Vision-Language Attribute Disentanglement and Reinforcement for Lifelong Person Re-Identification

Supplementary Material

In our supplementary materials, we provide additional ablation studies, implementation details, and visualization results for the proposed Vision-Language Attribute Disentanglement and Reinforcement (VLADR) approach. Furthermore, we include additional quantitative and qualitative results that demonstrate the effectiveness of our method against the state-of-the-art approaches.

The supplementary materials are organized as follows:

- Ablation studies and implementation details of the local attribute extraction module.
- Pseudo-code of the VLADR method and algorithm description.
- Comprehensive overview of the LReID benchmark datasets and their configuration.
- Extensive quantitative comparisons under additional LReID benchmarks.
- Qualitative visualization of retrieval results.

1. Ablation Studies and Implementation Details for Local Attribute Extraction

1.1. Ablation Studies on the Components of the Local Attribute Queries

Table 1. Ablation study on different part query.

Head	Upper	Lower	Foot	Seen-Avg mAP	Seen-Avg R@1	UnSeen-Avg mAP	UnSeen-Avg R@1
✗	✓	✓	✓	70.1	79.5	77.6	72.0
✓	✗	✓	✓	70.1	79.5	77.8	71.9
✓	✓	✗	✓	70.0	79.4	77.7	71.8
✓	✓	✓	✗	69.9	79.4	77.7	72.0
✓	✓	✓	✓	70.4	79.9	77.8	72.0

In this paper, four general human body regions, **i.e.**, the head, upper body, lower body, and feet, are selected as the candidate attributes. As shown in Table 1, incorporating all four attributes achieves the best overall performance, while removing any single attribute leads to performance degradation. This observation verifies that each body part provides complementary and discriminative information that benefits accurate person re-identification.

In addition, we experimented with introducing more fine-grained regions, such as the knee, hand, and arm. However, these attempts resulted in degraded performance. The primary reason is that such small parts are frequently occluded in the images, which introduces ambiguity and noise into the learning process, ultimately hindering effective attribute extraction.

1.2. Textual Attribute Prompts

To obtain accurate textual attributes for each person image, we design attribute-specific prompts that guide the generation of discriminative textual descriptions for the corresponding body regions. Table 2 summarizes the prompt formulations and formatting templates used in our method, where the prompt formulations correspond to the attribute-specific questions, and the formatting templates provide structured guidance for producing the textual responses from the pretrained VLM.

2. VLADR Algorithm Description

The pseudo-code of the proposed VLADR framework is provided in Algorithm 1. When the new domain data D_t is given, our method consists of two major processes: (1) multi-grain text attribute disentanglement (MTAD) and (2) inter-domain cross-modal attribute reinforcement (ICAR).

In MTAD, local textual attributes corresponding to different body parts are extracted, and a global textual representation is optimized through learnable prompt refinement. In ICAR, visual and textual attribute features are jointly optimized through multi-modal alignment. Besides, the inter-domain alignment is conducted to achieve attribute knowledge transfer. This design effectively mitigates catastrophic forgetting and enables continuous accumulation of discriminative attribute knowledge.

3. LReID Benchmark Dataset Configuration

The LReID benchmark [14] adopted in our main paper incorporates 12 person re-identification datasets: Market-1501 [26], LPW [24], CUHK-SYSU [19], MSMT17-V2 [18], CUHK03 [8], CUHK01 [7], CUHK02 [6], VIPeR [2], PRID [3], i-LIDS [1], GRID [13], and SenseReID [25].

To ensure a balanced and fair evaluation, we preprocess the datasets as follows. CUHK-SYSU is originally designed for person search. Following previous works [16, 21], we adapt it for ReID by cropping person images using ground-truth bounding boxes and retaining only identities with at least four annotations. LPW [15] is a video-based person re-identification dataset, which we adopt as a replacement for DukeMTMC-reID, given that both originate from video data. Following the protocol in [24], we convert LPW into a

Table 2. Prompts for generating local attribute descriptions from different body parts.

