CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification

Yiyu Chen, Zheyi Fan, Zhaoru Chen, Yixuan Zhu

Beijing Institute of Technology, China

{yiyuchen1998, Cchenzhaoru}@gmail.com, {funye, zhuyixuan}@bit.edu.cn

Abstract

Person re-identification (re-ID) is a challenging task that aims to learn discriminative features for person retrieval. In person re-ID, Jaccard distance is a widely used distance metric, especially in re-ranking and clustering scenarios. However, we discover that camera variation has a significant negative impact on the reliability of Jaccard distance. In particular, Jaccard distance calculates the distance based on the overlap of relevant neighbors. Due to camera variation, intra-camera samples dominate the relevant neighbors, which reduces the reliability of the neighbors by introducing intra-camera negative samples and excluding inter-camera positive samples. To overcome this problem, we propose a novel camera-aware Jaccard (CA-Jaccard) distance that leverages camera information to enhance the reliability of Jaccard distance. Specifically, we design camera-aware k-reciprocal nearest neighbors (CK-RNNs) to find k-reciprocal nearest neighbors on the intra-camera and inter-camera ranking lists, which improves the reliability of relevant neighbors and guarantees the contribution of inter-camera samples in the overlap. Moreover, we propose a camera-aware local query expansion (CLQE) to mine reliable samples in relevant neighbors by exploiting camera variation as a strong constraint and assign these samples higher weights in overlap, further improving the reliability. Our CA-Jaccard distance is simple yet effective and can serve as a general distance metric for person re-ID methods with high reliability and low computational cost. Extensive experiments demonstrate the effectiveness of our method. Code is available at https://github.com/chen960/CA-Jaccard/.

1. Introduction

Person re-identification (re-ID) aims to retrieve persons across non-overlapping camera views. It has drawn wide attention due to the growing demand for intelligent surveillance systems. Thanks to the advancement of deep learning, supervised re-ID methods [6, 25, 30, 33, 50] have achieved remarkable performance. However, these methods rely on
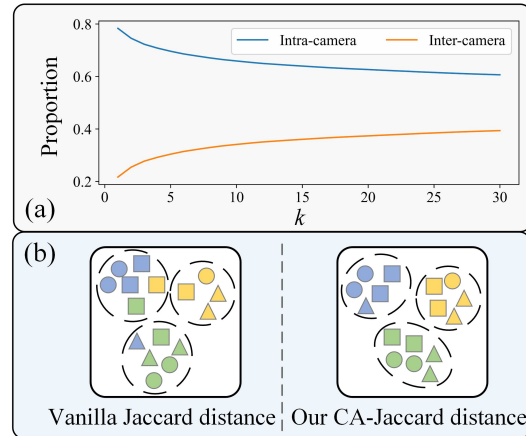

Figure 1. (a) Illustration of the average proportion of intra-camera and inter-camera samples in k-nearest neighbors of all samples. Due to camera variation, the average proportion of intra-camera samples in all samples' k-nearest neighbors is significantly higher than that of inter-camera samples. (b) Comparison of the feature spaces of using Jaccard distance and our CA-Jaccard distance. Different colors represent different identities and different shapes indicate different camera labels.

sufficient person identity annotations, limiting their application in real world scenarios. Hence, recent studies have focused on unsupervised re-ID, seeking to learn discriminative features using unlabeled data.

Recently, most state-of-the-art unsupervised re-ID methods, *i.e.* clustering-based re-ID methods [10, 13, 15, 44, 46], generally employ a iterative two-stage training procedure: 1) generating pseudo labels based on the Jaccard distance [51] between all training samples using a clustering algorithm [3, 11, 23]; 2) training the re-ID model with the generated pseudo labels. Despite their effectiveness, these methods still suffer from label noise. To overcome the above problem, numerous approaches [8, 14, 15, 17, 38, 43, 45] have been proposed. These approaches focus on improving or refining pseudo labels after clustering.

Moreover, there also have some re-ranking methods [24, 28, 51] are proposed to further improve the performance of re-ID methods. K-reciprocal re-ranking [51] is a popular

re-ranking method, which utilizes Jaccard distance to re-calculate the distance.

