Supplementary Material: 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

1. Details about Datasets

We evaluate the proposed method on four widely-used person re-ID datasets: Market1501[7], DukeMTMC-reID[5], CUHK03[4] and MSMT[6].

Market1501 contains 32,217 images of 1,501 identities captured by six cameras, where 12,936 images of 751 identities are for training, the remaining 19,281 images of 750 identities are for testing. The test set is further divided into a query set consisting of 3,368 images and a gallery set consisting of 15,913 images.

DukeMTMC-reID is collected in campus with eight cameras, which contains 36,411 images of 1,404 identities in total. Each identity is captured by at least two cameras. The train set contains 16,522 images of 702 identities. The rest data are for testing where the query set contains 2,228 images and the gallery set contains 17,661 images.

CUHK03 contains 28,193 images of 1,467 identities captured by two cameras. 26,263 images of 1,367 identities are used for training. The query set and gallery set have 200 and 1,730 images of the remaining 100 identities, respectively.

MSMT consists of 126,441 images of 4,101 identities captured by 15 cameras. The train set contains 32,621 images of 1,041 identities. The remaining images of 3,060 identities are assigned to query set (11,659 images) and gallery set (82,161 images).

2. Details of MDIF inference procedure

Fig. 1 shows the graph structure of MDIF module in inference mode. Features are calculated according to this structure, and we fetch the $\hat{\mathcal{H}}$ as the fused features. With respect to the graph information, (1) as explained in Section 3.3.1 in our paper, we use the recorded moving average values of domain-agent-nodes in inference. Therefore, domain-agent-nodes will not be affected by target samples, and edges among domain-agent-nodes are also fixed. (2) While please note that the whole graph structure is not necessary to be fixed. As for Q, *i.e.* the batch size in inference, it can be set to any number within the limitation of hard devices. The adjacency matrix $\mathcal{A}^{(l)}$ depicts connection relationships among graph nodes, it can be automatically tuned according to Q. In addition, the feature of a certain image will not be affected by other images since there is no connection between two instance nodes.

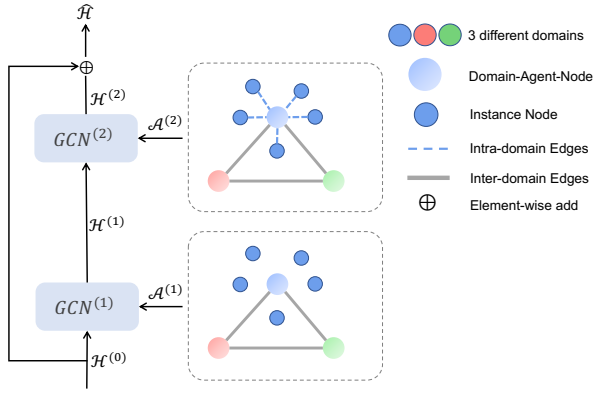

Figure 1. The illustration of GCN-based MDIF module in inference mode. For simplicity, we present two source domains (red and green) and one target domain (blue) here.

Table 1. Ablation studies on RDSBN module under another source-target setting.

Methods	Market+CUHK→Duke			
	mAP	R-1	R-5	R-10
ResNet50-*BN* (baseline)	58.3	72.8	84.4	88.2
ResNet50-*BIN*	59.2	73.1	84.1	88.0
ResNet50-*IBN*	61.9	75.9	87.6	90.6
ResNet50-*SN*	56.7	73.5	84.0	87.7
ResNet50-*DSBN*	60.4	76.2	86.1	90.1
ResNet50-*RDSBN*	**63.5**	**77.3**	**88.3**	**91.6**

3. More Ablation Study Experiments

To comprehensively validate the effectiveness and robustness of the proposed method, more ablation study experiments under another source-target setting are conducted and shown in this section. In accordance with previous experiments, two source datasets and one target dataset are employed for ablation study.

3.1. Effectiveness of RDSBN

As can be seen from Table 1, the observations are consistent with previous results, *i.e*. the proposed RDSBN module is obviously superior to other normalization methods. Specifically, our method outperforms DSBN [1], a recent and powerful domain adaptation work, by 3.1% mAP and 1.1% rank-1 accuracy. This again verifies the effectiveness of our RDSBN module.

Practically, the feature distance is usually used to measure whether two samples belong to the same identity or not. Thus we further validate the effect of RDSBN module by analyzing inter-class and intra-class feature distances. The ability of *pushing inter-class samples away* and *pulling intra-class samples close* is a widely-used criterion in representation learning [8, 3]. To prove the *'pushing away'* ability, we use the mean feature of samples to represent the corresponding identity, and calculate inter-class distance by averaging the pair-wise distances of all persons. The calculation can be formulated as follows.

$$\overline{D} = \frac{1}{M} \sum_{i,j \in Y} d(i,j), \tag{1}$$

where $\overline{D}$ denotes the inter-class distance, $d(i,j)$ represents the Euclidean feature distance between identity i and j, Y is the label set and M is the number of pair-wise identity distances. A larger inter-class distance indicates that the *'pushing away'* ability is stronger and identities are easier to distinguish. To prove the *'pulling close'* ability, we first calculate the feature variance of each identity, then measure the intra-class variance by averaging these variances. This process can be formulated by

$$\sigma_i^2 = \sum_{k=1}^{K} (X_{i,k} - \mu_i)^T (X_{i,k} - \mu_i), \tag{2}$$

$$\overline{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2, \tag{3}$$

where $X_{i,k}$, σ_i^2, μ_i and K denote the k-th feature, feature variance, mean feature and sample number of identity i, $\overline{\sigma^2}$ denotes the intra-class variance, N is the number of identities here. A smaller intra-class variance indicates that the *'pulling close'* ability is stronger and intra-class samples are more compact in feature space.

