Unsupervised Person Re-ID via Loose-Tight Alternate Clustering

Bo Li[1], Tianbao Liang[1], Jianming Lv[1(✉)], Shengjing Chen[2], and Hongjian Xie[2]

[1] South China University of Technology, Guangzhou, China
jmlv@scut.edu.cn
[2] Guangzhou Forsafe Digital Technology Co., Ltd., Guangzhou, China
{shengjing_chen,hongjian_xie}@for-safe.cn

Abstract. Recently developed clustering-guided unsupervised methods have shown their superior performance in the person re-identification (re-ID) problem, which aims to match the surveillance images containing the same person. However, the performance of these methods is usually very sensitive to the change of the hyper-parameters in the clustering methods, such as the maximum distance of the neighbors and the number of clusters, which determine the quality of the clustering results. Tuning these parameters may need a large-scale labeled validation set, which is usually not applicable in unlabeled domain and hard to be generalized to different datasets. To solve this problem, we propose a Loose-Tight Alternate Clustering method without using any sensitive clustering parameter for unsupervised optimization. Specifically, we address the challenge as a multi-domain clustering problem, and propose the Loose and Tight Bounds to alleviate two kinds of clustering errors. Based on these bounds, a novel Loose-Tight alternate clustering strategy is adopted to optimize the visual model iteratively. Furthermore, a quality measurement based learning method is proposed to mitigate the side-effects of the pseudo-label noise by assigning lower weight to those clusters with lower purity. Extensive experiments show that our method can not only outperform state-of-the-art methods without manual exploration of clustering parameters, but also achieve much higher robustness against the dynamic changing of the target domain.

Keywords: Clustering-guided · Person re-identication · Unsupervised optimization · Multi-domain clustering · Loose-Tight Alternate

1 Introduction

Person re-identification (re-ID) aims to match the surveillance images which contain the same person. The recently developed supervised algorithms [15,17,26] have achieved impressive performance based on convolutional neural networks. However, the extremely high cost of labeling the dataset limits the scalability of these methods. How to effectively learn a discriminative model on massive unlabeled data has become a hot research topic in this field.

G. Memmi et al. (Eds.): KSEM 2022, LNAI 13369, pp. 29–42, 2022.
https://doi.org/10.1007/978-3-031-10986-7_3

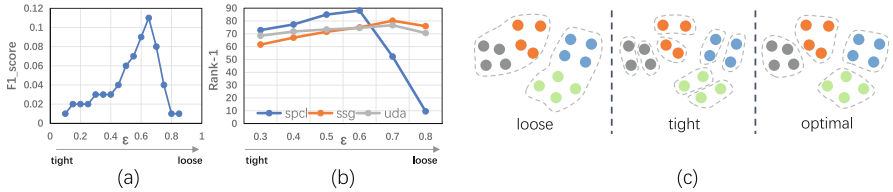

Fig. 1. The influence of the clustering parameter on clustering-guided unsupervised person re-ID tested on Market1501. (a): The F1 score of clustering results under different ϵ of DBSCAN. (b): The Rank-1 score of three methods under different ϵ of DBSCAN. The ϵ corresponds to τ in UDA [14] and d in SPCL [6]. (c): Three types of clustering criteria and the corresponding clustering results. The points in same color means the samples share with the same identity.

Recently, some unsupervised domain adaptive re-ID solutions (UDA) [4,12, 14,24,25,30,31] have been proposed by transferring the prior knowledge learned in labeled source datasets to another unlabeled target dataset. However, these UDA methods still require a large amount of manually annotated source data. In a more extreme configuration, some fully unsupervised methods [10,11,16,23] are proposed to learn a high-performance discriminative model on the unlabeled target dataset without using any labeled source data.

Most of state-of-the-art unsupervised methods [4,6,14,24,25] utilize clustering methods to obtain supervised signals from unlabeled instances. However, according to our observation as shown in Fig. 1, the performance of these methods is usually very sensitive to the change of the clustering parameters, which determine the quality of clustering results. For example, the maximum distance between neighbors, ϵ, is the most important parameter in DBSCAN [2], which affects the clustering results seriously. As shown in Fig. 1(a), the F1 score of the clustering result is very sensitive to the change of ϵ. Figure 1(c) shows the clustering results intuitively. Large ϵ corresponds to a loose clustering criterion which may form large groups of instances, while small ϵ may cause small and tight groups on the other hand. The key of these methods is to find the optimal clustering result closest to the ground truth. Figure 1(b) further shows the performance of the state-of-the-art unsupervised person re-ID methods SPCL [6], SSG [4] and UDA [14], which are all based on DBSCAN [2]. It clearly shows that their performance is very sensitive to ϵ. In particular, the changes of ϵ may lead to the collapse of the SPCL [6], which has much higher peak accuracy than the others. These methods [4,6,14] usually report the best performance using the optimal ϵ, which actually needs large labeled validation set for careful tuning and is difficult to be generalized to different datasets.

