Divide-and-Regroup Clustering for Domain Adaptive Person Re-identification

Zhengdong Hu[1,2] *, **Yifan Sun**[2], **Yi Yang**[3], **Jianguang Zhou**[1] †

[1] Research Center for Analytical Instrumentation, Institute of Cyber-Systems and Control, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China
[2] Baidu Research, China
[3] ReLER, Centre for Artificial Intelligence, University of Technology Sydney, Australia
zhengdonghu@zju.edu.cn, sunyf15@tsinghua.org.cn, yi.yang@uts.edu.au, jgzhou@zju.edu.cn

Abstract

Clustering is important for domain adaptive person re-identification (re-ID). A majority of unsupervised domain adaptation (UDA) methods conduct clustering on the target domain and then use the generated pseudo labels for adaptive training. Albeit important, the clustering pipeline adopted by current literature is quite standard and lacks consideration for two characteristics of re-ID, *i.e.*, 1) a single person has various feature distribution in multiple cameras. 2) a person's occurrence in the same camera are usually temporally continuous. We argue that the multi-camera distribution hinders clustering because it enlarges the intra-class distances. In contrast, the temporal continuity prior is beneficial, because it offers clue for distinguishing some look-alike person (who are temporally far away from each other). These two insights motivate us to propose a novel Divide-And-Regroup Clustering (DARC) pipeline for re-ID UDA. Specifically, DARC divides the unlabeled data into multiple camera-specific groups and conducts local clustering within each camera. Afterwards, it regroups those local clusters potentially belonging to the same person into a unity. Through this divide-and-regroup pipeline, DARC avoids directly clustering across multiple cameras and focuses on the feature distribution within each individual camera. Moreover, during the local clustering, DARC uses the temporal continuity prior to distinguish some look-alike person and thus reduces false positive pseudo labels. Consequentially, DARC effectively reduces clustering errors and improves UDA. Importantly, experimental results show that DARC is compatible to many pseudo-label-based UDA methods and brings general improvements. Based on a recent UDA method, DARC advances the state of the art (e.g, 85.1% mAP on MSMT-to-Market and 83.1% mAP on PersonX-to-Market).

Introduction

This paper considers unsupervised domain adaptive person re-identification (re-ID) task, which employs Unsupervised Domain Adaptation (UDA) to improve (cross-domain) re-ID. Basically, re-ID aims to retrieve all the images of the same identity in the database, given a query image of interest. (Luo et al. 2020; Wang et al. 2019; Luo et al. 2019a; Quan et al. 2019) When the training data and the testing data are collected from different domains, the underlying domain gap stands out as a prominent challenge. An effective approach to mitigate the domain gap is Unsupervised Domain Adaptation (UDA). UDA adapts the deep model to recognize the unlabeled target domain and thus improves cross-domain re-ID without additional annotations.

A popular family (Fu et al. 2019; Zhang et al. 2019; Zhai et al. 2020; Fan et al. 2018; Ge, Chen, and Li 2020; Ge et al. 2020; Dai et al. 2021a; Zheng et al. 2021; Dai et al. 2021b) of the UDA methods is based on pseudo label learning. Using

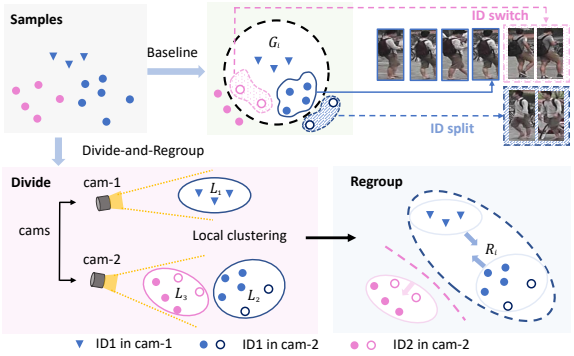

Figure 1: The *global clustering baseline* (in the first row) is prone to two types of clustering errors. "ID split": the multi-camera distribution enlarges the intra-class distances and sometimes causes a single person split into multiple clusters. "ID switch": some inherently look-alike person in the same camera are merged into a single cluster, resulting in ID switch. The proposed *DARC* (in the second row) reduces ID split and ID switch through a two-stage clustering pipeline. "Divide": it first divides the features into camera-specific groups and conducts a respective local clustering within each group. Local clustering avoids multi-camera distribution and utilizes the temporal continuity to separate look-alike (but different) person. "Regroup": it compares local clusters against each other and regroups those closely-distributed clusters into a single one.

*Zhengdong Hu makes his part of work during internship in Baidu Research.
†Corresponding author.

the already-learned deep embedding, they cluster the unlabeled samples on the target domain and assign each cluster with a pseudo label. These pseudo labels are then used for training (or fine-tuning) on the target domain.

In such a pseudo-label-based UDA pipeline, the clustering quality is vital. We notice that current literature adopt a quite standard "global" clustering pipeline. Specifically, they calculate the pairwise distance between all the sample pairs in the target domain to get an $N \times N$ distance matrix (N is the total number of samples), and use the entire distance matrix for clustering. This global clustering baseline lacks consideration for two characteristics of the re-ID task: 1) Multi-camera distribution: a single person has various feature distribution in multiple cameras. 2) Temporal Continuity: a person's occurrence in the same camera are usually temporally continuous.