Local Parts	Prompts	Formatting Template
Head	*Gender:* What is the gender of the person? *Hair Color*: What color is the person's hair? *Hair Length*: How long is the person's hair? *Accessories*: What accessory is the person wearing?	A {*Gender*} with {*Hair Color*}, {*Hair Length*} hair, wearing {*Accessories*}.
Upper Body	*Clothing*: What is the person wearing on the upper body?. *Color*: What color is the upper body clothing? *Movements*: What is the person doing?	Wearing {*Color*} {*Clothing*}, {*Movements*}.
Lower Body	*Clothing*: What is the person wearing on the lower body? *Color*: What color is the lower body clothing?	Wearing {*Color*} {*Clothing*}.
Foot	*Shoes*: What shoes is the person wearing? *Color*: What is the color of the shoes?	Wearing {*Color*} {*Shoes*}.

static ReID format by sampling one frame every 15 frames. To mitigate data imbalance, we uniformly select 500 identities from each dataset to construct the standard benchmark. The detailed dataset statistics are summarized in Table 3.

4. Experiments

4.1. Additional LReID Benchmark

Considering that some existing works report their results on the original LReID* benchmark, which includes the DukeMTMC-reID dataset, we also provide comparisons on this benchmark. Specifically, two dataset orders are adopted: Training Order-1: Market→SYSU→DukeMTMC→MSMT17→CUHK03, and an alternative Training Order-2: DukeMTMC→MSMT17→Market→SYSU→CUHK03.

Evaluation metrics include mean Average Precision (mAP) and Rank-1 accuracy (R@1), computed independently on seen and unseen domains. Seen-Avg denotes the average performance across all training domains, whereas UnSeen-Avg measures the domain generalization capability on unseen domains.

4.2. Results on Training Order-1

Under Training Order-1, VLADR achieves strong performance and establishes new state-of-the-art results. In particular, VLADR attains the highest Seen-Avg scores among all existing methods, surpassing the second-best approach (PAEMA†) by **1.0%** Seen-Avg mAP and **1.1%** Seen-Avg R@1. Regarding unseen-domain generalization, VLADR further achieves at least 0.1% UnSeen-Avg mAP and 0.2% UnSeen-Avg R@1 improvements over existing methods. These results validate the effectiveness of the proposed attribute disentanglement and reinforcement mechanism.

4.3. Results on Training Order-2

Results under Training Order-2 further verify the superiority and robustness of VLADR. Under the seen domains, VLADR surpasses the state-of-the-art DASK† by **3.4%** in Seen-Avg mAP and **2.7%** in Seen-Avg R@1. For unseen domains, VLADR achieves the highest UnSeen-Avg performance, reaching 78.2% UnSeen-Avg mAP and 72.1% UnSeen-Avg R@1, and outperforming the state-of-the-art DKUA by **1.7%** in UnSeen-Avg mAP and **0.9%** in UnSeen-Avg R@1. These improvements indicate that the proposed attribute disentanglement and inter-domain alignment schemes enable more effective cross-domain knowledge accumulation, leading to stronger adaptation on seen domains and superior generalization on unseen domains. The consistent gains across both dataset orders highlight the stability and practical applicability of VLADR in real-world lifelong learning scenarios.

5. Person Retrieval Results

Figure 1 presents the qualitative comparison of person retrieval results produced by VLADR and the state-of-the-art baseline DASK† across both seen and unseen domains. For each query, the retrieved results are ranked in descending order of similarity scores from left to right, where green boxes denote correct matches, red boxes indicate incorrect matches, and black boxes highlight the query images. The visualization demonstrates that VLADR consistently achieves more accurate retrieval under diverse challenging scenarios, including substantial pose variations, cross-viewpoint changes, and complex illumination conditions. Benefiting from the attribute-based reasoning mechanism, VLADR captures more fine-grained and semantically discriminative cues, leading to more robust feature learning and noticeably improved matching accuracy compared to DASK†.

