Online Pseudo Label Generation by Hierarchical Cluster Dynamics for Adaptive Person Re-identification

Yi Zheng*[1] Shixiang Tang*[3] Guolong Teng[1] Yixiao Ge[4] Kaijian Liu[1]
Jing Qin[5] Donglian Qi[1] Dapeng Chen†[2]
[1]Zhejiang University [2]AARC, Huawei [3]The University of Sydney
[4]The Chinese University of Hong Kong [5]The Hong Kong Polytechnic University
{21910099,qidl}@zju.edu.cn stan3906@uni.sydney.edu.au chendapeng8@huawei.com

Abstract

Adaptive person re-identification (adaptive ReID) targets at transferring learned knowledge from the labeled source domain to the unlabeled target domain. Pseudo-label-based methods that alternatively generate pseudo labels and optimize the training model have demonstrated great effectiveness in this field. However, the generated pseudo labels are inaccurate and cannot reflect the true semantic meaning of the unlabeled samples. We consider such inaccuracy stems from both the lagged update of the pseudo labels as well as the simple criterion of the employed clustering method. To tackle the problem, we propose an online pseudo label generation by hierarchical cluster dynamics for adaptive ReID. In particular, hierarchical label banks are constructed for all the samples in the dataset, and we update the pseudo labels of the sample in each coming mini-batch, performing the model optimization and the label generation simultaneously. A new hierarchical cluster dynamics is built for the label update, where cluster merge and cluster split are driven by a possibility computed by the label propagation. Our method can achieve better pseudo labels and higher reid accuracy. Extensive experiments on Market-to-Duke, Duke-to-Market, MSMT-to-Market, MSMT-to-Duke, Market-to-MSMT, and Duke-to-MSMT verify the effectiveness of our proposed method.

1. Introduction

Person re-identification (ReID) aims at retrieving the same person's images across different cameras [45, 13, 46, 22, 15, 17]. Despite the great advances of ReID with deep learning models [23, 34, 37, 44, 3], ReID still remains a challenge due to the various distributions of images captured by different camera systems, which is known as domain shift. Unsupervised Adaptive Person Re-identification (Adaptive ReID) is therefore proposed to transfer the learned knowledge from the labeled source domain (dataset) to correctly measure the inter-instance affinities in the unlabeled target domain (dataset).

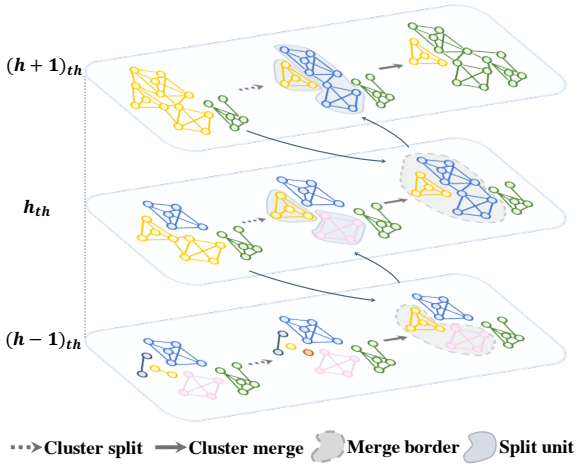


Figure 1. Schematic representation of online pseudo label refinement by hierarchical cluster dynamics. Different colors represent different clusters. Hierarchical labels are refined from low level to high level. In each level, clusters undergo the split and merge process in sequence (from left to right). For cluster split in h-th level, the clusters in $(h-1)$-th level serve as the indivisible split units. For cluster merge in h-th level, the cluster in $(h+1)$-th level provides a merge border which restricts the cluster candidates that can be merged. Label propagation is then leveraged to split and merge clusters in a unified framework.

Clustering-based methods [30, 54, 55, 14, 58] dominated the state-of-the-art performances for both adaptive and unsupervised ReID tasks. They adopted an iterative two-stage training scheme, where the pseudo labels are generated offline by clustering the extracted features before each training epoch. Although the clustered pseudo labels can

* These authors contributed equally.
† Corresponding author.

roughly capture the global distributions, we argue that the supervision signals by these labels are sub-optimal, since the offline label generation scheme fails to capture the varying feature distributions along with the network optimization on-the-fly. Furthermore, existing methods treated all the data points equally with the same clustering criterion, ignoring the fine-grained distributions of the complex real-world data. For example, person images captured by the same camera (domain) should be imposed a more strict criterion on clustering than those across cameras (domains).

To tackle the challenges above, we propose an online pseudo label generation strategy by exploring how to dynamically and progressively update the pseudo labels under a bottom-up framework. With the bottom-up framework, we can capture the fine-grained distributions of complex real-world data. With the well-designed dynamics for cluster split and merge, we can capture the instant semantic variation of the feature space. With the proposed method, the pseudo label generation and the feature learning can be conducted simultaneously in each training iteration.

Specifically, to enable the online pseudo label generation, we build a feature bank together with a hierarchical label bank for storing and updating the features and pseudo labels of all the unlabeled data, respectively. Given the encoded features of the samples in a mini-batch, the relevant sample features in the feature bank are momentum updated, and the corresponding labels are refined by merging to an existing cluster or splitting into a new cluster. The merge-and-split operation is achieved by propagating the pseudo labels among a set of clusters in terms of cluster-to-sample or sample-to-sample affinities, dubbed as "cluster dynamics" in our paper. Cluster dynamics is conducted iteratively to form a hierarchical pseudo label progression (as illustrated in Figure 1). The hierarchical label refinement structure also proves to be effective for capturing complex feature distribution of real images. Similar to mutualism in the ecosystem, the introduced cluster dynamics and the hierarchically progressed pseudo labels can properly capture the instant and diverse distribution variations in the feature space, leading to better performance of adaptive ReID.

The contributions of our work are summarized as threefolds. (1) We for the first time introduce to generate and update the pseudo labels online for adaptive ReID, where feature learning and pseudo label generation proceed simultaneously. (2) A novel cluster dynamics method is proposed to enable pseudo label progression and refinement iteratively in a bottom-to-up hierarchical structure. (3) Extensive experiments on multiple adaptive ReID benchmarks demonstrate the superiority of properly capturing the varying and fine-grained feature distributions by our method.

