# Deep Credible Metric Learning for Unsupervised Domain Adaptation Person Re-identification

Guangyi Chen[1,2,3], Yuhao Lu[1,5], Jiwen Lu[1,2,3] ⋆, and Jie Zhou[1,2,3,4]

[1] Department of Automation, Tsinghua University, China
[2] State Key Lab of Intelligent Technologies and Systems, China
[3] Beijing National Research Center for Information Science and Technology, China
[4] Tsinghua Shenzhen International Graduate School, Tsinghua University, China
[5] School of Computer Science, Beijing University of Posts and Telecommunications, China
chen-gy16@mails.tsinghua.edu.cn, luyuhao998@gmail.com,
{lujiwen,jzhou}@tsinghua.edu.cn

**Abstract.** The trained person re-identification systems fundamentally need to be deployed on different target environments. Learning the cross-domain model has great potential for the scalability of real-world applications. In this paper, we propose a deep credible metric learning (DCML) method for unsupervised domain adaptation person re-identification. Unlike existing methods that directly finetune the model in the target domain with pseudo labels generated by the source pre-trained model, our DCML method adaptively mines credible samples for training to avoid the misleading from noise labels. Specifically, we design two credibility metrics for sample mining including the k-Nearest Neighbor similarity for density evaluation and the prototype similarity for centrality evaluation. As the increasing of the pseudo label credibility, we progressively adjust the sampling strategy in the training process. In addition, we propose an instance margin spreading loss to further increase instance-wise discrimination. Experimental results demonstrate that our DCML method explores credible and valuable training data and improves the performance of unsupervised domain adaptation.

**Keywords:** Credible learning, Metric learning, Unsupervised domain adaptation, Person re-identification

## 1 Introduction

Person re-identification (ReID) aims at identifying a query individual from a large set of candidates under the non-overlapping camera views. As an essential role in various applications of security and surveillance, lots of attempts and dramatic improvements have been witnessed in recent years [22, 23, 37, 48, 58].

Despite the satisfactory performance obtained by the supervised deep learning model and some label annotations in the single domain, it is still a challenge

⋆ Corresponding author

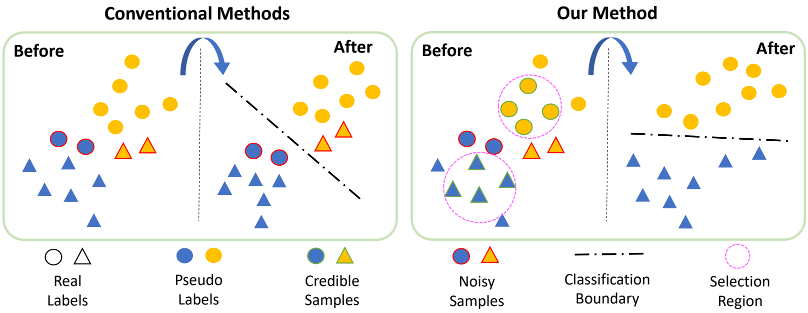

**Fig. 1.** Difference between our DCML method and conventional methods. The left part shows that conventional metric learning methods treat all samples equally to train the model and thus are easy to be misled by the noise labels. The right part shows that our method adaptively mines credible samples to train the model, which can avoid the damage from these low-quality samples. Best viewed in color.

to deploy the trained person ReID models on different target environments. It is due to the domain bias between the training and deploying environments, e.g., the model trained on one university dataset need to be applied for airport or underground station. One of the common methods is finetuning the deep model by the image data of the target domain and pseudo labels generated by the source pre-trained model (e.g., clustering [12, 34, 52], reference comparison [51], or nearest neighborhood [61] ). However, the predicted pseudo labels might involve much noise, which misleads the training process in the target domain. As shown in Fig. 1, the noisy labels might generate opposite gradients which undermine the model discrimination.

To address this problem, we propose a deep credible metric learning (DCML) method to avoid the damage from noise pseudo labels by adaptively exploring credible and valuable training samples. Specifically, our DCML method consists of two parts, including adaptively credible anchor sample mining and instance margin spreading. The former is proposed to explore credible samples, which are effective for learning the intra-class compact embeddings. We propose two credibility metrics including the k-Nearest Neighbor similarity and the prototype similarity. We implement two different similarity metrics to demonstrate the generality of the credible anchor sample mining strategy. The k-Nearest Neighbor similarity measures the neighborhood density of the sample by calculating the maximum distance (minimum similarity) between itself and k nearest neighbors. While the prototype similarity calculates the similarity between the sample and class prototype, which denotes the sample's centrality. Using these credibility metrics, we can select samples with higher credibility as anchors. As the training iterations increasing, the credibility of pseudo labels continues to increase too. We therefore, progressively reduce the limitation of anchor sample mining to select more credible training samples. In addition, we propose an instance margin spreading (IMS) loss to increase the instance-wise discrimination, due to the initial embeddings of target samples are always confusing and in-discriminative without supervised training. We regard each sample as an independent individual and learn a spreading embedding apace by pushing the samples away from each other by a large margin. We summarize the contributions of this work as follows:

1) We propose a deep credible metric learning (DCML) method for unsupervised domain adaptation person ReID, which adaptively and progressively mines credible and valuable training samples to avoid the damage from the noise of predicted pseudo labels.
2) We design an instance margin spreading method loss to encourage the instance-wise discrimination by spreading the embeddings of samples with a large margin.
3) We conduct extensive experiments to demonstrate the superiority of our method, and achieve the state-of-the-art performance on several large scale datasets including Market-1501 [57], DukeMTMC-reID [30], and CUHK03 [21].

