ASYMMETRIC DOMAIN KNOWLEDGE DISTILLATION FOR UNSUPERVISED CROSS-DOMAIN PERSON RE-IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unsupervised domain adaptive person re-identification (UDA-ReID) is plagued by noisy clustering-derived pseudo labels, inherent feature extraction errors from cross-domain style discrepancies, underutilized outliers, and ineffective source domain knowledge transfer in target fine tuning. Existing methods either focus on clustering or pseudo label refinement while ignoring cross-domain feature bias, or use contrastive learning hindered by unreliable clustering and marginal value of cross-domain negative samples. To address these issues in this paper, UDA-ReID is formulated as an asymmetric domain distillation task, distinct from traditional single domain teacher-student distillation by transferring identity-related feature clustering from labeled pre-trained source to unlabeled target domain. An asymmetric domain distillation method is proposed, where identity and domain-related features are decoupled and recoupled to mitigate domain gaps. And a graph network based feature transfer structure is designed for cross-domain transformation, along with a multi-domain pseudo label refinement strategy leveraging complementary clustering information to generate accurate and noise robust labels. Extensive experiments validate the effectiveness of our proposed framework.

1 INTRODUCTION

Unsupervised domain adaptive person re-identification (UDA-ReID) transfers knowledge from a labeled source domain to an unlabeled, stylistically distinct target domain (Dai et al., 2021; Jiang et al., 2020; Feng et al., 2021). Cluster-based methods (Fan et al., 2018a; Zhai et al., 2020; Yang et al., 2020; Ge et al., 2020a) have advanced this field by iteratively generating pseudo labels (via K-Means and DBSCAN algorithms) and fine-tuning models on the target domain. However, clustering performance limits the quality of pseudo labels, and label noise inevitably affects the final results.

To improve model training, many studies (Zheng et al., 2021; Xiong et al., 2025; Cao et al., 2026; Tian et al., 2026; Gao et al., 2025) aim to reduce noise interference. Chen et al. (2020) proposed PLMC model to use multi-scale clustering with various thresholds for robust results. Li et al. (2021) introduced AdaDC model, which adaptively employs different clustering algorithms to reduce errors through their complementarity. Cho et al. (2022) presented PPLR model to refine labels by exploring complementary global and local features. However, the clustering features in target domain are extracted by source-trained model. Due to domain style differences, such features contain inherent errors determined by the model, which cannot be fully eliminated by clustering algorithms. In addition, outliers often appear during clustering (Fig. 1(a)), and most methods directly discard them to ensure the precision of labels. In fact, for target domain, these may be difficult but valuable samples, and simple discarding will affect the final training performance (Ge et al., 2020b).

In contrast, contrastive learning-based methods exploit both source and target domains via hybrid memory, storing class centroids and unclustered outliers to measure feature similarity (Fig .1(b)) (Wu et al., 2022; Zhao et al., 2022; Liu et al., 2022b; Xu et al., 2022; Cheng et al., 2022b), where unclustered outliers are treated as independent classes to supervise discriminative learning. Although achieving better performance, clustering reliability remains critical in feature learning. Ge et al. (2020b) start with reliable cluster samples and gradually add unclustered instances during training; Song & Jin (2021) reduce cluster noise by minimizing prediction variance between primary and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

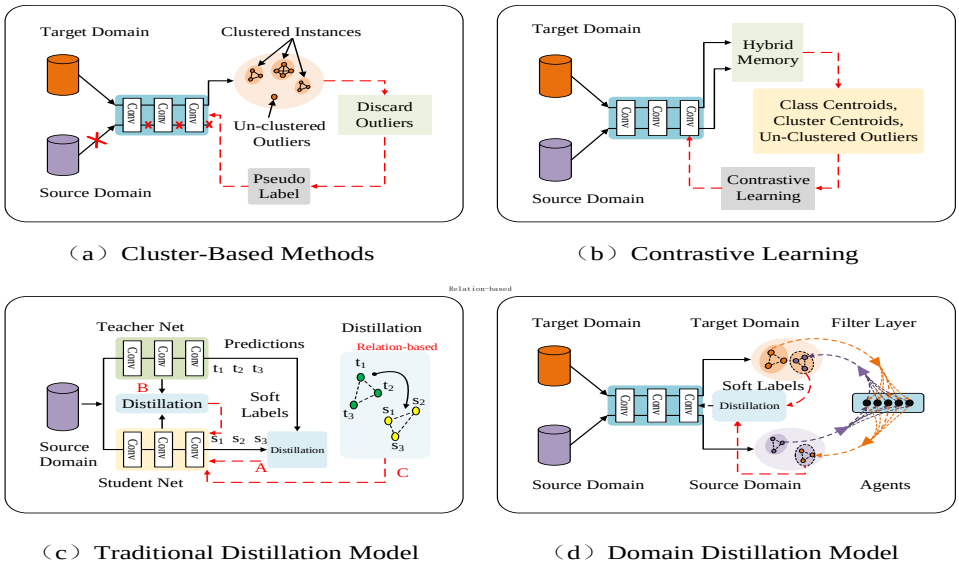


Figure 1: (a) Cluster-based Methods; (b) Contrastive learning; (c) Traditional distillation model, where A, B and C is the response-based, feature-based and relation-based knowledge distillation methods respectively; (d) Domain distillation model

auxiliary classifiers. Unlike methods using only source domain for pre-training and discarding it in fine-tuning, these frameworks take source domain as supervision for target representation learning. However, cross-domain negative samples are easily distinguishable due to large domain gaps, being simply pushed away from target queries and contributing little to model optimization.

Essentially, domain adaptive can be viewed as a process of knowledge distillation and transfer. Unlike traditional knowledge distillation(Hinton et al., 2015; Romero, 2014; Park et al., 2019; Yu et al., 2019; Jin et al., 2020), which distill knowledge between teacher and student models on a single domain, domain adaptation transfers knowledge from the source domain to the target domain by a single model. As cross-domain samples are asymmetric with mismatched categories and styles, domain adaptation does not follow the synchronous point-to-point feature transfer paradigm of traditional structures (Fig .1(c)). Instead, it transfers feature clustering behavior rather than the features themselves. An intuitive strategy is to distill shared identity-related features between different domains. It is foreseeable that the clustering precision of features in the pre-trained source domain will be higher than that in the untrained target domain. Meanwhile, labeled source domain data can be transferred to the unlabeled target domain as supplementary training supervision.

Therefore, an asymmetric domain distillation method is proposed. Theoretically, cross-domain feature distillation and transformation correspond to the decoupling and recoupling of identity-related and domain-related factors. Foremost, a dedicated layer is designed to disentangle domain factors, which contains shared agent elements from both domains. Each agent is learned to represent an identity sub-factor by measuring its similarity to the original feature. As shown in Fig .1(d), when a sample transfers across domains, its feature is decomposed and reconstructed via similarity coefficients with all agents. The reconstructed feature is then recoupled with a new domain factor and projected into the target embedding space. To this end, a domain transformation model is proposed, where the dentity-related and domain-related features are mapped into a special graph network(Kipf & Welling, 2016; Schlichtkrull et al., 2018; Hamilton et al., 2017) to rebuilt the domain spaces. In this way, features across domains can be mutually transformed for complementary clustering information. In summary, this work's major contributions for UDA-ReID are as follows:

1. An asymmetric domain distillation method is proposed, bridging the domain gap and transferring clustering knowledge from the labeled source to unlabeled target domain.
2. GCN-based feature transfer structure is introduced to reconstruct the feature space and enable effective cross-domain embedding transfer.

3. Multi-domain pseudo-label refinement approach is presented to exploit complementary clustering information and suppress label noise.

