

Unified Representation Causal Prompt Distillation for Re-Inference-Free Lifelong Person Re-Identification

Jiaqi Zhao^{1,2}, Jie Luo^{1,2}, Yong Zhou^{1,2,*}, Wen-Liang Du^{1,2}, Xixi Li^{1,2}, Rui Yao^{1,2}

¹School of Computer Science and Technology / School of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116, China

²Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China
{jiaqizhao, jieluo, yzhou, wldu, xixi, ruiyao}@cumt.edu.cn

Abstract

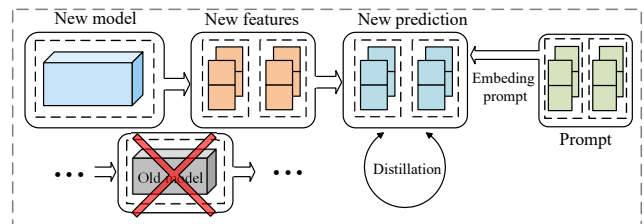
Lifelong person re-identification (LReID) aims to retrieve the target person from sequentially collected data. Due to significant domain gaps between datasets and the continuous increase of training data from different scenarios, weak inter-domain generalization and catastrophic forgetting issues have remained major challenges for LReID. To tackle these issues, a novel LReID method called Unified Representation Causal Prompt Distillation (URCPD) is proposed. Specifically, to reduce domain gaps among different scene datasets and improve model inter-domain generalization capability, a Feature Decoupling Style Transfer module (FDST) is proposed to map new features into a unified feature space. Furthermore, to reduce the accumulated forgetting of old knowledge during the training stage, a Causal Prompt Distillation module (CPD) is introduced. This module eliminates the re-inference process for distillation and embeds memory prompts to combat catastrophic forgetting. Extensive experiments on five classic LReID seen datasets and seven unseen datasets demonstrate that our method significantly outperforms state-of-the-art methods.

Code — <https://github.com/roger404com/URCPD>

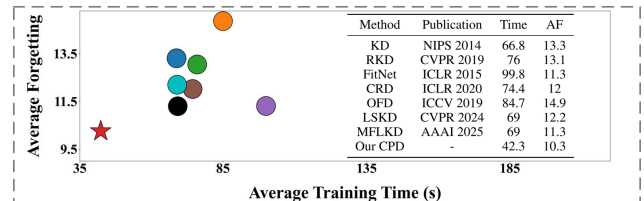
Introduction

Person re-identification (ReID) is a traditional downstream task in computer vision, aimed at retrieving target person from continuous data collected by surveillance cameras across different viewpoints. Currently, ReID has achieved remarkable performance on static datasets (Zhu et al. 2020; Fang, Yang, and Fu 2023). Building upon this foundation, variants such as multi-modal ReID (Zhao et al. 2023a,b; Yang, Chen, and Ye 2023), clothing-changing ReID (Bansal, Foresti, and Martinel 2022), occluded ReID (Liu et al. 2025b; Yuan et al. 2025), and text-based (Fu et al. 2025; Zhao et al. 2024) ReID have been proposed to address long-standing challenges in conventional ReID, such as illumination variations, low image resolution, and partial occlusions.

However, real-world scenarios present challenges beyond these problems. With the continuous improvement of public video surveillance systems, the volume of training data from



(a) Method Comparison



(b) Performance Comparison

Figure 1: (a) Our Causal Prompt Distillation eliminates the inference process of the old model in the new stage and embeds memory prompts to combat catastrophic forgetting. (b) Comparison of the average time consumption and average forgetting of our CPD with other distillation methods.

diverse scenarios has increased significantly. It is essential to consider not only reducing the training cost for new data, but also maintaining memory of old knowledge and respecting the access restrictions associated with old data. Consequently, increasing research efforts have shifted towards lifelong person re-identification (LReID) (Cui et al. 2024; Huang et al. 2022), which offers better adaptability to data, enhanced memory capabilities, and lower training costs. The AKA (Pu et al. 2021) algorithm provides a guiding paradigm for LReID. Nevertheless, due to the substantial environmental gaps between datasets and the minor differences among samples within a dataset, LReID continues to suffer from more severe catastrophic forgetting and weak generalization. To address these issues, we need to balance the plasticity and the stability of model, while simultaneously reducing inter-domain discrepancies. Most existing LReID methods mitigate catastrophic forgetting primarily through data replay and knowledge distillation. However, data replay requires storing historical samples, which conflicts with data privacy concerns inherent in LReID tasks. Regarding knowl-

*Corresponding author.

edge distillation, existing LReID methods typically employ three types: feature distillation, logit distillation and relation distillation. Crucially, all distillation methods necessitate retaining the old model and using it to perform inference on new data during each new learning phase. We call this process re-inference. To reduce computational overhead, we propose a novel LReID paradigm called re-inference-free LReID, which eliminates the re-inference process inherent in knowledge distillation (as shown in Fig. 1 (a)). We compare the average time consumption (The five training stages of LReID) and average forgetting (AF, the lower the better) (Chaudhry et al. 2018) of different knowledge distillation methods (as shown in Fig. 1 (b)). The comparison methods include: KD (Hinton 2015), RKD (Park et al. 2019a), FitNets (Romero et al. 2015), CRD (Tian, Krishnan, and Isola 2022), OFD (Heo et al. 2019), LSKD (Sun et al. 2024) and MFLKD (Gao et al. 2025).