## 2 Related Work

**Supervised Deep Person ReID:** Most existing person ReID methods obtain excellent performance by the supervised deep learning model and a number of label annotations. Some methods are devoted to designing more effective networks by part-based model [3,6,36,37,41] or attention model [1,2,11,22,31,47]. Other methods focus on capture more prior knowledge or supervisory signals, including body structure [18,19,53,54], human pose [29,35], attribute labels [39,55], and other loss functions [4,15,56]. Despite the recent progress in the supervised manner, the deployment of trained models for different target environments is still a challenge due to the large domain bias.

**Unsupervised Domain Adaptation Person ReID:** To address the above problem, Some works [24,49] study purely unsupervised learning to learn from unlabelled data for Re-ID. However, the performance is limited without any labeled data. Furthermore, many works attempt to learn the unsupervised domain adaptation person ReID model, which leverages the labeled source domain data and unlabeled target domain data. Many existing works [5,7,44] apply the generative model (e.g., GAN) to transform the images of source domain into the target domain as the training data, aiming to reduce the domain bias from data. While other works finetune the deep model with the target domain data and pseudo labels generated by the source pre-trained model. The clustering methods [12,34,52] and reference comparison [51] are widely used to generate the supervisory signal from pre-trained models. Besides, some unsupervised domain adaptation person ReID methods explore other human prior knowledge or auxiliary supervisory signals to improve the adaptation and generalization ability from the source domain to the target domain. EANet [16] employs the human parsing results to assist feature alignment. While TJ-AIDL attempts to learn a joint attribute-identity space which improves the model generalization ability with transferred attribute knowledge. Our work is related to PAST, which randomly selects the positive and negative samples from top k neighbors and k-2k neighbors respectively with all samples as the anchors and employs a cross-entropy loss as the promoting stage. However, PAST applies the fixed sampling strategy for all anchors in the whole training process which ignores the initial low-quality and continuous improvement of pseudo labels. Our DCML method adaptively selects credible anchors by measuring the credibility of each sample and progressively adjusts the sampling strategy for the different stages of the training process.

**Deep Metric Learning:** Deep metric learning aims to learn the discriminative feature embedding space instead of the final classifier, which generalizes better to the unseen environment [4]. Existing deep metric learning methods mainly focus on design effective loss functions or develop efficient sampling strategies. The loss designing methods focus on utilizing higher order relationships [26, 40, 42], global information [27, 33], or the margin maximum [8, 38, 50]. While sampling-based methods are devoted to mining the hard negative samples for training efficiency improvement. For instance, TriNet [15] samples the most negative samples in the batch for fast convergence. Harwood et al. [13] found the negative samples from an increasing search space defined by the nearest neighbor distance. However, these mining strategies tend to select the harder samples due to the larger gradient from violating triplet relation defined by the annotations, which is confused with the noise labels, especially for pseudo labels. To address this issue, we adaptively and progressively select the credible anchor samples, which is appropriate for the low-quality predicted pseudo labels.

## 3 Deep Credible Metric Learning

The goal of our deep credible metric learning method is adaptively and progressively discovering the credible samples to reduce the damage from noise labels. In this section, we will introduce our DCML method from two parts, including adaptively credible sample mining and instance margin spreading.

### 3.1 Problem Formulation

For the unsupervised domain adaptation person ReID problem, we have a source dataset $\mathcal{S} = \{\mathcal{X}^{\mathcal{S}}, \mathcal{Y}^{\mathcal{S}}\}$, where $\mathcal{X}^{\mathcal{S}}$ denotes the image data and $\mathcal{Y}^{\mathcal{S}}$ is the corresponding labels. Besides, we have another dataset in the deployed environment without any annotations, which is called target dataset $\mathcal{X}^{\mathcal{T}} = \{x_i^t\}_1^N$. The cross-domain person ReID system aims to learn the robust and generalizable representations in the target domain with the supervised source dataset and unsupervised target one. A popular solution for the unsupervised domain adaptation person ReID problem is finetuning the pre-trained model in the target domain with the predicted pseudo labels. Support we have predicted pseudo labels $\hat{\mathcal{Y}}^{\mathcal{T}} = \mathcal{P}(\mathcal{X}^{\mathcal{T}}; \mathcal{X}^{\mathcal{S}}, \mathcal{Y}^{\mathcal{S}})$ generated by the pre-trained model from the source domain, we learn feature embeddings with a convolutional neural network (CNN) $\mathcal{F}_\theta$ as $f_i = \mathcal{F}_\theta(x_i^t)$ with the objective function which is formulated as:

$$\theta = \arg\min_{\theta} \mathcal{L}(\theta; \mathcal{X}^{\mathcal{T}}, \hat{\mathcal{Y}}^{\mathcal{T}}), \tag{1}$$

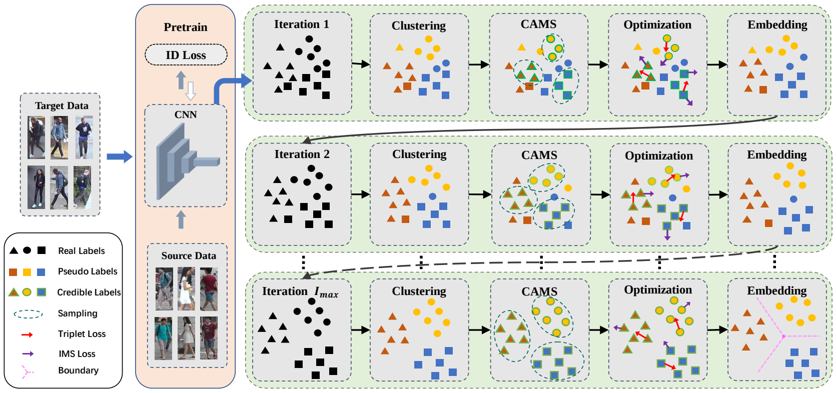

**Fig. 2.** Illustration of the deep credible metric learning method. The DCML method starts with learning a pre-trained CNN network with the source labeled data. In each iteration, we extract the embeddings of unlabeled target images and generate pseudo labels with the clustering method. To avoid the misleading of noise pseudo labels, we adaptively mine credible samples as the anchor data and optimize the model with these samples. The gradients come from two objective functions including the triplet loss with red arrows and the IMS loss with purple arrows. In addition, we progressively adjust the anchor sample mining strategy to select more anchor samples as iteration increases. Best viewed in color

where the objective is to learn CNN $\mathcal{F}_\theta$ by using pseudo labels as a supervisory signal. However, the performance of this objective function entirely depends on the properties of generated labels without a stable guarantee. The generated labels are always noisy due to the large domain bias between the source and target datasets. These noise labels always mislead the training process by providing wrong gradients. This inevitably leads to the necessity of adaptively credible samples mining for more reliable model learning.

