Unsupervised Domain Adaptation with Noise Resistible Mutual-Training for Person Re-identification

Fang Zhao[1], Shengcai Liao[1⋆], Guosen Xie[1], Jian Zhao[2], Kaihao Zhang[3], and Ling Shao[1,4]

[1] Inception Institute of Artificial Intelligence, Abu Dhabi, UAE
{fang.zhao,shengcai.liao,guosen.xie,ling.shao}@inceptioniai.org
[2] Institute of North Electronic Equipment, Beijing, China
zhaojian90@u.nus.edu
[3] Tencent AI Lab, Shenzhen, China
super.khzhang@gmail.com
[4] Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

Abstract. Unsupervised domain adaptation (UDA) in the task of person re-identification (re-ID) is highly challenging due to large domain divergence and no class overlap between domains. Pseudo-label based self-training is one of the representative techniques to address UDA. However, label noise caused by unsupervised clustering is always a trouble to self-training methods. To depress noises in pseudo-labels, this paper proposes a Noise Resistible Mutual-Training (NRMT) method, which maintains two networks during training to perform collaborative clustering and mutual instance selection. On one hand, collaborative clustering eases the fitting to noisy instances by allowing the two networks to use pseudo-labels provided by each other as an additional supervision. On the other hand, mutual instance selection further selects reliable and informative instances for training according to the peer-confidence and relationship disagreement of the networks. Extensive experiments demonstrate that the proposed method outperforms the state-of-the-art UDA methods for person re-ID.

Keywords: Unsupervised domain adaptation, person re-identification, collaborative clustering, mutual instance selection

1 Introduction

Person re-identification (re-ID), which aims at retrieving images of the same person from the database given a person image, has advanced considerably relying on the power of deep learning technology in recent years [58, 50, 51, 34, 29, 32, 35, 48, 53, 19]. However, due to the problem of domain shift [17], the performance of a deep re-ID model that performs well in a source domain may drop significantly

⋆ Corresponding author.

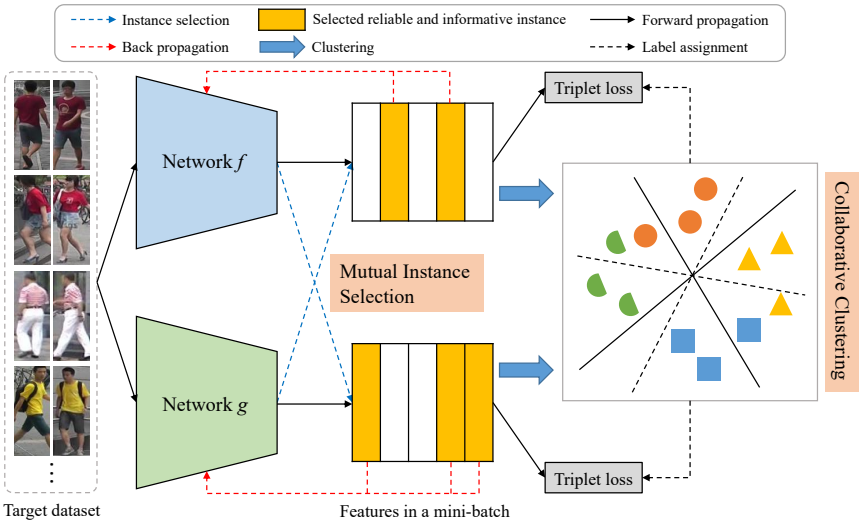

Fig. 1. Overview of the proposed Noise Resistible Mutual-Training (NRMT). NRMT maintains two networks during training, which performs collaborative clustering to ease the fitting to noisy instances and mutual instance selection to further select reliable and informative instances for the network update.

when applied to a target domain. Besides, it is usually not easy to obtain labels of target data in practice, which hinders supervised fine-tuning of the deep model on the target data.

To learn a deep re-ID model which generalizes well in the target domain without using labels from this domain, unsupervised domain adaptation (UDA) approaches are proposed given labeled source data and unlabeled target data [24, 56, 21, 45, 57, 5]. Different from the traditional setting of UDA which assumes that the source and target domains share the same classes, UDA in person re-ID is under an open-set scenario, i.e., the two domains have totally different person identities (classes). Thus, it is a more challenging task.

Self-training is an effective strategy for UDA in person re-ID [8, 31, 49, 11], which performs clustering with the pre-trained source model to assign pseudo-labels to samples of the target dataset, then alternately updates the model with the pseudo-labels on target data and re-assigns the labels with the updated model to make the model adapt to the target data progressively. In the early stage of training, pseudo-labels assigned by clustering usually contain lots of noises due to the divergence between the source and target domains. The model can correct some of them by learning from clean labels. However, as the number of training iteration increases, some noisy instances are fitted by the model and cannot be corrected anymore. These noises eventually harm the self-training model performance on the target data.

In order to address the problem mentioned above, we propose *Noise Resistible Mutual-Training* (NRMT) to effectively reduce the impact of noisy instances

throughout the training process by leveraging dual networks with information interaction. As shown in Fig. 1, NRMT maintains two networks during training, which performs collaborative clustering to ease the fitting to noisy instances and mutual instance selection to further select reliable and informative instances for the network update. We argue that there always exist some noisy instances that the single network cannot distinguish by itself in the iteration process of self-training. Inspired by deep learning with noisy labels [22, 14], we use another network with different learning ability to assist in correcting pseudo-label errors.

