REGULAR PAPER

Domain-adaptive person re-identification via domain alignment and mutual pseudo-label refinement

Songhao Zhu[1] · Tao Luo[1]

Abstract
Unsupervised domain-adaptive person re-identification refers to transferring knowledge from labeled to unlabeled datasets, thus alleviating the need for large amounts of labeled data. Existing methods address this problem using clustering methods to generate pseudo-labels. However, the pseudo-labels generated by current existing methods may be unstable and noisy, which will significantly degrade the performance of the method. In this paper, we propose a novel domain-adaptive person re-identification method via domain alignment and mutual pseudo-label refinement. First, we extract discriminative feature from the augmented data using a two-branch structure to enrich the feature diversity; second, we design a distributed adversarial domain alignment module to minimize domain differences; finally, we propose a consistency between local features and global features to refine pseudo-labels predicted by global features to exploit the complementary relationship between local and global features, and thus the noise generated by pseudo-label clustering is effectively reduced. Extensive experiments demonstrate that the proposed method can achieve remarkable results on popular benchmark datasets for domain-adaptive person re-identification.

Keywords Person re-identification · Domain adaptation · Domain alignment · Label refinement

1 Introduction

Person re-identification refers to finding the same person from different cameras, and it has a wide range of applications in finding missing persons, absconding criminal suspects, and other person-related scenarios. With the rapid development of deep learning, person re-identification methods have reached a certain height, among which supervised person re-identification methods as in [1–3] have achieved satisfactory accuracy. However, supervised methods require a large number of label annotations, which greatly increases the manpower material and financial resources. In such a situation, unsupervised methods have received increasing attention, and currently unsupervised methods mainly include fully unsupervised method and unsupervised domain-adaptive (UDA) method that will be utilized in this paper.

Currently, there are three major categories of unsupervised domain-adaptive person re-recognition methods. The first method utilizes clustering algorithm to assign pseudo-labels to the samples in unlabeled target domain, which can achieve the best performance so far, where SPCL in [4] is proposed to cluster the samples using the result of mutual distillation of two networks and then assign pseudo-labels to unlabeled target domain images. The second method utilizes generative adversarial networks to transform source domain characteristic distribution into target domain characteristic distribution, and then learns some domain-invariant characteristics, where SPGAN in [5] transforms source domain images into target domain images to keep the identity invariance between two domains using self-designed generative adversarial networks, and then utilizes the generated images to fine-tune the networks, and HHL in [6] learns the camera style invariant features in the style-shifted images. However, the performance of these algorithms is heavily dependent on the quality of generated images using generative adversarial networks, and does not sufficiently consider the relationship

✉ Songhao Zhu
zhush@njupt.edu.cn

Tao Luo
alphaluo@foxmail.com

[1] College of Automation and Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, China

between source domain samples and target domain samples and the relationship between target domain samples themselves. The third method tries to optimize the neural network by computing the similarity between reference images, and these reference images have soft labels different from pseudo-labels. However, the third method also does not consider the relationship that exists between source domain samples and the target samples, and the relationship between different samples of the same person in target domain. Based on the above analysis, the first method will be here utilized to deal with the person re-identification problem.

For the first category of the unsupervised domain-adaptive methods, the pseudo-labels generated by clustering its inherent noise, to a certain extent, reduce the performance of the method. To deal with noisy labels, mutual teaching in [7, 8] is designed to train pairwise networks and correct each other. However, these two networks are prone to overfitting each other. Recently, MEBNet in [9] utilizes multiple networks with different architectures to enhance feature diversity and attempts to reduce noisy labels by brainstorming training strategies. However, such an approach requires iteratively training multiple networks, which is time-consuming.

In non-domain alignment tasks, due to sampling variability, the label space of samples from the source and target domains differs within each mini-batch. This discrepancy can result in the generation of outliers and negatively impact the generalization performance of the model, leading to reduced recognition ability. In domain alignment tasks, this disparity can be leveraged to adjust the classifier to align the energy distribution of the target domain with that of the source domain, thereby mitigating the effects of domain shift and reducing the domain gap to better accommodate random sampling variability. This adjustment enhances the model's recognition ability.

To address the above problem, a novel domain-adaptive person re-identification method based on domain alignment and mutual pseudo-label refinement is here proposed. First, an instance-level domain-aligned module is designed to map features from two domains to a common feature space to learn domain-invariant feature representations by minimizing the distribution difference between both domain images. Furthermore, a pseudo-label refinement module is designed to mutually guide the relationship between global and local features, where local features are utilized to refine the pseudo-labels generated by global features and the generated pseudo-labels are fine-tuned by computing the consistency of different clustering results. The above pseudo-label refinement process is divided into the following stages: the features in different branches are firstly utilized to cluster the images in target domain; second, due to the diversity of features, the clustering results vary; third, the reliability of which category each sample comes from is evaluated by computing the intersection of the different clusters; finally,

the negative impact of pseudo-labels can be mitigated by selecting samples with reliable labels and incorporating the reliability into the re-identification loss.

Our contributions can be summarized as follows:

1. We propose a dual-branch structure for mutually guided learning, which utilizes instance-level domain alignment modules and mutually guided prediction methods to enhance the quality of pseudo-labels.
2. We propose a novel instance-level domain alignment module, which learns domain-invariant features and reduces domain gaps by minimizing the discrepancy in feature distribution between domains.
3. We propose a novel pseudo-label refinement module, which reduces pseudo-label noise by constructing the correlation between global features and local features.
4. Experiments show that our method achieves a tradeable effect and surpasses most state-of-the-art methods by large margins on multiple benchmarks of unsupervised domain-adaptive Re-ID.

2 Related work

2.1 Deep person re-identification

With the development of deep learning and large-scale image benchmarking, person re-identification has recently become a popular research topic. Existing supervised person re-identification models can be divided into two-step and one-step frameworks. Two-step framework typically consists of separately trained person detection and person re-recognition models, where different combinations of detection and re-identification models are systematically evaluated in [10] and [11] address the inconsistency between detection and re-identification tasks. The one-step framework designs a unified model to jointly solve person detection and person re-identification tasks in an end-to-end manner to make the pipeline more efficient, where [12] introduces a graphical model to explore the impact of contextual information on identity matching, [13] decomposes human representations into norms and angles to eliminate cross-task conflicts, and [14] develops a sequential structure to reduce low-quality proposals.