### 3.2 Adaptively Credible Sample Mining

The adaptively credible sample mining strategy aims to select the more credible samples to avoid the damage from noise labels. For one target sample and corresponding pseudo label $(x_i^t, \hat{y}_i^t)$, we define a credibility metric $\mathcal{C}(x_i^t, \hat{y}_i^t)$ to evaluate whether a label is credible enough as a supervisory signal. Given a threshold $\tau$, we select the more credible samples as the training data:

$$\mathcal{X}_C^{\mathcal{T}} = \{x_i^t \in \mathcal{X}^{\mathcal{T}} | \mathcal{C}(x_i^t, \hat{y}_i^t) > \tau\}, \tag{2}$$

where $\mathcal{X}_C^{\mathcal{T}}$ denotes selected credible dataset in which each sample is credible as an anchor sample to train the model. In the following subsections, we will introduce that the threshold $\tau$ is adaptive with the learning process, which reduces the threshold when the pseudo labels are more credible. The main problem is how to evaluate the credibility of samples. The basic assumption of our anchor sample mining strategy is that the central and dense samples are credible for training. Thus we design two credibility metrics including the k-Nearest Neighbor distance and the prototype distance to measure the neighborhood density and class centrality of samples.

**Prototype Similarity:** In the prototype similarity, we define the credibility of one sample with the similarity between it and the class prototype. Inspired by the prototypical network [32], we assume all support data points of the same "class" lie in a manifold, and calculate the class prototype as the center of class:

$$\mathcal{P}_k = \frac{1}{|\mathcal{M}_k|} \sum_{x_i^t \in \mathcal{M}_k} \mathcal{F}_\theta(x_i^t), \tag{3}$$

where $\mathcal{M}_k = \{x_i^t \in \mathcal{X}^\mathcal{T} | \hat{y}_i^t = k\}$ denotes the set of examples labeled with class $k$, and $\hat{y}_i^t$ is the pseudo label of $x_i^t$. Then the intra-class centrality can be calculated with the Euclidean distance as:

$$\mathcal{C}_P(x^t, \hat{y}^t) = -||x^t - \mathcal{P}_{\hat{y}^t}||_2. \tag{4}$$

The larger $\mathcal{C}_P(x^t, \hat{y}^t)$ values correspond to more intra-class consistent samples. When the intra-class centrality $\mathcal{C}_P(x^t)$ is large, the sample $x^t$ is close to the class prototype, which means that its representation as a class is trustworthy. On the contrary, the samples with small credibility values might be mislabeled since these samples are always close to the uncredited classification-plane.

**KNN Similarity:** Different from prototype similarity measuring the intra-class sample centrality, the KNN similarity calculates the local density by the neighborhood information. For a sample $x^t$, the neighborhood set $\mathcal{N}(x^t)$ consists of k samples whose distance is nearest with the $x^t$. The neighborhood set denotes the local neighborhood information of samples, which can be employed to describe the density. We define the KNN distance as

$$\mathcal{C}_N(x^t) = - \max_{x_i^t \in \mathcal{N}(x^t)} d(x^t, x_i^t), \tag{5}$$

where $d(\cdot, \cdot)$ is a distance metric, e.g., the Euclidean distance. We employ the minimal similarity among the k nearest neighborhoods to denote the local density. All the samples in the neighborhood set $\mathcal{N}(x^t)$ are more compact as KNN similarity $\mathcal{C}_N(x^t)$ is large, which denotes that the $x^t$ resides in a high-density region. When the samples are dense in the neighborhood set and far away from other samples, the neighborhood-based pseudo label generation method, e.g., clustering, will give a more reliable result. When the samples are dense and indistinguishable, they are also necessary to pay more attention. Thus, we select the samples with higher KNN similarity as training data.

**Progressively Learning:** In the whole training stage, we iteratively generate the pseudo labels with the embedding model and train the embedding model with pseudo labels. In each iteration, we first extract the embeddings with current model $\mathcal{F}_\theta$ and cluster on the embedding space to generate the pseudo labels. Then, we apply the pseudo labels as supervisory signal to train and update the embedding model. Though this iterative learning process, the pseudo labels become more and more credible and embeddings become more and more discriminative. In our DCML method, we progressively adjust the anchor sample mining strategy to select more anchor samples by reducing the selection threshold as iteration increases, since the pseudo labels are more credible as the model is finetuned. When the pseudo labels are credible enough, we tend to employ all the data in the target domain to train our model. Specifically, we design a linear threshold adaptation strategy, which progressively reduce the threshold $\tau$ with the iterations $r$. We formulate the threshold adaptation strategy with iterations $r$ as follows:

$$\tau = \arg\min_{\tau} |\mathcal{X}_c^{\mathcal{T}}| \geq (\gamma_0 + r \times \mathit{\Delta}\gamma)|\mathcal{X}^{\mathcal{T}}| \tag{6}$$

where $|\mathcal{X}_c^{\mathcal{T}}|$ and $|\mathcal{X}^{\mathcal{T}}|$ respectively denote the number of samples in the selected and original datasets. $\gamma_0$ and $\mathit{\Delta}\gamma$ are the hyperparameters of algorithm which respectively denote the initial sampling rate of anchor samples and the increment in each iteration. The basic goal of this strategy is adapting an appropriate threshold $\tau$ to select sufficient credible anchor samples. The number of selected samples progressively increases with the assuming that the credibility of pseudo labels increase as training iterations.