Specifically, for each iteration, collaborative clustering allows the two networks to not only learn by their respective pseudo-labels but also exploit the ones provided by each other as an additional supervision. For one network, its peer network can provide various labels for instances due to different learning ability. Although there also exists noises in these labels, they still can be used to reduce the effect of label errors of the single network because deep neural networks tend to fit easy (more likely to be correct) instances first [1]. For each mini-batch, mutual instance selection is introduced to further filter out noisy instances while keeping informative instances. Here the reliability of a triplet of instances is assessed for one network according to the prediction confidence of its peer network on this triplet. Informative instances are also important for improving the network performance. Thus, we further measure the amount of information of the triplet by the relationship disagreement of the predictions across the networks. Combining collaborative clustering at each iteration and mutual instance selection within each mini-batch, the proposed NRMT can effectively depress noises in pseudo-labels and improve the performance of both the two networks.

Our main contributions can be summarized as follows: 1) We present a novel noise resistible mutual-training method for unsupervised domain adaptation in person re-ID, which exploits dual network interaction to depress noises in pseudo-labels of unsupervised iterative training on the target data. 2) We introduce a collaborative clustering to ease the fitting to noisy instances by the memorization effects of deep networks. 3) We propose a mutual instance selection based on the peer-confidence and relationship disagreement of networks on triplets of instances to select reliable and informative instances in a mini-batch.

2 Related Work

Unsupervised domain adaptation. Our work is related to unsupervised domain adaptation (UDA) [36, 3, 37, 28]. Some methods are proposed to match distributions between the source and target domains [20, 33]. Long *et al.* [20] embed features of task-specific layers in a reproducing kernel Hilbert space to explicitly match the mean embeddings of different domain distributions. Sun *et al.* [33] propose to learn a linear transformation that aligns the second-order statistics of feature distributions between the two domains. There are also several works that learn domain-invariant features [12, 37]. Ganin *et al.* [12] introduce a gradient reversal layer to learn features invariant to domain via an adversarial loss. The

aforementioned methods only consider the closed-set scenario. Recently, some works are introduced to address the problem of open set domain adaptation [23, 27, 10], where several classes are unknown in the two domains (or in the target domain). However, classes of the two domains are entirely different for UDA in person re-ID, which presents a greater challenge.

UDA for person re-ID. There are many research works that have been proposed for unsupervised cross-domain person re-ID [24, 38, 56, 31, 30, 57, 41, 46, 40, 42, 5, 25, 44]. Some of them focus on image-level domain invariance. Wei *et al.* [39] propose a person transfer generative adversarial network to bridge the domain gap, which considers the style transfer and person identity keeping. Deng *et al.* [7] generate target image samples through the coordination between a CycleGAN and an Siamese network. Several works also try to improve the model generalization from the view of feature learning. Wang *et al.* [38] establish an identity-discriminative and attribute-sensitive feature representation space transferable to any new (unseen) target domain. Qi *et al.* [25] develop a camera-aware domain adaptation to reduce the discrepancy across sub-domains in cameras and utilize the temporal continuity in each camera to provide discriminative information.

Recently, some methods are developed based on the self-training framework. Fu *et al.* [11] present a self-similarity grouping to explore the potential similarities by both global and local appearance cues. Zhang *et al.* [49] propose a self-training method with progressive augmentation framework to offer complementary data information by different learning strategies for self-training. In contrast, our method provides complementary information through dual network interaction. Ge *et al.* [13] present a mutual mean-teaching framework to softly refine the pseudo-labels in the target domain. Note that our method and [13] are complementary and can be combined.

Deep learning with noisy labels. There exist several works that aim at improving the training of deep models with noisy labels. Decoupling [22] trains two networks simultaneously, and then updates models only using the instances that have different predictions from these two networks. Co-teaching [14] proposes to select small-loss instances of each network as the useful knowledge and transfer such useful instances to its peer network for the further training. Yu *et al.* [47] combine the disagreement strategy with Co-teaching, which trains two deep neural networks with the disagreement-update step (data update) and the cross-update step (parameters update). These methods mainly focus on the classification problem, which cannot be directly applied to the metric learning problem in our task.

3 Our Method

Given a labeled training dataset $\{\mathbf{X}^s, \mathbf{Y}^s\}$ from the source domain and an unlabeled training dataset $\mathbf{X}^t$ from the target domain where identities of persons are different from the ones in the source domain, we aim to learn discriminative feature representations for target testing dataset. In this section, we present the

proposed Noise Resistible Mutual-Training (NRMT) method, which incorporates the interaction of dual networks to depress noises in pseudo-labels produced by unsupervised clustering in a self-training process. Now, we proceed to explain each component of our NRMT in details.

3.1 Self-Training with Clustering

Since the ground truth labels of the target person images are not available, one way to fine-tune the target model is to consider the target labels as latent variables that can be inferred in the learning process. Thus, a typical self-training framework for unsupervised domain adaptation aims to minimize the following loss function:

$$\min_{\hat{\mathbf{Y}}^t, \mathbf{W}} \mathcal{L}(\hat{\mathbf{Y}}^t, f(\mathbf{X}^t; \mathbf{W})), \tag{1}$$

where $\hat{\mathbf{Y}}^t$ denotes the estimated target labels, $\mathbf{X}^t$ is the set of target images and f denotes the target model parameterized by $\mathbf{W}$.

