Unsupervised Multi-Source Domain Adaptation for Person Re-Identification

Zechen Bai[1], **Zhigang Wang**[1], **Jian Wang**[1], **Di Hu**[2,3*], **Errui Ding**[1*]
[1]Department of Computer Vision Technology (VIS), Baidu Inc., China,
[2]Gaoling School of Artificial Intelligence, Renmin University of China, Beijing 100872, China,
[3]Beijing Key Laboratory of Big Data Management and Analysis Methods
zechenbai.baidu@outlook.com, dihu@ruc.edu.cn,
{wangzhigang05,wangjian33,dingerrui}@baidu.com

Abstract

Unsupervised domain adaptation (UDA) methods for person re-identification (re-ID) aim at transferring re-ID knowledge from labeled source data to unlabeled target data. Although achieving great success, most of them only use limited data from a single-source domain for model pre-training, making the rich labeled data insufficiently exploited. To make full use of the valuable labeled data, we introduce the multi-source concept into UDA person re-ID field, where multiple source datasets are used during training. However, because of domain gaps, simply combining different datasets only brings limited improvement. In this paper, we try to address this problem from two perspectives, i.e. *domain-specific view and domain-fusion view. Two constructive modules are proposed, and they are compatible with each other. First, a rectification domain-specific batch normalization (RDSBN) module is explored to simultaneously reduce domain-specific characteristics and increase the distinctiveness of person features. Second, a graph convolutional network (GCN) based multi-domain information fusion (MDIF) module is developed, which minimizes domain distances by fusing features of different domains. The proposed method outperforms state-of-the-art UDA person re-ID methods by a large margin, and* ***even achieves comparable performance to the supervised approaches without any post-processing techniques****.*

1. Introduction

Person re-identification (re-ID) aims at retrieving images of a specified person across different cameras. Recently, supervised person re-ID methods [34, 37, 47] have made impressive progress. However, when testing on unseen datasets, they usually suffer from dramatic performance degradation. Collecting and annotating enough training data for a specified scenario is expected, but it is labor-intensive and expensive. Therefore, unsupervised domain adaptation (UDA) for person re-ID has attracted an increasing attention.

Numerous UDA person re-ID methods have been proposed. Among them, the pseudo-label-based branch [33, 41, 10] dominates the state-of-the-art methods. They usually consists of two steps: (1) obtaining a pre-trained model by supervised training on the source domain; (2) training on the target domain by pseudo-label prediction and fine-tune iteratively. Although pseudo-label-based methods are proved effective, most of them only use limited data from a single-source domain for model pre-training, making the rich labeled data insufficiently exploited. This is undoubtedly a great waste of resources.

To make full use of the existing abundant labeled data, we first introduce the multi-source concept into UDA person re-ID field, where multiple source datasets are used in both model pre-training and fine-tuning stages. For the latter, the ground-truth labels of source domain and pseudo labels of target domain together provide supervisions. However, we find simply combining different datasets brings limited improvement or even negative effect. This is because different datasets usually have different characteristics such as color, illumination, camera views, *etc*. This problem is widely known as domain gaps.

In this paper, we try to address this problem from two perspectives, *i.e*. domain-specific view and domain-fusion view. As for the former, several successful UDA methods [1, 45, 2] use specific network components to capture and eliminate incompatible characteristics of different domains. Nevertheless, most of these methods are originally designed for a close-set problem, *i.e*. general classification which has the following properties. (1) Multiple domains share all or part of the label set. (2) Samples have relatively small intra-class distances and large inter-class distances. Due to these two reasons, it is easy for these methods to correctly aggregate intra-class samples even they come from different do-

*Co-Corresponding Auther.

mains. On the contrary, person re-ID is a fine-grained open-set problem where person images present subtle differences and identities/classes are completely disjoint. Directly applying existing domain-specific methods to multi-domain person re-ID task may improperly aggregate identities from different domains. To tackle this problem, a rectification domain-specific batch normalization (RDSBN) module is proposed in this paper. Inspired by [2], we use individual BN branches to capture and reduce domain-specific information. Moreover, we exploit a rectification procedure that adaptively tunes BN parameters for each instance to enhance identity-related information and make features more discriminative.

As for the domain-fusion view, some works [21, 43, 22, 23] attempt to minimize domain distances by fusing feature distributions. Unfortunately, they also depend on the close-set setting. For re-ID problem, we develop a more flexible GCN-based multi-domain information fusion (MDIF) module to reduce domain distances. There are two main differences between previous GCN-based re-ID methods and our MDIF module. First, from motivation perspective, previous works [32, 14] leverage GCN to enhance feature representation, while our MDIF module aims at reducing domain gaps. Second, from working mechanism perspective, previous works mainly follow the formulation of ordinary GCN [17]. While in MDIF module, a domain-agent-node concept is proposed which weighted combines features of the same domain as a global representation. Then, our MDIF module enables information to propagate among domain-agent-nodes and instances to conduct domain fusion. During testing, domain-agent-nodes come from the recorded moving average values like BN, thus no extra computation is needed.

The contributions of this work can be summarized as three-fold. 1) We introduce the multi-source concept into UDA person re-ID field. To the best of our knowledge, this is a pioneering work to study the multi-source UDA problem in the person re-ID community. 2) We propose a rectification domain-specific batch normalization (RDSBN) module which can simultaneously reduce domain-specific information and improve the distinctiveness of person features. 3) We develop a GCN based multi-domain information fusion (MDIF) module to pull different domains close in feature space, and shed new light on the effect of GCN on reducing domain gaps.

In our experiments, we find the proposed two modules are compatible with each other, and the combination of them can outperform state-of-the-art methods by a large margin.

