Feature diversity learning with sample dropout for unsupervised domain adaptive person re-identification

Chunren Tang[1] · Dingyu Xue[1] · Dongyue Chen[1,2]

Abstract

Clustering-based approach has proved effective in dealing with unsupervised domain adaptive person re-identification (ReID) tasks. However, existing works along this approach still suffer from noisy pseudo labels and the unreliable representation ability during the whole training process. In order to solve these problems, in this paper, we propose a new approach to learn the feature representation with better generalization ability through limiting noisy pseudo labels. At first, we propose a Sample Dropout (SD) method to prevent the training of the model from falling into the vicious circle caused by samples that are frequently assigned with noisy pseudo labels, our method can correct the noisy labels and boost the representation ability. In addition, we put forward a new method referred as to Feature Diversity Learning (FDL) under the classic mutual-teaching architecture, which can significantly improve the generalization ability of the feature representation on the target domain in an unsupervised fashion. Experimental results show that our proposed FDL-SD achieves the state-of-the-art performance on multiple well-known benchmark datasets.

Keywords Unsupervised domain adaptation · Person re-identification · Cross domain · Feature learning

1 Introduction

Person re-identification (ReID) aims to match person images across multiple non-overlapping cameras, which has achieved attention from both industry and academia. Based on the development of deep learning technology, many complicated visual tasks can be solved with the help of powerful representation ability through deep neural networks [1, 2, 6, 15, 46]. Fortunately, the person ReID tasks can also get better results using the deep

✉ Dongyue Chen
  1910316@stu.neu.edu.cn

1   College of Information Science and Engineering, Northeastern University, Heping District, Shenyang, 110819, China

2   Foshan Graduate School of Innovation, Northeastern University, Foshan 528311, Guangdong, China

convolution networks [7, 35, 36, 54, 55]. Most of the existing person ReID models along the supervised approach [30, 31, 41–43] have achieved satisfactory performance. The models of these methods are optimized using the annotated labels, through which the models can learn the knowledge from the data distribution under the accurate supervised signal. The produced features from the well trained models can be utilized to recognize the target pedestrian. However, these models generally perform less well in real applications because they have never been trained to adapt to the application scenes. To address this issue, a new problem referred as to unsupervised domain adaptation becomes a hot topic in ReID task [33, 56, 66], which focuses on how to adapt a pretrained model from a labeled source domain to an unlabeled target domain.

The main challenge of unsupervised domain adaptive person ReID lies in learning feature representation with unlabeled target domain data. To solve this challenge, one major line attempts to assign pseudo labels for target samples based on the pretrained model trained with labeled source samples [9, 17, 60, 66], and then fine-tunes the model using the target samples with pseudo labels. Obviously, following this approach, the person ReID performance highly depends on the quality of pseudo labels. Therefore some works focus on obtaining highly dependable pseudo labels. Some of these works concentrate on the clustering process [9, 17, 29], in which target samples are assigned with pseudo labels based on different metrics. This process can improve the clustering result. Other works aim at how to effectively utilize target samples based on the clustering results [60, 61]. During the training process of these methods, the pseudo labels with different reliability are assigned with different weights.

However, due to the clustering results are unsatisfactory, all of the existing methods suffer from noisy pseudo labels. To observe and analyze this phenomenon, we conducted the experiments to track and tag the noisy labels that happened during the training phase. Experimental results revealed that a small proportion of the samples are assigned with wrong pseudo labels frequently. These samples can be regarded as hard samples for the unsupervised domain adaptive person ReID task. As is well known, the performance of a well-trained model relies more on hard samples than easy samples. For the same reason, hard samples with stubborn wrong pseudo labels will limit the performance of the ReID model heavily. Unfortunately, existing unsupervised domain adaptive ReID methods can hardly solve this problem. Meanwhile, most of these models only apply supervised loss functions such as the Cross Entropy and Triplet loss based on the pseudo labels to train the model, but neglect unsupervised feature learning without ground-truth labels or pseudo labels. Appropriate unsupervised guidelines can relieve the absence of ground-truth labels, and it can provide robust optimization for model training, which can enhance the generalization ability of the learned feature representation.

To address these two problems, we propose a novel solution FDL-SD to resist hard samples and learn features with better generalization ability in an united framework. In order to limit the negative influence of these hard samples, we propose a simple but powerful method which is called Sample Dropout (SD) to smooth the distribution of noisy pseudo labels, it can make the model escape from the local minimum trap and enable the model to explore better solutions. For most existing clustering-based unsupervised domain adaptive ReID works, assigning pseudo labels for all target samples is employed before each fine-tuning iterator, the model has no chance to get out of the local minimum solution in such a way. But in this paper, we propose a novel clustering method, in which a proportion of target samples are randomly discarded before each training epoch, through which the vicious circle of iterative training caused by hard samples can be broken. In addition to the proposed SD, we also present a new architecture to realize Feature Diversity Learning (FDL)

in an unsupervised way. The unsupervised fashion Feature Diversity Learning we proposed can optimize the model training in a stable way without the presence of the labels, it can increase the feature diversity. Therefore the representation ability of the model is improved, and our method is believed to suppress the ill effect of wrong pseudo labels and enhance the generalization ability of the feature representation. Due to the powerful feature representation ability, our proposed Feature Diversity Learning can also be used in many other fields [5, 13, 26, 38], in which the feature plays an import role to realize the specific tasks. Overall, the main contributions of this paper can be summarized in three aspects:

(1) We propose the Sample Dropout (SD) method to reduce the adverse effect of hard samples on domain adaptation, which can prevent some hard samples from being assigned with wrong pseudo labels all the time, thus breaking the vicious circle caused by these hard samples.
(2) We propose the Feature Diversity Learning (FDL) and embed it into a dual-branch architecture to learn feature diversity representation in an unsupervised fashion, which can learn stable feature representation and boost the generalization ability of model on target domain.
(3) Extensive experiments on multiple benchmark datasets show that our proposed FDL-SD achieves the state-of-the-art performance, which demonstrates the effectiveness of our proposed approach.

2 Related work

2.1 Unsupervised domain adaptive person ReID

Unsupervised domain adaptive person ReID aims at transferring the knowledge learned from a labeled source domain to an unlabeled target domain, which can make the model produce discriminate feature representation for target domain. Existing works can be roughly divided into two categories. For the first category, it attempts to reduce the domain gap between the labeled source domain and unlabeled target domain, the models of these methods have better generalization. Some methods reduce the discrepancy between two domains by aligning the feature distribution[23, 34, 39], aligning the distribution in representation space can reduce the shift between the source domain and target domain. Some other methods adopt Generative Adversarial Network (GAN) technology to transfer the person images from source style to target style [10, 12, 50, 68], these methods alleviate the domain bias in image level. However, the identify information of the target domain is ignored in these methods, where the target samples are not fully utilized to train the model. For the second category, it assigns pseudo labels for target samples based on the pretrained model trained with labeled source domain data, and then fine-tunes the model in supervised fashion [16, 17, 32, 57, 58, 60]. These methods are widely utilized because of their superior performance. BUC [32], HCT [57] and SpCL [17] directly fine-tune the model relying on the iteration of pseudo-label mining, the target domain samples with pseudo labels are used to train the model in the following phase. Some other methods adopt mutual-training to cluster target samples and train the model, such as MMT [16], MEB-Net [58] and NRMT [60], these methods can achieve better clustering results and powerful feature representation. But the wrong pseudo labels are inevitable in these approaches, and some hard samples may even seriously damage the model.