We argue that these two characteristics respectively bring negative / positive impact on clustering accuracy and can be avoided / utilized for better clustering. Specifically, the multi-camera distribution hinders clustering because it enlarges the intra-class distances. Thus, the instances of a single person sometimes are split into multiple clusters, resulting in the "*ID split*" error, as illustrated in Fig. 1 (in the first row). In contrast, the temporal continuity is beneficial for clustering because it offers clue for distinguishing some look-alike person. As illustrated in Fig. 1 (in the first row), two inherently look-alike person in the same camera are merged into a single cluster, resulting in the "*ID switch*" error. However, if we have the temporal prior, *i.e.*,these two people are temporally far away from each other, we may easily separate them and avoid the ID switch error.

These two insights motivate us to propose a novel Divide-And-Regroup Clustering (DARC) pipeline for re-ID UDA. Different from the global clustering, DARC has two stages, *i.e.*, local clustering and cross-camera regrouping. We explain these two stages as follows:

1) *Dividing and local clustering*. DARC first divides all the unlabeled samples into multiple camera-specific groups. Each group contains the samples from an individual camera. Within each group, DARC conducts a respective local clustering. The local clustering is critical to DARC and has two advantages, as illustrated in Fig. 1 (in the second row): First, it reduces the ID switch errors through cooperating temporal continuity. Second, it reduces the ID split errors within each individual camera by focusing on the local feature distribution. The detailed reasons and evidences for these two advantages are illustrated in Methods.

2) *Cross-camera regrouping*. Local clustering alone does not suffice the requirement for assigning pseudo labels. It is because the samples of a single person are scattered across multiple cameras, while local clustering does not facilitate cross-camera association. To regroup the local clusters potentially belonging to a single identity, DARC compares local clusters against each other and then regroups the closely-distributed local clusters into a single one, as illustrated in Fig. 1 (in the second row). Although the cross-camera comparison incurs the multi-camera distribution problem (as in the global clustering baseline), we show that the regrouping operation is robust to this problem in Methods.

Through the divide-and-regroup procedure, DARC effectively reduces both ID switch and ID split errors, and consequentially improves re-ID UDA. Importantly, DARC is capable to accommodate many pseudo-label-based UDA methods, and brings general improvement to these methods in a plug-and-play manner. We conduct comprehensive experiments under seven domain adaptive re-ID scenarios and demonstrate consistent improvement on several popular UDA methods. Based on a recent UDA method (Dai et al. 2021b), DARC advances the state of the art with 85.1% mAP on MSMT-to-Market and 83.1% mAP on PersonX-to-Market

Our main contributions are summarized as follows:

- We propose a novel Divide-and-Regroup Clustering (DARC) pipeline for domain adaptive re-ID. Different from the global clustering pipeline, DARC takes two characteristics of re-ID task into consideration and increases clustering accuracy.
- We show that DARC is compatible to many pseudo-label-based re-ID UDA methods and brings general improvement to these methods in a plug-and-play manner.
- We validate the effectiveness of the proposed DARC with extensive experiments on multiple domain adaptive re-ID datasets. On all these datasets, DARC achieves performance on par with the state of the art.

Related Works

Deep person Re-ID. Most existing Re-ID methods adopt deep network in the past few years. (Sun et al. 2018; Zhang et al. 2018; Luo et al. 2019b; Wang et al. 2018) adopt stripe-based methods to split the image into different parts and extract the local feature of each part. (Sarfraz et al. 2018; Wei et al. 2017) adopt pose-based methods to extract the pose feature of each person and estimate the relevance of pose between different images. Strong baseline (Luo et al. 2020) adopts effective training tricks for person ReID and proposes the BNNeck structure to cooperate ID loss and triplet loss in a better way. Though these methods achieve promising results in the labeled datasets, the performance drops sharply when the learned embedding is directly transformed to unlabeled datasets.

Clustering method for adaptive re-ID. UDA for person re-ID is an open-set problem where the source and target domain do not share the label space. Existing state of the art (Fu et al. 2019; Zhai et al. 2020; Fan et al. 2018; Ge, Chen, and Li 2020) generally pre-train the model on the source domain and then use the already-learned deep embedding to cluster the unlabeled samples on target domain. The generated pseudo labels are then used for training on the target domain. SSG (Fu et al. 2019) exploits the potential similarity of unlabeled samples to build multiple clusters automatically and then be assigned with pseudo labels. AD-Cluster (Zhai et al. 2020) augments person clusters in target domains to improve the ability of distinction with the augmented clusters. SPCL (Ge et al. 2020) proposes a unified framework to incorporate all available information from source and target domains. IDM (Dai et al. 2021b) generates intermediate domains' representations to bridge the link between the source and target domain. Different