In the case of person re-ID, source and target domains do not share the common label space. Thus, one cannot directly apply the classifier trained on the source dataset to estimate the target identities. Similar with [31, 8], we perform clustering on CNN features to assign pseudo-labels to instances with the most confident predictions and assume that they are mostly correct. Once the target model is updated with these pseudo-labels, the remaining instances with less confidence are continuously explored by the model adapted better to the target domain. Therefore, to minimize the loss function in Eq. (1), we firstly initialize the model parameters $\mathbf{W}$ on the source data $\{\mathbf{X}^s, \mathbf{Y}^s\}$ and then apply an alternating block coordinate descent algorithm: 1) Fix $\mathbf{W}$ and minimize the loss w.r.t $\hat{\mathbf{Y}}^t$ through clustering. 2) Fix $\hat{\mathbf{Y}}^t$ and optimize the loss w.r.t $\mathbf{W}$ by stochastic gradient descent.

3.2 Mutual-Training with Collaborative Clustering

The problem of self-training based models [31, 8] is that the quality (correctness) of pseudo-labels generated by unsupervised clustering on the target data heavily affects the model performance. Although the deep learning model in self-training can avoid fitting noisy instances in the early stage of training due to the memorization effects of deep neural networks [1] and improve the performance progressively as more and more instances with high confidence are explored, there inevitably exist some label errors that cannot be corrected and would be overfitted as the training proceeds. These accumulated errors eventually impede the performance growth.

In order to reduce the label error accumulation throughout the training process, the proposed NRMT maintains two neural networks f parameterized by $\mathbf{W}_f$ and g parameterized by $\mathbf{W}_g$ simultaneously during training, and allows them to share clustering information by collaborative clustering at each iteration to reduce the effect of their respective label errors.

To make f and g have different learning abilities, we use different random seeds to pre-train f and g on the source dataset $\mathbf{X}^s$ with labels $\mathbf{Y}^s$ by the triplet loss and the Softmax loss [31]. Here f and g have the same network architecture to facilitate the deployment. Because deep neural networks are highly non-convex models, different initializations can still lead to different local optima even with the same architecture and optimization algorithm [14]. Then, we use the pre-trained f and g to extract features on the target dataset $\mathbf{X}^t$ and obtain two sets of pseudo-labels $\hat{\mathbf{Y}}_f^t$ and $\hat{\mathbf{Y}}_g^t$ through applying clustering to the features. Since the target domain has classes different from the source domain, we drop the Softmax loss and fine-tune the networks on the target data only using the triplet loss with the pseudo-labels. To share clustering information, f and g consider both their own pseudo-labels and the ones of their peer networks. Thus, we have a joint loss function for each network:

$$\mathcal{L}_f = \mathcal{L}_{tri}(\hat{\mathbf{Y}}_f^t, f(\mathbf{X}^t; \mathbf{W}_f)) + \mathcal{L}_{tri}(\hat{\mathbf{Y}}_g^t, f(\mathbf{X}^t; \mathbf{W}_f)), \tag{2}$$

$$\mathcal{L}_g = \mathcal{L}_{tri}(\hat{\mathbf{Y}}_g^t, g(\mathbf{X}^t; \mathbf{W}_g)) + \mathcal{L}_{tri}(\hat{\mathbf{Y}}_f^t, g(\mathbf{X}^t; \mathbf{W}_g)), \tag{3}$$

where $\mathcal{L}_{tri}$ is the batch-sampling triplet loss [16].

Different from self-training where the network assigns new pseudo-labels to the training instances at each iteration only according to its own parameter update, in NRMT, to make the learning more robust, the two networks f and g collaboratively assign pseudo-labels, i.e., each instance has two pseudo-labels from f and g, respectively. The study on memorization in deep networks [1] suggests that deep networks tend to prioritize learning easy patterns. Usually noisy instances caused by clustering are relatively hard examples, thus if one instance is assigned two labels, the networks will fit the clean (easy) one first to become robust and the error may be eliminated at the next iteration. The joint loss functions in Eq. (2) and Eq. (3) are similar to Co-training [2] where classifiers are trained on two views (two independent sets of features). However, here we have two networks but only have a single view, and we utilize the memorization effect of deep networks to handle the error in labels.

3.3 Mutual Instance Selection

Although collaborative clustering across networks is able to ease the fitting to noisy instances for each iteration, these noisy instances still have impact on the network training in a mini-batch, especially in the advanced stage of training. To further select reliable and informative instances in a mini-batch, we introduce a mutual instance selection strategy by considering both the peer-confidence and relationship disagreement of the two networks.

Reliable Instance Selection by Peer-Confidence. In order to select reliable instances for training, we consider using the prediction confidence of the peer network to measure the reliability of instances for one network. We argue that in the metric learning, the relationship of one pair of instances with other pairs

in the feature space can provide more information about the network prediction than the distance between one instance and another one. Thus, we compute the prediction confidence based on the relationship of a triplet of instances.