### 3.3 Instance Margin Spreading

The pre-trained embeddings on the target domain are always confusing and in-discriminative. It is difficult to cluster these in-discriminative samples and generate credible pseudo labels. In order to increase the inter-class discrimination, we propose an instance margin spreading (IMS) loss which spreads the embeddings by pushing the samples a large margin apart from each other for a discriminative embeddings space. Inspirited by the instance discrimination learning [46] which assumes each instance is a independent class, we aim to learn a spreading metric space where the distances between each instance pair are over a large margin. Different from conventional margin-based losses (e.g., triplet loss), our IMS loss doesn't require any labels, which learns the embedding space only by the instance-wise discrimination. The basic formulation of this margin constraint is as follows:

$$\mathcal{L}_{ims}(x_a^t) = \sum_{i \neq a} \max\big(0, m - d_{a,i}\big) \tag{7}$$

where $x_a$ denotes the random selected sample, $d_{a,i}$ denotes the distance between the sample pair $d(x_a^t, x_i^t)$ and $i \neq a$ represents all other samples in the dataset except itself. The $m$ is a margin which denotes the lower bound of distances between each sample pair. As shown in [4] and [33], we can obtain the equivalent loss function by replacing the $\max(0, x)$ with a continuous exponential function

**Algorithm 1 :** DCML

```
Require: Source dataset $\mathcal{S}$; target dataset $\mathcal{T}$; maximal iterative number $R_{max}$.
Ensure: The parameters $\theta$ of embedding network $\mathcal{F}_\theta$.
1: Obtain the target-style dataset $\mathcal{S}'$ by a GAN;
2: Initialize $\theta$ by pre-training on the target-style source dataset $\mathcal{S}'$ ;
3: for $r = 1, 2, \ldots, R_{max}$ do
4:     Extract embedding features of training data by $\mathcal{F}_\theta$ ;
5:     Generate pseudo labels $\hat{\mathcal{Y}}^{\mathcal{T}}$ by clustering with extracted features;
6:     Adjust sampling threshold $\tau$ with the number of iterations $r$ as (6)
7:     Mine credible sample set $\mathcal{X}_C^{\mathcal{T}}$ as (2)
8:     Update $\mathcal{F}_\theta$ with credible sample set $\mathcal{X}_C^{\mathcal{T}}$ and generated pseudo labels $\hat{\mathcal{Y}}^{\mathcal{T}}$ as (9)
9: end for
10: return $\theta$
```

and a logarithmic function, which is formulated as:

$$\begin{aligned}\mathcal{L}_{ims}(x_a^t) &= \log\big(1 + \sum_{i \neq a} e^{m-d_{a,i}}\big) \\ &= -\log \frac{e^{-d_{a,a}}}{e^{-d_{a,a}} + \sum_{i \neq a} e^{m-d_{a,i}}} \\ &= -\log \frac{e^{-d_{a,a}}}{\sum_{i=1}^{N} e^{m_a - d_{a,i}}},\end{aligned} \tag{8}$$

where $m_a$ is an adaptive margin. For the same instance, $m_a$ is zero. For others, $m_a$ is large. In this formulation, we assume that the distance between the sample and itself is zero, i.e., $d_{a,a} = 0$. Different from other instance discrimination learning methods (e.g., [46], [61]), we learn a spreading metric space with a large margin. This metric space encourages an inter-class discrimination by the margin constraint, which is beneficial for robust clustering and credible sample mining.

### 3.4 Objective Function

Given the anchor sample set $\mathcal{X}_C^{\mathcal{T}}$ discovered by our adaptively credible sample mining strategy, we train our embedding model $\mathcal{F}_\theta$ with the objective function combining the proposed instance margin spreading loss and conventional metric learning loss:

$$\mathcal{L} = \sum_{x_i^t \in \mathcal{X}_C^{\mathcal{T}}} \mathcal{L}_{tri}(x_i^t) + \lambda \mathcal{L}_{ims}(x_i^t), \tag{9}$$

where $\mathcal{L}_{tri}(x_i^t)$ is the common metric learning loss: Triplet Loss [15], and $\lambda$ denotes the hyper-parameter that balance the importance of different objectives. The triplet loss aims to learn an embedding space in which an anchor sample is closer to its positive sample than other negative ones by a large margin. We formulated it as follows:

$$\mathcal{L}_{tri}(x_i^t) = \big[||f_i - f_i^+||_2^2 - ||f_i - f_i^-||_2^2 + m_{tri}\big]_+, \tag{10}$$

**Table 1.** The basic statictics of all datasets in experiments.

| **Datasets** | Identities | Images | Cameras | Train IDS | Test IDS | Labeling |
|---|---|---|---|---|---|---|
| **Market-1501** | 1501 | 32668 | 6 | 751 | 750 | Hand/DPM |
| **DukeMTMC-reID** | 1812 | 36411 | 8 | 702 | 1110 | Hand |
| **CUHK03** | 1467 | 14096 | 2 | 767 | 700 | DPM |

where $[\cdot]_+$ indicates the max function $\max(0, \cdot)$ which denotes that gradients will disappear when the difference between the intra-class and inter-class distances is large enough. $f_i, f_i^+, f_i^-$ respectively denote as features of the anchor, positive and negative sample in a triplet. The positive and negative samples selection strategy follows [15] that only uses the hardest positive and negative points in the mini-batch. $m_{tri}$ is a margin to enhance the discriminative ability, which is similar with $m_a$ in the instance margin spreading loss. For more clear explanation, we provide the Algorithm 1 to introduce the learning process of our DCML method in detail.