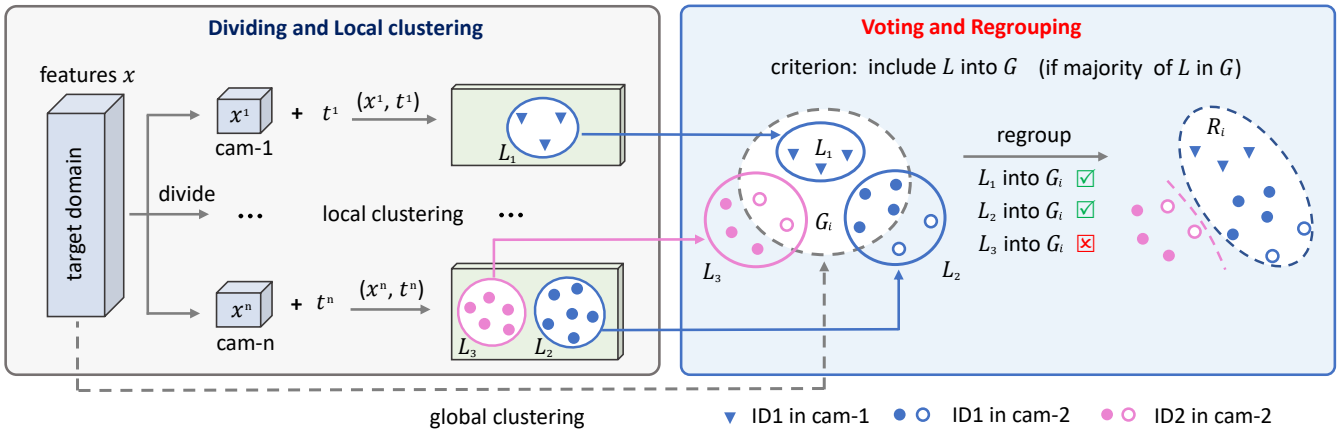

Figure 2: Pipeline of the proposed DARC. "Dividing and local clustering": Given the deep features x of the target domain, DARC first divides all the deep features into n groups (n is the total camera number) and performs independent local clustering within each camera. The local clustering has two advantages: 1) it uses temporal continuity of timestamp t to reduces ID switch errors and 2) it focuses on the local feature distribution within each individual camera and reduces ID split. "Voting and regrouping": Due to the dividing operation, each identity is naturally split into multiple local clusters. Therefore, DARC has to regroup the local clusters potentially belonging to the same identity. To this end, DARC performs another round of global clustering. If the majority instances in two local clusters appear in the same global cluster, we consider these two local clusters belonging to a single identity, yielding the voting-to-regroup criterion.

from the above methods, which adopt the standard clustering pipeline, we propose a novel plug-and-play clustering method, which can be compatible to many pseudo-label-based UDA methods.

Learn to refine the pseudo labels. Under cross-domain scenarios, the noise of pseudo labels is the dominant challenge to hinder the performance of UDA re-ID. Many methods propose to utilize the label-smoothing and solve the uncertainty to refine the soft pseudo labels. MMT (Ge, Chen, and Li 2020) learns features from the target domain via off-line refined hard pseudo labels and on-line refined soft pseudo labels in an alternative training manner. UNRN (Zheng et al. 2020) incorporates the uncertainty to re-weight the samples' contribution. NRMT (Zhao et al. 2020) proposes two networks during training to perform collaborative clustering. Different from all the above methods, we consider two essential characteristics of re-ID task and propose a novel two-stage clustering method, which reduces two major types of pseudo label noises, *i.e.* ID switch from local clustering and the ID split from cross-camera grouping, respectively.

Proposed Methods

Overview

Before clustering, we input the unlabeled samples (on the target domain) into the deep model and extract the deep features. Given the deep features, DARC performs a divide-and-regroup clustering pipeline, as illustrated in Fig. 2.

"*Dividing and local clustering*": DARC first divides all the deep features into K groups (K is the total number of cameras). Within each group, we conduct local clustering based on two distances, *i.e.*, the feature distances and the temporal distances. Local clustering has two advantages: 1) it utilizes temporal continuity to reduce ID switch and 2) it focuses on the feature distribution in each individual camera to reduce within-camera ID split (*i.e.*, splitting the images of a same person in a single camera into multiple clusters). The details are to be elaborated later.

"*Voting and regrouping*": Local clustering does not suffice the requirement for assigning pseudo labels. It is because the dividing operation naturally scatters the samples of each person across multiple cameras. Therefore, DARC needs to associate the local clusters which potentially belong to a same person. To this end, DARC merges the local clusters with the global clustering results for cross-camera association. Specifically, if the majority instances in two local clusters appear in a same global cluster, we consider these two local clusters as belonging to a single identity. In another word, during the regrouping operation, a local cluster is dominated by its majority samples, yielding a voting effect. We show that this voting procedure facilitates cross-camera association and still maintains the benefits of local clustering (low ID switch and within-camera ID split). The details are to be elaborated later.

With the above two stages, DARC fulfills the complete clustering procedure and assigns pseudo labels to the unlabeled samples. Afterwards, we may cooperate the pseudo labels with any existing pseudo-label-based adaptive training methods.

Local Clustering

Local clustering combines two types of distances, *i.e.*, the Euclidean distance D_{euc} between feature vectors and the temporal distance D_{temp} between the corresponding timestamps, to infer the overall distance $D_{overall}$ between sample

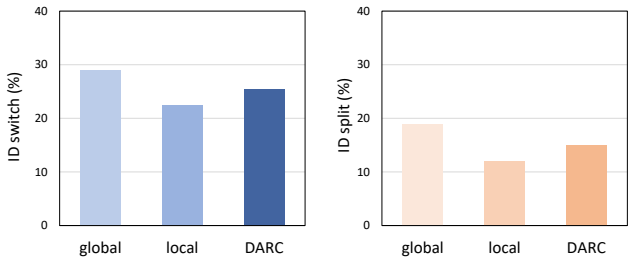

Figure 3: We compare two clustering errors (*i.e.*, ID switch and ID split) in global clustering, local clustering and the proposed DARC. The features are extracted on Market1501 using a deep model trained on PersonX.

pairs.

