
Uncertainty-aware Clustering for Unsupervised Domain Adaptive Object Re-identification

Pengfei Wang¹ Changxing Ding^{1*} Wentao Tan¹ Mingming Gong² Kui Jia¹ Dacheng Tao³

¹ South China University of Technology

² University of Melbourne

³ JD Explore Academy, China

{eepengfei.wang, eewentaotan}@mail.scut.edu.cn, {chxding, kuijia}@scut.edu.cn,
mingming.gong@unimelb.edu.au, taodacheng@jd.com

Abstract

Unsupervised Domain Adaptive (UDA) object re-identification (Re-ID) aims at adapting a model trained on a labeled source domain to an unlabeled target domain. State-of-the-art object Re-ID approaches adopt clustering algorithms to generate pseudo-labels for the unlabeled target domain. However, the inevitable label noise caused by the clustering procedure significantly degrades the discriminative power of Re-ID model. To address this problem, we propose an uncertainty-aware clustering framework (UCF) for UDA tasks. First, a novel hierarchical clustering scheme is proposed to promote clustering quality. Second, an uncertainty-aware collaborative instance selection method is introduced to select images with reliable labels for model training. Combining both techniques effectively reduces the impact of noisy labels. In addition, we introduce a strong baseline that features a compact contrastive loss. Our UCF method consistently achieves state-of-the-art performance in multiple UDA tasks for object Re-ID, and significantly reduces the gap between unsupervised and supervised Re-ID performance. In particular, the performance of our unsupervised UCF method in the MSMT17→Market1501 task is better than that of the fully supervised setting on Market1501. The code of UCF is available at <https://github.com/Wang-pengfei/UCF>.

1 Introduction

The goal of object re-identification (Re-ID) is to retrieve object images belonging to the same identity across different camera views. Due to its broad range of potential applications, (*e.g.*, smart retail), Re-ID research has experienced explosive growth in recent years [66, 24, 45, 56, 30, 53, 6, 57, 22, 51, 15, 48, 67, 64]. Most existing approaches achieve remarkable performance when the training and testing data are drawn from the same domain. However, due to the presence of significant domain gaps, Re-ID models trained on source datasets typically exhibit clear performance drops when directly applied to the target datasets. Unsupervised Domain Adaptive (UDA) object Re-ID is therefore proposed to adapt the model trained on the source image domain with identity labels to the target image domain without the need for identity annotations. Unlike the traditional UDA setting, which assumes that both domains share the same classes, UDA in object Re-ID is a more challenging open-set problem, in that the two domains have totally different identities (classes).

State-of-the-art methods [43, 11, 62, 9, 65, 49, 68, 23, 13] adopt clustering algorithms to generate pseudo-labels for the target domain. At the beginning of each epoch, a clustering algorithm is applied on the features extracted from the current model to generate pseudo-labels for each image. The

*Corresponding Author.

current model is then updated via retraining with the pseudo-labels. These two steps alternate so that the model gradually adapts to the target data. While pseudo-label approaches have achieved promising results, there are still two major challenges to deal with. First, due to the domain gap, the current model is not an optimal feature extractor for the target domain; second, the unsupervised nature of the clustering makes it difficult to obtain the real identity labels, even given the optimal feature extractor. The obtained pseudo-labels therefore usually contain a certain level of noise, which undermines the final Re-ID performance.

In this paper, we propose an uncertainty-aware clustering framework (UCF) to handle the above problem from two perspectives. First, we identify and decompose unreliable clusters using a novel hierarchical clustering algorithm. Due to the domain shift, the Re-ID model has limited discriminative power in the target domain; as a result, inter-class distances may vary dramatically. This means that images of visually similar identities may be grouped into the same cluster, the size of which tends to be large. To handle this problem, we first adopt a clustering algorithm, such as DBSCAN [7], to perform coarse clustering. We then calculate the silhouette coefficient [42], which measures both the tightness and separation of each cluster. For clusters with small silhouette coefficient, we further perform fine-grained clustering within the cluster. In this way, unreliable clusters can be decomposed into several smaller ones.

Second, we identify images with unreliable pseudo-labels using a novel uncertainty-aware collaborative instance selection method. Specifically, we adopt a deep network and its temporally averaged model, *i.e.*, the mean-Net [47], to cluster images in the target domain, respectively. Since these two models have different learning capabilities, their clustering results will be different. We then evaluate whether each instance is located in similar clusters across the two networks. If a large number of overlapping samples exist in the two clusters, the clustering result of this instance is considered to be reliable. Finally, we only adopt instances with reliable labels for model training, which reduces the impact of noise in the pseudo-labels.

Through joint hierarchical clustering and reliable sample selection, our UCF framework can effectively reduce the adverse effects of noisy pseudo-labels. We further propose a compact contrastive loss for UDA Re-ID. Recent approaches [13, 73, 23] typically adopt contrastive loss for model training. However, these losses require all image features in the target domain to be stored in the memory bank. This may result in two problems: first, this strategy consumes a lot of memory; second, only the features of a small number of images are updated in each iteration. These problems become especially serious for large-scale Re-ID datasets, such as MSMT17 [52]. To solve the above mentioned problems, we propose an improved contrastive loss using a class-level memory bank, which stores one single feature vector for each class rather than the features of all images.

Our main contributions can be summarized as follows: 1) We propose a strong baseline that adopts an improved contrastive loss using compact class-level memory banks; 2) We design a hierarchical clustering scheme to improve the quality of clustering, which decomposes unreliable clusters from coarse to fine; 3) We introduce a novel collaborative clustering method to identify images with unreliable pseudo-labels, which significantly relieves the impact of noise in pseudo-labels; 4) Our approach outperforms state-of-the-art methods by large margins on many UDA tasks for Re-ID.

2 Related Works

We review the literature in three parts: 1) unsupervised domain adaptive (UDA) object Re-ID, 2) contrastive learning, and 3) deep learning with noisy labels.

2.1 UDA Object Re-ID

Existing UDA approaches for object Re-ID can be roughly divided into two categories: pseudo-label-based methods [43, 9, 65, 11, 62, 73, 60, 49] and domain translation-based methods [5, 52, 2, 14]. Domain translation-based methods transfer labeled images in the source domain to the style of the target domain images, then use these transferred images and the inherited ground-truth labels for model training. However, a gap inevitably arises between the translated image and the real target domain image, which affects the performance of these approaches. Pseudo-label-based methods group unannotated images using clustering algorithms and then train the network with pseudo-labels generated by clustering. For example, Li *et al.*[23] employed both visual and temporal similarity cues

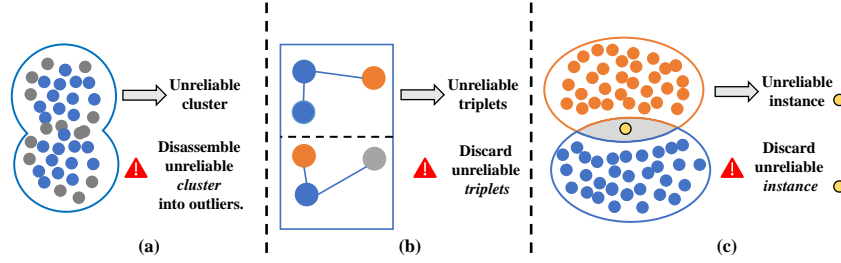


Figure 1. (a) SpCL [11] regards all instances in one unreliable cluster as outliers. (b) NRMT [68] removes unreliable triplets. (c) Our UCF measures the uncertainty of each instance, which is more fine-grained. (Best viewed in color.)

to promote the quality of pseudo-labels. However, existing approaches typically ignore the noise remaining in pseudo-labels.