2. Related Work

2.1. Unsupervised Domain Adaptation for Person Re-ID

Mainstream unsupervised domain adaptation methods for person re-ID can be categorized into two branches. The first branch employs generative adversarial networks (GANs) to transfer the style of labeled images from a source domain to the target domain, and uses transferred images for training. Based on this pipeline, SPGAN [3, 4] propose to preserve similarity of images before and after image translation. Similarly, PTGAN [38] leverages semantic segmentation to constrain the consistency of human body regions during style transfer. Unfortunately, the performances of these methods are not satisfactory enough.

On the contrary, the pseudo-label-based branch dominates the state-of-the-art methods in recent years [6, 33, 7, 46, 41, 44, 10, 52, 19, 20]. PUL [6] iteratively obtains pseudo labels by clustering algorithms on the target domain and fine-tunes the model using generated labels. Song *et al*. [33] follow this paradigm and provide more theoretical analysis. SSG [7] uses both global and local features, and assign pseudo labels for them separately. PAST [46] progressively selects more reliable samples into training set. ACT [41] designs an asymmetric co-teaching framework to resist label noise generated by clustering algorithms. MMT [10] utilizes both hard and sort labels via a mutual mean-teaching framework. DG-Net++ [52] extracts and focuses on identity-related features by a disentangling module. Although these works have achieved great successes, they use only limited data from a single-source domain for model pre-training, making rich labeled data insufficiently exploited.

2.2. Multi-Source Domain Adaptation

Recent years, more multi-source domain adaptation methods have been studied for practical applications. Ganin *et al*. [9] tries to solve domain adaptation problem by adversarial learning. CMSS [42] learns a dynamic curriculum to decide which domains are best for aligning to the target. Peng *et al*. [29] aims to transfer knowledge by dynamically aligning moments of multi-domain feature distributions. DSBN [2] narrows domain gaps by incorporating different BN layers for different domains. However, most of them assume there are label overlappings among multiple domains, making them unsuitable for the open-set person re-ID task.

2.3. Feature Normalization

Recently, simultaneously fusing multiple normalization techniques becomes popular to boost generalization of netwroks. IBN-Net [28] manually designs fusion policy for batch normalization (BN) and instance normalization (IN).

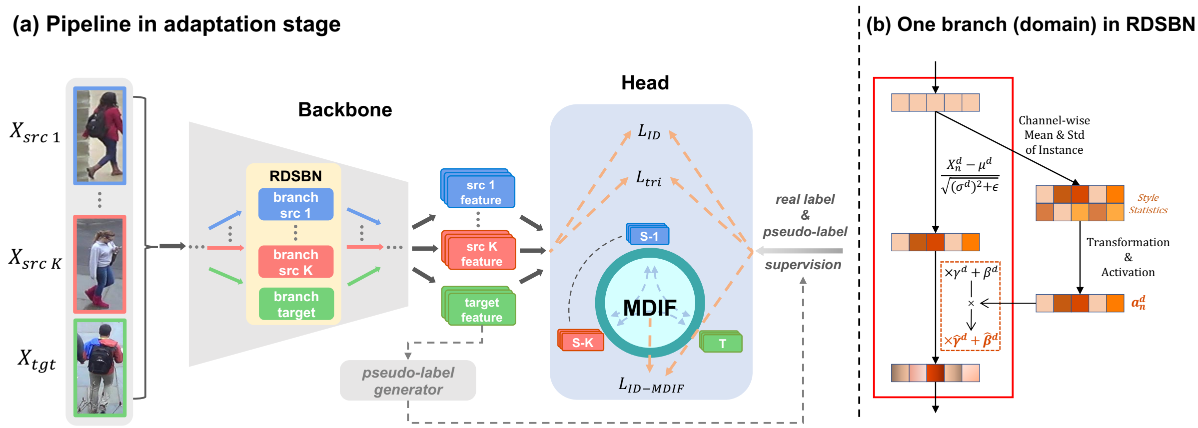

Figure 1. **(a)** The illustration of the proposed framework, including pseudo-label generator, backbone equipped with RDSBN and head equipped with MDIF. **(b)** The rectification operation in RDSBN. Best viewed in color.

BIN [27] learns adaptive gate weights to assign different importances to BN and IN. Switchable Normalization (SN) [25] extends this strategy to multiple normalization techniques. Whereas, when applying these methods to UDA person re-ID task, the performance is still unsatisfactory.

2.4. Graph Convolutional Network Related Methods

Several works [17, 35] concentrate on designing better GCN architectures to address graph-structured problems. Other works extend GCN to different applications, *e.g.* action recognition [40], anomaly detection [50], recommendation system [26] and supervised person re-ID [32]. Great successes are witnessed in these methods, while few works exploit the effect of GCN on reducing domain gaps.

3. Methodology

3.1. Overview

Under the context of multi-source unsupervised domain adaptive re-ID, we have K fully labeled source datasets $\mathcal{S}=\{\{Z^1,Y^1\},\{Z^2,Y^2\}, ... ,\{Z^K,Y^K\}\}$. Z^k and Y^k denote data samples and ground-truth labels in the k-th source domain, respectively. In addition, we have an unlabeled target dataset $\mathcal{T}=\{Z^{K+1}\}$. Our goal is to leverage labeled samples in $\mathcal{S}$ and unlabeled samples in $\mathcal{T}$ to learn a re-ID model that generalizes well in the target domain. Note that dataset and domain can be used interchangeably in this paper.

