

# MULTI-DOMAIN SYNCHRONOUS REFINEMENT NETWORK FOR UNSUPERVISED CROSS-DOMAIN PERSON RE-IDENTIFICATION

Sikai Bai, Junyu Gao\*, Qi Wang\*, Xuelong Li

School of Computer Science and School of Artificial Intelligence, Optics and Electronics  
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China  
{whitesk1973, gjy3035, crabwq}@gmail.com, xuelong\_li@nwpu.edu.cn.

## ABSTRACT

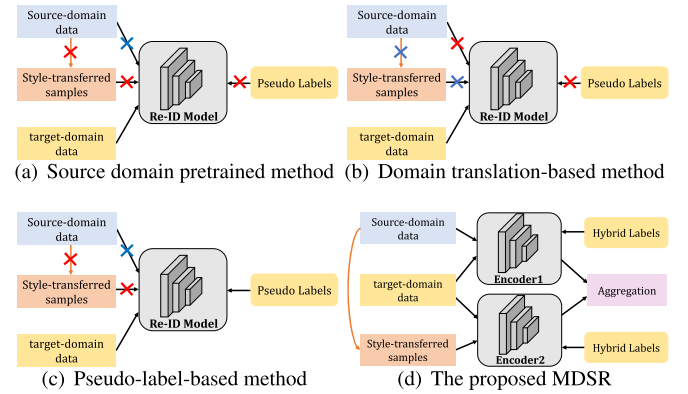
Unsupervised cross-domain person re-identification (re-ID) is a challenging task, because it is an open-set problem with completely unknown person identities in the target domain. Existing methods attempt to tackle the challenge by transferring image style across domains or generating pseudo labels in the target domain, whereas the valuable information in multiple domains (*i.e.*, source domain, style-transferred data, and target domain) is not taken fully into consideration. To this end, we propose a novel multi-domain synchronous refinement (MDSR) network, where valuable knowledge from multiple domains is sufficiently exploited and refined to enforce the discriminative ability of the model. MDSR network contains two complementary modules dedicated to source-to-target domain adaptation and style-transferred data to the target domain adaptation, respectively. The domain adaptive knowledge from two modules is aggregated in the final stage. Extensive experiments verify our method achieves significant improvements over the state-of-the-art approaches on multiple unsupervised domain adaptive person re-ID tasks.

**Index Terms**— Person re-identification, unsupervised domain adaptation, synchronous refinement

## 1. INTRODUCTION

Person re-identification (re-ID) aims at matching the same person from images captured by multiple cameras. It has received more and more attention due to its wide applications, such as video surveillance and public safety. Encouraged by the success of deep learning and the availability of large-scale labeled datasets, supervised re-ID methods [1], [2], [3] have made impressive progress. However, as shown in Fig. 1(a), re-ID models pre-trained on label-rich datasets (source domain) are directly applied to other unlabeled datasets (target domain), and the performance is decreased by a big margin due to domain gaps between different datasets. Moreover,

\*Qi Wang and Junyu Gao are the corresponding authors. This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.



**Fig. 1.** Cross-domain person re-ID methods considered in this paper. The orange arrow represents the domain translation operation for generating style-transferred samples. The red cross indicates that the input data is simply discarded, and the blue cross denotes that the corresponding input data is only used during pre-training.

it is impractical to continuously annotate person identities in various new target domains.

To mitigate the domain gap, a large number of methods have been proposed to focus on unsupervised domain adaptation (UDA) for person re-ID. They can be mainly categorized as two directions: 1) Domain translation-based methods [4], [5], [6], [7]. These methods focus on the translation from source domain to the target domain to create style-transferred samples by using generative adversarial networks (GAN). However, as shown in Fig.1(b), source domain data is simply discarded after translation, and there are no pseudo-labels for target domain data in these approaches. 2) Pseudo-label-based methods [8], [9], [10], [11]. These methods first pre-train re-ID models on the labeled source domain, then finetune the pre-trained models with pseudo-labeled data on the target domain, where the pseudo labels can be predicted by some cluster algorithms. Although this pipeline dominates current state-of-the-art performance, noisy pseudo-labels and large domain gaps between source and target domains have

negative effects on the performance. Besides, as shown in Fig. 1(c), the useful knowledge from style-transferred samples is not considered in these methods.