To address above problems, we propose a *Loose-Tight Alternate Clustering* (LTAC) framework to learn from noisy pseudo-labels and alleviate the sensitivity of clustering parameters. Distinct from traditional DBSCAN based clustering method, we do not configure the optimized ϵ at first, which is usually hard to tune. We go another way by modeling the challenge as a multi-domain clustering problem and define the loose and tight bounds of the clustering criteria to reduce one kind of clustering errors respectively. Then a novel Loose-Tight Alternate

Clustering strategy is proposed to run the loose and tight clustering alternately to optimize the visual model gradually. Moreover, we propose a *Quality Measurement based Learning* method to further reduce the side-effects of the clustering errors.

Main contributions of this paper are as follows:

- We propose the *Loose and Tight Bounds* of the clustering criteria in the multi-domain clustering problem to reduce two kinds of clustering errors.
- We propose a novel *Loose-Tight Alternate Clustering strategy* (LTAC) for unsupervised person re-ID by generating two types of pseudo-labels alternately based on the Loose and Tight Bounds to optimize the visual model gradually.
- A *Quality Measurement based Learning* method is proposed to reduce the side-effect of the pseudo-label noise by assigning smaller weight to those clusters with lower purity.
- Comprehensive experiments are conducted and show that our method can not only outperform state-of-the-art methods without manually configured sensitive clustering parameters, but also achieve much higher robustness against dynamic change of target domain.

The rest of this paper is organized as follows. In Sect. 2, we discuss some related work. In Sect. 3, we introduce the details of the Loose-Tight Alternate Clustering method for unsupervised person re-ID namely LTAC. After that, we provide the experimental evaluations of the proposed method in Sect. 4. Finally, we conclude our work in Sect. 5.

2 Related Work

2.1 Clustering-Guided Unsupervised Person re-ID

One of the most popular way to tackle unsupervised person re-ID is the clustering-guided framework, which utilizes pseudo-labels based on clustering results. PUL [3] selects samples close to the cluster centroid for training gradually. BUC [10] proposes a bottom-up clustering approach to gradually merge samples into one identity. HCT [23] improves the distance measurement of BUC by using an unweighted pair-group method with arithmetic means. However, the changes of merging steps significantly impact the final performance of BUC and HCT. MMT [5] softly refines the pseudo-labels via mutual mean-teaching, which needs auxiliary models and is sensitive to the k value of K-means. Some methods have verified the effectiveness of DBSCAN [2] in clustering. UDA [14] proposes a vanilla self-training framework with DBSCAN. SSG [4] generates multiple clusters from global body to local parts using DBSCAN. SPCL [6] creates more reliable clusters gradually by tuning the maximum neighbor distance ϵ of DBSCAN to a tight or loose criterion manually to mitigate the effects of noisy pseudo-labels. However, these methods are sensitive to the ϵ. Most of the above clustering-guided methods are somewhat sensitive to the parameters of the clustering algorithms they use. Our method chooses the time-tested clustering algorithm DBSCAN for clustering and tries to alleviate the sensitivity to the ϵ.

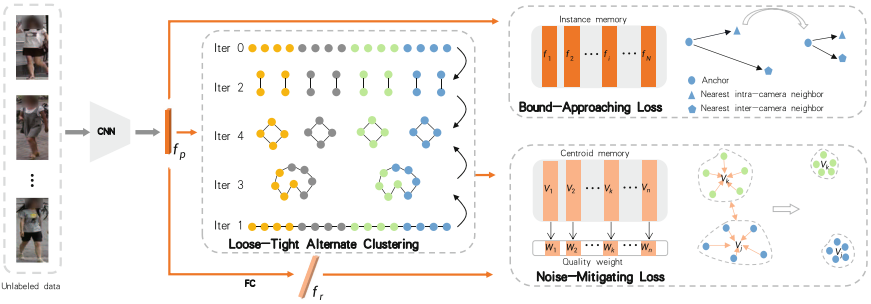

Fig. 2. The overview of the *Loose-Tight Alternate Clustering* method. The *Loose Bound* and *Tight Bound* are applied as clustering criteria alternately. The *Bound-Approaching Loss* narrows the gap between these two bounds, and the *Noise-Mitigating Loss* assigns higher weight to purer clusters during training to reduce the side-effects of clustering errors.

