Unsupervised Multi-Source Domain Adaptation for Person Re-Identification

Zechen Bai¹, Zhigang Wang¹, Jian Wang¹, Di Hu²,³*, Errui Ding¹*¹
1Department of Computer Vision Technology (VIS), Baidu Inc., China,
2Gaoling School of Artificial Intelligence, Renmin University of China, Beijing 100872, China,
3Beijing 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-
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 networks. IBN-Net [28] manually designs fusion policy for batch normalization (BN) and instance normalization (IN).
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 toUDA 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 $S = \{Z^1, Y^1\}, \{Z^2, Y^2\}, \ldots, \{Z^K, Y^K\}$. $Z^k$ 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 $T = \{Z^{K+1}\}$. Our goal is to leverage labeled samples in $S$ and unlabeled samples in $T$ 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 $\gamma \in \mathbb{R}^C$ and bias parameter $\beta \in \mathbb{R}^C$. Let $X_n \in \mathbb{R}^{C \times H \times W}$ 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

$$
BN(X_n; \gamma, \beta) = \gamma \cdot \frac{X_n - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta, \quad (1)
$$

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

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

$$
\mu^t = (1 - \alpha) \hat{\mu}^{t-1} + \alpha \mu^t, \quad (2)
$$

$$
\sigma^t = (1 - \alpha) \hat{\sigma}^{t-1} + \alpha \sigma^t. \quad (3)
$$

---

**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.
\[(\sigma^t)^2 = (1 - \alpha)(\sigma^{t-1})^2 + \alpha(\sigma^t)^2. \]

The final \( \mu \) and \( \sigma \) 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 \( \mathbf{X}_n^d \in \mathbb{R}^{C \times HW} \) denote a feature map of the \( n \)-th sample in \( d \)-th domain, RDSBN module can be written as

\[
\text{RDSBN}(\mathbf{X}_n^d; \gamma^d, \beta^d, \alpha^d_n) = \alpha^d_n \cdot \gamma^d \cdot \frac{\mathbf{X}_n^d - \mu^d_n}{\sqrt{(\sigma^d_n)^2 + \epsilon}} + \alpha^d_n \cdot \beta^d.
\]

(4)

For simplicity, the moving average version is not introduced again. The superscript \( d \) indicates that calculations are carried out within a specific domain. \( \alpha^d_n \in \mathbb{R}^C \) is a channel-wise weight that rectifies \( \gamma^d \) and \( \beta^d \) 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 \( \mu^d_n \) and standard deviation \( \sigma^d_n \) of each instance are adopted to estimate \( \alpha^d_n \).

\[
\alpha^d_n = \text{Sigmoid}(1^M \ast (r^d \ast [\mu^d_n, \sigma^d_n])),
\]

(5)

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

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

\[
\text{agt}^d = \sum_{n=1}^{Q} w_n^d \cdot \mathbf{X}_n^d,
\]

(6)

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

\[
(\text{agt}^d)^t = (1 - \alpha)(\text{agt}^d)^{t-1} + \alpha(\text{agt}^d)^t,
\]

(7)

where \( t \) and \( \alpha \) have the same meanings with Eq. 2 and Eq. 3. Similar to BN, the recorded \( \text{agt}^d \) 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 matrices \( \mathcal{A}^{(l)} \in \mathbb{R}^{(3Q+3) \times (3Q+3)} \) 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

![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.](image-url)
the intra-connections are omitted in $A^{(1)}$, 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, $A^{(i)}$ can be formulated as

$$A^{(i)}_{i,j} = \begin{cases} 1 & i = j \text{ or node } i \text{ connects node } j \\ 0 & \text{otherwise} \end{cases},$$

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

$$\tilde{A}^{(i)} = \left(D^{(i)}\right)^{-\frac{1}{2}} A^{(i)} \left(D^{(i)}\right)^{-\frac{1}{2}},$$

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

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

where $W^{(l)} \in \mathbb{R}^{C \times C}$ is a learnable transformation matrix and $\rho$ denotes a non-linear function that is LeakyReLU 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

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

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.

$$L_{src} = L_{ID} + L_{tri}. $$

Due 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

$$L_{ada} = L_{ID} + L_{ID-MDIF} + L_{tri}.$$
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.

Figure 3. Pair-wise Euclidean distances among three domains: Market, Duke and CUHK03.
**Table 5.** Comparison with state-of-the-art methods about unsupervised domain adaptive re-ID. (*) indicates the implementation is based on the authors’ code.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Duke→Market</th>
<th>Market→Duke</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mAP</td>
<td>R-1</td>
</tr>
<tr>
<td><strong>Single-Source UDA re-ID methods:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAST [46]</td>
<td>ICCV’19</td>
<td>54.6</td>
</tr>
<tr>
<td>SSG [7]</td>
<td>ICCV’19</td>
<td>58.3</td>
</tr>
<tr>
<td>ECN++ [51]</td>
<td>TPAMI’20</td>
<td>63.8</td>
</tr>
<tr>
<td>MMCL [36]</td>
<td>CVPR’20</td>
<td>60.4</td>
</tr>
<tr>
<td>SNR [15]</td>
<td>CVPR’20</td>
<td>61.7</td>
</tr>
<tr>
<td>AD-Cluster [44]</td>
<td>CVPR’20</td>
<td>68.3</td>
</tr>
<tr>
<td>DG-Net++ [52]</td>
<td>ECCV’20</td>
<td>61.7</td>
</tr>
<tr>
<td>NRMT [48]</td>
<td>ECCV’20</td>
<td>71.7</td>
</tr>
<tr>
<td>MMT [10] (DBSCAN)*</td>
<td>ICLR’20</td>
<td>74.6</td>
</tr>
<tr>
<td>Baseline</td>
<td>-</td>
<td>58.7</td>
</tr>
<tr>
<td>Baseline+Ours</td>
<td>-</td>
<td>68.4</td>
</tr>
<tr>
<td>MMT+Ours</td>
<td>-</td>
<td><strong>81.5</strong></td>
</tr>
<tr>
<td><strong>Multi-Source UDA methods (source+CUHK03+MSMT):</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMT+GRL* [8]</td>
<td>ICML’15</td>
<td>77.1</td>
</tr>
<tr>
<td>MMT+MomentMatching* [30]</td>
<td>ICCV’19</td>
<td>78.1</td>
</tr>
<tr>
<td>MMT+DSBN* [2]</td>
<td>CVPR’19</td>
<td>81.1</td>
</tr>
<tr>
<td>MMT+Ours</td>
<td>-</td>
<td><strong>86.0</strong></td>
</tr>
<tr>
<td><strong>Supervised methods:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCB [34]</td>
<td>ECCV’18</td>
<td>81.6</td>
</tr>
<tr>
<td>bag-of-tricks [24]</td>
<td>CVPRW’19</td>
<td>85.9</td>
</tr>
</tbody>
</table>

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
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 Duke-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