To tackle the aforementioned problem, we propose an unsupervised multi-source domain adaptation framework for person re-ID. As shown in Fig. 1 (a), the proposed framework can be decoupled into three parts, *i.e.* pseudo-label generator, backbone and head. Following most methods [33, 41], the clustering algorithm DBSCAN [5] is employed as our pseudo-label generator which takes all target features as input and outputs clusters as pseudo-labels before each epoch. The backbone is responsible for extracting domain-invariant features from images. It can be chosen from various popular networks. Afterwards, the regular BN layers are replaced by our RDSBN module that mitigates domain-specific characteristics and enhances distinctiveness of person features. Given extracted features, the network head with MDIF module is designed to perform information fusion among multiple domains to further reduce domain gaps. The details of RDSBN module, MDIF module and optimization of the proposed framework are described as follows.

3.2. Rectification Domain-Specific Batch Normalization

3.2.1 Batch Normalization Revisit

A BN layer normalizes features within a mini-batch and performs affine transformation using scale parameter $\boldsymbol{\gamma} \in \mathbb{R}^C$ and bias parameter $\boldsymbol{\beta} \in \mathbb{R}^C$. Let $\boldsymbol{X}_n \in \mathbb{R}^{C\times HW}$ denote the feature map of the n-th sample, where C, H and W indicate the number of channels, height and width, respectively. BN is formulated as

$$\mathbf{BN}(\boldsymbol{X}_n;\boldsymbol{\gamma},\boldsymbol{\beta}) = \boldsymbol{\gamma}\cdot\frac{\boldsymbol{X}_n-\boldsymbol{\mu}}{\sqrt{\boldsymbol{\sigma}^2+\epsilon}}+\boldsymbol{\beta}, \tag{1}$$

where mean value $\boldsymbol{\mu} \in \mathbb{R}^C$ and standard deviation $\boldsymbol{\sigma} \in \mathbb{R}^C$ are calculated with respect to a mini-batch, ϵ is a small constant avoiding divide-by-zero, $(\cdot)$ and $\sqrt{\cdot}$ are both channel-wise operations here.

Actually, during training, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are estimated through moving average operation with the update momentum α. Formally, we donote them by $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\sigma}}$. Given the t-th mini-batch, corresponding mean $\bar{\boldsymbol{\mu}}^t$ and standard deviation $\bar{\boldsymbol{\sigma}}^t$ are calculated by

$$\bar{\boldsymbol{\mu}}^t = (1-\alpha)\bar{\boldsymbol{\mu}}^{t-1}+\alpha\boldsymbol{\mu}^t, \tag{2}$$

$$(\bar{\boldsymbol{\sigma}}^t)^2 = (1-\alpha)(\bar{\boldsymbol{\sigma}}^{t-1})^2 + \alpha(\boldsymbol{\sigma}^t)^2. \tag{3}$$

The final $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\sigma}}$ are used for normalization in the testing stage.

3.2.2 Design of RDSBN

The regular BN is performed upon the whole mini-batch. However, sharing parameters across multiple domains is inappropriate due to the existence of domain gap [2]. Inspired by [2], the proposed RDSBN module also normalizes each domain separately using individual BN branches. Differently, we design a rectification procedure as shown in Fig. 1 (b). Let $\boldsymbol{X}_n^d \in \mathbb{R}^{C\times HW}$ denote a feature map of the n-th sample in d-th domain, RDSBN module can be written as

$$\mathbf{RDSBN}(\boldsymbol{X}_n^d; \boldsymbol{\gamma}^d, \boldsymbol{\beta}^d, \boldsymbol{a}_n^d) = \boldsymbol{a}_n^d \cdot \boldsymbol{\gamma}^d \cdot \frac{\boldsymbol{X}_n^d - \boldsymbol{\mu}^d}{\sqrt{(\boldsymbol{\sigma}^d)^2 + \epsilon}} + \boldsymbol{a}_n^d \cdot \boldsymbol{\beta}^d. \tag{4}$$

For simplicity, the moving average version is not introduced again. The superscript d indicates that calculations are carried out within a specific domain. $\boldsymbol{a}_n^d \in \mathbb{R}^C$ is a channel-wise weight that rectifies $\boldsymbol{\gamma}^d$ and $\boldsymbol{\beta}^d$ to rearrange features. This is expected to explicitly enhance the identity-related style information. In particular, inspired by [13], the style-related statistics, *i.e.* channel-wise mean $\boldsymbol{\mu}_n^d$ and standard deviation $\boldsymbol{\sigma}_n^d$ of each instance are adopted to estimate $\boldsymbol{a}_n^d$.

$$\boldsymbol{a}_n^d = Sigmoid(\mathbf{1}^M * (\boldsymbol{r}^d * [\boldsymbol{\mu}_n^d, \boldsymbol{\sigma}_n^d])), \tag{5}$$

where $Sigmoid(\cdot)$ means regular sigmoid activation function, $*$ indicates matrix multiplication, $\mathbf{1}^M \in \mathbb{R}^{1\times M}$ is an all-one matrix for dimension reduction, $\boldsymbol{r}^d \in \mathbb{R}^{M\times 2}$ is a learnable parameter, $[\cdot]$ denotes concatenate operation that combines $\boldsymbol{\mu}_n^d$ and $\boldsymbol{\sigma}_n^d$ into a two-dimension matrix, *i.e.* $[\boldsymbol{\mu}_n^d, \boldsymbol{\sigma}_n^d] \in \mathbb{R}^{2\times C}$.

The proposed rectification procedure is simple yet very effective, which plays an important role in increasing the distinctiveness of person features after reducing domain gaps. This is critical for the fine-grained open-set person re-ID problem. Experiments show the superiority of our RDSBN module. See section 4.3.2 for details.