2.2 Camera-Aware Unsupervised Person re-ID

A key challenge in unsupervised person re-ID is the cross-camera scene variation. HHL and ECN [29,30] generate new fake images in the style of each camera for each sample and then enforce the camera-invariance to each person image and its corresponding camera-style transferred images. [12] imposes the neighborhood invariance constraint for inter-camera matching and intra-camera matching separately to improve the vanilla neighborhood invariance. [19] proposes a camera-aware similarity consistency loss which imposes the pairwise similarity distribution invariance upon cross-camera matching and intra-camera matching to alleviate the cross-camera scene variation. These methods play exemplary roles in leveraging the camera information. Our method further explores the integration of camera information with clustering-guided framework.

3 Methodology

3.1 Problem Definition

Under the setting of fully unsupervised person re-ID, a dataset X_t is provided that contains N_t images without any identity annotations. In addition, the number of cameras N_c and the camera-ID of each image (i.e. $C = \{c_i\}_{i=1}^{N_t}, c_i \in [0, N_c)$) is available. The goal is to learn a re-ID model on X_t, which aims to search for the images containing the same person as that in the query image.

As a popularly used unsupervised technique, clustering based pseudo-label methods achieve state-of-the-art performance in unsupervised person re-ID [4,6,14]. Performing a certain clustering mechanism such as DBSCAN on the pedestrian images collected from multiple cameras is the key step in this kind of methods. Due to the visual diversity of different cameras, the distances of inter-camera and intra-camera image pairs vary a lot. The images from the same camera with similar background and lightness tend to have much smaller distance

than those from different cameras. The diversity of the distance brings new challenges to the traditional clustering method, which usually cheats each instance equally. We address the challenge of clustering multi-camera pedestrian images as a ***Multi-domain Clustering*** problem, where each camera can be viewed as a different domain with specific visual style. How to effectively integrate the diversity of domains into the clustering method is one core task addressed in this paper.

3.2 Loose and Tight Clustering Bounds

Most of state-of-the-art unsupervised person re-ID algorithms [4,6,14,24] are based on the DBSCAN [2] clustering algorithm to achieve great performance advances. However, as the most important parameter in DBSCAN, the maximum distance between neighbors (ϵ) affects the clustering results seriously. As shown in Fig. 1(b), the performance of the models are very sensitive to the changing of ϵ.

There are two kinds of errors while conducting DBSCAN: the *Mix Error* and *Split Error* as shown in Fig. 1(c). In particular, applying a large ϵ may yield a loose clustering introducing the *Mix Error*, where the images with large distance from different identities are mixed into one cluster. On the contrary, using a small ϵ may yield a tight clustering causing the *Split Error*, where the images with the same identities may be separated. Most of the DBSCAN based pseudo-label methods need a large labeled validation set for careful tuning of ϵ and are difficult to be generalized to different datasets.

Since the optimal ϵ is not easy to decide according to unlabeled data, we go another way to solve the problem by seeking the proper bounds of clustering criteria. As observed in the research [27], the most similar image pairs in the same camera are very possible from the same person, which are usually sampled from the continuous frames of the camera. On the other hand, while only considering the cross-camera image pairs, the most similar ones also tend to be from the same person who walks across different cameras. Based on these observations, we define two bounds of distance, the ***Loose Bound*** and ***Tight Bound***, for clustering criteria. Specifically, the *Tight Bound* is defined as the average distance of intra-camera nearest neighbors:

$$\epsilon_T = \frac{1}{N_t} \sum_{i=1}^{N_t} \min d_r(x_i, x_{ia}), \forall x_{ia} \in C_i, x_{ia} \neq x_i \tag{1}$$

C_i is the set of images that are captured from the same camera with x_i and the $d_r(\cdot)$ is the popular and effective jaccard distance computed with k-reciprocal encoding [27]. On the other hand, the *Loose Bound* is defined as the average of the smallest distance between cross-camera images.

$$\epsilon_L = \frac{1}{N_t} \sum_{i=1}^{N_t} \min d_r(x_i, x_{ie}), \forall x_{ie} \notin C_i \tag{2}$$

While adopted as the hyper-parameter of the maximum neighboring distance in DBSCAN, the *Tight Bound* ϵ_T is small enough to tightly group those positive intra-camera matchings, leading to small *Mix Error* and large *Split Error*. On the other hand, the *Loose Bound* ϵ_L is large enough for DBSCAN [2] to group the instance loosely, which may lead to small *Split Error* and large *Mix Error*. How to balance these two opposite bounds makes up the core task in the following presented Loose-Tight Alternate Clustering model.