2 Related Work

Domain translation-based methods use GANs (Goodfellow et al., 2014; Deng et al., 2018; Koch et al., 2015; Zhu et al., 2017; Wei et al., 2018b) to transfer the style of source domain images to target domain for model training. Choi et al. (2018) presented StarGAN, a unified multi-domain image translation framework that achieves flexible cross-domain mapping with a single generator and discriminator. Li et al. (2019) proposed PDA-Net for pose disentanglement and adaptation. It learns a cross-domain pose-invariant feature space to focus on identity-related features and alleviate pose-variations during image translation. Zheng et al. (2019) introduced DG-Net, which decomposes latent space into three subspaces and uses cycle-consistent generation to disentangle identity-related features, ensuring generation quality and consistency. Such methods based on image-to-image translation place high demands on generative models, and there remains substantial room for improvement in the quality of generated images for practical person re-identification tasks.

Clustering-based approaches (Fan et al., 2018b; Yang et al., 2021; Zhang et al., 2021b; Xiahou et al., 2023) typically generate pseudo labels by clustering algorithms. Cheng et al. (2022a) introduced a dynamic contrastive learning and probability distillation framework, which extracts effective self-supervised signals to facilitate robust representations. Lee et al. (2023) presented a camera-driven curriculum learning strategy, transferring knowledge progressively from source to target domain with guidance of camera labels. Gao et al. (2024b) formulated a mutual distillation framework that employs bidirectional distillation to narrow domain gap between different datasets. Yang et al. (2025b) constructed an inter-domain equilibrium network, balancing domain distributions to mitigate cross-domain shifts. Tian et al. (2025b) devised a part based feature complementary denoising method, which enhances feature learning via part wise denoising and complementation. Wang et al. (2025) engineered a noise perception self-supervised learning framework, focusing on modeling feature noise to boost the robustness and discriminability of representations. These approaches seek to exploit the source domain data to boost model training on the target domain. However, weakened supervision arising from the substantial inter-domain discrepancy still poses a pressing challenge.

3 Methodology

This section elaborates on the proposed asymmetric domain knowledge distillation method (ADKD) and the corresponding multi-task training strategy. An overview is first provided, followed by the core components—the domain style transformation model and the label assignment and sample selection—and a detailed description of the objective functions and optimization process.

3.1 Overview

Built upon ResNet-50 (He et al., 2016) backbone with domain-specific BN (Chang et al., 2019a), as shown in Fig. 2, feature disentanglement and reconstruction module based on a graph convolutional network (DT-GCN) is employed to complete the domain transformation of images. At its core, the model deploys multi-domain complementary clustering to adaptively adjust the label assignment and sample selection, thereby improving the training efficiency on unseen target domains.

3.2 Domain Style Transformation Model

In Fig. 2(a), a feature decomposition module is designed after the backbone network to model domain-related and identity-related features across domains. A specialized Filter Layer (Agents) is deployed between domains, where pedestrian features are filtered and reconstructed by identity agent sub-factors encoding identity-related features to eliminate domain-related information.

Identity-related Features: let $I = \{a_i \mid i \in [1, N_a]\}$ be the set of identity agent sub-factors, $S = \{(x_{s,i}, y_{s,i}) \mid i \in [1, N_s]\}$ be the labeled source domain samples, and $T = \{(x_{t,i}, y_{t,i}) \mid i \in [1, N_t]\}$ be the target domain samples, where N_a, N_s and N_t are their respective numbers. $x_{s,i}$ and $x_{t,i}$ are features of source and target domains, $y_{s,i}$ is the label of source domain, and $y_{t,i}$ is the

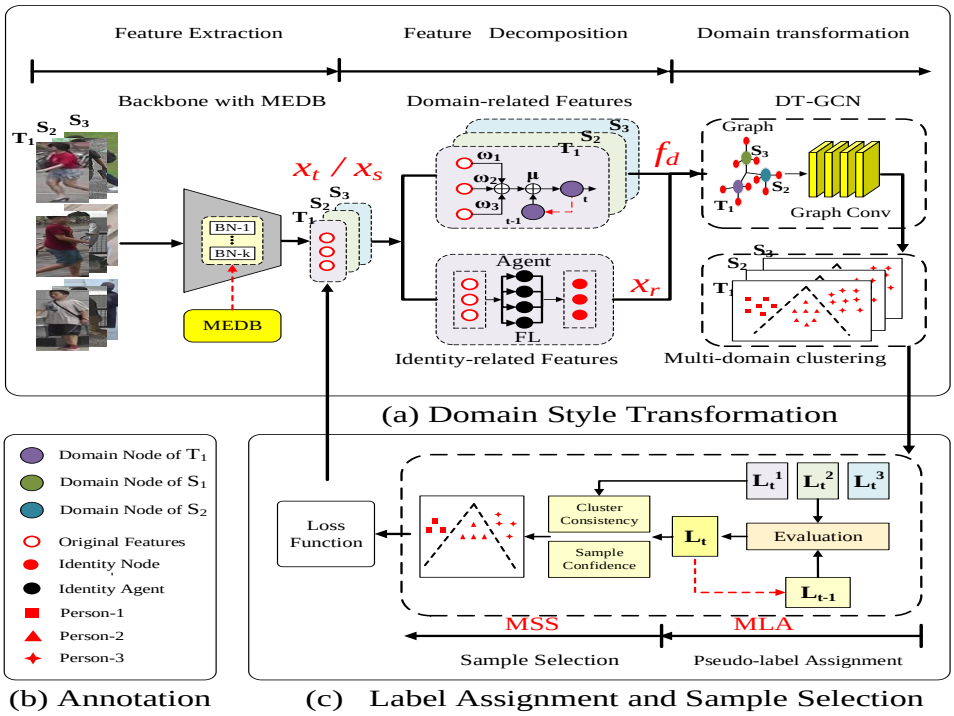


Figure 2: Domain distillation model (ADKD), where S_1, S_2 and T denote 2 source and 1 target domains respectively, and DS-BN stands for domain specific batch normalization.

generated pseudo label of target domain. When feature x is fed into the filter layer, the similarity matrix between x and each identity agent sub-factor is computed as:

$$\begin{aligned} [\gamma_{x,1}, \ldots, \gamma_{x,N_a}] &= \frac{1}{\sum_{a_k \in I} e^{(a_k \cdot x)}} \exp\big((a_1 \cdot x), \ldots, (a_{N_a} \cdot x)\big) \\ \text{s.t.} \quad 0 < \gamma_{x,k} < 1; &\quad \sum_{k=1}^{N_a} \gamma_{x,k} = 1; \quad x \in S \cup T. \end{aligned} \tag{1}$$

where $\gamma_{x,k}$ is the similarity between x and a_k, as well as the proportion of identity information described by a_k in x relative to the total identity features. By this similarity matrix, appropriate sub-factors are selected to reconstruct identity features and eliminate domain style factors:

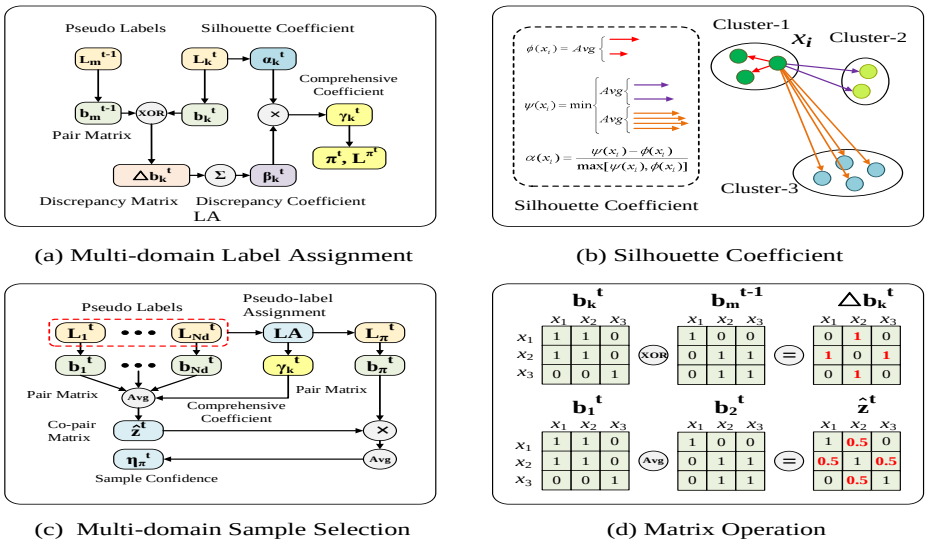
$$\begin{aligned} x_r &= \gamma_{x,1} \cdot a_1 + \gamma_{x,2} \cdot a_2 + \cdots + \gamma_{x,N_a} \cdot a_{N_a} \\ \text{s.t.} \quad x_r \in R; &\quad R = \{(x_{r,i}, y_{r,i}) \mid i \in [1, N_s + N_t]\} \end{aligned} \tag{2}$$

where x_r is the reconstructed feature of x via filter layer; R is the set of reconstructed features for source/target domains; $y_{r,i}$ is the identity label.

To ensure x_r is free of domain-related elements, these sub-factors are trained with joint identity classification and domain confusion:

$$\begin{aligned} L(x, \theta_F, \theta_I, \theta_D) &= -\frac{1}{N_x} \sum_{i=1}^{N_x} y_i \cdot \log P(y_i | x_i; \theta_I) + L_D(x, \theta_F, \theta_D), \\ \text{s.t.} \quad &\begin{cases} \theta_I = \theta_S; & \text{if } (y_i, x_i) \in S, \\ \theta_I = \theta_T; & \text{if } (y_i, x_i) \in T, \\ \theta_I = \theta_R; & \text{if } (y_i, x_i) \in R. \end{cases} \end{aligned} \tag{3}$$

where the first term is identity classification loss; N_x is batch size; θ_F is the parameters of backbone; θ_S, θ_T, θ_R are identity classifier parameters for source, target, and reconstructed samples; θ_D is



(a) Multi-domain Label Assignment (b) Silhouette Coefficient

(c) Multi-domain Sample Selection (d) Matrix Operation

Figure 3: Multi-domain Label Assignment and Sample Selection.

domain classifier parameter; $L_D(x;\theta_F,\theta_D)$ is domain confusion loss:

$$L_D(x,\theta_F,\theta_D)=\begin{cases}-\dfrac{1}{N_x}\sum_{i=1}^{N_x} d_i\cdot\log P(d_i|x_i;\theta_F,\theta_D) & \text{if } x_i\in S\cup T,\\ -\dfrac{1}{N_x}\sum_{i=1}^{N_x}\dfrac{1}{N_d}\sum_{j=1}^{N_d}\log P(d_j|x_{r,i};\theta_D) & \text{if } x_{r,i}\in R.\end{cases} \tag{4}$$

where N_d is domain numbers; d_i is domain label; $P(d_i|x_i;\theta_D)$ is the probability of x_i belonging to domain d_i. Filter layer updates parameters via batch-wise adversarial training: (1) train backbone and domain classifier to optimize θ_F and θ_D; (2) fix θ_F and θ_D and feed $x_{r,i}$ into the domain classifier for domain confusion classification to train identity agent filter layer, where $x_{r,i}$ is predicted to each domain class with uniform probability $(1/N_d)$ to obtain domain-invariant features.

Domain-related Feature: the global style of a domain, obtained by weighting its all samples:

$$f_d^k=\omega_1\cdot x_{k,1}+\omega_2\cdot x_{k,2}+\cdots+\omega_{N_k}\cdot x_{k,N_k},$$
$$\text{s.t.}\quad\begin{cases}0<\omega_i<1;\quad\sum_{i=1}^{N_k}\omega_i=1,\\ \omega_i=f(x_{k,1},\ldots,x_{k,N_k})\end{cases} \tag{5}$$

where f_d^k is domain style feature of domain k and $x_{k,i}$ $(i\in[1,N_k])$ are its samples; $f(\cdot)$ is fully connected layer to generate weights ω_i. During training, one batch only covers part data of one domain, so f_d^k estimated from successive batches are iteratively updated via moving average:

$$(f_d^k)^t=(1-\varepsilon)\cdot(f_d^k)^{t-1}+\varepsilon\cdot(f_d^k)^t,\quad\text{s.t.}\quad 0<\varepsilon<1 \tag{6}$$

where ε is adjustment coefficient and t is iteration number. Identity node x_r from filter layer and domain node f_d^k are then fed into a 4-layers graph network (DT-GCN) as interconnected nodes for coupling, yielding domain specific spaces. In Fig. 2(a), each identity node is connected to different domain node for space transformation, where domain nodes are interconnected with each other, thereby forming a unified multi-domain space with both distinct separation and reasonable association to avoids information loss caused by large domain gaps during domain transformation.

3.3 Label Assignment and Sample Selection

Given N_d domains: $\Phi_1,\ldots,\Phi_{N_d}$, the domain distillation model (ADKD) aims to transfer feature x into different domains to find an optimal clustering space π^t for pseudo labels L^{π^t} (Fig. 3.(a)) and select reliable samples (Fig. 3.(c)) by leveraging inter-domain complementarity.

Multi-domain Label Assignment: pseudo label quality relies on clustering reliability within each domain. Intra-cluster similarity and inter-cluster divergence are measured to evaluate sample's correlation with its category vs. others, which can be quantified by the silhouette coefficient (Fig. 3(b)):

$$\alpha_k^t = \frac{1}{N_k} \cdot \sum_{i=1}^{N_k} g_k^t(x_i) = \frac{1}{N_k} \cdot \sum_{i=1}^{N_k} \frac{\psi_k^t(x_i) - \phi_k^t(x_i)}{\max(\psi_k^t(x_i), \phi_k^t(x_i))}$$
$$\text{s.t.} \quad \begin{cases} \psi(x_i) = \frac{1}{|C_i|-1} \sum_{x_v \in C_i, x_v \neq x_i} \rho(x_i, x_v) \\ \phi(x_i) = \min_{k \neq i} \frac{1}{|C_i|} \sum_{x_v \in C_k} \rho(x_i, x_v) \end{cases} \tag{7}$$

where α_k^t is the silhouette coefficient of domain k at iteration t and N_k is its sample number; C_i is the i-th category; $\psi(x_i)$ is the intra-cluster similarity (average distance between sample x_i and same-category samples); $\phi(x_i)$ is the inter-cluster divergence (average distance between x_i and adjacent-category samples); $\rho(x_i, x_v)$ is the featuere distance between x_i and x_v.