The temporal distance D_{temp} is defined as:

$$D_{temp}(i,j) = \frac{|t_i - t_j|}{\sqrt{\sum_{k=1}^{m}(t_i - t_k)^2}}, \tag{1}$$

where $D_{temp}(i,j)$ is the temporal distance between *i-th* and *j-th* samples, t_i, t_j denotes their timestamps, m denotes the number of samples in a single camera. We normalize the D_{temp} to make the distance quadratic sum equals 1 in each camera.

Intuitively, if the temporal distance D_{temp} between two occurrences is large, these two occurrences are likely to belong to different identities. We model this intuition into an overall distance $D_{overall}$ by:

$$D_{overall} = D_{euc} + \alpha D_{temp}, \tag{2}$$

where α is a hyper-parameter.

Local clustering reduces ID switch. We recall that ID switch happens when two inherently look-alike people are merged into a single cluster. With $D_{overall}$, if two samples are inherently similar (small D_{euc}) but are temporally far away from each other (large D_{temp}), local clustering may still distinguish them with relatively large $D_{overall}$.

Local clustering reduces within-camera ID split. We further analyze how local clustering reduces the ID split within the same camera. Our analysis is based on the commonly-adopted Jaccard distance. Specifically, we need to transform the $D_{overall}$ into the Jaccard distance. We revisit the standard pipeline of such a transform as follows:

1) Given a random sample p, we achieve its k nearest neighbors $\mathbb{N}(p,k)$ by ranking its $D_{overall}$. Afterwards, we obtain the corresponding k-reciprocal nearest neighbors $\mathbb{R}(p,k)$, which is formulated as:

$$\mathbb{R}(p,k) = \{q \mid (q \in \mathbb{N}(p,k)) \wedge (p \in \mathbb{N}(q,k))\} \tag{3}$$

2) The Jaccard distance between two samples p and q is calculated by the Jaccard metric of their k-reciprocal sets as:

$$D_j(p,q) = 1 - \frac{|\mathbb{R}(p,k) \cap \mathbb{R}(q,k)|}{|\mathbb{R}(p,k) \cup \mathbb{R}(q,k)|} \tag{4}$$

Let us consider the ranking list of a random sample p. We assume that q is a positive sample to p and p, q are in the same camera. After we replace the global clustering with the local clustering, the position of q in the ranking list should be raised. There are two reasons: First, some false-positive (look-alike but different) samples become farther from p. Second, the samples from other cameras are removed from the ranking list of p. Consequentially, the k-reciprocal nearest neighbor sets of p and q are prone to larger overlap (Zhong et al. 2017a; Bai and Bai 2016; Ye et al. 2016), which reduces the Jaccard distance between them. In another word, the Jaccard distance (within a single camera) between two positive samples becomes smaller, compared with the global Jaccard distance, resulting in higher within-class compactness and thus reducing the risk of ID split.

Voting and Regrouping

After local clustering, each local cluster only contains samples from a single camera. To regroup the samples potentially belonging to a same identity, we first perform a global clustering and use the global clusters as the clues for associating local clusters.

We recall that the global clustering requires comparison across different cameras and thus involves the multi-camera distribution problem. To alleviate the problem, we devise a "vote-to-regroup" strategy, as illustrated in Fig. 2. The key idea of the vote-to-regroup strategy is: if most samples of a local cluster belong to a specified global cluster G_k, we include this entire local cluster into G_k. Similarly, if the majority of samples in two local clusters belong to the same global cluster G_k, we associate these two local clusters as belonging to a single identity.

We formulate the above-described criterion as follows. Let us assume a local cluster L_i and a global cluster G_k. L_i is consisted of several samples, *i.e.*, $L_i = \{g_1^i, g_2^i, ...\}$. Given the local cluster L_i and global cluster G_k, we evaluate their overlap degree $P(L_i \to G_k)$ by:

$$P(L_i \to G_k) = \frac{|L_i \cap G_k|}{|L_i|} \tag{5}$$

where $|\cdot|$ denotes the number of samples in the set (local cluster). We use $P(L_i \to G_k)$ as the indicator for deciding whether L_i should be included into G_k, which is formulated as:

$$\begin{aligned} &G_k \leftarrow L_i \\ &s.t. \quad P(L_i \to G_k) \geq \delta \end{aligned} \tag{6}$$

where δ is a threshold.

When multiple local clusters satisfy Eq. 6, we merge these local clusters into a single G_k. After the regrouping process, the global cluster G_k becomes different from the original global clustering results, as shown in Fig. 2. Specifically, some local clusters with small overlap degree are excluded from G_k, while some local clusters with large overlap degree are included into G_k. Therefore, we mark the regrouped cluster as R_k.

The vote-to-regroup operation brings two benefits. On the one hand, it treats each local cluster as an inseparable unit

thus and inherits the advantage of high purity (from the local clustering). On the other hand, the voting procedure benefits the cross-camera aggregation with strong resistance against camera bias. Consequentially, DARC maintains high purity (low ID switch) and meanwhile reduces ID split.

Discussions on the clustering errors. Fig. 3 visualizes a statistic of ID split and ID switch errors on Market-1501 dataset (Zheng et al. 2015). We use the strong baseline (Ge et al. 2020) trained on PersonX(Sun and Zheng 2019) to extract the deep features. We note that during local clustering, each identity is separated into multiple local clusters. To evaluate the local clustering, we adopt an ideal assumption that local clusters are compared only within each camera. It does not take the error of cross cameras into consideration and thus indicates the lower error bound of DARC. After we regroup these local clusters, the clustering errors inevitably increase over the local clustering. However, DARC still maintains superiority (lower ID switch and ID split) against the global clustering.

