

RULER: Source-Free Domain Adaptive Person Re-identification via Uncertain Label Refinery

Aihua Zheng¹, Zhihao Fei¹, Yuhe Ding², Chenglong Li^{1*}
and Bin Luo²

¹Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui Provincial Key Laboratory of Security Artificial Intelligence, School of Artificial Intelligence, Anhui University, Hefei 230601, China.

²Anhui Provincial Key Laboratory of Multi-modal Cognitive Computation, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Abstract

Source-free domain adaptive person re-identification (ReID) aims to address the cross-domain person ReID task with a well-trained source model, which solves the limitations of data privacy and transmission costs in real-world scenarios. Existing methods mainly generate pseudo labels for target data, which are unreliable because of distribution shifts. First, the ubiquitous difficult samples may lead to the ambiguity of the model prediction. Second, the source model may have a bias towards certain classes. To alleviate these two problems, we propose a source-free domain adaptive person ReID method via the Uncertain LabEl Refinery (RULER). RULER consists of Uncertainty-aware Pseudo-Labeling Refinery (UPLR) and Frequency-weighted Contrastive Learning (FCL). To reduce the ambiguity of predictions caused by sample label uncertainty, UPLR generates pseudo labels by clustering samples after multiple random dropouts and then integrates the results to obtain high-confidence pseudo labels. Moreover, FCL defines the frequency of each class as the sample weight and introduces a frequency-weighted contrastive loss to alleviate the class biases of the model. RULER improves the quality of pseudo labels and mitigates the source model's bias towards

*Corresponding author

certain classes. We achieve competitive results compared to state-of-the-art methods on both real-to-real and synthetic-to-real source-free domain adaptation scenarios, validating the effectiveness of RULER.

Keywords: Person Re-identification, Source-Free Domain Adaptation, Cluster, Contrastive Learning, Transfer Learning

1 Introduction

Person re-identification (ReID) aims to retrieve person images of the same ID as the query person image across non-overlapping cameras [1]. In recent years, fully supervised person ReID methods [2–4] and cross-modal person ReID methods[5–7] have achieved great progress. However, these methods require considerable time and resources for labeling images of individuals, making them challenging to implement in real-world scenarios. Furthermore, model performance is significantly degraded when a model is trained on the labeled source domain, while tested directly on the unlabeled target domain. This decline is attributed to a domain shift between the source and target domains, which results in poor model transferability. Consequently, recent research [8–10] focuses on employing domain adaptation methods to mitigate the issue of domain shift. In particular, research on domain adaptive person ReID [11–16] seeks to address the performance degradation of deep learning models caused by mismatches in the data distribution between the source and target domains.

Traditional unsupervised domain adaptation (UDA) methods for person ReID rely on source data, which can lead to potential information leakage and compromised privacy in the source domain. Additionally, utilizing source domain data increases data transmission costs. To solve this problem, Liang *et al.* [17] first introduces the source-free domain adaptation (SFDA) paradigm. SFDA eliminates the necessity for access to source domain data and achieves target domain adaptation solely through a well-trained source model. In the source-free domain adaptive person ReID task, methods can be broadly divided into two categories: data generation techniques and pseudo-label-based methods. Among these methods, Qu *et al.* [18–20] employ a generation module to produce source-similar images, thereby compensating for the absence of source domain data during domain adaptation. However, it introduces additional parameters and requires multiple models to complete the adaptation process in the target domain, thereby increasing the actual deployment cost and complexity of the model. Currently, pseudo-label-based methods [21–24] dominate the field of source-free domain adaptive person ReID. These methods typically employ K-means [25] or DBSCAN [26] clustering algorithms to generate pseudo labels for unlabeled target domain data, and then related methods are designed to utilize the pseudo labels.

However, the pseudo labels obtained through these methods are often unreliable. We analyze the unreliability from two perspectives. First, these

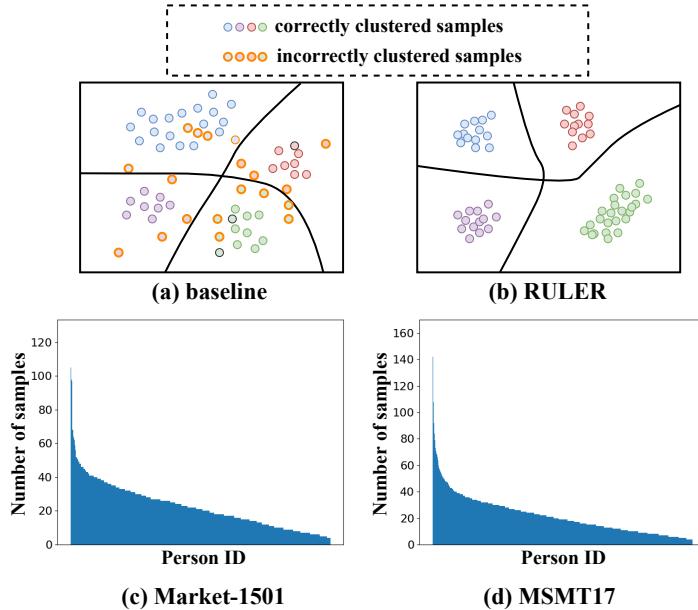


Fig. 1 Existing methods contain sample label uncertainty and class imbalance problems. Figure (a) illustrates numerous samples with clustering errors before adaptation. Figure (b) demonstrates that our proposed method achieves clearer classification boundaries and more compact clusters, thereby effectively improving clustering accuracy. Figures (c) and (d) show the sample number distributions for different Person IDs after clustering on the unlabeled target datasets Market-1501 and MSMT17, respectively. The horizontal axes of these bar graphs represent Person IDs, while the vertical axes indicate the number of samples corresponding to each ID.

methods neglect the uncertainty of sample labels. Specifically, as shown in Fig. 1 (a), difficult samples lead to unclear classification boundaries between different IDs. Many ambiguous samples near these boundaries increase the likelihood of incorrect clustering. This uncertainty leads to unreliable pseudo labels, negatively impacting the adaptation of the source model to the target domain. Second, the pseudo labels generated by existing methods exhibit class imbalance problems. For example, as shown in Fig. 1 (c) and (d), after clustering the two widely used person ReID task datasets, Market-1501 [27] and MSMT17 [28], and then analyzing the distribution of the number of samples for each person ID, we can observe that both significant class imbalance problem [29]. This model bias towards certain classes leads to a class imbalance in pseudo labels, which will affect the generalization of the model.

As a result, we propose a method termed source-free domain adaptive person ReID via Uncertain LabEl Refinery (RULER). RULER consists of two key components, Uncertainty-aware Pseudo-Labeling Refinery (UPLR) and Frequency-weighted Contrastive Learning (FCL). First, to address the problem of poor clustering effects due to sample label uncertainty, we design the UPLR scheme for clustering. UPLR employs a strategy of randomly dropping target

4 *RULER: Source-Free Domain Adaptive Person Re-identification via ...*

domain features multiple times to increase clustering effectiveness. Specifically, we cluster global features to obtain global pseudo labels, while also clustering the randomly dropped target domain features to obtain subset pseudo labels. Then, to resolve label inconsistencies across different clustering iterations, we design a label mapping scheme to match the global pseudo labels and subset pseudo labels. Finally, we utilize a voting mechanism to obtain more confident pseudo labels for each sample based on the results from multiple clustering iterations. As shown in Fig. 1 (b), our method achieves clearer classification boundaries and more compact clusters, effectively improving clustering performance. We further validate the effectiveness of UPLR through extensive experiments and t-SNE visualization.

Second, we address the problem of model bias towards a certain class and class imbalance problem. We design the FCL scheme to implicitly align the target domain with the unseen source domain. This approach uses a memory bank to store cluster center features for each class and optimizes model training through contrastive learning loss with new features extracted during training. Additionally, after clustering each subset result obtained in the UPLR stage, we calculate the frequency of each class and then convert it into the corresponding class weight. These weights are used to further optimize the model through frequency-weighted contrastive loss. FCL effectively mitigates the model's bias toward certain classes and enhances the pseudo labels refinement process. This enables the adapted target model to learn more discriminative features of the person in the unlabeled target domain, significantly improving the model's generalization ability.