As mentioned above, Jaccard distance [51] is widely used in person re-ID. However, Jaccard distance overlooks the detrimental impact of camera variation (*e.g.* viewpoint, illumination and background), which substantially contributes to label noise in clustering scene and performance degradation in re-ranking scene. Specifically, based on the original distance matrix (*i.e.* Euclidean distance or cosine distance), Jaccard distance measures the distance between samples based on the overlap of their relevant neighbors, which means the accuracy of relevant neighbors determines the reliability of Jaccard distance. The relevant neighbors are found by applying k-reciprocal nearest constraint and encoded into a weighted unit vector called weighted neighbors vector. Higher weights are assigned to closer neighbors to reflect their greater contribution to the overlap calculation. However, as shown in Fig. 1 (a), due to camera variation, intra-camera samples dominate the k-nearest neighbors, resulting high proportion and weight of intra-camera samples in the weighted relevant neighbors vectors. It undermines the reliability of Jaccard distance by introducing many high weight intra-camera negative samples and hindering informative inter-camera positive samples into weighted relevant neighbors vectors. Moreover, Jaccard distance utilizes local query expansion to expand weighted relevant neighbors vector of a sample by averaging the weighted relevant neighbors vectors of its k-nearest neighbors. Since k-nearest neighbors mainly consist of intra-camera samples, the proportion and weight of intra-camera samples are further increased, while the reliability of Jaccard distance is further decreased.

To address these problems, we propose camera-aware Jaccard (CA-Jaccard) distance, a simple yet effective distance metric that enhances the reliability of the Jaccard distance [51] with camera information for more accurate pseudo label generation, which is shown in Fig. 1 (b). In particular, our approach modifies the robust k-reciprocal nearest neighbors (KRNNs) and local query expansion (LQE) of Jaccard distance in a camera-aware manner to increase the accuracy of relevant neighbors. We discover that inter-camera samples have more information and reliability. Therefore, to include more inter-camera samples into relevant neighbors and restrain the proportion and weight of intra-camera samples under camera variation, we propose camera-aware k-reciprocal nearest neighbors (CK-RNNs) for more accurate relevant neighbors. CKRNNs impose the k-reciprocal nearest constraint separately for the intra-camera and inter-camera ranking lists with different k values, and then combine the neighbors obtained from both. Additionally, to further improve the accuracy of relevant neighbors, we propose camera-aware local query expansion (CLQE) to obtain weighted expanded neighbors vectors by averaging the weighted CKRNNs vectors of intra-camera and inter-camera k-nearest neighbors. CLQE exploits camera variation as a strong constraint to mine reliable samples that frequently appear in the relevant neighbors of both intra-camera and inter-camera k-nearest neighbors, and enlarges their weights for greater contribution in overlap.

Our contributions can be summarized as follows:

(1) We propose a novel camera-aware Jaccard (CA-Jaccard) distance that leverages camera-aware k-reciprocal nearest neighbors (CKRNNs) and camera-aware local query expansion (CLQE) to enhance the reliability of Jaccard distance.

(2) Our CA-Jaccard distance is simple yet effective, with higher reliability and lower computational cost than Jaccard distance, and can serve as a general distance metric for person re-ID.

(3) Extensive experiments on different person re-ID scenarios demonstrate the effectiveness of our CA-Jaccard distance.

2. Related work

2.1. Clustering for Unsupervised Person Re-ID

In unsupervised person re-ID, datasets lack identity label information. Many works utilize clustering [12, 13, 19, 21, 41, 42, 44] and k-nearest neighbors [32, 39, 52] to generate pseudo labels. Clustering-based methods [10, 13, 44] demonstrate their superiority by achieving state-of-the-art performance. They generally leverage the Jaccard distance [51] to compute the distance matrix and then adopt the DB-SCAN clustering [11] algorithm for pseudo label generation. However, the generated pseudo labels inevitably contain label noise, which severely affects the performance. Recent methods tackle this problem using robust clustering techniques [15, 42], label refinement procedures[8, 17, 45], co-teaching algorithms [14, 38, 43, 47]. Although these methods strive to reduce label noise, they neglect the label noise caused by unreliable Jaccard distance. RIDES [7] improves the original distance by reducing the distance of reliable inter-camera sample pairs, which improves the accuracy of relevant neighbors and the reliability of Jaccard distance implicitly and limitedly. Different from [7], our method enhances Jaccard distance directly, and improves the accuracy of relevant neighbors significantly and stably. In this paper, our method brings more reliable pseudo labels in clustering scene.