### 3.5 Discussion

Some methods (e.g., PUL [10], UDA [34], PAST [52], and SSG [12]) also apply the clustering algorithm to generate pseudo labels of target domain. However, the pseudo labels might **involve much noise**, which misleads the training process in the target domain. To solve this problem, our DCML method develops a credible sample mining strategy in the metric learning to avoid the noisy labels. PUL [10] have proposed a reliable objective function to regulate the sparsity of samples, and then simultaneously optimized the objective of the discriminative model and the regulation term of the number of samples. However, this regulation term may disturb the original discriminative learning since the valuable samples in the optimization process tend to be removed. Different from PUL, our DCML method proposes a credible sample mining strategy which is inspired by the hard negative mining in the metric learning. The credible data sampling is separated from the metric learning process, **without the disturbance**. As far as we know, DCML is the first metric learning method to adaptively select credible samples, which does not break the discriminative learning.

## 4 Experiment

In this section, we evaluated our DCML method on three large-scale person ReID datasets: Market-1501 [57], DukeMTMC-reID [30], and CUHK03 [21]. Quantitatively, we compared our DCML method with other state-of-the-art unsupervised domain adaptation person ReID approaches and conducted ablation studies to analyze each component. Besides, we visualized the embedding space to qualitatively analyze our method.

**Table 2.** Ablation studies show the influences of design choices on mAP and Rank-1,5,10(%), with Market-1501 as the source dataset and DukeMTMC-reID as the target dataset and vice versa. The † denotes that this method is reproduced by ourself with the same backbone and hyperparameters.

| **Method** | **M → D** | | | | **D → M** | | | |
|---|---|---|---|---|---|---|---|---|
| | **mAP** | **R1** | **R5** | **R10** | **mAP** | **R1** | **R5** | **R10** |
| UDA† | 54.4 | 72.7 | 82.1 | 85.6 | 56.5 | 78.4 | 86.5 | 89.5 |
| UDA† + GAN | 60.4 | 76.3 | 85.8 | 88.4 | 70.5 | 85.8 | 93.2 | 95.1 |
| UDA† + CAMS+ IMSLoss | 60.2 | 75.9 | 84.0 | 86.7 | 69.2 | 85.4 | 92.8 | 94.8 |
| UDA† + GAN + IMSLoss | 62.2 | 76.9 | 85.9 | 88.8 | 71.3 | 86.9 | 92.9 | 95.1 |
| DCML (KNN ) | 63.3 | 79.1 | **87.2** | 89.4 | **72.6** | 87.9 | **95.0** | **96.7** |
| DCML (Prototype ) | **63.5** | **79.3** | 86.7 | **89.5** | 72.3 | **88.2** | 94.9 | 96.4 |

### 4.1 Datasets and Experimental Settings

**Datasets:** Our experiments are conducted on three large-scale datasets including Market-1501 [57],DukeMTMC-reID [30], and CUHK03 [21]. Although all the above datasets are collected from the natural real-world scene of the university environment, there still is a large domain shift among them such as background, illumination, and clothing style. For example, the persons in the Market-1501 and DueMTMC-reID datasets mainly come from Asia and America respectively. For all datasets, we share the same experiment settings with the standard cross-domain person ReID experimental setups in the baseline method UDA [34] and PAST [52]. Specifically, we follow the source/target selection strategy, training/testing ID splitting strategy, and evaluation measuring protocols. For Market-1501 and DukeMTMC-reID datasets, we evaluated our method in the single query mode. While for the CUHK03 dataset, we only use the DPM detected images and choose the new train/test evaluation protocol in [59] for a fair comparison. The detailed information of the datasets are shown in Table 1.

**Evaluation Protocol:** In our experiments, we employed the standard metrics including cumulative matching characteristic (CMC) curve and the mean average precision (mAP) score to evaluate the performance of the person reID methods. We reported rank-1, rank-5 and rank-10 accuracy and mAP score in our experiments. Note that post-processing methods, e.g., re-ranking [59], are **not** applied for the final evaluation.

### 4.2 Implementation Details

**Source Domain Pre-training:** Leveraging the labeled source domain images, we pre-train a CNN model in a supervised manner by following the training strategy described in [2]. Specifically, we use the ImageNet pre-trained ResNet50 [14] without any attention model as the backbone of our model for fairness. The original $stride = 2$ convolution layer in the last block is replaced by a $stride = 1$ one to preserve the image resolution. For image preprocessing, we attempt to use the generative images by the SPGAN [7] and adopt the random horizontal flipping, random cropping, and random erasing data augmentation methods for image diversity. The supervisory signals in the source domain training consist of label smooth cross-entropy loss and triplet loss. Besides, other hyperparameters including image resolution, batch size, learning rate, weight decay factor, learning rate decay strategy, and max epochs are the same as [2].

**Pseudo Label Generation:** We adopt the DBSCAN clustering method [9] to generate pseudo labels, which is the same as the baseline UDA method [34]. The input of DBSCAN algorithm is the reranked distance matrix of the target domain samples and the output is the clustering result. We give each image cluster containing more than two samples a pseudo-label and then discard the individual images.

**DCML:** In the process of target domain adaptation, we train our model for 8 iterations and 30 epochs are required in each iteration. For the credible sample mining strategy, we set $\gamma_0 \approx 0.75$ and $\Delta\gamma \approx 0.05$ to update the sample selection threshold. Taking the DukeMTMC-reID datasets as an example, we select 12000 anchor samples in the first iteration and increase 1000 samples each iteration. For objective function, we respectively set the margins $m_a = 0.1$ and $m_{tri} = 0.3$ for instance margin spreading loss and triplet loss. The rate of loss weighting is set as $\lambda = 0.01$. In each mini-batch, we randomly select 224 samples from the credible sample set, in which each individual contains 16 images. We use Adam optimizer with an initial learning rate of 0.0005 and the weight decay of 0.001. The initial learning rate is reduced to 0.1 at 3th and 6th iterations, and in each iteration, it is temporarily reduced in the last 10 epochs. We conducted All our experiments on 4 Nvidia GTX 1080Ti GPUs with PyTorch 1.2.