2. Related Work

Adaptive person re-identification. Adaptive person re-identification (adaptive ReID) has attracted much attention since it can save human efforts to image annotations. Generally, it can be divided into two categories, *i.e.,* pseudo-label-based methods [41, 14, 58, 15, 55, 64, 53] and domain-transfer-based methods [7, 17, 24]. Domain-transfer learning targets to learn domain-invariant features from style-transferred source domain images. SPGAN [9] and PT-GAN [48] transformed source domain images to match the image styles of the target domain while maintaining the original person identities. The style-transferred images and their identity labels were then used to fine-tune the model. HHL [64] learned camera-invariant features with camera style transferred images. The performance of domain-transferred methods largely relied on the quality of generated images and all of them ignored the valuable dependencies of images in the target domain.

Pseudo-label-based methods long stayed as state-of-the-arts for adaptive ReID. The first work [12] in this stream proposed to cluster features and optimize the network with pseudo labels alternately. SSG [14] introduced to assign hard pseudo labels for both global and local features, whose performance was largely hindered by the hard label noise. MMT [15] then focused on label refinery and proposed a mutual learning scheme for better pseudo labels. AD-Cluster [55] refined the clusters with augmented and generated images. The most related work to ours were SpCL [16] and BUC [30]. SpCL [16] applied a self-paced learning scheme to progressively generate more reliable clusters for the proposed unified contrastive loss. However, the pseudo labels in SpCL were updated after each epoch and thus could not capture the instant label variation of the extracted features, which became a bottleneck for pseudo label refinement. Our method proposes an online clustering framework, which could capture the varying distributions in the feature space. BUC [30] also proposed an online hierarchical clustering framework when generating pseudo labels. However, it only defined cluster merge criterion and no split mechanism was proposed. Therefore, there was no dynamical bi-directional cluster adjustment, *i.e.,* split and merge, in BUC. Additionally, BUC only leveraged the traditional criterion of hierarchical clustering which failed to consider the data distribution in the feature space.

Pseudo label generation Typical pseudo label generation can be divided into two categories: offline pseudo label generation and online pseudo label generation. Traditional clustering methods such as K-Means [33], spectral clustering [40], and DBSCAN [11] all belong to offline pseudo label generation and rely on certain assumptions on data distribution, such as convex shape, similar size, or same density of clusters. More recently, some offline but neural-network-based methods [5, 19, 52, 51, 18] were proposed to

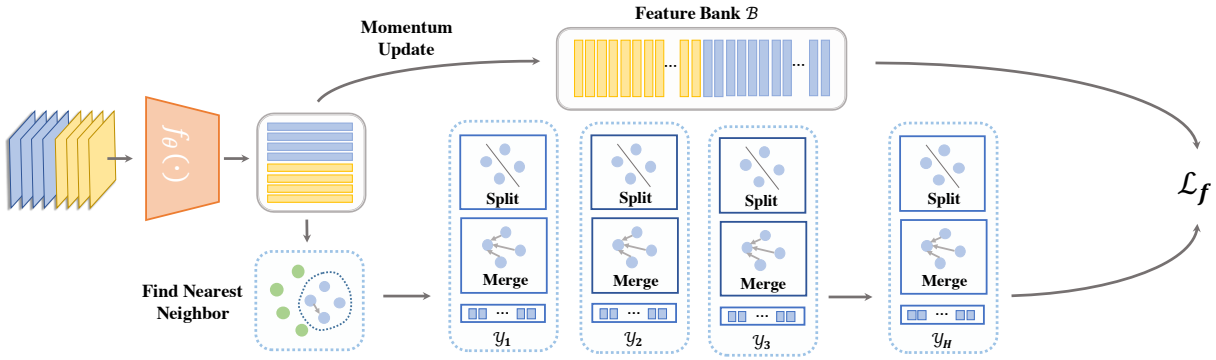


Figure 2. Illustration of the proposed framework. A mini-batch consists of the same number of source images (yellow) and target images (blue). The feature bank $\mathcal{B}$ is updated by extracted features in a momentum way. For online pseudo label generation, we first assign the label of the extracted feature as the label of the nearest neighbor and try to split/merge clusters containing both the new feature and its nearest neighboring feature (blue solid circles surrounded by dash blue circle). We implement hierarchical label refinement by split-merge sequence in each level. The pseudo labels generated in the top label bank are fed into contrastive loss.

perceive the data distribution from the training samples and thus they improved the traditional algorithms by relaxing above strong manual assumptions. While offline clustering alone achieved the state-of-the-art for features in the stable space, they were not the optimal solution for adaptive ReID, especially considering the varying feature distribution by network optimization.

To capture the varying feature distribution in adaptive ReID, online pseudo label generation might be a better solution because it assigned the label immediately after seeing a new sample [29]. Zhan *et.al.* [57] leveraged the online K-means clustering in training, achieving higher performance and efficiency for unsupervised learning. Swav [2] constructed an optimization problem for online clustering and incorporated it into contrastive learning. However, the number of classes in [57, 2] was a pre-defined hyper-parameter and was fixed during training, which was not practical in adaptive ReID. In contrast, our method sets the similarity threshold as a hyper-parameter and the number of clusters was adaptive. Therefore, our method can take the advantage of online clustering and get rid of the inappropriate class number in previous works.

3. Preliminary

Problem formulation. The goal of the adaptive person ReID is to learn a feature extractor f_θ that can generalize well on the target domain with labeled source domain data. The task provides two kinds of training data. One is the labeled data of source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_s}$, where each example $(\mathbf{x}_i^s, y_i^s)$ is composed of an image $\mathbf{x}_i^s \in \mathcal{X}^s$ and a label $y_i^s \in \mathcal{Y}^s$. The other is the unlabeled data of the target domain $\mathcal{D}_t = \{\mathbf{x}_i^t\}_{i=1}^{N_t}$. Here, N_s and N_t denote the number of images in the source domain and target domain, respectively. The source domain and the target domain usually have very different data distributions.