Specifically, we introduce a new re-inference-free training strategy for LReID and propose a corresponding method named Unified Representation Causal Prompt Distillation (URCPD). We design a Feature Decoupling Style Transfer (FDST) module. This module transfers the style of new data into a unified style space, achieving unified cross-domain feature representation and thereby enhancing the model’s generalization ability. Furthermore, we propose a Causal Prompt Distillation (CPD) module, which generates memory prompts and leverages causal relationships to enable re-inference-free knowledge distillation. Within this module, we train a set of small parameters specifically as prompt parameters for memorizing old tasks. These prompt parameters are optimized using a separate optimizer, concurrently with the overall model training. During the distillation process, parameters elsewhere in the model are frozen, only the prompt parameters remain active. These prompts are generated and embedded into the image features to significantly boost the model’s resistance to forgetting. In general, our contributions are four folds:

- A new LReID paradigm called Unified Representation Causal Prompt Distillation (URCPD) for re-inference-free lifelong person re-Identification is introduced. This paradigm eliminates the reasoning process of the old model in knowledge distillation.
- A Feature Decoupling Style Transfer module (FDST) is proposed to improve model inter-domain generalization capability by transferring new data styles to a unified style space.
- A Causal Prompt Distillation module (CPD) is designed to reduce model catastrophic forgetting by consolidating old knowledge through embedding memory prompt. It also eliminates the re-inference process for distillation to reduce computational overhead.
- We demonstrate that URCPD has achieved state-of-the-art performances on five classic LReID seen datasets and seven unseen datasets (including both RFL-ReID task and LReID task).

Related Work

Re-Identification

Person re-identification (ReID) technology refers to the process of retrieving specific pedestrians within non-overlapping camera networks using computer vision techniques. From the initial traditional ReID (Dou et al. 2023; Zou et al. 2023) to contemporary multi-modal ReID (Fang, Yang, and Fu 2023; Yang, Chen, and Ye 2023), ReID technology has progressively addressed increasingly challenging and complex scenarios. To tackle large intra-sample variations, clothing-changing ReID (Bansal, Foresti, and Martinel 2022) and occluded ReID (Liu et al. 2025b; Yuan et al. 2025) have emerged as prominent research topics with substantial progress. However, conventional ReID approaches have not adequately considered the data constraints inherent in real-world deployment. Surveillance footage constitutes streaming data, where training data arrives sequentially over time.

Lifelong Person Re-Identification

Lifelong person re-identification (LReID) requires models to continuously learn new identity features while retaining old knowledge upon sequentially receiving pedestrian data from different distributions. Similar to other continuous learning tasks, balancing model plasticity and stability remains a core challenge for LReID. In the work of Pu et al. (Pu et al. 2021), continuous learning (Zhao et al. 2021; Wan et al. 2022) methods were first introduced to ReID, proposing the AKA framework, which became a pioneer in LReID research. Subsequently, to address data privacy concerns, Cui et al. (Cui et al. 2024) proposed rehearsal-free LReID techniques, while Xu et al. (Xu et al. 2025a) and Li et al. (LiQiwei et al. 2024) introduced exemplar-free LReID techniques. To further mitigate catastrophic forgetting and reduce computational overhead, this paper proposes a novel LReID paradigm that eliminates the inference process of old models during new training phases, directly performs label distillation.

Knowledge Distillation

Knowledge distillation (KD) (Hinton 2015) is a model compression and knowledge transfer technique. In lifelong learning, knowledge distillation is frequently employed to combat catastrophic forgetting. Existing LReID methods typically utilize three types of Knowledge distillation: feature distillation (Romero et al. 2015), logit distillation (Hinton 2015) and relation distillation (Park et al. 2019b). Feature distillation learns intermediate feature representations, significantly improving feature quality. Logit distillation learns the final probability distribution, enhancing model generalization. Relational distillation learns the relationships or global structure between samples. To reduce the cumulative forgetting of old knowledge during training and alleviate catastrophic forgetting, this paper will generate a set of memory prompts. These prompts will be embedded to consolidate old knowledge. In addition, we will eliminate the re-inference process in knowledge distillation to reduce computational overhead.

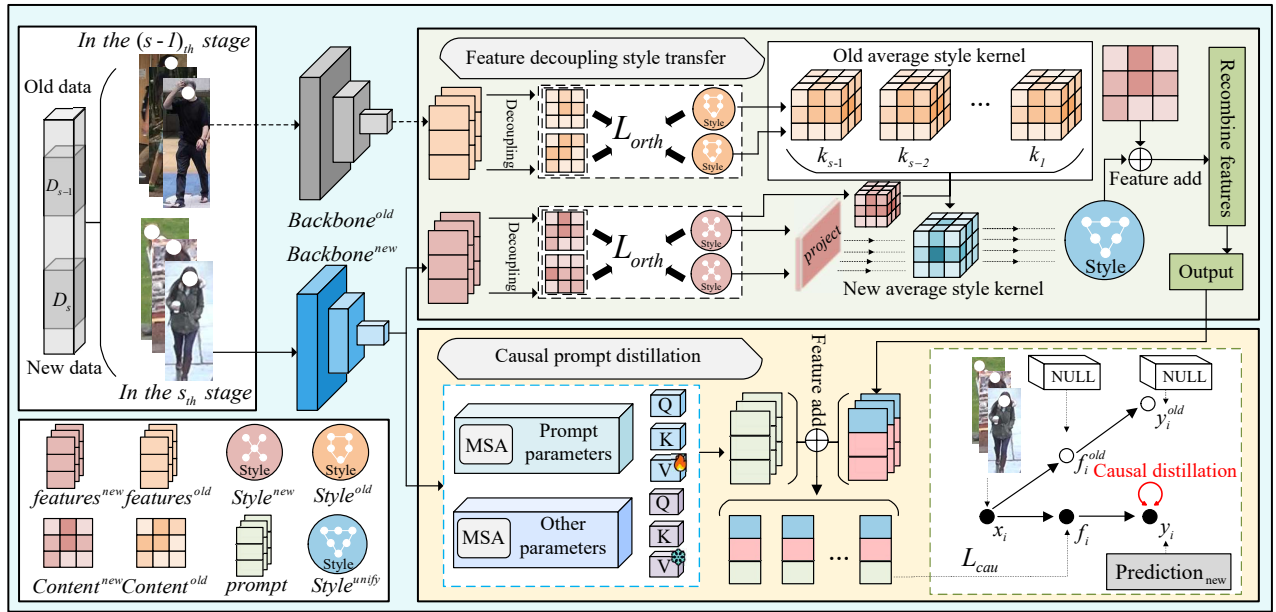


Figure 2: Overview of our proposed Unified Representation Causal Prompt Distillation (URCPD) method. The FDST module decouples the features into content factor and style factor. It then computes an average style kernel by aggregating the style factors across all previous stages. Within the new stage, FDST projects the style factor of the new features onto this average style kernel and recombines the content with the projected style. The CPD module trains a set of prompt parameters for memorizing old knowledge. During prompt generation, all other parameters are frozen, only the prompt parameters remain active. The generated prompt will be embedded in image feature and then perform re-inference-free distillation.