Recently, weakly supervised settings without accessible personal identity labels have been proposed, where [15–17] introduce a strip-based method to segment an image into different patches and extract local features of each patch, and strong baseline in [18] employs an effective training technique for person re-identification and proposes the BNNeck structure to match identity loss and triple loss. Although these methods achieve satisfactory results on labeled datasets, the results are poor when trained on unlabeled datasets.

2.2 Domain adaptation

The goal of domain adaptation is to transfer the knowledge acquired from a well-labeled source domain to a target domain. Typically, these two domains possess distinct feature distributions, known as the domain gap, creating a challenge for performance improvement. The majority of domain adaptation algorithms [55–57] can be broadly classified into two categories: feature level and sample level. For instance, MDD [58] tackles inter-domain divergence and intra-class density by minimizing the former and maximizing the latter at the feature level. On the sample level, [59] introduces a symmetric mapping among domains to reconstruct target images resembling the source domain. Recent research emphasizes the significance of both feature level and sample level adaptations in unsupervised domain adaptation tasks. Consequently, [60] suggests a holistic approach that integrates feature adaptation with distribution matching and sample adaptation with landmark selection. However, the general domain adaptation pipeline, assuming identical classes between domains, is unsuitable for person re-identification tasks due to differing identities in the two person re-identification domains. Therefore, developing domain adaptation algorithms specifically tailored for person re-identification becomes imperative.

2.3 Unsupervised re-identification

Current unsupervised person re-identification methods contain the following two categories, fully unsupervised and unsupervised domain-adaptive. For fully unsupervised person re-identification, a dataset without any labels is utilized to train network model, and clustering is utilized to generate pseudo-labels. HCT in [19] presents hierarchical clustering to generate pseudo-labels and utilizes PK sampling in the training process. MMCL in [20] predicts pseudo-labels using similarity calculation and circular consistency. LReid in [21] formulates lifelong person re-identification as a domain adaptation problem, and designs a pseudo-task transformation module to map the features of the new task into the feature space of the old tasks. Group Sampling in [22] highlights the shortcomings involved in triplet sampling, and further proposes a novel group sampling strategy for unsupervised person re-identification, which addresses the negative effect of deteriorated overfitting and enhances statistical stability related to the unsupervised model.

Compared with the fully unsupervised person re-identification approach, unsupervised domain-adaptive person re-identification approach differs in that its source domain is annotated with labels. Unsupervised domain adaption is performed by transferring the knowledge learned in the labeled source domain to the unlabeled target domain and then fine-tuning learned knowledge on the target domain. Currently,

unsupervised domain-adaptive methods can be divided into generative adversarial network-based method and pseudo-label-based method. PTGAN in [23] introduces generative adversarial network to match source domain images with target domain images. MMT in [24] proposes a framework for mutual learning of teacher–student model, MEB-Net in [9] designs three networks for mutual average learning, UNRN in [25] presents a method to estimate the reliability of pseudo-labels, AWB in [26] integrates a novel light-weight module into the dual networks of mutual learning to enhance the complementarity to depress the noise in pseudo-labels, and DARC in [27] utilizes a novel divide-and-regroup clustering pipeline to take two characteristics of re-identification task into consideration to increase the clustering accuracy. MCM in [28] proposes a multi-centroid memory to alleviate the label noise problem in previous UDA re-identification methods, where the impact of label noises can be reduced by selecting reliable positive and negative centroids from MCM for each input query [29]. Proposes a plug-and-play intermediate domain module to smoothly bridge the source and target domains, which will better adapt between the two extremes to ease the UDA person re-identification task. However, these methods often ignore the domain gap problem caused by cross-domain. To reduce the domain gap, a novel instance-level domain alignment strategy is here proposed. Figure 1 shows an example of different domain adaptation strategies.

2.4 Pseudo-label refining

Due to challenges in obtaining high-quality labels across various real-world scenarios, there has been a growing emphasis on robust training methods that can handle noisy labels [61]. The objective of robust loss design is to identify a function that demonstrates resilience to noisy labels. [62] finds that the mean absolute error loss is effective in handling noisy labels. Loss adjustment approaches employ various techniques, such as correction through the noise transition matrix [63] or utilizing a sample re-weighting scheme based on label reliability [64], aiming to alleviate the impact of noisy labels [30]. This technique also proposes an end-to-end framework designed to measure observation noise and mitigate negative effects for improved network optimization. Additionally, [31] estimates the correctness of pseudo-labels in semantic segmentation predictions. In the context of clustering-based unsupervised domain-adaptive person re-identification, EUG in [32] and GLT in [33] are uncertainty-based approaches. Specifically, EUG utilizes the distance between samples and cluster centroids in feature space to assess the reliability of samples, while GLT explicitly corrects noisy labels to select reliable pseudo-labels for progressive model training. However, these methods typically require a sufficient number of clean labels to estimate

Domain Adaptation

(a)

(b)
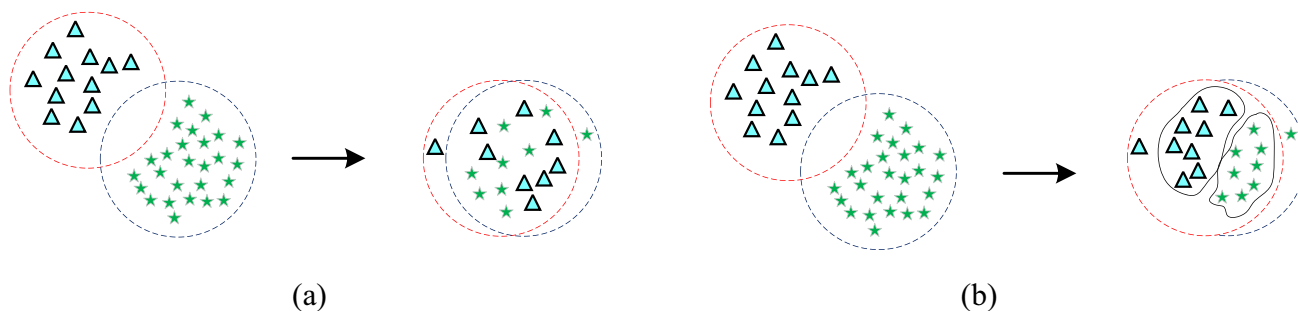
Fig. 1 Comparison between existing domain adaptation strategies and our proposed domain adaptation strategy, where (**a**) Existing strategies usually ignore the domain gap issues arising from cross-domain scenarios, (**b**) an instance-level domain alignment strategy is introduced to address this challenge

the degree of noise, making them less suitable for unsupervised person re-identification scenarios. In situations where pseudo-labels exhibit exceptionally high noise levels at the beginning of training, these approaches become impractical. Therefore, we propose a mutually refined model, which aims to reduce uncertainty and reduce the noise of pseudo-labels through the complementarity between global features and local features, thereby improving the accuracy of pseudo-label prediction.