Given an instance x, its corresponding positive instance x_p and negative instance x_n from a mini-batch, we encode the relationship of the triplet $\{x, x_p, x_n\}$ by the difference between the Euclidean distances of the positive and negative pairs in the feature space:

$$\mathcal{D}(x, x_p, x_n; f) = ||f(x) - f(x_p)||_2 - ||f(x) - f(x_n)||_2, \tag{4}$$
$$\mathcal{D}(x, x_p, x_n; g) = ||g(x) - g(x_p)||_2 - ||g(x) - g(x_n)||_2, \tag{5}$$

where $f(x)$ and $g(x)$ is the features extracted by the networks f and g, respectively. The smaller the difference is, the higher the confidence is. If the difference of the peer network g (resp. f) of f (resp. g) for the triplet $\{x, x_p, x_n\}$ is smaller than a threshold T_c:

$$\mathcal{D}(x, x_p, x_n; g) < T_c, \tag{6}$$
$$\text{resp. } \mathcal{D}(x, x_p, x_n; f) < T_c, \tag{7}$$

we call $\{x, x_p, x_n\}$ as a peer-confident triplet of instances for f (resp. g) and use this peer-confident triplet to update f (resp. g). Because the two networks have different learning abilities, we expect that they can filter out various noisy instances [14] to make up for each other's mistakes.

Informative Instance Selection by Relationship Disagreement. The peer-confidence of the network can pick up reliable (clean) instances in a mini-batch, but these instances usually contain lots of easy instances which provide limited information for the network performance improvement. To further select more informative instances, we propose to use the relationship disagreement between one network and its peer network to measure the amount of information on the basis of the peer-confidence.

Similar to the peer-confidence, we compute the relationship disagreement on a triplet of instances. We first define the prediction inconsistency of the two networks f and g combined with Eq. (4) and Eq. (5) as:

$$\mathcal{I}(x, x_p, x_n; f, g) = \mathcal{D}(x, x_p, x_n; f) - \mathcal{D}(x, x_p, x_n; g). \tag{8}$$

Larger absolute value of the inconsistency indicates that the triplet of instances has larger amount of information. It can be considered that there is the relationship disagreement between the predictions of two networks for the triplet $\{x, x_p, x_n\}$ if the absolute value of the prediction inconsistency is larger than a threshold T_d:

$$|\mathcal{I}(x, x_p, x_n; f, g)| > T_d \tag{9}$$

The networks are only updated on the mini-batch data with the relationship disagreement. Furthermore, when combined with the peer-confidence, Eq. (9)

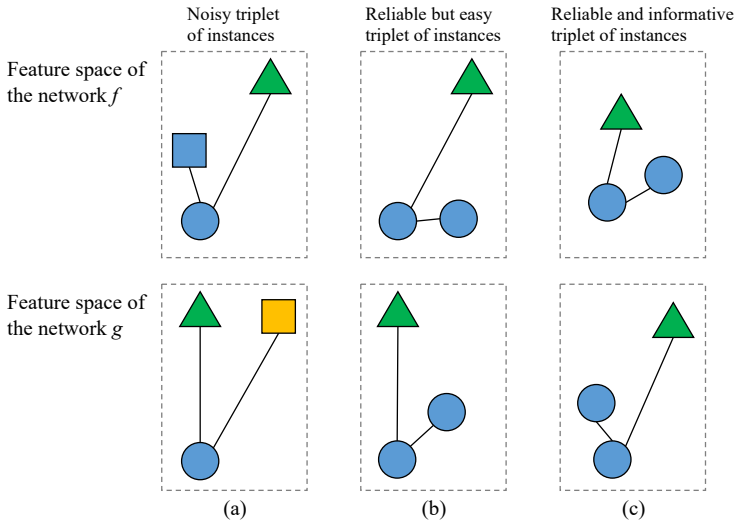

Fig. 2. Three types of triplets of instances obtained by the proposed mutual instance selection strategies. Different shapes (circle, triangle and square) denote different ground truth class labels and different colors (blue, green and yellow) denote different pseudo-labels. (a) Noisy triplet of instances obtained by $\mathcal{D}(x, x_p, x_n; g) \geq T_c$; (b) Reliable but easy triplet of instance obtained by $\mathcal{D}(x, x_p, x_n; g) < T_c$ but $\mathcal{I}(x, x_p, x_n; f, g) \leq T_d$; (c) Reliable and informative triplet of instances obtained by $\mathcal{D}(x, x_p, x_n; g) < T_c$ and $\mathcal{I}(x, x_p, x_n; f, g) > T_d$. (Best viewed in color).

can be rewritten with the absolute symbol removed:

$$\mathcal{I}(x, x_p, x_n; f, g) > T_d, \tag{10}$$

$$\mathcal{I}(x, x_p, x_n; g, f) > T_d. \tag{11}$$

The intuition is that, for the item within the absolute symbol in Eq. (9) which is smaller than $-T_d$, because T_d is not less than zero and $\{x, x_p, x_n\}$ meets the peer-confidence condition in Eq. (6) or Eq. (7), we have

$$\mathcal{D}(x, x_p, x_n; f) < \mathcal{D}(x, x_p, x_n; g) - T_d < \mathcal{D}(x, x_p, x_n; g) < T_c, \tag{12}$$

$$\text{or } \mathcal{D}(x, x_p, x_n; g) < \mathcal{D}(x, x_p, x_n; f) - T_d < \mathcal{D}(x, x_p, x_n; f) < T_c. \tag{13}$$

As a result, when T_c is set to a proper small value, for the network f or g, the triplet $\{x, x_p, x_n\}$ is actually an easy instance that can be ignored during training. Fig. 2 illustrates three types of triplets of instances obtained by the proposed mutual instance selection strategies, where we consider instances selection for the network f according to the prediction of the network g.