The corresponding experimental results are shown in Table 2, in which the Combined Dataset denotes the combination of Market, Duke and CUHK03. Compared to DSBN, the proposed RDSBN module achieves larger inter-class distance and smaller intra-class variance in all datasets. This result demonstrates that our RDSBN module can extract better features for re-ID task. Note that MDIF module is not used in this experiment.

3.2. Effectiveness of MDIF.

Performances of the proposed MDIF module under different model settings are reported in Table 4. We can observe that placing MDIF module on different base models always boosts performance significantly. Applying MDIF module to baseline achieves 4.4% mAP and 4.1% rank-1 accuracy improvement. Even building upon our RDSBN module, MDIF also brings 1.5% mAP and 1.8% rank-1 accuracy improvement.

The effect of domain-agent-node is also validated in Market+CUHK→Duke task. As shown in Table 5, the proposed domain-agent-node, especially 'Weighted Mean Agent Node', outperforms other variants stably. In addition, the inferior result of 'intra-connections in $A^{(1)}$' verifies again that it is reasonable to omit intra-connections in the first GCN layer of MDIF module.

3.3. Effectiveness of the Proposed Method on Different Base Frameworks

To further confirm the effectiveness and robustness of the proposed method, we apply it to the ordinary baseline framework and the state-of-the-art MMT [2] framework, respectively. Results are shown in Table 3. We can see that our method achieves 9.3% mAP and 5.4% rank-1 accuracy improvement on baseline, and 9.9% mAP and 4.6% rank-1 accuracy improvement on MMT. The proposed method also outperforms their variants equipped with DSBN . These stable improvements demonstrate the superiority of our method.

References

[1] 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. 2

[2] 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. 2

[3] Jiabo Huang, Qi Dong, Shaogang Gong, and Xiatian Zhu. Unsupervised deep learning by neighbourhood discovery. *arXiv preprint arXiv:1904.11567*, 2019. 2

Table 2. Analysis about inter-class distance and intra-class variance.

Metrics & Methods	Market	Duke	CUHK03	Combined Dataset
inter-class distance *DSBN*	1.80	1.83	1.84	1.92
inter-class distance *RDSBN*	**1.88**	**1.85**	**1.91**	**1.95**
intra-class variance *DSBN*	0.35	0.36	0.32	0.34
intra-class variance *RDSBN*	**0.32**	**0.33**	**0.30**	**0.32**

Table 3. Experiments on different domain adaptation frameworks.

Methods	Duke+CUHK→Market				Market+CUHK→Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
Baseline	70.1	86.7	95.0	96.5	58.3	72.8	84.4	88.2
Baseline+*DSBN*	74.1	89.2	96.2	97.9	60.4	76.2	86.1	90.1
Baseline+*Ours*	**79.4**	**92.1**	**97.1**	**98.1**	**65.0**	**79.1**	**88.5**	**91.6**
MMT	75.3	89.6	96.5	97.6	62.6	76.2	87.1	90.2
MMT+*DSBN*	81.2	92.7	97.5	98.4	65.7	79.0	89.0	92.1
MMT+*Ours*	**85.2**	**94.2**	**98.0**	**98.8**	**69.0**	**81.2**	**90.3**	**93.2**

Table 4. Ablation studies on MDIF module under another source-target setting.

Methods	Market+CUHK→Duke			
	mAP	R-1	R-5	R-10
ResNet50-*BN* (baseline)	58.3	72.8	84.4	88.2
ResNet50-*BN-MDIF*	**62.7**	**76.9**	**87.8**	**91.2**
ResNet50-*DSBN*	60.4	76.2	86.1	90.1
ResNet50-*DSBN-MDIF*	**63.3**	**78.1**	**88.2**	**91.5**
ResNet50-*RDSBN*	63.5	77.3	88.3	91.6
ResNet50-*RDSBN-MDIF*	**65.0**	**79.1**	**88.5**	**91.6**

Table 5. Analysis of MDIF module design under another source-target setting.

Methods	Market+CUHK→Duke			
	mAP	R-1	R-5	R-10
baseline	58.3	72.8	84.4	88.2
Instance Graph	61.6	76.1	86.8	90.0
Mean Agent Node	62.5	76.9	87.0	90.9
Weighted Mean Agent Node	**62.7**	**77.3**	**87.8**	**91.2**
intra-connections in $A^{(1)}$	61.9	76.0	87.0	90.2

[4] 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. 1

[5] 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. 1

[6] 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. 1

[7] 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. 1

[8] Chengxu Zhuang, Alex Lin Zhai, and Daniel Yamins. Local aggregation for unsupervised learning of visual embeddings. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6002–6012, 2019. 2