3 Proposed method

3.1 Overview

Unsupervised domain-adaptive person re-identification task focuses on transferring the knowledge learned on the source domain with label annotations $D_s = \left\{ \left(x_i^s, y_i^s \right)|_{i=1}^{N_s} \right\}$ to the target domain without any label annotations $D_t = \left\{ \left(x_i^t \right)|_{t=1}^{N_t} \right\}$, where N_s represents the number of samples on the labeled source domain and N_t represents the number of samples on the unlabeled target domain. Furthermore, x_i^s represents the identity information of each image in the source domain, y_i^s represents the true label of each image in the source domain, and x_i^t represents the identity information of each image in the unlabeled target domain.

Figure 2 illustrates the framework of the proposed dual-branch model for unsupervised domain-adaptive person re-identification, which aims to reduce the domain gap through domain alignment module and to reduce the noise of pseudo-labels through the global and local feature extraction module. Specifically, global features are refined through local features, which in turn benefit from refinement guided by global features, which helps solve the problem of domain-adaptive pseudo-label noise and improve the accuracy of pseudo-label prediction. The specific training process is described as follows: First, the proposed model is trained with labeled source domain images; second, the trained model is iteratively trained with unlabeled target domain images, and the target domain images are clustered to generate pseudo-labels in each iteration; finally, the consistency of clustering results is utilized to assess the reliability of clustering samples, which can effectively reduce the noise in pseudo-labels.

### 3.2 Supervised training in source domain

The proposed dual-branch model aims to transfer the knowledge from labeled source domains to unlabeled target domains. Specifically, the proposed DAMPR model can output two features $D_{i,k}^n$ and predict probabilities $q(y_i^s|x_{i,k}^n)$, where $x_{i,k}^n$ is the ith sample's the nth augmented image inputted into the kth branch.

The cross-entropy loss of the proposed dual-branch model can be formulated as follows:

$$\mathcal{L}_{ce}^n = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{2} \log q_j(y_i^s|x_{ik}^n) \tag{1}$$

where $q(y_i^s|x_{i,k}^n)$ is the predicted probability of the sample x_i in the kth branch.

The SoftMax triplet loss of the proposed dual-branch model can be formulated as follows:
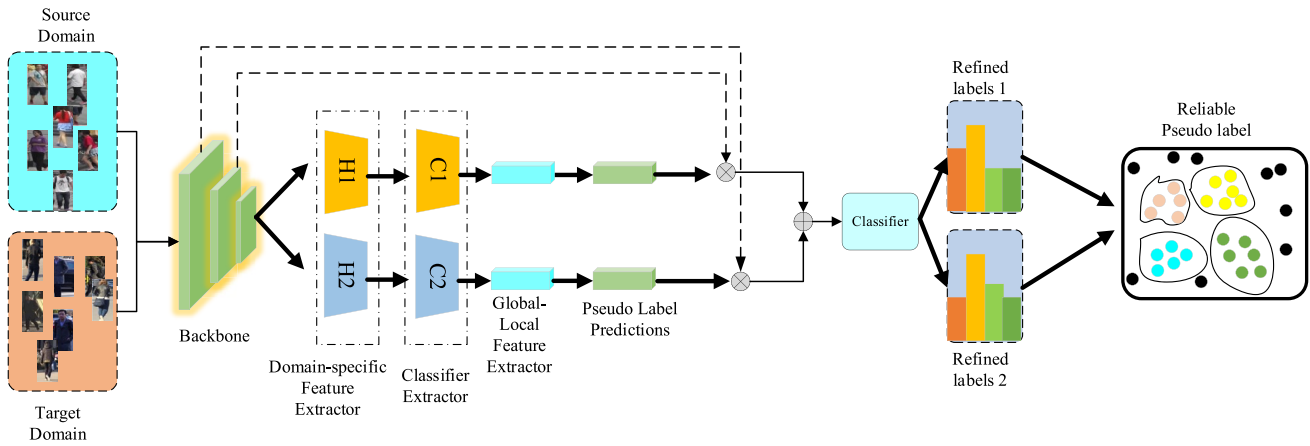
Fig. 2 The overall framework of the proposed domain alignment and mutual pseudo-label refinement (DAMPR) network, which is composed of a backbone network, a domain alignment module including feature extractors and classifier extractors, a global and local feature extraction module, a pseudo-label prediction module, and a mutual refinement module of pseudo-labels

$$\mathcal{L}_{tri}^n = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{2} log \frac{e^{\|D_{i+,k}^n - D_{i-,k}^n\|}}{e^{\|D_{i,k}^n - D_{i-,k}^n\|} + e^{\|D_{i,k}^n - D_{i+,k}^n\|}} \quad (2)$$

where $D_{i,k}^n$ is the feature for the source domain sample x_i^s in the kth branch, and $D_{i+,k}^n$ and $D_{i-,k}^n$ mean the positive and negative samples for the ith sample respectively.

A dual-branch network architecture can be utilized to obtain different features and probability predictions, where the consistency among different features can be utilized to evaluate the sample's reliability to reduce noisy samples.

3.3 Domain alignment module

To achieve aligned the characteristic distribution between source domain and target domain, maximum mean difference (MMD) is here utilized to map the characteristic distribution in source domain and target domain to another characteristic space. Through a mapping $\phi(\bullet)$, the MMD between the features from different batches can be obtained using the following formula:

$$MMD(X_S, X_T) = \|\frac{1}{|X_S|} \sum_{x_i^s} \in X_S \phi(x_i^s) - \frac{1}{|X_T|} \sum_{x_i^t} \in X_T \phi(x_i^t)\| \quad (3)$$

where X_S represents the number of randomly selected samples on the labeled source domain, and X_T represents the number of randomly selected samples on the unlabeled target domain.