For the clarity, the training process of NRMT is summarized in Algorithm 1. It is worth noting that we only maintain two networks in the stage of training and the performance of the two networks can be aligned to the similar level via the information interaction. Thus, we can use any one of the two networks for the deployment in practice.

Algorithm 1 Noise Resistible Mutual-Training (NRMT)

Input: Deep networks f and g, labeled source training dataset $\{\mathbf{X}^s, \mathbf{Y}^s\}$, unlabeled target training dataset $\mathbf{X}^t$, maximal number of updates N_{max}, maximal number of iterations I_{max}.
Output: f and g.

```
Pre-train f and g on {X^s, Y^s} with different random seeds, respectively;
for I = 1 to I_max do
    Extract features f(x) and g(x) on X^t;
    Perform clustering on f(x) and g(x) to generate pseudo-labels Ŷ_f^t and Ŷ_g^t;
    for N = 1 to N_max do
        Sample mini-batches M(Ŷ_f^t) and M(Ŷ_g^t) from X^t with Ŷ_f^t and Ŷ_g^t;
        Obtain M_f(Ŷ_f^t) and M_f(Ŷ_g^t) by Eq. (6) and Eq. (10);
        Obtain M_g(Ŷ_f^t) and M_g(Ŷ_g^t) by Eq. (7) and Eq. (11);
        Update f with both M_f(Ŷ_f^t) and M_f(Ŷ_g^t) as in Eq. (2);
        Update g with both M_g(Ŷ_f^t) and M_g(Ŷ_g^t) as in Eq. (3);
    end for
end for
```

4 Experiments

In this section, we evaluate the proposed NRMT using three large-scale person re-ID datasets, *i.e.*, Market-1501 [52], DukeMTMC-reID [54, 26] and MSMT17 [39] and the performance evaluations are presented in term of Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) under the single-query setting.

4.1 Datasets

Market-1501 [52] contains 32,668 labeled images of 1,501 identities. 12,936 images of 751 identities form the training set. 3,368 query images from the other 750 identities and 19,732 gallery images (with 2,793 distractors) are used as the test set. The bounding boxes of persons are generated by Deformable Part Model (DPM) [9]. **DukeMTMC-reID** [54, 26] includes 36,411 labeled images of 1,404 identities. 702 identities are randomly selected for training and the rest is used for testing. There are 16,522 training images, 2,228 query images and 17,661 gallery images. **MSMT17** [39] is the largest re-ID dataset consisting of 126,441 bounding boxes of 4,101 identities taken by 12 outdoor and 3 indoor cameras. 32,621 images of 1,041 identities are used for training.

4.2 Implementation Details

We adopt ResNet-50 [15] as the architectures of the two networks and initialize them with the parameters pre-trained on ImageNet [6]. All images are resized to 256×128. Random horizontal flipping and random erasing [55] are employed for training data augmentation. We use the Softmax and triplet losses to pre-train

Table 1. Evaluation on different values of the threshold T_c. Results of the two networks f and g are reported, respectively.

T_c	Duke → Market		Market → Duke	
	mAP	R1	mAP	R1
0	71.0/70.2	86.9/86.5	61.1/60.9	76.9/76.6
0.5	71.5/70.6	87.3/87.0	61.7/61.4	77.5/77.0
1.0	**72.2/71.1**	**88.0/87.5**	**62.3/62.0**	**78.1/77.5**
1.5	72.0/71.0	87.7/87.3	62.0/61.8	78.0/77.2
2.0	71.7/70.7	87.4/87.0	61.7/61.5	77.7/77.0

Table 2. Evaluation on different values of the threshold T_d. Results of the two networks f and g are reported, respectively.

T_d	Duke → Market		Market → Duke	
	mAP	R1	mAP	R1
0.3	71.2/70.4	87.3/86.8	61.3/61.0	77.0/76.7
0.4	71.6/70.7	87.6/87.2	61.8/61.5	77.5/77.1
0.5	**72.2/71.1**	**88.0/87.5**	**62.3/62.0**	**78.1/77.5**
0.6	72.0/70.8	87.7/87.3	62.1/61.7	77.8/77.3
0.7	71.5/70.4	87.3/86.9	61.6/61.3	77.2/76.8

the two networks on the source dataset with different random seeds, respectively. The margin m in the triplet loss is 0.5. For each mini-batch, we randomly sample 32 identities and 4 images per identity. The SGD optimizer with a momentum of 0.9 is used to train the networks and the learning rate is 6e-5.

The peer-confidence threshold T_c is set to 1.0 and the relationship disagreement threshold T_d is set to 0.5. The HDBSCAN clustering algorithm [4] is adopted to produce pseudo-labels for each iteration, which does not require the number of clusters as prior parameter. The number of minimum samples for each cluster is set to 8. The maximal number of iterations is 30. At the first half of the iterative process, we train the networks only using collaborative clustering. Then we add mutual instance selection to further select clean and informative data in mini-batches for the network update.