In this paper, we propose a multi-domain synchronous refinement (MDSR) network for unsupervised domain adaptive person re-ID. MSDR network learns the valuable knowledge from two designed modules, and aims to reduce the negative impacts of domain gap, preserve original inter-sample relations in the source domain and make full use of the useful data in multiple domains. Specifically, we first propose a united style-similarity hybrid learning (SSHL) module to reduce the negative impacts of the domain gap. SSHL consists of a domain translator, a feature encoder and a cluster generator. The domain translator aims at generating style-transferred samples with a similar style to the target domain. The feature encoder is used to perform joint training for the style-transferred samples and target-domain data. The cluster generator aims to predict reliable pseudo labels on the target domain by using the extracted features from above encoder. Meanwhile, a relation-invariant hybrid learning (RIHL) module is devised for preserving inter-sample relations in the source domain. RIHL contains a feature encoder and a cluster generator. The feature encoder in RIHL aims at joint training for the labeled source domain data and the pseudo-labeled target data. The cluster generator in RIHL utilizes the extracted features from the feature encoder in RIHL to predict reliable pseudo labels. Furthermore, to construct a multi-domain synchronous refinement network, the two above modules are synchronously trained to sufficiently exploit the valuable knowledge from multiple domains. Finally, the discriminative ability of the model is further refined by aggregating the learned knowledge from two modules.

The major contribution of our work can be summarized as three-fold. (1) In response to the negative impacts of domain gaps and the loss of original inter-sample relations during translation, this work develops a style-similarity hybrid learning (SSHL) module and a relation-invariant hybrid learning (RIHL) module, respectively. (2) In order to sufficiently exploit the knowledge from multiple domains, this work proposes a multi-domain synchronous refinement (MDSR) network by synchronously training the two above modules, and the ability of the model is further refined by aggregating the learned knowledge from two modules. (3) The proposed MSDR network achieves consistent performance gain over the state-of-the-arts on multiple unsupervised domain adaptation tasks of person re-ID.

## 2. RELATED WORK

### 2.1. Unsupervised Domain Adaptation

Unsupervised domain adaptation (UDA) attempts to generalize the learned knowledge from the labeled source domain to the target domain with unknown classes well. Some ap-

proaches aim at aligning the feature distributions between the source and target domains [12], [13]. Correlation Alignment (CORAL) [12] mapped the covariance and mean of two distributions, and CYCADA [13] transferred samples across domains at both pixel- and feature-level. Moreover, some other methods try to find domain-invariant feature spaces [14], [15], [16], [17]. Long *et al.* [14] and Tzeng *et al.* [15] employed the Maximum Mean Discrepancy (MMD). Ganin *et al.* [16] and Ajakan *et al.* [17] built domain confusion loss for this purpose. However, the above methods only focus on conventional domain adaptation, where the source and target domains share the same set of class labels.