Table 3. Detailed statistics of datasets used in the Lifelong Person Re-Identification benchmark.

Domain Type	Dataset Name	Original Configuration			LReID Benchmark		
		Train	Query	Gallery	Train	Query	Gallery
Seen	CUHK03 [8]	767	700	700	500	700	700
	Market-1501 [26]	751	750	751	500	751	751
	LPW [15]	875	876	876	500	876	876
	CUHK-SYSU [19]	942	2900	2900	500	2900	2900
	MSMT17-V2 [18]	1041	3060	3060	500	3060	3060
UnSeen	i-LIDS [1]	243	60	60	-	60	60
	VIPeR [2]	316	316	316	-	316	316
	GRID [13]	125	125	126	-	125	126
	PRID [3]	100	100	649	-	100	649
	CUHK01 [7]	485	486	486	-	486	486
	CUHK02 [6]	1677	239	239	-	239	239
	SenseReID [25]	1718	521	1718	-	521	1718

Table 4. Comparison of seen-domain anti-forgetting and unseen-domain generalization on Training Order-1 of LReID* benchmark.

Net.	Method	Pub.	Market		SYSU		LPW		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
ResNet	LwF[9]	*TPAMI 2017*	40.6	70.8	66.2	69.7	55.3	73.2	12.3	32.5	27.6	26.6	40.4	54.6	52.5	45.1
	DKP[20]	*CVPR 2024*	60.3	80.6	83.6	85.4	71.2	78.9	19.7	41.8	43.6	44.2	55.7	66.2	59.2	51.6
	LSTKC[21]	*AAAI 2024*	54.7	76.0	81.1	83.4	68.5	75.3	20.0	43.2	44.7	46.5	53.8	64.9	57.0	49.9
	DKP++[27]	*TPAMI 2025*	63.9	83.6	83.7	85.3	72.1	79.4	26.5	52.6	46.5	47.6	58.5	71.1	65.7	58.5
	LSTKC++[23]	*TPAMI 2025*	65.3	83.2	85.8	87.5	73.2	80.1	21.0	43.5	47.7	48.9	58.6	70.0	63.2	56.3
	DASK[22]	*AAAI 2025*	61.2	82.3	81.9	83.7	70.5	76.2	29.1	57.6	46.2	48.1	57.8	70.0	65.3	58.4
	PrR[17]	*arXiv 2025*	67.1	84.9	84.1	85.2	75.3	81.5	32.9	61.2	57.2	59.3	63.3	74.6	74.5	65.2
ViT	PAEMA[4]	*IJCV 2024*	68.9	71.7	91.3	90.0	79.1	82.3	55.3	31.4	50.8	50.0	69.1	75.1	72.7	61.8
	DRE[11]	*TNNLS 2025*	69.4	85.2	85.6	86.8	72.5	79.6	21.5	43.7	54.7	58.1	60.7	71.9	74.2	62.5
	DCR[10]	*TCSVT 2025*	75.9	85.7	87.3	88.5	76.2	81.4	25.3	50.1	60.5	61.3	67.0	75.4	76.3	69.2
	DKUA[12]	*arXiv 2025*	82.5	91.3	90.1	91.5	78.9	84.7	25.9	51.4	61.2	62.8	71.7	80.7	**77.4**	70.3
CLIP	LwF[9]	*TPAMI 2017*	63.4	82.2	79.4	81.4	70.2	76.5	15.5	35.7	41.9	42.6	54.1	63.7	58.5	51.6
	DKP†[20]	*CVPR 2024*	79.3	90.3	89.5	90.6	74.8	81.2	36.9	62.7	59.0	60.1	70.1	78.3	72.9	65.1
	LSTKC†[21]	*AAAI 2024*	79.1	91.5	92.0	92.7	75.4	82.6	38.3	64.2	60.7	62.5	71.0	80.6	74.8	68.0
	PAEMA†[4]	*IJCV 2024*	81.1	91.7	92.2	92.8	77.5	84.2	45.4	71.0	61.7	63.6	**71.7**	**82.1**	77.2	**71.2**
	DKP++†[27]	*TPAMI 2025*	77.5	89.1	87.0	87.8	72.6	80.3	37.2	63.1	57.9	59.1	68.4	77.2	73.3	66.7
	LSTKC++†[23]	*TPAMI 2025*	69.4	85.9	90.1	91.4	76.8	83.5	38.9	65.0	70.0	72.1	69.0	80.5	74.0	67.9
	DASK†[22]	*AAAI 2025*	75.7	88.3	90.5	91.1	74.3	81.7	40.0	66.2	63.3	66.1	70.8	80.7	74.9	68.2
	VLADR	*This Paper*	79.7	90.9	91.9	92.6	78.2	85.1	46.0	71.6	67.6	69.9	**72.7**	**83.2**	**77.5**	**71.4**