In summary, RULER effectively addresses clustering errors caused by sample label uncertainty, improves clustering accuracy, and mitigates the impact of unbalanced data distributions to a certain extent. Our contributions are summarized below:

- We propose the Uncertainty-aware Pseudo-Labeling Refinery (UPLR) for clustering optimization. By randomly dropping features for label refinement multiple times during the pseudo labels generation process, we obtain more confident pseudo labels, effectively reducing the impact of sample label uncertainty and improving clustering quality.
- We propose a Frequency-weighted Contrastive Learning (FCL) to alleviate the impact of model bias towards certain classes and class imbalance problems. By computing class weights through label refinement and optimizing the model via frequency-weighted contrastive loss, FCL improves the adaptation of the source model to the target domain.
- Our method achieves significant performance improvements compared to state-of-the-art methods on both real-to-real and synthetic-to-real source-free domain adaptation scenarios.

2 Related work

2.1 Domain Adaptive Person ReID

Domain adaptive person ReID methods migrate source domain data information to the target domain to improve the model's ReID performance in the target domain, which has attracted widespread attention in recent years. These methods can be divided into two categories, GAN transferring and joint training. The first one uses a method based on GAN transferring, using a generative model to transform the source domain image so that the source domain matches the image style of the target domain. Typically taking a source image as input, a generative adversarial network (GAN) [30] is used to generate a target stylized person image for image translation [31]. For example, use GAN to transfer image styles across domains [28, 32] or decompose features into features related to ID or irrelevant [33]. Since GAN training is unstable, methods utilizing generative models usually involve many heuristics and require a large number of parameters. The second method adopts a method based on joint training, which uses joint training between the source domain and the target domain to narrow the distribution difference between the source domain and the target domain. Dai *et al.* [12] propose an intermediate domain module to dynamically generate the representation of the intermediate domain by using two domain factors to mix the hidden representations of the source and target domains to promote a more effective transfer of source knowledge to the target domain. Wu *et al.* [15] design a multi-center memory and only compare samples in the same domain through a domain-specific contrastive learning mechanism to fully explore the information in the domain. Ge *et al.* [34] propose a self-paced contrastive learning framework SpCL with hybrid memory, which dynamically generates source domain class-level, target domain cluster-level, and non-cluster instance-level supervision signals through hybrid memory for learning feature representations. Zhang *et al.* [35] propose a novel local correlation ensemble model that focuses on the diversity of intra-class information and the reliability of class centers. Finally, Lee *et al.* [16] propose a camera-driven curriculum learning framework CaCL, which utilizes the camera labels of images to gradually transfer knowledge from the source domain to the target domain. However, these existing domain adaptive person ReID methods all require access to source domain data. Due to factors such as privacy protection and storage overhead, source domain data is difficult to obtain in practical applications, which makes traditional domain adaptation methods unusable.

2.2 Source-Free Domain Adaptive Person ReID

To solve the domain adaptation task when the source domain is absent, Liang *et al.* [17, 36] find that the trained source model hides rich source domain knowledge, which can be migrated to the target domain when the source domain data is absent, and thus led to the model adaptive concept.

Now, for source-free domain adaptive person ReID, the main methods can be divided into two categories: data generation techniques and pseudo-label-based approaches. The first method uses data generation techniques, using the generation module to generate source-similar samples to make up for the lack of source domain data. Qu *et al.* [18–20] handle inter-domain person appearance style differences using GAN-based domain-style diversity augmentation and intradomain individual style misalignment through adversarial mutual teaching-learning. However, due to the unstable training of the GAN, this method cannot ensure the quality of the generated source-similar images and introduce additional parameters. The second method uses pseudo-label-based approaches, usually using a well-trained source model to initialize the target model, and then uses the DBSCAN [26] or K-means [25] clustering algorithm to generate pseudo labels for unlabeled target domain data for supervised training of the target model. Ge *et al.* [24] propose a Mutual Mean-Teaching(MMT) to conduct pseudo label refinery with the off-line refined hard pseudo labels and online refined soft pseudo labels in a collaborative training manner. Chen *et al.* [23] explore using an asymmetric structure inside the neural network to address self-ensembled teacher-student networks that quickly converge to a consensus that leads to a local minimum. Han *et al.* [22] model the probabilistic uncertainty by measuring the inconsistency between the predicted and ideal distributions of pseudo labels to determine and purify wrong labels of target domain samples for UDA person ReID. Chen *et al.* [21] design a model that integrates a data augmentation method with a multi-label assignment strategy to achieve semantic feature decoupling in the source domain. However, pseudo-label-based approaches ignore DBSCAN [26] as a major clustering method. The uncertainty of sample labels may lead to incorrect sample allocation to clusters and unclear boundaries between clusters, impacting clustering accuracy and resulting in unreliable pseudo labels. This uncertainty can adversely affect the adaptation of the source model to the target domain.

3 Method

Our method aims to utilize a model trained on the source domain to adapt to the target domain, particularly in cases where source domain data is unavailable. As shown in Fig. 2, we propose RULER to improve clustering accuracy and facilitate the adaptation of the source model to the target domain. This approach enables the network to generate more reliable pseudo labels for training, thereby enhancing knowledge transfer from the source domain. As a result, the model performance on the target domain is significantly improved.

3.1 Overview

For the ordinary unsupervised domain adaptive person ReID task, we obtain n_s labeled samples $\{(x_s^i, y_s^i)\}_{i=1}^{N_s}$ from the source domain D_s , where $x_s^i \in X_s, y_s^i \in Y_s$, and n_t unlabelled samples $\{(x_t^i)\}_{i=1}^{N_t}$ from the target domain D_t , where $x_t^i \in X_t$. The goal of source-free domain adaptive person ReID is to obtain

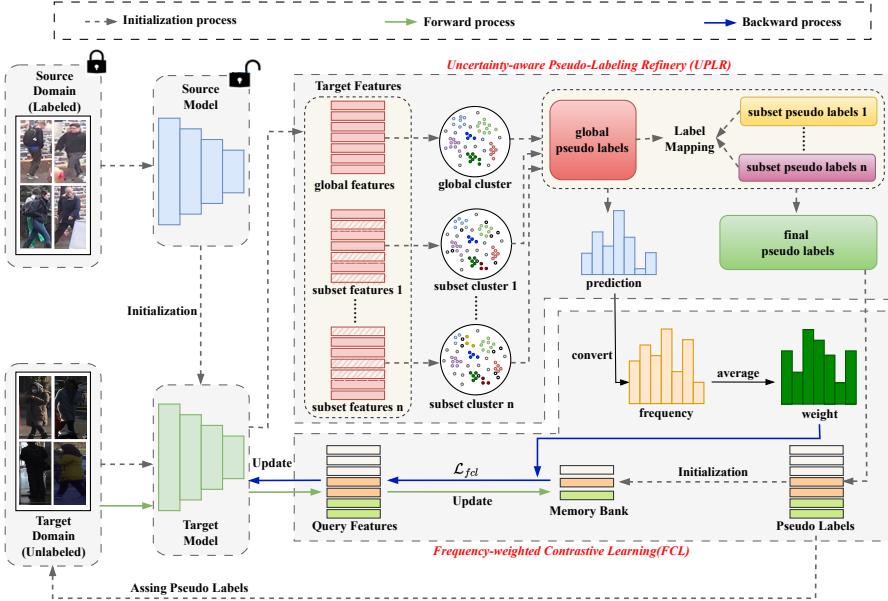


Fig. 2 Source-Free Domain Adaptive Person ReID via the Uncertain LabEl Refinery (RULER). RULER consists of two key components, the Uncertainty-aware Pseudo-Labeling Refinery (UPLR) and Frequency-weighted Contrastive Learning (FCL). UPLR mitigates the impact of sample label uncertainty and improves clustering quality by randomly dropping features for label refinement multiple times during the pseudo labels generation process. FCL addresses model bias towards certain classes and class imbalance by employing frequency-weighted contrastive loss optimization.

pseudo labels for the target domain by clustering $\{(\hat{y}_t^i)\}_{i=1}^{N_t}\}$. Subsequently, the network uses both the source domain and the target domain with pseudo labels to complete the target domain adaptation process. Here, the goal of source-free domain adaptive person ReID is to learn the target function $f_t : \mathcal{X}_t \rightarrow \mathcal{Y}_t$, and only $\{(x_t^i)\}_{i=1}^{N_t}\}$ and the source function $f_s : \mathcal{X}_s \rightarrow \mathcal{Y}_s$ are available.