4.3 Parameter Analysis

We first study impacts of some important parameter settings in the proposed NRMT, including the peer-confidence threshold T_c, the relationship disagreement threshold T_d and the number of minimum samples in the HDBSCAN clustering algorithm.

Peer-confidence threshold T_c. To analyze the impact of T_c in Eq. (6) and Eq. (7), we fix the relationship disagreement threshold $T_d = 0.5$ in all experiments. The results are listed in Table 1. We can observe that a proper value of T_c is important for NRMT to filter out noisy instances, which provides a reasonable assessment of the noise confidence. The best performance is achieved when T_c is set to 1.0.

Table 3. Evaluation on different numbers of the minimum samples for each cluster in HDBSCAN. Results of the two networks f and g are reported, respectively.

Min. Samp.	Duke → Market		Market → Duke	
	mAP	R1	mAP	R1
6	71.5/70.7	86.8/86.5	61.7/61.6	77.1/77.0
8	**72.2/71.1**	**88.0/87.5**	**62.3/62.0**	**78.1/77.5**
10	71.9/71.1	87.5/87.1	61.8/61.4	77.7/ 76.9

Table 4. Performance evaluation of components in the proposed NRMT on Market-1501 and DukeMTMC-reID. **Separate Training:** Train the two networks separately. **CC:** Collaborative clustering. **SC:** Instance selection by the peer-confidence. **SD:** Instance selection by the relationship disagreement. Results of the two networks f and g are reported, respectively.

Methods	DukeMTMC-reID → Market-1501			
	mAP	R1	R5	R10
Direct Transfer	33.0/32.3	63.3/62.3	77.2/76.5	82.3/81.9
Separate Training	54.2/53.0	76.4/75.6	88.3/87.8	92.2/91.8
Ours w/ CC	68.9/68.2	85.9/85.5	94.0/94.1	96.2/96.1
Ours w/ CC+SC	70.9/70.1	86.8/86.3	94.3/94.2	96.3/96.1
Ours w/ CC+SC+SD	**72.2/71.1**	**88.0/87.5**	**94.7/94.5**	**96.5/96.4**
Methods	Market-1501 → DukeMTMC-reID			
	mAP	R1	R5	R10
Direct Transfer	30.2/30.2	47.3/46.5	61.9/61.8	68.2/68.1
Separate Training	48.7/48.2	67.2/66.5	80.3/80.0	84.3/84.1
Ours w/ CC	59.1/58.7	75.8/75.4	85.5/85.2	88.2/88.3
Ours w/ CC+SC	60.6/60.2	76.6/76.3	86.1/85.9	88.9/88.7
Ours w/ CC+SC+SD	**62.3/62.0**	**78.1/77.5**	**87.0/86.8**	**89.7/89.2**

Relationship disagreement threshold T_d. We also conduct experiments to investigate the impact of T_d in Eq. (10) and Eq. (11). In all experiments, we fix the peer-confidence threshold $T_c = 1.0$. As reported in Table 2, when $T_d = 0.5$, we can obtain the best results. When T_d is set to a larger value, fewer instances are selected for update, which is likely to discard instances that are actually informative. Too small values of T_d will allow most of the instances to be involved in update, which may contain too many easy instance and thus cannot provide effective information for improving the network.

Number of minimum samples. To evaluate the influence of the number of minimum samples in HDBSCAN, we report the results of {6, 8, 10} minimum samples in Table 3. As we can see, the number 8 yields the superior performance. Note that our NRMT is not very sensitive to this prior clustering parameter.

4.4 Ablation Study

We further validate the effectiveness of each component in the proposed NRMT, including collaborative clustering, instance selection by the peer-confidence and

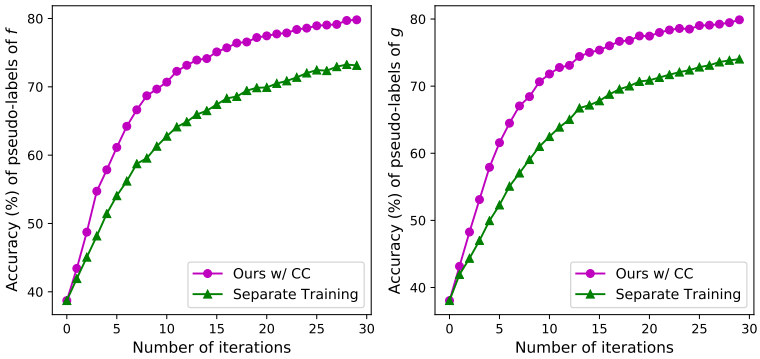

Fig. 3. Comparison on the accuracy of pseudo-labels in the iteration process for DukeMTMC-reID → Market-1501.

relationship disagreement on Market-1501 and DukeMTMC-reID. The results are shown in Table 4. As we can see, by sharing clustering information between two networks on the whole dataset, "Ours w/ CC" improves the performance of both the two networks compared with "Separate Training". This demonstrates that the collaborative clustering is able to ease the fitting to noisy instances caused by unsupervised clustering by exploiting different learning abilities of two networks and the memorization effect of deep networks. "Ours w/ CC+SC" and "Ours w/ CC+SC+SD" further obtain better results by prediction information interaction between the networks in mini-batches, which can pick up clean and informative instances to update the networks.