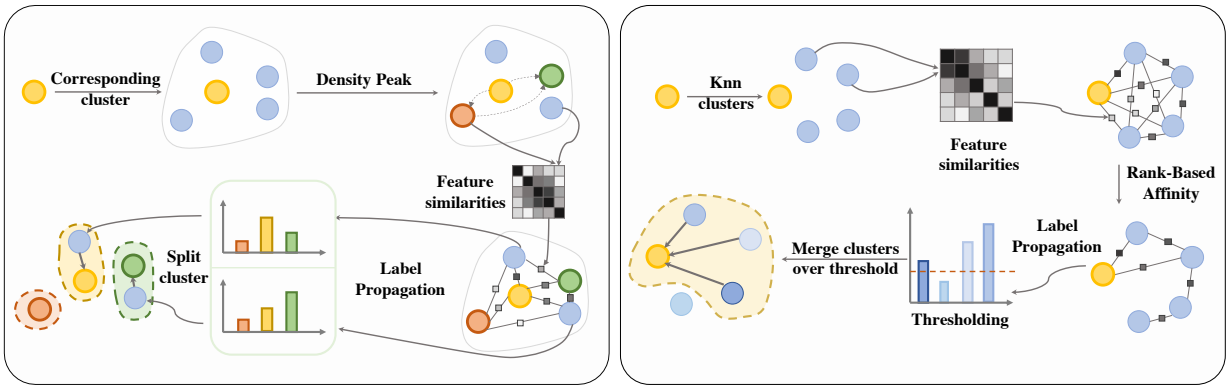
Pseudo-label-based approach by contrastive loss. Current pseudo-label-based approaches adopt an iterative and alternative two-stage pipeline for adaptive ReID. Specifically, they construct a feature bank $\mathcal{B} = \{\mathbf{v}_1^s, \mathbf{v}_2^s, ..., \mathbf{v}_{N_s}^s, \mathbf{v}_1^t, \mathbf{v}_2^t, ..., \mathbf{v}_{N_t}^t\}$, where $\mathbf{v}_i^s$ and $\mathbf{v}_i^t$ represent the stored feature of a source domain image and a target domain image, respectively. Before the training of each epoch, it performs offline clustering on the target domain data to generate the pseudo labels $\mathcal{Y} = \{1, 2, ..., N_t\}$. The cluster center $\mathbf{c}_k$ is the feature average of samples with pseudo label k. Considering the source domain data with ground truth labels and the target domain data with the pseudo labels, a contrastive loss function can be defined for each feature $\mathbf{f} = f_\theta(\mathbf{x})$ following SpCL [16] ,

$$\mathcal{L}_\mathbf{f} = -\log \frac{\exp(\langle \mathbf{f}, \mathbf{z}^+ \rangle / \tau)}{\sum_{k=1}^{n_s} \exp(\langle \mathbf{f}, \mathbf{w}_k \rangle / \tau) + \sum_{k=1}^{n_t} \exp(\langle \mathbf{f}, \mathbf{c}_k \rangle / \tau)}, \quad (1)$$

where $\mathbf{z}^+$ denotes the positive class prototype of $\mathbf{f}$. If $\mathbf{f}$ is a source domain feature, $\mathbf{z}^+ = \mathbf{w}_k$ is the center of the source domain class k that $\mathbf{f}$ belongs to. If $\mathbf{f}$ belongs to the k-th target domain cluster, $\mathbf{z}^+ = \mathbf{c}_k^+$ is the k-th cluster center. Besides, τ is the temperature empirically set to be 0.05, n_s is the number of source domain classes, n_t is the number of target domain clusters. The feature bank is initialized by the features extracted by ImageNet-pretrained models and is updated in a momentum way [49, 66, 6, 20, 42, 16].

4. Online Pseudo Label Generation

We incorporate a new online pseudo label generation mechanism into the contrastive learning framework, where the pseudo label generation is performed simultaneously with the model optimization. The overall framework is illustrated in Figure 2. In particular, the training samples in each batch are sampled from both the source domain $\mathcal{D}_s$ and the target domain $\mathcal{D}_t$. The newly extracted features of these samples are used to update the feature bank $\mathcal{B}$ in a momentum way. The target domain samples in a mini-batch are



(a) Cluster Split **(b) Cluster Merge**

Figure 3. Illustration of cluster split and merge in Dynamic Cluster Refinement. Yellow circles represent the unit containing the extracted features. (a) Cluster split in h-th level. Blue circles represent other split units in $(h-1)$-th level that belong to the same cluster with the yellow circle in h-th level. Red, yellow, and green circles mean the anchor units by density peak selection. The arrow in the light yellow and light green in the split cluster stage means the non-anchor units merging into the corresponding anchor unit. (b) Cluster merge in h-th level. Blue circles represent different units in h-th level but belong to the same cluster in $(h+1)$-th level, *i.e.,* units within the merge border. We merge clusters if the probability obtained by Eq. 2 is larger than a threshold σ.

fed into a *hierarchical label bank* $\mathcal{H}$. With the feature embeddings provided in $\mathcal{B}$, the coming-in samples undergo a series of merge-and-split operations in each level of the hierarchical bank, denoted by the *cluster dynamics*, to finally obtain reliable pseudo labels. Based on the features in $\mathcal{B}$ and the labels in $\mathcal{H}$, we optimize the loss function in Eq. 1. The pseudo-code of our proposed method is presented in the supplementary materials.

4.1. Hierarchical Label Bank

The hierarchical label bank $\mathcal{H} = \{\mathcal{Y}_1, \mathcal{Y}_2, ..., \mathcal{Y}_H\}$ consists of H levels and clusters the coming samples in a bottom-up manner. During clustering, it obeys a **cluster preserve property**. That is, the samples belonging to the same class in the low level should also belong to the same class in higher levels.

Based on this property, we perform a sequence of cluster-split and cluster-merge operations across different levels of the hierarchical label bank. Specifically, to split a cluster in the $(h+1)$-th level, we use the h-th level clusters after merge operation as the basic units and determine whether they still belong to the same cluster by label propagation. In this way, a cluster in the $(h+1)$-th level will be broken if the basic units are not so robust to be in the same cluster. Similarly, to merge the clusters in the h-th level, we only try to conflate those clusters belonging to one cluster in the $(h+1)$-th level. The clusters that have a high possibility of belonging to the same cluster after label propagation will be merged into one cluster. The update of the pseudo labels on h-th level bank $\mathcal{Y}_h$ is closely related to the clustering results of the adjacent levels, *i.e.,* $\mathcal{Y}_{h-1}$ and $\mathcal{Y}_{h+1}$, ensuring the robustness of the online pseudo label generation.

4.2. Cluster Dynamics

We now delve into the details of cluster dynamics in the h-th level, where the cluster split and cluster merge are performed by label propagation in a unified framework.

Label propagation. Label propagation in our method is used to compute the possibility of two samples/clusters belonging to the same class. It is an iterative algorithm and has a closed-form solution:

$$\mathbf{P}^* = (\mathbf{I} - \alpha \mathbf{S})^{-1} \mathbf{Y}_0, \tag{2}$$

where $\mathbf{P}^* = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n) \in \mathbb{R}^{n \times K}$ is the result of the label propagation. n is the number of units to be assigned label and K is the number of classes that n units may belong to. For the i-th unit, the results $\mathbf{p}_i = (p_{i,1}, ..., p_{i,K})$, where $p_{i,j}$ is the probability that the i-th unit takes label j. $\mathbf{Y}_0 \in \mathbb{R}^{n \times K}$ is the initial label probabilities that we will specifically define in cluster merge and cluster split. $\mathbf{S} \in \mathbb{R}^{n \times n}$ is the affinity matrix after normalization [32].