In this paper, we adopt the second line clustering-based methods to solve the unsupervised domain adaptive person ReID task, but the difference compared with the existing works is that our work is motivated to reduce the ill effect caused by hard samples and explore feature diversity representation.

2.2 Learning with noisy labels

Some remarkable methods have been devoted to handling noisy labels, which can be categorized into loss correction, sample reweighting and label correction. Loss correction methods engage to design special loss functions against noisy labels [18, 48, 59], with the help of the designed loss functions, the noise can be reduced and the optimization of the model can get better solutions. Sample reweighting methods assign various weights to different samples [11, 19, 40, 61], the samples with noisy labels can be suppressed in these methods, therefore their ill influence for the model training is alleviated. For example, UNRN [61] proposed to re-weight samples based on the uncertainty of their pseudo labels, the reliable samples are used better and the samples with low certainty are assigned with small weights. Label correction methods [9, 24, 27, 29, 45] focus on direct correction of noisy labels. DCML [9] proposed to gradually increase the usage of pseudo labels as the training process goes on and the pseudo label reliability of these samples is boosted, which helps the model learn knowledge from easy to hard. ADTC [24] proposed to use a two-stage clustering strategy to assign pseudo labels for target samples, it uses the kmeans method to generate the centroids of clusters at first and then assign pseudo labels with the metric of k-reciprocal Jaccard distance, which can improve the clustering quality. However, all of these methods can't solve the problem that some hard samples with noisy labels hurt the training of the model seriously and are difficult to detect or correct in the iterative optimization process. Therefore, this paper adopts a novel Sample Dropout method to weaken the ill influence caused by theses hard samples.

2.3 Unsupervised feature learning

There are no ground-truth labels in some classification tasks, because it is expensive and time consuming to annotate the lables for the large amounts of data. Therefore, the unsupervised methods are adopted in many works. MMCL [49] designed a memory-based multi-label classification loss which integrates multi-label classification and single-label classification in a unified framework. Similar to MMCL, Xiao et al. [53] introduced a parameter-free Online Instance Matching loss with a memory dictionary scheme, which trains feature encoder directly instead of needing to learn a big classifier matrix and it can speed up the model convergence. In order to mitigate the effects of noisy pseudo labels, MMT [16] introduced a novel soft softmax-triplet loss to support learning with soft pseudo labels and it enhances the reliable feature learning. To solve the problem that previous contrastive losses [8, 20, 51] only focused on separating instances without considering any ground-truth classes or pseudo-class labels, SpCL [17] proposed a unified contrastive loss jointly distinguishes source domain classes, clusters and un-clustered instances of target domain, the utilization of all the data enables the model to learn better knowledge expression. However, all of these methods need to adopt pseudo labels as supervised signals, but the wrong pseudo labels are inevitable. In this paper, we propose the Feature Diversity Learning (FDL), which does not need labels as supervision that avoids the noisy labels affecting the model training. At the same time, it can boost the model's representation ability.

3 Method

3.1 Overall framework

For the unsupervised domain adaptive person ReID task, we have a labeled training dataset $D^s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_s}$ collected from the source domain, where $\mathbf{x}_i^s$ and y_i^s denote the i-th source sample and its corresponding person identity label, N_s is the number of all the source samples for training. The unlabeled target domain samples are denoted as $D^t = \{\mathbf{x}_i^t\}_{i=1}^{N_t}$. The general UDA person ReID task is "A to B", where A is the source domain dataset with annotated labels and the B is the target domain dataset without annotated labels. In the UDA task, we need to learn discriminate feature representation for target domain using the knowledge from source domain dataset. Before each iterative training epoch, the clustering algorithm DBSCAN [3] will be used to assign pseudo labels for target samples. In order to reduce the ill influence of hard samples in the target domain and boost the generalization ability of the model, we propose a novel framework that contains Sample Dropout and Feature Diversity Learning, as shown as Fig. 1.

Our model adopts the dual-branch structure which consists of feature encoders F_1 and F_2. Correspondingly, we leverage momentum update mechanism to construct and update two mean feature encoders $\tilde{F}_1$ and $\tilde{F}_2$ respectively. During the training process, two mini-batches are randomly selected from source domain and target domain, and they are fed into the two branches. As the Fig. 1 shown, the same input is passed through the mutual teaching architecture F_1, F_2, $\tilde{F}_1$ and $\tilde{F}_2$ at the same time. The feature encoders F_1 and F_2 are optimized by the backpropagation fashion through the gradients produced by our designed loss functions in this paper, and the parameters of mean feature encoders $\tilde{F}_1$ and $\tilde{F}_2$ are updated by the momentum update fashion according to the parameters of F_1 and F_2. Therefore, the mean feature encoders can smooth the negative effect caused by noisy labels, and the features produced by the mean feature encoders provide more reliable features to
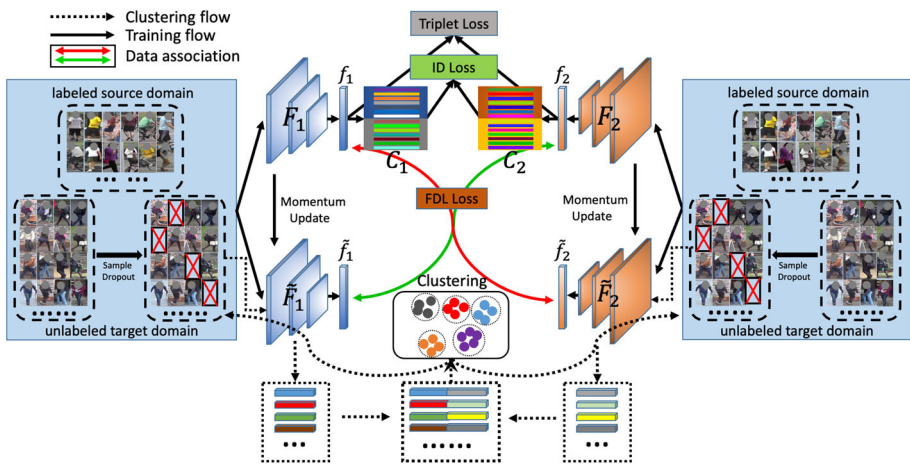

Fig. 1 An overview of the proposed architecture. Our framework adopts a dual-branch structure which is consisted of feature encoders F_1, F_2 and their corresponding mean feature encodes $\tilde{F}_1$, $\tilde{F}_2$. Classifiers C_1 and C_2 follow behind F_1 and F_2. The labeled source domain in the figure means that the labels of the source domain are available during the training phase, and the unlabeled target domain represents that there are no annotated labels for the target domain to guide the model learning

guide the model learning of F_1 and F_2. A point to note is that the proposed SD step has been applied in the target domain before each training epoch. Along each branch, the output feature vector $\mathbf{f}_i$ of the encoder F_i will be sent into the classifier C_i.

As shown in Fig. 1, the classic Cross Entropy loss is deployed based on the ID predictions given by the classifier C_i, the Triplet loss is calculated based on the feature vector $\mathbf{f}_i$, and the proposed FDL loss is embedded into the mutual teaching architecture based on the encoder's output $\mathbf{f}_i$ and mean feature encoder's output $\tilde{f}_j$, $i \neq j$. At last, the feature vectors of two mean encoders will be concatenated into a united vector $\tilde{f} = [\tilde{f}_1; \tilde{f}_2]$ to represent each sample in the testing stage, which will be also used by the clustering algorithm to produce pseudo labels.