Recently, some methods have been proposed that attempt to solve the label noise problem. Ge *et al.* [11] proposed generating more robust soft labels via mutual mean-teaching. However, the classifier trained with noisy labels forms the foundation for soft label generation, which hinders the improvement of Re-ID performance. Ge *et al.* [13] further proposed the SpCL approach. As shown in Fig. 1(a), it identifies and regards all instances in one unreliable cluster as outliers with reference to their proposed reliability criterion. However, removing all images in a cluster may waste samples with reliable pseudo-labels. Similarly, Zhao *et al.* [68] introduced the Noise Resistible Mutual-Training (NRMT) approach, as shown in Fig. 1(b), which removes triplets that are considered to be unreliable. The reliability of a triplet is measured with reference to the distance between the features of the triplet samples extracted by two networks. Unlike the above works, our approach is more fine-grained, as it first improves the clustering quality and then removes unreliable instances rather than complete clusters or triplets.

2.2 Contrastive Learning

As a promising paradigm of unsupervised learning, contrastive learning has lately achieved state-of-the-art performance in unsupervised visual representation learning. Recently, contrastive learning methods combined with data augmentation strategies achieved great successes, such as SimCLR [1], MoCo [18], and BYOL [16]. These methods treat each instance as a class represented by a feature vector and data pairs are constructed through data augmentations. These methods treat each instance as a class, which yields poor results for the domain adaptive object Re-ID task, because the intra- and inter-class similarity on the unlabeled target domain cannot be measured accurately. Some recent works [49, 13, 23, 73] have introduced improved contrastive loss to domain adaptation. For example, the SpCL approach [13] includes a unified contrastive loss, which jointly distinguishes source-domain classes, target-domain clusters, and un-clustered instances. One common drawback of these methods is the need to store all instance features, which requires a large amount of memory. To solve this problem, we propose a new contrastive loss with a compact class-level memory bank, which resolves these issues by storing a single feature vector for each cluster rather than all instance features.

2.3 Deep Learning with Noisy Labels

Many studies have attempted to effectively train deep neural networks in the presence of noisy labels for close-set classification problems. Some recent works have introduced a sample selection approach that selects data with reliable labels for training [31, 33]. Notably, the small loss trick, which regards samples with small training loss as clean, has demonstrated powerful ability. However, the small loss trick is not suitable to select clean samples in UDA object Re-ID task. This is because the number of target domain clusters (classes) changes through re-clustering during the training process. Moreover, recent studies suggest various ways in which additional performance gain can be achieved by maintaining two networks to avoid accumulating sampling bias [17, 61]. For example, Co-teaching [17] works by training two deep models simultaneously, where each network selects the small-loss instances as reliable samples for the other one. These methods focus primarily on

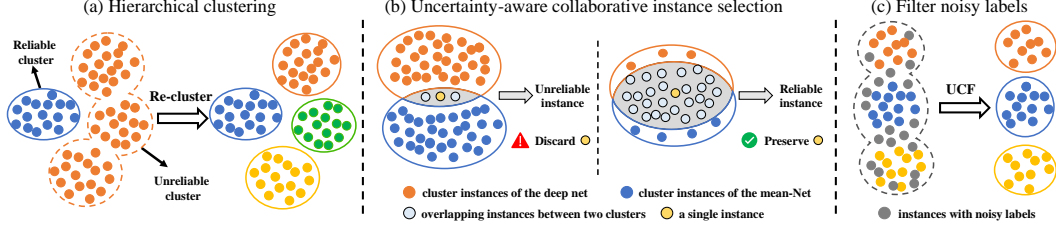


Figure 2. (a) Illustration of the proposed hierarchical clustering (HC) method. (b) Illustration of the proposed uncertainty-aware collaborative instance selection (UCIS) method. (c) Illustration of the overall clustering results by UCF. (Best viewed in color.)

the close-set problems with pre-defined classes, which cannot be generalized to our open-set object Re-ID task with completely unknown classes on the target domain.

3 Methodology

In this section, we present the details of our uncertainty-aware clustering framework (UCF), which reduces the effects of the noisy pseudo-labels in clustering-based Unsupervised Domain Adaptation (UDA). Our key idea is to select samples with reliable pseudo-labels in the target domain for model training purposes. To this end, we propose hierarchical clustering and uncertainty-aware collaborative instance selection methods to reduce the adverse effects of noisy pseudo-labels, and therefore improves the ability of model to learn cross-domain discriminative representations. In addition, we propose a strong baseline with a new contrastive loss using compact class-level memory banks.

Formally, we denote the source domain data as $\mathbb{D}_s = \{(\mathbf{x}_i^s, y_i^s) |_{i=1}^{N_s}\}$, where \mathbf{x}_i^s and y_i^s denote the i -th training instance and its annotation, respectively. The target-domain data without ground-truth labels are denoted as $\mathbb{D}_t = \{\mathbf{x}_i |_{i=1}^{N_t}\}$. N_s and N_t denote the sample size in the source and target domains, respectively.

3.1 Supervised Pre-training for Source Domain

In the first stage of UCF, we train the Re-ID model $F(\cdot | \theta)$ with the labeled source dataset \mathbb{D}_s using the cross-entropy loss and the triplet loss [20]; here, θ denotes parameters of the deep network. The pre-trained Re-ID model has the basic discriminability for domain adaptation. We then adopt this pre-trained network $F(\cdot | \theta)$ to extract the features of the target domain images. Following the existing clustering-based UDA methods [65, 13, 23], we use DBSCAN [7] and Jaccard distance to cluster the extracted features into K clusters before each epoch. We consider each cluster as a class and assign the same pseudo label for the instances belonging to the same cluster.

3.2 Uncertainty-aware Clustering Framework

Hierarchical clustering As explained in Section 1, images of visually similar identities may be grouped into the same cluster, which introduces significant noise to the pseudo-labels. Recent method [13] simply regards all instances in the unreliable clusters as outliers. However, this strategy may result in a large number of informative instances being lost. In the following, we handle this problem using a hierarchical clustering (HC) method that conducts fine-grained clustering in these clusters.

Intuitively, a reliable cluster should be compact and independent from other clusters. This means that the distances between instances in the same cluster should be small, and the distance between different clusters should be large. To measure the reliability of one cluster, we first calculate the silhouette coefficient [42] for each of its instances. Specifically, the silhouette coefficient for the i -th instance in the k -th cluster is formulated as follows:

$$\mathcal{S}(\mathbf{f}_k^i) = \frac{b(\mathbf{f}_k^i) - a(\mathbf{f}_k^i)}{\max(a(\mathbf{f}_k^i), b(\mathbf{f}_k^i))} \in [-1, 1], \quad (1)$$