Inspired by re-ranking [62], the normalized affinity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ is not only based on cosine similarity but also based on the ranking order. Specifically, we denote $\mathcal{N}_k(\mathbf{c}_i)$ as the k-nearest neighbors of $\mathbf{c}_i$, where $\mathbf{c}_i$ is the feature center of the i-th unit. The affinities are computed between two units that are mutual neighbors of each other, which mathematically can be presented as

$$\hat{\mathbf{S}}_{ij} = \begin{cases} \langle \mathbf{c}_i, \mathbf{c}_j \rangle & \text{if } \mathbf{c}_i \in \mathcal{N}_k(\mathbf{c}_j) \text{ and } \mathbf{c}_j \in \mathcal{N}_k(\mathbf{c}_i), \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The normalized affinity matrix $\mathbf{S}$ in Eq. 2 is then obtained by $\hat{\mathbf{S}}$ following [32]. Because the sample pair with positive affinities are with high confidence to belong to the same class, Eq. 3 is more strict than the common cosine similarity, which ensures the clustering precision.

Cluster split. In the h-th level, cluster split divides one cluster $\mathcal{C}_i^h$ into small clusters. The split units are the clusters in the $(h-1)$-th level that contain the samples in $\mathcal{C}_i^h$. We denote these clusters as $\mathcal{O}_i^{h-1} = \{\mathcal{C}_{i,1}^{h-1}, \mathcal{C}_{i,2}^{h-1}, ..., \mathcal{C}_{i,n}^{h-1}\}$, where n is the cluster number. The corresponding center features are collected in $\mathbf{o}_i^{h-1} = \{\mathbf{c}_{i,1}^{h-1}, \mathbf{c}_{i,2}^{h-1}, ..., \mathbf{c}_{i,n}^{h-1}\}$. To split the cluster $\mathcal{C}_i^h$, we first select K anchor samples from $\mathbf{o}_i^{h-1}$, then leverage label propagation to assign labels of $\{1, .., K\}$ to the centers in $\mathbf{o}_i^{h-1}$. Therefore the clusters in $\mathcal{O}_i^{h-1}$ will have different labels. The samples in $\mathcal{C}_i^h$ finally borrow the labels in $\mathcal{O}_i^{h-1}$ to finish the cluster split.

In the process of the cluster split, we leverage the density peak selection [39] to choose K label anchors. With the label anchors, we initialize $(\mathbf{Y}_0^h)_{ij} = 1$ if $\mathbf{c}_i^{h-1}$ is the j-th label anchor, and $(\mathbf{Y}_0^h)_{ij} = 0$ otherwise. The affinity matrix is computed according to Eq. 3. After we get the closed-form solution $(\mathbf{P}^h)^*$ by Eq. 2, the k-th cluster $\mathcal{C}_{i,k}^{h-1}$ will be assigned a label by

$$y_k^h \leftarrow \arg\max \mathbf{p}_i^h. \tag{4}$$

In the label bank, we usually choose a new label index to indicate y_k^h in order to avoid label overlapping.

Cluster merge. At the h-th level, cluster merge aims to conflate small clusters that are highly possible to belong to the same semantic class. To obey the cluster preserve property, the clusters to be merged in h-th level should belong to the same cluster in the $(h+1)$-th level. Suppose there are n clusters in h-th level belonging to the same cluster $\mathcal{C}_i^{h+1}$ in the $(h+1)$-th level. These clusters are denoted by $\mathcal{O}_i^h = \{\mathcal{C}_{i,1}^h, \mathcal{C}_{i,2}^h, ..., \mathcal{C}_{i,n}^h\}$ and the corresponding center features are denoted by $\mathbf{o}^h = \{\mathbf{c}_{i,1}^h, \mathbf{c}_{i,2}^h, ..., \mathbf{c}_{i,n}^h\}$. To merge the clusters in $\mathcal{O}_i^h$, we first leverage label propagation to compute the possibility that two units belong to the same class and then use a threshold σ to determine whether they should be merged. Therefore some clusters in $\mathcal{O}_i^h$ will obtain the same label to finish the cluster merge.

In the process of cluster merge, the number of units is equal to the number of classes, *i.e.,* $n = K$, therefore we initialize $\mathbf{Y}_0^h = \mathbf{I}_{n\times n}$. With the same affinity matrix definition in cluster merge and the closed-form solution $(\mathbf{P}^h)^*$, we merge clusters $\mathcal{C}_u^h$ and $\mathcal{C}_v^h$, *i.e.,* $y_u^h \leftarrow y_v^h$, if $p_{u,v}^h > \sigma$. In the label bank, we usually re-organize labels to $\{1, 2, ..., N\}$, where N is the number of classes after merge in order to avoid label discontinuity.

5. Experiment

5.1. Experimental Setup

Dataset. We evaluate our method on three large-scale person re-ID datasets: Market-1501 [61], DukeMTMC-reID [38] and MSMT17 [48]. *Market-1501* consists of 32,668 annotated images of 1,501 identities taken by 6 cameras, for which 12,936 images of 751 identities are used for training and 19,732 images of 750 identities are in the test set. *DukeMTMC-reID* contains 36,411 labeled images belonging to 1,404 identities. It has 16,522 person images of 702 identities for training and the remaining images of other identities for testing. All images are collected from 8 cameras. *MSMT17* is the largest and most challenging dataset. It consists of 126,441 bounding boxes of 4,101 identities shot by 15 cameras, for which 32,621 images of 1,041 identities are used for training.

Evaluation protocol. We adopt Mean average precision(mAP) and cumulative matching characteristic (CMC) to evaluate the methods' performance on target domain datasets. No post-processing techniques such as re-ranking [62] or multi-query fusion [61] are implemented.