Fig. 2 illustrates the RULER framework proposed in this paper. We solve the source-free domain adaptive person ReID task in three steps. First, we generate a simple source model from the source data using a ResNet network and a standard ReID loss. Second, we initialize the target model using the source model, transferring the source domain knowledge to the target domain without accessing the source data. We then extract features from the unlabeled target domain data using the target model for clustering and generating pseudo labels. To optimize this crucial clustering process, we design an innovative Uncertainty-aware Pseudo-Labeling Refinery (UPLR) scheme that ensures the extraction of highly reliable pseudo labels. Simultaneously, we initialize a memory bank to store cluster center features as the representation of each cluster. To address the class imbalance problem, we introduce Frequency-weighted Contrastive Learning (FCL) to enhance adaptive performance. Finally, training is supervised by pseudo labels, and the cluster center features stored in

the memory bank are updated in each batch. The last two steps are repeated in each epoch to enhance the source model's adaptation to the target domain and enable the model to learn discriminative features more effectively in the target domain.

3.2 Source Model Generation

We aim to develop a deep neural network and train the source model $f_s : \mathcal{X}_s \rightarrow \mathcal{Y}_s$ by minimizing the cross-entropy loss and triplet loss [37]:

$$\mathcal{L}_{id}^s = \frac{1}{N_s} \sum_{i=1}^{N_s} y_s^i \log \hat{y}_s^i, \quad (1)$$

$$\begin{aligned} \mathcal{L}_{tri}^s = \frac{1}{N_s} \sum_{i=1}^{N_s} & \max(0, f_s(x_s^i) - f_s(x_s^{i,p}) \\ & - f_s(x_s^i) + f_s(x_s^{i,n}) + m), \end{aligned} \quad (2)$$

where $\hat{y}_s = \delta_k(f_s(x_s))$, the overall loss function of the pre-trained source model can be described as follows:

$$\mathcal{L}_{ReID}^s = \mathcal{L}_{id}^s + \mathcal{L}_{tri}^s. \quad (3)$$

3.3 Uncertainty-aware Pseudo-Labeling Refinery (UPLR)

We aim to optimize the DBSCAN [26] clustering process, address clustering errors arising from sample label uncertainty, and obtain more confident pseudo labels. Inspired by Monte-Carlo Dropout [38], we design the Uncertainty-aware Pseudo-Labeling Refinery (UPLR) to achieve higher confidence in pseudo labels, thereby improving clustering accuracy. Unlike traditional dropout techniques, UPLR implements random dropout at the feature level rather than at the neuron level within the network model. This dropping process is random and non-repetitive in each iteration, resulting in the generation of different subset feature vectors each time. Through this perturbation of the feature space, UPLR enhances the model's diversity and robustness. Additionally, a voting mechanism is employed to derive pseudo labels from multiple subset feature vectors, thereby increasing the reliability and accuracy of the final pseudo labels. Specifically, the source model f_s is used to initialize the target model f_t . Subsequently, features of each sample are extracted from the target domain dataset D_t using the target model, resulting in the feature set $U_t = \{u_t^1, u_t^2, \dots, u_t^{N_t}\}$.

The DBSCAN [26] clustering algorithm is conducted in each epoch, and UPLR is performed every ϵ epoch based on the activation parameter ϵ . This optimization enhances the DBSCAN [26] clustering algorithm, ensuring the generation of more confident pseudo labels. The UPLR consists of four specific steps. First, samples are randomly dropped from the global feature collection U_t via the drop ratio ρ . This process is iterated for n drop iterations. Each

iteration drops distinct samples to generate n different subset feature vectors $U_t^{(r)}$. Specifically, we use the random function $rand(\cdot)$, which follows a uniform distribution, to randomly select a set of feature indices to be dropped, based on the predetermined drop ratio ρ applied to the indices of the global feature vectors. Subsequently, the subset feature vectors are generated by randomly selecting and dropping features from the global feature vectors according to the selected indices, as follows:

$$U_t^{(r)} = U_t \odot rand(\rho)^{(r)}, \quad r = 1, 2, \dots, n. \quad (4)$$

Second, we apply DBSCAN [26] clustering to each of the n subset feature vectors, resulting in n different subset clustering results, defined as:

$$\begin{aligned} \mathbf{C}_{all} = & \{(U_t, C_{global}), (U_t^{(1)}, C_{subset}^{(1)}), \\ & (U_t^{(2)}, C_{subset}^{(2)}), \dots, (U_t^{(n)}, C_{subset}^{(n)})\}. \end{aligned} \quad (5)$$

Third, we aim to obtain more reliable pseudo labels for each sample based on multiple different subset clustering results. As the pseudo label assignment rules change with each iteration of the DBSCAN [26] clustering algorithm, the assigned labels may vary. In other words, when clustering the same sample across different subset feature vectors, the expectation is that they should have the same ID. However, inconsistent labels are assigned due to limitations in the pseudo label assignment rules. Using this type of pseudo labels directly for subsequent operations is not conducive to obtaining the final pseudo labels with higher confidence. Therefore, we use the Hungarian algorithm [39] to match pseudo labels across different subset clusters. This process involves label mapping between the pseudo labels of the global cluster and those of the subset clusters, followed by re-labeling to address the issue of inconsistent label assignments. Specifically, we employ the label mapping function km to derive the new pseudo labels after matching each clustering result from the subset feature vectors:

$$\bar{Y}_{subset}^{(r)} = km(Y_{global}, Y_{subset}^{(r)}), \quad r = 1, 2, \dots, n, \quad (6)$$

where Y_{global} represents the pseudo labels obtained by clustering the global features, and $Y_{subset}^{(r)}$ denotes the pseudo labels obtained by clustering each subset feature vectors.

Finally, pseudo labels with higher confidence for each sample are derived from the results of multiple clustering via a voting mechanism. Specifically, based on Eq. (6), we first obtain the pseudo labels $Y_{subset}^{(r)}$ of each subset feature vector through label mapping, which subsequently results in the subset pseudo labels \mathbf{L} . Next, we iterate through the set of \mathbf{L} to calculate the pseudo labels result matrix \mathbf{F} , where $\delta(\mathbf{L}, Y_{global})$ adds the score of the pseudo label in \mathbf{L} that is predicted to be the same as Y_{global} , defined as follows:

$$\mathbf{L} = \{\bar{Y}_{subset}^{(1)}, \bar{Y}_{subset}^{(2)}, \dots, \bar{Y}_{subset}^{(n)}\}, \quad (7)$$

$$\mathbf{F}_{ij} = \delta(l_i^{(m)}, j), \quad (8)$$

where, $0 \leq m < n$, $0 \leq i < N_t$, $0 \leq j < n_t$, where n represents the number of iterations for clustering different subset features, N_t represents the number of samples in the target domain, and n_t denotes the number of classes obtains through clustering. Next, we compute the softmax probability for each sample across all classes and designate the pseudo labels with the highest probability as the final pseudo labels for the sample. At this point, we achieve pseudo labels \hat{y}_t with higher confidence.

$$\hat{y}_t^i = \operatorname{argmax} \left(\frac{\exp(\mathbf{F}_{ij})}{\sum_{j=1}^{n_t} \exp(\mathbf{F}_{ij})} \right). \quad (9)$$

Finally, we obtain a target domain $D'_t = \left\{ (x_t^i, \hat{y}_t^i) \mid_{i=1}^{N_t'} \right\}$ containing K classes and pseudo labels with higher confidence, to complete the adaptation of the source model to the target domain, where N_t' represents the number of samples in the target domain without noise labels.

3.4 Frequency-weighted Contrastive Learning (FCL)

After obtaining pseudo labels for each sample, we use the target domain data with the assigned pseudo labels to complete the adaptation of the source model on the target domain. Here, we use the same settings as Cluster-Contrast [40]. Specifically, it encompasses the process of memory initialization, neural network training, and memory updating. To mitigate potential class imbalance problems, we design a frequency-weighted contrastive learning scheme for training optimization.

Memory initialization. Unlike the instance-level memory bank [34, 41, 42], the RULER does not store the features of each sample. Instead, it stores the representation of each cluster c_1, \dots, c_K in the memory-based feature bank. We initialize the cluster representation via the mean feature vector of each cluster. Initially, feature representations for each sample are extracted via Eq. (10), and subsequently, we employ Eq. (11) to compute the mean feature vector of samples within the same cluster, which serves as the cluster representation:

$$u_i = f_t(x_t^i), \quad (10)$$

$$c_k = \frac{1}{|\mathcal{H}_k|} \sum_{u_i \in \mathcal{H}_k} u_i, \quad (11)$$

where \mathcal{H}_k denotes the k -th cluster set and $|\cdot|$ indicates the number of instances per cluster. \mathcal{H}_k contains all the feature vectors in the cluster k .