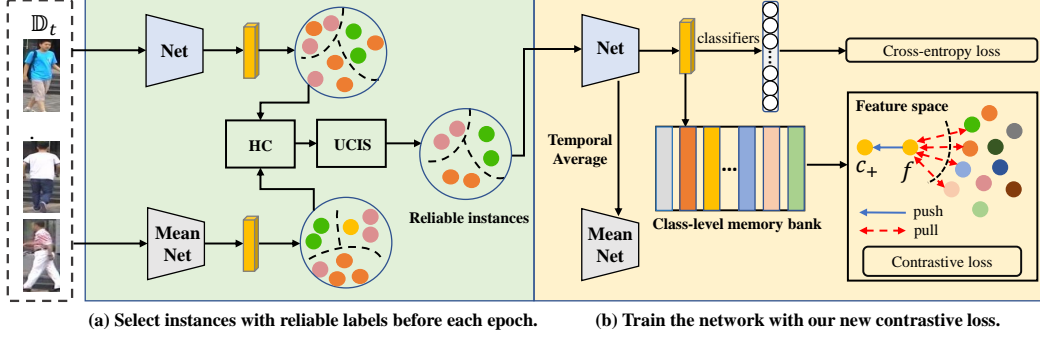


Figure 3. Model architecture of UCF in the training stage. UCF adopts a deep network and its temporally averaged model, *i.e.*, the mean-Net, to cluster images in the target domain. After that, the proposed novel hierarchical clustering scheme and the uncertainty-aware collaborative instance selection method are used to select images with reliable labels for model training. This effectively reduces the impact of noisy labels. Step (a) and step (b) are performed alternately. Note that the parameters of mean-Net model are not updated during back propagation. In the testing stage, the mean-Net model is adopted for inference. (Best viewed in color.)

where \mathbf{f}_k^i denotes the feature of the instance. $a(\mathbf{f}_k^i)$ represents the average distance between the i -th instance and all the other instances in the k -th cluster. Moreover, $b(\mathbf{f}_k^i)$ represents the average distance between the instance and all instances in the nearest cluster, which can be calculated as follows:

$$a(\mathbf{f}_k^i) = \frac{1}{|\mathcal{I}_k| - 1} \sum d_J(\mathbf{f}_k^i, \mathbf{f}_k^j), \mathbf{f}_k^j \in \mathcal{I}_k, i \neq j, \quad (2)$$

$$b(\mathbf{f}_k^i) = \min_{l \neq k} \left\{ \frac{1}{|\mathcal{I}_l|} \sum d_J(\mathbf{f}_k^i, \mathbf{f}_l^j) \right\}, \mathbf{f}_l^j \in \mathcal{I}_l, \quad (3)$$

where $d_J(\cdot, \cdot)$ represents the Jaccard distance, \mathcal{I}_k (\mathcal{I}_l) denotes the set of samples belonging to the k (l)-th cluster, and $|\cdot|$ denotes the number of features in a cluster. Since the Jaccard distance between each pair of samples has been calculated during DBSCAN clustering, this step hardly increases time consumption. Finally, we calculate the average silhouette coefficient for the k -th cluster:

$$\mathcal{S}(\mathcal{I}_k) = \frac{1}{|\mathcal{I}_k|} \sum \mathcal{S}(\mathbf{f}_k^i), \mathbf{f}_k^i \in \mathcal{I}_k. \quad (4)$$

When $\mathcal{S}(\mathcal{I}_k) < 0$, the intra-class distance surpass the inter-class distance. This usually indicates unreliable clustering from an object Re-ID perspective. We adopt a threshold of α to select these unreliable clusters. As shown in Fig. 2(a), we do not change the reliable ($\mathcal{S}(\mathcal{I}_k) > \alpha$) clusters, but we decompose an unreliable cluster into several smaller ones. In more detail, we use DBSCAN with the maximum neighbor distance d for coarse clustering and then measure the reliability of each cluster. Within each unreliable cluster, we use DBSCAN with the maximum neighbor distance of $2/3d$ for fine-grained clustering. Since the number of samples in each unreliable cluster is limited, this step only adds a small amount of time consumption.

Uncertainty-aware collaborative instance selection Although hierarchical clustering improves the quality of clustering, there are still inevitably noisy pseudo-labels in many clusters. In order to identify individual instances with noisy pseudo-labels, we propose an uncertainty-aware collaborative instance selection (UCIS) method, which adopts a deep network and its temporally averaged (mean-Net) [47] model to cluster the samples in the target domain separately. The parameters of the two models at iteration T are denoted as θ and $E^{(T)}[\theta]$, respectively. $E^{(T)}[\theta]$ is obtained as follows:

$$E^{(T)}[\theta] = \sigma E^{(T-1)}[\theta] + (1 - \sigma)\theta, \quad (5)$$

where σ is a temporal ensemble momentum coefficient whose value is within the range $[0, 1)$.

After hierarchical clustering, we obtain the fine-grained clustering results of the two models. We regard the clustering result of one instance as reliable if it is located in two similar clusters across the two models. The similarity of the two clusters is evaluated according to their overlap. More

specifically, we propose the following metric to measure the clustering uncertainty of one instance \mathbf{x}_i :

$$\mathcal{U}(\mathbf{x}_i) = \frac{|\mathcal{I}_k \cap \mathcal{I}_{\text{mean}l}|}{|\mathcal{I}_k|} \in [0, 1], \quad (6)$$

where \mathcal{I}_k and $\mathcal{I}_{\text{mean}l}$ denote the clusters containing \mathbf{x}_i by the two models, respectively. A larger $\mathcal{U}(\mathbf{x}_i)$ indicates larger overlap between \mathcal{I}_k and $\mathcal{I}_{\text{mean}l}$.

The value of $\mathcal{U}(\mathbf{x}_i)$ is therefore able to reflect the reliability of the pseudo-label for \mathbf{x}_i . We set $\beta \in [0, 1]$ as a threshold to select instances with reliable pseudo labels. As shown in Fig. 2(b), in each epoch, we only preserve instances for model training where $\mathcal{U}(\mathbf{x}_i)$ is larger than β .

Here we adopt mean-Net to select instances with reliable pseudo labels in offline clustering. Some methods train two networks together for close-set UDA problems [17], where the two networks select reliable samples for each other. This strategy may not work well in our framework. This is because UCF selects samples deemed reliable by both networks based on uncertainty, which requires the two networks to have different discriminative power. However, as empirically proved in [11], the two networks will obtain similar discriminative power if they are trained with exactly the same supervision signals. Therefore, mean-Net is a better choice in our framework.

3.3 A Strong Baseline for Clustering-based UDA

Fig. 3 illustrates the structure of our method. Aside from the commonly used cross-entropy loss, we propose to use the following contrastive loss. Given the feature \mathbf{f} of one target domain instance, our proposed contrastive loss is formulated as follows:

$$\mathcal{L}_c^t(\theta) = -\log \frac{\exp(\langle \mathbf{f}, \mathbf{c}^+ \rangle / \tau)}{\sum_{k=1}^K \exp(\langle \mathbf{f}, \mathbf{c}_k \rangle / \tau)}, \quad (7)$$

where \mathbf{c}^+ stands for the positive class prototype corresponding to \mathbf{f} , τ is a temperature factor, and $\langle \cdot, \cdot \rangle$ denotes the inner product between two feature vectors. The loss value is low when \mathbf{f} is similar to \mathbf{c}^+ and dissimilar to all the other cluster prototypes.

Memory initialization Each cluster is regarded as one class. The class-level memory bank $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ is initialized with the mean feature of each cluster. Formally,

$$\mathbf{c}_k = \frac{1}{|\mathcal{I}_k|} \sum_{\mathbf{f}_k^i \in \mathcal{I}_k} \mathbf{f}_k^i. \quad (8)$$

Memory update During training, \mathbf{c}_k is continuously updated within each epoch, according to all instances in the k -th cluster:

$$\mathbf{c}_k \leftarrow m^t \mathbf{c}_k + (1 - m^t) \mathbf{f}_k^i, \mathbf{f}_k^i \in \mathcal{I}_k, \quad (9)$$

where $m^t \in [0, 1]$ is the momentum coefficient for updating the target-domain class prototypes.