In addition, performance improvement between adjacent iterations is another criterion for pseudo label selection. Keeping a reasonable difference enables model to learn more diverse information.

$$\beta_k^t = \sum_{i=1}^{N_k} \sum_{v=i+1}^{N_k} \Delta b_k^t(x_i, x_v) = \sum_{i=1}^{N_k} \sum_{v=i+1}^{N_k} \left[b_k^t(x_i, x_v) \oplus b_m^{t-1}(x_i, x_v) \right]$$
$$\text{s.t.} \quad b_k(x_i, x_v) = \begin{cases} 1 & \text{if} \quad L_k(x_i) = L_k(x_v) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where β_k^t is label difference between iteration t and $t-1$; $L_k(x_i)$ is pseudo label of sample x_i in domain k; $b_k(x_i, x_v)$ is indicator matrix for whether x_i and x_v are in the same cluster; $\oplus$ is XOR operation (Fig. 3(d)); Δb_k^t is difference matrix between labels generated by domain k and those by domain m selected in the previous iteration $t-1$. Finally, the comprehensive coefficient γ_k^t can be calculated by α_k^t and β_k^t to select the optimal clustering space π^t for current iteration t:

$$\pi^t = \arg\max_k \gamma_k^t \quad \text{s.t.} \quad \gamma_k^t = \alpha_k^t \cdot \beta_k^t, \quad k \in [1, N_d] \tag{9}$$

Multi-domain Sample Selection: incorrectly clustered samples is a major factor impairing model training. Sample pairs with the same label across domains are likely to be correctly clustered (positive), while those with different labels tend to be incorrectly clustered (negative). Thus, cross domain comprehensive clustering confidence is computed to select the most reliable positive pairs:

$$\eta_\pi^t(x_i) = \frac{1}{\sum_{\nu=1}^{N_k} b_\pi^t(x_i, x_\nu)} \cdot \sum_{\nu=1}^{N_k} \left[\hat{z}^t(x_i, x_\nu) \cdot b_\pi^t(x_i, x_\nu) \right]$$
$$\text{s.t.} \quad \begin{cases} z^t(x_i, x_\nu) = \frac{1}{N_d} \cdot \sum_{k=1}^{N_d} b_k(x_i, x_\nu) \\ \hat{z}^t(x_i, x_\nu) = \frac{1}{\sum_{k=1}^{N_d} \gamma_k^t} \cdot \sum_{k=1}^{N_d} \left[\gamma_k^t \cdot b_k(x_i, x_\nu) \right] \end{cases} \tag{10}$$

where N_d is domain number; $b_k(x_i, x_\nu)$ is the relational matrix proposed above that describes whether the labels of x_i and x_ν are consistent in domain k; $z^t(x_i, x_\nu)$ is the average confidence of whether the labels of x_i and x_ν are consistent across different domains; $\hat{z}^t(x_i, x_\nu)$ is the confidence matrix by weighting $b_k(x_i, x_\nu)$ with the comprehensive coefficient γ_k^t considering the importance of different domains; $\eta_\pi^t(x_i)$ is the final confidence of sample x_i by combining the confidence matrix with the relational matrix of the selected domain π in the current iteration cycle t. Then, samples are sorted and the top-q are collected in a gradually expanding manner to train the model:

$$q^t = \lambda_{\text{start}} + \frac{t}{n_t} \cdot (\lambda_{\text{end}} - \lambda_{\text{start}}) \tag{11}$$

where n_t is the total training iterations; the q is initial set to λ_{start} and increases linearly to λ_{end} with training iterations, gradually enhancing the learning of negative pairs from easy to hard.

Table 1: Comparison of ADKD with other methods.

Methods	Reference	Duke → Market				Market → Duke			
		mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
MCRN (Wu et al., 2022)	AAAI 22	83.8	93.8	97.5	98.5	71.5	84.5	91.7	93.8
CACHE (Liu et al., 2022a)	TCSVT 23	83.1	93.4	97.5	98.2	71.7	83.5	91.4	93.9
DREAMT (Tao et al., 2022)	TMM 23	81.4	93.3	98.0	98.7	69.8	82.3	90.9	93.6
DKD-MPL (Zhu et al., 2024)	TMM 24	86.0	94.5	98.0	98.5	38.0	68.1	79.0	81.7
Dual-GCL (Zhang et al., 2024)	TCDS 24	82.7	93.5	98.0	99.0	39.3	70.5	81.0	83.6
IIDS (Xuan & Zhang, 2022)	TPAMI 24	78.0	91.2	96.2	97.7	68.7	82.1	90.8	93.7
DMD (Gao et al., 2024a)	TMM 25	78.0	91.2	96.2	97.7	68.7	82.1	90.8	93.7
IDE (Yang et al., 2025a)	TIP 25	83.6	93.3	97.6	98.4	31.9	60.6	71.5	75.9
OAE (Li et al., 2026)	TMM 26	83.0	93.4	97.4	98.6	-	-	-	-
ADKD	**Ours**	**86.4**	**94.9**	**98.7**	**99.4**	**73.3**	**84.8**	**92.9**	**94.4**

Methods	Reference	Duke→MSMT				Market→MSMT			
		mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
HGA (Zhang et al., 2021a)	AAAI 21	26.8	58.6	64.7	69.2	25.5	55.1	61.2	65.5
RDSBN+MDIF (Bai et al., 2021)	CVPR 21	33.6	64.0	75.6	79.6	30.9	61.2	73.1	77.4
MCRN (Wu et al., 2022)	AAAI 22	35.7	67.5	77.9	81.6	32.8	64.4	75.1	79.2
CACHE (Liu et al., 2022a)	TCSVT 23	31.0	57.8	70.4	75.3	31.3	58.0	69.8	74.5
DREAMT (Tao et al., 2022)	TMM 23	30.3	58.0	70.5	75.3	25.3	51.6	64.3	69.7
DKD-MPL (Zhu et al., 2024)	TMM 24	38.0	68.1	79.0	81.7	-	-	-	-
IDE (Yang et al., 2025a)	TIP 25	-	-	-	-	31.9	60.6	71.5	75.9
BFFT (Tian et al., 2025a)	TIFS 25	37.5	65.7	78.8	82.0	36.1	66.2	76.9	80.5
OAE (Li et al., 2026)	TMM 26	-	-	-	-	35.3	65.3	76.9	81.0
ADKD	**Ours**	**40.6**	**70.9**	**81.3**	**83.2**	**37.3**	**67.9**	**78.7**	**81.6**

4 EXPERIMENTS

4.1 IMPLEMENTATION DETAILS AND EVALUATION SETTINGS

The ResNet-50 is employed as backbone and initialized with the weights pretrained on ImageNet (Deng et al., 2009). Experiments are conducted on 2 NVIDIA-P100 GPUs. The Adam optimizer is utilized to optimize the network with a weight decay of 0.0005. All images are uniformly resized to 256×128, with the batch size set to 128. The hyperparameters (ε, λ_{start}, λ_{end}) in equation 6 and equation 11 are set to 0.3, 0.5, and 1.0. Several data augmentation strategies are adopted, including random flipping, cropping and erasing (Nie et al., 2024). The model is first pre-trained on source domains, and the initial learning rate is set to 0.00035 then decreased to 1/10 of its original value at the 40-th and 70-th epochs with a total training duration of 80 epochs. Subsequently, the model is fine-tuned on target domain with learning rate fixed at 0.00035 for a total of 100 epochs.

The model is evaluated on 4 widely used datasets: Market-1501 (Zheng et al., 2015), DukeMTMC-ReID (Zheng et al., 2017), CUHK03 (Li et al., 2014) and MSMT17 (Wei et al., 2018a). For simplicity, Market, Duke, CUHK and MSMT are used as abbreviations. Evaluation is performed based on the standard protocol by the mean average precision (mAP) (Everingham et al., 2010) and cumulative matching characteristic (CMC) curve (Bolle et al., 2005) represented by the Rank-1 (R-1), Rank-5 (R-5) and Rank-10 (R-10) accuracy metrics.