### 2.2. Cross-domain Person Re-Identification

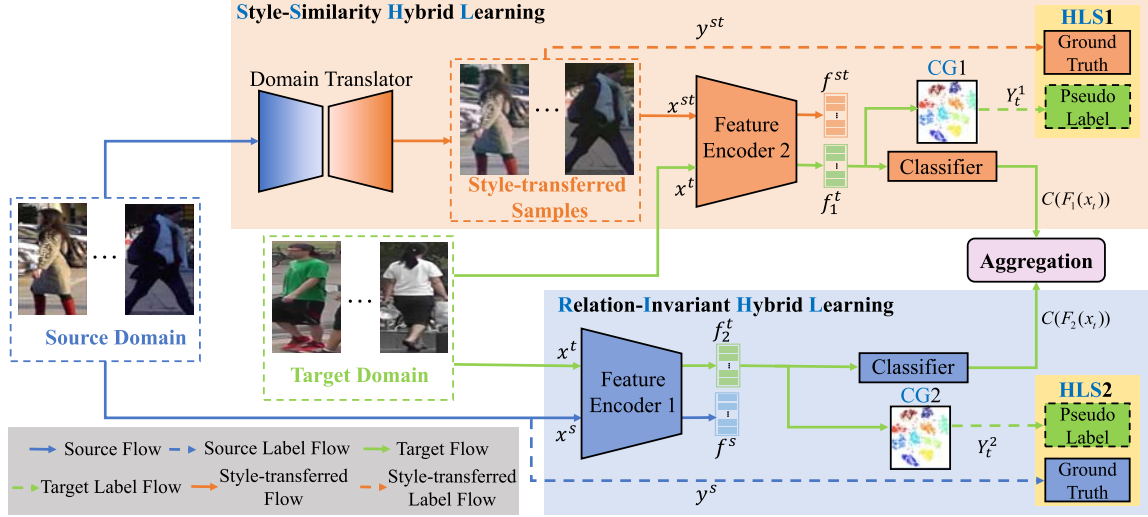
Cross-domain person re-identification is an open-set problem with disjoint person identities between different domains. There are many works focusing on unsupervised methods include domain translation and pseudo-label generation. Specifically, the domain translation-based methods [4], [6], [7] used the re-ID model pre-trained in labeled style-transferred samples to perform unsupervised finetuning in the target domain. SPGAN [4] and PTGAN [6] maintained the ID-related features invariant by identity-based regularization. SDA [7] used online relationship consistency regularization term to maintain inter-sample relations. The pseudo-label-based approaches [18], [19], [20], [21] enforce feature learning in the target domain by utilizing the pseudo labels, where the labels are commonly generated by cluster algorithms (e.g. Kmeans or DBSCAN). SSG [20] assigned multi-scale pseudo labels by introducing human local features. Besides, in order to further optimize the predicted pseudo labels, MMT [21] and SPCL [18] introduced mutual mean-teaching and self-paced learning mechanism, respectively. Nevertheless, the above methods are unable to consider the valuable knowledge from style-transferred samples. And most of these methods suffer from the adverse impacts of domain gap between source and target domain and noisy pseudo labels.

## 3. PROPOSED METHOD

The overview architecture of the proposed multi-domain synchronous refinement (MDSR) network is illustrated in Fig. 2. In this section, we first give some essential symbolic definitions, then introduce the details of the proposed SSHL module and RIHL module. Finally, the proposed MSDR network is presented by integrating the above two modules to sufficiently exploit the valuable knowledge from multiple domains.

### 3.1. Notation

To facilitate understanding, some symbolic definitions are given here. Given the labeled source domain data  $X^s$  and un-



**Fig. 2.** The overview architecture of our multi-domain synchronous refinement network. CG1 and CG2 denote two constructed cluster generators. HLS1 and HYL2 denote two devised hybrid-label systems.

labeled target domain data  $X^t$ , where  $X^s = \{(x_i^s, y_i^s) | i=1 \dots N_s\}$  contains  $N_s$  person images with corresponding ground-truth identity labels, and  $X^t = \{(x_i^t, \hat{y}_i^t) | i=1 \dots N_t\}$  involves  $N_t$  target images with pseudo labels predicted by the clustering algorithm. Furthermore, we denote style-transferred samples as  $X^{st} = \{(x_i^{st}, y_i^{st}) | i=1 \dots N_{st}\}$ , in which  $x_i^{st}$  and  $y_i^{st}$  are generated from the source domain data.