The fundamental idea expressed in Eq. (3) is that if the feature distributions are identical, the associated statistics will also be identical. Formally, the Maximum Mean Discrepancy defines the variance metric, as represented by the following formula:

$$D_H = \|E[\phi(x_i^s)] - E[\phi(x_i^t)]\| \quad (4)$$

where H is the reproducing kernel Hillbert space (RKHS), which is endowed with the kernel trick $k(x^s, x^t) = (\phi(x^s), \phi(x^t))$ where $(\cdot, \cdot)$ represents inner product of vectors. Furthermore, $\phi(\cdot)$ denotes some feature map to map the original samples to RKHS.

The MMD loss is formulated as follows:

$$L_{mmd} = \frac{1}{N} \sum_{i=1}^{N} F(H_i(F(X_s)), H_i(F(X_t))) \quad (5)$$

where $F(\cdot)$ is a feature extractor, which is utilized to map domain feature space into a common feature space to learn the domain-invariant representation in source and target domains, and $H(\cdot)$ is a feature extractor, which is utilized to obtain common latent feature representations from source domain and target domain. Since target domain samples near class boundary are prone to be misclassified by the classifier learned from source domain samples, Eq. (5) is here minimized to reduce the difference between target domain samples and source domain samples.

In this paper, the absolute value of the difference between prediction probabilities for target domain samples is here utilized as the difference loss, as represented by the following formula:

$$L_{dsca} = \frac{2}{N} \sum_{j=1}^{N_{cs}} \sum_{i=1}^{N_{ct}} E\left[\left|C_i(H_i(F(X_s))) - C_j(H_i(F(X_t)))\right|\right] \quad (6)$$

where $C(\cdot)$ is the classifier, and N_{cs} and N_{ct} are the number of classifier in source domain and target domain respectively. By minimizing Eq. (6), the prediction probabilities of the

classifiers are similar, thus reducing the difference between two domains.

3.4 Pseudo-label refinement

Figure 3 illustrates the process of refining pseudo-labels by introducing local features. The following two interesting phenomena can be drawn from Fig. 3a that (1) if only global features are considered, there will be a lot of detail missing and two different images will be clustered into one category, (2) using incorrect clustering results as a supervised signal to train network model will lead to poor performance. Compared with the above situation, if local features are introduced, such as the specific features of upper body in Fig. 3a, the differences in detail will be emphasized so that these two people can be easily distinguished. In this way, local features can be utilized to refine the clustering results of global features. Similarly, global features can be utilized to refine the clustering results of local features. As shown in Fig. 3b, clustering only the local features of the lower body cannot easily distinguish two similar images. If we use global features, it is easy to distinguish the difference between the two pictures.

To improve the prediction of global features by learning local features, label smoothing in [54] is here utilized to refine pseudo-labels of each body part according to the corresponding cross-protocol score reflecting the reliability of global clustering result for each body part. Given the pseudo-label y_t of the target image x_t, the label smoothing for local feature is formulated as below:

$$y_t^{an} = (1 - \gamma_t^{an})p + \gamma_t^{an}y_t \tag{7}$$

where γ_t^{an} is a weight determining the strength of label smoothing, and p is a uniform vector. Different from conventional label smoothing that employs a constant weight for γ_t^{an}, the weight for each part γ_t^{an} is here dynamically adjusted according to the cross-protocol score. Given the refined pseudo-labels γ_t^{an}, the cross-entropy loss is formulated as below:

$$L_{ls} = \frac{1}{N}\sum_{t=1}^{N} (1 - \gamma_t^{an})D_{KL}(p||\beta_t^{an}) + (\gamma_t^{an}U(y_t, \beta_t^{an})) \tag{8}$$

where $U(\cdot)$ and $DKL(\cdot)$ are cross-entropy and KL divergence respectively, and two terms are balanced by γ_t^{an} with the value of the cross-protocol score. Furthermore, $\beta_t^{an} = h_\varnothing(f_i^{an})$ is the prediction vector of the nth local feature a_n, where $h_\varnothing(\cdot)$ is the global feature classifier consisting of a fully connected layer and a SoftMax function.

The core idea of mutual refinement of pseudo-labels of global features and local features is to retain only those samples, whose pseudo-labels are consistent with that of the rest. A mutually guided pseudo-label refinement strategy is here proposed to predict pseudo-labels of local features to refine pseudo-labels of global features, simultaneously for pseudo-labels of local features, pseudo-labels of global features will be utilized to refine it. Considering that less differentiated fragments in local features may generate misleading insights, cross-protocol scores are utilized to aggregate the predictions of pseudo-labels of local features to improve the accuracy of pseudo-labels. The mutual-guided refined label g_t^q obtained by Eq. (9) is here considered as a pseudo-label for the global feature:

Fig. 3 An example of local–global feature mutual refinement module, where (**a**) represents the refinement of global features through the fine-grained local features, and (**b**) represents the refinement of local features through the comprehensiveness of global features

(a)　　　　　　　　　　(b)

$$g_t^q = \left(y_t - \sum_{t=1}^{N} u_t \beta_t^{an} \right) + (1 - \delta)u_t \beta_t^{an} \qquad (9)$$
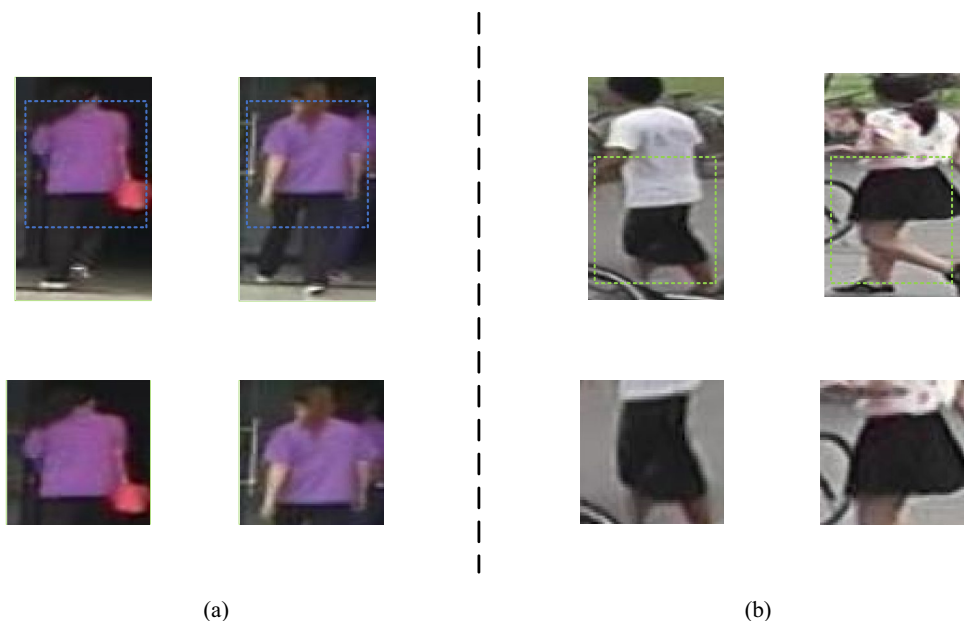
where u_t is the ensemble weight, and $\delta \in [0, 1]$ is the weight controlling the ratio of pseudo-label and ensembled prediction. Given the obtained refined pseudo-label g_t^q, the pseudo-labels are mutually refined using the following formulation:

$$L_{mgrp} = - \sum_{t=1}^{N} g_t^q log\left(a_t^q\right) \qquad (10)$$

where a_t^q is the prediction vector by the global feature. This mutually guided label refinement can help the model to fully exploit the knowledge of global and local features.

Total loss for the target domain can be formulated using the following equation:

$$L_{target} = L_{dsca} + L_{ls} + L_{mgrp} \qquad (11)$$

4 Experimental results

4.1 Datasets and evaluation protocols

We evaluate our method on three large-scale re-identification datasets: Market-1501 dataset in [34], DukeMTMC-ReID dataset in [35] and MSMT17 dataset in [23].

Market-1501 dataset contains 1501 identities with 32,668 images, which was captured by 6 different cameras. The training set contains 751 identities with 12,936 images, and the testing set contains 750 identities with 19,732 images, where the query set contains 3368 images and the gallery set contains 16,364 images.

DukeMTMC-ReID dataset is a sub-dataset of Duke-MTMC, which contains 1812 identities with 36,411 images, which was captured by 8 high-definition cameras. These 36,411 images are divided into 16,522 training images, 2228 query images, and 17,661 gallery images.

MSMT17 dataset is a large-scale dataset, which contains 4101 identities with 126,441 images. The training set contains 1041 identities and testing set contains 3060 identities.

Cumulative Matching Characteristic (CMC) and Mean Average Precision (mAP) are utilized to evaluate the model performance. All experiment results are obtained under the single-query setting, and no post-processing is applied.

4.2 Implementation details

The input images are resized to 256×128, and padding, random flip, and random crop are employed as data augmentation in both source domain pre-training and target domain fine-tuning.

ResNet-50 [36] pretrained on ImageNet [37] is adopted as the backbone, which is trained a total of 80 epochs where each epoch contains 400 iterations. Adam [53] with weight decay of 5×10^{-4} is adopted for training, and the initial learning rate is set to be 3.5×10^{-4} and is divided by 10 at the 40th and 60th epoch, in a total 80 epochs. We utilize the DBSCAN clustering algorithm, and the Jaccard distance with k-reciprocal nearest neighbor is used as the distance metric. The eps in DBSACN is set to be 0.6.

4.3 Experimental results

We compare our method with other unsupervised re-ID methods on Market-1501, MSMT17, DukeMTMC-ReID, and all the results are shown in Table 1. Our method obtains the performance of 82.3% on mAP and 93.2% on rank-1 when transferring DukeMTMC-ReID to Market-1501.

Among existing methods for UDA person re-identification, SSG in [38], MMT, MEBNet, and UNRN are all clustering-based methods. SSG employs both global body and local body part features for clustering and evaluation. We construct the baseline based on P^{2}LR which introduces probabilistic uncertainty of pseudo-labels for UDA person re-identification. Compared to the baseline P^{2}LR, our proposed DAMPR significantly improves the UDA re-identification accuracy with 1.3%, 0.5%, 3.4%, and 4.3% mAP improvements on four UDA re-identification settings. Compared to MEBNet which establishes three networks to perform mutual mean learning, we increase the mAP by 6.3%, 5.2% with a simpler architecture design. Notably, UNRN and GLT leverage source data during target fine-tuning stage and build an external support memory to mine hard pairs. Our DAMPR still achieves 4.2% and 2.2% mAP gains to UNRN, 2.8% and 2.1% mAP gains to GLT on the public dataset.

4.4 Ablation study

In this section, we evaluate each components of our proposed framework by conducting ablation studies on Duke-MTMC-ReID → Market-1501, Market-1501 → Duke-MTMC-ReID, DukeMTMC-ReID → MSMT17 and Market-1501 → MSMT17 tasks. The experimental results are shown in Table 2.

Effectiveness of instance-level alignment: To validate the effectiveness of our task-sensitive instance-level alignment design, we compare it with normal domain alignment conduct instance alignment on both head networks without balancing between them. As observed in Table 2, the task-sensitive design successfully alleviates the inner task conflicts and outperforms normal strategy by a large margin.

Table 1 Performance comparison of the proposed method and state-of-the-art methods for domain adaptation on DukeMTMC-ReID, Market-1501, and MSMT17 datasets