Net. denotes the visual encoder architecture. † indicates integration with CLIP-ReID [5].

6. Key Findings and Insights

The comprehensive experimental analysis across multiple dataset orders, visual encoders, and baseline methods reveals several important insights about our VLADR approach:

1. **Effective Attribute Disentanglement**: By systematically decomposing person representations into complementary local body part attributes and holistic global features, VLADR captures both fine-grained discriminative details and global semantic consistency, leading to more robust and interpretable representations.
2. **Superior Cross-Modal Alignment**: Joint optimization of vision-language features enforces semantic consistency across modalities, resulting in more stable and generalizable representations compared to vision-only approaches. This cross-modal alignment mechanism significantly enhances both anti-forgetting on seen do-

Table 5. Comparison results on Training Order-2 of LReID* benchmark.

Net.	Method	Pub.	DukeMTMC		MSMT17		Market		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
ResNet	LwF[9]	*TPAMI 2017*	54.2	72.8	10.3	28.2	47.1	71.7	77.9	80.4	28.8	27.8	43.7	56.2	53.3	45.9
	DKP[20]	*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	51.3	47.8
	LSTKC[21]	*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++[27]	*TPAMI 2025*	55.7	73.9	22.2	47.8	65.9	85.6	84.4	86.5	49.5	51.7	55.5	69.1	64.9	57.9
	LSTKC++[23]	*TPAMI 2025*	55.3	70.6	15.7	34.5	65.6	82.7	86.3	88.0	47.1	47.6	54.0	64.0	62.1	54.7
	DASK[22]	*AAAI 2025*	55.7	74.4	25.2	51.9	71.6	87.7	84.8	86.2	48.4	49.8	57.1	70.0	65.5	57.9
	PrR[17]	*arXiv 2025*	55.9	73.3	29.2	57.8	71.7	87.2	85.0	86.5	55.5	57.7	59.4	72.8	71.3	62.8
ViT	PAEMA[4]	*IJCV 2024*	79.8	67.2	49.4	26.0	85.8	69.8	91.0	89.9	49.7	49.3	62.5	69.4	71.1	60.4
	DRE[11]	*TNNLS 2025*	59.7	74.2	18.7	34.8	65.4	82.7	84.8	86.7	51.9	53.2	56.1	66.3	74.0	63.5
	DCR[10]	*TCSVT 2025*	64.1	77.2	25.4	44.9	70.6	84.5	86.1	88.2	54.2	58.7	60.1	70.7	75.8	70.1
	DKUA[12]	*arXiv 2025*	69.5	81.8	31.8	57.6	73.7	87.9	87.6	88.1	59.5	60.9	64.4	75.3	**76.5**	**71.2**
CLIP	LwF[9]	*TPAMI 2017*	59.4	75.1	12.3	30.0	57.9	77.5	81.5	83.3	39.7	40.9	50.2	61.3	57.1	49.1
	DKP†[20]	*CVPR 2024*	71.0	83.8	33.4	59.3	77.2	89.8	89.7	90.4	60.2	62.6	66.3	77.2	75.2	68.9
	LSTKC†[21]	*AAAI 2024*	69.8	82.9	33.5	59.4	72.1	87.2	92.1	93.0	60.6	62.6	65.6	77.0	73.5	66.7
	PAEMA†[4]	*IJCV 2024*	71.7	84.0	42.2	68.5	80.5	91.7	92.9	93.4	62.7	64.1	66.8	77.8	74.0	66.9
	DKP++†[27]	*TPAMI 2025*	67.5	80.3	32.6	58.4	73.6	88.5	86.8	87.7	55.6	57.6	63.2	74.6	73.0	66.1
	LSTKC++†[23]	*TPAMI 2025*	60.5	75.5	32.0	56.8	71.4	86.3	92.0	93.0	70.2	72.6	65.2	76.9	75.6	69.4
	DASK†[22]	*AAAI 2025*	68.8	82.3	38.4	64.8	75.1	88.4	90.4	90.8	63.7	65.9	**67.3**	**78.4**	74.7	67.7
	VLADR	*This Paper*	69.7	82.6	41.6	68.0	80.4	91.2	92.8	93.3	68.8	70.4	**70.7**	**81.1**	**78.2**	**72.1**