### 3.2. Style-Similarity Hybrid Learning

Previous UDA methods for person re-ID commonly attempt to reduce domain gap by two ways, *i.e.*, generating style-transferred samples for source domain data or predicting pseudo labels in the target domain. Different from these works, the designed style-similarity hybrid learning (SSHL) module considers domain adaptation in two aspects, where it integrates the processes of domain translation and pseudo label prediction, while performing joint training with the generated style-transferred samples and target domain data. Specifically, as shown in Fig. 2, there are three sub-modules in SSLH, namely, the domain translator  $\mathcal{T}$ , the feature encoder  $\mathcal{E}_1$ , and the cluster generator  $\mathcal{G}_1$ . They are parameterized with  $\theta_t$ ,  $\theta_{e1}$  and  $\theta_{g1}$ , respectively.

**Domain Translator.** The domain translator focuses on translating the annotated source domain data to style-transferred samples, which is implemented by a generative adversarial network, where the samples enjoy ground-truth identity labels derived from the source domain data and have a similar style with the target domain data. The constructed domain translator only need to maintain identity-related information invariant during translation. Given a source data pair  $(x^s, y^s)$  as input, the corresponding style-transferred sample

pair  $(x^{st}, y^{st})$  generated by the domain translator  $\mathcal{T}$  can be formally expressed as:

$$x^{st} = \mathcal{T}(x^s; \theta_t), y^{st} = y^s. \quad (1)$$

**Feature Encoder.** The feature encoder in SSLH is designed to provide more discriminative target features by performing joint training with the style-transferred samples and target domain data. Initially, the style-transferred samples enjoy the ground-truth identity labels while the target domain data has no identity labels. It is worth noting that our feature encoder does not require additional pre-training in the source domain or style-transferred samples. Formally, given a style-transferred image  $x^{st}$  and target domain image  $x^t$ , the corresponding output features  $f^{st}$  and  $f^t$  are obtained by the following mapping:

$$f^{st} = \mathcal{E}_1(x^{st}; \theta_{e1}), f_1^t = \mathcal{E}_1(x^t; \theta_{e1}). \quad (2)$$

**Cluster Generator.** To refine joint training for the above feature encoder, we first predict the pseudo labels for the target domain data by using DBSCAN cluster algorithm. Then a hybrid-label system HLS1 is constructed by concatenating the predicted pseudo labels and the source ground-truth identity labels. Furthermore, we refine the clustering algorithm by a novel clustering criterion and take it as our cluster generator, which encourages to preserve more reliable data in clusters by measuring the reliability of each feature point during clustering. Concretely, given the feature  $f^t$  in the target domain, our clustering criterion can be represented as:

$$\begin{cases} r_1(f^t) = \max(0, \frac{|\gamma(f^t) \cap \gamma_s(f^t)|}{|\gamma(f^t) \cup \gamma_s(f^t)|} + \frac{|\gamma(f^t) \cap \gamma_e(f^t)|}{|\gamma(f^t) \cup \gamma_e(f^t)|} - \alpha), \\ r_2(f^t) = \max(0, \frac{|\gamma(f^t) \cap \gamma_s(f^t)|}{|\gamma(f^t) \cup \gamma_s(f^t)|} - \frac{|\gamma(f^t) \cap \gamma_e(f^t)|}{|\gamma(f^t) \cup \gamma_e(f^t)|} - \beta), \end{cases} \quad (3)$$

where  $\gamma(f^t)$  denotes the cluster set containing the feature  $f^t$ ,  $\gamma_s(f^t)$  and  $\gamma_e(f^t)$  are the clusters including  $f^t$  when shrinking and enlarging the distance threshold in the DBSCAN algorithm, respectively. Two dynamic thresholds (*i.e.*,  $\alpha$  and  $\beta$ ) and two obtained clustering rules (*i.e.*,  $r_1(f^t)$  and  $r_2(f^t)$ ) focus on identifying reliable feature points in the clusters on-the-fly. Overall, we compare the initial clustering results of feature points with the results when shrinking and enlarging the threshold distance respectively, to analyze the clustering reliability of feature points from target domain data. Finally, given the output feature  $f^t$  of the above feature encoder, the corresponding pseudo-label  $\hat{y}_i^t$  can be obtained by the following mapping:

$$\hat{y}_2^t = \begin{cases} \mathcal{G}_1(f^t; \theta_{g1}), & c_1(f^t) > 0, c_2(f^t) > 0, \\ -1, & \text{otherwise}, \end{cases} \quad (4)$$

when  $\hat{y}_i^t = -1$ , the corresponding samples are divided into discrete instances with individual person class labels. Therefore, the number of pseudo labels equals  $n_c^t + n_o^t$ , where  $n_c^t$  and  $n_o^t$  represent the number of reliable clusters and discrete instances, respectively.