Method

Preliminaries

Given a streaming data set $D = \{D^1, D^2, \dots, D^S\}$. Every $D^S = \{T^S, G^S\}$ contains a training set $T^S = \{x_i^S, y_i^S\}$ and a corresponding gallery set $G^S = \{x_i^S, y_i^S\}$. x_i^S, y_i^S are person images and identity labels. The sample size of each dataset D^S is N^S . The total sample size of all datasets D is $N = N^1 + N^2 + \dots + N^S$. We use the LReID baseline PEAME (LiQiwei et al. 2024), including feature extraction module $\phi(\cdot)$, which is used to extract the image features of the input person. Classifier head $\psi(\cdot)$, which is used to predict the probability of a person’s identity. The base model extracts the feature f_i for each person x_i and predicts the identity probability y_i of the person.

Overview of URCPD

The re-inference-free LReID we proposed does not require going through the inference process of the old model at each stage. As shown in Fig. 2, Unified Representation Causal Prompt Distillation (URCPD) is mainly divided into two parts. Feature Decoupling Style Transfer module (FDST) is proposed to improve model inter-domain generalization capability by transferring new data styles to a unified style space; Causal Prompt Distillation module (CPD) is designed to reduce model catastrophic forgetting by consolidating old knowledge through embedding memory prompt. It also eliminates the re-inference process for distillation to reduce computational overhead.

Feature Decoupling Style Transfer

To reduce domain gaps among different scene datasets and improve model inter-domain generalization capability, a Feature Decoupling Style Transfer module is proposed to map new features into a unified feature space. Specific methods are as follows.

Person features encompass both content information (e.g., body posture, gender, appearance) and style information (e.g., clothing style, environmental conditions, camera characteristics). During person retrieval, the more stable content information often holds greater value than the mutable style information. Consequently, some prior work (Jin et al. 2020) directly filters out style information, retaining only content information. However, this approach is overly aggressive. The goal of decoupling is to maximally separate style and content. Nevertheless, achieving perfect decoupling in practice is fundamentally impossible. Blindly discarding the entire style factor risks losing implicit content potentially embedded within it.

Given a new feature f_i , an orthogonal constraint loss is applied to enforce orthogonality between the content and style encoding spaces. This strictly decouples f_i into content factor c_i and style factor s_i . The orthogonal constraint loss function is:

$$\mathcal{L}_{orth} = |c^T s|_F^2, \quad (1)$$

where F represents the frobenius norm of the matrix, c^T is the transpose matrix of c .

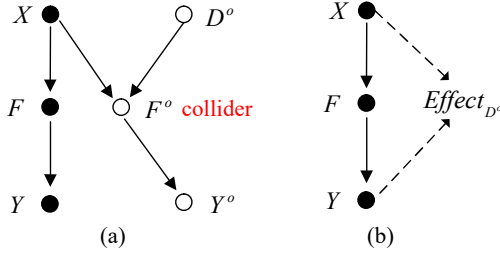


Figure 3: The causal graph in the LReID task.

After obtaining the style factor s , calculate the average style μ^S of all samples in the S_{th} stage and the average style kernel μ from stage 1_{th} to S_{th} :

$$\mu^S = \frac{1}{N^S} \sum_{i=1}^{N^S} s_{i,\mu} = \frac{1}{S} \sum_1^S \mu^S, \quad (2)$$

where, N^S is the number of samples in stage S_{th} , f_i is the feature of the person x_i , S represents the current training stage.

To constrain the style factors, we cluster the style factors of the same task towards the average style core center:

$$\mathcal{L}_{cent} = |s - \mu|^2. \quad (3)$$

After that, project s onto the uniform style space. We need to calculate the similarity degree α^S between s and μ through Eq. (4), and then calculate the final unified style s_{unify} through Eq. (5):

$$\alpha^S = \frac{\exp(\langle s, \mu \rangle / \tau)}{\sum_{j=1}^S \exp(\langle s, \mu \rangle / \tau)}, \quad (4)$$

$$s_{unified} = \alpha^S \cdot \mu. \quad (5)$$

In Eq. (4), τ is the temperature coefficient, which controls the sharpness of the distribution. Finally, we reorganized the content c and style $s_{unified}$ to obtain the features after unifying the style $f_{unified}$.

In summary, the overall loss function of the Feature Decoupling Style Transfer module is:

$$\mathcal{L}_{FDST} = \mathcal{L}_{orth} + \mathcal{L}_{cent}. \quad (6)$$

Causal Prompt Distillation

To reduce the accumulated forgetting of old knowledge during the training stage, a Causal Prompt Distillation module (CPD) is introduced. We will train a set of prompt vectors, which are a few parameters learned specifically for memorizing old tasks. Furthermore, to reduce computational overhead, we have eliminated the inference process of the old model in the new stage of knowledge distillation. The specific methods are as follows.