3.3. Multi-Domain information Fusion

3.3.1 Domain-Agent-Node

We also attempt to fuse domain information by GCN to reduce domain distances. One naive way to apply GCN here is building a graph that connects instances from different domains. However, directly fusing instance-level features is prone to incur identity-related noises. Therefore, we propose a domain-agent-node as the global domain representation. It is obtained by weighted combining features of the same domain. Formally, the domain-agent-node is formulated as below

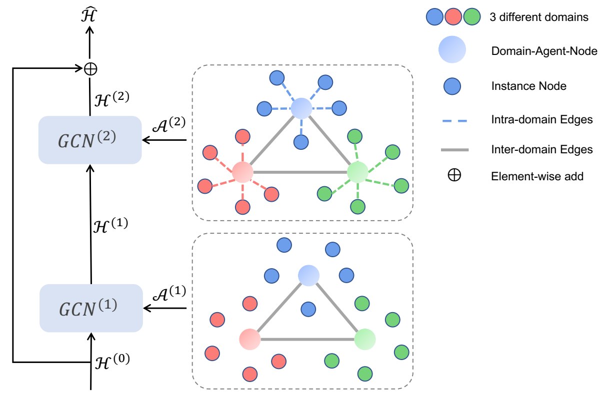

Figure 2. The illustration of GCN-based Multi-Domain Information Fusion module. For convenience, we present two source domains and one target domain here. Best viewed in color.

$$\boldsymbol{agt}^d = \sum_{n=1}^{Q} w_n^d \cdot \boldsymbol{X}_n^d, \tag{6}$$

where $\boldsymbol{agt}^d$ denotes the agent-node of the d-th domain, Q means the number of samples belonging to the same domain within a mini-batch, $w_n^d = f(\boldsymbol{X}_n^d)$ is a learnable scalar weight and normalized by $w_n^d = w_n^d / \sum_{m=1}^{Q} w_m^d$. $f(\cdot)$ indicates a fully connected layer.

In order to estimate the domain-agent-node more stably, we adopt the moving average technique during training like BN

$$(\bar{\boldsymbol{agt}}^d)^t = (1-\alpha)(\bar{\boldsymbol{agt}}^d)^{t-1} + \alpha(\bar{\boldsymbol{agt}}^d)^t, \tag{7}$$

where t and α have the same meanings with Eq. 2 and Eq. 3. Similar to BN, the recorded $\bar{\boldsymbol{agt}}^d$ is used for inference in the testing stage.

3.3.2 Graph construction

Let $G(V, E)$ denote a graph where V is the vertex (node) set and E is the edge set. As shown in Fig. 2, we stack two GCN layers to construct $G(V, E)$, in which the first layer is responsible for fusing global domain representations, and the second layer ensures all instances are capable of receiving information from other domains. Without loss of generality, assume there are 3 domains in a mini-batch, each domain contains Q samples and 1 domain-agent-node. Thus there are totally $3 \times (Q+1)$ nodes. We denote $\mathcal{H}^{(l)} \in \mathbb{R}^{(3Q+3)\times C}$ as node features where the superscript (l) indicates the l-th GCN layer. As for the edge set E, it can be represented by adjacency matrixes $\mathcal{A}^{(l)} \in \mathbb{R}^{(3Q+3)\times(3Q+3)}$ of two GCN layers. For inter-domain edges, the connections only exist among 3 domain-agent-nodes in both $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$. For intra-domain edges, each instance is connected to its affiliated domain-agent-node in $\mathcal{A}^{(2)}$. Note that the intra-connections are omitted in $\mathcal{A}^{(1)}$, since in the first layer of GCN, domain-agent-nodes have not incorporated information from each other and cannot provide fused features for corresponding intra-domain nodes.

Based on the above connection relationship, $\mathcal{A}^{(l)}$ can be formulated as

$$A_{i,j}^{(l)} = \begin{cases} 1 & i = j \text{ } or \text{ } node \text{ } i \text{ } connects \text{ } node \text{ } j \\ 0 & otherwise \end{cases}, \quad (8)$$

where i and j represent row index and column index of $\mathcal{A}^{(l)}$, respectively. $\mathcal{A}^{(l)}$ is then normalized by its degree matrix $D^{(l)}$

$$\tilde{\mathcal{A}^{(l)}} = (D^{(l)})^{-\frac{1}{2}} \mathcal{A}^{(l)} (D^{(l)})^{-\frac{1}{2}}, \quad (9)$$

where $D_{i,i}^{(l)} = \sum_j \mathcal{A}_{i,j}^{(l)}$. Afterwards, each GCN layer can be written as a non-linear transformation

$$\mathcal{H}^{(l)} = \rho(\tilde{\mathcal{A}}^{(l)} \mathcal{H}^{(l-1)} \mathcal{W}^{(l)}), \quad (10)$$

where $\mathcal{W}^{(l)} \in \mathbb{R}^{C \times C}$ is a learnable transformation matrix and ρ denotes a non-linear function that is LeakyReLU [39] in our method. Let $\mathcal{H}^{(0)} \in \mathbb{R}^{3Q \times C}$ denote the input features to our MDIF module, the corresponding $\mathcal{H}^{(2)}$ without domain-agent-node features is output and treated as residuals to be added to original features

$$\hat{\mathcal{H}} = \mathcal{H}^{(0)} + \mathcal{H}^{(2)}. \quad (11)$$

Finally, we use the fused features $\hat{\mathcal{H}}$ as sample representations. For more detailed explanation of the MDIF inference procedure, please refer to **Supplementary**.