Implementation details. We implement our framework on PyTorch [36] and adopt ResNet50 [21] pretrained on ImageNet [8] as the backbone of extractor f_θ. We use Adam [27] with a weight decay of 0.0005 to optimize our network. The initial learning rate is set to 3.5×10^{-4} and decrease to 0.1 of its previous value every 8,000 iterations in total 24,000 iterations. Following SpCL [16], the temperature τ in Eq. 1 is set as 0.05 and the feature bank is updated with momentum equals 0.2. During training, each mini-batch contains 64 source domain images and 64 target domain images. The input images are resized to 256×128. We employ random flipping, random crop, and random erasing [63] for data augmentation. To save the computation and memory cost, we adopt a three-level hierarchical structure. For cluster dynamics, α in Eq. 2 is set to 0.99 for both cluster merge and cluster split. We adopt 30 k-nearest neighbors in Eq. 3 for all datasets and use a fixed threshold of 0.25 for cluster merge on all levels. During the split process, we select up to 8 anchor samples for each cluster.

5.2. Experimental Results

Adaptive person ReID. We compare our proposed framework with state-of-the-art methods on adaptive ReID in Table 1. Our method outperforms all current state-of-the-arts with a plain ResNet50 backbone. We compare our framework with the most competing method SpCL [16]. Similar to our method, SpCL also employs both feature bank and label bank for contrastive loss. However, it adopts the simple clustering algorithm, *i.e.,* DBSCAN, and only updates pseudo labels after each epoch, which may fail to capture the diverse and varying feature distribution. In contrast, our method can capture the varying feature distribution because it refines the labels and optimizes the feature extractor simultaneously, *i.e.,* refining labels for each coming mini-batch. For capturing diverse distributions, our method refines labels from bottom to up and leverages label propagation to dynamically split and merge clusters in each level by considering the neighboring information in feature space. Based on the differences above, our method leads to

Table 1. Comparison with state-of-the-art methods on domain adaptive person re-ID. "D", "M" and "MSMT" stand for DukeMTMC-ReID, Market-1501 and MSMT17, respectively. "Rk" denotes rank-k accuracy(%). "mAP" denotes mean average precision(%). "-" denotes not reported. The top-two results are highlighted with bold and underlined fonts, respectively.

Methods		D → M		M → D		D → MSMT		M → MSMT		MSMT → M		MSMT → D	
		mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
D-MMD [35]	ECCV'20	48.8	70.6	46.0	63.5	15.3	34.4	13.5	29.1	75.1	89.5	62.7	79.3
PCB-PAST [58]	ICCV'19	54.6	78.4	54.3	72.4	-	-	-	-	-	-	-	-
SSG [14]	ICCV'19	58.3	80.0	53.4	73.0	13.3	32.2	13.2	31.6	-	-	-	-
MMCL [43]	CVPR'20	60.4	84.4	51.4	72.4	16.2	43.6	15.1	40.8	-	-	-	-
ACT [50]	AAAI'20	60.6	80.5	54.5	72.4	-	-	-	-	-	-	-	-
SNR [26]	CVPR'20	61.7	82.8	58.1	76.3	-	-	-	-	-	-	-	-
DG-Net++ [67]	ECCV'20	61.7	82.1	63.8	78.9	22.1	48.8	22.1	48.4	64.6	83.1	58.2	75.2
ECN++ [65]	TPAMI'20	63.8	84.1	54.4	74.0	16.0	42.5	15.2	40.4	-	-	-	-
DAAM [25]	AAAI'20	67.8	86.4	63.9	77.6	21.6	46.7	20.8	44.5	-	-	-	-
AD-Cluster [55]	CVPR'20	68.3	86.7	54.1	72.6	-	-	-	-	-	-	-	-
JVTC+ [28]	ECCV'20	67.2	86.8	66.5	80.4	27.5	52.9	25.1	48.6	-	-	-	-
NRMT [59]	ECCV'20	71.7	87.8	62.2	77.8	20.6	45.2	19.8	43.7	-	-	-	-
DCML [4]	ECCV'20	72.6	87.9	63.5	79.3	-	-	-	-	-	-	-	-
MMT [15]	ICLR'20	73.8	89.5	65.1	78.0	25.1	52.9	24.0	50.1	-	-	-	-
MEB-Net [56]	ECCV'20	76.0	89.9	66.1	79.6	-	-	-	-	-	-	-	-
SpCL [16]	NeurIPS'20	77.5	89.7	-	-	-	-	26.8	53.7	-	-	-	-
UNRN [60]	AAAI'21	78.1	**91.9**	69.1	82.0	26.2	54.9	25.3	52.4	-	-	-	-
Ours		**80.0**	91.5	**70.1**	**82.2**	**29.3**	**56.1**	**28.4**	**54.9**	**80.2**	**91.4**	**71.2**	**83.1**

Table 2. Comparison with state-of-the-art methods on the unsupervised person re-ID task without the labeled source domain data. The notations are the same as those in Table 1.

Methods		Market-1501				DukeMTMC-reID				MSMT17			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
BUC [30]	AAAI'19	38.3	66.2	79.6	84.5	27.5	47.4	62.6	68.4	-	-	-	-
SSL [31]	CVPR'20	37.8	71.7	83.8	87.4	28.6	52.5	63.5	68.9	-	-	-	-
MMCL [43]	CVPR'20	45.5	80.3	89.4	92.3	40.2	65.2	75.9	80.0	11.2	35.4	44.8	49.8
HCT [54]	CVPR'20	56.4	80.0	91.6	95.2	50.7	69.6	83.4	87.4	-	-	-	-
CycAs [47]	ECCV'20	64.8	84.8	-	-	60.1	77.9	-	-	26.7	50.1	-	-
SpCL [16]	NeurIPS'20	73.1	88.1	95.1	97.0	-	-	-	-	19.1	42.3	55.6	61.2
Ours		**78.1**	**91.1**	**96.4**	**97.7**	**65.6**	**79.8**	**88.6**	**91.6**	**26.9**	**53.7**	**65.3**	**70.2**

2.5% improvements in mAP on Duke-to-Market and up to 3% gains in mAP on more challenging tasks like Duke-to-MSMT and Market-to-MSMT.

Unsupervised person ReID. Unsupervised person ReID focuses on training the ReID model without any labeled data, *i.e*, excluding source domain data from the training set. Our method can be easily generalized to such a setting by excluding $\mathbf{w}_k$ in Eq. 1. As shown in Table 2, our method surpasses all current state-of-the-arts by about 5% in mAP on Market-1501 and DukeMTMC-reID. We compared our method with the most similar BUC [30] method except SpCL. BUC proposed a bottom-up clustering approach, but it only merged two cluster pairs at each step without considering the distribution of clusters around. Moreover, it did not include the split step and keep the balance of clustering by setting a fixed speed of merging. The lack of split mechanism may take the BUC less effective than our method. Our method outperforms BUC by 39.8% and 38.1% in terms of mAP on Market-1501 and DukeMTMC-reID.