In Fig. 3 (a), X are person samples at the new stage, F are features of persons, and Y are identities of persons. D^o is the old dataset, F^o are X 's features extracted by

the old model, Y^o are X 's identities inferred by the old model. Based on the causal path diagram in Fig. 3 (a), we can use $do(\cdot)$ function (Pearl and Judea 2014; Pearl, Glymour, and Jewell 2016) to represent the causal effects between variables. For example, $do(D^o = X^o)$ forces X^o to be assigned to D^o , and the influence of D^o on Y can be expressed as $P(Y|do(D^o = X^o)) - P(Y|do(D^o = 0))$. Here, $do(D^o = 0)$ indicates no intervention in D^o . We simplify the influence of D^o on Y . Then, for each new stage of LReID, the causal effect of the old data D^o on the new prediction Y is:

$$Effect_{D^o} = P(Y = Y^S | D^o = X^o) - P(Y = Y^S | D^o = 0). \quad (7)$$

However, the path between D^o and Y is blocked. Therefore, we need to introduce a collider. By controlling the collider, we can make two originally unrelated variables related, thereby generating a collider bias. For example, a causal graph $Environmental\ impact \rightarrow Myopia \leftarrow Genetic\ factors$ indicates that the factors influencing whether a person is nearsighted are the environment and genetics. Obviously, environment and genetics are two unrelated concepts. But if a person has non-hereditary myopia, it is very likely that this person has been influenced by the environment, and vice versa. At this point, two unrelated variables become related due to the existence of the collider. Therefore, under the influence of the collider, Eq. (7) can be written as:

$$Effect_{D^o} = \sum_X P(Y|X, F^o) (P(X|F^o, D^o = X^o) - P(X|F^o, D^o = 0)), \quad (8)$$

where, $P(Y|X, F^o)$ implies the prediction of $X = x_i$, it can be written as $P(Y|X = x_i)$. ($P(X|F^o, D^o = X^o) - P(X|F^o, D^o = 0)$) is the probability that feature $F^o = f_i^o$ is the i_{th} person, it can be written as $P(x_i, F^o, D^o)$. For a new dataset D^S containing N samples, the causal effect formula can be written from Eq. (8) as:

$$Effect_{D^o} = \sum_{x_i \in D^S} P(Y|X = x_i) P(x_i, F^o, D^o), \quad (9)$$

where, $P_1 + P_2 + \dots + P_N = 1$. Therefore, the final causal effect formula is:

$$Effect_{D^o} = \sum_{x_i \in D^S} P(Y|X = x_i), \quad (10)$$

By using Eq. (10), the causal effects of the old data can be directly deduced and calculated using Y and X (as shown in Fig. 3 (b)). By directly distilling the causal effect, the reasoning process of the old model in the new stage can be eliminated. This re-inference-free distillation can reduce computational overhead. In addition, to combat catastrophic forgetting, we have specially trained a set of prompt parameters for memorizing old knowledge. In the backbone, x_i will be input into the multi-head self-attention (MSA) layer:

$$MSA(Q, K, V) = Concat(h_1, \dots, h_m) W^O, \quad (11)$$

where, m is the number of attention heads, the i_{th} head is $h_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$. Q, K, V are input

Tsak	Method	Publication	Market-1501		CUHKSYSU		DukeMTMC		MSMT17-V2		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
LReID	JointTrain	-	82.3	92.3	92.3	92.9	75.9	86.0	44.8	67.6	65.1	64.4	71.9	80.8	70.9	64.3
	LwF	T-PAMI 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	WACV 2021	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	CVPR 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	44.3	40.4
	PatchKD	ACM MM2022	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	T-PAMI 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
	ConRFL	PR 2023	59.2	78.3	82.1	84.3	45.6	61.8	12.6	30.4	<u>51.7</u>	<u>53.8</u>	50.2	61.7	57.4	52.3
	LSTKC	AAAI 2024	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
	DKP	CVPR 2024	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
	DASK	AAAI 2025	61.2	82.3	81.9	83.7	58.5	74.6	<u>29.1</u>	57.6	46.2	48.1	55.4	69.3	65.3	58.4
	LSTKC++	T-PAMI 2025	65.3	83.2	85.8	87.5	56.0	70.5	21.0	43.5	47.7	48.9	55.2	66.7	63.2	56.3
	DCR	TCSVT 2025	75.6	<u>87.9</u>	87.3	88.5	60.1	71.9	25.3	50.1	60.5	61.3	61.8	<u>71.9</u>	60.8	58.3
	BaseLine*	IJCV 2024	75.0	88.7	89.3	<u>90.6</u>	63.6	75.9	27.3	51.2	47.6	48.6	60.5	71.0	67.2	<u>59.4</u>
Ours	-	71.3	86.7	89.8	91.1	66.4	78.7	31.4	<u>55.4</u>	50.2	50.4	61.8	72.4	68.9	61.4	
RFL-ReID	LwF	T-PAMI 2017	39.1	58	40.0	40.7	7.8	15.3	2.6	7.1	23.3	23.9	22.6	29	-	-
	AKA	CVPR 2021	36.1	52.2	38.6	37.6	7.6	13.8	3.1	8.3	26.5	26.5	22.4	27.7	-	-
	CVS	CVPR 2022	38.8	55.6	49	49.7	19.3	30.0	4.6	11.5	24.7	24.7	27.3	34.3	-	-
	PatchKD	ACM MM2022	61.4	78.4	57.8	59.0	20.8	34.4	5.1	12.8	36	37.6	36.2	44.4	-	-
	C2R	CVPR 2024	62.7	79.7	64.4	66.3	26.7	41.7	6.8	15.7	37.2	37.6	39.5	48.2	-	-
	DASK*	AAAI 2025	43.1	58.3	60.5	60.3	<u>53.2</u>	70.7	27.4	52.6	<u>49.6</u>	<u>50.7</u>	46.8	58.6	-	-
	BaseLine*	IJCV 2024	76.0	88.5	<u>87.1</u>	88.9	69.3	<u>80.0</u>	<u>36.0</u>	<u>58.9</u>	43.2	42.6	<u>62.3</u>	<u>71.8</u>	-	-
	Ours	-	<u>70.6</u>	<u>84.4</u>	87.4	<u>88.7</u>	69.3	81.4	38.0	60.5	59.6	60.9	65.0	75.2	-	-

Table 1: Training order-1: Market-1501 \rightarrow CUHK-SYSU \rightarrow DukeMTMC \rightarrow MSMT17-V2 \rightarrow CUHK03.

query, key, and value of the MSA layer, and the first input of (Q, K, V) is x_i . W_i^Q , W_i^K , W_i^V and W^O are projection matrices. Embed the memory prompt parameter p that needs to be learned into V to obtain $MSA(Q, K, V; p)$ and train it together with the backbone network using a separate optimizer.