In summary, there are three main steps during each epoch of the iterative training: 1) Sample Dropout in the target domain, 2) assigning pseudo labels for target domain samples, 3) training the model with source samples and target samples. Through the iterative optimization between feature representations and pseudo labels, the performance of the proposed method will be effectively improved.

3.2 Sample dropout

Noisy pseudo labels are harmful to the model but inevitable for unsupervised domain adaptive ReID tasks. However, we found that a small part of samples often account for a large proportion of noisy pseudo labels during the whole training process, which we define as hard samples. Compared with the general samples which are assigned with wrong pseudo labels only a few times during the whole training process, hard samples will continuously mislead the training process to wrong directions and irreversibly damage the model. In order to solve this problem, we propose Sample Dropout to smooth the distribution of noisy pseudo labels, which can reduce the adverse effect caused by hard samples. Besides, the correction on noisy pseudo labels can also make the model learn better feature representation, which can further improve the model's performance.

Sample Dropout is adopted before the clustering step of each iterative training epoch. In the beginning of k-th iterative epoch, a proportion of samples are randomly selected from the target dataset and denoted as $D_k^t = \left\{ \mathbf{x}_{r(j,k)}^t \right\}_{j=1}^{M_t}$, where $M_t = (1 - \rho)N_t$, ρ represents the Sample Dropout rate. Function $r(j, k)$ indicates the j-th random sample from the target dataset D^t in the k-th epoch. Then only the selected samples D_k^t will be assigned with pseudo labels based on their clustering results, and the residual target samples will be dropped out from the current training epoch.

Though the proposed SD seems extremely simple and naïve, it proves very powerful in dealing with noisy pseudo labels because it can prevent the model from accumulating degradation caused by hard samples. As is shown in Fig. 2, in upper row, the sample h is a hard sample which is always assigned with wrong pseudo labels. As mentioned above, such a hard sample with wrong labels can easily lead to the vicious circle of iterative training between wrong label and bad feature so that the feature space will be forced to gather samples with different person IDs while scattering the samples with the same ID, just as shown as the top row of Fig. 2. Considering that the query in a traditional ReID task is always based on the feature distance between two samples, such a distorted feature space will noticeably degrade the performance of the model. In contrast, as shown as Fig. 2(b), the employment of SD can occasionally stop these hard samples from participating in the clustering and training steps and consequently prevent the training process from falling into the local minimum trap caused by these hard samples. Our proposed SD method can reduce
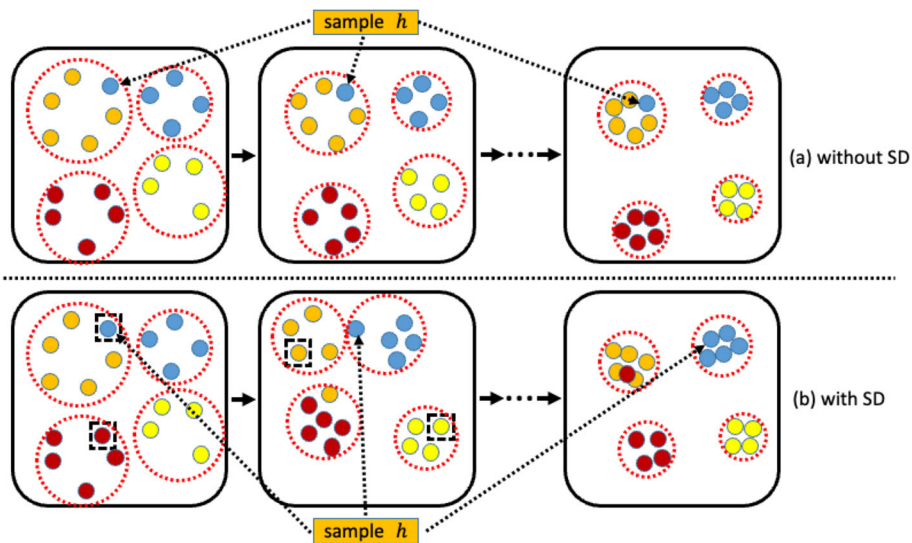
Fig. 2 Illustration of the clustering results before each fine-tuning epoch. Circles with the same color represent samples with the same person ID. Samples in the same red dotted circle belong to the same cluster. The upper row (a) represents a typical clustering process without SD, and the bottom row (b) represents the process with SD, in which the black dotted boxes mark the discarded samples in SD step

the hard noisy labels significantly, and these hard samples have the chance to be corrected in the next training epoch. Without the disturbance of hard samples, the feature space can be pulled back from the wrong direction to a better direction in the next epoch. As a result, there is larger probability to assign these hard samples with correct pseudo labels, as shown as the second plot in the bottom row of Fig. 2. By repeatedly implementing SD, the negative impact of hard samples on the whole training process can be effectively suppressed, and the final performance of the model will be dramatically improved.

3.3 Feature diversity learning

Most unsupervised domain adaptive ReID methods employ only pseudo labels to guide the training of the model. On one hand, noisy pseudo labels could hurt the model severely. On the other hand, the generalization ability of the feature representation has not been sufficiently boosted because only the pseudo-label-based losses are applied in the training stage. Therefore, we present a new approach referred as to Feature Diversity Learning (FDL), in which the model can learn better knowledge from the data distribution without the annotated labels and pseudo labels, and the generalization ability can be effectively enhanced.

Our model adopts a dual-branch structure to produce two feature streams which will be concatenated together at the final stage to form a united feature representation. The diversity between the two streams can be regarded as a kind of regularization term. It is believed to benefit the generalization ability of the feature representation. The main idea of FDL is about how to make both the two streams serve to the same ReID task while keeping their diversity to each other. The biggest challenge lies on the trade-off of the similarity and the diversity between the two feature streams. Higher similarity helps to speed up the convergence and reduce the empirical error but take higher risk of overfitting. In contrast, keeping

appropriate feature diversity generally indicates better generalization ability but makes the training difficult to converge. To get a stable and proper balance, as shown as Fig. 1, we build a mutual teaching architecture between the two streams by constructing a mean feature encoder $\tilde{F}_i$ for each stream and updating it with the momentum of the corresponding feature encoder, which can be demonstrated as Eq. 1:

$$\tilde{\boldsymbol{\theta}}_i(T) = \alpha\tilde{\boldsymbol{\theta}}_i(T-1) + (1-\alpha)\boldsymbol{\theta}_i(T) \tag{1}$$

where $\boldsymbol{\theta}_i$ and $\tilde{\boldsymbol{\theta}}_i$ are the parameters of F_i and $\tilde{F}_i$ respectively, T and $(T-1)$ represent the current statement and previous iterative statement, and α is the momentum coefficient. When the parameters of the feature encoder $F_i, i = 1, 2$ are updated through the backpropagation fashion in a mini-batch training, the parameters of mean feature encoder $\tilde{F}_i, i = 1, 2$ are updated at the same time. The error amplification can be avoided in such a way, that's because the temporally average model can produce more reliable feature representation. In addition, the FDL loss function $\mathcal{L}_{\text{FDL}}$ is put forward to guide the feature diversity learning within the proposed architecture, as shown in Eq. 2:

$$\mathcal{L}_{\text{FDL}} = S(\boldsymbol{f}_1^{\text{T}}\tilde{\boldsymbol{f}}_2) + S(\boldsymbol{f}_2^{\text{T}}\tilde{\boldsymbol{f}}_1) \tag{2}$$

where $S(x)$ is a soft-plus function.