2.2. Re-ranking for Person Re-ID

Re-ranking is a post-processing technique to improve the original retrieval results by the information of near neighbors. In [29], k-nearest neighbors are first used for re-ranking. Many works [28, 40, 51] further discover more potential information based on k-nearest neighbors. To re-
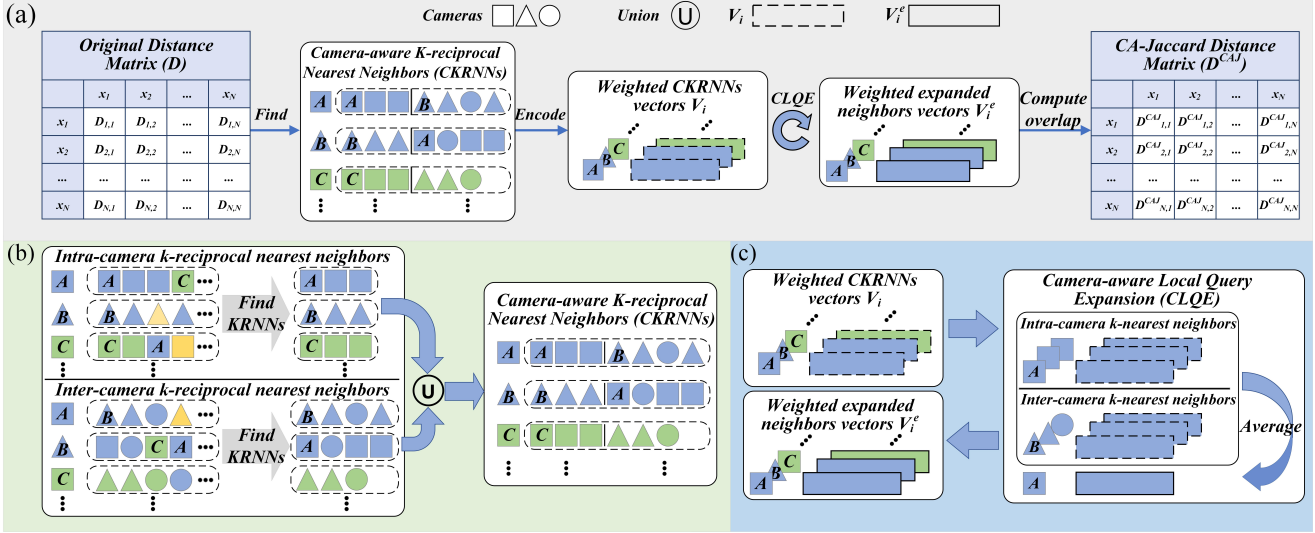
Figure 2. Schematic diagram of CA-Jaccard distance. (a) Overview of CA-Jaccard distance computation. Given the original distance matrix, find CKRNNs and encode them as vectors, then use CLQE to generate weighted expanded neighbors vectors. Finally, calculate the overlap between these vectors to obtain the CA-Jaccard distance matrix. (b) CKRNNs. CKRNNs find reliable relevant neighbors by applying the k-reciprocal nearest constraint on intra-camera and inter-camera ranking lists. (c) CLQE. CLQE averages the weighted CKRNNs vectors of intra-camera and inter-camera k-nearest neighbors to obtain weighted expanded neighbors.

duce the false positives in the top-k of original ranking lists, k-reciprocal nearest neighbors [18, 26] are introduced into person re-ID by K-reciprocal(KR) re-ranking[51]. Inspired by sparse contextual activation (SCA) encoding [1] and average query expansion (AQE) [9], KR re-ranking [51] searches k-reciprocal neighbors and computes the Jaccard distance with k-reciprocal encoding and local query expansion. ECN [28] improves the original pairwise distance by aggregating the distances between expanded neighbors of image pairs. These methods can not handle the large camera variation in ranking lists well, which significantly hinders the performance of re-ranking. To solve this problem, we attempt to aggregate the camera information into re-ranking.

3. Methodology

In this section, we first revisit Jaccard distance. Then, we elaborate on the details of our camera-aware Jaccard (CA-Jaccard) distance, which enhances the Jaccard distance by using camera-aware k-reciprocal nearest neighbors (CKRNNs) and camera-aware local query expansion (CLQE).

3.1. Preliminary

Our goal is to compute general and reliable distances between samples for different person re-ID scenarios. The computation procedures of CA-Jaccard distance in the re-ranking and clustering scene are very similar. Therefore, for similarity, we introduce our CA-Jaccard distance within the clustering scenario of clustering-based unsupervised re-

ID methods. In this case, we are provided with an unlabeled re-ID training dataset $X = \{x_i\}_{i=1}^{N}$ with N images, where x_i denotes the i-th image. Each image x_i is associated with its camera label c_i.