### 3.3. Relation-Invariant Hybrid Learning

Although the style-transferred samples have similar styles to the target domain and reduce the negative impact of the domain gap, it is found in [7] that these samples only retain identity-related information while losing the original inter-sample relations from the source domain during translation.

To maintain original inter-sample relations invariant and implement joint learning for source domain and target domain data, we devise a relation-invariant hybrid learning (RIHL) module. As illustrated in Fig. 2, RIHL contains a feature encoder  $\mathcal{E}_2$  and a cluster generator  $\mathcal{G}_2$ , which have the same architecture details as sub-modules in SSHL, but the parameters are not shared with them. We parameterize two sub-modules in RIHL with  $\theta_{e2}$  and  $\theta_{g1}$  respectively, and build a new hybrid-label system HLS2 by concatenating the source ground-truth and the target pseudo label set. These parameterized sub-modules are iteratively trained to strengthen the discriminative of target features, which is implemented by simultaneously using the source domain data with original inter-sample relations and the pseudo labeled target domain data as the input of the module. Therefore, the output feature of the feature encoder  $\mathcal{E}_2$  in RIHL can be denoted as:

$$\begin{cases} f_2^t = \mathcal{E}_2(x^t; \theta_{e2}), & x^t \in X^t, \\ f^s = \mathcal{E}_2(x^s; \theta_{e2}). & x^s \in X^s. \end{cases} \quad (5)$$

### 3.4. Multi-domain synchronous refinement network

In order to perform multiple domains synchronous training and refinement, we construct a novel network by integrating the proposed SSHL module and RIHL module, referred as

**Table 1.** Data statistics (identities/images) for experiments on Market-1501, DukeMTMC-ReID and MSMT17.

Dataset	#Train	#Test	#All
Market-1501 [22]	751/12936	1501/19732	1501/32688
DukeMTMC-ReID [23]	702/16522	702/19899	1812/36411
MSMT17 [6]	1041/32621	3060/93820	4101/126411

MDSR network. We use the abundant data from the source domain and target domain as the MSDR network input for synchronous training. Specifically, for source domain data, the MSDR network reduces the negative impact of the domain gap by generating style-transferred samples in SSHL, and the original inter-sample relations invariant is considered by directly utilizing the rich source information in RIHL. For target domain data, the valuable complementary knowledge is extracted from the two designed modules, where we refine the predicted pseudo labels based on the designed clustering criterion. At the last stage, we perform aggregation for the learned knowledge from two modules by a weighted factor  $\lambda$  to further refine the discriminative ability of the model, and the aggregated results are regarded as the final outputs of the MSDR network. Therefore, given target domain data  $x^t$ , the final output  $y_o$  of the MSDR network can be given by:

$$\begin{aligned} y_o &= \lambda \mathcal{C}_1(f_1^t) + (1 - \lambda) \mathcal{C}_2(f_2^t) \\ &= \lambda \mathcal{C}_1(\mathcal{F}_1(x^t; \theta_{e1})) + (1 - \lambda) \mathcal{C}_2(\mathcal{F}_2(x^t; \theta_{e2})), \end{aligned} \quad (6)$$

where  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are two learnable classifiers from SSHL and RIHL to identify person class.

## 4. EXPERIMENTS

### 4.1. Datasets and Implementation Details

**Datasets.** The proposed approach is evaluated on three wide-used person re-ID datasets, including Market-1501 [22], DukeMTMC-reID [23] and MSMT17 [6]. Data statistics are shown in Table 1, DukeMTMC-reID considers additional 408 identities as distractors, and MSMT17 is the most challenging and large-scale dataset.