$$S(x) = \ln(1 + \exp(x)) \tag{3}$$

It is easy to understand that the more diverse the vectors $\mathbf{f}_i$ and $\tilde{\boldsymbol{f}}_j, i \neq j$ are, the smaller the FDL loss $\mathcal{L}_{\text{FDL}}$ is. When the model is trained under such an unsupervised fashion, the model can not only learn robust feature representation but also boost the discriminate ability. With the proposed mutual teaching structure embedded with the FDL loss, stable and proper diversity between the two branches can be expected. Experimental results prove that the proposed Feature Diversity Learning is effective. We believe this is because the two branches will converge to different suboptimal solutions under the guidance of the FDL loss, which helps lower the risk of overfitting to the outlier samples or unreliable pseudo labels and therefore enhances the generalization ability of the feature representation and boost the model's recognition ability.

3.4 Overall loss

In addition to the FDL loss, we also adopt the widely used Cross Entropy (CE) loss and the Triplet loss to train the model.

The CE loss is assigned based on the predictions given by both the classifiers C_1 and C_2. For the source mini-batch B^{s} and target mini-batch B^{t}, the CE loss $\mathcal{L}_{\text{CE}}$ is defined as follows:

$$\mathcal{L}_{\text{CE}} = -[\frac{1}{|B^{\text{s}}|} \sum_{\boldsymbol{x}_i \in B^{\text{s}}} (\log(\hat{y}_{i,*}^{(1)}) + \log(\hat{y}_{i,*}^{(2)})) + \frac{1}{|B^{\text{t}}|} \sum_{\boldsymbol{x}_j \in B^{\text{t}}} (\log(\hat{y}_{j,\#}^{(1)}) + \log(\hat{y}_{j,\#}^{(2)}))] \tag{4}$$

where $\hat{y}_{i,*}^{(1)}$ and $\hat{y}_{i,*}^{(2)}$ denote the predicted probabilities of the source sample $\mathbf{x}_i^{\text{s}}$ on its true label, which are obtained by the classifiers C_1 and C_2 respectively. Similarly, $\hat{y}_{j,\#}^{(1)}$ and $\hat{y}_{j,\#}^{(2)}$ are those of the target sample $\mathbf{x}_j^{\text{t}}$ on the pseudo label.

The standard Triplet loss for a training sample $\mathbf{x}_a$ (the anchor sample) and its feature vector $F(\mathbf{x}_a)$ can be calculated according to Eq.5:

$$\mathcal{L}_{\text{t}}(\boldsymbol{x}_a; F) = [\tau + \|F(\boldsymbol{x}_a) - F(\boldsymbol{x}_p))\| - \|F(\boldsymbol{x}_a) - F(\boldsymbol{x}_n))\|]_+ \tag{5}$$

where x_p and x_n are the hardest positive and negative samples for the anchor x_a in the same mini-batch. By applying the Triplet loss $\mathcal{L}_t$ to both the source and target mini-batches in two branches, the overall Triplet loss can be written as (6). It should be noted especially that the hardest positive and negative samples of a target sample are selected according to their pseudo labels.

$$
\begin{aligned}
\mathcal{L}_{\text{TRI}} = & \frac{1}{|B^{\text{s}}|} \sum_{x_i \in B^{\text{s}}} (\mathcal{L}_t(x_i; F_1) + \mathcal{L}_t(x_i; F_2)) + \\
& \frac{1}{|B^{\text{t}}|} \sum_{x_j \in B^{\text{t}}} (\mathcal{L}_t(x_j; F_1) + \mathcal{L}_t(x_j; F_2))
\end{aligned}
\tag{6}
$$

At last, the CE loss, Triplet loss and the proposed FDL loss are combined together to train the model in an end-to-end fashion. The overall loss function can be written as:

$$
\mathcal{L} = \beta \mathcal{L}_{\text{CE}} + \gamma \mathcal{L}_{\text{TRI}} + \delta \mathcal{L}_{\text{FDL}}
\tag{7}
$$

where β, γ and δ are the coefficients of the three losses. A point to note is that none of the target samples are involved in the pretraining on the source domain. In other words, the mini-batch B^{t} can be viewed as an empty set in the pretraining stage.

According to the relative research, the CE loss helps to cluster each ID class in the global scale, and the Triplet loss does well in correcting the boundaries between neighboring classes. Coupled with the proposed FDL loss, both the empirical error and the generalization error of the model can be properly balanced through the iterative training based on the overall loss given by (7). Above all, the training algorithm of the proposed approach can be summarized as follows.

Input: Original source domain dataset D^{s}, target domain dataset D^{t}, the initialized model parameters θ_1, θ_2, $\tilde{\theta}_1$ and $\tilde{\theta}_2$ for feature encoders F_1, F_2, mean feature encoders $\tilde{F}_1$ and $\tilde{F}_2$ respectively. The Sample Dropout rate ρ. And the hyperparameters α, β, γ, δ and τ in Eq.1, Eq.7 and Eq.5.

1: Pretrain model F_1 and F_2 on the dataset D^{s} based on Eq.7, and update the parameters
　　of $\tilde{F}_1$ and $\tilde{F}_2$ according to Eq.1 at the same time.

2: For $k = 1, 2, ..., K$, **do**:

3:　　Randomly select a subset D_k^{t} from D^{t} for clustering and training according to the
　　　　Sample Dropout algorithm for k-th training epoch;

4:　　Extract the united features $\tilde{f}$ for the dataset D_k^{t} using the mean feature encoders $\tilde{F}_1$
　　　　and $\tilde{F}_2$;

5:　　Assign pseudo labels for the dataset D_k^{t} using the DBSCAN [3] clustering algorithm
　　　　based on their features $\tilde{f}$;

6:　　Train F_1 and F_2 on the source domain dataset D^{s} and the target domain subset D_k^{t}
　　　　with pseudo labels to minimize the overall loss in Eq.7, and update the parameters
　　　　of $\tilde{F}_1$ and $\tilde{F}_2$ according to Eq.1 at the same time;

7: Until the model F_1, F_2, $\tilde{F}_1$ and $\tilde{F}_2$ converged.

Return Final model F_1, F_2, $\tilde{F}_1$ and $\tilde{F}_2$ with optimal parameters.

Algorithm 1 Training process.

Person images in Market-1501   Person images in DukeMTMC-reID   Person images in MSMT17

Fig. 3 The person images in Market-1501, DukeMTMC-reID and MSMT17

4 Experiments

4.1 Datasets

The proposed approach is evaluated on three popular datasets Market-1501[63], DukeMTMC-reID[37] and MSMT17[50]. The Market-1501 dataset contains 32,668 labeled images of 1501 identities from 6 disjoint cameras, in which the training set includes 751 identities and 12936 images, the gallery set contains 19732 images from 750 identities, and the query set contains 3368 images from 750 identities. The DukeMTMC-reID dataset includes 36411 images with 1402 identities, in which 16522 images of 702 identities are for training. The gallery and query sets contain 17,661 and 2228 images of another 702 identities. MSMT17 contains 126441 images of 4101 identities, of which 1041 identities and 3060 identities are used for training and testing respectively. The person images in these datasets are shown in Fig. 3. When the image is fed into the feature encoder, through the convolution layers, activation layers and the pooling layers, the feature vector will be produced as the Fig. 4 shown.