3.2. Revisit Jaccard Distance

The core idea of Jaccard distance is that if two images are similar, their relevant neighbors should also be similar. Based on this assumption, Jaccard distance measures the distance between samples according to the overlap of their relevant neighbors. Jaccard distance incorporates the robust k-reciprocal nearest neighbors (KRNNs) into relevant neighbors and then expands them using local query expansion (LQE). Since neighbors sets treat each neighbor equally and overlap computation of sets is time-consuming, Jaccard distance encodes the neighbors sets of samples into weighted unit vectors and transforms set comparison problem into pure vector computation. The detailed calculation steps of Jaccard distance are as follows.

Original distance computation. The original distance matrix D is obtained by applying either the cosine or Euclidean distance based on the features extracted by the model $f_\theta(\cdot)$ from all samples.

Robust k-reciprocal nearest neighbors. For sample x_i, the ranking list $L_i = \{x_1^i, x_2^i, ..., x_N^i\}$ can be obtained by arranging samples according to the original distance between x_i and all training samples. The k-nearest neighbors $N(x_i, k)$ of x_i are defined as the top-k samples of ranking list L_i:

$$N(x_i, k) = L_i[1:k]. \quad (1)$$

Then the KRNNs $R(x_i, k_1)$ can be found:

$$R(x_i, k_1) = \{x_j | x_i \in N(x_j, k_1) \wedge x_j \in N(x_i, k_1)\}. \quad (2)$$

To recall some positive samples may be excluded from the KRNNs, robust KRNNs are computed as follows:

$$R^*(x_i, k_1) \leftarrow R(x_i, k_1) \cup R\left(x_j, \tfrac{1}{2}k_1\right)$$
$$s.t. \left|R(x_i, k_1) \cap R\left(x_j, \tfrac{1}{2}k_1\right)\right| \geq \tfrac{2}{3}\left|R\left(x_j, \tfrac{1}{2}k_1\right)\right| \quad (3)$$
$$\forall \, x_j \in R(x_i, k_1),$$

where $|\cdot|$ denotes the number of samples in the set. This operation employs a strict constraint to ensure that most of the recalled samples are positive samples.

Vectorization of neighbors. To reduce the computational complexity and increase the discriminability of neighbors, the robust KRNNs of sample x_i are encoded into a weighted robust KRNNs vector $V_i = [V_{i,1}, V_{i,2}, ..., V_{i,N}]$, where V_i is a N-dimension unit vector and $V_{i,j}$ is computed according to the original distance between x_i and x_j if x_j is within the robust k-reciprocal nearest neighbor of x_i, otherwise it is zero:

$$V_{i,j} = \begin{cases} \dfrac{e^{-D_{i,j}}}{\sum\limits_{x_l \in R^*(x_i, k_1)} e^{-D_{i,l}}} & if \ x_j \in R^*(x_i, k_1) \\ 0 & otherwise, \end{cases} \quad (4)$$

where $D_{i,j}$ is the original distance between x_i and x_j.

Local query expansion. Considering similar samples may share similar features and neighbors, LQE is adopted to generate the weighted expanded neighbors vector V_i^e by averaging the weighted robust KRNNs vectors of x_i's k-nearest neighbors:

$$V_i^e = \frac{1}{|N(x_i, k_2)|} \sum_{x_j \in N(x_i, k_2)} V_j, \quad (5)$$

where $k_2 < k_1$ because there are noise in k-nearest neighbors, V_j denotes the weighted robust KRNNs vector of x_j.

Overlap computation. The Jaccard distance $D_{i,j}^J$ between x_i and x_j are computed by vectorized overlap computation:

$$D_{i,j}^J = 1 - \frac{\sum\limits_{l=1}^{N} \min\left(V_{i,l}^e, V_{j,l}^e\right)}{\sum\limits_{l=1}^{N} \max\left(V_{i,l}^e, V_{j,l}^e\right)}, \quad (6)$$

where min and max can be regarded as the intersection and union operation in vector form.

The Jaccard distance is widely used in many methods, but it still has drawbacks. Camera variation makes it difficult for robust KRNNs and LQE to obtain reliable relevant neighbors for overlap computation, which hinders the reliability of Jaccard distance. Therefore, the key motivation of our method is to improve the reliability of relevant neighbors. To achieve this goal, we propose CKRNNs and CLQE to make efforts from different aspects.

3.3. Camera-aware K-reciprocal Nearest Neighbors

Although robust KRNNs utilize some constraints to find relevant neighbors, the neighbors are still unreliable. Camera variation causes intra-camera samples to have a high proportion and low ranks in k-nearest neighbors. Consequently, they have a high proportion in robust KRNNs. Negative samples from the same camera are heavily included in robust KRNNs vectors with large weight, while informative and reliable inter-camera samples are hardly included, reducing the reliability of the neighbors. To find more inter-camera relevant samples and restrain the proportion and weight of intra-camera samples, we propose camera-aware k-reciprocal nearest neighbors (CKRNNs), as shown in Fig. 2(b).