Neural network training. Contrastive learning [43–45] has a wide range of applications across various fields. It is used to train the feature extractor

of deep neural networks, obtaining representations with strong discriminative abilities. In our method, we use contrastive learning loss to optimize the adaptation process of the source model to the target domain. Specifically, outside of the UPLR stage, we utilize the ClusterNCE [40, 46] loss as defined by Eq. (12):

$$\mathcal{L}_{cl}^t = -\frac{1}{N_t'} \sum_{i=1}^{N_t'} \log \frac{\exp(q \cdot c_+/\tau)}{\sum_{k=0}^K \exp(q \cdot c_k/\tau)}, \quad (12)$$

where q is the image input by the current mini-batch, τ is a temperature hyper-parameter [47], c_+ is the cluster representation of the positive sample corresponding to the current image, and c_k is the unique representation of each cluster.

Algorithm 1 Algorithm of the proposed RULER.

Input: Target dataset \mathcal{X}_t ; a well-trained source model f_s ;

Parameter: ϵ , n , ρ , m and τ ;

Output: New target model f_t ;

Initialize the target model f_t with f_s ;

for i in epochs **do**

Extract global feature vectors U_t from \mathcal{X}_t by f_t ;

Cluster U_t into K clusters with UPLR;

for r in n **do**

Randomly drop features in U_t by Eq. (4);

Cluster subset feature vectors $U_t^{(r)}$ by Eq. (5);

Label map between Y_{global} and $Y_{subset}^{(r)}$ by Eq. (6);

Final pseudo labels \hat{y}_t^i by voting mechanism by Eq. (9);

end

Initialize memory bank by Eq. (11);

Calculate the class weights w based on \mathbf{F} by Eq. (13);

Compute FCL loss by Eq. (14);

Update cluster feature by Eq. (15);

Update the encoder f_t by optimizer;

end

In the UPLR stage, we extract weight information for each class through multiple clustering iterations. To mitigate degradation in model adaptation resulting from class imbalance, we assign varying weights to each class. Specifically, according to the \mathbf{F} obtained by the UPLR scheme, we define the prediction result as $p = \mathbf{F}_{ij}$ for each sample. Then, convert it to the corresponding frequency $f = \frac{\exp(p)}{\sum \exp(p)}$. Finally, the weight of each class is calculated through Eq. (13), and the frequency-weighted contrastive loss is designed based on the calculated class weights for further optimization of the model, as shown

in Eq. (14):

$$w = \text{AVG} \left(\frac{\exp(\mathbf{F}_{ij})}{\sum_{j=1}^{n_t} \exp(\mathbf{F}_{ij})} \right), \quad (13)$$

$$\mathcal{L}_{fcl}^t = -\frac{1}{N_t'} \sum_{i=1}^{N_t'} (1-w) \log \frac{\exp(q \cdot c_+ / \tau)}{\sum_{k=0}^K \exp(q \cdot c_k / \tau)}. \quad (14)$$

Memory updating. To more effectively leverage the latest feature information and facilitate the adaptation process of the source model to the target domain, we update the mean features used for cluster representation in the memory bank in real time within each mini-batch:

$$c_k \leftarrow m c_k + (1-m)q, \quad (15)$$

where q is the image input by the current mini-batch and m is the momentum updating factor. The overall algorithm is shown in Algorithm 1.

4 Experiments

4.1 Setup

4.1.1 Datasets

We utilize four datasets to evaluate the effectiveness of RULER in source-free domain adaptive person ReID tasks.

Market-1501 [27] serves as a widely recognized benchmark for the field of person ReID. It contains images of 1,501 IDs captured by 6 cameras, which contain 12,936 images of 751 IDs for training and 19,732 images of 750 IDs for testing.

DukeMTMC-reID [48] is a subset of the DukeMTMC [49] dataset, which was taken from the campus of Duke University. The original dataset contains 34,183 images collected from 8 cameras. This dataset includes images of 1404 IDs captured by 8 cameras, including 16,522 images of 702 IDs for training and 17,661 images of 702 IDs for testing.

MSMT17 [28] is a more challenging person ReID dataset that collects a large number of identities, bounding boxes, and cameras under the multi-scene multi-time. It contains 126,441 images obtained from 15 cameras, including 32,621 images with 1,041 IDs and 93,820 images with 6,120 IDs, which are used for training and testing, respectively.

Unreal [50] is a dataset consisting of virtual synthesized images, which provides synthesized images and corresponding ID tags, and 130,244 images for training.

4.1.2 Evaluation metric

We apply our method in six experimental settings, encompassing source-free domain adaptive person ReID tasks that include mid-to-mid adaptation tasks

and mid-to-large adaptation tasks. These experiments include real-to-real and synthetic-to-real scenarios to comprehensively evaluate the effectiveness of the proposed method. We evaluate the performance of the proposed method via the widely used mean average precision (mAP) and Rank-1/5/10 values of cumulative matching features.

4.2 Implementation Details

We adopt ResNet-50 [51] without the last classification layer as the backbone network to conduct all the experiments. All the codes are implemented in Pytorch. In the source model training stage. We adopt ResNet-50 as the backbone network. The network is initialized with ImageNet [52] pre-trained weights. Given the mini-batch of images, the network parameters are updated independently by optimizing Eq. (3). We adopt the Adam [53] optimizer to train the source model with a weight decay of 5e-4. The initial learning rate is set to 3.5e-4 and is decreased to 1/10 of its previous value on the 40th and 70th epoch in the total 80 epochs. In the source model adaptation stage. We use Cluster-Contrast [40] as the backbone of the baseline for our method and initialize the network parameters of this baseline using the source model to serve as the target model. All the images are resized to 256×128 , and random horizontal flipping, padding with 10 pixels, random cropping, and random erasing [54] are used for data augmentation. Each mini-batch contains 256 images of 16 pseudo-person identities (16 instances for each person). We adopt the Adam optimizer to train the ReID model with a weight decay of 5e-4. The initial learning rate is set to 3.5e-4 and is reduced to 1/10 of its previous value every 20 epochs in 60 epochs. The momentum factor m is set to 0.1 in the memory bank update, and the loss temperature τ is set to 0.05 for all datasets. For DBSCAN [26], we use the same parameter settings in all the experiments, with a maximum distance d between two samples of 0.6 and a minimal number of neighbors in a core point of 4. For the three hyper-parameters ϵ , n , and ρ of UPLR, in the mid-to-mid adaptation tasks, we set the parameters as 3, 10, and 0.2, respectively. For the mid-to-large adaptation tasks, the corresponding settings are 3, 10, and 0.5, respectively.

4.3 Comparison in real-to-real scenarios

We evaluate our method with state-of-the-art UDA methods across four domain adaptive person ReID tasks, including mid-to-mid adaptation tasks and more challenging mid-to-large adaptation tasks. The experimental results are shown in Table 1. In the third column of each table, the types are categorized as Data and Mod.. The Data category represents source non-free domain adaptive person ReID methods that require source domain data, while the Mod. category refers to source-free domain adaptive person ReID methods that utilize only the source pre-trained model.

Mid-to-Mid adaptation tasks. Our method achieves the best or second-best performance across two different domain adaptive person ReID tasks.

Table 1 Quantitative comparisons with state-of-the-art UDA person ReID methods in real-to-real scenarios. **RED** and **GREEN** indicate the best and second-best performances, respectively. ‘Data’ and ‘Mod.’ represent the source non-free and the source-free domain adaptive person ReID methods, respectively.