4.2 Settings

Implementation details: The feature encoders F_1, F_2 and their mean encoders $\tilde{F}_1$ and $\tilde{F}_2$ adopt the ResNet-50 [21] as backbone. They are initialized with the parameters pretrained on the ImageNet. Compared with the original ResNet-50, the fully connected layers of these encoders are discarded. The corresponding classifiers C_1 and C_2 will be trained directly with the given datasets. In the training process, each input image is uniformly resized to 256×128. Horizontal flipping, random cropping, and random erasing are performed to generate the augmented data [25, 44]. The mini-batch size is set as $|B^s| = |B^t| = 60$, in which each identity contains 4 different samples. Hyperparameters α, β, γ, δ and τ in (1), (7) and (5) are set as 0.999, 1, 1, 0.5 and 0.3 respectively. The training is implemented by the Adam optimizer with the learning rate schedule where $\eta = 0.00035$ at beginning and is divided by
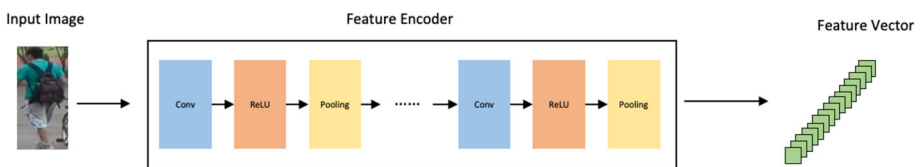

Fig. 4 The illustration for the feature extraction

10 after every 20 epochs. The whole training process is finished after 55 epochs, in which the first epoch is used to pretrain the model with only the source domain dataset and the other epochs are used to train the model with both the source and target datasets.

Evaluation metrics: In the testing process, cumulative matching characteristics at Rank-1, Rank-5, Rank-10 and mAP are applied to evaluate the performance of our method.

4.3 Comparison with the state-of-the-art methods

We compare our proposed FDL-SD method with other SOTA unsupervised domain adaptive ReID works on the Market-1501, DukeMTMC-reID and MSMT17 datasets. The comparison results are shown in Table 1.

At first, we compare the ReID performance on Market-1501 and DukeMTMC-reID. Some typical works along the approach of domain bias reduction, including PTGAN [50], SPGAN [12] and CR-GAN [10] which are GAN-based, DAAM [22], D-MMD [34] and SADA [47] which are based on feature alignment are compared with our approach. The results show that our model significantly outperforms the best of these models by about 15-23% (mAP: 71.3% vs 55.8% on the Market-to-Duke task, and 83.0% vs 59.8% on the Duke-to-Market task). Comparing the traditional clustering-based methods like SSG [14] and ECN[66], our model is still more superior in all the metrics on both the adaptation tasks. There are also a lot of latest models tackling the noisy problem such as DRDL [28], HDS [65], GLT [62] and ACMA [67] in the comparison. Even these methods have achieved better performance in comparison with the above approaches, our method can still beat them. Especially, our method outperforms ACMA [67] by 1.5%/1.0% on mAP/Rank-1 scores for the Market-to-Duke task, and also takes an advantage of 1.9% for the mAP score on the Duke-to-Market task. For the most challenging tasks of Market-to-MSMT and Duke-to-MSMT, our approach also has significant improvement, especially for the mAP score on Duke-to-MSMT task, our work outperforms the existing works by a large margin. The result shows that the university of our approach is better than other methods and we provide a solid and universal baseline for future research on UDA person ReID tasks. All of these strongly support the superiority of our method over the SOTA competing approaches.

4.4 Analysis for sample dropout rate ρ

It has been found from the experimental results that the Sample Dropout rate ρ has a significant impact on the representation ability of the model and affect the final performance of the model. To clarify this influence, a group of tests are carried out to analyse the underlying relationship between the parameter ρ, the ReID metrics and the clustering results.

The influence on performance metrics: We have uniformly sampled the values of ρ from 0 to 0.8. The metric scores under different values of ρ are shown in Tables 2 and 3. It can be observed that the performance on all the tasks has been boosted clearly as the parameter ρ increases from 0. It reaches the peak when $\rho = 0.4$ and falls back when ρ continues increasing. The results prove that the proposed SD method plays a notable role to affect the training results, and a proper value of ρ can significantly improve the performance of the model, which is consistent with our purpose.

The influence on clustering result: To deeply dig out the underlying mechanism of the proposed SD method, we design a group of experiments to observe the distributions of noisy pseudo labels for different values of ρ. According to the proposed clustering-based method, a target sample x_i^t in the k-th clustering and training process will be assigned with a noisy

Table 1 Performance (%) comparison with some popular SOTA unsupervised domain adaptive ReID methods on Market-1501, DukeMTMC-reID and MSMT17

Methods	Market to Duke				Duke to market			
	mAP	R1	R5	R10	mAP	R1	R5	R10
PTGAN [50]	–	27.4	–	50.7	–	38.6	–	66.1
SPGAN [12]	22.3	41.1	56.6	63.0	22.8	51.5	70.1	76.8
ECN [66]	40.4	63.3	75.8	80.4	43.0	75.1	87.6	91.6
CR-GAN [10]	48.6	68.9	80.2	84.7	54.0	77.7	89.7	92.7
SSG [14]	53.4	73.0	80.6	83.2	58.3	80.0	90.0	92.4
D-MMD [34]	46.0	63.5	78.8	83.9	48.8	70.6	87.0	91.5
DAAM [22]	48.8	71.3	82.4	86.3	53.1	77.8	89.9	93.7
SADA [47]	55.8	74.5	85.3	88.7	59.8	83.0	91.8	94.1
NRMT [60]	62.2	77.8	86.9	89.5	71.7	87.8	94.6	96.5
MMT [16]	65.1	78.0	88.8	92.5	71.2	87.7	94.9	96.9
SpCL [17]	68.8	82.9	90.1	92.5	76.7	90.3	96.2	97.7
RDSBN [4]	66.6	80.3	89.1	92.6	81.5	92.9	97.6	98.4
MSPGI [52]	65.8	79.2	90.1	93.4	74.4	90.1	96.4	97.8
CTFRN [64]	67.2	79.8	90.3	93.4	76.1	90.0	96.8	97.9
UNRN [61]	69.1	82.0	90.7	93.5	78.1	91.9	96.1	97.8
GLT [62]	69.2	82.0	90.2	92.8	79.5	92.2	96.5	97.8
DRDL [28]	68.6	82.6	91.5	93.6	82.2	92.7	97.2	98.2
HDS [65]	68.7	80.8	90.7	93.7	81.3	92.5	97.4	98.1
ACMA [67]	69.8	82.6	90.8	93.0	81.1	93.5	97.5	98.0
Ours	71.3	83.6	91.2	93.4	83.0	93.2	97.6	98.3
Methods	Market to MSMT				Duke to MSMT			
	mAP	R1	R5	R10	mAP	R1	R5	R10
ECN [66]	8.5	25.3	36.3	42.1	10.2	30.2	41.5	46.8
SSG [14]	13.2	31.6	–	49.6	13.3	32.2	–	51.2
DAAM [22]	20.8	44.5	–	–	21.6	46.7	–	–
NRMT [60]	19.8	43.7	56.5	62.2	20.6	45.2	57.8	63.3
RDSBN [4]	20.7	46.9	60.0	65.0	21.3	47.7	60.8	66.4
MMT [16]	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
UNRN [61]	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
ACMA [67]	25.9	53.4	66.4	72.1	26.8	54.6	70.9	73.0
CTFRN [64]	26.0	53.5	67.6	72.7	27.5	56.7	70.1	75.3
GLT [62]	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
HDS [65]	27.1	52.8	65.2	70.6	29.4	56.8	69.7	74.7
Ours	26.6	53.7	67.0	72.3	31.2	60.1	72.6	77.4

pseudo label when its true label is not equal to the dominant true label of the corresponding cluster. The more times the noisy pseudo labels are assigned in the iterative epochs for a sample, the harder the sample is. According to the results, the curves of four metrics against the increasing values of ρ are plotted in Fig. 5.