3.3 Loose-Tight Alternate Clustering

While performing clustering according to the *Tight Bound* ϵ_T and the *Loose Bound* ϵ_L respectively, two types of pseudo-labels can be achieved based on the tight and loose clustering results. The re-ID model can learn useful knowledge from both types of supervised signals. However, overfitting to any kind of pseudo-labels will limit the final performance of the re-ID model. As shown in Fig. 2, we train the re-ID model with these two types of pseudo-labels alternately to avoid the model being biased towards each kind of pseudo-labels. The detail of the alternate clustering method is shown in Algorithm 1.

Algorithm 1. Loose-Tight Alternate Clustering

Input: unlabeled dataset X_t.

Initialize : T : total iterations , I : current number of iterations , Ψ: the visual model needed to optimize, E: learning epochs in each iteration.
$I \leftarrow 0$
while $I < T$ **do**
  Compute the *Tight Bound* ϵ_T according to Eq. (1)
  Compute the *Loose Bound* ϵ_L according to Eq. (2)
  if $I\%2 == 0$ **then**
    $\overline{M} \leftarrow$ DBSCAN with ϵ_T
  else
    $\overline{M} \leftarrow$ DBSCAN with ϵ_L
  end if
  Train Ψ based on $\overline{M}$ by minimizing L_t (Eq. (6)) for E epochs
  $I \leftarrow I + 1$
end while
return Ψ

In particular, the clustering result in each iteration of clustering is defined as:

$$\overline{M} = \{\overline{M}_k | 0 \leq k < n\} \tag{3}$$

where n is the number of clusters, k is the cluster-ID, and $\overline{M}_k$ is the k^{th} cluster. By assigning the pseudo-label of each sample as its cluster-ID, the re-ID model can be trained with the cross-entropy loss, which is formulated as follows:

$$L_{tc} = -\sum_{k=0}^{n-1} \sum_{x_i \in \overline{M}_k} log(\frac{e^{V_k f_i}}{\sum_{j=0}^{n-1} e^{V_j f_i}}) \tag{4}$$

where V_k is the centroid vector of the k^{th} cluster $\overline{M}_k$ and f_i is the feature vector of the instance x_i. After learning from an instance x_i, the centroid of the k^{th} cluster will be updated by $V_k \leftarrow (V_k + f_i)/2$.

By minimizing the loss L_{tc}, the visual feature of an instance is dragged to the centroid of the cluster it belongs to, and pushed away from other clusters. In this way, while using the *Loose Bound* as the clustering criterion, the feature vectors of the instances, which have lower possibility to come from the same person, are pushed away. Meanwhile, the cross-camera images with relative smaller distance are mixed together, which may help to reduce the cross-domain diversity of visual styles. Furthermore, while using the *Tight Bound*, the feature vectors of intra-camera images with smaller distance are dragged together, which have high possibility to share the same identities. By alternately using these two kinds of bounds, the model can compress the *Split Error* and *Mix Error* alternately.

The *Loose Bound* and the *Tight Bound* are re-calculated in each iteration (Line 4 and 5 of Algorithm 1) and adapt to the visual diversity of different cameras. Larger gap between these two bounds indicates the larger domain diversity of this classic Multi-domain Clustering problem. Thus, reducing the gap may reduce the diversity and help improve the accuracy of the clustering. Motivated by this analysis, we propose a simple *Bound-Approaching Loss* (BAL) to narrow the gap between two bounds by minimizing the difference between the intra-camera nearest neighbor distance and the inter-camera nearest neighbor distance:

$$L_{ba} = \sum_i \max(\min_{x_i \circ x_j} d(x_i, x_j) - \min_{x_i \bullet x_k} d(x_i, x_k), 0) \tag{5}$$

where $d(\cdot)$ is the simple Euclidean distance. $x_i \circ x_j$ denotes that x_i and x_j are from different cameras, while $x_i \bullet x_k$ denotes that they are from the same camera.

The cross-entropy loss L_{tc} and the *Bound-Approaching Loss* L_{ba} are combined together as follows to optimize the visual model, as used in the Line 11 of Algorithm 1.

$$L_t = L_{tc} + L_{ba} \tag{6}$$

Furthermore, to facilitate calculating the distance of intra-camera and inter-camera image pairs, we maintain an instance memory bank $\mathcal{I}$ that stores the feature of each sample. During the back-propagation in each iteration, we update the memory bank for the training sample x_i through

$$\mathcal{I}[i] = (\mathcal{I}[i] + f_i)/2 \tag{7}$$

where $\mathcal{I}[i]$ is the memory of x_i in the i-th slot, f_i is the feature of x_i.