Methods	References	Type	Mid-to-Mid adaptation tasks									
			Market-1501 → DukeMTMC-reID				DukeMTMC-reID → Market-1501					
			mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10		
MMCL [41]	CVPR 2020	Data	51.4	72.4	82.9	85.0	60.4	84.4	92.8	95.0		
SpCL [34]	NeurIPS 2020	Data	68.8	82.9	90.1	92.5	76.7	90.3	96.2	97.7		
GCL [11]	CVPR 2021	Data	67.6	81.9	88.9	90.6	75.4	90.5	96.2	97.1		
HCD [55]	ICCV 2021	Data	70.1	82.2	-	-	80.0	91.5	-	-		
IDM [12]	ICCV 2021	Data	70.5	83.6	91.5	93.7	82.8	93.2	97.5	98.1		
IcJO [13]	KBS 2022	Data	65.7	81.5	89.8	92.6	78.1	91.0	96.3	97.6		
DREAMT [14]	TMM 2022	Data	69.8	82.3	90.9	93.6	81.4	93.3	98.0	98.7		
MCRN [15]	AAAI 2022	Data	71.5	84.5	91.7	93.8	83.8	93.8	97.5	98.5		
MMT [24]	ICLR 2020	Mod.	65.1	78.0	88.8	92.5	71.2	87.7	94.9	96.9		
ABMT [23]	WACV 2021	Mod.	69.1	82.0	-	-	78.3	92.5	-	-		
P ² LR [22]	AAAI 2022	Mod.	70.8	82.6	90.8	93.7	81.0	92.6	97.4	98.3		
MDJL [21]	PR 2023	Mod.	62.8	78.6	86.6	88.7	59.8	80.3	87.4	89.9		
S2ADAP [20]	KBS 2023	Mod.	71.8	83.7	91.3	94.1	82.0	93.5	98.0	98.8		
IAMT [19]	TOMM 2024	Mod.	71.3	83.4	-	-	82.8	93.6	-	-		
AAMT [18]	TMM 2024	Mod.	74.7	85.7	92.8	94.7	83.9	94.2	98.4	98.9		
source-only	-	Mod.	30.5	49.4	64.0	70.5	25.2	55.2	71.5	77.9		
baseline	-	Mod.	73.7	85.0	91.7	93.6	83.2	92.5	96.9	97.7		
Ours	-	Mod.	74.1	85.4	92.5	94.8	84.2	93.5	97.0	98.0		
Methods	References	Type	Mid-to-Large adaptation tasks									
			Market-1501 → MSMT17				DukeMTMC-reID → MSMT17					
			mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10		
MMCL [41]	CVPR 2020	Data	15.1	40.8	51.8	56.7	16.2	43.6	54.3	58.9		
SpCL [34]	NeurIPS 2020	Data	26.8	53.7	65.0	69.8	26.5	53.1	65.8	70.5		
GCL [11]	CVPR 2021	Data	27.0	51.1	63.9	69.9	29.7	54.4	68.2	74.2		
HCD [55]	ICCV 2021	Data	28.4	54.9	-	-	29.3	56.1	-	-		
IDM [12]	ICCV 2021	Data	33.5	61.3	73.9	78.4	35.4	63.6	75.5	80.2		
IcJO [13]	KBS 2022	Data	26.0	51.3	63.4	68.9	27.3	54.8	67.3	72.4		
DREAMT [14]	TMM 2022	Data	25.3	51.6	64.3	69.7	30.3	58.0	70.5	75.3		
MCRN [15]	AAAI 2022	Data	32.8	64.4	75.1	79.2	35.7	67.5	77.9	81.6		
CaCL [16]	ICCV 2023	Data	36.5	66.6	75.3	80.1	-	-	-	-		
MMT [24]	ICLR 2020	Mod.	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8		
ABMT [23]	WACV 2021	Mod.	23.2	49.2	-	-	26.5	54.3	-	-		
P ² LR [22]	AAAI 2022	Mod.	29.0	58.8	71.2	76.0	29.9	60.9	73.1	77.9		
MDJL [21]	PR 2023	Mod.	13.4	34.4	44.5	50.6	17.1	40.3	51.2	56.3		
S2ADAP [20]	KBS 2023	Mod.	27.0	53.8	64.5	70.2	29.7	58.5	70.6	75.8		
IAMT [19]	TOMM 2024	Mod.	29.5	59.3	-	-	30.2	61.7	-	-		
AAMT [18]	TMM 2024	Mod.	33.2	61.5	72.0	77.2	35.2	64.8	75.7	78.6		
source-only	-	Mod.	6.5	18.8	29.1	34.5	8.1	25.0	37.0	42.3		
baseline	-	Mod.	34.5	61.9	72.7	76.7	36.2	64.0	74.9	79.0		
Ours	-	Mod.	40.0	67.7	78.4	82.2	42.9	70.3	80.4	83.7		

Specifically, it outperforms source-free domain adaptive person ReID methods such as S2ADAP [20] and IAMT [19], as well as the latest source non-free domain adaptive person ReID methods such as CaCL [16] and MCRN[15], which rely on additional source domain data. Although our method only achieves the second-best performance compared to the latest AAMT [18] in mAP and Rank-1/5. However, the AAMT [18] uses the source model to generate source-style adversarial examples, which undoubtedly increases the risk of source domain information leakage. This operation also increases the cost of data transmission and does not meet the requirements of the SFDA task for protecting source domain information and reducing data transmission costs. In particular, the mAP of our method is still the best comparing the DukeMTMC-reID → Market-1501 tasks. While the Rank-1/5/10 results are slightly lower

than those of AAMT [18], IAMT [19], and S2ADAP [20], our method is more straightforward. It does not introduce additional parameters, and the mAP of our method is 0.3% higher than AAMT [18]. Since our method does not use source domain data as supervisory information in the domain adaptation stage, our Rank-5 and Rank-10 results may be slightly inferior to those of the source non-free adaptive person ReID methods.

Mid-to-Large adaptation tasks. As shown in Table 1, our method outperforms both state-of-the-art source-free and source non-free domain adaptive person ReID methods. Specifically, it surpasses the performance of AAMT [18] by 6.8% and 7.7% in the mAP in two different mid-to-large adaptation tasks, respectively. Additionally, the Rank-1/5/10 also achieves significant improvement compared to AAMT [18]. In these mid-to-large adaptation tasks, the performance of AAMT [18] is worse than our method. This can be attributed to the MSMT17 [28] dataset, which presents greater challenges due to its collection of numerous identities, bounding boxes, and cameras across diverse scenes and times. Consequently, the uncertainty associated with sample labels in MSMT17 [28] is higher than that in Market-1501 [27] and DukeMTMC-reID [48], leading to reduced accuracy in the obtained pseudo labels. Moreover, the class imbalance problem is more obvious. Our method effectively addresses the uncertainty in sample labels, resulting in higher-quality pseudo labels and mitigating the excessive prediction bias towards certain classes caused by this imbalance.

4.4 Comparison in synthetic-to-real scenarios

Additionally, we evaluate both the large-to-mid adaptation task and the large-to-mid adaptation task in synthetic-to-real scenarios. The experimental results are shown in Table 2. We compare this method with two previous domain adaptive person ReID methods, which require source domain data on large-to-mid and large-to-large tasks, respectively.

Table 2 Quantitative comparisons with the state-of-the-art UDA person ReID methods in synthetic-to-real scenarios. RED and GREEN indicate the best and second-best performances, respectively. ‘Data’ and ‘Mod.’ represent the source non-free and the source-free domain adaptive person ReID methods, respectively.

Methods	References	Type	Large-to-Mid adaptation tasks				Large-to-Large adaptation tasks			
			Unreal → Market-1501				Unreal → MSMT17			
			mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
IDM [12]	ICCV 2021	Data	83.2	92.8	97.3	98.5	38.3	67.3	78.4	82.6
CaCL [16]	ICCV 2023	Data	84.0	93.0	97.6	98.5	40.3	70.0	80.5	84.0
source-only	-	Mod.	33.6	61.2	77.1	82.4	10.2	29.0	42.3	48.1
baseline	-	Mod.	84.8	93.1	97.0	97.7	40.1	67.2	78.1	81.5
Ours	-	Mod.	85.5	93.5	97.5	98.2	45.1	72.2	82.0	85.0

Large-to-Mid adaptation tasks. Our method outperforms CaCL [16] by 1.5% and 0.5% on mAP and Rank-1, respectively. Although it is not as good as CaCL [16] in Rank-5 and Rank-10, CaCL [16] is a source non-free domain adaptive person ReID method that cannot protect the privacy of source

domain information. Furthermore, utilizing source domain data significantly increases the data transmission costs for large-to-mid adaptation tasks.

Large-to-Large adaptation tasks. Our method outperforms CaCL [16] and IDM [12] in all the metrics and achieves significant performance improvements over CaCL [16]. By effectively addressing the uncertainty between sample labels, our method achieves outstanding results on large-level datasets. Additionally, we observe that the source model trained with extensive source domain data yields better domain adaptation performance in the same target domain.

These results indicate that our method can effectively transfer the knowledge learned from the source domain to the target domain. It performs well in both common real-to-real scenarios and synthetic-to-real scenarios.

4.5 Ablation study

Table 3 Ablation study of the proposed components. UPLR and FCL address the problems of model prediction ambiguity due to ubiquitous difficult samples and source model bias for certain classes, respectively.