### 4.3 Ablation Study

To analyze the effectiveness of individual components in our DCML approach, we conducted comprehensive ablation experiments on the M→ D and D→ M settings, where M → D denotes that the source dataset is Market-1051 and the target dataset is DukeMTMC-reID. We reproduced the UDA [34] method with the same backbone and hyperparameters of our method as the baseline, and applied the proposed credible anchor mining strategy, instance margin spreading loss, and the GAN based image style transfer on it. Table 2 We exhibited the comparison results in different settings in Table 2 and analyzed different components as follows.

**Credible Anchor Mining Strategy:** As shown in Table 2, CAMS denotes our credible anchor mining strategy. Compared the performance under the setting of $UDA\dagger + GAN + IMSLoss$ and the full DCML method, we can observe the obvious decline when the CAMS is removed. It illustrates that progressively and adaptively mining credible samples assists the target domain training by discarding samples with noise labels. In addition, we compared the effectiveness of different credibility similarity methods. The KNN similarity and prototype similarity are comparable to evaluate the credibility, which indicates our sample mining strategy is robust for different credibility evaluation methods.

**Table 3.** Performance comparisons with SOTA unsupervised domain adaptation person Re-ID methods from Market-1501 to DukeMTMC-reID and vice versa.

| **Method** | **M → D** | | | | **D → M** | | | |
|---|---|---|---|---|---|---|---|---|
| | **mAP** | **R1** | **R5** | **R10** | **mAP** | **R1** | **R5** | **R10** |
| PTGAN [44] | - | 27.4 | - | 50.7 | - | 38.6 | - | 66.1 |
| SPGAN [7] | 22.3 | 41.1 | 56.6 | 63.0 | 22.8 | 51.5 | 70.1 | 76.8 |
| SPGAN+LMP [7] | 26.2 | 46.4 | 62.3 | 68.0 | 26.7 | 57.7 | 75.8 | 82.4 |
| HHL [60] | 27.2 | 46.9 | 61.0 | 66.7 | 31.4 | 62.2 | 78.8 | 84.0 |
| DA2S [17] | 30.8 | 53.5 | - | - | 27.3 | 58.5 | - | - |
| CR-GAN [5] | 48.6 | 68.9 | 80.2 | 84.7 | 54.0 | 77.7 | 89.7 | 92.7 |
| TJ-AIDL [43] | 23.0 | 44.3 | 59.6 | 65.0 | 26.5 | 58.2 | 74.8 | 81.1 |
| TAUDL [20] | 43.5 | 61.7 | - | - | 41.2 | 63.7 | - | - |
| UCDA [28] | 45.6 | 64.0 | - | - 49.6 | 73.7 | - | - | |
| EANet [16] | 48.0 | 78.0 | - | - | 51.6 | 78 | - | - |
| PUL [10] | 16.4 | 30.0 | 43.4 | 48.5 | 20.5 | 45.5 | 60.7 | 66.7 |
| MAR [51]* | 48.0 | 67.1 | 79.8 | - | 40.0 | 67.7 | 81.9 | |
| CASCL [45]* | 37.8 | 59.3 | 73.2 | 77.8 | 35.5 | 65.4 | 80.6 | 86.2 |
| ENC [61] | 40.4 | 63.3 | 75.8 | 80.4 | 43.0 | 75.1 | 87.6 | 91.6 |
| UDA [34] | 49.0 | 68.4 | 80.1 | 83.5 | 53.7 | 75.8 | 89.5 | 93.2 |
| PAST [52] | 54.3 | 72.4 | - | - | 54.6 | 78.4 | - | - |
| SSG++ [12] | 60.3 | 76.0 | 85.8 | 89.3 | 68.7 | 86.2 | 94.6 | 96.5 |
| DCML (KNN ) | 63.3 | 79.1 | **87.2** | 89.4 | **72.6** | 87.9 | **95.0** | **96.7** |
| DCML (Prototype ) | **63.5** | **79.3** | 86.7 | **89.5** | 72.3 | **88.2** | 94.9 | 96.4 |

**Table 4.** Performance comparisons with other methods from CUHK03 to DukeMTMC-reID and Market-1501.

| Methods | C → D | | C → M | |
|---|---|---|---|---|
| | **mAP** | **Rank-1** | **mAP** | **Rank-1** |
| PUL [10] | 12.0 | 23.0 | 18.0 | 41.9 |
| PTGAN [44] | - | 17.6 | - | 31.5 |
| HHL [60] | 23.4 | 42.7 | 29.8 | 56.8 |
| EANet [16] | 26.4 | 45.0 | 40.6 | 66.4 |
| PAST [52] | 51.8 | 69.9 | 57.3 | **79.5** |
| DCML(KNN) | **56.9** | **73.7** | 58.0 | 78.7 |
| DCML(Prototype) | 54.6 | 72.2 | **59.5** | 78.7 |

**Instance Margin Spreading Loss:** The proposed IMS Loss aims to increase inter-class discrimination by enlarging the margin between the instances. We conducted the ablation studies about IMS Loss on the both "UDA" and "UDA+GAN" baselines, and obtained consistent improvement. Besides, we observed that the improvement on the stronger baseline (GAN+UDA) is lower than the original UDA method. This might be due to the generative images with GAN have a lower domain shift than the original images. The embedding space pre-trained with generative images is more spreading.

**Image Style Transfer:** In our final system, we employed the domain adaptation generative images with SPGAN [7] to pre-train the model on the source domain. The generator transfers the style of source domain images to the target domain style, which reduces the domain shift between source and target datasets. With the generative images pre-train, the baseline UDA method achieves a large improvement, which demonstrates that the quality of predicted pseudo labels is important for target domain finetuning. It also motivates us to additionally enhance the quality of pseudo labels.