6. Ablation Study

To investigate the contribution of hierarchical clustering and cluster dynamics, we carry out component analysis of our approach in Table 3 on adaptive ReID tasks. "*Src.* only" only uses source domain images with ground-truth IDs for training and tests on target domain. "*Src.* + *tgt.* instance" treats each target domain sample as a distinct class. "*Src.* + *tgt.* HC *w/o* LP" adopts typical hierarchical clustering structure with average linkage criterion. "*Src.* + *tgt.* HC (merge) *w* LP" is the framework that removes cluster split dynamics and only allows cluster merge by label propagation. "*Src.* + *tgt.* LP *w/o* HC" leverages label propagation on 1-level

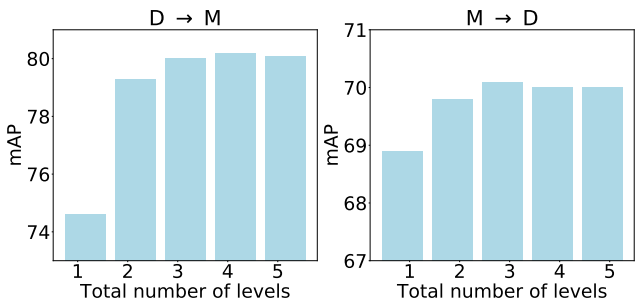


Figure 4. The performance in terms of mAP with different total number of levels, *i.e., H* in Sec. 4, on DukeMTMC-reID to Market-1501 and Market-1501 to DukeMTMC-reID.

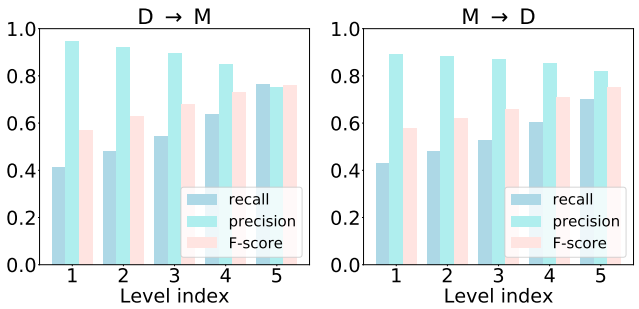


Figure 5. Performance comparison at different level indices. Blue, green, and pink results represent average recall, average precision, and F-score at each level, respectively.

Table 3. Ablation studies of our method on individual components.

Methods	D → M		M → D	
	mAP	R1	mAP	R1
Src. only	19.6	44.4	15.7	29.4
Src. + *tgt.* instance	6.0	14.8	4.7	10.2
Src. + *tgt.* HC *w/o* LP	53.3	74.8	59.0	75.0
Src. + *tgt.* HC (merge) *w* LP	75.0	90.0	61.7	78.0
Src. + *tgt.* LP *w/o* HC	74.6	89.0	68.9	82.2
Src. + *tgt.* HC *w* LP	**80.0**	**91.5**	**70.1**	**82.2**

Table 4. Comparison with different label updating intervals. The experiments are implements with $H = 3$. We report the results of every experiment in terms of mAP, R1, and R5.

Interval	D → M			M → D		
	mAP	R1	R5	mAP	R1	R5
0	80.0	91.5	96.3	70.1	82.2	89.7
1	74.2	89.3	95.5	67.4	80.6	89.0
2	68.3	86.3	93.8	64.0	77.8	87.3
3	64.2	82.5	91.7	62.2	76.6	86.5

bank. "*Src.* + *tgt.* HC *w* LP" is the proposed method that both hierarchical clustering and cluster dynamics (split and merge) by label propagation are adopted.

6.1. Effectiveness of Different Components

Effectiveness of hierarchical label generation. We first explore the necessity of clustering and then the effectiveness of hierarchical label generation. By comparing "*Src.* only" and "*Src.* + *tgt.* instance", we conclude replacing feature centers $\mathbf{c}_k$ by instance features $\mathbf{v}_k$ in Eq. 1 leads to lower performance than "*Src.* only". The result demonstrates that directly generalizing typical contrastive loss in unsupervised image classification to ReID is inapplicable and it is necessary to provide pseudo labels since ReID aims to explore intra-class and inter-class relations. For hierarchical label generation, we compare "*Src.* + *tgt.* LP *w/o* HC" and "*Src.* + *tgt.* HC *w* LP", concluding that hierarchical label generation provides better pseudo labels by improving the model 5.4% and 1.2% in mAP on Duke-to-Market and Market-to-Duke.

Effectiveness of cluster dynamics. Cluster dynamics can provide better pseudo labels by merging and splitting clusters automatically, while traditional clustering methods only include samples/clusters merge. To verify the effectiveness of the split mechanism, we compare results between "*Src.* + *tgt.* HC (merge) *w* LP" and "*Src.* + *tgt.* HC *w* LP". We can see cluster split improves 5.0% and 8.4% in terms of mAP on Duke-to-Market and Market-to-Duke, respectively, which proves the effectiveness of cluster split.

Effectiveness of label propagation. Cluster merge and cluster split use label propagation to update labels dynamically. To illustrate the effectiveness of label propagation, we conduct experiments that use average linkage as the criterion for cluster merge and split. With the comparison between "*Src.* + *tgt.* HC *w/o* LP" and "*Src.* + *tgt.* HC *w* LP", we can see that label propagation achieves 26.7% and 11.1% improvements in terms of mAP on Duke-to-Market and Market-to-Duke, respectively.

6.2. Hierarchical Label Generation

Number of levels. The design of hierarchical label generation aims to capture diverse feature distributions. In order to explore the influence of the number of hierarchical label bank levels, we conduct experiments on Duke-to-Market and Market-to-Duke with hierarchical label banks from 1 level to 5 levels. From Figure 4, we can observe that by changing the number of levels from 1 to 2, the result has a significant jump. When the number of levels continuously increases, the mAP only marginally increases and gradually comes to convergence. It demonstrates that for person ReID, increasing number of levels can get better pseudo labels and achieve an improvement in terms of mAP. As for the stagnant growth after 2 levels, we attribute this phenomenon to the simplicity of shooting conditions for academic datasets. In academic datasets, people in images with the similar background tend to have similar postures, which

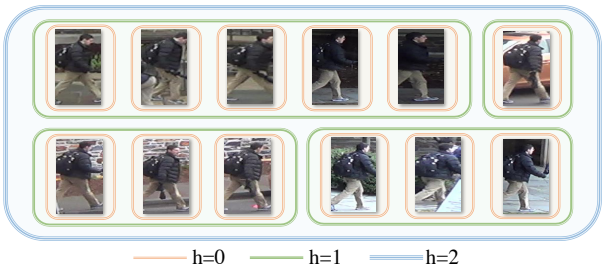


Figure 6. Visualization of hierarchical clustering results of one identity in DukeMTMC-reID.

makes the two-level structure have a sufficiently high effect.