Table 2 The performance (%) of the model on Market-to-Duke and Duke-to-Market tasks when different values of ρ are applied

Value of ρ	Market to Duke				Duke to Market			
	mAP	R1	R5	R10	mAP	R1	R5	R10
0	65.3	79.5	89.5	92.5	74.1	89.8	95.5	97.1
0.1	67.4	80.7	90.0	92.8	77.4	90.5	96.1	97.6
0.2	69.4	83.3	91.2	93.2	81.1	92.7	96.8	98.0
0.3	70.8	83.4	91.7	94.1	80.4	92.1	96.9	97.7
0.4	71.3	83.6	91.2	93.4	83.0	93.2	97.6	98.3
0.5	71.0	83.1	91.2	93.9	81.1	92.3	97.3	98.4
0.6	70.0	83.0	90.8	92.8	79.7	91.7	97.2	98.1
0.7	65.9	79.7	88.9	91.7	75.1	88.8	95.3	97.1
0.8	52.8	70.4	81.2	84.6	47.3	71.3	82.8	86.4

The clustering error rate is the ratio of noisy labels to all the labels when we implement the clustering algorithm on the fully converged model. The corresponding pink curves in Fig. 5(a), (b), (c) and (d) illustrate that the clustering error rate increases with the growing value of ρ. This result coincides with our intuition because higher SD rate indicates that more samples are not fully utilized in the training. The blue and green curves in Fig. 5 record the relative error rates of the 10% and 20% hardest samples against the parameter ρ, respectively. By 10% or 20% hardest, we mean the top 10% or 20% samples that are most frequently assigned with noisy pseudo labels. The corresponding relative error rates are defined as the ratio of the noisy labels ever assigned to these hardest samples to all the noisy labels ever generated in the whole training process. It can be easily discovered that the relative error rates of the 10% or 20% hardest samples fall steadily as the value of ρ grows from 0 to 0.8. This result reveals a truth that the proposed SD method can effectively prevent noisy pseudo labels from concentrating on a minority of target samples, namely the

Table 3 The performance (%) of the model on Market-to-MSMT and Duke-to-MSMT tasks when different values of ρ are applied

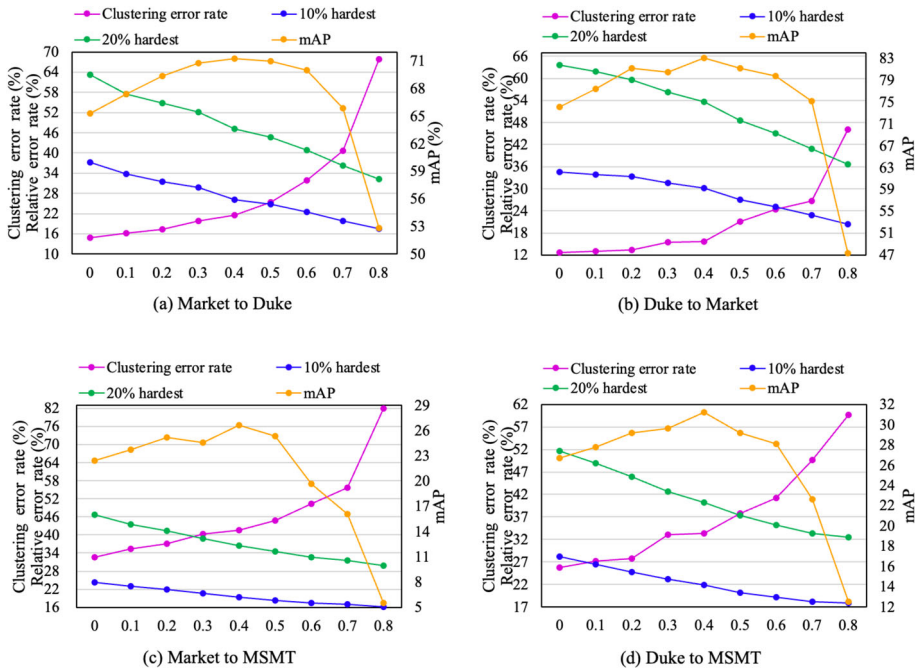| Value of ρ	Market to MSMT				Duke to MSMT			
	mAP	R1	R5	R10	mAP	R1	R5	R10
0	22.4	49.0	61.7	67.5	26.7	55.5	68.4	73.8
0.1	23.7	50.9	63.6	69.3	27.8	56.7	69.3	74.6
0.2	25.2	52.2	65.3	79.8	29.2	57.8	70.7	75.7
0.3	24.1	50.9	64.1	69.5	29.6	58.6	71.4	76.2
0.4	26.6	53.7	67.0	72.3	31.2	60.1	72.6	77.4
0.5	25.3	51.6	65.0	70.3	29.2	57.7	70.4	74.9
0.6	19.7	42.7	55.1	60.8	28.1	55.7	68.4	73.8
0.7	16.1	36.3	47.8	53.4	22.6	46.9	60.3	65.7
0.8	5.5	13.6	21.5	25.4	12.5	30.0	40.8	45.9

Fig. 5 Curves of clustering error rate, relative error rate of the 10% and 20% hardest samples and the mAP plotted along with growing values of the parameter ρ. Similar to (a), the left vertical axes of (b), (c) and (d) represent the clustering error rate and relative error rate, the right axes indicate the mAP scores

hardest samples we mentioned above. Our method can alleviate the damage for the model caused by these hard samples.

Considering the synthetic influence of the hyperparameter ρ on both aspects of the clustering error rate of noisy labels and the relative error rate of the hardest samples, a clear conclusion can be drawn that there must be a most proper value of ρ that can keep the optimal balance between the two factors. The mAP curve in Fig. 5 also provides solid evidence to support this point, in which the mAP score reaches the peak at $\rho = 0.4$ for the four tasks. All above experimental results and the corresponding analysis prove that the proposed SD method can significantly suppress the ill influence caused by hard samples and consequently improve the reliability of the model.