3.4 Quality Measurement Based Learning

During the training in each iteration, the cross-entropy loss L_{tc} of Eq. (4) is used to optimize the visual model based on the pseudo-labels, which are the cluster IDs achieved by the clustering algorithm. The quality of the clustering results

determines the correctness of the pseudo-labels. In order to make the model learn more knowledge from more reliable pseudo labels, we further extend this loss to the following *Noise-Mitigating Loss* to consider the quality of each cluster and assign higher weight to the pseudo-labels of purer clusters:

$$L_{nm} = -\sum_{k=0}^{n-1} W_k \sum_{x_i \in \overline{M}_k} log(\frac{e^{V_k f_i}}{\sum_{j=0}^{n-1} e^{V_j f_i}}) \tag{8}$$

where W_k indicates the quality measurement of the k^{th} cluster $\overline{M}_k$. To obtain W_k, we first compute the intra-cluster dissimilarity a_i and the inter-cluster dissimilarity b_i for $\forall x_i \in \overline{M}_k$ by:

$$a_i = \frac{1}{N_k - 1} \sum_{\substack{x_j \in \overline{M}_k \\ j \neq i}} d_r(x_i, x_j) \tag{9}$$

$$b_i = \frac{1}{N_t - N_k} \sum_{x_o \notin \overline{M}_k} d_r(x_i, x_o) \tag{10}$$

where N_k is the size of the cluster $\overline{M}_k$, N_t is the size of the whole dataset, $d_r(\cdot)$ is the jaccard distance computed with k-reciprocal encoding [27]. Then the quality score of $\overline{M}_k$ is then defined as the average silhouette coefficient of the samples within $\overline{M}_k$, which is formulated as follows:

$$Q_k = \frac{1}{N_k} \sum_{x_i \in \overline{M}_k} \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{11}$$

Furthermore, we normalize the quality score of each cluster via the exp maximum and minimum normalization:

$$W_k = \frac{e^{Q_k} - \min\limits_{j=1..n}(e^{Q_j})}{\max\limits_{j=1..n}(e^{Q_j}) - \min\limits_{j=1..n}(e^{Q_j})} + \alpha \tag{12}$$

where α is the positive constant to prevent the weight of the cluster with the lowest quality score from being set to zero. α is set as 0.01 in all experiments. By using this quality measurement W_k, the clusters that have higher intra cohesion and outer separation from other clusters will be assigned with higher weight when updating the parameters of the visual model. In this way, the negative effects of the noise in pseudo-labels can be further mitigated.

By combining the *Bound-Approaching Loss* L_{ba} with the *Noise-Mitigating Loss* L_{nm}, the complete loss function is defined as follows:

$$L'_t = L_{ba} + L_{nm}, \tag{13}$$

which aims to narrow the gap between the Loose and Tight Bounds and mitigate the pseudo-labels noise. Accordingly, the loss function L_t in Line 11 of Algorithm 1 can be replaced with L'_t here to enhance the performance of the learnt model.

4 Experiments

4.1 Datasets and Evaluation Protocol

Market1501. Market1501 [9] is a large scale person re-ID benchmark that contains 32,688 images of 1501 identities captured by 6 cameras. Specifically, 12,936 images of 751 identities are provided for training and the rest 19,732 images of 750 identities are for testing.

MSMT17. MSMT17 [18] is a newly released benchmark that contains 126,411 images of 4,101 identities collected from 15 non-overlapping camera views. It contains 32,621 images of 1,041 identities for training. The query contains 11,659 images of 3,060 identities and the gallery includes 126,441 images.

Evaluation Protocol. We utilize the Cumulative Matching Characteristic (CMC) curve and the mean average precision (mAP) to evaluate the performance of the proposed method. Furthermore, we report the Rank-1, Rank-5, Rank-10 scores to represent the CMC curve.

4.2 Implementation Details

We adopt the ResNet-50 [7] pre-trained on ImageNet [1] as the backbone of our model. The input image is resized to 256×128. The mini-batch size is 64. Random cropping, flipping, and random erasing [28] are adopted as data augmentation strategies. The SGD optimizer is used with the learning rate as 3.5×10^{-3}. Furthermore, each re-ID model is trained for 60 iterations. During each iteration, 800 epochs are executed.

4.3 Ablation Studies

Effectiveness of Alternate Clustering. To prove the necessity and importance of clustering with the *Loose Bound* ϵ_L and *Tight Bound* ϵ_T, we conduct experiments which only utilize ϵ_L or ϵ_T to cluster. The re-ID model is trained based on the vanilla cross-entropy loss L_{tc} (Eq. (4)). The experimental results are reported in the Table 1. When clustering only with ϵ_L, the clustering criterion is too loose, resulting in a lot of samples being grouped into one cluster. In this case, the *Mix Error* of pseudo-labels is pretty high, leading to the collapse of the re-ID model. When clustering only with ϵ_T, the clustering criterion is tight and the samples in each cluster are possibly fewer. In this case, the clustering accuracy will be higher, and the re-ID model can learn more useful knowledge from these kinds of pseudo-labels. However, the tight clustering criterion may not be able to group those positive inter-camera matchings into the same cluster, limiting the further improvement of the re-ID model. By training with these two types of pseudo-labels alternately (LTAC), the re-ID model is able to learn more useful knowledge and avoids being biased towards either of the two kinds of pseudo-label noise. Furthermore, we illustrate the number of clusters during training

