space, we employ t-SNE to visualize features of selected identities. As shown in Fig. 6, two identities from each of the five seen datasets are selected. Comparing Fig. 6(a) (baseline) with Fig. 6(b) (PKA), we observe that the PKA model yields more compact and separable clusters. In the baseline, clusters are scattered and overlapping, making inter-class distinction difficult. In contrast, PKA produces tighter and more distinct clusters, indicating that it learns more discriminative features beneficial for classification and analysis.



(a)The weight of L_{pkp} and L_{tri} (b)Noise coefficient scale

Figure 5: Ablation studies on hyperparameters.



(a) (b)

Figure 6: t-SNE results of our PKA compared with the Baseline. Different colors represent different identities, highlighting the class-wise aggregation and the distinction between different categories. The visualization demonstrates how our PKA improves intra-class cohesion and inter-class separation.

5 Conclusion

In this paper, we tackle the challenging task of LReID by introducing PKA, a novel non-exemplar-based approach. PKA mitigates catastrophic forgetting through Prototype-guided Knowledge Propagation (PKP), which utilizes prototypes and triplet loss to preserve and transfer knowledge, and Adaptive Parameter Evolution (APE) to enable dynamic model updates for task adaptation. Extensive experiments on five benchmark datasets demonstrate PKA's effectiveness, achieving notable improvements in mean Average Precision (mAP) and rank-1 accuracy, along with enhanced generalization. These results highlight the potential of PKA as a robust, privacy-preserving, and practical solution for LReID.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China under Grant (62176188, 62361166629, 62225113) and Key Research and Development Project of Hubei Province (2022BAD175). The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Contribution Statement

Zhijie Lu and Wuxuan Shi contribute equally to this work.

References

[Asadi *et al.*, 2023] Nader Asadi, MohammadReza Davari, Sudhir Mudur, Rahaf Aljundi, and Eugene Belilovsky. Prototype-sample relation distillation: towards replay-free continual learning. In *International Conference on Machine Learning*, pages 1093–1106. PMLR, 2023.

[Branch, 2006] Home Office Scientific Development Branch. Imagery library for intelligent detection systems (i-lids). In *2006 IET conference on crime and security*, pages 445–448. IET, 2006.

[Dai *et al.*, 2018] Pingyang Dai, Rongrong Ji, Haibin Wang, Qiong Wu, and Yuyu Huang. Cross-modality person re-identification with generative adversarial training. In *IJCAI*, volume 1, page 6, 2018.

[Ge *et al.*, 2022] Wenhang Ge, Junlong Du, Ancong Wu, Yuqiao Xian, Ke Yan, Feiyue Huang, and Wei-Shi Zheng. Lifelong person re-identification by pseudo task knowledge preservation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 688–696, 2022.

[Goswami *et al.*, 2024] Dipam Goswami, Yuyang Liu, Bartłomiej Twardowski, and Joost van de Weijer. Fecam: Exploiting the heterogeneity of class distributions in exemplar-free continual learning. *Advances in Neural Information Processing Systems*, 36, 2024.

[Gray and Tao, 2008] Douglas Gray and Hai Tao. 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*, pages 262–275. Springer, 2008.

[Hirzer *et al.*, 2011] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17*, pages 91–102. Springer, 2011.

[Huang *et al.*, 2022] Zhipeng Huang, Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, and Zheng-jun Zha. Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14288–14297, 2022.

[Huang *et al.*, 2023] Jinze Huang, Xiaohan Yu, Dong An, Yaoguang Wei, Xiao Bai, Jin Zheng, Chen Wang, and Jun Zhou. Learning consistent region features for lifelong person re-identification. *Pattern Recognition*, 144:109837, 2023.

[Leng *et al.*, 2019] Qingming Leng, Mang Ye, and Qi Tian. A survey of open-world person re-identification. *IEEE TCSVT*, 30(4):1092–1108, 2019.

[Li and Hoiem, 2017] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

[Li and Wang, 2013] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3594–3601, 2013.

[Li *et al.*, 2013] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11*, pages 31–44. Springer, 2013.

[Li *et al.*, 2014] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 152–159, 2014.

[Li *et al.*, 2024] Zongyi Li, Jianbo Li, Yuxuan Shi, Hefei Ling, Jiazhong Chen, Runsheng Wang, and Shijuan Huang. Cross-modal generation and alignment via attribute-guided prompt for unsupervised text-based person retrieval. In *Proceedings of the International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization*, pages 1047–1055, 2024.