**Implementation Details.** We adopt the ImageNet-pretrained IBN-ResNet50 as the backbone of our feature encoders (*i.e.*,  $\mathcal{E}_1$  and  $\mathcal{E}_2$ ), and utilize eSPGAN [5] as our domain translator  $\mathcal{T}$  in SSHL. The training parameters include 50 epochs, batch size 64, momentum 0.2, weight decay 0.0005. We initialize the learning rate to be 0.00035 and decrease it by 10 every 20 epochs. For two dynamic thresholds (*i.e.*,  $\alpha$  and  $\beta$ ) in clustering criterion, inspired by [18], they are set to the sum and difference of the top-90%  $\frac{|\gamma(f^t) \cap \gamma_s(f^t)|}{|\gamma(f^t) \cup \gamma_s(f^t)|}$  in the first epoch and the maximum  $\frac{|\gamma(f^t) \cap \gamma_e(f^t)|}{|\gamma(f^t) \cup \gamma_e(f^t)|}$  in the cluster of each epoch. And the weight factor  $\lambda$  in our MSDR network

**Table 2.** Comparison of our MSDR network with state-of-the-art methods. The best results are highlighted with bold fonts.

Methods		DukeMTMC-reID $\rightarrow$ Market-1501				Market-1501 $\rightarrow$ DukeMTMC-reID				DukeMTMC-reID $\rightarrow$ MSMT17			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
ECN [24]	CVPR'19	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4	10.2	30.2	41.5	46.8
SSG [20]	ICCV'19	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2	13.3	32.2	-	51.2
MMCL [19]	CVPR'20	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.5	16.2	43.6	54.3	58.9
ECN++ [25]	TPAMI'20	63.8	84.1	92.8	95.4	54.4	74.0	83.7	87.4	16.0	42.5	55.9	61.5
DG-Net++ [26]	ECCV'20	63.8	78.9	87.8	90.4	61.7	82.1	90.2	92.7	22.1	48.8	60.9	65.9
JVTC+ [27]	ECCV'20	67.2	86.8	95.2	97.1	66.5	80.4	89.9	92.2	27.5	52.9	70.5	75.9
GPR [28]	ECCV'20	71.5	88.1	94.4	96.2	65.2	79.5	88.3	91.4	24.3	51.7	64.0	68.9
SDA [7]	arXiv'20	74.3	89.7	95.9	97.4	66.7	79.9	89.1	92.7	30.3	59.6	71.7	76.2
MMT [21]	ICLR'20	76.5	90.9	96.4	97.9	67.3	80.8	90.3	93.0	29.7	58.8	71.0	76.1
SPCL [18]	NeurIPS'20	79.2	91.5	96.9	98.0	69.9	83.4	91.0	93.1	31.8	58.9	70.4	75.2
<b>MSDR(Ours)</b>		<b>81.2</b>	<b>92.3</b>	<b>97.1</b>	<b>98.1</b>	<b>72.7</b>	<b>84.8</b>	<b>92.0</b>	<b>93.4</b>	<b>34.2</b>	<b>60.3</b>	<b>72.0</b>	<b>76.3</b>

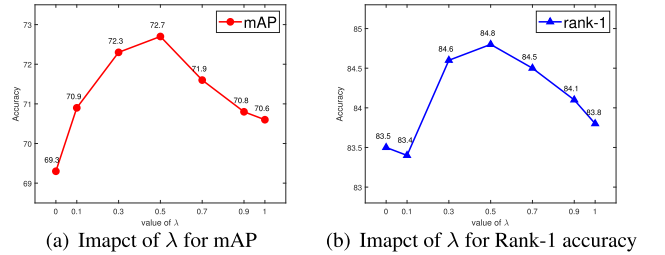
**Table 3.** Components performance of MSDR network.