Methods	DukeMTMC to Market-1501				Market-1501 to DukeMTMC			
	mAP	R1	R5	R10	mAP	R1	R5	R10
PUL [39]	20.5	45.5	60.7	66.7	16.4	30.0	43.4	48.5
SPGAN+LMP [5]	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
BUC [50]	38.3	66.2	79.6	84.5	27.5	47.4	62.6	68.4
ECN [40]	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
PDA-Net [41]	47.6	75.2	86.3	90.2	45.1	63.2	77.0	82.5
PCB-PAST [42]	54.6	78.4	–	–	54.3	72.4	–	–
MPLP [43]	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
AD-Cluster [44]	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
MMT [24]	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
NRMT [51]	71.7	87.8	94.6	96.5	62.2	77.8	86.9	89.5
HGA [45]	70.3	89.5	–	–	67.1	80.4	–	–
B-SNR+GDS-H [46]	72.5	89.3	–	–	59.7	76.7	–	–
SpCL [4]	76.7	90.3	96.1	97.6	68.8	82.9	90.1	92.5
MEB-Net [9]	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
Dual-Refinement [47]	78.0	90.9	96.4	97.7	67.7	82.1	90.1	92.5
UNRN [25]	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5
GLT [33]	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8
P^{2}LR [48]	81.0	92.6	97.4	98.3	70.8	82.6	90.8	93.7
CMT [49]	81.1	92.9	97.1	98.0	70.1	82.7	90.8	93.8
Ours	82.3	93.2	97.7	98.4	71.3	83.9	91.3	93.8
Methods	DukeMTMC to MSMT17				Market-1501 to MSMT17			
	mAP	R1	R5	R10	mAP	R1	R5	R10
ECN [40]	8.5	23.5	36.3	42.1	10.2	30.2	41.5	46.8
SSG [38]	13.2	31.6	–	49.6	13.3	32.2	–	51.2
MMT [24]	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
JVTC+ [52]	27.5	52.9	70.5	75.9	25.1	48.6	65.3	68.2
SPCL [4]	26.5	53.1	65.8	70.5	26.8	53.7	65.0	69.8
Dual-refinement [47]	26.9	55.0	68.4	73.2	25.1	53.3	66.1	71.5
UNRN [25]	26.2	54.9	67.3	70.6	25.3	52.4	64.7	69.7
GLT [33]	27.7	59.5	70.1	74.2	26.5	56.6	67.5	72.0
P^{2}LR [48]	29.9	60.9	73.1	77.9	29.0	58.8	71.2	76.0
CMT [49]	28.7	59.3	71.7	76.5	28.2	59.8	70.4	75.1
Ours	33.3	64.5	74.7	79.6	33.3	65.2	74.7	79.3

Effectiveness of mutual-guided label refinement: To verify the effectiveness of MGRP, we evaluate other label refinement techniques. One way is to refine labels with the prediction of global features by the mean-teacher model [24]. We further investigate MGRP without cross-protocol scores by averaging the predictions of part features. As shown in Table 2, our MGRP significantly outperforms other label refinement methods. It demonstrates the superiority of MGRP and the effectiveness of the cross-protocol score. The refined pseudo-label by MGRP captures reliable fine-grained information that cannot be achieved by considering only global features, and it helps to generate more effective refined labels.

Comparisons with supervised learning: In Table 3, we compare the performance of supervised learning, direct transfer, and DAMPR. The fully supervised learning utilizes the ground-truth label to train the model and thus gets the best performance. When directly transferring the model from Market-1501 to DukeMTMC-ReID, the performance of mAP drops from 85.6% to 28.2%, which means there is a large domain gap between the two datasets. Our method

Table 2 Ablation study on the effectiveness of components in DAMPR method

Methods	DukeMTMC to Market-1501		Market-1501 to DukeMTMC	
	mAP	R1	mAP	R1
Model pre-training	30.3	58.6	53.4	79.8
Baseline	81.0	91.6	70.8	82.6
Baseline+DA	81.4	91.9	70.9	82.7
Baseline+MGRP	81.6	92.8	71.1	83.4
All	82.3	93.2	71.3	83.9

DA domain alignment module, *MGRP* mutual refinement of pseudo-labels of global features and local features

Table 3 Comparisons with supervised learning, direct transfer, and single branch

Methods	DukeMTMC to Market-1501		Market-1501 to DukeMTMC	
	mAP	R1	mAP	R1
Fully supervised	85.6	93.9	76.0	87.2
Direct Transfer	28.2	57.4	27.0	47.1
Single Branch	80.2	92.0	71.1	82.8
Ours	82.3	93.2	71.3	83.9

improves the mAP from 81.0% to 82.3% compared with the baseline. And even use a single branch in the inference stage, our method can also achieve 80.2% mAP and 93.2% rank-1, which is superior to other methods.

4.5 Parameter analysis

We analyze the impact of parameter in our method δ being the weighting parameter for mutual-guided label refinement. We tune the value of parameter while keeping the others fixed, and the results are in Fig. 4. We can find that when δ <0.3 or δ>0.7, the performance decreases. This is because, with a small δ, samples with noisy pseudo-label cannot be found. But when δ is too large, less sample can be selected for training. The predictions of the initial training stage usually output uniform distributions, so the labels refined by MGRP also collapse to uniform distributions, providing noisy training signals. Based on these experimental results, we set $\delta = 0.5$.

Visualization: We present the visualization results to validate the effectiveness of domain alignment and mutual pseudo-label refinement for domain-adaptive person Re-Identification. Figure 5 demonstrates 3 pairs of ranking results, where the ranking results of other methods in Fig. 5a and the ranking results of our method in Fig. 5b. We can observe that the wrong matching results in Fig. 5a generally have similar clothing or background, which share similar styles. Our method can reduce the mismatches by complementing local and global features each other. This phenomenon confirms the effectiveness of our approach.

5 Conclusion

In this paper, we have proposed a domain alignment and mutual pseudo-label refinement for domain-adaptive person re-identification. We design a novel domain alignment module to learn domain-invariant representations by minimizing the domain distribution differences to reduce the domain
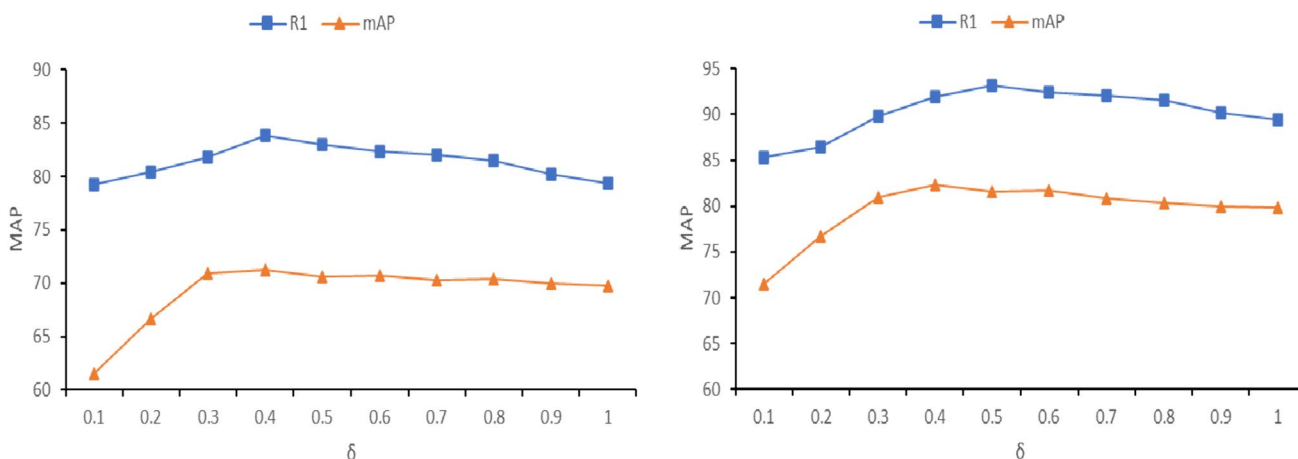
Fig. 4 Performance comparison with different weighting parameter δ, left is Duke to market, right is Market to Duke