4.2 COMPARISON WITH OTHER METHODS

Comparison with other methods on single-source domain is conducted on 4 cross-domain person ReID tasks (Table 1), the ADKD achieves 73.3%/84.8% of mAP/R-1 metrics on the Market→Duke task and 86.4%/94.9% on the Duke→Market task, outperforming all competitors. Specifically, The OAE is surpassed by 3.4% mAP and 1.5% R-1 on the challenging Duke→Market task. And for MSMT, 37.3%/67.9% and 40.6%/70.9% of mAP/R-1 are obtained on Market→MSMT and Duke→MSMT task, exceeding the recent methods: BFFT, IDE and OAE,which show that our method delivers stable and substantial improvements.

Comparison with other methods on multi-source domains is conducted on Table 2. The core of our method lies in effectively exploiting the complementary effects across multiple distinct domains. The table shows that our method consistently outperforms all competing multi-domain adaptation methods across all tasks. Notably, when targeting the large scale MSMT dataset, our method

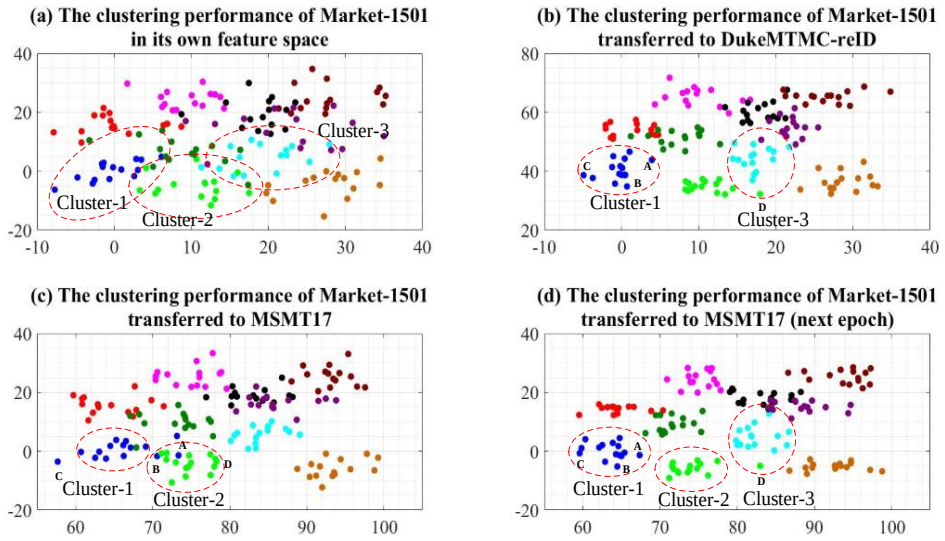


Figure 4: The analysis of domain clustering performance.

Table 2: Comparison of ADKD with other multi-domain methods.

Methods	Market+CUHK+ Duke→ MSMT				Duke+CUHK+ Market → MSMT				MSMT+CUHK+ Duke → Market				MSMT+CUHK+ Market → Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
MMT+GRL (Ganin & Lempitsky, 2015)	22.6	46.3	59.8	65.8	22.6	46.3	59.8	65.8	77.1	90.4	96.8	97.8	64.3	77.6	88.1	91.3
MMT+MM (Peng et al., 2019)	22.7	47.1	60.4	65.8	22.7	47.1	60.4	65.8	78.1	91.2	96.8	98.0	64.2	77.6	87.9	91.4
MMT+DSBN (Chang et al., 2019b)	22.6	49.2	62.1	67.4	22.6	49.2	62.1	67.4	81.1	92.8	97.3	98.5	65.6	79.6	89.1	92.1
RDSBN+MDIF (Bai et al., 2021)	34.9	64.7	76.2	80.2	34.9	64.7	76.2	80.2	86.0	94.8	97.9	98.6	68.9	82.1	90.4	93.0
ADKD	**43.7**	**72.5**	**82.8**	**85.5**	**43.7**	**72.5**	**82.8**	**85.5**	**87.2**	**95.7**	**98.8**	**99.5**	**75.7**	**86.2**	**94.0**	**95.8**

achieves significant improvements over the second best results with the margins of up to 8.8% mAP and 7.8% R-1 accuracy, which demonstrates its clear superiority in the multi-source setting.

4.3 Ablation Study

Ablation study on different components is conducted on Table 3a, where FL is the filter layer, MLA is the multi-domain label assignment and MSS is the multi-domain sample selection. Taking the Duke→Market task as an example, adding FL improves mAP by 4.3% and R-1 by 1.4%. The subsequent MLA further increases mAP by 1.6% and R-1 by 2.3%. The full module with integrated MSS achieves the best performance of 86.4% mAP and 94.9% R-1, corresponding to an additional 2.3% mAP and 1.1% R-1 improvement. These results validate the distinct and complementary roles of FL, MLA and MSS in boosting cross-domain performance.

Ablation study on domain-related nodes is conducted on Table 3b, where the weighted moving average node is compared with the average and weighted average nodes. In particular, the average node directly averages all in-domain samples, while the weighted one uses learned weights. From this table, performance of the weighted average node outperforms the unweighted one, and the weighted moving average node achieves the best result, as it leverages historical global information to counteract randomness in the current single batch and obtain a robust domain-node feature, which validates its effectiveness for robust domain feature integration.

Exploration with visual analytics is presented in Fig. 4 and Fig. 5. Fig. 4 shows the clustering performance in their un-trained native space (Fig. 4(a)) and other pre-trained source domains (Fig. 4(b) and (c)). Obviously, at the initial stage, the performance in their un-trained native space is markedly inferior to that following their transfer to pre-trained source domain spaces, and some samples show distinct clustering behaviors in different spaces. For instance, in the MSMT space, samples A and B originally belonging to cluster-1 are assigned to cluster-2 and C becomes an outlier, whereas in the Duke space, D from cluster-2 is assigned to cluster-3. Accordingly, when the Duke

Table 3: Ablation study on different components and domain-related nodes of ADKD.

Methods	Duke→Market				Market→Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
ResNet-50	68.1	85.6	94.3	96.4	57.4	73.5	84.2	88.1
+DS-BN	75.2	89.0	95.1	97.9	66.6	78.8	87.0	89.9
+FL	79.5	90.4	95.7	97.9	68.2	80.5	88.4	91.2
+DT-GAN	82.5	91.5	95.7	97.9	70.2	81.5	89.8	92.5
+MLA	84.1	93.8	96.1	98.0	72.7	83.1	91.7	93.6
+MSS	86.4	94.9	98.7	99.4	73.3	84.8	92.9	94.4

(a) Effectiveness of different components.

Methods	Duke→Market				Market→Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
ResNet-50 (He et al., 2016) +	68.1	85.6	94.3	96.4	57.4	73.5	84.2	88.1
Average Domain Node	83.6	92.1	97.4	98.2	70.5	80.9	91.0	93.5
Weighted Average Node	85.3	93.5	98.1	98.7	72.1	83.5	92.3	94.0
Weighted Moving Average Node	86.4	94.9	98.7	99.4	73.3	84.8	92.9	94.4

(b) Effectiveness of domain-related nodes.