Adaptive Training with Pseudo Labels

Given the final pseudo labels, DARC conducts a following adaptive training on the target domain. Let us assume the loss associated with a single sample x_i is $\mathscr{L}(x_i)$. We use a weighted sum of all the losses *w.r.t.* to the samples in the mini-batch, which is formulated as:

$$\mathscr{L}_{batch} = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot \mathscr{L}(x_i), \tag{7}$$

where N is the batch size, w_i is the weight factor corresponding to x_i. Please note that the detailed loss function for $\mathscr{L}$ can be implemented with any popular loss functions for re-ID, *e.g.*, the cross-entropy loss, the contrastive loss, etc. w_i denotes the weight of samples. We recall that during the vote-to-regroup procedure, a regrouped cluster R_k might absorb some samples which originally do not belong to the global cluster G_k. We assign a relatively small weight factor by setting $w_i = P(L \to G_k)$ (the overlap degree in Eq. 5) for these samples. As for the other samples, we set their weight factor to 1.

Experiments

Dataset

We evaluate the proposed DARC on different cross-domain scenes with two real person datasets, *i.e.*, Market1501 (Zheng et al. 2015), MSMT17 (Wei et al. 2018) and two synthetic person dataset PersonX (Sun and Zheng 2019), UnrealPerson (Zhang et al. 2021).

Implementation Details

We adopt ResNet-50 (He et al. 2016) pretrained on ImageNet (Deng et al. 2009) as the backbone. We construct each mini-batch with 64 source images (from 16 identities) and 64 target images (from 16 pseudo identities). Correspondingly, the batch size is 128. We resize the image size to 256×128 and utilize random flipping, random padding and random erasing (Zhong et al. 2017b) for data augmentation.

Methods		MA → MS		MS → MA	
		mAP	top-1	mAP	top-1
MMT	Oracle	43.6	69.3	82.4	92.3
	Orig	22.9	49.2	75.6	89.3
	DARC	24.2	50.4	78.5	90.9
SPCL	Oracle	46.7	72.7	83.5	93.1
	Orig	26.8	53.7	77.5	89.7
	DARC	27.3	54.9	80.9	91.7
IDM	Oracle	54.3	78.7	86.5	94.5
	Orig	33.5	61.3	82.1	92.4
	DARC	35.2	64.5	85.1	94.1

Table 1: DARC brings consistent improvement on popular re-ID UDA methods. "MA" : "Market", "MS" : "MSMT".

The essential clustering method for the local clustering and global clustering in DARC is DBSCAN (Ester et al. 1996). The training optimizer is Adam with 5×10^{-4} weight decay.

The Effectiveness of DARC

DARC brings general improvement. We validate the applicability of DARC on several popular UDA methods *i.e.*, MMT (Ge, Chen, and Li 2020), SPCL (Ge et al. 2020) and IDM (Dai et al. 2021b). For each method, we compare three modes: 1)"Oracle" uses the ground-truth labels for adaptive training. 2) "orig" uses the original clustering pipeline in the corresponding literature to assign pseudo labels. 3) "DARC" uses the proposed divide-and-regroup clustering pipeline to assign pseudo labels. The results are summarized in Table 1. It clearly shows that DARC consistently improves all three UDA methods. For example, on "MSMT→Market", DARC improves MMT, SPCL and IDM with +2.9%, +3.4% and +3.0% mAP, respectively. We note that IDM is a very recent UDA method and achieves competitive re-ID accuracy under cross-domain scenario. Using IDM as the baseline, DARC still gains considerable improvement and the achieved performance is very competitive (*e.g.*, 85.1% mAP and 94.1% top-1 on "MSMT→Market").

Comparison with the state of the art on real → real re-ID datasets. We compare the proposed DARC with state-of-the-art methods using two real-person datasets, Market1501 and MSMT17. The results are summarized in Table 2. It is observed that DARC achieves competitive cross-domain re-ID accuracy under all the two settings. Specifically, based on a relatively early method SPCL, DARC brings substantial improvement and achieves competitive performance compared with the state of the art. Based on a more recent method IDM, DARC still brings considerable improvement and advances the state of the art. In this paper, we report 85.1% and 94.1% top-1 accuracy on "MSMT → Market". Moreover, on the more challenging "Market→ MSMT", DARC achieves 64.5% top-1 accuracy.

Comparison with the state of the art on synthetic → real re-ID datasets. We conduct cross-domain experiments using the synthetic PersonX and UnrealPerson for training and two real-person datasets for testing. The comparison results between DARC and several state-of-the-art methods are sum-

Methods		Market → MSMT			
		mAP	top-1	top-5	top-10
MMCL (Wang and Zhang 2020)	CVPR20	15.1	40.8	51.8	56.7
MMT (Ge, Chen, and Li 2020)	ICLR20	22.9	49.2	63.1	68.8
DG-Net++ (Zou et al. 2020)	ECCV20	22.1	48.4	60.9	66.1
SPCL (Ge et al. 2020)	NIPS20	26.8	53.7	65.0	69.8
Dual-Refinement (Dai et al. 2021a)	TIP21	25.1	53.3	66.1	71.5
UNRN (Zheng et al. 2020)	AAAI21	25.3	52.4	64.7	69.7
GLT (Zheng et al. 2021)	CVPR21	26.5	56.6	67.5	72.0
IDM (Dai et al. 2021b)	ICCV21	33.5	61.3	73.9	78.4
SPCL + DARC	Ours	27.3	54.9	67.0	72.0
IDM + DARC	**Ours**	**35.2**	**64.5**	**76.2**	**80.4**