Net. denotes the visual encoder architecture. † indicates integration with CLIP-ReID [5].

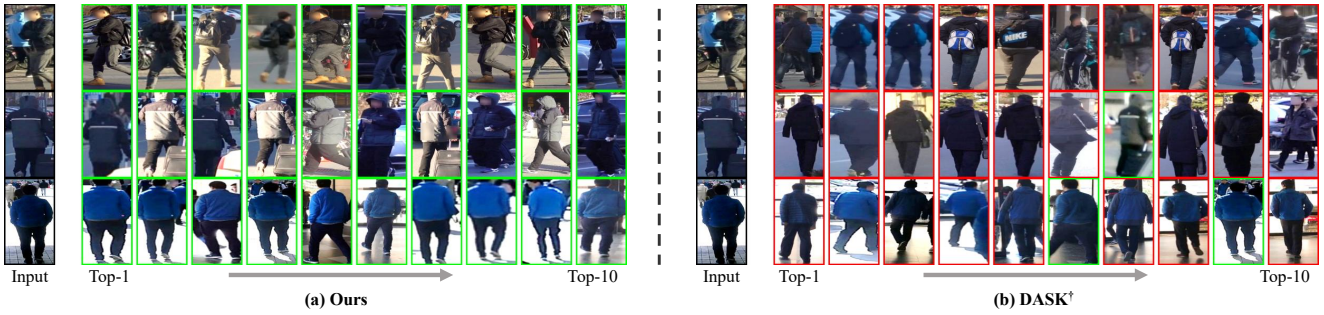

Figure 1. Visualization of person retrieval results. Green boxes indicate correct matches, red boxes indicate incorrect matches, and black boxes denote query images.

mains and generalization on unseen domains.

3. **Strong Improvement of Knowledge Consolidation**: The inter-domain attribute knowledge transfer mechanism effectively preserves and reinforces knowledge from previously learned domains during continual learning, substantially reducing performance degradation on seen domains while maintaining strong performance on new tasks.
4. **Consistent Domain Generalization**: The persistent improvements on unseen domains across both dataset orders demonstrate that our attribute-based representation learning provides superior generalization capability, validating the effectiveness of VLADR for practical lifelong person re-identification scenarios where data from new domains may lack training samples.

References

[1] 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. 1, 3

[2] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275. Springer, 2008. 1, 3

[3] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and dis-

Algorithm 1 VLADR: Vision-Language Attribute Disentanglement and Reinforcement

Input: Current batch $D_t = \{(x_i, y_i)\}_{i=1}^{n_t}$, Previous Model $\mathbf{M}_{t-1}$, Query $\boldsymbol{q}_{vis}^{t-1}$, Parameters $\boldsymbol{\theta}_{t-1}$.
Output: Updated model $\mathbf{M}_t$, Updated query $\boldsymbol{q}_{vis}^{t}$, Updated parameters $\boldsymbol{\theta}_t$