### 4.4 Comparison with State-of-the-art Methods

We compared our method with other SOTA unsupervised domain adaptation person ReID methods on the Market-1501, DukeMTMC-ReID and CUHK03 datasets. Specifically, we conducted the experiments following evaluation settings in [52] including M→ D, D→ M, C→ D, and C→ M tasks, where M, D, C respectively denote Market-1501, DukeMTMC-ReID and CUHK03 datasets. As shown in Table 3 and 4, the bottom groups summarize the performance of methods generating pseudo superiority signal to train the model on the target domain, while the top and middle groups respectively show these methods using GAN or other auxiliary attributes. Our DCML achieved consistent improvement over other comparing methods, which indicates the effectiveness of our credible sample mining strategy and instance margin spreading loss.

**M→ D and D→ M:** As shown in Table 3, we compare our results with 7 methods finetuning meodel by pseudo superiority signal, 5 methods reducing the domain shift with GAN and 4 methods using auxiliary clues. The * in the tables denotes that the method whose source dataset is MSMT17 [44], which is the largest re-ID dataset with large-scale images and multiple cameras. We achieve the state-of-the-art results for both settings.

**C→ D and C→ M:** We also evaluated our DCML method using CUHK03 [21] as the source dataset. The results of our DCML method and other state-of-the-art methods are summarized in Table 4. Our DCML method improved PAST [52] by adaptively and mining credible anchors and progressively adjusting the mining strategy, which avoids the misleading from noise labels. Note that we don't use the complex part model like PCB [37] in our DCML method.

### 4.5 Qualitative Analysis

To validate the effectiveness of our DCML method, we qualitatively examined the learned embeddings. As shown in Fig. 3, we visualize the Barnes-Hut t-SNE [25] map of our learned embeddings of the gallery dataset in DukeMTMC-ReID. To observe the details, we magnify several regions in the corners. Despite the large intra-class variations such as illumination, backgrounds, viewpoints and human poses, our DCML method still groups similar individuals on the target domain in an unsupervised manner.

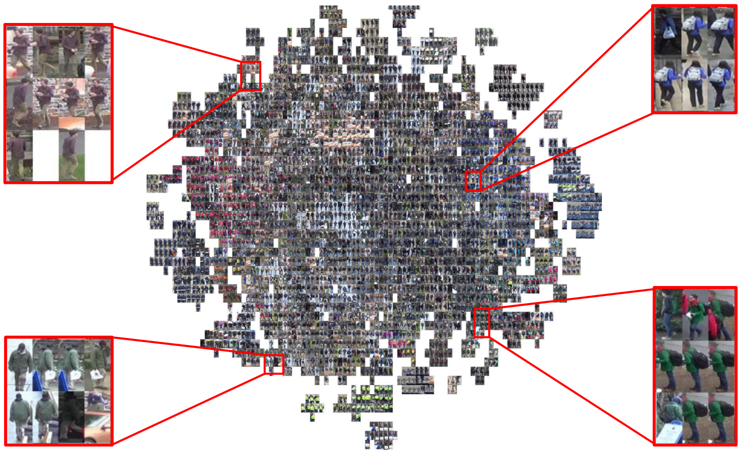

**Fig. 3.** Barnes-Hut t-SNE visualization [25] of the proposed DCML method on the gallery set of DukeMTMC-ReID, where we zoom in several areas for a clear view.

## 5 Conclusion

In this paper, we have proposed a deep credible metric learning method for unsupervised domain adaptation person re-identification, which adaptively mines credible samples to train the network and progressively adjusts the sample mining strategy with the learning process. It is due to that the generated pseudo labels are always unreliable and the noise will mislead the model training. We present two similarity metrics for the goal of measuring the credibilities of pseudo labels, including the k-Nearest Neighbor distance for density evaluation and the prototype distance for centrality evaluation. With the training process, we progressively reduce the limitation to select more samples. In addition, we propose an instance margin spreading loss to further increase the inter-class discrimination. We have conducted extensive experiments to demonstrate the effectiveness of our DCML method. In the future, we will attempt to design a credible negative mining strategy to further improve the cross-domain metric learning.

## Acknowledge

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFA0700802, in part by the National Natural Science Foundation of China under Grant 61822603, Grant U1813218, Grant U1713214, and Grant 61672306, in part by Beijing Natural Science Foundation under Grant No. L172051, in part by Beijing Academy of Artificial Intelligence (BAAI), in part by a grant from the Institute for Guo Qiang, Tsinghua University, in part by the Shenzhen Fundamental Research Fund (Subject Arrangement) under Grant JCYJ20170412170602564, and in part by Tsinghua University Initiative Scientific Research Program.