Methods		MSMT → Market			
		mAP	top-1	top-5	top-10
CASCL (Wu, Zheng, and Lai 2019)	ICCV19	35.5	65.4	80.6	86.2
DG-Net++ (Zou et al. 2020)	ECCV20	64.6	83.1	91.5	94.3
D-MMD++ (Mekhazni et al. 2020)	ECCV20	50.8	72.8	88.1	92.3
MMT (Ge, Chen, and Li 2020)	ICLR20	75.6	89.3	95.8	97.5
SPCL (Ge et al. 2020)	NIPS20	77.5	89.7	96.1	97.6
IDM (Dai et al. 2021b)	ICCV21	82.1	92.4	97.5	98.4
SPCL + DARC	Ours	80.9	91.7	97.0	98.2
IDM + DARC	**Ours**	**85.1**	**94.1**	**97.6**	**98.7**

Table 2: Comparison with the state-of-the-arts on two benchmarks Market-1501 and MSMT17.

marized in Table 3. Based on IDM, DARC advances the new state of the art (e.g, 83.1% mAP on "PersonX-to-Market"). It indicates that DARC is competent for adapting the knowledge learned from synthetic data to real-world testing data.

Comparison with the state of the art on unsupervised re-ID. Some methods (*i.e.* OIM (Xiao et al. 2017), BUC (Lin et al. 2019), SSL (Lin et al. 2020), HCT (Zeng et al. 2020), SPCL (Ge et al. 2020) evaluate on the unlabeled target domain without access to any source domain data, yielding the strict unsupervised re-ID. So far as we know, IDM does not offer compatibility to unsupervised re-ID. Therefore, we use SPCL as the baseline and summarize the unsupervised re-ID performance in Table 4. We observe that DARC achieves 79.4% mAP and 90.6% top-1 accuracy on Market-1501, surpassing the prior methods by a large margin.

Analysis on the time consumption. DARC only adds small time consumption to re-ID UDA. It is because each training epoch for UDA typically consists of a clustering procedure and a sequential adaptive training, and clustering only occupies a small proportion of the total time consumption. For example, on "PersonX→Market", a training epoch in the SpCL baseline costs about 196s (adaptive training) + 26s (global clustering). DARC increase the clustering time to 37s and maintains the adaptive training time unchanged. Therefore, DARC only incurs about +5% additional time consumption (222s→233s).

Ablation Study

Besides the novel divide-and-regroup framework, DARC has several good practices, which jointly contribute to its superiority. We investigate the contributions of the following factors:

- Using temporal continuity for local clustering.
- Using the overlap degree as the soft weight factor of training loss from individual samples.

The effectiveness of temporal continuity for local clustering. The temporal information in each camera is considered to suppress the ID switch. To be intuitive, the trajectory of the same person is likely to be consecutive in each single camera. It is effective to distinguish negative samples when different look-alike people appear in the same camera view at different time. Specially, the temporal continuity is only performed for local clustering.

We use **T** to denote the temporal information. As listed in Table 5, the temporal information improves the DARC (e.g, +1.3% mAP on "PersonX→Market") and achieves performance (e.g, 77.6% mAP on "PersonX→Market").

The effectiveness of soft weight factors. During adaptive training, the overall training loss in a mini-batch is a weighted sum of the losses from individual samples. The weight factor is the overlap degree (Eq. 5). We use **S** to denote soft weight. In Table 5, the soft weight factor improves the DARC (e.g, +1.1% mAP on "PersonX→Market") and achieves performance (e.g, 77.4% mAP on "PersonX→Market").

Analysis on Hyper-parameters

We analyze the impact of two important hyper-parameters, *i.e.*, α in the overall distance (Eq. 2) and δ in the criterion for regrouping local clusters (Eq. 6). We employ "IDM + DARC" to evaluate the performance on "PersonX → Market".

Methods		PersonX → Market				PersonX → MSMT17			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
MMT (Ge, Chen, and Li 2020)	ICLR20	71.0	86.5	94.8	97.0	17.7	39.1	52.6	58.5
SPCL (Ge et al. 2020)	NIPS20	73.8	88.0	95.3	96.9	22.7	47.7	60.0	65.5
IDM (Dai et al. 2021b)	ICCV21	81.3	92.0	97.4	98.2	30.3	58.4	70.7	75.5
SPCL + DARC	Ours	78.9	90.6	96.6	98.1	24.5	50.8	63.0	68.0
IDM + DARC	**Ours**	**83.1**	**93.1**	**97.7**	**98.5**	**32.3**	**61.3**	**73.5**	**78.2**

Methods		Unreal → Market				Unreal → MSMT17			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
JVTC (Li and Zhang 2020)	ECCV20	78.3	90.8	-	-	25.0	53.7	-	-
IDM (Dai et al. 2021b)	ICCV21	83.2	92.8	97.3	98.2	38.3	67.3	78.4	82.6
IDM + DARC	**Ours**	**85.1**	**93.8**	**98.0**	**98.7**	**39.1**	**68.6**	**79.6**	**83.7**

Table 3: Comparison with the state-of-the-arts on synthetic to real benchmarks.