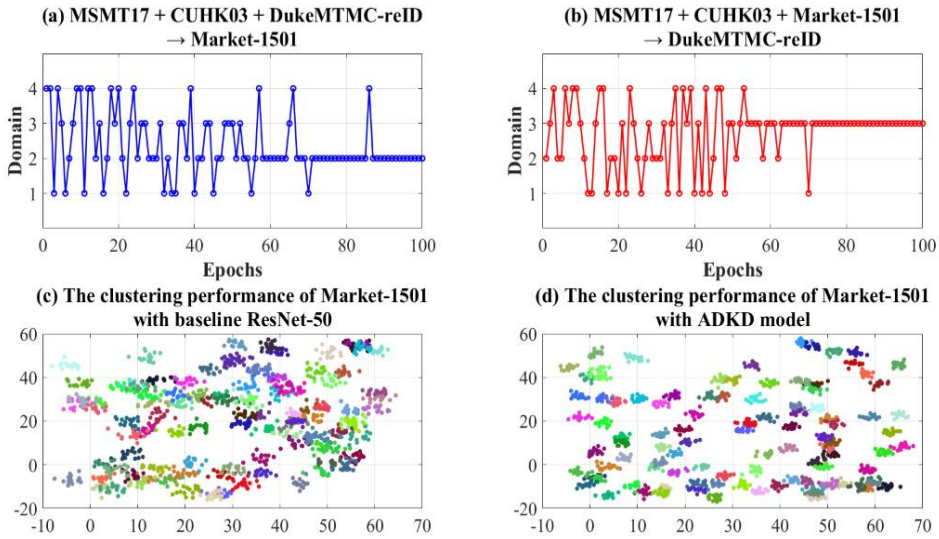


Figure 5: (a-b) are the process of selecting clustering spaces, where 1, 2, 3, 4 denote the domains: CUHK, Market, Duke, MSMT; (c-d) are the t-SNE visualizations of the model with or without the ADKD module, where the samples are selected from 100 pedestrians in the Market-1501.

space is chosen as the label-generation space for the current training epoch, A, B and C are corrected to cluster-2 in the next epoch (Fig. 4(d)); conversely, D falls into a wrong cluster in the MSMT space due to the incorrect label generated by the Duke space. Hence, iterative selection of diverse clustering spaces during training is required to progressively refine sample labels.

Fig. 5 illustrates the process of adaptively selecting clustering spaces for label generation during model training. In Fig. 5(a-b), at the initial stage with domains 2 (Market) and 3 (Duke) acting as target domains, the model tends to adopt pseudo labels derived from the source domains. As training proceeds, the target domain evolves toward an organized state. Correspondingly, the clustering performance of its samples within the native space surpasses that of other spaces and gradually stabilizes. Fig. 5(c-d) describe the clustering quality with and without ADKD module. After introducing the ADKD, the inter-class distances increase significantly compared with those of the baseline, which further verifies the effectiveness of our method.

5 Conclusion

This paper addresses the core challenges of noisy pseudo-labels, cross-domain feature bias, and inefficient knowledge transfer in unsupervised domain-adaptive person re-identification. It formulates UDA-ReID as an asymmetric domain distillation task, decoupling and recoupling identity and domain features to alleviate domain gaps. A graph-based feature transfer structure and multi-domain pseudo-label refinement strategy are proposed to generate accurate and noise-robust target pseudo-labels. Extensive experiments verify the effectiveness of the framework in mitigating existing limitations and boosting cross-domain person re-identification performance.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and grammatical refinement of the manuscript text. We have not used generative AI tools for research idea formulation, algorithm design, experimental design, data analysis, or drawing conclusions. Other disclosure tasks are not applicable to this work. We have manually reviewed and revised all AI-polished text. Two authors carefully proof-read the AI-revised paragraphs to ensure consistency with our original technical statements and to avoid unintended semantic alterations. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Zechen Bai, Zhigang Wang, Jian Wang, Di Hu, and Errui Ding. Unsupervised multi-source domain adaptation for person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12914–12923, 2021.

Ruud M Bolle, Jonathan H Connell, Sharath Pankanti, Nalini K Ratha, and Andrew W Senior. The relation between the roc curve and the cmc. In *Fourth IEEE workshop on automatic identification advanced technologies (AutoID'05)*, pp. 15–20. IEEE, 2005.

Ge Cao, Qing Tang, Vo Xuan Thuy, Adri Priandana, and Kang-Hyun Jo. Optimal proxy mining contrastive network for unsupervised person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(2):685–696, 2026. doi: 10.1109/TCSVT.2026.3548923.

Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7346–7354, 2019a. doi: 10.1109/CVPR.2019.00753.

Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 7354–7362, 2019b.

Shuni Chen, Zheyi Fan, and Jianyuan Yin. Pseudo label based on multiple clustering for unsupervised cross-domain person re-identification. *IEEE Signal Processing Letters*, 27:1460–1464, 2020.

De Cheng, Jingyu Zhou, Nannan Wang, and Xinbo Gao. Hybrid dynamic contrast and probability distillation for unsupervised person re-id. *IEEE Transactions on Image Processing*, 31:3334–3346, 2022a.

De Cheng, Jingyu Zhou, Nannan Wang, and Xinbo Gao. Hybrid dynamic contrast and probability distillation for unsupervised person re-id. *IEEE Transactions on Image Processing*, 31:3334–3346, 2022b.

Yoonki Cho, Woo Jae Kim, Seunghoon Hong, and Sung-Eui Yoon. Part-based pseudo label refinement for unsupervised person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7308–7318, 2022.

Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8789–8797, 2018.

Yongxing Dai, Jun Liu, Yan Bai, Zekun Tong, and Ling-Yu Duan. Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification. *IEEE Transactions on Image Processing*, 30:7815–7829, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 994–1003, 2018.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338, 2010.

Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 14(4):1–18, 2018a.

Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 14(4):1–18, 2018b.

Hao Feng, Minghao Chen, Jinming Hu, Dong Shen, Haifeng Liu, and Deng Cai. Complementary pseudo labels for unsupervised domain adaptation on person re-identification. *IEEE Transactions on Image Processing*, 30:2898–2907, 2021.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pp. 1180–1189. PMLR, 2015.

Xingyu Gao, Zhenyu Chen, Jianze Wei, Rubo Wang, and Zhijun Zhao. Deep mutual distillation for unsupervised domain adaptation person re-identification. *IEEE Transactions on Multimedia*, 27: 1059–1071, 2024a.

Xingyu Gao, Zhenyu Chen, Jianze Wei, Rubo Wang, and Zhijun Zhao. Deep mutual distillation for unsupervised domain adaptation person re-identification. *IEEE Transactions on Multimedia*, 27: 1059–1071, 2024b.

Xingyu Gao, Zhen Chen, Jianwei Ru, Rubo Wang, and Zhijun Zhao. Deep mutual distillation for unsupervised domain adaptation person re-identification. *IEEE Transactions on Multimedia*, 27: 1095–1106, 2025. doi: 10.1109/TMM.2025.3512345.

Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. *arXiv preprint arXiv:2001.01526*, 2020a.

Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *Advances in neural information processing systems*, 33:11309–11321, 2020b.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Kongzhu Jiang, Tianzhu Zhang, Yongdong Zhang, Feng Wu, and Yong Rui. Self-supervised agent learning for unsupervised cross-domain person re-identification. *IEEE Transactions on Image Processing*, 29:8549–8560, 2020.

Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Uncertainty-aware multi-shot knowledge distillation for image-based object re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11165–11172, 2020.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, pp. 1–30. Lille, 2015.

Geon Lee, Sanghoon Lee, Dohyung Kim, Younghoon Shin, Yongsang Yoon, and Bumsub Ham. Camera-driven representation learning for unsupervised domain adaptive person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11453–11462, 2023.

Longfei Li, Qing Han, Weidong Min, Qi Wang, Meng Zhu, Zheng Huang, Ziyang Deng, and Ruibin Chen. Outliers adaptation exploration and centroids matching label refinement for unsupervised person re-identification. *IEEE Transactions on Multimedia*, 28:3905–3920, 2026. doi: 10.1109/TMM.2026.3654461.

Shihua Li, Mingkuan Yuan, Jie Chen, and Zhilan Hu. Adadc: Adaptive deep clustering for unsupervised domain adaptation in person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6):3825–3838, 2021.

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*, pp. 152–159, 2014.