## References

1. Chen, B., Deng, W., Hu, J.: Mixed high-order attention network for person re-identification. In: ICCV (October 2019)
2. Chen, G., Lin, C., Ren, L., Lu, J., Jie, Z.: Self-critical attention learning for person re-identification. In: ICCV (2019)
3. Chen, G., Lu, J., Yang, M., Zhou, J.: Spatial-temporal attention-aware learning for video-based person re-identification. TIP **28**(9), 4192–4205 (2019)
4. Chen, G., Zhang, T., Lu, J., Zhou, J.: Deep meta metric learning. In: ICCV (October 2019)
5. Chen, Y., Zhu, X., Gong, S.: Instance-guided context rendering for cross-domain person re-identification. In: ICCV. pp. 232–242 (2019)
6. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: CVPR. pp. 1335–1344 (2016)
7. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: CVPR. pp. 994–1003 (2018)
8. Duan, Y., Lu, J., Zhou, J.: Uniformface: Learning deep equidistributed representation for face recognition. In: CVPR. pp. 3415–3424 (2019)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd. vol. 96, pp. 226–231 (1996)
10. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: Clustering and fine-tuning. TOMM **14**(4), 83 (2018)
11. Fang, P., Zhou, J., Roy, S.K., Petersson, L., Harandi, M.: Bilinear attention networks for person retrieval. In: ICCV (October 2019)
12. Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: ICCV (October 2019)
13. Harwood, B., Kumar, B., Carneiro, G., Reid, I., Drummond, T., et al.: Smart mining for deep metric learning. In: ICCV. pp. 2821–2829 (2017)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
15. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv (2017)
16. Huang, H., Yang, W., Chen, X., Zhao, X., Huang, K., Lin, J., Huang, G., Du, D.: Eanet: Enhancing alignment for cross-domain person re-identification. arXiv preprint arXiv:1812.11369 (2018)
17. Huang, Y., Wu, Q., Xu, J., Zhong, Y.: Sbsgan: Suppression of inter-domain background shift for person re-identification. In: ICCV (October 2019)
18. Kalayeh, M.M., Basaran, E., Gökmen, M., Kamasak, M.E., Shah, M.: Human semantic parsing for person re-identification. In: CVPR. pp. 1062–1071 (2018)
19. Li, D., Chen, X., Zhang, Z., Huang, K.: Learning deep context-aware features over body and latent parts for person re-identification. In: CVPR (2017)
20. Li, M., Zhu, X., Gong, S.: Unsupervised person re-identification by deep learning tracklet association. In: ECCV. pp. 737–753 (2018)
21. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR. pp. 152–159 (2014)

22. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: CVPR. p. 2 (2018)
23. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR. pp. 2197–2206 (2015)
24. Lin, Y., Dong, X., Zheng, L., Yan, Y., Yang, Y.: A bottom-up clustering approach to unsupervised person re-identification. In: AAAI. vol. 33, pp. 8738–8745 (2019)
25. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. JMLR **9**(Nov), 2579–2605 (2008)
26. Oh Song, H., Jegelka, S., Rathod, V., Murphy, K.: Deep metric learning via facility location. In: CVPR. pp. 5382–5390 (2017)
27. Oh Song, H., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: CVPR. pp. 4004–4012 (2016)
28. Qi, L., Wang, L., Huo, J., Zhou, L., Shi, Y., Gao, Y.: A novel unsupervised camera-aware domain adaptation framework for person re-identification. In: ICCV (October 2019)
29. Qian, X., Fu, Y., Xiang, T., Wang, W., Qiu, J., Wu, Y., Jiang, Y.G., Xue, X.: Pose-normalized image generation for person re-identification. In: ECCV. pp. 650–667 (2018)
30. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: ECCV. pp. 17–35 (2016)
31. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: CVPR. pp. 5363–5372 (2018)
32. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: NeurIPS. pp. 4077–4087 (2017)
33. Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective. In: NeurIPS. pp. 1857–1865 (2016)
34. Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., Wang, X.: Unsupervised domain adaptive re-identification: Theory and practice. arXiv preprint arXiv:1807.11334 (2018)
35. Su, C., Li, J., Zhang, S., Xing, J., Gao, W., Tian, Q.: Pose-driven deep convolutional model for person re-identification. In: ICCV (2017)
36. Sun, Y., Xu, Q., Li, Y., Zhang, C., Li, Y., Wang, S., Sun, J.: Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. In: CVPR (June 2019)
37. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: ECCV. pp. 480–496 (2018)
38. T Ali, M.F., Chaudhuri, S.: Maximum margin metric learning over discriminative nullspace for person re-identification. In: ECCV. pp. 122–138 (2018)
39. Tay, C.P., Roy, S., Yap, K.H.: Aanet: Attribute attention network for person re-identifications. In: CVPR. pp. 7134–7143 (2019)
40. Ustinova, E., Lempitsky, V.: Learning deep embeddings with histogram loss. In: NeurIPS. pp. 4170–4178 (2016)
41. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACMMM. pp. 274–282 (2018)
42. Wang, J., Zhou, F., Wen, S., Liu, X., Lin, Y.: Deep metric learning with angular loss. In: ICCV. pp. 2593–2601 (2017)
43. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR. pp. 2275–2284 (2018)

44. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: CVPR. pp. 79–88 (2018)
45. Wu, A., Zheng, W.S., Lai, J.H.: Unsupervised person re-identification by camera-aware similarity consistency learning. In: ICCV (October 2019)
46. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: CVPR. pp. 3733–3742 (2018)
47. Xia, B.N., Gong, Y., Zhang, Y., Poellabauer, C.: Second-order non-local attention networks for person re-identification. In: ICCV (October 2019)
48. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: CVPR. pp. 1249–1258 (2016)
49. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. In: CVPR. pp. 3415–3424 (2017)
50. Yu, B., Tao, D.: Deep metric learning with tuplet margin loss. In: ICCV. pp. 6490–6499 (2019)
51. Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: CVPR. pp. 2148–2157 (2019)
52. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: ICCV (October 2019)
53. Zhang, Z., Lan, C., Zeng, W., Chen, Z.: Densely semantically aligned person re-identification. In: CVPR. pp. 667–676 (2019)
54. Zhao, H., Tian, M., Sun, S., Shao, J., Yan, J., Yi, S., Wang, X., Tang, X.: Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In: CVPR (2017)
55. Zhao, Y., Shen, X., Jin, Z., Lu, H., Hua, X.s.: Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In: CVPR. pp. 4913–4922 (2019)
56. Zheng, F., Deng, C., Sun, X., Jiang, X., Guo, X., Yu, Z., Huang, F., Ji, R.: Pyramidal person re-identification via multi-loss dynamic training. In: CVPR. pp. 8514–8522 (2019)
57. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV. pp. 1116–1124 (2015)
58. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., Tian, Q., et al.: Person re-identification in the wild. In: CVPR. vol. 1, p. 2 (2017)
59. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR (2017)
60. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: ECCV. pp. 172–188 (2018)
61. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: Exemplar memory for domain adaptive person re-identification. In: CVPR. pp. 598–607 (2019)