Accuracy of pseudo labels at different levels. To get a deeper insight into the structure of hierarchies, we present average recall, average precision, and BCubed F-score [1] at different level indices with 5 levels in total, *i.e.,* $H = 5$, in Figure 5. We can observe that the average recall increases significantly but the average precision only drops slightly from 1_{st} level to 5_{th} level. The increasing average recall is mainly due to the capture of semantic information at high-level hierarchies while the slight drop in average precision is the unavoidable price for recall improving. We also report BCubed F-score to evaluate the overall performance of each level. The result shows that the BCubed F-score increases along with the level index, which indicates that hierarchical clustering improves the label accuracy effectively.

Visualization. In order to present the proposed hierarchical label generation more intuitively, we visualize the clustering results at different levels. As shown in Figure 6, at the low level ($h = 1$), samples in the same cluster share high similarities, such as the same shooting angle and the same background. At the high level ($h = 2$), samples of the same identity with larger variations in human posture and background are gathered.

6.3. Cluster Dynamics

Label updating interval. Compared with other state-of-the-arts utilizing offline clustering, the advantage of our online pseudo label generation is to capture the varying feature space by network optimization and pseudo label refinement simultaneously. We conduct experiments to explore the influence of label updating intervals. The more interval is, the greater the inconsistent with network optimization and pseudo labels is. As shown in Table 4, we can observe that the testing result drops as the update interval increases. This trend illustrates the importance of updating pseudo labels and features simultaneously.

Anchor sample selection methods in cluster split. Anchor samples serve as seeds for cluster split. We consider better seeds for clusters will benefit the final performance. In order to explore the influence of the anchor sample selection method, we conduct experiments with different methods including random choosing, farthest point sampling [10], and density peak [39]. As shown in Table 5, density peak surpasses other methods when tested on Market-to-Duke and Duke-to-Market. This might be attributed to the fact that high-density samples are more representative, which will increase the accuracy of split results.

Table 5. Performance comparison with different selection methods. The total number of levels H is set to 3 in these experiments. Random, FPS, and density peak stand for cluster split dynamics use random choosing, farthest point sampling, and density peak to select anchor samples in cluster split procedure, respectively.

Methods	D → M			M → D		
	mAP	R1	R5	mAP	R1	R5
Random	75.0	89.4	95.4	69.2	82.2	90.5
FPS	75.4	89.1	95.5	69.6	82.0	90.3
Density peak	80.0	91.5	96.3	70.1	82.2	89.7

Table 6. Comparison with different values of split α_S and merge α_M in Eq. 2. We report mAP on DukeMTMC-reID to Market-1501 and Market-1501 to DukeMTMC-reID. The total number of levels H is 3 in these experiments.

α_M	D → M	M → D	α_S	D → M	M → D
0.1	66.9	56.4	0.3	76.1	68.8
0.3	68.0	58.5	0.5	76.3	69.0
0.5	69.2	63.3	0.7	76.4	69.2
0.7	70.7	64.6	0.9	77.1	69.2
0.9	78.9	69.0	0.95	77.9	69.5
0.99	80.0	70.1	0.99	80.0	70.1

Alpha. To investigate the influence of α in Eq. 2, we compare the performance in terms of mAP with different merge α_M and split α_S. As shown in Table 6, our method achieves the best performance when both α_M and α_S are set to 0.99. The results suggest that regardless of cluster merge and cluster split, the performance improves when more information about surrounding neighbors is considered.

7. Conclusion

Pseudo label noise is one of the bottlenecks for further improvements of clustering-based adaptive person re-identification methods. In this paper, we consider such label noise stems from the lagged update of the pseudo labels and the simple criterion of the employed clustering algorithms. We propose an online label generation with hierarchical cluster dynamics for adaptive person re-identification. With online label generation, we refine the labels of samples in the coming mini-batch and optimize the network simultaneously. With hierarchical cluster dynamics, we split and merge clusters in a bottom-up framework, capturing diverse and complex feature distributions. Extensive experiments on various datasets of person re-identification verify the effectiveness of our proposed method.

References

[1] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo. Erratum to a comparison of extrinsic clustering evaluation metrics based on formal constraints (inf retrieval, (2009), 12, (461-486), 10.1007/s10791-008-9066-8). *Information Retrieval*, 12(5), 2009. 8

[2] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020. 3

[3] Dapeng Chen, Zejian Yuan, Badong Chen, and Nanning Zheng. Similarity learning with spatial constraints for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1268–1277, 2016. 1

[4] Guangyi Chen, Yuhao Lu, Jiwen Lu, and Jie Zhou. Deep credible metric learning for unsupervised domain adaptation person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 643–659, Cham, 2020. Springer International Publishing. 6

[5] Jie Chen, Tengfei Ma, and Cao Xiao. Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv preprint arXiv:1801.10247*, 2018. 2

[6] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 3

[7] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Instance-guided context rendering for cross-domain person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 232–242, 2019. 2

[8] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 5

[9] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 994–1003, 2018. 2

[10] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi. The farthest point strategy for progressive image sampling. *IEEE Transactions on Image Processing*, 6(9):1305–1315, 1997. 8

[11] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996. 2

[12] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):1–18, 2018. 2

[13] Hao Feng, Minghao Chen, Jinming Hu, Dong Shen, Haifeng Liu, and Deng Cai. Complementary pseudo labels for unsupervised domain adaptation on person re-identification. *IEEE Transactions on Image Processing*, 30:2898–2907, 2021. 1

[14] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, U. Uiuc, and T. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6111–6120, 2019. 1, 2, 6

[15] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations*, 2020. 1, 2, 6

[16] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 11309–11321. Curran Associates, Inc., 2020. 2, 3, 5, 6

[17] Yixiao Ge, Feng Zhu, Rui Zhao, and Hongsheng Li. Structured domain adaptation with online relation regularization for unsupervised person re-id. *arXiv e-prints*, pages arXiv–2003, 2020. 1, 2