After training the memory prompt parameters, use $MSA(Q, K, [freeze(V); p])$ (freeze V and retain the p parameter) to infer the prompt word prompt, and embed the prompt into the $f_{unified}$ to obtain f_{final} .

Finally, we use a new causal chain $D^o \rightarrow collider \leftarrow X \rightarrow F(f_{final}) \rightarrow Y(y_{final})$, and calculate $Effect_{D^o}$ with y_{final} by Eq. (10) to perform Causal Prompt Distillation $\arg \min_{model} (-\log(Effect_{D^o})) \rightarrow model$.

Experiment

Datasets and Evaluation Metrics

Training datasets. We set the training dataset according to the classic research on lifelong person re-identification (Pu et al. 2021), including Market-1501 (Zheng et al. 2015), CUHKSYSU (Xiao et al. 2017), DukeMTMC (Zheng, Zheng, and Yang 2017), MSMT17-V2 (Wei et al. 2018), and CUHK03 (Li et al. 2014). For lifelong person re-identification, we need to simulate the real data generation process with streaming data, set according to the data order of C2R (Sun and Mu 2022), including *order-1*: Market-1501 \rightarrow CUHK-SYSU \rightarrow DukeMTMC \rightarrow MSMT17-V2 \rightarrow

CUHK03 and *order-2*: DukeMTMC-ReID \rightarrow MSMT17-V2 \rightarrow Market-1501 \rightarrow CUHK-SYSU \rightarrow CUHK03.

Test datasets. According to the conventional test method of LReID, we split the test dataset into the seen test set and the unseen test set. The seen test set is the same as the training set, and the unseen test set consists of seven ReID datasets outside the training set, including CUHK01 (Li, Zhao, and Wang 2013), CUHK02 (Li and Wang 2013), VIPeR (Gray and Tao 2008), i-LIDS (Zheng, Gong, and Xiang 2009), PRID (Hirzer et al. 2011), GRID (Loy, Xiang, and Gong 2010), and SenseReID (Zhao et al. 2017).

Evaluation Metrics. We use mAP (mean Average Precision) and Rank-1 to evaluate the performance of all comparison methods on each data set. In addition, for each comparison method, we calculated the average mAP and average Rank-1 of all datasets to evaluate the overall performance of the method. The higher the value of mAP and Rank-1, the higher the accuracy of the model.

Implementation Details

For image feature extraction, we use the same backbone and optimizer as (LiQiwei et al. 2024). The batch size is set to 64. The backbone’s learning rate is set to 8×10^{-3} and the prompt’s learning rate is set to 5×10^{-3} . In the first stage, the number of iterations is 80, the learning rate decays by $\times 0.1$ at the 40_{th} and 70_{th} epochs. For the other stage, the number of epochs is 60, the learning rate decays by $\times 0.1$ at the 30_{th} epoch. Our experimental environment

Tsak	Method	Publication	DukeMTMC		MSMT17-V2		Market-1501		CUHKSYSU		CUHK03		Average		UnSeen-Avg	
			mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
LReID	JointTrain	-	75.9	86.0	44.8	67.6	82.3	92.3	92.3	92.9	65.1	64.4	71.9	80.8	70.9	64.3
	LwF	T-PAMI 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	47.2	42.6
	CRL	WACV 2021	43.5	63.1	4.8	13.7	35.0	59.8	70.0	72.8	34.5	36.8	37.6	49.2	47.8	43.5
	AKA	CVPR 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	44.3	40.4
	PatchKD	ACM MM2022	58.3	74.1	6.4	17.4	43.2	67.4	74.5	76.9	33.7	34.8	43.2	54.1	49.1	45.4
	MEGE	T-PAMI 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	47.7	44.0
	ConRFL	PR 2023	34.4	51.3	7.6	20.1	61.6	80.4	82.8	85.1	49.0	50.1	47.1	57.4	57.9	53.4
	LSTKC	AAAI 2024	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
	DKP	CVPR 2024	53.4	70.5	14.5	33.3	60.6	81.0	83.0	84.9	45.0	46.1	51.3	63.2	59.0	51.6
	DASK	AAAI 2025	55.7	74.4	25.2	<u>51.9</u>	71.6	87.7	84.8	86.2	48.4	<u>49.8</u>	57.1	70.0	65.5	57.9
	LSTKC++	T-PAMI 2025	55.3	70.6	15.7	34.5	65.6	82.7	86.3	88.0	47.1	47.6	54.0	64.7	62.1	54.7
	DCR	TCSVT 2025	64.1	77.2	<u>25.4</u>	44.9	<u>70.6</u>	<u>84.5</u>	86.1	88.2	54.2	58.7	60.1	<u>70.7</u>	61.6	59.2
BaseLine*	IJCV 2024	71.5	82.6	23.6	46.8	64.3	81.9	89.0	90.3	45.5	46.5	58.8	69.6	<u>67.3</u>	<u>60.4</u>	
Ours	-	<u>69.1</u>	<u>81.2</u>	29.4	53.5	61.6	80.0	<u>88.7</u>	<u>89.9</u>	<u>49.4</u>	<u>49.3</u>	<u>59.6</u>	70.8	68.7	62.4	
RFL-ReID	LwF	T-PAMI 2017	15.0	22.9	1.2	3.2	9.5	19.4	38.8	37.5	20.2	19.6	16.9	20.5	-	-
	AKA	CVPR 2021	11.1	15.1	1.3	3.2	13.4	27.3	35.9	34.7	25.2	25.6	17.4	21.2	-	-
	CVS	CVPR 2022	29.0	41.9	3.5	9.4	30.7	49.6	60.0	61.2	28.5	29.9	30.3	38.4	-	-
	PatchKD	ACM MM2022	46.5	60.9	4.0	10.4	31.1	50.5	63.0	64.0	35.8	36.6	36.1	44.5	-	-
	C2R	CVPR 2024	48.4	63.6	6.2	14.9	37.0	55.6	67.4	68.4	39.2	39.5	39.7	48.4	-	-
	DASK*	AAAI 2025	32.6	47.3	20.6	12.5	72.4	86.4	<u>82.8</u>	84.0	52.3	53.5	52.1	62.8	-	-
	BaseLine*	IJCV 2024	63.7	80.0	<u>23.1</u>	<u>45.6</u>	41.8	67.6	82.5	86.1	<u>55.3</u>	<u>55.5</u>	<u>53.3</u>	<u>66.9</u>	-	-
	Ours	-	<u>62.4</u>	<u>77.9</u>	25.7	48.6	<u>42.5</u>	<u>67.7</u>	83.7	<u>85.3</u>	58.4	59.4	54.5	67.8	-	-