Yu-Jhe Li, Ci-Siang Lin, Yan-Bo Lin, and Yu-Chiang Frank Wang. Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7919–7929, 2019.

Yuxuan Liu, Hongwei Ge, Liang Sun, and Yaqing Hou. Complementary attention-driven contrastive learning with hard-sample exploring for unsupervised domain adaptive person re-id. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):326–341, 2022a.

Yuxuan Liu, Hongwei Ge, Liang Sun, and Yaqing Hou. Complementary attention-driven contrastive learning with hard-sample exploring for unsupervised domain adaptive person re-id. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):326–341, 2022b.

Ren Nie, Jin Ding, Xue Zhou, and Xi Li. Rethinking normalization layers for domain generalizable person re-identification. In *European conference on computer vision*, pp. 267–284. Springer, 2024.

Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. Relational knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3967–3976, 2019.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1406–1415, 2019.

Adriana Romero. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550*, 2014.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.

Xulin Song and Zhong Jin. Robust label rectifying with consistent contrastive-learning for domain adaptive person re-identification. *IEEE Transactions on Multimedia*, 24:3229–3239, 2021.

Yusheng Tao, Jian Zhang, Jiajing Hong, and Yuesheng Zhu. Dreamt: Diversity enlarged mutual teaching for unsupervised domain adaptive person re-identification. *IEEE Transactions on Multimedia*, 25:4586–4597, 2022.

Qing Tian, Xiang Liu, Jixin Sun, Jun Wan, and Zhen Lei. Unsupervised domain adaptation person re-identification: Bridged by feature fusion transitional domain. *IEEE Transactions on Information Forensics and Security*, 2025a.

Qing Tian, Bin Wang, Xiang Liu, Jiashuo Shen, Keyang Cheng, Weihua Ou, and Zhen Lei. Part-based feature complementary denoising for unsupervised person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025b.

Qing Tian, Bin Wang, Xiang Liu, Jiashu Shen, Keyang Cheng, Weihua Ou, and Zhen Lei. Part-based feature complementary denoising for unsupervised person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(2):1371–1382, 2026. doi: 10.1109/TCSVT.2026.3549012.

Jingya Wang, Jianfeng Wen, Weiping Ding, Chunlin Yu, Xiatian Zhu, and Zhiyong Wang. Noise perception self-supervised learning for unsupervised person re-identification. *IEEE transactions on emerging topics in computational intelligence*, 2025.

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*, pp. 79–88, 2018a.

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*, pp. 79–88, 2018b.

Yuhang Wu, Tengteng Huang, Haotian Yao, Chi Zhang, Yuanjie Shao, Chuchu Han, Changxin Gao, and Nong Sang. Multi-centroid representation network for domain adaptive person re-id. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 2750–2758, 2022.

Yaotao Xiahou, Ning Li, and Xiaochao Li. Identity consistency multi-viewpoint generative aggregation for person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(3):1441–1455, 2023.

Mingfu Xiong, Kaikai Hu, Hongyuan Wang, Zhongyuan Hu, Xiankang Yang, and Feng Liu. Adaptive clustering and weighted regularized contrastive learning framework for unsupervised person re-identification. *IEEE Transactions on Multimedia*, 27:1–12, 2025. doi: 10.1109/TMM.2025.3548623.

Mingyuan Xu, Haiyun Guo, Yuheng Jia, Zhitao Dai, and Jinqiao Wang. Pseudo label rectification with joint camera shift adaptation and outlier progressive recycling for unsupervised person re-identification. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):3395–3406, 2022.

Shiyu Xuan and Shiliang Zhang. Intra-inter domain similarity for unsupervised person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3):1711–1726, 2022.

Fengxiang Yang, Ke Li, Zhun Zhong, Zhiming Luo, Xing Sun, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Shaozi Li. Asymmetric co-teaching for unsupervised cross-domain person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 12597–12604, 2020.

Mingliang Yang, Jing Zhao, Da Huang, and Ji Wang. Progressive unsupervised domain adaptation for image-based person re-identification. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 7730–7736. IEEE, 2021.

Xi Yang, Wenjiao Dong, Gu Zheng, Nannan Wang, and Xinbo Gao. Idenet: an inter-domain equilibrium network for unsupervised cross-domain person re-identification. *IEEE Transactions on Image Processing*, 2025a.

Xi Yang, Wenjiao Dong, Gu Zheng, Nannan Wang, and Xinbo Gao. Idenet: an inter-domain equilibrium network for unsupervised cross-domain person re-identification. *IEEE Transactions on Image Processing*, 2025b.

Zhengxu Yu, Zhongming Jin, Long Wei, Jishun Guo, Jianqiang Huang, Deng Cai, Xiaofei He, and Xian-Sheng Hua. Progressive transfer learning for person re-identification. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4220–4226, 2019. doi: 10.24963/ijcai.2019/586.

Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9021–9030, 2020.

Lin Zhang, Ran Song, Yifan Wang, Qian Zhang, and Wei Zhang. Dual-graph contrastive learning for unsupervised person reidentification. *IEEE Transactions on Cognitive and Developmental Systems*, 16(4):1352–1363, 2024.

Minying Zhang, Kai Liu, Yidong Li, Shihui Guo, Hongtao Duan, Yimin Long, and Yi Jin. Unsupervised domain adaptation for person re-identification via heterogeneous graph alignment. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 3360–3368, 2021a.

Xiao Zhang, Yixiao Ge, Yu Qiao, and Hongsheng Li. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3436–3445, 2021b.

Xinpeng Zhao, Xiao Dou, and Xiaowei Zhang. Cluster-aware diversity samples mining for unsupervised person re-identification. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1707–1712. IEEE, 2022.

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*, pp. 1116–1124, 2015.

Yi Zheng, Shixiang Tang, Guolong Teng, Yixiao Ge, Kaijian Liu, Jing Qin, Donglian Qi, and Dapeng Chen. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8371–8381, 2021.

Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE international conference on computer vision*, pp. 3754–3762, 2017.

Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2138–2147, 2019.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

Wenjie Zhu, Bo Peng, and Wei Qi Yan. Dual knowledge distillation on multiview pseudo labels for unsupervised person re-identification. *IEEE Transactions on Multimedia*, 26:7359–7371, 2024.

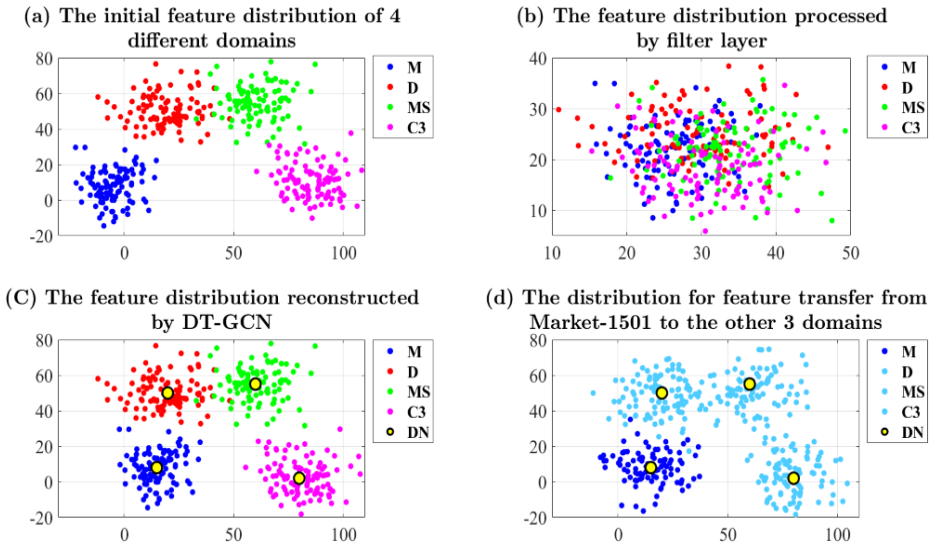


Figure 6: Feature distribution via FL filtering layer and DT-GCN domain transformation layer, where M, D, MS, C3 represent the Market-1501, DukeMTMC-reID, MSMT17 and CUHK03 domain spaces respectively, and DN denotes the corresponding domain node.