3.4. Optimization

3.4.1 Two-Stage Training

Following the popular pipeline, we adopt a two-stage training scheme including source-model pre-training and domain adaptation fine-tuning.

For the first stage, a person re-ID model with DSBN layers is trained by optimizing ID loss (cross-entropy loss) and triplet loss [12]

$$\mathcal{L}_{src} = \mathcal{L}_{ID} + \mathcal{L}_{tri}. \quad (12)$$

Duo to the limitation of page-length and generality of these two losses, we omit their detailed formulation. Please refer to corresponding references for details.

For the second stage, the re-ID model is further equipped with RDSBN and MDIF modules: replacing DSBN with RDSBN and stacking MDIF module on top of the backbone as shown in Fig 1. Both ground-truth labels and pseudo labels provide supervisions in this stage. We select ID loss to train MDIF module. The original losses building upon the input features of MDIF module are retained to preserve their representativeness of identities. So that in the second stage, the total loss can be written as

$$\mathcal{L}_{ada} = \mathcal{L}_{ID} + \mathcal{L}_{ID-MDIF} + \mathcal{L}_{tri}. \quad (13)$$

4. Experiment

4.1. Datasets and Evaluation Metrics

We evaluate the proposed method on four widely-used person re-ID datasets: Market1501[49], DukeMTMC-reID[31], CUHK03[18] and MSMT[38]. Mean average precision (mAP) and CMC rank-1, rank-5, rank-10 accuracies are adopted to evaluate the performances.

4.2. Implementation Details

In both source-model pre-training and domain adaptation fine-tuning stages, ResNet50 [11] is adopted as the backbone network. Experiments are carried out on two NVIDIA-P100 GPUs. We follow the experimental settings of [10], details are listed as below. Adam optimizer [16] is utilized to optimize networks with a weight decay of 0.0005. The input image is uniformly resized to 256×128. Several data augmentation techniques are performed such as random flipping and random crop. Random erasing is only used in fine-tuning stage.

Stage1: source-model pre-training. In this stage, the ResNet50-DSBN network is initialized with ResNet50 weights pre-trained on ImageNet. Each training mini-batch contains 32 person images of 8 identities (4 for each identity). The learning rate is initialized as 0.00035 and is decreased to 1/10 of its previous value on the 40-th and 70-th epoch in total 80 epochs.

Stage2: domain adaptation fine-tuning. The proposed network ResNet50-RDSBN-MDIF is initialized with weights of the above ResNet50-DSBN. For each training mini-batch, we select $32 \times (K+1)$ person images (8 identities in each domain, K source datasets and 1 target dataset). The learning rate is fixed to 0.00035 for overall 40 training epochs. During testing, we only use the target domain RDSBN branch for feature extraction and fetch the fused features after MDIF module for inference.

4.3. Ablation Studies

4.3.1 Necessity of Multiple Source Domains

A key problem to this work is the necessity of introducing multiple source domains. Experiments results with respect to the number of source datasets are shown in Table 1. Along the 'Methods' column, 'Direct Transfer' means directly using the pre-trained source-model for testing. 'MMT (DBSCAN)' follows the setting of [10] where one or more source datasets are only used for pre-training. 'MMT-with-Source' further brings source data into fine-tuning stage. We observe that increasing the number of source domains brings stable performance improvements in 'Direct Transfer'. It indicates that introducing more source data is beneficial to the generalization ability of a re-ID model. However, simply combining multi-

Table 1. Ablation studies on different number of source domains with Market1501 as target domain.

Methods	1-source: Duke				2-source: Duke+CUHK				3-source: Duke+CUHK+MSMT			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
Direct Transfer	27.7	56.2	71.8	76.6	36.0	63.5	78.5	83.3	40.3	67.5	82.0	86.4
MMT (DBSCAN)	74.6	88.4	96.2	97.8	75.3	89.5	96.6	98.0	74.8	89.3	96.2	97.7
MMT-with-Source	68.9	87.1	95.0	96.7	75.3	89.6	96.5	97.6	77.7	90.8	96.9	97.9
MMT + Ours	**81.5**	**92.9**	**97.6**	**98.4**	**85.2**	**94.2**	**98.0**	**98.8**	**86.0**	**94.8**	**97.9**	**98.6**

Table 2. Ablation studies on RDSBN module. Due to space limitation, we only show a portion of the results here and more comparisons can be found in **Supplementary**.

Methods	Duke+CUHK→Market			
	mAP	R-1	R-5	R-10
ResNet50-*BN* (baseline)	70.1	86.7	95.0	96.5
ResNet50-*BIN*	70.5	87.8	94.9	96.5
ResNet50-*IBN*	70.7	88.0	95.3	97.3
ResNet50-*SN*	65.8	84.8	93.6	95.8
ResNet50-*DSBN*	74.1	89.2	96.2	**97.9**
ResNet50-*RDSBN*	**78.2**	**91.0**	**96.8**	**97.9**

ple source datasets ('MMT (DBSCAN)' and 'MMT-with-Source') brings only limited improvements or even negative effect. In comparison, our method can boost performance significantly when using more datasets. In a nutshell, introducing multiple source datasets is beneficial, and our method can effectively leverage these data.