Table 2: Training order-2: DukeMTMC → MSMT17-V2 → Market-1501 → CUHK-SYSU → CUHK03.

is: pytorch1.13.1, python3.7, GPU is a single NVIDIA RTX A6000.

Comparison with State-of-the-arts

Comparison method. For LReID, we compare LwF (Li et al. 2017), CRL (Zhao et al. 2021), AKA (Pu et al. 2021), PatchKD (Shao et al. 2023), MEGE (Pu et al. 2023), ConRFL (Huang et al. 2023), LSTKC (Xu, Zou, and Zhou 2024), DKP (Xu et al. 2024), DASK (Xu et al. 2025a), LSTKC++ (Xu et al. 2025b), DCR (Liu et al. 2025a) and baseline PAEMA (LiQiwei et al. 2024).

In addition, we also conducted comparative experiments on the re-indexing free lifelong person re-identification (RFL-ReID) (Cui et al. 2024) task. The comparative methods included: LwF (Li et al. 2017), AKA (Pu et al. 2021), CVS (Wan et al. 2022), PatchKD (Shao et al. 2023), C^2R (Cui et al. 2024), DASK (Xu et al. 2025a) and baseline PAEMA (LiQiwei et al. 2024).

Comparison of experimental results. The experimental results are shown in Table 1 and Table 2. The bold part is the best experimental result and the underlined part is the second best result. * indicates that we reproduce the experiment in our experimental environment. JointTrain is the upper limit of precision for each dataset, which is specifically set in the PatchKD (Shao et al. 2023). Overall, the accuracy of our method is the highest on most data sets.

In *order-1*. On the LReID tasks, the average mAP and Rank-1 of our URCPD method are respectively higher than

the baseline 1.3% and 1.4% on the seen datasets, and are respectively 1.7% and 2% higher on the invisible dataset. On RFL-ReID, the average mAP and Rank-1 are 2.2% and 3.4% higher than the baseline respectively.

In *order-2*. On the LReID tasks, the average mAP and Rank-1 of our URCPD method are respectively higher than the baseline 0.8% and 1.2% on the seen datasets, and are respectively 1.4% and 2% higher on the invisible dataset. On RFL-ReID, the average mAP and Rank-1 are 1.2% and 0.9% higher than the baseline respectively.

In *order-1*, the performance of our method reached the optimal or suboptimal level in 85% of the indicators. In *order-2*, the performance of our method reached the optimal or suboptimal level in 88% of the indicators.

Comparison of anti-forgetting tendency. Fig. 4 draws the anti-forgetting tendency of each method on *order-1*. Specifically, we use $model^S$ to test dataset $D^1 \sim D^S$, where $S \in [1, 5]$. Then record the test results of each stage. It can be seen that our method is superior to all existing methods in the accuracy of MAP and rank-1 at each stage.

Ablation Studies

In order to further study each module of our method (URCPD), we conducted the ablation experiment on *order-1*, and the experimental results are shown in Table 3.

Effectiveness of FDST. We add the Feature Decoupling Style Transfer module (FDST) to the base. For LReID, the average mAP and Rank-1 are improved by 0.7% and 0.8%

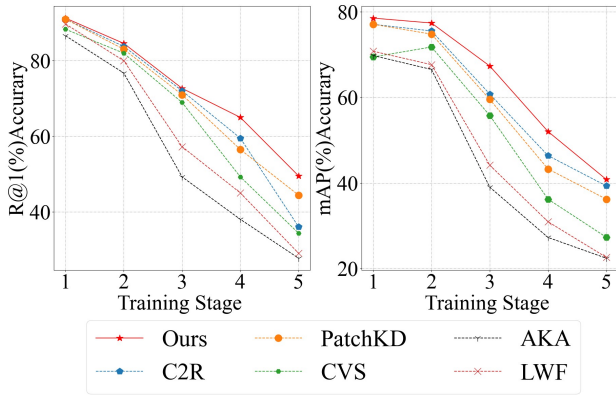


Figure 4: Anti-forgetting tendency on order-1.

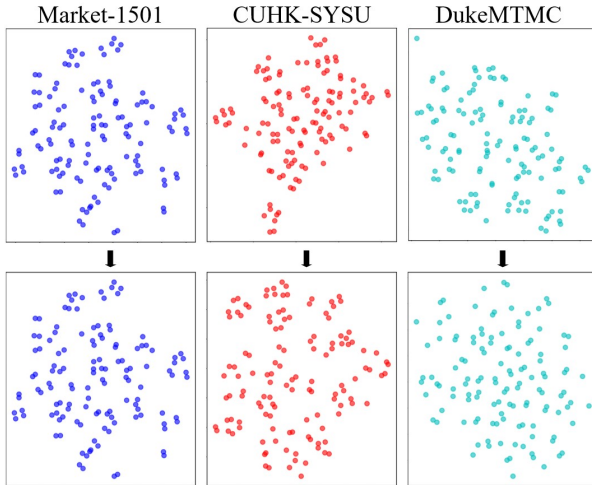


Figure 5: Visualization of the feature style distribution. The new domain distribution tends to resemble the old domain distribution.

compared with the base respectively on seen datasets, 0.8% and 1.4% compared with the base respectively on unseen datasets. Moreover, for RFL-ReID, the average mAP and Rank-1 increased by 1.0% and 1.2%. This indicates that the FDST module’s feature decoupling and style unification are effective in improving the model’s performance.