Discussion Compared with existing methods [55, 18, 1, 37], our proposed contrastive loss has two advantages. First, we only need to store class prototypes in the memory rather than the features of all samples, meaning that our approach has less memory cost. Second, each feature in the memory bank can be updated frequently within one epoch, which enables accurate loss computation in Eq. 7. In comparison, each feature in an instance-level memory bank can be updated only once per epoch, which may bring error in contrastive loss computation.

4 Experiments

4.1 Datasets and Evaluation Protocol

Following [13], we conduct extensive experiments on multiple large-scale Re-ID benchmarks, including two real-world person datasets and one synthetic person dataset, as well as two real-world vehicle datasets and one synthetic vehicle dataset. We evaluate our proposed method on both the mainstream real \rightarrow real adaptation tasks and the more challenging synthetic \rightarrow real adaptation tasks in person and vehicle Re-ID problems. The details of these datasets are summarized in Table 1.

Table 1. Statistics of the datasets used for training and evaluation

Dataset	# type	# train IDs	# train images	# test IDs	# query images	# cameras	# total images
Market-1501 [71]	real	751	12,936	750	3,368	6	32,217
DukeMTMC-ReID [41]	real	702	16,522	702	2,228	8	36,411
MSMT17 [52]	real	1,041	32,621	3,060	11,659	15	126,441
PersonX [44]	synthetic	410	9,840	856	5,136	6	45,792
VeRi-776 [27]	real	575	37,746	200	1,678	20	51,003
VehicleID [26]	real	13,164	113,346	800	5,693	-	221,763
VehicleX [36]	synthetic	1,362	192,150	-	-	11	192,150

Person Re-ID datasets Market-1501 [71], DukeMTMC-ReID [41], and MSMT17 [52] are real-world person image datasets that are widely used in domain adaptive tasks. MSMT17 includes more images that were captured in more challenging scenarios. The synthetic PersonX database [44] was constructed based on the Unity tool [40] with manually designed challenges, including random occlusion, resolution and illumination changes.

Vehicle Re-ID datasets To verify the generalization ability of our method on different kinds of objects, we conduct experiments with the real-world VeRi-776 [27], VehicleID [26], and the synthetic VehicleX datasets. VehicleX [36] is also generated by the Unity engine [59, 46] and further translated to the real-world style by SPGAN [5].

Evaluation protocol In our experiments, only ground-truth IDs of the source-domain datasets are provided for training. Experiments are conducted in line with the official evaluation protocol for each database. We adopt the widely used top-1/5/10 and mean Average Precision (mAP) as evaluation metrics. Moreover, following [11, 63, 69], the mean-Net is adopted for inference for both the baseline and our UCF method.

4.2 Implementation Details

We implement our framework in PyTorch [38]. We adopt ResNet-50 [19] as the backbone of the feature extractor and initialize the model with the parameters pre-trained on ImageNet [4]. After Layer4 of the ResNet-50 model, we add one Generalized-Mean (GeM) pooling [39] layer, one 1-D batch normalization [21] layer, and one L2-normalization layer. The L2-normalization layer produces 2048-dimensional feature vectors. Following [29], we perform data augmentation via random erasing, cropping, and flipping. For both source-domain pre-training and target-domain fine-tuning, we consistently construct a mini-batch with 64 person images of 16 identities. The person and vehicle images are resized to 256×128 and 224×224 pixels, respectively. To achieve faster convergence, we adopt embeddings of cluster centroids to initialize the weights of the classifiers. The momentum coefficients in Eq. 9 and Eq. 5 are set to 0.2 and 0.999, respectively. For DBSCAN, following [65, 13, 23], the hyper-parameter d is set to 0.6 and the minimal number of neighbors in a core point is set to 4. Following [13, 73], the temperature τ in Eq. 7 is set as 0.05. The threshold α in hierarchical clustering is set to 0.0. The uncertainty threshold β is set to 0.8. The ADAM method is adopted for optimization. The initial learning rate is set to 0.00035 and is decreased by multiplying by 0.1 on the 50-th epoch. The training lasts until the 80-th epoch.

4.3 Comparison with State-of-the-Art Methods

We compare the performance of UCF with state-of-the-art methods on multiple domain adaptation tasks, including real→real and more challenging synthetic→real tasks. The performance of these methods is tabulated in Table 2, Table 3, and Table 4, respectively. “Oracle” stands for the Re-ID performance in the fully supervised setting. It is clear that UCF significantly outperforms all state-of-the-art methods on both person and vehicle datasets with a plain ResNet-50 backbone.

Results on real→real UDA person Re-ID tasks We compare the performance of UCF with state-of-the-art methods on six UDA settings in Table 2. It is clear that UCF consistently outperforms existing approaches by large margins on all these benchmarks. In particular, UCF outperforms MMT [11] by 12.4%, 6.4%, 11.9%, 11.4%, 9.9%, and 8.2% in terms of mAP on these six tasks. It is worth noting that both UCF and MMT adopt mean-Net during the training stage. Moreover, UCF surpasses SpCL [13] by as much as 8.0% and 12.4% in terms of mAP and top-1 accuracy

Table 2. Comparison with state-of-the-art UDA re-ID methods on real \rightarrow real tasks

Methods	Reference	DukeMTMC-ReID \rightarrow Market-1501				Market-1501 \rightarrow DukeMTMC-ReID			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
PUL [8]	TOMM 2018	20.5	45.5	60.7	66.7	16.4	30.0	43.4	48.5
TJ-AIDL [50]	CVPR 2018	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
SPGAN+LMP [5]	CVPR 2018	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
HHL [72]	ECCV 2018	31.4	62.2	78.8	84.0	27.2	46.9	61.0	66.7
ECN [73]	CVPR 2019	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
PDA-Net [25]	ICCV 2019	47.6	75.2	86.3	90.2	45.1	63.2	77.0	82.5
PCB-PAST [65]	ICCV 2019	54.6	78.4	-	-	54.3	72.4	-	-
SSG [10]	ICCV 2019	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
MMCL [49]	CVPR 2020	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
ECN-GPP [74]	TPAMI 2020	63.8	84.1	92.8	95.4	54.4	74.0	83.7	87.4
JVTC+ [23]	ECCV 2020	67.2	86.8	95.2	97.1	66.5	80.4	89.9	92.2
AD-Cluster [62]	CVPR 2020	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
MMT [11]	ICLR 2020	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
CAIL [28]	ECCV 2020	71.5	88.1	94.4	96.2	65.2	79.5	88.3	91.4
NRMT [68]	ECCV 2020	71.7	87.8	94.6	96.5	62.2	77.8	86.9	89.5
MEB-Net [63]	ECCV 2020	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
SpCL [12]	NeurIPS 2020	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5
Dual-Refinement [3]	TIP 2021	78.0	90.9	96.4	97.7	67.7	82.1	90.1	92.5
UNRN [69]	AAAI 2021	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5
GLT [70]	CVPR 2021	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8
Ours		83.6	93.7	97.7	98.5	71.5	83.7	91.4	93.5
Oracle		82.7	94.1	97.9	98.8	71.3	84.5	92.2	94.2