(a)   (b)

Fig. 5 Examples of ranking results on Market-1501. The green and red boxes indicate the correct matchings and the wrong matchings, respectively. **a** is the results of other methods, and **b** is the results of our methods

gap, and design a novel pseudo-label refinement module to reduce the noisy pseudo-labels by constructing the correlation of global features and local features. Our method achieves superior performance on benchmark dataset.

Declarations

Conflict of interest This declaration is not applicable.

Ethical approval This declaration is not applicable.

References

1. Wei Li, Rui Zhao, Tong Xiao, Xiaogang Wang. Deep ReID: deep filter pairing neural network for person re-identification. IEEE conference on computer vision and pattern recognition, 2014: 152–159.
2. Wei Li, Xiatian Zhu, Shaogang Gong. Harmonious attention network for person re-identification. IEEE conference on computer vision and pattern recognition, 2018: 2285–2294.
3. Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, Qi Tian. Glad: global-local-alignment descriptor for pedestrian retrieval. ACM international conference on multimedia, 2017: 420–428.
4. Yixiao Ge, Dapeng Chen, Feng Zhu, Rui Zhao, Hong sheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-Id. Annual conference on neural information processing systems, 2020: 1–13.
5. Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. IEEE conference on computer vision and pattern recognition, 2018: 994–1003.
6. Zhun Zhong, Liang Zheng, Shaozi Li,Yi Yang. Generalizing a person retrieval model hetero and homogeneously. European conference on computer vision, 2018: 172–188.
7. Ying Zhang, Tao Xiang, Timothy M. Hospedales, Huchuan Lu. Deep mutual learning. IEEE conference on computer vision and pattern recognition, 2018: 4320–4328.
8. Shi, Y., Ling, H., Lei, Wu., Shen, J., Li, P.: Learning refined attribute-aligned network with attribute selection for person re-identification. Neurocomputing **402**, 124–133 (2020)
9. Yunpeng Zhai, Qixiang Ye, Shijian Lu, Mengxi Jia, Rongrong Ji, Yonghong Tian. Multiple expert brainstorming for domain adaptive person re-identification. European conference on computer vision, 2020: 594–611.
10. Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, Yi Yang, Qi Tian. Person re-identification in the wild. IEEE conference on computer vision and pattern recognition, 2017: 3346–3355.
11. Cheng Wang, Bingpeng Ma, Hong Chang, Shiguang Shan, Xilin Chen. TCTS: A task-consistent two stage framework for person search. IEEE conference on computer vision and pattern recognition, 2020: 11949–11958.
12. Yichao Yan, Qiang Zhang, Bingbing Ni, Wendong Zhang, Minghao Xu, Xiaokang Yang. Learning context graph for person search. IEEE conference on computer vision and pattern recognition, 2019: 2158–2167.
13. Di Chen, Shanshan Zhang, Jian Yang, Bernt Schiele. Norm-aware embedding for efficient person search. IEEE conference on computer vision and pattern recognition, 2020: 12612–12621.
14. Zhengjia Li, Duoqian Miao. Sequential end-to-end network for efficient person search. AAAI conference on artificial intelligence, 2021: 2011–2019.
15. Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, Shengjin Wang. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). European conference on computer vision, 2018: 501–518.

16. Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, Jian Sun. Aligned ReID: surpassing human-level performance in person re-identification. CoRRabs/1711.08184, 2017.

17. Luo, H., Jiang, W., Zhang, X., Fan, X., Qian, J., Zhang, C.: Aligned ReID++: dynamically matching local information for person re-identification. Pattern Recogn. **94**, 53–61 (2019)

18. Luo, H., Jiang, W., Youzhi, Gu., Liu, F., Liao, X., Lai, S., Jianyang, Gu.: A strong baseline and batch normalization neck for deep person re-identification. IEEE Trans. Multimedia **22**(10), 2597–2609 (2020)

19. Kaiwei Zeng, Munan Ning, Yaohua Wang, Yang Guo. Hierarchical clustering with hard-batch triplet loss for person re-identification. IEEE conference on computer vision and pattern recognition, 2020: 13657–13665.

20. Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. IEEE conference on computer vision and pattern recognition, 2020: 10981–10990.

21. Wenhang Ge, Junlong Du, Ancong Wu, Yuqiao Xian, Ke Yan, Feiyue Huang, Weishi Zheng. Lifelong person re-identification by pseudo task knowledge preservation. AAAI conference on artificial intelligence, 2022: 688–696.

22. Han, X., Xuehui, Yu., Li, G., Zhao, J., Pan, G., Ye, Q., Jiao, J., Han, Z.: Rethinking sampling strategies for unsupervised person re-identification. IEEE Trans. Image Process. **32**, 29–42 (2023)

23. Longhui Wei, Shiliang Zhang, Wen Gao, Qi Tian. Person transfer GAN to bridge domain gap for person re-identification. IEEE conference on computer vision and pattern recognition, 2018: 79–88.

24. Yixiao Ge, Dapeng Chen, Hongsheng Li. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. International conference on learning representations, 2020: 1–15.

25. Kecheng Zheng, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, Zhengjun Zha. Exploiting sample uncertainty for domain adaptive person re-identification. AAAI conference on artificial intelligence, 2021: 3538–3546.

26. Wang, W., Zhao, F., Liao, S.: Attentive wave block: complementarity-enhanced mutual networks for unsupervised domain adaptation in person re-identification and beyond. IEEE Trans. Image Process. **31**, 1532–2154 (2022)

27. Zhengdong Hu, Yifan Sun, Yi Yang, Jianguang Zhou. Divide-and-regroup clustering for domain adaptive person re-identification AAAI conference on artificial intelligence, 2022: 980–988.