Methods	Market-1501 $\rightarrow$ DukeMTMC-reID			
	mAP	Rank-1	Rank-5	Rank-10
Baseline	32.9	61.7	77.1	82.0
SSHL	69.3	83.5	90.5	92.6
RIHL	70.6	83.8	91.5	93.1
<b>MSDR</b>	<b>72.7</b>	<b>84.8</b>	<b>92.0</b>	<b>93.4</b>

is 0.5. Besides, we adopt mean average precision (mAP) and CMC Rank-1, Rank-5, Rank-10 accuracies to evaluate the performance of models. All experiments are conducted on 4 GTX-1080TI GPUs, and it is worth noting that no post-processing technology (*e.g.*, multi-query fusion [22] or re-ranking [29]) is used in our experiments.

#### 4.2. Comparison with State-of-the-arts.

We compare the proposed approach with state-of-the-art methods on three unsupervised domain adaptive re-ID tasks. As shown in Table 2, our MSDR network significantly outperforms all existing methods in terms of mAP and CMC Rank-1, Rank-5, Rank-10 accuracies. Specifically, compared with domain translation-based approaches, the proposed MSDR network surpass the state-of-the-art method SDA by a big margin, obtains relative improvements of 6.9% (81.2% *vs.* 74.3%), 6% (72.7% *vs.* 66.7%) and 3.9% (34.2% *vs.* 30.3%) mAP on Market-1501  $\rightarrow$  DukeMTMC-reID, Market-1501  $\rightarrow$  DukeMTMC-reID and DukeMTMC-reID  $\rightarrow$  MSMT17 tasks, respectively. And comparing with pseudo-label-based methods, although these methods have previously achieved the optimal performance, MSDR still outperforms the state-of-the-art approach SPCL by 2% (81.2% *vs.* 79.2%), 2.8% (72.7% *vs.* 69.9%) and 2.4% (34.2% *vs.* 31.8%) mAP on three domain adaptation tasks. These results show that the effectiveness of our MSDR network by synchronously refining the knowledge from multiple domains is impressive, and it yields

**Fig. 3.** Performance analysis of the proposed MSDR with different values of  $\lambda$  on Market-1501  $\rightarrow$  DukeMTMC-reID.

a new record on the cross-domain person re-id problem.

#### 4.3. Ablation Study

**Components Effectiveness.** To evaluate the effectiveness of each component in our MSDR network (*i.e.*, SSL and RIHL), we create a baseline method that directly apply the pretrained IBN-ResNet50 in the labeled source domain to conduct fine-tuning on the unlabeled target domain, while comparing it with the single modules and the combination of both modules (*i.e.*, MSDR) on Market-1501  $\rightarrow$  DukeMTMC-reID task. As shown in Table 3, SSL utilizes style-transferred samples and the pseudo labeled target data to jointly train the model, and it gains 36.4% mAP improvement. RIHL considers original inter-sample relations on source domain data and predicts pseudo labels on the target domain, and it obtains 37.7% mAP improvements. Furthermore, better performance can be provided by the weighted aggregation for SSL and RIHL together.

**Parameter Analysis.** Finally, we analyze the impacts of the different values of the hyperparameter  $\lambda$  on performance, when SSL and RIHL are weighted aggregated in the final stage of the model. As shown in Fig. 3, our MSDR network achieves the optimal performance when setting the weighted factor  $\lambda$  as 0.5 in Eq. 6.

## 5. CONCLUSION

In this paper, we propose a multi-domain synchronous refinement network for unsupervised cross-domain person re-ID, which sufficiently exploits the valuable knowledge from multiple domains by two designed modules (*i.e.*, style-similarity hybrid learning module and relation-invariant hybrid learning module). Extensive experiments demonstrate the superiority of our method over recent state-of-the-art approaches.