Methods	Reference	Market-1501 \rightarrow MSMT17				DukeMTMC-ReID \rightarrow MSMT17			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
ECN [73]	CVPR 2019	8.5	25.3	36.3	42.1	10.2	30.2	41.5	46.8
SSG [10]	ICCV 2019	13.2	31.6	-	49.6	13.3	32.2	-	51.2
ECN-GPP [74]	TPAMI 2020	15.2	40.4	53.1	58.7	16.0	42.5	55.9	61.5
MMCL [49]	CVPR 2020	15.1	40.8	51.8	56.7	16.2	43.6	54.3	58.9
NRMT [68]	ECCV 2020	19.8	43.7	56.5	62.2	20.6	45.2	57.8	63.3
CAIL [28]	ECCV 2020	20.4	43.7	56.1	61.9	24.3	51.7	64.0	68.9
MMT [11]	ICLR 2020	22.9	49.2	63.1	68.8	23.3	50.1	63.9	69.8
JVTC+ [23]	ECCV 2020	25.1	48.6	65.3	68.2	27.5	52.9	70.5	75.9
SpCL [12]	NeurIPS 2020	26.8	53.7	65.0	69.8	26.5	53.1	65.8	70.5
Dual-Refinement [3]	TIP 2021	25.1	53.3	66.1	71.5	26.9	55.0	68.4	73.2
UNRN [69]	AAAI 2021	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
GLT [70]	CVPR 2021	26.5	56.6	67.5	72.0	27.7	59.5	70.1	74.2
Ours		34.8	66.1	76.6	80.6	34.7	66.5	77.0	80.9
Oracle		45.1	74.5	84.8	88.0	45.1	74.5	84.8	88.0

Methods	Reference	MSMT17 \rightarrow Market-1501				MSMT17 \rightarrow DukeMTMC-reID			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
CASCL [54]	ICCV 2019	35.5	65.4	80.6	86.2	37.8	59.3	73.2	77.8
MAR [60]	CVPR 2019	40.0	67.7	81.9	87.3	48.0	67.1	79.8	84.2
PAUL [58]	CVPR 2019	40.1	68.5	82.4	87.4	53.2	72.0	82.7	86.0
DG-Net++ [75]	ECCV 2020	64.6	83.1	91.5	94.3	58.2	75.2	73.6	86.9
D-MMD [35]	ECCV 2020	50.8	72.8	88.1	92.3	51.6	68.8	82.6	87.1
MMT [11]	ICLR 2020	75.6	89.3	95.8	97.5	63.3	77.4	88.4	91.7
SpCL [12]	NeurIPS 2020	77.5	89.7	96.1	97.6	69.3	82.9	91.0	93.0
Ours		85.5	94.6	97.9	98.8	71.5	84.1	91.6	93.6
Oracle		82.7	94.1	97.9	98.8	71.3	84.5	92.2	94.2

Table 3. Comparison with state-of-the-art UDA re-ID methods on synthetic \rightarrow real tasks

Methods	Reference	PersonX \rightarrow MSMT17				PersonX \rightarrow Market1501				PersonX \rightarrow DukeMTMC-reID			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
MMT [11]	ICLR 2020	17.7	39.1	52.6	58.5	71.0	86.5	94.8	97.0	60.1	74.3	86.5	90.5
SpCL [13]	NeurIPS 2020	22.7	47.7	60.0	65.5	73.8	88.0	95.3	96.9	67.2	81.8	90.2	92.6
Ours		28.3	58.2	69.7	74.3	80.5	92.1	97.1	98.2	70.7	84.8	91.7	94.1
Oracle		45.1	74.5	84.8	88.0	82.7	94.1	97.9	98.8	70.9	84.5	92.2	94.2

Table 4. Performance comparisons with state-of-the-art UDA vehicle Re-ID methods

Methods	Reference	VehicleID \rightarrow VeRi-776				VehicleX \rightarrow VeRi-776			
		mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
MMT [11]	ICLR 2020	35.3	74.6	82.6	87.0	35.6	76.0	83.1	87.4
SpCL [13]	NeurIPS 2020	38.9	80.4	86.8	89.6	38.9	81.3	87.3	90.0
Ours		40.5	85.2	88.7	90.9	40.6	84.4	88.4	91.5
Oracle		71.9	93.6	96.9	98.3	71.9	93.6	96.9	98.3

Table 5. Ablation studies on each key component of UCF

Methods	Market1501→MSMT17				PersonX→MSMT17			
	mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
Source Pretrain	8.4	22.6	32.9	38.1	2.7	8.8	14.8	18.3
Cross-entropy loss	28.5	57.5	69.6	74.3	23.5	51.4	64.3	69.5
Contrastive loss	26.8	55.0	66.5	71.5	22.4	48.8	60.3	64.9
Strong baseline	31.6	61.7	72.3	76.4	26.2	55.0	67.0	71.6
Baseline w/ HC	33.8	63.4	74.5	78.9	27.6	57.1	68.8	73.3
Baseline w/ UCIS	33.3	64.6	75.0	79.0	27.6	57.4	68.6	73.3
Ours(full)	34.8	66.1	76.6	80.6	28.3	58.2	69.7	74.3

Table 6. Performance comparison between our class-level contrastive loss and instance-level contrastive loss

Methods	Market1501→MSMT17				PersonX→MSMT17			
	mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
Instance-level contrastive loss	24.3	52.5	64.1	69.1	19.1	45.3	56.2	61.4
Class-level contrastive loss (ours)	26.8	55.0	66.5	71.5	22.4	48.8	60.3	64.9

in the Market1501→MSMT17 task, respectively. Finally, UCF significantly outperforms one very recent method named GLT [70] by 8.3% and 7.0% in terms of the mAP accuracy, for Market1501→MSMT17 and DukeMTMC-ReID→MSMT17 tasks, respectively. The above experimental results clearly demonstrates the effectiveness of UCF.

UCF also significantly bridges the gap between the unsupervised and fully-supervised settings. For example, UCF achieves 94.6% top-1 accuracy and 85.5% mAP on the MSMT17→Market1501 task, meaning that it surpasses the the performance of “Oracle” on the Market-1501 database by 0.5% in top-1 accuracy and 2.8% in mAP, respectively. In addition to the reliable pseudo labels generated by UCF, another possible reason is that MSMT17 is larger than Market1501; therefore, supervised pre-training on MSMT17 provides better model initialization before domain adaptation.

Results on synthetic→real UDA person Re-ID tasks Compared with the real→real UDA re-ID tasks, the synthetic→real UDA tasks are usually more challenging due to the dramatic domain gap. As shown in Table 3, UCF outperforms state-of-the-art methods by large margins. For example, UCF beats the SpCL [12] method by 4.1% in terms of top-1 accuracy and 6.7% in terms of mAP on the PersonX→Market-1501 task. It is also worth noting that the performance of UCF in synthetic→real tasks still exceeds that of SpCL in real→real tasks. Specifically, UCF achieves 92.1% top-1 accuracy and 80.5% mAP on the PersonX→Market1501 task, which outperform SpCL [12] on the MSMT17→Market-1501 task by 2.4% in terms of top-1 accuracy and 3.0% in terms of mAP.

Although these results are promising, there is still a clear gap between UCF and “Oracle” on large-scale datasets such as MSMT17. This motivates us to develop more robust clustering and pseudo label generation methods in the future.

Results on vehicle Re-ID datasets As Table 4 shows, the performance of UCF surpasses that of SpCL by 4.8% in top-1 accuracy and 1.6% in mAP on the VehicleID→VeRi-776 task. Moreover, UCF outperforms SpCL by 3.1% in top-1 accuracy and 1.7% in mAP on the VehicleX→VeRi-776 task. These experimental results further demonstrate the effectiveness of UCF for object Re-ID.