4.3.2 Effectiveness of RDSBN

To evaluate the effectiveness of our RDSBN module, we compare it with the most related DSBN [2] and several powerful normalization techniques in Table 2. For simplicity, two source datasets and one target dataset are used for verification. All methods in this experiment are built upon the most ordinary pseudo-label-based UDA pipeline (Sec. 1), and all of them use source data in domain adaptation fine-tuning stage for fair comparison. We first create a baseline model equipped with regular BN, which has already achieved a good result. BIN [27] and IBN [28] further boost the performance due to their abilities of increasing generalization. However, the improvement is not significant, SN [25] even causes performance degradation. DSBN outperforms the above methods, which shows advantage of domain-specific design in dealing with multi-domain problem. However, DSBN is not designed for the fine-grained open-set re-ID problem. In comparison, our RDSBN module considerably improves the results, which outperforms DSBN by 4.1% mAP and 1.8% top-1 accuracy. This result demonstrates the importance and effectiveness of the proposed rectification procedure.

Table 3. Ablation studies on MDIF module. See **Supplementary** for more results.

Methods	Duke+CUHK→Market			
	mAP	R-1	R-5	R-10
ResNet50-*BN* (baseline)	70.1	86.7	95.0	96.5
ResNet50-*BN-MDIF*	73.2	88.4	95.2	97.1
ResNet50-*DSBN*	74.1	89.2	96.2	97.9
ResNet50-*DSBN-MDIF*	78.3	91.3	96.7	97.9
ResNet50-*RDSBN*	78.2	91.0	96.8	97.9
ResNet50-*RDSBN-MDIF*	79.4	92.1	97.1	98.1

Table 4. Analysis of MDIF module design. See **Supplementary** for more results.

Methods	Duke+CUHK→Market			
	mAP	R-1	R-5	R-10
baseline	70.1	86.7	95.0	96.5
Instance Graph	72.4	87.6	94.2	96.8
Mean Agent Node	72.9	88.0	94.9	96.8
Weighted Mean Agent Node	**73.2**	**88.4**	**95.2**	**97.1**
intra-connections in $A^{(1)}$	73.1	87.7	95.1	97.0

$$
\underset{\text{without MDIF}}{\begin{bmatrix} 0 & 1.37 & 1.41 \\ 1.37 & 0 & 1.63 \\ 1.41 & 1.63 & 0 \end{bmatrix}} \longrightarrow \underset{\text{with MDIF}}{\begin{bmatrix} 0 & 1.07_{(-0.3)} & 1.16_{(-0.25)} \\ 1.07_{(-0.3)} & 0 & 1.15_{(-0.48)} \\ 1.16_{(-0.25)} & 1.15_{(-0.48)} & 0 \end{bmatrix}}
$$

Figure 3. Pair-wise Euclidean distances among three domains: Market, Duke and CUHK03.

4.3.3 Effectiveness of MDIF and Structure Analysis

Performances of the proposed MDIF module under different model settings are reported in Table 3. We can observe that placing MDIF module on different base models always bring obvious improvements. This shows the effectiveness and robustness of the MDIF module.

Another experiment is conducted to directly validate whether MDIF module can decrease domain distances. We use the average feature of the whole domain as its representation. Pair-wise Euclidean distances among three domains are calculated and shown in Fig. 3. It can be seen that the proposed MDIF module reduces domain distances to a large extent.

Table 5. Comparison with state-of-the-art methods about unsupervised domain adaptive re-ID. (*) indicates the implementation is based on the authors' code.

Methods		Duke→Market				Market→Duke			
		mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
Single-Source *UDA re-ID methods:*									
PAST [46]	ICCV'19	54.6	78.4	-	-	54.3	72.4	-	-
SSG [7]	ICCV'19	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
ECN++ [51]	TPAMI'20	63.8	84.1	92.8	95.4	54.4	74.0	83.7	87.4
MMCL [36]	CVPR'20	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
SNR [15]	CVPR'20	61.7	82.8	-	-	58.1	76.3	-	-
AD-Cluster [44]	CVPR'20	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
DG-Net++ [52]	ECCV'20	61.7	82.1	90.2	92.7	63.8	78.9	87.8	90.4
NRMT [48]	ECCV'20	71.7	87.8	94.6	96.5	62.2	77.8	86.9	89.5
MMT [10] (DBSCAN)*	ICLR'20	74.6	88.4	96.2	97.8	61.0	75.1	87.3	91.2
Baseline	-	58.7	82.4	90.7	94.2	49.2	67.5	80.8	86.2
Baseline+**Ours**	-	68.4	85.9	94.3	96.5	55.8	72.9	84.4	88.1
MMT+**Ours**	-	**81.5**	**92.9**	**97.6**	**98.4**	**66.6**	**80.3**	**89.1**	**92.6**
Multi-Source *UDA methods (source+CUHK03+MSMT):*									
MMT+GRL* [8]	ICML'15	77.1	90.4	96.8	97.8	64.3	77.6	88.1	91.3
MMT+MomentMatching* [30]	ICCV'19	78.1	91.2	96.8	98.0	64.2	77.6	87.9	91.4
MMT+DSBN* [2]	CVPR'19	81.1	92.8	97.3	98.5	65.6	79.6	89.1	92.1
MMT+**Ours**	-	**86.0**	**94.8**	**97.9**	**98.6**	**68.9**	**82.1**	**90.4**	**93.0**
Supervised *methods:*									
PCB [34]	ECCV'18	81.6	93.8	97.5	98.5	69.2	83.3	-	-
bag-of-tricks [24]	CVPRW'19	85.9	94.5	98.2	99.0	76.4	86.4	93.9	96.1