 # Stage 1: Multi-grain Text Attribute Disentanglement
$\mathcal{T}_{x_i} = \{\mathcal{B}(x_i, T_j^Q)\}_{j=1}^{N_a}$; *# Extract local textual attributes*
$\boldsymbol{f}_{text}^{g} = \mathcal{T}(\text{“A photo of a } [X]_1[X]_2 \cdots [X]_M\text{”})$; *# Global textual features*
Optimize prompt parameters $\{[X]_i\}_{i=1}^{M}$ with loss $\mathcal{L}_{prompt}$;

 # Stage 2: Inter-domain Cross-modal Attribute Reinforcement
Initialize $\mathbf{M}_t^0 = \mathbf{M}_{t-1}$, $\boldsymbol{q}_{vis}^0 = \boldsymbol{q}_{vis}^{t-1}$, $\boldsymbol{\theta}_t^0 = \boldsymbol{\theta}_{t-1}$
for $e = 1$ **to** N_{epoch} **do**
  # Coarse-grained Global Representation Learning
  for (x_i, y_i) **in** D_t **do**
    $\boldsymbol{f}_t^i = \mathbf{M}_t^{e-1}(x_i)$; *# Extract vision tokens*
    $\boldsymbol{f}_{vis}^{g,i} = [\text{CLS}]$ *(token from visual features)*;
    $\boldsymbol{f}_{text}^{g,i} = \mathcal{T}([X]_{y_i})$; *# Extract textual tokens*
    $\mathcal{L}_{ReID} = \mathcal{L}_{Tri} + \mathcal{L}_{ce}$; *# Triplet + cross-entropy loss*
    $\mathcal{L}_{global} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp(s(\boldsymbol{f}_{vis}^{g,i}, \boldsymbol{f}_{text}^{g,i}))}{\sum_{k=1}^{K} \exp(s(\boldsymbol{f}_{vis}^{g,i}, \boldsymbol{f}_{text}^{g,k}))}$;

    # Fine-grained Local Attribute Mining
    $F_{vis}^{l,t} = \boldsymbol{\theta}_t^{e-1}(\boldsymbol{q}_{vis}, \boldsymbol{f}_t^i, \boldsymbol{f}_t^i)$; *# Visual local attributes*
    $F_{text}^{l,t} = \mathcal{T}(\mathcal{T}_{x_i})$; *# Textual local attributes*

    $\mathcal{L}_{MAlign} = \frac{1}{N_a}\sum_{p=1}^{N_a} \mathcal{L}_{cons}^{p}$; *# Multi-modal alignment loss*

    # Inter-domain Attribute Knowledge Transfer
    $F_{vis}^{l,t-1} = \boldsymbol{\theta}_{t-1}(\boldsymbol{q}_{vis}^{t-1}, \boldsymbol{f}_{t-1}^i, \boldsymbol{f}_{t-1}^i)$; *# Historical attributes*
    $\boldsymbol{R}_t^p = \gamma(\boldsymbol{v}^{p,t}(\boldsymbol{v}^{p,t})^\top)$ and $\boldsymbol{R}_{t-1}^p = \gamma(\boldsymbol{v}^{p,t-1}(\boldsymbol{v}^{p,t-1})^\top)$; *# Relation matrices*

    $\mathcal{L}_{DAlign} = \frac{1}{N_a}\sum_{p=1}^{N_a} \mathcal{L}_{Rel}^{p}$; *# Domain alignment loss*
  end for

  Optimize with loss: $\mathcal{L} = \mathcal{L}_{ReID} + \mathcal{L}_{global} + \alpha\mathcal{L}_{MAlign} + \beta\mathcal{L}_{DAlign}$;
end for
Return updated models: $\mathbf{M}_t$, $\boldsymbol{q}_{vis}^t$, $\boldsymbol{\theta}_t$;

criminative classification. In *Image Analysis*, pages 91–102. Springer, 2011. 1, 3