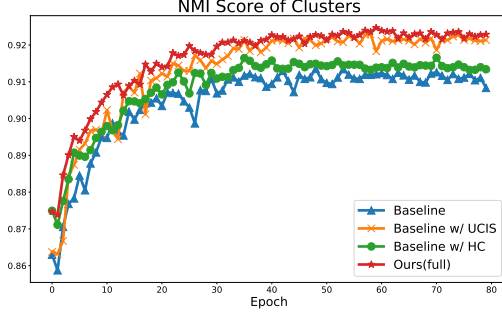
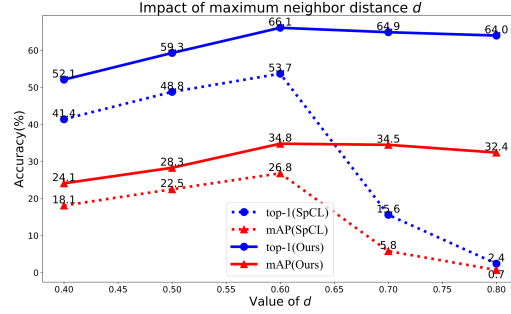
4.4 Ablation Studies

We systematically investigate the effectiveness of each key component of UCF: namely, the strong baseline, hierarchical clustering (HC), and uncertainty-aware collaborative instance selection (UCIS), respectively. Experiments are conducted on real→real and more challenging synthetic→real adaptation tasks, specifically Market1501→MSMT17 and PersonX→MSMT17. The results are summarized in Table 5. “Source Pretrain” represents the Re-ID model trained in the source domain and tested directly in the target domain.

Effectiveness of the strong baseline We build our baseline with the cross-entropy loss and our new contrastive loss, both of which are described in Section 3.3. We first evaluate the performance when only classification loss or contrastive loss is used. As shown in Table 5, the two settings achieve 28.5% and 26.8% mAP respectively for the Market1501→MSMT17 task. In addition, as shown in Table 6, our new contrastive loss outperforms the conventional instance-level contrastive loss by 2.5%

Table 7. Performance comparison between UCF and SpCL [13] with our strong baseline

Methods	Market1501→MSMT17				PersonX→MSMT17			
	mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
Strong baseline	31.6	61.7	72.3	76.4	26.2	55.0	67.0	71.6
Strong baseline+SpCL [12]	33.3	63.5	74.0	78.6	27.3	56.8	68.0	73.4
Ours(full)	34.8	66.1	76.6	80.6	28.3	58.2	69.7	74.3

**Figure 4.** Comparisons on the Normalized Mutual Information (NMI) scores of clusters during the training process on the Market1501→MSMT17 task.**Figure 5.** Performance comparison between UCF and SpCL [13] with different values of hyper-parameter d .

and 3.3% mAP on the two UDA tasks, respectively. When the two loss functions are used together, we obtain a strong baseline. For example, compared with “Source Pretrain” in Table 5, our baseline promotes the top-1 accuracy by 39.1% and 46.2%, as well as mAP by 23.2% and 23.5%, on the two UDA tasks, respectively. These results prove that our baseline is simple but effective.

Effectiveness of the hierarchical clustering Compared with our baseline, the hierarchical clustering method consistently yields performance gains. For example, “Baseline w/ HC” outperforms the baseline in terms of top-1 accuracy by 1.7% and 2.1%, as well as mAP by 2.2% and 1.4%, on Market1501→MSMT17 and PersonX→MSMT17 tasks, respectively. This is because the hierarchical clustering improves the quality of pseudo-labels, meaning that the deep model can learn more discriminative features.

Effectiveness of the uncertainty-aware collaborative instance selection When the baseline is equipped with the UCIS module, the performance of both UDA tasks is promoted. In particular, UCIS improves the top-1 accuracy of the baseline by 2.9% and 2.4%, as well as mAP by 1.7% and 1.4%, on the two tasks, respectively. The above results demonstrate the necessity of reducing the impact of noisy labels, as well as the effectiveness of our method.

Effectiveness of the UCF framework Finally, with both the HC and UCIS modules, our full model achieves better performance than using either of the modules alone. The above comparisons justify the effectiveness of each key component in our framework.

Furthermore, we test the performance of SpCL [12] based on our strong baseline. We equip SpCL with a hybrid memory to save target-domain cluster centroids and target-domain un-clustered instance features. Experimental results are summarized in Table 7. It is shown that UCF still outperforms SpCL by 2.6% and 1.5% in terms of top-1 accuracy and mAP on Market1501→MSMT17 task, respectively. The above experimental results justify the effectiveness of UCF.

Analysis of the quality of pseudo labels In Fig. 4, we illustrate the improvement in the quality of pseudo labels. Following SpCL [13], we illustrate the Normalized Mutual Information (NMI) [34] scores of clusters during training on the Market1501→MSMT17 task. NMI [34] is an index that measures the accuracy of the clustering results. It can accordingly be observed that, compared with the baseline, the quality of the pseudo-labels is significantly improved when the proposed techniques are applied.

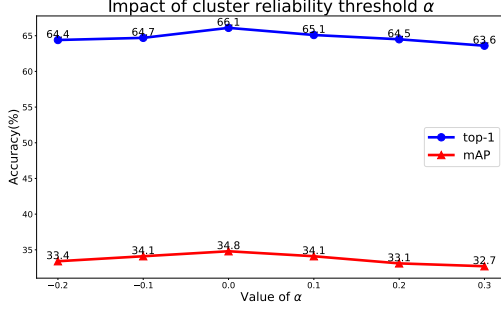


Figure 6. Performance of UCF with different values of α .

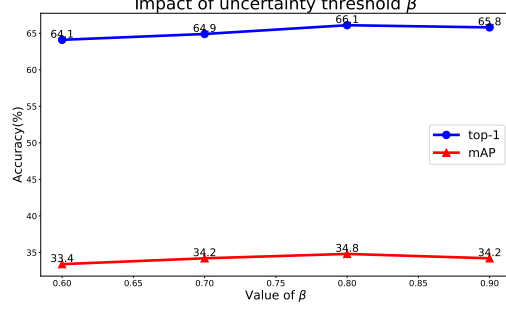


Figure 7. Performance of UCF with different values of β .

4.5 Parameter Analysis

We tune the hyper-parameters on the Market-1501→MSMT17 task, then directly apply the chosen hyper-parameters to all the other tasks.

Maximum neighbor distance d for DBSCAN. DBSCAN is one of the most popular clustering algorithms in the UDA Re-ID literature. For DBSCAN, the maximum neighborhood distance d is an important hyperparameter. As demonstrated in Fig. 5, we find that the value of d may considerably affect the performance of state-of-the-art methods. In particular, a larger value of d may result in a dramatic performance drop; this is because the pseudo-labels will contain more noise as the value of d increases. In comparison, the performance of UCF is significantly more robust. This is because UCF successfully improves the clustering quality and removes samples with unreliable pseudo-labels.

Cluster reliability threshold α for hierarchical clustering α is a threshold on $\mathcal{S}(\mathcal{I}_k)$. According to the definition of $\mathcal{S}(\mathcal{I}_k)$, a negative value of $\mathcal{S}(\mathcal{I}_k)$ means that intra-class distance surpasses inter-class distance. This usually indicates unreliable clustering from an object Re-ID perspective. As demonstrated in Fig. 6, our framework achieves the optimal performance when α is set to 0.0 on the MSMT17→Market-1501 task, which is consistent with our above analysis. When α is larger than 0.0, the top-1 accuracy and mAP gradually decrease. This is because some reliable clusters will be forced to be decomposed, resulting in more noisy pseudo-labels and therefore performance degradation.

Uncertainty threshold β for collaborative instance selection As described in Section 3, we require an uncertainty threshold β to select samples with reliable pseudo-labels. In Fig. 7, we investigate the effect of different values of β . As can be seen from Fig. 7, the performance of UCF is generally robust to the value of β while the best performance is achieved when β is set to 0.8. The performance of UCF reduces when β is set to a smaller value, such as 0.6; this may be because samples with noisy pseudo-labels cannot be identified when the threshold is low.