Methods	Mid-to-Mid adaptation tasks				Mid-to-Large adaptation tasks			
	DukeMTMC-reID → Market-1501				DukeMTMC-reID → MSMT17			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
baseline	83.2	92.5	96.9	97.7	36.2	64.0	74.9	79.0
+ UPLR	83.7	93.0	97.1	98.2	40.7	67.8	78.7	82.3
+ UPLR + FCL	84.2	93.5	97.0	98.0	42.9	70.3	80.4	83.7

Effectiveness of each component. To verify the contributions of the proposed components in our method, we implement the ablation study of RULER on the mid-to-mid adaptation task and mid-to-large adaptation task in Table 3. In the table, source-only refers to testing the source model directly on the target domain. The results of this direct adaptation are suboptimal due to the domain shifts, particularly in mid-to-large adaptation tasks. Compared with the baseline, the proposed UPLR scheme significantly enhances the adaptation performance of the source model in the target domain. The performance improvement is due to UPLR rendering the pseudo labels obtained through clustering more confident than those obtained by the baseline. Especially in mid-to-large adaptation tasks, our method clearly improves performance. This is because large-level datasets such as MSMT17 [28] has a higher sample label uncertainty than mid-level datasets such as Market-1501 [27] and DukeMTMC-reID [48]. Previous methods do not consider the problem from this perspective, while our UPLR innovatively solves the problem of low clustering accuracy caused by sample label uncertainty in mid-level and large-level datasets. Furthermore, the proposed FCL scheme enhances the adaptability of the source model in the target domain. It effectively mitigates the prediction bias toward certain classes resulting from class imbalance, thereby further improving the source model’s adaptability in the target domain. Overall, UPLR and FCL

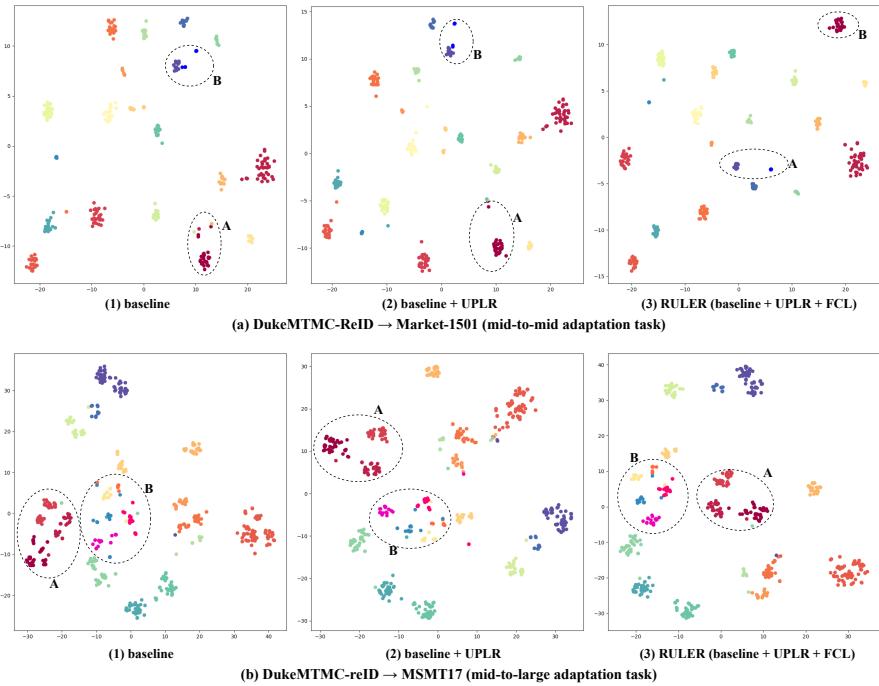


Fig. 3 The t-SNE visualizations for (a) DukeMTMC-reID → Market-1501 (mid-to-mid adaptation task) and (b) DukeMTMC-reID → MSMT17 (mid-to-large adaptation task), with the target features of (1) baseline, (2) baseline + UPLR, and (3) RULER. The different colors represent different IDs. For clarity, we randomly select 20 IDs from the target domain dataset for t-SNE visualization.

address the problems of model prediction ambiguity due to ubiquitous difficult samples and source model bias for certain classes, respectively.

t-SNE visualization. To evaluate the effectiveness of RULER, we present the t-SNE [56] visualization results of target features on the mid-to-mid adaptation task and mid-to-large adaptation task, as shown in Fig. 3. For clarity, we randomly select 20 IDs from the target domain dataset for t-SNE visualization. In the mid-to-mid adaptation task, (1) in Fig. 3 (a) shows the results of the baseline method, which shows an overall acceptable clustering effect. However, the clustering of some features remains suboptimal. For example, the feature clusters in areas A and B have unclear classification boundaries, causing some samples to be mistakenly clustered into other clusters. After adding the proposed UPLR scheme, as shown in (2) in Fig. 3 (a), this problem is effectively alleviated, the classification boundaries of feature clusters in areas A and B are clearer, and the clustering errors are reduced. Additionally, the proposed FCL scheme further enhances the formation of distinct classification boundaries, as shown in (3) in Fig. 3 (a). The clustering effect of the baseline is very poor on the more challenging mid-to-large task of DukeMTMC-reID → MSMT17. As shown in (1) in Fig. 3 (b), the classification boundaries between

most feature clusters are almost invisible, resulting in many incorrectly clustered samples. In contrast, the proposed method significantly enhances the clustering accuracy, as shown in (2) and (3) in Fig. 3 (b). Taking the feature clusters in areas A and B as examples, our method effectively separates different clusters. RULER benefits from both the UPLR and FCL schemes, leading to more compact feature clusters and clearer classification boundaries. Furthermore, it improves the quality of pseudo labels and reduces bias in the source model toward certain classes.

Table 4 Experimental results of the DBSCAN and K-means clustering algorithms on two different adaptation tasks of mid-to-mid adaptation tasks and mid-to-large adaptation tasks.

Methods	Mid-to-Mid adaptation tasks				Mid-to-Large adaptation tasks			
	DukeMTMC-reID → Market-1501				DukeMTMC-reID → MSMT17			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
RULER(DBSCAN)	84.2	93.5	97.0	98.0	42.9	70.3	80.4	83.7
RULER(K-means)	83.7 ↓	92.6 ↓	96.8 ↓	97.8 ↓	39.8 ↓	67.7 ↓	78.5 ↓	82.3 ↓

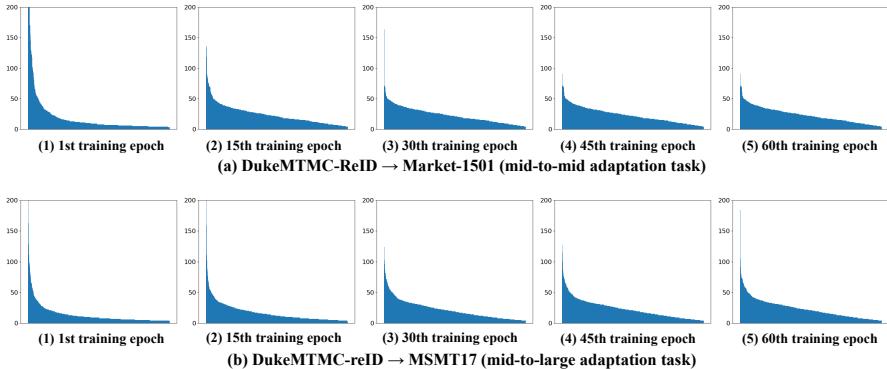


Fig. 4 Class distributions for (a) DukeMTMC-reID → Market-1501 (mid-to-mid adaptation task) and (b) DukeMTMC-reID → MSMT17 (mid-to-large adaptation task), with the class distributions of pseudo labels obtained by RULER during the 1st, 15th, 30th, 45th, and 60th training epochs, respectively. The horizontal axis represents the pseudo labels, while the vertical axis represents the number of samples corresponding to each class. For clarity, specific pseudo labels are not shown in the figures.

Analysis of different clustering algorithms. In our method, we compare the commonly used DBSCAN [26] and K-means [25] clustering algorithms in domain-adaptive person ReID. The results presented in Table 4 indicate that the performance of K-means is significantly lower than that of DBSCAN. Furthermore, the K-means algorithm requires the number of classes to be pre-specified (we use the same setup as the MMT [24], setting the number of pseudo classes to 500 and 1,000 for the mid-to-mid and mid-to-large adaptation tasks, respectively), which is inconsistent with the requirements of domain adaptation in real-world scenarios, as the number of classes in the target domain is

typically unknown. In contrast, DBSCAN is a density-based clustering algorithm that does not necessitate pre-specifying the number of classes. Therefore, we choose to utilize DBSCAN for clustering to obtain pseudo labels.