To explore the ability of correcting label errors of collaborative clustering, Fig. 3 illustrates the accuracy of pseudo-labels generated by clustering in the iteration process. It can be seen that the pseudo-label accuracy of the two networks f and g trained with collaborative clustering are both improved significantly compared with the networks trained separately. This shows that sharing clustering information between two networks on the whole dataset can effectively correct label errors at each iteration and reduce the accumulation of noises during training.

In Fig. 4, we show some examples of clean and informative, noisy and easy triplets of instances obtained by the proposed mutual instance selection strategy. We can observe that the clean and informative triplets selected by our strategy contains negative examples with similar appearances and positive examples with large variations. Meanwhile, our strategy can filter out not only noisy triplets but also easy triplets. This indicates that our strategy is able to act as a robust online hard example mining for the triplet loss in training with noisy labels.

4.5 Comparison with State-of-the-art Methods

In this section, we compare the proposed NRMT with the state-of-the-art unsupervised person re-ID methods on the transfers between DukeMTMC-reID and

Fig. 4. Examples of (a) clean and informative, (b) noisy and (c) easy triplets of instances obtained by the proposed mutual instance selection strategy in a mini-batch. Only the clean and informative triplets are used for the network update. For each triplet, the first two ones are positive examples and the last one is negative example.

Market-1501 and the transfers from DukeMTMC-reID/Market-1501 to MSMT17. Here we reports the averaged performance of the two networks f and g in NRMT.

Table 5 shows the results on the transfers between DukeMTMC-reID and Market-1501. We first compare the proposed NRMT with two hand-crafted features, i.e., LOMO [18] and Bag-of-Words (BoW) [52]. We can see that deep learning features can significantly improve the performance. Three unsupervised methods including UMDL [24], PUL [8] and DECAMEL [45] are compared. Our method surpasses these methods by a large margin by adapting to the target data from the source data progressively. We also compare with the unsupervised domain adaptation methods, including UDAP [31], MAR [46], ECN [57], PCB-R-PAST [49], SSG [11], ACT [43], etc. our method still achieves the best performance. Especially, our NRMT outperforms PCB-R-PAST [49], which also focuses on the improvement of label quality, by 17.1%/9.4% on mAP/Rank-1 accuracy for DukeMTMC-reID $\rightarrow$ Market-1501 and by 7.9%/5.4% for Market-1501 $\rightarrow$ DukeMTMC-reID. This demonstrates the effectiveness of information interactions between dual networks for noise reduction. Moreover, our NRMT also exceeds the second best method ACT [43] by clear margins.

We also evaluate our NRMT on transfers from DukeMTMC-reID and Market-1501 to MSMT17 in Table 6 . The results obtained by NRMT are 20.6%/45.2% on mAP/R1 accuracy for DukeMTMC-reID $\rightarrow$ MSMT17 and 19.8%/43.7% for Market-1501 $\rightarrow$ MSMT17, which all exceed the second best method, i.e., SSG [11]. This further demonstrates the superiority of our NRMT on the large-scale dataset.

Table 5. Comparison with the state-of-the-art UDA methods on Market-1501 and DukeMTMC-reID. The averaged performance of the two networks f and g is reported.

Methods	Market-1501				DukeMTMC-reID			
	mAP	R1	R5	R10	mAP	R1	R5	R10
LOMO [18]	8.0	27.2	41.6	49.1	4.8	12.3	21.3	26.6
BoW [52]	14.8	35.8	52.4	60.3	8.3	17.1	28.8	34.9
UMDL [24]	12.4	34.5	52.6	59.6	7.3	18.5	31.4	37.6
PUL [8]	20.5	45.5	60.7	66.7	16.4	30.0	43.4	48.5
DECAMEL [45]	32.4	60.2	-	-	-	-	-	-
PTGAN [39]	-	38.6	-	66.1	-	27.4	-	50.7
SPGAN+LMP [7]	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
TJ-AIDL [38]	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
HHL [56]	31.4	62.2	78.8	84.0	27.2	46.9	61.0	66.7
ARN [17]	39.4	70.3	80.4	86.3	33.4	60.2	73.9	79.5
UDAP [31]	53.7	75.8	89.5	93.2	49.0	68.4	80.1	83.5
MAR [46]	40.0	67.7	81.9	-	48.0	67.1	79.8	-
ECN [57]	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
CR-GAN+LMP [5]	33.2	64.5	79.8	85.0	33.3	56.0	70.5	74.6
PCB-R-PAST [49]	54.6	78.4	-	-	54.3	72.4	-	-
SSG [11]	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
ACT [43]	60.6	80.5	-	-	54.5	72.4	-	-
NRMT	**71.7**	**87.8**	**94.6**	**96.5**	**62.2**	**77.8**	**86.9**	**89.5**

Table 6. Comparison with the state-of-the-arts on transfers from DukeMTMC-reID and Market-1501 to MSMT17.