[18] Senhui Guo, Jing Xu, Dapeng Chen, Chao Zhang, Xiaogang Wang, and Rui Zhao. Density-aware feature embedding for face clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6698–6706, 2020. 2

[19] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216*, 2017. 2

[20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020. 3

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 5

[22] Xuanyu He, Wei Zhang, Ran Song, and Xiangyuan Lan. Take more positives: A contrastive learning framework for unsupervised person re-identification. *arXiv preprint arXiv:2101.04340*, 2021. 1

[23] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 1

[24] Fa-Ting Hong, Wei-Hong Li, and Wei-Shi Zheng. Learning to detect important people in unlabelled images for semi-supervised important people detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4146–4154, 2020. 2

[25] Yangru Huang, Peixi Peng, Yi Jin, Yidong Li, and Junliang Xing. Domain adaptive attention learning for unsupervised person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:11069–11076, 04 2020. 6

[26] X. Jin, C. Lan, W. Zeng, Z. Chen, and L. Zhang. Style normalization and restitution for generalizable person re-identification. In *2020 IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (CVPR), pages 3140–3149, 2020. 6

[27] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014. 5

[28] Jianing Li and Shiliang Zhang. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 483–499, Cham, 2020. Springer International Publishing. 6

[29] Edo Liberty, Ram Sriharsha, and Maxim Sviridenko. An algorithm for online k-means clustering. In *2016 Proceedings of the eighteenth workshop on algorithm engineering and experiments (ALENEX)*, pages 81–89. SIAM, 2016. 3

[30] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8738–8745, 2019. 1, 2, 6

[31] Y. Lin, L. Xie, Y. Wu, C. Yan, and Q. Tian. Unsupervised person re-identification via softened similarity learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3387–3396, 2020. 6

[32] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sungju Hwang, and Yi Yang. LEARNING TO PROPAGATE LABELS: TRANSDUCTIVE PROPAGATION NETWORK FOR FEW-SHOT LEARNING. In *International Conference on Learning Representations*, 2019. 4

[33] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982. 2

[34] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1

[35] Djebril Mekhazni, Amran Bhuiyan, George Ekladious, and Eric Granger. Unsupervised domain adaptation in the dissimilarity space for person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 159–174, Cham, 2020. Springer International Publishing. 6

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 5

[37] Ruijie Quan, Xuanyi Dong, Yu Wu, Linchao Zhu, and Yi Yang. Auto-reid: Searching for a part-aware convnet for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3750–3759, 2019. 1

[38] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set formulti-target, multi-camera tracking. In Gang Hua and Hervé Jégou, editors, *Computer Vision – ECCV 2016 Workshops*, pages 17–35, Cham, 2016. Springer International Publishing. 5

[39] Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *science*, 344(6191):1492–1496, 2014. 5, 8

[40] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000. 2

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

[42] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019. 3

[43] D. Wang and S. Zhang. Unsupervised person re-identification via multi-label classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10978–10987, 2020. 6

[44] Guangcong Wang, Jianhuang Lai, Peigen Huang, and Xiaohua Xie. Spatial-temporal person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8933–8940, 2019. 1

[45] Menglin Wang, Baisheng Lai, Jianqiang Huang, Xiaojin Gong, and Xian-Sheng Hua. Camera-aware proxies for unsupervised person re-identification. *arXiv preprint arXiv:2012.10674*, 2020. 1

[46] Wenhao Wang, Shengcai Liao, Fang Zhao, Cuicui Kang, and Ling Shao. Domainmix: Learning generalizable person re-identification without human annotations. *arXiv preprint arXiv:2011.11953*, 2020. 1

[47] Zhongdao Wang, Jingwei Zhang, Liang Zheng, Yixuan Liu, Yifan Sun, Yali Li, and Shengjin Wang. Cycas: Self-supervised cycle association for learning re-identifiable descriptions. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 72–88, Cham, 2020. Springer International Publishing. 6

[48] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79–88, 2018. 2, 5

[49] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, 2018. 3

[50] Fengxiang Yang, Ke Li, Zhun Zhong, Zhiming Luo, Xing Sun, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Shaozi Li. Asymmetric co-teaching for unsuper-

vised cross-domain person re-identification. In *AAAI*, pages 12597–12604, 2020. 6

[51] Lei Yang, Dapeng Chen, Xiaohang Zhan, Rui Zhao, Chen Change Loy, and Dahua Lin. Learning to cluster faces via confidence and connectivity estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13369–13378, 2020. 2

[52] Lei Yang, Qingqiu Huang, Huaiyi Huang, Linning Xu, and Dahua Lin. Learn to propagate reliably on noisy affinity graphs. *arXiv preprint arXiv:2007.08802*, 2020. 2

[53] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2148–2157, 2019. 2

[54] Kaiwei Zeng, Munan Ning, Yaohua Wang, and Yang Guo. Hierarchical clustering with hard-batch triplet loss for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13657–13665, 2020. 1, 6

[55] Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9021–9030, 2020. 1, 2, 6

[56] Yunpeng Zhai, Qixiang Ye, Shijian Lu, Mengxi Jia, Rongrong Ji, and Yonghong Tian. Multiple expert brainstorming for domain adaptive person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 594–611, Cham, 2020. Springer International Publishing. 6

[57] Xiaohang Zhan, Jiahao Xie, Ziwei Liu, Yew-Soon Ong, and Chen Change Loy. Online deep clustering for unsupervised representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6688–6697, 2020. 3

[58] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8222–8231, 2019. 1, 2, 6

[59] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 526–544, Cham, 2020. Springer International Publishing. 6

[60] Kecheng Zheng, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, and Z. Zha. Exploiting sample uncertainty for domain adaptive person re-identification. *ArXiv*, abs/2012.08733, 2020. 6

[61] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015. 5

[62] Z. Zhong, L. Zheng, D. Cao, and S. Li. Re-ranking person re-identification with k-reciprocal encoding. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3652–3661, 2017. 4, 5

[63] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020. 5

[64] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 598–607, 2019. 2

[65] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang. Learning to adapt invariance in memory for person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020. 6

[66] Chengxu Zhuang, Alex Lin Zhai, and Daniel Yamins. Local aggregation for unsupervised learning of visual embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6002–6012, 2019. 3

[67] Yang Zou, Xiaodong Yang, Zhiding Yu, B. V. K. Vijaya Kumar, and Jan Kautz. Joint disentangling and adaptation for cross-domain person re-identification. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 87–104, Cham, 2020. Springer International Publishing. 6