Analysis of class distribution. As shown in Fig. 4, we analyze the class distribution of the pseudo labels produced by RULER during the 1st, 15th, 30th, 45th, and 60th training epochs for the mid-to-mid adaptation and mid-to-large adaptation tasks, respectively. It is evident that a significant class imbalance problem arises in the initial training phase of the target model, adversely affecting the adaptation of the source model to the target domain. However, FCL effectively mitigates the impact of class imbalance on model predictions by dynamically adjusting class weights during training and constraining model training with Eq. (14), thereby enhancing the model's domain adaptation ability.

4.6 Analysis of clustering quality

Table 5 Comparison of cluster accuracy between the baseline and RULER in real-to-real scenarios and synthetic-to-real scenarios.

adaptation task	Accuracy (%)	
	baseline	RULER
Market-1501 → DukeMTMC-reID	76.5	77.7
DukeMTMC-reID → Market-1501	78.4	80.3
Market-1501 → MSMT17	60.0	64.2
DukeMTMC-reID → MSMT17	61.2	66.3
Unreal → Market-1501	79.7	81.7
Unreal → MSMT17	63.6	67.6

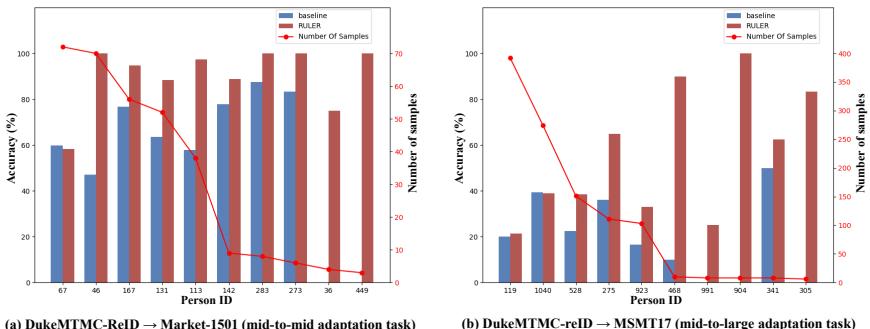


Fig. 5 Clustering accuracy between the baseline and our method (RULER) for person IDs in (a) DukeMTMC-reID → Market-1501 (mid-to-mid adaptation task) and (b) DukeMTMC-reID → MSMT17 (mid-to-large adaptation task).

As shown in Table 5, we compare the clustering accuracy results of the baseline and RULER in different domain adaptation tasks. Our method achieves significant improvements over the baseline, particularly in the mid-to-large

task. This is because the direct clustering in the baseline does not solve the problem of the poor clustering effect caused by sample label uncertainty, which is more prominent in the large-level dataset MSMT17 [28]. To further analyze the effect of clustering, as shown in Fig. 5, we randomly select 10 person IDs and analyze the clustering accuracy of each ID on the baseline and RULER. Notably, the first 5 person IDs belong to the head class, while the last 5 person IDs belong to the tail class. The results indicate that RULER achieves significant improvement in clustering accuracy on different adaptation tasks compared with the baseline. For certain person IDs, such as 36 and 449 in the DukeMTMC-reID \rightarrow Market-1501 task, and 991, 904, and 305 in the DukeMTMC-reID \rightarrow MSMT17 task, RULER consistently delivers better clustering outcomes, outperforming the baseline. While the baseline shows relatively better clustering performance for the head class, its effectiveness diminishes significantly for the tail class. RULER benefits from the UPLR scheme, which better extracts sample information and mitigates the adverse effects of sample label uncertainty on clustering. Additionally, the FCL scheme not only addresses the problem of model prediction bias towards certain classes but also enhances the clustering accuracy of our method, particularly for the tail class.

4.7 Analysis on Hyper-parameters

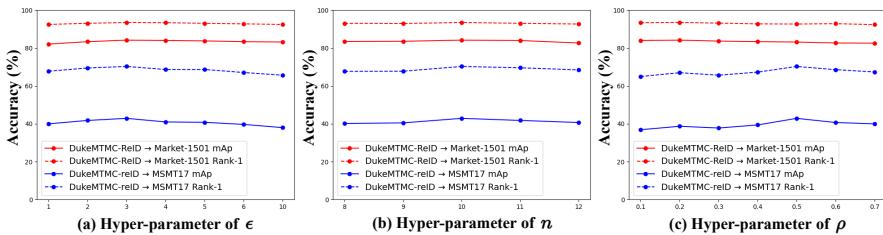


Fig. 6 Effects of the three hyper-parameter ϵ , n , and ρ ablation experiments in the DukeMTMC-reID \rightarrow Market-1501 (mid-to-mid adaptation task) and DukeMTMC-reID \rightarrow MSMT17 (mid-to-large adaptation task).

We verify the effects of the three hyper-parameters ϵ , n , and ρ in the proposed method for both the mid-to-mid adaptation task and the mid-to-large adaptation task. As shown in Fig. 6 (a), (b), and (c), respectively. Through extensive ablation experiments on the mid-to-mid and mid-to-large adaptation tasks, the optimal hyper-parameters ϵ , n , and ρ are 3, 10, 0.2, and 3, 10, 0.5, respectively. Our method effectively mitigates the impact of sample label uncertainty on clustering results. The reliability of the pseudo labels is gradually enhanced to obtain pseudo labels with higher confidence and improve the adaptability of the source model in the target domain. Generally, the hyper-parameters are not sensitive in our method.

5 Conclusion

In this paper, we propose source-free domain adaptive person ReID via the Uncertain LabEl Refinery (RULER) to address the challenge of poor clustering effectiveness in unlabeled target domains within source-free domain adaptive person ReID tasks. RULER incorporates a novel Uncertainty-aware Pseudo-Labeling Refinery (UPLR) scheme. This approach intervenes in the pseudo labels generation process by randomly dropping target domain features to mitigate the low confidence in pseudo labels arising from sample label uncertainty. It aims to increase confidence and improve the clustering accuracy of the obtained pseudo labels. Furthermore, we design a Frequency-weighted Contrastive Learning (FCL) scheme to address model bias toward certain classes, thereby implicitly aligning the target domain with the unseen source domain. The proposed method effectively adapts the source model to the target domain, and experimental results from six source-free domain adaptive person ReID tasks validate its effectiveness. However, fully leveraging camera information enhances clustering accuracy and improves the model's cross-domain generalization ability. Additionally, the use of black-box source models (i.e., only network predictions are available) helps protect the privacy of source domain data. These strategies represent key directions for future improvements.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grants 62372003 and 62376004, the Natural Science Foundation of Anhui Province under Grants 2308085Y40 and 2208085J18, and the University Synergy Innovation Program of Anhui Province under Grant GXXT-2022-036.

References

- [1] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(6), 2872–2893 (2021)
- [2] Zhou, K., Xiang, T.: Torchreid: A library for deep learning person re-identification in pytorch. *arXiv preprint arXiv:1910.10093* (2019)
- [3] Yu, Z., Zhao, Y., Hong, B., Jin, Z., Huang, J., Cai, D., Hua, X.-S.: Apparel-invariant feature learning for person re-identification. *IEEE Transactions on Multimedia* **24**, 4482–4492 (2021)
- [4] Yan, P., Liu, X., Zhang, P., Lu, H.: Learning convolutional multi-level transformers for image-based person re-identification. *Visual Intelligence* **1**(1), 24 (2023)

- [5] Niu, K., Huang, T., Huang, L., Wang, L., Zhang, Y.: Improving inconspicuous attributes modeling for person search by language. *IEEE Transactions on Image Processing* **32**, 3429–3441 (2023)
- [6] Niu, K., Huang, L., Long, Y., Huang, Y., Wang, L., Zhang, Y.: Comprehensive attribute prediction learning for person search by language. *IEEE Transactions on Image Processing* (2024)
- [7] Zheng, A., Pan, P., Li, H., Li, C., Luo, B., Tan, C., Jia, R.: Progressive attribute embedding for accurate cross-modality person re-id. In: *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 4309–4317 (2022)
- [8] Li, A., Zhang, K., Wang, L.: Zero-shot fine-grained classification by deep feature learning with semantics. *Machine Intelligence Research* **16**, 563–574 (2019)
- [9] Kutbi, M., Peng, K.-C., Wu, Z.: Zero-shot deep domain adaptation with common representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7), 3909–3924 (2021)
- [10] Mirza, M.J., Micorek, J., Possegger, H., Bischof, H.: The norm must go on: Dynamic unsupervised domain adaptation by normalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14765–14775 (2022)
- [11] Chen, H., Wang, Y., Lagadec, B., Dantcheva, A., Bremond, F.: Joint generative and contrastive learning for unsupervised person re-identification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2004–2013 (2021)
- [12] Dai, Y., Liu, J., Sun, Y., Tong, Z., Zhang, C., Duan, L.-Y.: Idm: An intermediate domain module for domain adaptive person re-id. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11864–11874 (2021)
- [13] Sun, J., Li, Y., Chen, H., Zhu, X., Peng, Y., Peng, Y.: Inter-cluster and intra-cluster joint optimization for unsupervised cross-domain person re-identification. *Knowledge-Based Systems* **251**, 109162 (2022)
- [14] Tao, Y., Zhang, J., Hong, J., Zhu, Y.: Dreamt: Diversity enlarged mutual teaching for unsupervised domain adaptive person re-identifcation. *IEEE Transactions on Multimedia* (2022)
- [15] Wu, Y., Huang, T., Yao, H., Zhang, C., Shao, Y., Han, C., Gao, C., Sang, N.: Multi-centroid representation network for domain adaptive person re-id. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.