28. Yuhang Wu, Tengteng Huang, Haotian Yao, Chi Zhang, Yuanjie Shao, Chuchu Han, Changxin Gao, Nong Sang. Multi-centroid representation network for domain adaptive person re-ID. AAAI conference on artificial intelligence, 2022: 2750–2758.

29. Yongxing Dai, Jun Liu, Yifan Sun, Zekun Tong, Chi Zhang, LingYu Duan. IDM: an intermediate domain module for domain adaptive person re-ID. IEEE conference on computer vision, 2021:11864–11874.

30. Alex Kendall, Yarin Gal:. What uncertainties do we need in bayesian deep learning for computer vision? Annual conference on neural information processing systems, 2017 5574–5584.

31. Zheng, Z., Yang, Yi.: Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. Int. J. Comput. Vision **129**(4), 1106–1120 (2021)

32. Yu, Wu., Lin, Y., Dong, X., Yan, Y., Bian, W., Yang, Yi.: Progressive learning for person re-identification with one example. IEEE Trans. Image Process. **28**(6), 2872–2881 (2019)

33. Kecheng Zheng, Wu Liu, Lingxiao He, Tao Mei, Jiebo Luo, and Zhengjun Zha. Group-aware label transfer for domain adaptive person re-identification. IEEE conference on computer vision and pattern recognition, 2021: 5310–5319.

34. Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: a benchmark. IEEE conference on computer vision, 2015: 1116–1124.

35. Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. European conference on computer vision workshops, 2016: 17–35.

36. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition, 2016: 770–778.

37. Jia Deng, Wei Dong, Richard Socher, Lijia Li, Kai Li, and Feifei Li. Imagenet: a large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition, 2009: 248–255.

38. Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, Huang Thomas. Self-similarity Grouping: a simple unsupervised cross domain adaptation approach for person re-identification. IEEE conference on computer vision, 2019: 6111–6120.

39. Fan, H., Zheng, L., Yan, C., Yang, Yi.: Unsupervised person re-identification: clustering and fine-tuning. ACM Trans Multimedia Comput, Commun, Appl **14**(4), 1–18 (2018)

40. Zhan Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, Yi Yang. Invariance matters: exemplar memory for domain adaptive person re-identification. IEEE conference on computer vision and pattern recognition, 2019: 598–607.

41. Yujhe Li, Cisiang Lin, Yanbo Lin, Yuchiang Frank Wang. Cross dataset person re-identification via unsupervised pose disentanglement and adaptation. IEEE conference on computer vision, 2019: 7918–7928.

42. Xinyu Zhang, Jiewei Cao, Chunhua Shen, Mingyu You. Self training with progressive augmentation for unsupervised cross-domain person re-identification. IEEE conference on computer vision, 2019: 8221–8230.

43. Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. IEEE conference on computer vision and pattern recognition, 2020: 10978–10987.

44. Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, Yonghong Tian. ad-cluster: augmented discriminative clustering for domain adaptive person re-identification. IEEE conference on computer vision and pattern recognition, 2020: 9018–9027.

45. Minying Zhang, Kai Liu, Yidong Li, Shihui Guo, Hongtao Duan, Yimin Long, Yi Jin. Unsupervised domain adaptation for person re-identification via heterogeneous graph alignment. AAAI conference on artificial intelligence, 2021: 3360–3368.

46. Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen. Global distance distributions separation for unsupervised person re-identification. European conference on computer vision, 2020: 735–751.

47. Dai, Y., Liu, J., Bai, Y., Tong, Z., Duan, L.: Dual-refinement: joint label and feature refinement for unsupervised domain adaptive person re-identification. IEEE Trans. Image Process. **30**, 7815–7829 (2021)

48. Han, J., Li, Y., Wang, S.: Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification. AAAI Conf Artif Intell **36**(1), 790–798 (2022)

49. Shengming Yu, Shengjin Wang. Consistency mean-teaching for unsupervised domain adaptive person re-identification. International conference on image and graphics processing, 2022: 159–166.

50. Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. AAAI conference on artificial intelligence, 2019: 8738–8745.

51. Zhao, F., Liao, S., Xie, G.-S., Zhao, J., Zhang, K., Shao, L. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. European conference on computer vision, 2020: 526–544. Springe.

52. Jianing Li, Shiliang Zhang. Unsupervised domain adaptive person re-identification. European conference on computer vision, 2020: 483–499.

53. Diederik P Kingma, Jimmy Ba. Adam: a method for stochastic optimization. International conference on learning representations, 2015: 1–15.

54. Michal Lukasik, Srinadh Bhojanapalli, Aditya Menon, and Sanjiv Kumar. Does label smoothing mitigate label noise? In conference on learning representation, 2015: 1–15.

55. Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In proceedings of the 30th international conference on neural information processing systems, pages 136–144, 2016.

56. Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. International conference on machine learning, 2015: 1180–1189.

57. Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. IEEE conference on computer vision and pattern recognition, 2018: 3723–3732.

58. Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, Heng Tao Shen. Maximum density divergence for domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 2021, 43(11): 3918–3930.

59. Paolo Russo, Fabio M Carlucci, Tatiana Tommasi, Barbara Caputo. From source to target and back: symmetric bi-directional adaptive gan. IEEE conference on computer vision and pattern recognition, 2018: 8099–8108.

60. Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, Heng Tao Shen. Locality preserving joint transfer for domain adaptation. IEEE transactions on image processing, 2019, 28(12):6103–6115.

61. Song, H., Kim, M., Park, D., Shin, Y., Lee, J.-G.: Learning from noisy labels with deep neural networks: a survey. IEEE Trans Neural Netw Learning Syst **34**(11), 8135–8153 (2023)

62. Aritra Ghosh, Himanshu Kumar, and P. S. Sastry. Robust loss functions under label noise for deep neural networks. AAAI Conference on Artificial Intelligence, 2017: 1919–1925.

63. Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. Using trusted data to train deep networks on labels corrupted by severe noise. Annual conference on neural information processing systems, 2018: 10477–10486.

64. Haw-Shiuan Chang, Erik G. Learned-Miller, Andrew McCallum. Active bias: training more accurate neural networks by emphasizing high variance samples. Annual conference on neural information processing systems, 2017: 1002–1012.