Methods	MS → MA			
	mAP	top-1	mAP	top-1
BUC	38.3	66.2	79.6	84.5
SSL	37.8	71.7	83.8	87.4
MMCL	45.5	80.3	89.4	92.3
HCT	56.4	80.0	91.6	95.2
SPCL	73.1	88.1	95.1	97.0
SPCL+DARC	**79.4**	**90.6**	**96.6**	**98.1**

Table 4: DARC on unsupervised re-ID.

Methods	MS → MA		PX → MA	
	mAP	top-1	mAP	top-1
Baseline (SPCL)	77.5	89.7	73.8	88.0
+ DARC(w/o S&T)	79.0	90.7	76.3	89.0
+ DARC(w/o T)	79.4	91.0	77.4	89.5
+ DARC(w/o S)	79.7	91.1	77.6	89.8
DARC (Full)	**80.9**	**91.7**	**78.9**	**90.6**

Table 5: Ablation studies on two components. "S": using soft weight factor for training losses. "T": adding temporal distance (into the overall distance) for local clustering. "MS" denotes "MSMT", "PX" denotes "PersonX", "MA" denotes "Market".

In Fig.4 (a), we evaluate the impact of hyper-parameter α, which controls the weight of temporal distance in Eq. 2. We observe that the re-ID accuracy first increases (when α increases from 0 to 0.2) and then decreases (when α further increases to 1.0). We set $\alpha = 0.2$ as the weight factor.

In Fig.4 (b), we evaluate the impact of hyper-parameter δ, which denotes the threshold of overlap degree in Eq. 6. When the overlap score $P > \delta$, we consider that the majority of samples are within the global cluster and then include the entire local cluster into the global one. We set δ to vary from 0.4 to 1. It is observed that the re-ID accuracy undergoes an increase (when δ increases from 0.4 to 0.7) and then a decrease (when $\delta > 0.7$). Therefore, we set $\delta = 0.7$ as the optimized threshold for regrouping.

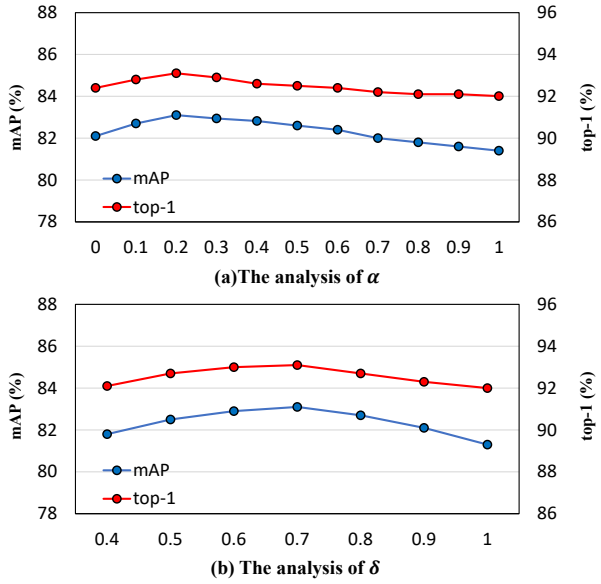

Figure 4: The analysis on Hyper-parameters α and δ.

Conclusion

This paper proposes a novel divide-and-regroup clustering (DARC) pipeline for UDA re-ID. We notice two characteristics of the re-ID task: *i.e.* multi-camera distribution and temporal continuity, which can respectively bring negative / positive impact on clustering accuracy. In response, DARC first divides all the target samples into multiple camera-specific groups to perform local clustering. The local clustering gains two benefits. First, it suppresses ID switch by cooperating temporal continuity. Second, it reduces ID split within each individual camera by focusing on the local feature distribution. Afterwards, DARC employs a vote-to-regroup strategy to associate local clusters across multiple cameras. Consequentially, the proposed DARC reduces both the ID switch and ID split errors. Experimental results confirm that DARC brings general improvements to many UDA methods and achieves state-of-the-art performance.

References

Bai, S.; and Bai, X. 2016. Sparse contextual activation for efficient visual re-ranking. *IEEE Transactions on Image Processing*, 25(3): 1056–1069.

Dai, Y.; Liu, J.; Bai, Y.; Tong, Z.; and Duan, L.-Y. 2021a. Dual-Refinement: Joint Label and Feature Refinement for Unsupervised Domain Adaptive Person Re-Identification. *IEEE Transactions on Image Processing*, 30: 7815–7829.

Dai, Y.; Liu, J.; Sun, Y.; Tong, Z.; Zhang, C.; and Duan, L.-Y. 2021b. IDM: An Intermediate Domain Module for Domain Adaptive Person Re-ID. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Deng, J.; Dong, W.; Socher, R.; Li, L.; Kai Li; and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, 226–231.

Fan, H.; Zheng, L.; Yan, C.; and Yang, Y. 2018. Unsupervised Person Re-Identification: Clustering and Fine-Tuning. *ACM Trans. Multimedia Comput. Commun. Appl.*, 14(4).

Fu, Y.; Wei, Y.; Wang, G.; Zhou, Y.; Shi, H.; and Huang, T. S. 2019. Self-Similarity Grouping: A Simple Unsupervised Cross Domain Adaptation Approach for Person Re-Identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Ge, Y.; Chen, D.; and Li, H. 2020. Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification. arXiv:2001.01526.