4.5 Ablation study on FDL

To validate the effectiveness of our proposed Feature Diversity Learning, a group of ablation tests are designed to show the mAP and Rank-1 scores of the model when the FDL is or is not applied. As shown as Fig. 6, against different values of ρ in the horizontal axes, the red and green bars represent the mAP scores with or without the FDL, and the blue and pink bars reflect the Rank-1 scores when the FDL is or is not applied, respectively. It can be clearly seen that the FDL achieves 2.8% and 1.2% improvements on mAP and Rank-1 scores in dealing with the Market-to-Duke task when $\rho = 0.4$, and it also shows 2.7% and 1.4% superiority of mAP and Rank-1 scores on the Duke-to-Market task. This advantage can also be validated on both of the Market-to-MSMT and Duke-to-MSMT tasks, the performance
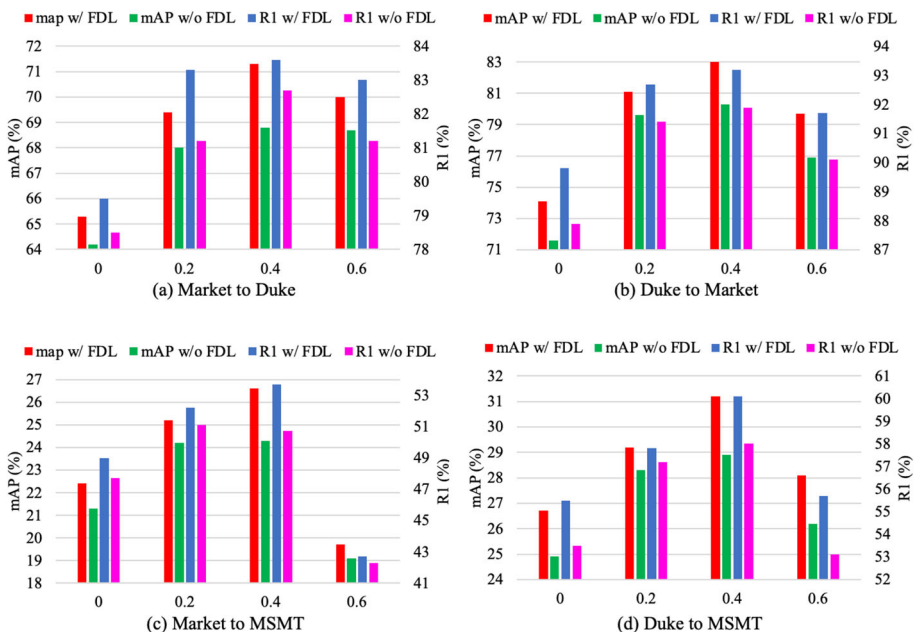
Fig. 6 The mAP and Rank-1 scores for ablation tests under different values of ρ on four tasks, where "w/" means the FDL is applied to the model while "w/o" means the opposite

has significant improvement when the FDL loss is adopted. The results under different values of ρ on all the tasks indicate the same conclusion that the proposed FDL certainly improves the generalization ability of the model on the unsupervised domain adaptive ReID task, which supports our viewpoint that the diversity between the two feature streams helps to prevent the model from falling into the overfitting trap.

4.6 The application of our method

Person ReID task has important application value in real-world, it can maintain public order and serve the society. We have implemented the intelligent video system, the system can recognize and query the target person. During the training phase, due to the absence of the annotated labels of the persons in the target scene, we have to use the UDA person ReID technology to implement the specific functions. Just as we have mentioned above, there are some problems that the existing works can not solve, but our proposed method can meet the demands. Therefore, our method has important real-world application value.

5 Conclusion

Noisy pseudo label is one of the most challenging problem for clustering-based unsupervised domain adaptive ReID models, which is the main problem we have addressed in this paper. With our proposed Sample Dropout method, the hard samples can be effectively suppressed and consequently the uneven distribution of noisy labels can be smoothed, which is proved helpful for breaking the vicious circle between noisy label and bad feature and

improving the training results. The proposed Feature Diversity Learning provides a new approach for the well-known mutual teaching architecture, which provides an unsupervised fashion to make the model learn knowledge from the data distribution and focuses on enhancing the diversity of the two feature streams. Ablation study shows that FDL has a stable positive impact on the generalization ability of the model. With the above two improvements, it is proved by the comparison results that our proposed FDL-SD outperforms most state-of-the-art methods on the unsupervised domain adaptive ReID task. And our method has been applied in the real-world scene, it has significant value in video system.

In spite of the advantages of performance, further room around the proposed SD and FDL methods needs much more exploration, including the intrinsic mechanism about the dropout rate ρ, and the underlying relationship between momentum-based mean feature vector and the FDL scheme. We will remain focusing on these problems in our future work.

Declarations

Ethics approval and consent to participate Compliance with ethical standards.

Consent for Publication The authors declare that they consent to publication.

Conflict of Interests The authors declare that they have no conflict of interests.

References

1. Aggarwal AK, Jaidka P (2022) Segmentation of crop images for crop yield prediction. Int J Biol Biomed 7
2. Ahuja U, Singh S, Kumar M, Kumar K, Sachdeva M (2022) Covid-19: social distancing monitoring using faster-rcnn and yolov3 algorithms. Multimed Tools Appl 82(5):7553–7566
3. Bäcklund H, Hedblom A, Neijman N (2011) A density-based spatial clustering of application with noise. Data Mining TNM033 :11–30
4. Bai Z, Wang Z, Wang J, Hu D, Ding E (2021) Unsupervised multi-source domain adaptation for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12914–12923
5. Bansal M, Kumar M (2021) Kumar, M: 2d object recognition: a comparative analysis of sift, surf and orb feature descriptors. Multimed Tools Appl 80:18839–18857
6. Bansal M, Kumar M, Sachdeva M, Mittal A (2021) Transfer learning for image classification using vgg19: Caltech-101 image data set. J Amb Intell Hum Comput :1–12
7. Behera N, Sa PK, Bakshi S (2020) Person re-identification for smart cities: state-of-the-art and the path ahead. Pattern Recognition Letters **138**(5):
8. Chen T, Kornblith S, Norouzi M, Hinton G (2020) A simple framework for contrastive learning of visual representations. In: International conference on machine learning, PMLR, pp 1597–1607
9. Chen G, Lu Y, Lu J, Zhou J (2020) Deep credible metric learning for unsupervised domain adaptation person re-identification. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part VIII 16, Springer, pp 643–659
10. Chen Y, Zhu X, Gong S (2019) Instance-guided context rendering for cross-domain person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 232–242
11. Cheng H, Zhu Z, Li X, Gong Y, Sun X, Liu Y (2020) Learning with instance-dependent label noise: a sample sieve approach. arXiv:2010.02347
12. Deng W, Zheng L, Ye Q, Kang G, Yang Y, Jiao J (2018) Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 994–1003