Methods		Market→MSMT				Duke→MSMT			
		mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
Single-Source *UDA re-ID methods:*									
SSG [7]	ICCV'19	13.2	31.6	-	49.6	13.3	32.2	-	51.2
ECN++ [51]	TPAMI'20	15.2	40.4	53.1	58.7	16.0	42.5	55.9	61.5
MMCL [36]	CVPR'20	15.1	40.8	51.8	56.7	16.2	43.6	54.3	58.9
DG-Net++ [52]	ECCV'20	22.1	48.4	60.9	66.1	22.1	48.8	60.9	65.9
MMT [10] (DBSCAN)*	ICLR'20	25.7	51.9	65.3	70.9	28.1	56.1	68.9	74.3
Baseline	-	8.1	19.9	29.6	35.0	9.3	23.4	33.5	39.4
Baseline+**Ours**	-	20.7	46.9	60.0	65.0	21.3	47.7	60.8	66.4
MMT+**Ours**	-	30.9	61.2	73.1	77.4	33.6	64.0	75.6	79.6
Multi-Source *UDA methods (Market+Duke+CUHK03):*									
MMT+GRL* [8]	ICML'15	22.6	46.3	59.8	65.8	22.6	46.3	59.8	65.8
MMT+MomentMatching* [30]	ICCV'19	22.7	47.1	60.4	65.8	22.7	47.1	60.4	65.8
MMT+DSBN* [2]	CVPR'19	22.6	49.2	62.1	67.4	22.6	49.2	62.1	67.4
MMT+**Ours**	-	**34.9**	**64.7**	**76.2**	**80.2**	**34.9**	**64.7**	**76.2**	**80.2**

The effect of domain-agent-node is also analyzed. As shown in Table 4, we compare it with the 'Instance Graph' that directly constructs graph based on all instances within a mini-batch. The adjacency matrix of 'Instance Graph' is constructed by calculating cosine similarities of instances. Moreover, we explore two variants of the domain-agent-node, *i.e*. 'Mean Agent Node' and 'Weighted Mean Agent Node'. The former directly averages node features , while the latter weighted combines node features. Both of them outperform 'Instance Graph', indicating the proposed domain-agent-node can better represent a domain and promote domain information fusion. Finally, the better one, *i.e*. 'Weighted Mean Agent Node' is selected in our method. Based on 'Weighted Mean Agent Node', a detailed experiment is conducted to verify the irrationality of intra-connections in the first GCN layer of MDIF module. The inferior result validates our assumption, namely, domain-agent-nodes have not been fused at this time and should not

Table 6. Evaluate domain adaptation model on source domains with market1501 as target dataset.

Methods	Dukemtmc				CUHK03				MSMT			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
PCB[34]	69.2	83.3	-	-	57.5	63.7	-	-	-	-	-	-
bag-of-tricks[24]	76.4	86.4	93.9	96.1	-	-	-	-	-	-	-	-
Pre-trained model	65.4	79.5	89.6	92.6	61.4	64.0	77.0	87.5	33.1	60.4	74.6	79.9
MMT (DBSCAN)	16.2	30.0	43.9	51.0	-	-	-	-	-	-	-	-
MMT (3 Sources + 1 target)	72.5	82.8	91.9	95.0	82.9	85.0	93.0	96.0	46.0	71.2	83.1	86.8
Ours (3 Sources + 1 target)	**78.1**	**88.1**	**94.4**	**96.3**	**87.3**	**90.5**	**94.0**	**97.0**	**54.2**	**78.3**	**87.8**	**90.8**

propagate information to their intra-domain nodes.

4.4. Comparison with state-of-the-arts

To further prove the superiority of the proposed method, we compare it with several state-of-the-art methods on four domain adaptation tasks as MMT [10]: Market-to-Duke, Duke-to-Market, Market-to-MSMT and Duke-to-MSMT. The results are shown in Table 5. For fair comparison, we first conduct single-source domain adaptation experiments. Even though, our method significantly improves the baseline. When combining our method with the state-of-the-art work MMT, it outperforms all comparison UDA person re-ID works by a large margin. Specifically, our method achieves 81.5% and 66.6% mAP on Market-to-Duke and Duket-to-Market tasks, which outperforms MMT by 6.9% and 5.6%, respectively. This result indicates that our method can bring stable improvements even on the state-of-the-art method.

When extending our method to the multi-source version, another round of performance boost can be seen. Since no previous work studies multi-source UDA re-ID problem, we apply several general multi-source UDA methods to the state-of-the-art work MMT to compare with ours. It can be seen that our method is still superior to these comparison methods under the multi-source setting. In particular, when selecting the large scale dataset MSMT as target, our method outperforms them by 12+% mAP and 15+% top-1 accuracy. Moreover, in some tasks, our unsupervised results even **achieve comparable results with some popular fully-supervised methods** such as PCB [34] and bag-of-tricks [24]. This result again verifies the effectiveness of our method. Note that no extra post-processing technique is adopted.

4.5. Further Improvements on Source Domain

Finally, we study the model performance on source-domain testsets after domain adaptation, and list related results in Table 6, where 'Pre-trained model' is obtained by supervised training on the corresponding single-source dataset. From the rows of 'Pre-trained model' and 'MMT (DBSCAN)', we can see that state-of-the-art UDA person re-ID methods usually forget the source-domain knowledge after fine-tuning the pre-trained model on the target domain. Employing source data into domain adaptation fine-tuning stage is much better as shown in 'MMT (3 Sources + 1 target)'. In comparison, the proposed method adapt to multiple domains more effectively, which further improves the result. Our domain adaption model is obviously superior to pre-trained model, and even outperforms popular supervised re-ID methods [34, 24]. Such a phenomenon indicates that our method could also be applied to improve the supervised training.

5. Discussion about the Module Compatibility

In aforementioned experiments, RDSBN and MDIF modules show good compatibility. We consider that they complement each other. Specifically, on one hand, RDSBN module cannot completely eliminate domain-specific information, MDIF helps to further reduce domain gaps by fusing features. On the other hand, directly applying MDIF module is prone to incur noisy features since the original domain gaps are very large. RDSBN module may just establish a favourable condition for MDIF module.