Methods	DukeMTMC-reID → MSMT17				Market-1501 → MSMT17			
	mAP	R1	R5	R10	mAP	R1	R5	R10
PTGAN [39]	3.3	11.8	-	27.4	2.9	10.2	-	24.4
ECN [57]	10.2	30.2	41.5	46.8	8.5	25.3	36.3	42.1
SSG [11]	13.3	32.2	-	51.2	13.2	31.6	-	49.6
NRMT	**20.6**	**45.2**	**57.8**	**63.3**	**19.8**	**43.7**	**56.5**	**62.2**

5 Conclusions

This paper proposed a noise resistible mutual-training method (NRMT) for unsupervised domain adaptation (UDA) in person re-ID to effectively depress label noises in a self-training process. In NRMT, two networks are maintained during training. For each iteration, these two networks share clustering information to ease the fitting to noisy instances. For each mini-batch update, the networks also exchange prediction information to further select both reliable and informative instances. Extensive experimental results demonstrate that the proposed NRMT achieves the state-of-the-art performance for UDA in person re-ID.

Acknowledgments This work was supported by the National Natural Science Foundation of China under Grant 61702163.

References

1. Arpit, D., Jastrzebski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y.: A closer look at memorization in deep networks. In: International Conference on Machine Learning (ICML) (2017)
2. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on Computational learning theory (1998)
3. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. In: Advances in neural information processing systems (NeurIPS) (2016)
4. Campello, R.J., Moulavi, D., Sander, J.: Density-based clustering based on hierarchical density estimates. In: Pacific-Asia conference on knowledge discovery and data mining (PAKDD) (2013)
5. Chen, Y., Zhu, X., Gong, S.: Instance-guided context rendering for cross-domain person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (2009)
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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
8. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) **14**(4) (2018)
9. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model (2008)
10. Feng, Q., Kang, G., Fan, H., Yang, Y.: Attract or distract: Exploit the margin of open set. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
11. 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: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
12. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. The Journal of Machine Learning Research (JMLR) **17**(1) (2016)
13. Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In: International Conference on Learning Representations (ICLR) (2020)
14. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: Advances in neural information processing systems (NeurIPS) (2018)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (2016)
16. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)

17. Li, Y.J., Yang, F.E., Liu, Y.C., Yeh, Y.Y., Du, X., Frank Wang, Y.C.: Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
18. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (2015)
19. Liu, Z., Wang, J., Gong, S., Lu, H., Tao, D.: Deep reinforcement active learning for human-in-the-loop person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
20. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning (ICML) (2015)
21. Lv, J., Chen, W., Li, Q., Yang, C.: Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
22. Malach, E., Shalev-Shwartz, S.: Decoupling "when to update" from "how to update". In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
23. Panareda Busto, P., Gall, J.: Open set domain adaptation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
24. Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., Tian, Y.: Unsupervised cross-dataset transfer learning for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (2016)
25. 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: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
26. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking (2016)
27. Saito, K., Yamamoto, S., Ushiku, Y., Harada, T.: Open set domain adaptation by backpropagation. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
28. Shu, R., Bui, H.H., Narui, H., Ermon, S.: A dirt-t approach to unsupervised domain adaptation. In: Proceedings of the International Conference on Learning Representations (ICLR) (2018)
29. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
30. Song, J., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.M.: Generalizable person re-identification by domain-invariant mapping network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
31. 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)
32. Suh, Y., Wang, J., Tang, S., Mei, T., Mu Lee, K.: Part-aligned bilinear representations for person re-identification. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
33. Sun, B., Feng, J., Saenko, K.: Return of frustratingly easy domain adaptation. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
34. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)

35. 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: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
36. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Simultaneous deep transfer across domains and tasks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
37. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
38. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
39. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
40. Wu, A., Zheng, W.S., Lai, J.H.: Unsupervised person re-identification by camera-aware similarity consistency learning. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
41. Xie, G.S., Liu, L., Jin, X., Zhu, F., Zhang, Z., Qin, J., Yao, Y., Shao, L.: Attentive region embedding network for zero-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
42. Xie, G.S., Liu, L., Zhu, F., Zhao, F., Zhang, Z., Qin, J., Yao, Y., Shao, L.: Region graph embedding network for zero-shot learning. In: Proceedings of the European Conference on Computer Vision (ECCV) (2020)
43. Yang, F., Li, K., Zhong, Z., Luo, Z., Sun, X., Cheng, H., Guo, X., Huang, F., Ji, R., Li, S.: Asymmetric co-teaching for unsupervised cross-domain person re-identification. In: Thirtieth AAAI Conference on Artificial Intelligence (AAAI) (2020)
44. Yang, Q., Yu, H.X., Wu, A., Zheng, W.S.: Patch-based discriminative feature learning for unsupervised person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
45. Yu, H.X., Wu, A., Zheng, W.S.: Unsupervised person re-identification by deep asymmetric metric embedding. IEEE transactions on pattern analysis and machine intelligence (TPAMI) (2018)
46. Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
47. Yu, X., Han, B., Yao, J., Niu, G., Tsang, I., Sugiyama, M.: How does disagreement help generalization against label corruption? In: International Conference on Machine Learning (ICML) (2019)
48. Zhang, K., Luo, W., Ma, L., Liu, W., Li, H.: Learning joint gait representation via quintuplet loss minimization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
49. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019)
50. 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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

51. Zhao, L., Li, X., Zhuang, Y., Wang, J.: Deeply-learned part-aligned representations for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
52. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
53. Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., Kautz, J.: Joint discriminative and generative learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
54. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
55. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. arXiv preprint arXiv:1708.04896 (2017)
56. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
57. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: Exemplar memory for domain adaptive person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
58. Zhou, S., Wang, J., Wang, J., Gong, Y., Zheng, N.: Point to set similarity based deep feature learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)