13. Fekri-Ershad S (2021) Cell phenotype classification using multi threshold uniform local ternary patterns in fluorescence microscope images. Multimed Tools Appl 80(8):12103–12116
14. Fu Y, Wei Y, Wang G, Zhou Y, Shi H, Huang TS (2019) Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 6112–6121
15. Garg M, Ubhi JS, Aggarwal AK (2022) Neural style transfer for image steganography and destylization with supervised image to image translation. Multimed Tools Appl :1–18
16. Ge Y, Chen D, Li H (2020) Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. In: International conference on learning representations
17. Ge Y, Zhu F, Chen D, Zhao R, Li H (2020) Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: Advances in neural information processing systems
18. Ghosh A, Kumar H, Sastry P (2017) Robust loss functions under label noise for deep neural networks. In: Proceedings of the AAAI conference on artificial intelligence, vol 31
19. Han B, Yao Q, Yu X, Niu G, Xu M, Hu W, Tsang I, Sugiyama M (2018) Co-teaching: robust training of deep neural networks with extremely noisy labels. arXiv:1804.06872
20. He K, Fan H, Wu Y, Xie S, Girshick R (2020) Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9729–9738
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
22. Huang Y, Peng P, Li Y, Jin Y, Xing J (2020) Ge, s., Domain adaptive attention model for unsupervised cross-domain person re-identification. AAAI
23. Huang H, Yang W, Chen X, Zhao X, Huang K, Lin J, Huang G, Du D (2018)
24. Ji Z, Zou X, Lin X, Liu X, Huang T, Wu S (2020) An attention-driven two-stage clustering method for unsupervised person re-identification. In: Computer Vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part XXVIII 16, Springer, pp 20–36
25. Kaur A, Chauhan AS, Kumar Aggarwal A (2022) Prediction of enhancers in dna sequence data using a hybrid cnn-dlstm model. IEEE/ACM Transactions on Computational Biology and Bioinformatics
26. Kumar A, Kumar M, Kaur A (2021) Face detection in still images under occlusion and non-uniform illumination. Multimed Tools Appl 80:14565–14590
27. Lee K-H, He X, Zhang L, Yang L (2018) Cleannet: transfer learning for scalable image classifier training with label noise. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5447–5456
28. Li H, Xu K, Li J, Yu Z (2022) Dual-stream reciprocal disentanglement learning for domain adaptation person re-identification. Knowl-Based Syst 251:109315
29. Li J, Zhang S (2020) Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In: European conference on computer vision, Springer, pp 483–499
30. Li J, Zhang S, Tian Q, Wang M, Gao W (2019) Pose-guided representation learning for person re-identification IEEE transactions on pattern analysis and machine intelligence
31. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 152–159
32. Lin Y, Dong X, Zheng L, Yan Y, Yang Y (2019) A bottom-up clustering approach to unsupervised person re-identification. In: Proceedings of the AAAI conference on artificial intelligence, vol. 33, pp 8738–8745
33. Liu J, Zha Z-J, Chen D, Hong R, Wang M (2019) Adaptive transfer network for cross-domain person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7202–7211
34. Mekhazni D, Bhuiyan A, Ekladious G, Granger E (2020) Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: European conference on computer vision, Springer, pp 159–174
35. Ning X, Gong K, Li W, Zhang L (2021) Jwsaa: joint weak saliency and attention aware for person re-identification. Neurocomputing 453:801–811
36. Ning X, Gong K, Li W, Zhang L, Bai X, Tian S (2020) Feature refinement and filter network for person re-identification. IEEE Trans Circ Syst Video Technol 31(9):3391–3402
37. Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C (2016) Performance measures and a data set for multi-target, multi-camera tracking. In: European conference on computer vision, Springer, pp 17–35
38. Shaheed K, Mao A, Qureshi I, Kumar M, Hussain S, Ullah I, Zhang X (2022) Ds-cnn: a pre-trained xception model based on depth-wise separable convolutional neural network for finger vein recognition. Expert Syst Appl 191:116288

39. Shan L, Li H, Li CT, Kot AC (2018) Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. In: BMVC 2018

40. Shu J, Xie Q, Yi L, Zhao Q, Zhou S, Xu Z, Meng D (2019) Meta-weight-net: learning an explicit mapping for sample weighting. arXiv:1902.07379

41. Su C, Li J, Zhang S, Xing J, Gao W, Tian Q (2017) Pose-driven deep convolutional model for person re-identification. In: Proceedings of the IEEE international conference on computer vision, pp 3960–3969

42. Sun Y, Zheng L, Yang Y, Tian Q, Wang S (2018) Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) 480–496

43. Tay C-P, Roy S, Yap K-H (2019) Aanet: attribute attention network for person re-identifications. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7134–7143

44. Ubhi JS, Aggarwal AK et al (2022) Neural style transfer for image within images and conditional gans for destylization. J Vis Commun Image Represent 85:103483

45. Vahdat A (2017) Toward robustness against label noise in training deep discriminative neural networks. arXiv:1706.00038

46. Walia S, Kumar K, Kumar M (2023) Unveiling digital image forgeries using markov based quaternions in frequency domain and fusion of machine learning algorithms. Multimed Tools Appl 82(3):4517–4532

47. Wang G, Lai J-H, Liang W, Wang G (2020) Smoothing adversarial domain attack and p-memory reconsolidation for cross-domain person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10568–10577

48. Wang Y, Ma X, Chen Z, Luo Y, Yi J, Bailey J (2019) Symmetric cross entropy for robust learning with noisy labels. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 322–330

49. Wang D, Zhang S (2020) Unsupervised person re-identification via multi-label classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10981–10990

50. Wei L, Zhang S, Gao W, Tian Q (2018) Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 79–88

51. Wu Z, Xiong Y, Yu SX, Lin D (2018) Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3733–3742

52. Xi J, Zhou Q, Li X, Zheng S (2022) Momentum source-proxy guided initialization for unsupervised domain adaptive person re-identification. Neurocomputing 483:116–126

53. Xiao T, Li S, Wang B, Lin L, Wang X (2017) Joint detection and identification feature learning for person search. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3415–3424

54. Yan C, Pang G, Bai X, Liu C, Ning X, Gu L, Zhou J (2021) Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss. IEEE Trans Multimedia 24:1665–1677

55. Ye M, Shen J, Lin G, Xiang T, Shao L, Hoi SC (2021) Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell 44(6):2872–2893

56. Yu H-X, Zheng W-S, Wu A, Guo X, Gong S, Lai J-H (2019) Unsupervised person re-identification by soft multilabel learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2148–2157

57. Zeng K, Ning M, Wang Y, Guo Y (2020) Hierarchical clustering with hard-batch triplet loss for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 13657–13665

58. Zhai Y, Ye Q, Lu S, Jia M, Ji R, Tian Y (2020) Multiple expert brainstorming for domain adaptive person re-identification. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part VII 16, Springer, pp 594–611

59. Zhang Z, Sabuncu MR (2018) Generalized cross entropy loss for training deep neural networks with noisy labels. In: 32Nd conference on neural information processing systems (neurIPS)

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

61. Zheng K, Lan C, Zeng W, Zhang Z, Zha Z-J (2021) Exploiting sample uncertainty for domain adaptive person re-identification. In: AAAI 2021, AAAI

62. Zheng K, Liu W, He L, Mei T, Luo J, Zha Z-J (2021) Group-aware label transfer for domain adaptive person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5310–5319

63. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: Proceedings of the IEEE international conference on computer vision, pp 1116–1124

64. Zheng D, Xiao J, Chen K, Huang X, Chen L, Zhao Y (2022) Soft pseudo-label shrinkage for unsupervised domain adaptive person re-identification. Pattern Recogn 127:108615
65. Zheng D, Xiao J, Wei Y, Wang Q, Huang K, Zhao Y (2022) Unsupervised domain adaptation in homogeneous distance space for person re-identification. Pattern Recogn 132:108941
66. Zhong Z, Zheng L, Luo Z, Li S, Yang Y (2019) Invariance matters: exemplar memory for domain adaptive person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 598–607
67. Zhou H, Kong J, Jiang M, Liu T (2022) Heterogeneous dual network with feature consistency for domain adaptation person re-identification. International Journal of Machine Learning and Cybernetics :1–15
68. Zou Y, Yang X, Yu Z, Kumar BV, Kautz J (2020) Joint disentangling and adaptation for cross-domain person re-identification. In: Computer Vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part II 16, Springer, pp 87–104