Table 1. Ablation studies of the proposed method on Market-1501 and MSMT17. $LTAC\_L$ means the model only using the Loose Bound, and $LTAC\_T$ means only using the Tight Bound. L_{tc},L_{ba} and L_{nm} are the three loss functions described in Eqs. 4, 5 and 8 respectively.

Methods	Market-1501				MSMT17			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
$LTAC\_L + L_{tc}$	3.1	9.8	21.3	27.7	1.8	6.9	12.3	15.8
$LTAC\_T + L_{tc}$	61.8	81.4	89.7	92.6	15.0	40.1	50.9	55.2
$LTAC + L_{tc}$	69.5	86.6	93.6	95.8	17.7	42.7	53.3	57.9
$LTAC + L_{tc} + L_{ba}$	72.4	88.3	94.8	97.3	18.1	45.6	57.0	62.1
$LTAC + L_{nm}$	70.7	87.9	94.2	95.9	19.4	47.6	58.2	63.3
$LTAC + L_{nm} + L_{ba}$	**73.2**	**89.3**	**95.4**	**97.3**	**21.5**	**51.2**	**61.6**	**67.1**

Table 2. Comparison with state-of-the-arts fully unsupervised person re-ID methods on Market1501 and MSMT17. "None" means using the model pretrained on Imagenet. **Bold** indicates the best and underlined the runner-up. $*$ denotes using the back-bond method Resnet-50 like us.

Methods	Market-1501					MSMT17				
	Source	mAP	Rank-1	Rank-5	Rank-10	Source	mAP	Rank-1	Rank-5	Rank-10
OIM [20]	None	14.0	38.0	58.0	66.3	None	–	–	–	–
BUC [10]	None	38.3	66.2	79.6	84.5	None	–	–	–	–
SSL [11]	None	37.8	71.7	83.8	87.4	None	–	–	–	–
MMCL [16]	None	45.5	80.3	89.4	92.3	None	11.2	35.4	44.8	49.8
HCT [23]	None	56.4	80.0	91.6	95.2	None	–	–	–	–
IICS [21]$*$	None	67.1	85.5	–	–	None	–	–	–	–
SPCL [6]	None	73.1	88.1	95.1	97.0	None	19.1	42.3	55.6	61.2
Ours	None	**73.2**	**89.3**	**95.4**	**97.3**	None	**21.5**	**51.2**	**62.7**	**67.1**

on Market-1501 in Fig. 3(a). It can be observed that the quantity of clusters is closer to ground-truth identities when training with LTAC using both Loose and Tight Bounds.

Effectiveness of the Bound Approaching Loss. To evaluate the effectiveness of the *Bound-Approaching Loss*, we train the re-ID model in four different settings as reported in the last 4 rows of Table 1. It can be observed that no matter we train the re-ID model with the traditional cross-entropy loss (L_{tc}) or the quality weighted loss (L_{nm}), adding the *Bound-Approaching Loss* L_{ba} can lead to a further improvement on both two large-scale benchmarks. Furthermore, we illustrate the dynamic changes of the gap between the bounds during training on Market-1501 in Fig. 3(b). When we train the re-ID model with L_{ba}, the gap between ϵ_L and ϵ_T decreases to zero gradually. This proves that the *Bound-Approaching Loss* can reduce the visual diversity of different cameras significantly.

Effectiveness of Quality Measurement based Learning. We evaluate the effectiveness of the Noise-Mitigating Loss L_{nm} used in the *Quality Measurement based Learning* as described in Sect. 3.4 on Market-1501 and MSMT17. The experimental results are reported in the Table 1. It can be observed that training the re-ID model with the Noise-Mitigating Loss L_{nm} leads to a higher performance than training the re-ID model with the traditional cross-entropy loss (L_{tc}). Specifically, the mAP improves from 72.4% to 73.2% and 18.1% to 21.5% when training on Market-1501 and MSMT17. The improvement is more obvious on MSMT17, since it is more challenging and the scale of pseudo-labels noise is larger. This proves the effectiveness of the *Quality Measurement based Learning* to mitigate the negative effects of pseudo-label noise.