Effectiveness of CPD. We add the Causal Prompt Distillation (CPD) to the base. For LReID, the average mAP and Rank-1 are improved by 0.6% and 1.0% compared with the base respectively on seen datasets, 1.0% and 1.1% compared with the base respectively on unseen datasets. Moreover, for RFL-ReID, the average mAP and Rank-1 increased by 1.7% and 2.6%. This indicates that the CDP module can effectively alleviate catastrophic forgetting.

Overall, the proposed URCPD demonstrates excellent gains over the baseline. For LReID, there are performance improvements of 1.3% mAP and 1.4% Rank-1 on seen datasets, and 1.7% mAP and 2.0% Rank-1 on unseen datasets. For RFL-ReID, there are performance improvements of 2.7% mAP and 3.4% Rank-1.

base	FDST	CPD	LReID-Seen		LReID-UnSeen		RFL-ReID	
			mAP	R@1	mAP	R@1	mAP	R@1
✓	-	-	60.5	71.0	67.2	59.4	62.3	71.8
✓	✓	-	61.2	71.8	68.0	60.8	63.3	73.0
✓	-	✓	61.1	72.0	68.2	60.5	64.0	74.4
✓	✓	✓	61.8	72.4	68.9	61.4	65.0	75.2

Table 3: Ablation studies of each module in URCPD.

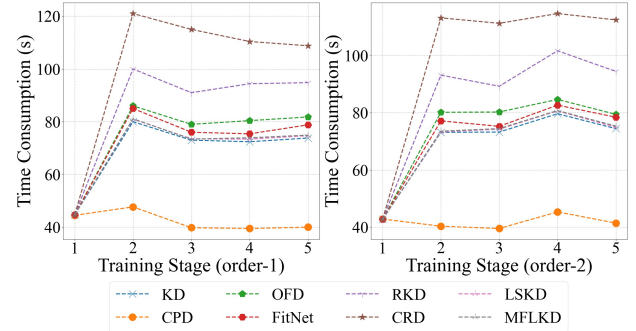


Figure 6: The time consumption for training one batch of each dataset.

Visualization

Fig. 5 is the t-SNE visualization of the feature style distribution of different domains. Here, we only present the style distribution changes in the three domains during the training process of *order-1*. The upper part of the figure shows the distribution without a unified style, while the lower part shows a unified style. Since the style always aligns with that of the old domain, there has been no change in Domain 1, while the styles of Domain 2 and Domain 3 have already moved closer to Domain 1.

Computing Consumption

To verify the assistance of our CPD module in reducing computational overhead, we conducted training consumption time test on *order-1*. The test results are shown in Fig. 6, we add six knowledge distillation methods to the base for comparison. The results show that our CPD reduces a significant amount of time consumption.

Conclusion

In our paper, a new LReID paradigm called Unified Representation Causal Prompt Distillation (URCPD) for re-inference-free lifelong person re-identification is introduced. This paradigm eliminates the reasoning process of the old model in knowledge distillation. We also designed a new ReID method for re-inference-free lifelong person re-identification. Specifically, a Feature Decoupling Style Transfer module (FDST) is proposed to improve model inter-domain generalization capability by transferring new data styles to a unified style space, and a Causal Prompt Distillation module (CPD) is designed to reduce model catastrophic forgetting by consolidating old knowledge through embedding memory prompt.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62272461, Grant 62172417, and Grant 62277046; in part by the Double First Class Project of China University of Mining and Technology for Independent Innovation and Social Service under Grant 2022ZZCX06; in part by the Six Talent Peaks Project in Jiangsu Province under Grant 2015-DZXX-010 and Grant 2018-XYDXX-044.

References

- Bansal, V.; Foresti, G. L.; and Martinel, N. 2022. Cloth-Changing Person Re-identification with Self-Attention. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 602–610.
- Chaudhry, A.; Dokania, P. K.; Ajanthan, T.; and Torr, P. H. 2018. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Proceedings of the European conference on computer vision (ECCV)*, 532–547.
- Cui, Z.; Zhou, J.; Wang, X.; Zhu, M.; and Peng, Y. 2024. Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16614–16623.
- Dou, Z.; Wang, Z.; Li, Y.; and Wang, S. 2023. Identity-seeking self-supervised representation learning for generalizable person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, 15847–15858.
- Fang, X.; Yang, Y.; and Fu, Y. 2023. Visible-infrared person re-identification via semantic alignment and affinity inference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11270–11279.
- Fu, A.; Zhao, J.; Zhou, Y.; Du, W.; Yao, R.; and El Saddik, A. 2025. Similarity Regulation and Calibration Alignment for Weakly Supervised Text-Based Person Re-Identification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 21(3).
- Gao, Z.; Han, S.; Zhang, X.; Xu, K.; Zhou, D.; Mao, X.; Dou, Y.; and Wang, H. 2025. Maintaining fairness in logit-based knowledge distillation for class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16763–16771.
- Gray, D.; and 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*, 262–275.
- Heo, B.; Kim, J.; Yun, S.; Park, H.; Kwak, N.; and Choi, J. Y. 2019. A Comprehensive Overhaul of Feature Distillation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1921–1930.
- Hinton, G. 2015. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*.
- Hirzer, M.; Beleznai, C.; Roth, P. M.; and 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*, 91–102.
- Huang, J.; Yu, X.; An, D.; Wei, Y.; Bai, X.; Zheng, J.; Wang, C.; and Zhou, J. 2023. Learning consistent region features for lifelong person re-identification. *Pattern Recognition*, 144: 109837.
- 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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14268–14277.
- Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; and Zhang, L. 2020. Style Normalization and Restitution for Generalizable Person Re-Identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3140–3149.
- Li, D.; Chen, X.; Zhang, Z.; and Huang, K. 2017. Learning deep context-aware features over body and latent parts for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 384–393.
- Li, W.; and Wang, X. 2013. Locally aligned feature transforms across views. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3594–3601.
- Li, W.; Zhao, R.; and Wang, X. 2013. Human Reidentification with Transferred Metric Learning. In Lee, K. M.; Matsushita, Y.; Rehg, J. M.; and Hu, Z., eds., *Computer Vision – ACCV 2012*, 31–44.
- Li, W.; Zhao, R.; Xiao, T.; and Wang, X. 2014. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 152–159.
- LiQiwei; XuKunlun; PengYuxin; and ZhouJiahuan. 2024. Exemplar-Free Lifelong Person Re-identification via Prompt-Guided Adaptive Knowledge Consolidation. *International Journal of Computer Vision*.
- Liu, S.; Fan, H.; Wang, Q.; Ren, W.; Tang, Y.; and Cong, Y. 2025a. Domain Consistency Representation Learning for Lifelong Person Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–12.
- Liu, X.; Guo, J.; Chen, H.; Miao, Q.; Xi, Y.; and Liu, R. 2025b. Adaptive Occlusion-Aware Network for Occluded Person Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(5): 5067–5077.
- Loy, C. C.; Xiang, T.; and Gong, S. 2010. Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision*, 90: 106–129.
- Park, W.; Kim, D.; Lu, Y.; and Cho, M. 2019a. Relational Knowledge Distillation. arXiv:1904.05068.
- Park, W.; Kim, D.; Lu, Y.; and Cho, M. 2019b. Relational Knowledge Distillation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3962–3971.
- Pearl; and Judea. 2014. Interpretation and identification of causal mediation. *Psychological Methods*, 19(4): 459–81.