- 36, pp. 2750–2758 (2022)
- [16] Lee, G., Lee, S., Kim, D., Shin, Y., Yoon, Y., Ham, B.: 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)
 - [17] Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning, pp. 6028–6039 (2020)
 - [18] Qu, X., Zhang, H., Zhu, L., Nie, L., Liu, L.: Aamt: Adversarial attack-driven mutual teaching for source-free domain-adaptive person reidentification. *IEEE Transactions on Multimedia* (2024)
 - [19] Qu, X., Liu, L., Zhu, L., Nie, L., Zhang, H.: Instance-level adversarial source-free domain adaptive person re-identification. *ACM Transactions on Multimedia Computing, Communications and Applications* (2024)
 - [20] Qu, X., Liu, L., Zhu, L., Nie, L., Zhang, H.: Source-free style-diversity adversarial domain adaptation with privacy-preservation for person re-identification. *Knowledge-Based Systems* **283**, 111150 (2024)
 - [21] Chen, F., Wang, N., Tang, J., Yan, P., Yu, J.: Unsupervised person re-identification via multi-domain joint learning. *Pattern Recognition* **138**, 109369 (2023)
 - [22] Han, J., Li, Y.-L., Wang, S.: Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 790–798 (2022)
 - [23] Chen, H., Lagadec, B., Bremond, F.: Enhancing diversity in teacher-student networks via asymmetric branches for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1–10 (2021)
 - [24] Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. arXiv preprint arXiv:2001.01526 (2020)
 - [25] MacQueen, J., *et al.*: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297 (1967). Oakland, CA, USA

- [26] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., *et al.*: A density-based algorithm for discovering clusters in large spatial databases with noise. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 96, pp. 226–231 (1996)
- [27] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1116–1124 (2015)
- [28] Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 79–88 (2018)
- [29] Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* **106**, 249–259 (2018)
- [30] Cortes, C., Lawarence, N., Lee, D., Sugiyama, M., Garnett, R.: Advances in neural information processing systems 28. In: Proceedings of the 29th Annual Conference on Neural Information Processing Systems (2015)
- [31] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2223–2232 (2017)
- [32] Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 994–1003 (2018)
- [33] Zou, Y., Yang, X., Yu, Z., Kumar, B.V., Kautz, J.: Joint disentangling and adaptation for cross-domain person re-identification. In: Proceedings of the European Conference on Computer Vision (2020). Springer
- [34] Ge, Y., Zhu, F., Chen, D., Zhao, R., *et al.*: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. *Advances in Neural Information Processing Systems* **33**, 11309–11321 (2020)
- [35] Zhang, Y., Zhang, F., Jin, Y., Cen, Y., Voronin, V., Wan, S.: Local correlation ensemble with gcn based on attention features for cross-domain person re-id. *ACM Transactions on Multimedia Computing, Communications and Applications* **19**(2), 1–22 (2023)
- [36] Liang, J., Hu, D., Wang, Y., He, R., Feng, J.: Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling

- transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(11), 8602–8617 (2021)
- [37] Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
- [38] Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059 (2016)
- [39] Kuhn, H.W.: The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2**(1-2), 83–97 (1955)
- [40] Dai, Z., Wang, G., Yuan, W., Zhu, S., Tan, P.: Cluster contrast for unsupervised person re-identification. In: Proceedings of the Asian Conference on Computer Vision, pp. 1142–1160 (2022)
- [41] Wang, D., Zhang, S.: Unsupervised person re-identification via multi-label classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10981–10990 (2020)
- [42] Chen, J., Gao, C., Sun, L., Sang, N.: Ccsd: cross-camera self-distillation for unsupervised person re-identification. *Visual Intelligence* **1**(1), 27 (2023)
- [43] Lu, H., Huo, Y., Ding, M., Fei, N., Lu, Z.: Cross-modal contrastive learning for generalizable and efficient image-text retrieval. *Machine Intelligence Research* **20**, 569–582 (2023)
- [44] Wang, W.-C., Ahn, E., Feng, D., Kim, J.: A review of predictive and contrastive self-supervised learning for medical images. *Machine Intelligence Research* **20**, 483–513 (2023)
- [45] She, D., Xu, K.: Contrastive self-supervised representation learning using synthetic data. *Machine Intelligence Research* **18**, 556–567 (2021)
- [46] Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
- [47] Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)
- [48] Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.

3754–3762 (2017)

- [49] Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Proceedings of the European Conference on Computer Vision, pp. 17–35 (2016)
- [50] Zhang, T., Xie, L., Wei, L., Zhuang, Z., Zhang, Y., Li, B., Tian, Q.: Unre-alperson: An adaptive pipeline towards costless person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11506–11515 (2021)
- [51] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
- [52] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
- [53] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- [54] Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13001–13008 (2020)
- [55] Zheng, Y., Tang, S., Teng, G., Ge, Y., Liu, K., Qin, J., Qi, D., Chen, D.: 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)
- [56] Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(11) (2008)



Aihua Zheng received B.Eng. degrees and finished Master-Doctor Combined Program in Computer Science and Technology from Anhui University of China in 2006 and 2008, respectively. And received Ph.D. degree in computer science from University of Greenwich of UK in 2012. She visited University of Stirling and Texas State University during June to September in 2013 and during September 2019 to August 2020 respectively. She is currently an Professor and PhD supervisor at the School of Artificial Intelligence, Anhui University. Her main research interests include vision based artificial intelligence and pattern recognition. Especially on person/vehicle re-identification, audio visual computing, and multi-modal intelligence.

E-mail: ahzheng214@foxmail.com

ORCID iD: 0000-0002-9820-4743



Zhihao Fei received his B.Eng. degree in 2022 and is currently pursuing the M.Eng degree in the School of Artificial Intelligence, Anhui University, Hefei, China. His research interests include Computer Vision, Domain Adaptation and Person Re-identification.

E-mail: zhfly_021@163.com

Yuhe Ding is currently a Ph.D. student in the School of Computer Science and Technology, Anhui University, Hefei, China. She received her B.Eng. degree from Anhui University in July 2019. From 2020 to 2022, she was a visiting student at



Institute of Automation, Chinese Academy of Sciences. Her research interests focus on transfer learning, pattern recognition, and computer vision.

E-mail: madao3c@foxmail.com



Chenglong Li received the M.S. and Ph.D. degrees from the School of Computer Science and Technology, Anhui University, Hefei, China, in 2013 and 2016, respectively. From 2014 to 2015, he worked as a Visiting Student with the School of Data and Computer Science, Sun Yatsen University, Guangzhou, China. He was a post-doctoral research fellow at the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He is currently an Professor and PhD supervisor at the School of Artificial Intelligence, Anhui University. His research interests include computer vision and deep learning. He was a recipient of the ACM Hefei Doctoral Dissertation Award in 2016.

E-mail: lcl1314@foxmail.com (Corresponding author)

ORCID iD: 0000-0002-7233-2739

Bin Luo received the B.Eng. degree in electronics and the M.Eng. degree in computer science from the Anhui University of China, Hefei, China, in 1984 and 1991, respectively, and the Ph.D. degree in computer science from the University of York, York, U.K., in 2002. He has authored or coauthored more than 200 papers in journals and refereed conferences. He is currently a Professor with the Anhui University



of China. His research interests include random graphbased pattern recognition, image and graph matching, and spectral analysis. He is currently the Chair of the IEEE Hefei Subsection. He was a peer Reviewer of international academic journals such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Pattern Recognition, Pattern Recognition Letters.

E-mail: ahu_lb@163.com

ORCID iD: 0000-0002-1414-3307