4.6 Qualitative Comparisons

In Fig. 8, we utilize t-SNE [32] to visualize the clustering results by “Baseline”, “Baseline w/ HC”, “Baseline w/ UCIS”, and the “UCF” model for the Market-1501→MSMT17 task, respectively. We have the following observations.

First, as illustrated in Fig. 8(a), due to the limited discriminative power of the Re-ID model in the target domain, many visually similar images may be grouped into the same cluster. The size of such clusters is often large. Second, as illustrated in Fig. 8(b), when the proposed hierarchical clustering (HC) method is utilized, the unreliable clusters in Fig. 8(a) are decomposed into multiple smaller ones. Third, as shown in Fig. 8(c), the uncertainty-aware collaborative instance selection (UCIS) method identifies instances with unreliable pseudo labels, which are represented using the gray color in the figure. Finally, combining UCIS and HC can achieve the best clustering results, which proves that the two modules are complementary. The above visualization results are consistent with the results in the experimentation section.

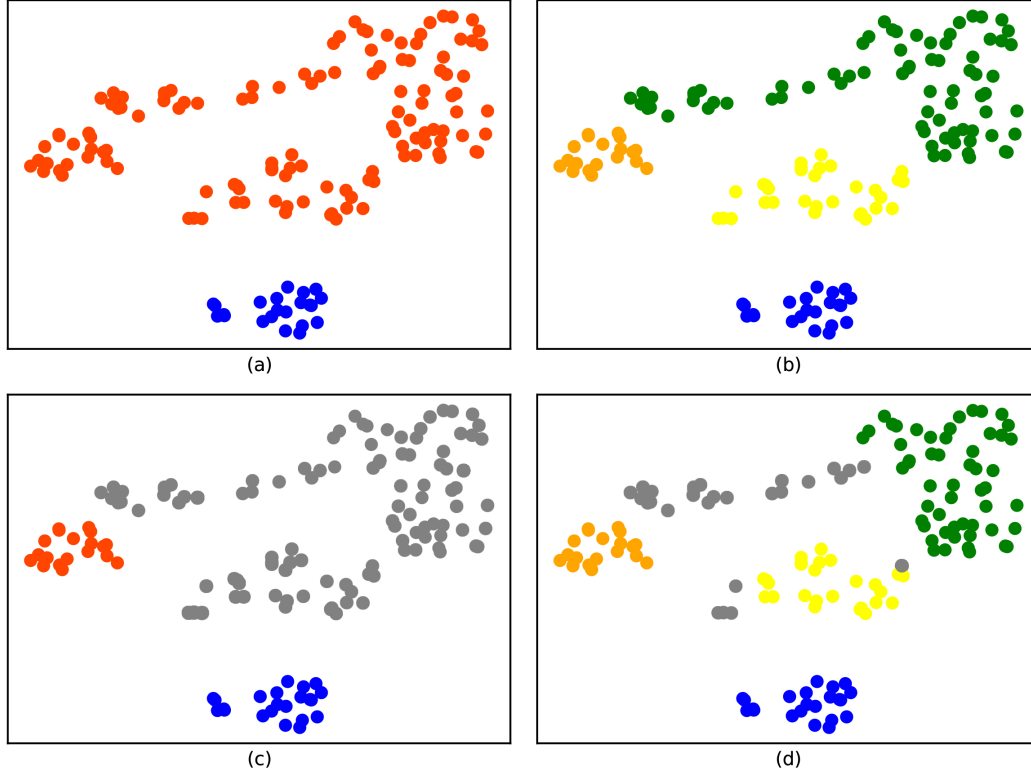


Figure 8. Visualization using t-SNE [32] for the clustering results on the target domain by four models: (a) “Baseline”, (b) “Baseline $w/$ HC”, (c) “Baseline $w/$ UCIS”, and (d) the “UCF model”. Different clusters are represented using different colors. Gray denotes outliers. (Best viewed in color.)

5 Conclusion

In this work, we propose an uncertainty-aware clustering framework (UCF) to tackle the problem of noisy pseudo labels in clustering-based UDA object Re-ID tasks. UCF handles the label noise problem on two levels. First, a novel hierarchical clustering scheme is proposed to promote the clustering quality; second, an uncertainty-aware collaborative instance selection method is introduced to select images with reliable labels for model training. These two techniques significantly relieve the noise in pseudo-labels and consequently improve the quality of deep feature learning. Our UCF method significantly outperforms state-of-the-art object Re-ID methods on many domain adaptation tasks.

References

- [1] Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proc. Int. Conf. Mach. Learn. pp. 1597–1607 (2020)
- [2] Chen, Y., Zhu, X., Gong, S.: Instance-guided context rendering for cross-domain person re-identification. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 232–242 (2019)
- [3] Dai, Y., Liu, J., Bai, Y., Tong, Z., Duan, L.Y.: Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification. IEEE Trans. Image Process (2021)
- [4] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 248–255 (2009)
- [5] Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 994–1003 (2018)
- [6] Ding, C., Wang, K., Wang, P., Tao, D.: Multi-task learning with coarse priors for robust part-aware person re-identification. IEEE Trans. Pattern Anal. Mach. Intell (2020)

- [7] Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proc. Data. Min. Knowl. Discov.* p. 226–231 (1996)
- [8] Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: Clustering and fine-tuning. *ACM Trans. Multimed. Comput. Commun. Appl.* **14**(4), 1–18 (2018)
- [9] Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: *Proc. IEEE Int. Conf. Comput. Vis.* pp. 6112–6121 (2019)
- [10] Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Huang, T.S.: Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: *ICCV*. pp. 6112–6121 (2019)
- [11] Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In: *Proc. Int. Conf. Learn. Represent.* pp. 1–15 (2020)
- [12] Ge, Y., Wang, H., Zhu, F., Zhao, R., Li, H.: Self-supervising fine-grained region similarities for large-scale image localization. In: *Proc. Eur. Conf. Comput. Vis.* pp. 369–386 (2020)
- [13] Ge, Y., Zhu, F., Chen, D., Zhao, R., Li, H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: *Proc. Adv. Neural Inf. Process. Syst.* (2020)
- [14] Ge, Y., Zhu, F., Zhao, R., Li, H.: Structured domain adaptation with online relation regularization for unsupervised person re-id. *arXiv preprint arXiv:2003.06650* (2020)
- [15] Gong, X., Yao, Z., Li, X., Fan, Y., Luo, B., Fan, J., Lao, B.: Lag-net: Multi-granularity network for person re-identification via local attention system. *IEEE Trans. Multimedia* (2021)
- [16] Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., Azar, M.G., et al.: Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733* (2020)
- [17] Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. In: *Proc. Adv. Neural Inf. Process. Syst.* pp. 8527–8537 (2018)
- [18] He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 9729–9738 (2020)
- [19] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 770–778 (2016)
- [20] Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017)
- [21] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proc. Int. Conf. Mach. Learn.* pp. 448–456 (2015)
- [22] Jiang, B., Wang, X., Zheng, A., Tang, J., Luo, B.: Ph-gcn: Person retrieval with part-based hierarchical graph convolutional network. *IEEE Trans. Multimedia* (2021)
- [23] Li, J., Zhang, S.: Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In: *Proc. Eur. Conf. Comput. Vis.* pp. 1–14 (2020)
- [24] Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 2285–2294 (2018)
- [25] Li, Y.J., Lin, C.S., Lin, Y.B., Wang, Y.C.F.: Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. In: *Proc. IEEE Int. Conf. Comput. Vis.* pp. 7919–7929 (2019)
- [26] Liu, H., Tian, Y., Yang, Y., Pang, L., Huang, T.: Deep relative distance learning: Tell the difference between similar vehicles. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 2167–2175 (2016)
- [27] Liu, X., Liu, W., Mei, T., Ma, H.: A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: *Proc. Eur. Conf. Comput. Vis.* pp. 869–884 (2016)
- [28] Luo, C., Song, C., Zhang, Z.: Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. In: *Proc. Eur. Conf. Comput. Vis.* pp. 224–241 (2020)
- [29] Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 1–8 (2019)
- [30] Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., Gu, J.: A strong baseline and batch normalization neck for deep person re-identification. *IEEE Trans. Multimedia* **22**(10), 2597–2609 (2019)
- [31] Lyu, Y., Tsang, I.W.: Curriculum loss: Robust learning and generalization against label corruption. *arXiv preprint arXiv:1905.10045* (2019)
- [32] Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**(11) (2008)