[4] Qiwei Li, Kunlun Xu, Yuxin Peng, and Jiahuan Zhou. Exemplar-free lifelong person re-identification via prompt-guided adaptive knowledge consolidation. *IJCV*, 132(11): 4850–4865, 2024. 3, 4

[5] Siyuan Li, Li Sun, and Qingli Li. Clip-reid: exploiting vision-language model for image re-identification without concrete text labels. In *AAAI*, pages 1405–1413, 2023. 3, 4

[6] Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *CVPR*, pages 3594–3601. IEEE, 2013. 1, 3

[7] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*, pages 31–44. Springer, 2012. 1, 3

[8] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, pages 152–159. IEEE, 2014. 1, 3

[9] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE TPAMI*, 40(12):2935–2947, 2018. 3, 4

[10] Shiben Liu, Huijie Fan, Qiang Wang, Weihong Ren, Yandong Tang, and Yang Cong. Domain consistency representation learning for lifelong person re-identification. *arXiv preprint arXiv:2409.19954*, 2024. 3, 4

[11] Shiben Liu, Huijie Fan, Qiang Wang, Xiai Chen, Zhi Han, and Yandong Tang. Diverse representations embedding for lifelong person re-identification. *IEEE Transactions on Neural Networks and Learning Systems*, 2025. 3, 4

[12] Shiben Liu, Mingyue Xu, Huijie Fan, Qiang Wang, Yandong Tang, and Zhi Han. Distribution-aware knowledge unification and association for non-exemplar lifelong person re-identification. *arXiv preprint arXiv:2508.03516*, 2025. 3, 4

[13] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 90(1):106–129, 2010. 1, 3

[14] Nan Pu, Wei Chen, Yu Liu, Erwin M Bakker, and Michael S Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *CVPR*, pages 7897–7906. IEEE, 2021. 1

[15] Guanglu Song, Biao Leng, Yu Liu, Congrui Hetang, and Shaofan Cai. Region-based quality estimation network for large-scale person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, 2018. 1, 3

[16] Zhicheng Sun and Yadong Mu. Patch-based knowledge distillation for lifelong person re-identification. In *ACM MM*, pages 696–707, 2022. 1

[17] Mingyu Wang, Haojie Liu, Zhiyong Li, and Wei Jiang. Pr2r: Information-fused and style-aware privacy-preserving replay for lifelong person re-identification. *arXiv preprint arXiv:2508.01587*, 2025. 3, 4

[18] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88. IEEE, 2018. 1, 3

[19] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search, 2016. 1, 3

[20] 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. IEEE, 2024. 3, 4

[21] Kunlun Xu, Xu Zou, and Jiahuan Zhou. Lstkc: Long short-term knowledge consolidation for lifelong person re-identification. In *AAAI*, pages 16202–16210, 2024. 1, 3, 4

[22] Kunlun Xu, Chenghao Jiang, Peixi Xiong, Yuxin Peng, and Jiahuan Zhou. Dask: Distribution rehearsing via adaptive style kernel learning for exemplar-free lifelong person re-identification. In *AAAI*, pages 8915–8923, 2025. 3, 4

[23] Kunlun Xu, Zichen Liu, Xu Zou, Yuxin Peng, and Jiahuan Zhou. Long short-term knowledge decomposition and consolidation for lifelong person re-identification. *IEEE TPAMI*, 2025. 3, 4

[24] Kunlun Xu, Fan Zhuo, Jiangmeng Li, Xu Zou, and Jiahuan Zhou. Self-reinforcing prototype evolution with dual-knowledge cooperation for semi-supervised lifelong person re-identification. In *ICCV*, 2025. 1

[25] 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 *CVPR*, pages 907–915. IEEE, 2017. 1, 3

[26] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124. IEEE, 2015. 1, 3

[27] Jiahuan Zhou, Kunlun Xu, Fan Zhuo, Xu Zou, and Yuxin Peng. Distribution-aware knowledge aligning and prototyping for non-exemplar lifelong person re-identification. *IEEE TPAMI*, 2025. 3, 4