Ge, Y.; Zhu, F.; Chen, D.; Zhao, R.; and Li, H. 2020. Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. arXiv:2006.02713.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, J.; and Zhang, S. 2020. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In *European Conference on Computer Vision*, 483–499. Springer.

Lin, Y.; Dong, X.; Zheng, L.; Yan, Y.; and Yang, Y. 2019. A bottom-up clustering approach to unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 8738–8745.

Lin, Y.; Xie, L.; Wu, Y.; Yan, C.; and Tian, Q. 2020. Unsupervised person re-identification via softened similarity learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3390–3399.

Luo, H.; Gu, Y.; Liao, X.; Lai, S.; and Jiang, W. 2019a. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; and Gu, J. 2020. A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification. *IEEE Transactions on Multimedia*, 22(10): 2597–2609.

Luo, H.; Jiang, W.; Zhang, X.; Fan, X.; Qian, J.; and Zhang, C. 2019b. AlignedReID++: Dynamically matching local information for person re-identification. *Pattern Recognition*, 94: 53 – 61.

Mekhazni, D.; Bhuiyan, A.; Ekladious, G.; and Granger, E. 2020. Unsupervised domain adaptation in the dissimilarity space for person re-identification. In *European Conference on Computer Vision*, 159–174. Springer.

Quan, R.; Dong, X.; Wu, Y.; Zhu, L.; and Yang, Y. 2019. Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Sarfraz, M. S.; Schumann, A.; Eberle, A.; and Stiefelhagen, R. 2018. A Pose-Sensitive Embedding for Person Re-Identification With Expanded Cross Neighborhood Re-Ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sun, X.; and Zheng, L. 2019. Dissecting person re-identification from the viewpoint of viewpoint. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 608–617.

Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. 2018. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Wang, D.; and Zhang, S. 2020. Unsupervised Person Re-identification via Multi-label Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10981–10990.

Wang, G.; Yuan, Y.; Chen, X.; Li, J.; and Zhou, X. 2018. Learning Discriminative Features with Multiple Granularities for Person Re-Identification. In *Proceedings of the 26th ACM International Conference on Multimedia*, 274–282. Association for Computing Machinery.

Wang, Z.; Jiang, J.; Yu, Y.; and Satoh, S. 2019. Incremental Re-Identification by Cross-Direction and Cross-Ranking Adaption. *IEEE Transactions on Multimedia*, 21(9): 2376–2386.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 79–88.

Wei, L.; Zhang, S.; Yao, H.; Gao, W.; and Tian, Q. 2017. GLAD: Global-Local-Alignment Descriptor for Pedestrian Retrieval. In *Proceedings of the 25th ACM International Conference on Multimedia*, 420–428. Association for Computing Machinery.

Wu, A.; Zheng, W.-S.; and Lai, J.-H. 2019. Unsupervised person re-identification by camera-aware similarity consistency learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6922–6931.

Xiao, T.; Li, S.; Wang, B.; Lin, L.; and Wang, X. 2017. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3415–3424.

Ye, M.; Liang, C.; Yu, Y.; Wang, Z.; Leng, Q.; Xiao, C.; Chen, J.; and Hu, R. 2016. Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing. *IEEE Transactions on Multimedia*, 18(12): 2553–2566.

Zeng, K.; Ning, M.; Wang, Y.; and Guo, Y. 2020. Hierarchical clustering with hard-batch triplet loss for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13657–13665.

Zhai, Y.; Lu, S.; Ye, Q.; Shan, X.; Chen, J.; Ji, R.; and Tian, Y. 2020. AD-Cluster: Augmented Discriminative Clustering for Domain Adaptive Person Re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, T.; Xie, L.; Wei, L.; Zhuang, Z.; Zhang, Y.; Li, B.; and Tian, Q. 2021. UnrealPerson: An Adaptive Pipeline towards Costless Person Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11506–11515.

Zhang, X.; Cao, J.; Shen, C.; and You, M. 2019. Self-Training With Progressive Augmentation for Unsupervised Cross-Domain Person Re-Identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Zhang, X.; Luo, H.; Fan, X.; Xiang, W.; Sun, Y.; Xiao, Q.; Jiang, W.; Zhang, C.; and Sun, J. 2018. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. arXiv:1711.08184.

Zhao, F.; Liao, S.; Xie, G.-S.; Zhao, J.; Zhang, K.; and Shao, L. 2020. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *European Conference on Computer Vision*, 526–544. Springer.

Zheng, K.; Lan, C.; Zeng, W.; Zhang, Z.; and Zha, Z.-J. 2020. Exploiting Sample Uncertainty for Domain Adaptive Person Re-Identification. *arXiv preprint arXiv:2012.08733*.

Zheng, K.; Liu, W.; He, L.; Mei, T.; Luo, J.; and Zha, Z.-J. 2021. Group-aware label transfer for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5310–5319.

Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. 2015. Scalable Person Re-Identification: A Benchmark. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Zhong, Z.; Zheng, L.; Cao, D.; and Li, S. 2017a. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1318–1327.

Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; and Yang, Y. 2017b. Random Erasing Data Augmentation. arXiv:1708.04896.

Zou, Y.; Yang, X.; Yu, Z.; Kumar, B.; and Kautz, J. 2020. Joint Disentangling and Adaptation for Cross-Domain Person Re-Identification. *arXiv preprint arXiv:2007.10315*.