- Pearl, J.; Glymour, M.; and Jewell, N. P. 2016. Causal Inference in Statistics: A Primer.
- Pu, N.; Chen, W.; Liu, Y.; Bakker, E. M.; and Lew, M. S. 2021. Lifelong person re-identification via adaptive knowledge accumulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7901–7910.
- Pu, N.; Zhong, Z.; Sebe, N.; and Lew, M. S. 2023. A memorizing and generalizing framework for lifelong person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Romero, A.; Ballas, N.; Kahou, S. E.; Chassang, A.; and Bengio, Y. 2015. FitNets: Hints for Thin Deep Nets. In *ICLR*.
- Shao, Z.; Zhang, X.; Ding, C.; Wang, J.; and Wang, J. 2023. Unified pre-training with pseudo texts for text-to-image person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11174–11184.
- Sun, S.; Ren, W.; Li, J.; Wang, R.; and Cao, X. 2024. Logit Standardization in Knowledge Distillation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15731–15740.
- Sun, Z.; and Mu, Y. 2022. Patch-based knowledge distillation for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, 696–707.
- Tian, Y.; Krishnan, D.; and Isola, P. 2022. Contrastive Representation Distillation. arXiv:1910.10699.
- Wan, T. S.; Chen, J.-C.; Wu, T.-Y.; and Chen, C.-S. 2022. Continual learning for visual search with backward consistent feature embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16702–16711.
- Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 79–88.
- Xiao, T.; Li, S.; Wang, B.; Lin, L.; and Wang, X. 2017. Joint Detection and Identification Feature Learning for Person Search. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3376–3385.
- Xu, K.; Jiang, C.; Xiong, P.; Peng, Y.; and Zhou, J. 2025a. DASK: Distribution Rehearsing via Adaptive Style Kernel Learning for Exemplar-Free Lifelong Person Re-Identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(9): 8915–8923.
- Xu, K.; Liu, Z.; Zou, X.; Peng, Y.; and Zhou, J. 2025b. Long Short-Term Knowledge Decomposition and Consolidation for Lifelong Person Re-Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–16.
- Xu, K.; Zou, X.; Peng, Y.; and Zhou, J. 2024. Distribution-Aware Knowledge Prototyping for Non-Exemplar Lifelong Person Re-Identification. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16604–16613.
- Xu, K.; Zou, X.; and Zhou, J. 2024. Lstkc: Long short-term knowledge consolidation for lifelong person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 16202–16210.
- Yang, B.; Chen, J.; and Ye, M. 2023. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11069–11079.
- Yuan, C.; Zhang, G.; Ma, C.; Zhang, T.; and Niu, G. 2025. From Poses to Identity: Training-Free Person Re-Identification via Feature Centralization. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24409–24418.
- Zhao, B.; Tang, S.; Chen, D.; Bilen, H.; and Zhao, R. 2021. Continual representation learning for biometric identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1198–1208.
- 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1077–1085.
- Zhao, J.; Fu, A.; Zhou, Y.; liang Du, W.; and Yao, R. 2024. Fine-grained semantic oriented embedding set alignment for text-based person search. *Image and Vision Computing*, 152: 105309.
- Zhao, J.; Wang, H.; Zhou, Y.; Yao, R.; Chen, S.; and Saddik, A. E. 2023a. Spatial-Channel Enhanced Transformer for Visible-Infrared Person Re-Identification. *IEEE Transactions on Multimedia*, 25: 3668–3680.
- Zhao, J.; Wang, H.; Zhou, Y.; Yao, R.; Zhang, L.; and El Saddik, A. 2023b. Context-aware and part alignment for visible-infrared person re-identification. *Image and Vision Computing*, 138: 104791.
- Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, 1116–1124.
- Zheng, W. S.; Gong, S.; and Xiang, T. 2009. Associating Groups of People. *Active Range Imaging Dataset for Indoor Surveillance*.
- Zheng, Z.; Zheng, L.; and Yang, Y. 2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE international conference on computer vision*, 3754–3762.
- Zhu, Z.; Jiang, X.; Zheng, F.; Guo, X.; Huang, F.; Sun, X.; and Zheng, W. 2020. Aware loss with angular regularization for person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 13114–13121.
- Zou, C.; Chen, Z.; Cui, Z.; Liu, Y.; and Zhang, C. 2023. Discrepant and multi-instance proxies for unsupervised person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11058–11068.