- [33] Malach, E., Shalev-Shwartz, S.: Decoupling" when to update" from" how to update". In: Proc. Adv. Neural Inf. Process. Syst. pp. 960–970 (2017)
- [34] McDavid, A.F., Greene, D., Hurley, N.: Normalized mutual information to evaluate overlapping community finding algorithms. arXiv preprint arXiv:1110.2515 (2011)
- [35] Mekhazni, D., Bhuiyan, A., Ekladios, G., Granger, E.: Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: Proc. Eur. Conf. Comput. Vis. pp. 1–14 (2020)
- [36] Naphade, M., Wang, S., Anastasiu, D.C., Tang, Z., Chang, M.C., Yang, X., Zheng, L., Sharma, A., Chellappa, R., Chakraborty, P.: The 4th ai city challenge. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. workshop. pp. 626–627 (2020)
- [37] Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
- [38] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: Proc. Adv. Neural Inf. Process. Syst. pp. 8026–8037 (2019)
- [39] Radenović, F., Tolias, G., Chum, O.: Fine-tuning cnn image retrieval with no human annotation. IEEE Trans. Pattern Anal. Mach. Intell. **41**(7), 1655–1668 (2018)
- [40] Riccitiello, J.: John riccitiello sets out to identify the engine of growth for unity technologies (interview). VentureBeat. Interview with Dean Takahashi. Retrieved January **18**, 3 (2015)
- [41] Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Proc. Eur. Conf. Comput. Vis. pp. 17–35 (2016)
- [42] Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. pp. 53–65 (1987)
- [43] Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., Wang, X.: Unsupervised domain adaptive re-identification: Theory and practice. Pattern Recognit. **102**, 107173 (2020)
- [44] Sun, X., Zheng, L.: Dissecting person re-identification from the viewpoint of viewpoint. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 608–617 (2019)
- [45] Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proc. Eur. Conf. Comput. Vis. pp. 480–496 (2018)
- [46] Tang, Z., Naphade, M., Birchfield, S., Tremblay, J., Hodge, W., Kumar, R., Wang, S., Yang, X.: Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 211–220 (2019)
- [47] Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proc. Adv. Neural Inf. Process. Syst. pp. 1195–1204 (2017)
- [48] Wan, C., Wu, Y., Tian, X., Huang, J., Hua, X.S.: Concentrated local part discovery with fine-grained part representation for person re-identification. IEEE Trans. Multimedia **22**(6), 1605–1618 (2019)
- [49] Wang, D., Zhang, S.: Unsupervised person re-identification via multi-label classification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 10981–10990 (2020)
- [50] Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 2275–2284 (2018)
- [51] Wang, K., Wang, P., Ding, C., Tao, D.: Batch coherence-driven network for part-aware person re-identification. IEEE Trans. Image Process **30**, 3405–3418 (2021)
- [52] Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 79–88 (2018)
- [53] Wei, L., Zhang, S., Yao, H., Gao, W., Tian, Q.: Glad: Global–local-alignment descriptor for scalable person re-identification. IEEE Trans. Multimedia **21**(4), 986–999 (2018)
- [54] Wu, A., Zheng, W.S., Lai, J.H.: Unsupervised person re-identification by camera-aware similarity consistency learning. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 6922–6931 (2019)
- [55] Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 3733–3742 (2018)
- [56] Xu, J., Zhao, R., Zhu, F., Wang, H., Ouyang, W.: Attention-aware compositional network for person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 2119–2128 (2018)
- [57] Yan, C., Pang, G., Bai, X., Liu, C., Xin, N., Gu, L., Zhou, J.: Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss. IEEE Trans. Multimedia (2021)

- [58] Yang, Q., Yu, H.X., Wu, A., Zheng, W.S.: Patch-based discriminative feature learning for unsupervised person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 3633–3642 (2019)
- [59] Yao, Y., Zheng, L., Yang, X., Naphade, M., Gedeon, T.: Simulating content consistent vehicle datasets with attribute descent. arXiv preprint arXiv:1912.08855 (2019)
- [60] Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 2148–2157 (2019)
- [61] Yu, X., Han, B., Yao, J., Niu, G., Tsang, I., Sugiyama, M.: How does disagreement help generalization against label corruption? In: Proc. Int. Conf. Mach. Learn. pp. 7164–7173 (2019)
- [62] Zhai, Y., Lu, S., Ye, Q., Shan, X., Chen, J., Ji, R., Tian, Y.: Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 9021–9030 (2020)
- [63] Zhai, Y., Ye, Q., Lu, S., Jia, M., Ji, R., Tian, Y.: Multiple expert brainstorming for domain adaptive person re-identification. In: Proc. Eur. Conf. Comput. Vis. pp. 594–611 (2020)
- [64] Zhang, P., Xu, J., Wu, Q., Huang, Y., Ben, X.: Learning spatial-temporal representations over walking tracklet for long-term person re-identification in the wild. IEEE Trans. Multimedia (2020)
- [65] Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 8222–8231 (2019)
- [66] Zhang, Z., Lan, C., Zeng, W., Chen, Z.: Densely semantically aligned person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 667–676 (2019)
- [67] Zhao, C., Lv, X., Zhang, Z., Zuo, W., Wu, J., Miao, D.: Deep fusion feature representation learning with hard mining center-triplet loss for person re-identification. IEEE Trans. Multimedia **22**(12), 3180–3195 (2020)
- [68] Zhao, F., Liao, S., Xie, G.S., Zhao, J., Zhang, K., Shao, L.: Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In: Proc. Eur. Conf. Comput. Vis. pp. 1–14 (2020)
- [69] Zheng, K., Lan, C., Zeng, W., Zhang, Z., Zha, Z.J.: Exploiting sample uncertainty for domain adaptive person re-identification. In: Proc. AAAI Conf. Artif. Intell. pp. 3538–3546 (2021)
- [70] Zheng, K., Liu, W., He, L., Mei, T., Luo, J., Zha, Z.J.: Group-aware label transfer for domain adaptive person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 5310–5319 (2021)
- [71] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proc. IEEE Int. Conf. Comput. Vis. pp. 1116–1124 (2015)
- [72] Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero-and homogeneously. In: Proc. Eur. Conf. Comput. Vis. pp. 172–188 (2018)
- [73] Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: Exemplar memory for domain adaptive person re-identification. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 598–607 (2019)
- [74] Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Learning to adapt invariance in memory for person re-identification. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
- [75] Zou, Y., Yang, X., Yu, Z., Kumar, B., Kautz, J.: Joint disentangling and adaptation for cross-domain person re-identification. In: Proc. Eur. Conf. Comput. Vis. pp. 1–14 (2020)