## 6. REFERENCES

- [1] X. Fan, W. Jiang, H. Luo, and M. Fei, “Sphered: Deep hypersphere manifold embedding for person re-identification,” *JVCIR*, vol. 60, pp. 51–58, 2019.
- [2] Y. Yuan, J. Zhang, and Q. Wang, “Deep gabor convolution network for person re-identification,” *Neurocomputing*, vol. 378, pp. 387–398, 2020.
- [3] M. Jiang, Y. Yuan, and Q. Wang, “Self-attention learning for person re-identification,” in *BMVC*, 2018, p. 204.
- [4] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, “Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification,” in *CVPR*, 2018, pp. 994–1003.
- [5] W. Deng, L. Zheng, Q. Ye, Y. Yang, and J. Jiao, “Similarity-preserving image-image domain adaptation for person re-identification,” *arXiv preprint arXiv:1811.10551*, 2018.
- [6] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” in *CVPR*, 2018, pp. 79–88.
- [7] Y. Ge, F. Zhu, R. Zhao, and H. Li, “Structured domain adaptation for unsupervised person re-identification,” *arXiv preprint arXiv:2003.06650*, 2020.
- [8] HX. Yu, WS. Zheng, A. Wu, X. Guo, S. Gong, and JH. Lai, “Unsupervised person re-identification by soft multilabel learning,” in *CVPR*, 2019, pp. 2148–2157.
- [9] X. Zhang, J. Cao, C. Shen, and M. You, “Self-training with progressive augmentation for unsupervised cross-domain person re-identification,” in *ICCV*, 2019, pp. 8222–8231.
- [10] H. Huang, X. Chen, and K. Huang, “Proxy task learning for cross-domain person re-identification,” in *ICME*, 2020, pp. 1–6.
- [11] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, and Y. Tian, “Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification,” in *CVPR*, 2020, pp. 9021–9030.
- [12] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” *arXiv preprint arXiv:1511.05547*, 2015.
- [13] J. Hoffman, E. Tzeng, T. Park, JY. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, “Cycada: Cycle-consistent adversarial domain adaptation,” in *ICML*, 2018, pp. 1989–1998.
- [14] M. Long, Y. Cao, J. Wang, and M. Jordan, “Learning transferable features with deep adaptation networks,” in *ICML*, 2015, pp. 97–105.
- [15] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep domain confusion: Maximizing for domain invariance,” *arXiv preprint arXiv:1412.3474*, 2014.
- [16] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *JMLR*, vol. 17, no. 1, pp. 2096–2030, 2016.
- [17] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, “Domain-adversarial neural networks,” *arXiv preprint arXiv:1412.4446*, 2014.
- [18] Y. Ge, D. Chen, F. Zhu, R. Zhao, and H. Li, “Self-paced contrastive learning with hybrid memory for domain adaptive object re-id,” *arXiv preprint arXiv:2006.02713*, 2020.
- [19] D. Wang and S. Zhang, “Unsupervised person re-identification via multi-label classification,” in *CVPR*, 2020, pp. 10981–10990.
- [20] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, and S.T. Thomas, “Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification,” in *ICCV*, 2019, pp. 6112–6121.
- [21] Y. Ge, D. Chen, and H. Li, “Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification,” pp. 1–14, 2020.
- [22] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in *CVPR*, 2015, pp. 1116–1124.
- [23] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in *ECCV*, 2016, pp. 17–35.
- [24] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Invariance matters: Exemplar memory for domain adaptive person re-identification,” in *CVPR*, 2019, pp. 598–607.
- [25] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, “Learning to adapt invariance in memory for person re-identification,” *TPAMI*, 2020.
- [26] Y. Zou, X. Yang, Z. Yu, BVK. Kumar, and J. Kautz, “Joint disentangling and adaptation for cross-domain person re-identification,” in *ECCV*, 2020, pp. 87–104.
- [27] J. Li and S. Zhang, “Joint visual and temporal consistency for unsupervised domain adaptive person re-identification,” pp. 1–14, 2020.
- [28] C. Luo, C. Song, and Z. Zhang, “Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup,” pp. 1–8, 2020.
- [29] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li, “Re-ranking person re-identification with k-reciprocal encoding,” in *CVPR*, 2017, pp. 1318–1327.