[Loy *et al.*, 2010] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision*, 90:106–129, 2010.

[Pu *et al.*, 2021] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7901–7910, 2021.

[Pu *et al.*, 2022] Nan Pu, Yu Liu, Wei Chen, Erwin M Bakker, and Michael S Lew. Meta reconciliation normalization for lifelong person re-identification. In *Proceedings of the 30th ACM international conference on multimedia*, pages 541–549, 2022.

[Pu *et al.*, 2023] Nan Pu, Zhun Zhong, Nicu Sebe, and Michael S Lew. A memorizing and generalizing framework for lifelong person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Ristani *et al.*, 2016] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European conference on computer vision*, pages 17–35. Springer, 2016.

[Schroff *et al.*, 2015] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[Shi and Ye, 2023] Wuxuan Shi and Mang Ye. Prototype reminiscence and augmented asymmetric knowledge aggregation for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1772–1781, 2023.

[Shi and Ye, 2024] Wuxuan Shi and Mang Ye. Prospective representation learning for non-exemplar class-incremental learning. In *NeurIPS*, 2024.

[Sun and Mu, 2022] Zhicheng Sun and Yadong Mu. Patch-based knowledge distillation for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 696–707, 2022.

[Tung and Mori, 2019] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1365–1374, 2019.

[Wei *et al.*, 2018] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79–88, 2018.

[Wu and Gong, 2021] Guile Wu and Shaogang Gong. Generalising without forgetting for lifelong person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 2889–2897, 2021.

[Wu *et al.*, 2025] Bin Wu, Wuxuan Shi, Jinqiao Wang, and Mang Ye. Synthetic data is an elegant gift for continual vision-language models. *arXiv preprint arXiv:2503.04229*, 2025.

[Xiao *et al.*, 2017] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3415–3424, 2017.

[Xiao *et al.*, 2023] Jia-Wen Xiao, Chang-Bin Zhang, Jiekang Feng, Xialei Liu, Joost van de Weijer, and Ming-Ming Cheng. Endpoints weight fusion for class incremental semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7204–7213, 2023.

[Xu *et al.*, 2024a] Kunlun Xu, Xu Zou, Yuxin Peng, and Jiahuan Zhou. Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification. In *CVPR*, pages 16604–16613, 2024.

[Xu *et al.*, 2024b] Kunlun Xu, Xu Zou, Yuxin Peng, and Jiahuan Zhou. Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16604–16613, 2024.

[Ye *et al.*, 2018] Mang Ye, Zheng Wang, Xiangyuan Lan, and Pong C Yuen. Visible thermal person re-identification via dual-constrained top-ranking. In *IJCAI*, volume 1, page 2, 2018.

[Ye *et al.*, 2021] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE TPAMI*, 44(6):2872–2893, 2021.

[Ye *et al.*, 2023] Mang Ye, Zesen Wu, Cuiqun Chen, and Bo Du. Channel augmentation for visible-infrared re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2299–2315, 2023.

[Ye *et al.*, 2024] Mang Ye, Shuoyi Chen, Chenyue Li, Wei-Shi Zheng, David Crandall, and Bo Du. Transformer for object re-identification: A survey, 2024.

[Yu *et al.*, 2023a] Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, and Jingya Wang. Lifelong person re-identification via knowledge refreshing and consolidation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3295–3303, 2023.

[Yu *et al.*, 2023b] Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, and Jingya Wang. Lifelong person re-identification via knowledge refreshing and consolidation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3295–3303, 2023.

[Zhang *et al.*, 2016] Yaqing Zhang, Xi Li, Liming Zhao, and Zhongfei Zhang. Semantics-aware deep correspondence structure learning for robust person re-identification. In *IJCAI*, pages 3545–3551, 2016.

[Zhang *et al.*, 2024] Wenxuan Zhang, Paul Janson, Rahaf Aljundi, and Mohamed Elhoseiny. Overcoming generic knowledge loss with selective parameter update. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24046–24056, 2024.

[Zhao *et al.*, 2017] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1077–1085, 2017.

[Zheng *et al.*, 2015] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015.

[Zhu *et al.*, 2021] Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5871–5880, 2021.

[Zhu *et al.*, 2022] Kai Zhu, Wei Zhai, Yang Cao, Jiebo Luo, and Zheng-Jun Zha. Self-sustaining representation expansion for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9296–9305, 2022.