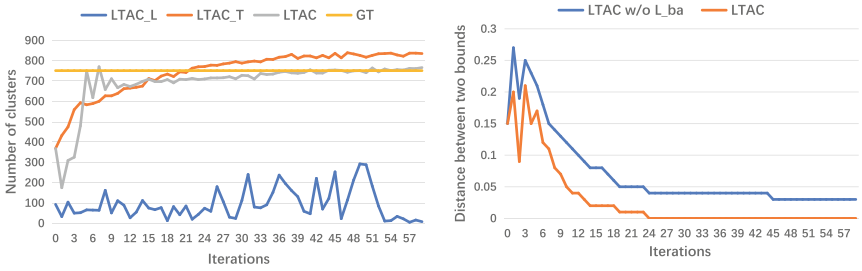

Fig. 3. Result of LTAC on Market-1501. (a): The dynamic changes of cluster numbers. $LTAC\_L$ means only using the Loose Bound ϵ_L and $LTAC\_T$ means only using the Tight Bound ϵ_T. GT indicates the ground-truth cluster number. (b): The distance between ϵ_L and ϵ_T. $LTAC\ w/o\ L\_ba$ means training without using L_{ba}.

4.4 Comparison with State-of-the-Art Methods

Our method is compared with state-of-the-art fully unsupervised re-ID methods in Table 2, which shows that $LTAC$ can achieve the best performance in all cases. It is interesting to observe that the superiority of our method is more obvious in the larger dataset MSMT17, which verifies the better generalization ability of our method. Furthermore, we also test the performance in the unsupervised domain adaptation (UDA) scenario, where the models are transferred from a labeled source domain to an unlabeled target domain. Table 3 shows the comparison results with state-of-the-art UDA methods. Our method outperforms all UDA methods using DBSCAN (e.g. SSG [4], MMT [5,6] SPCL [6]). More importantly, our method doesn't require any manual tuning of the sensitive clustering parameters, so it is more robust and competitive in real-world applications.

4.5 Robustness Evaluation

To further evaluate the robustness of our method, we design and implement several experiments to simulate the dynamic changing of the target domain. Specifically, we randomly select some different cameras in the dataset and augment

Table 3. Comparison with state-of-the-arts unsupervised domain adaptive person re-ID methods on Market1501 and MSMT17. **Bold** indicates the best and underlined the runner-up.

Methods	Market-1501					MSMT17				
	Source	mAP	Rank-1	Rank-5	Rank-10	Source	mAP	Rank-1	Rank-5	Rank-10
PAUL [22]	MSMT17	40.1	68.5	82.4	87.4	Market	–	–	–	–
ECN++ [31]	MSMT17	–	–	–	–	Market	15.2	40.4	53.1	58.7
SSG∗ [4]	MSMT17	–	–	–	–	Market	13.2	31.6	-	49.6
DG-Net++ [32]	MSMT17	64.6	83.1	91.5	94.3	Market	22.1	48.4	60.9	66.1
D-MMD [13]	MSMT17	50.8	72.8	88.1	92.3	Market	13.5	29.1	46.3	54.1
MMT-dbscan∗ [5,6]	MSMT17	75.6	89.3	95.8	97.5	Market	24.0	50.1	63.5	69.3
SPCL [6]	MSMT17	77.5	89.7	96.1	97.6	Market	**26.8**	53.7	65.0	69.8
Ours	MSMT17	**80.4**	**92.8**	**97.2**	**98.0**	Market	26.0	**56.1**	**67.5**	**72.4**

Table 4. Robustness comparison between our method and SPCL. "Supervised" means supervised learning as the upper bound. "Noise/x" indicates the noise is added to x cameras. "Improvement" means the improvement of our method relative to SPCL.

Methods	Market-1501							
	Noise/0		Noise/2		Noise/4		Noise/6	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Supervised	82.2	91.8	78.8	90.4	68.8	83.8	65.2	82.3
SPCL [6]	73.1	88.1	**65.9**	82.7	46.9	67.9	41.4	62.7
Ours	**73.2**	**89.3**	65.3	**84.5**	**49.4**	**73.4**	**44.6**	**68.8**
Improvement(%)	0.14↑	1.36↑	0.91↓	2.18↑	5.33↑	8.1↑	7.73↑	9.73↑
Methods	MSMT17							
	Noise/0		Noise/5		Noise/10		Noise/15	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
Supervised	44.5	70.5	30.3	56.4	21.2	42.6	17.4	39.8
SPCL [6]	19.1	42.3	9.1	21.3	4.8	11.4	4.8	11.8
Ours	**21.5**	**51.2**	**15.9**	**39.6**	**8.4**	**23.3**	**9.2**	**26.7**
Improvement(%)	12.57↑	21.04↑	74.73↑	85.92↑	75.00↑	104.39↑	91.67↑	126.27↑