6. Conclusion

In this work, we propose a multi-source framework for the unsupervised domain adaptive person re-ID task. Under the multi-source setting, simply combining different datasets brings limited improvement due to domain gaps. We propose to alleviate this problem from domain-specific view and domain-fusion view. Correspondingly, two constructive modules are developed, *i.e.* RDSBN module and MDIF module. The former can reduce domain-specific characteristics and enhance distinctiveness of person features. The latter explores domain-agent-nodes to represent global domain information and uses a GCN structure to fuse features from different domains. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art UDA person re-ID methods by a large margin, and even achieve comparable results to some popular fully supervised works.

References

[1] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 343–351, 2016. 1

[2] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 7354–7362, 2019. 1, 2, 4, 6, 7

[3] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, 2018. 2

[4] Weijian Deng, Liang Zheng, Qixiang Ye, Yi Yang, and Jianbin Jiao. Similarity-preserving image-image domain adaptation for person re-identification. *CoRR*, abs/1811.10551, 2018. 2

[5] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA*, pages 226–231, 1996. 3

[6] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Trans. Multim. Comput. Commun. Appl.*, 14(4):83:1–83:18, 2018. 2

[7] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 2, 7

[8] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015. 7

[9] Yaroslav Ganin and Victor S. Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37, pages 1180–1189, 2015. 2

[10] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations*, 2020. 1, 2, 5, 7, 8

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016. 5

[12] Alexander Hermans*, Lucas Beyer*, and Bastian Leibe. In Defense of the Triplet Loss for Person Re-Identification. *arXiv preprint arXiv:1703.07737*, 2017. 5

[13] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017. 4

[14] Bo Jiang, Xixi Wang, and Bin Luo. PH-GCN: person re-identification with part-based hierarchical graph convolutional network. *CoRR*, 2019. 2

[15] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. *CVPR*, 2020. 7

[16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 5

[17] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017. 2, 3

[18] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014. 5

[19] Yutian Lin, Yu Wu, Chenggang Yan, Mingliang Xu, and Yi Yang. Unsupervised person re-identification via cross-camera similarity exploration. *IEEE Transactions on Image Processing*, 29:5481–5490, 2020. 2

[20] Yutian Lin, Lingxi Xie, Yu Wu, Chenggang Yan, and Qi Tian. Unsupervised person re-identification via softened similarity learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3390–3399, 2020. 2

[21] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015. 2

[22] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *Advances in neural information processing systems*, pages 136–144, 2016. 2

[23] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017. 2

[24] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 7, 8

[25] Ping Luo, Jiamin Ren, Zhanglin Peng, Ruimao Zhang, and Jingyu Li. Differentiable learning-to-normalize via switchable normalization. *International Conference on Learning Representation (ICLR)*, 2019. 3, 6

[26] Chen Ma, Liheng Ma, Yingxue Zhang, Jianing Sun, Xue Liu, and Mark Coates. Memory augmented graph neural networks for sequential recommendation. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 5045–5052, 2020. 3

[27] Hyeonseob Nam and Hyo-Eun Kim. Batch-instance normalization for adaptively style-invariant neural networks. In *Ad-*

vances in Neural Information Processing Systems, 2018. 3, 6

[28] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 2, 6

[29] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 1406–1415, 2019. 2

[30] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1406–1415, 2019. 7

[31] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 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. 5

[32] Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen, and Xiaogang Wang. Person re-identification with deep similarity-guided graph neural network. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XV*, volume 11219, pages 508–526, 2018. 2, 3

[33] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, 102:107173, 2020. 1, 2, 3

[34] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and A strong convolutional baseline). In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV*, volume 11208, pages 501–518, 2018. 1, 7, 8

[35] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018. 3

[36] Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *CVPR*, 2020. 7

[37] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018*, pages 274–282, 2018. 1

[38] L. Wei, S. Zhang, W. Gao, and Q. Tian. Person transfer gan to bridge domain gap for person re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 79–88, 2018. 2, 5

[39] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015. 5

[40] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 7444–7452, 2018. 3

[41] Fengxiang Yang, Ke Li, Zhun Zhong, Zhiming Luo, Xing Sun, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Shaozi Li. Asymmetric co-teaching for unsupervised cross-domain person re-identification. In *AAAI*, pages 12597–12604, 2020. 1, 2, 3

[42] Luyu Yang, Yogesh Balaji, Ser-Nam Lim, and Abhinav Shrivastava. Curriculum manager for source selection in multi-source domain adaptation. *CoRR*, abs/2007.01261, 2020. 2

[43] Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, and Susanne Saminger-Platz. Central moment discrepancy (cmd) for domain-invariant representation learning. *arXiv preprint arXiv:1702.08811*, 2017. 2

[44] Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 9018–9027, 2020. 2, 7

[45] Weichen Zhang, Wanli Ouyang, Wen Li, and Dong Xu. Collaborative and adversarial network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3801–3809, 2018. 1

[46] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8222–8231, 2019. 2, 7

[47] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Relation-aware global attention for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[48] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. 2020. 7

[49] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, page 1116–1124, 2015. 5

[50] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H. Li, and Ge Li. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[51] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Learning to adapt invariance in memory for person re-identification. *TPAMI*, 2020. 7

[52] Yang Zou, Xiaodong Yang, Zhiding Yu, B. V. K. Vijaya Kumar, and Jan Kautz. Joint disentangling and adaptation for cross-domain person re-identification. *CoRR*, abs/2007.10315, 2020. 2, 7