A Appendix

A.1 Feature Visualization Analysis via FL filtering layer and DT-GCN

Fig. 6 represents the the feature distribution of 4 different domains via the via FL filtering layer and DT-GCN domain transformation layer. Fig. 6(a) is the feature distribution extracted by the backbone with domain-specific batch normalization (DS-BN), and Fig. 6(b) describes the feature distribution after the FL layer filters domain-style information and retains only identity-related factors. Fig. 6(c) shows the reconstructed features obtained by DT-GCN, which recouples identity-related nodes with their original domain nodes. Fig. 6(d) depicts the reconstructed features after spatial transformation, where DT-GCN recouples identity-related nodes with other domain nodes. The features from the Market-1501 space are transferred to the DukeMTMC-reID, MSMT17 and CUHK03 spaces.

A.2 Analysis of Domain Number and Comprehensive Evaluation Coefficient

Fig. 7 represents the effects of source-domain number and comprehensive evaluation coefficient γ. From Fig. 7(a-c), evidently, increasing the number of source domains facilitates the model to learn more diverse features and improves generalization performance. Meanwhile, the existence of multi-source-domain spaces helps exploit the complementarity among different clustering spaces, optimizes the generation of sample pseudo-labels, and exerts a positive effect on model training within the target domain. Fig. 7(d-f) illustrate the ablation study of comprehensive evaluation coefficient γ in equation 9, where α is the silhouette coefficient and β is the discrepancy coefficient of labels between iteration t and t-1. Notably, the silhouette coefficient focuses on the clustering performance of samples within the feature space. If one domain space outperforms others at the early training stage, the model tends to continuously adopt pseudo-labels generated from this space, falling into a vicious cycle and failing to exploit the complementarity of other domain spaces. In contrast, the discrepancy coefficient excessively emphasizes label differences between successive iterations while ignoring the importance of clustering quality for label accuracy. Instead, the comprehensive evaluation coefficient integrates both metrics. It takes into account the stability of feature clustering, and leverages the discrepancy coefficient to explore the diversity across different domain spaces. Through multi-domain cooperation, it addresses hard samples that cannot be resolved within a single domain, and consequently achieves the best performance.

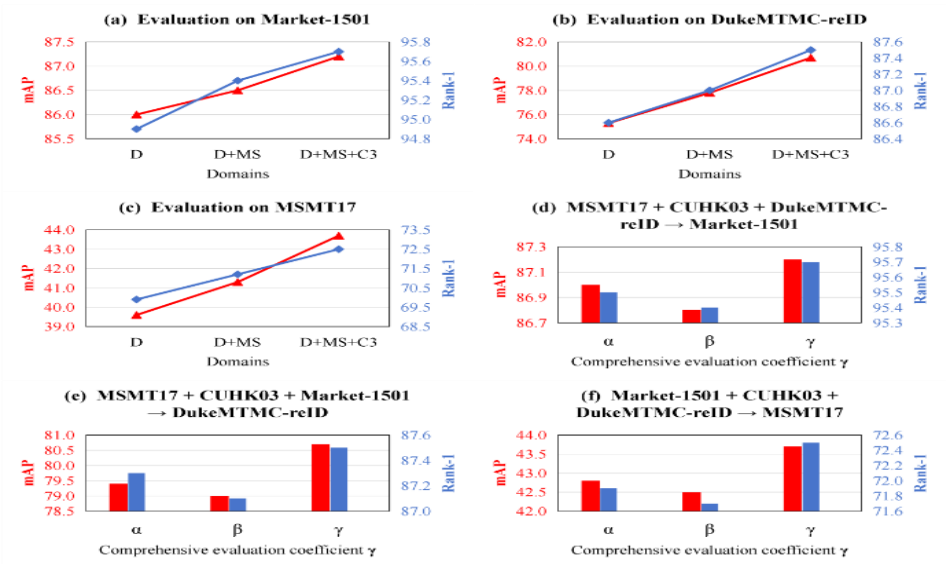


Figure 7: Effects of source-domain number and comprehensive evaluation coefficient γ, where M, D, MS, C3 represent the Market-1501, DukeMTMC-reID, MSMT17 and CUHK03.

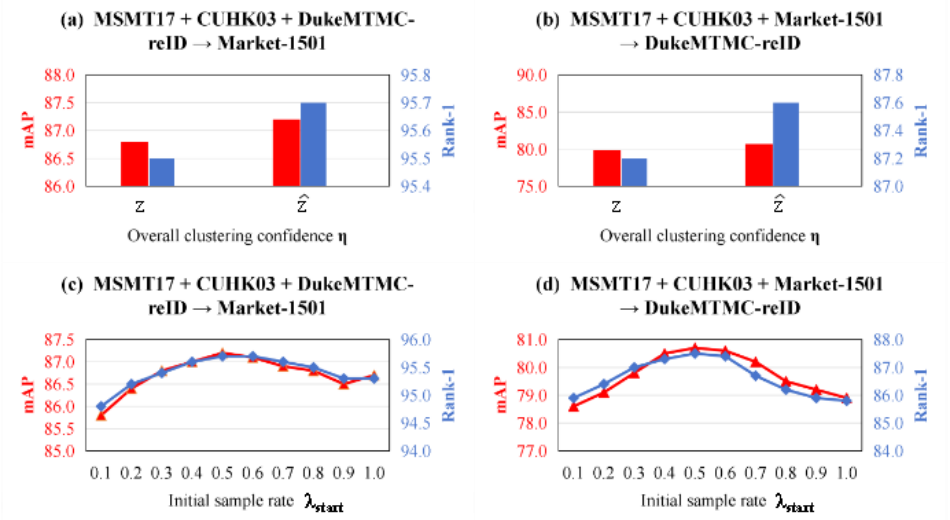


Figure 8: Effects of overall clustering confidence η and initial sampling rate λ_{start}.

A.3 Feature Visualization Analysis via FL filtering layer and DT-GCN

Fig. 8 represents effects of overall clustering confidence η and initial sampling rate λ_{start}. The clustering confidence η is the product of $b_\pi(x_i, x_\nu)$ and $\hat{z}^t(x_i, x_\nu)$ in equation 10, where $b_\pi(x_i, x_\nu)$ is the matrix that describes whether the labels of samples x_i and x_ν are consistent in the selected domain π; $z^t(x_i, x_\nu)$ is the average matrix of $b_k(x_i, x_\nu)$ across different domains; $\hat{z}^t(x_i, x_\nu)$ is the matrix by weighting $b_k(x_i, x_\nu)$ with the comprehensive coefficient γ_k^t. From the Fig. 8(a-b), $\hat{z}^t(x_i, x_\nu)$ is clearly superior to $z^t(x_i, x_\nu)$ due to its consideration of the effect of varying clustering quality on domain-wise confidence. Furthermore, Fig. 8(c-d) illustrate how the proportion λ_{start} of initial training samples selected by this confidence affects model training. As shown in the figure, the model training accuracy first increases and then decreases with the growth of this proportion. On the one hand, insufficient samples result in inadequate training. On the other hand, a higher proportion brings more erroneous samples and induces model performance degradation.