the images with randomly selected noise generated by the imgaug library [8]. We utilize four types of weather noise including clouds, fog, snow, and rain. Table 4 shows the comparison results with the state-of-the-art unsupervised method SPCL [6] under different experimental settings. As the number of polluted cameras increases, the performance of all methods declines. However, the performance of our method outperforms SPCL [6] with a large margin, especially in the case with the highest ratio of noise. In particular, when we randomly select six cameras of Market1501 for noise augmentation, our method achieves 68.8% Rank-1 precision while SPCL [6] only achieves 62.7% Rank-1 precision. When we randomly select five cameras of MSMT17 for noise augmentation, our method achieves 15.9% mAP and 39.6% Rank-1, which exceeds SPCL [6] by 6.8% and 18.3% respectively. The experimental results in Table 4 illustrate

that our method is more robust than SPCL [6]. As we calculate the clustering parameters based on the statistics of unlabeled data without manually setting clustering parameters, our method is more applicable to complex and dynamic realistic scenes.

5 Conclusion

In this paper, we proposed a *Loose-Tight Alternate Clustering* framework which explores the *Loose Bound* and *Tight Bound* in multi-domain clustering, and learns from two types of pseudo-labels alternately. The two bounds were obtained on the basis of the inter-camera nearest neighbor distance and the intra-camera nearest neighbor distance. A *Bound-Approaching Loss* was further proposed to narrow the gap between these two bounds to reduce the domain diversity. Furthermore, a *Quality Measurement based Learning* method was introduced to mitigate the negative effects of the pseudo-label noise. Experiments on two large benchmarks demonstrated the applicability, competitiveness and robustness of our method.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61876065), the Special Fund Project of Marine Economy Development in Guangdong Province([2021]35), and Guangzhou Science and Technology Program key projects (202007040002).

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: CVPR (2009)
2. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD (1996)
3. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person re-identification: clustering and fine-tuning. ACM (TOMM) (2018)
4. 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 (2019)
5. Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. In: ICLR (2020)
6. Ge, Y., Zhu, F., Chen, D., Zhao, R., Li, H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID. In: NeurIPS (2020)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
8. Jung, A.B., et al.: Imgaug (2020). https://github.com/aleju/imgaug. Accessed 01 Feb 2020
9. Liang, Z., Liyue, S., Lu, T., Shengjin, W., Jingdong, W., Qi, T.: Scalable person re-identification: a benchmark. In: ICCV (2015)
10. Lin, Y., Dong, X., Zheng, L., Yan, Y., Yang, Y.: A bottom-up clustering approach to unsupervised person re-identification. In: AAAI (2019)

11. Lin, Y., Xie, L., Wu, Y., Yan, C., Tian, Q.: Unsupervised person re-identification via softened similarity learning. In: CVPR (2020)
12. Luo, C., Song, C., Zhang, Z.: Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. In: ECCV (2020)
13. Mekhazni, D., Bhuiyan, A., Ekladious, G.S.E., Granger, E.: Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: ECCV (2020)
14. Song, L., et al.: Unsupervised domain adaptive re-identification: theory and practice. Pattern Recogn. (2020)
15. 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: ECCV (2018)
16. Wang, D., Zhang, S.: Unsupervised person re-identification via multi-label classification. In: CVPR (2020)
17. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACM MM (2018)
18. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: CVPR (2018)
19. Wu, A., Zheng, W., Lai, J.: Unsupervised person re-identification by camera-aware similarity consistency learning. In: ICCV (2019)
20. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. In: CVPR (2017)
21. Xuan, S., Zhang, S.: Intra-inter camera similarity for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11926–11935 (2021)
22. Yang, Q., Yu, H., Wu, A., Zheng, W.: Patch-based discriminative feature learning for unsupervised person re-identification. In: CVPR (2019)
23. Zeng, K., Ning, M., Wang, Y., Guo, Y.: Hierarchical clustering with hard-batch triplet loss for person re-identification. In: CVPR (2020)
24. Zhai, Y., et al.: Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification. In: CVPR (2020)
25. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: ICCV (2019)
26. Zheng, F., et al.: Pyramidal person re-identification via multi-loss dynamic training. In: CVPR (2019)
27. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR (2017)
28. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: AAAI (2020)
29. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero- and homogeneously. In: ECCV (2018)
30. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: exemplar memory for domain adaptive person re-identification. In: CVPR (2019)
31. Zhun, Z., Liang, Z., Zhiming, L., Shaozi, L., Yi, Y.: Learning to adapt invariance in memory for person re-identification. In: TPAMI (2020)
32. Zou, Y., Yang, X., Yu, Z., Kumar, B.V.K.V., Kautz, J.: Joint disentangling and adaptation for cross-domain person re-identification. In: ECCV (2020)