For sample x_i, we obtain the intra-camera ranking list L_i^{intra} and inter-camera ranking list L_i^{inter}:

$$L_i^{intra} = \{x_1^{i\_intra}, x_2^{i\_intra}, ...x_{N_{c_i}}^{i\_intra}\}, \quad (7)$$

$$L_i^{inter} = \{x_1^{i\_inter}, x_2^{i\_inter}, ...x_{N-N_{c_i}}^{i\_inter}\}, \quad (8)$$

where $x_j^{i\_intra}$ and $x_j^{i\_inter}$ represent the j-th sample in the intra-camera and inter-camera ranking list, and N_{c_i} means the number of samples share the same camera label c_i as x_i.

Then we find k-nearest neighbors in both ranking lists to obtain intra-camera k-nearest neighbors $N^{intra}(x_i, k_1^{intra})$ and inter-camera k-nearest neighbors $N^{inter}(x_i, k_1^{inter})$:

$$N^{intra}(x_i, k_1^{intra}) = L_i^{intra}[1:k_1^{intra}], \quad (9)$$

$$N^{inter}(x_i, k_1^{inter}) = L_i^{inter}[1:k_1^{inter}], \quad (10)$$

where k_1^{intra} and k_1^{inter} mean different k are used in intra-camera and inter-camera ranking lists.

Next, we impose the k-reciprocal nearest constraint on both intra-camera and inter-camera k-nearest neighbors, and union the obtained neighbors as CKRNNs $R^c(x_i, k_1^{intra}, k_1^{inter})$, which can be formulated as:

$$R^c(x_i, k_1^{intra}, k_1^{inter}) =$$
$$\{x_j | x_i \in N^{intra}(x_j, k_1^{intra}) \wedge x_j \in N^{intra}(x_i, k_1^{intra})\}$$
$$\cup \{x_j | x_i \in N^{inter}(x_j, k_1^{inter}) \wedge x_j \in N^{inter}(x_i, k_1^{inter})\}. \quad (11)$$

By using a smaller k_1^{intra}, we can include only intra-camera positive samples and exclude intra-camera negative samples. We discover that inter-camera samples are more informative and reliable in overlap computation. Thus, we use a large k_1^{inter} to find more inter-camera samples, increasing the proportion of inter-camera samples in CKRNNs. When CKRNNs are encoded into a weighted CKRNNs vector, although the weight of each intra-camera sample is relatively large due to their small original distance, a large amount of inter-camera samples in CKRNNs ensure the proportion and total weight of inter-camera samples, which enhances the reliability of neighbors.

Note that we do not use recall operation which has a great positive effect on the reliability of robust KRNNs and Jaccard distance. This is because the key of the recall operation is to recall more inter-camera positive samples, which is explicitly achieved in CKRNNs by applying the k-reciprocal nearest constraint on the inter-camera ranking list.

3.4. Camera-aware Local Query Expansion

LQE is used in Jaccard distance to incorporate more samples from the robust KRNNs of k-nearest neighbors and reassign weights of neighbors by averaging weighted robust KRNNs vectors of k-nearest neighbors. Due to camera variation, most k-nearest neighbors are intra-camera samples, which also have a high proportion of intra-camera samples with high weights in their weighted robust KRNNs vectors. As a result, LQE reassigns higher weights to intra-camera negative samples which frequently occur in robust KRNNs of k-nearest neighbors, while reassigning lower weights to inter-camera positive samples which have low proportion but are informative and reliable. In this case, the unreliability of relevant neighbors is further exacerbated.

Unlike LQE which reduces the reliability of neighbors, we propose camera-aware local query expansion (CLQE) to boost the reliability of neighbors in a clever way. As shown in Fig. 2(c), CLQE averages the weighted CKRNNs vectors of intra-camera and inter-camera k-nearest neighbors to obtain weighted expanded neighbors:

$$V_i^e = \frac{1}{|N^{intra}(x_i, k_2^{intra})| + |N^{inter}(x_i, k_2^{inter})|} \times$$
$$(\sum_{x_j \in N^{intra}(x_i, k_2^{intra})} V_j + \sum_{x_l \in N^{inter}(x_i, k_2^{inter})} V_l),$$
$$(12)$$

where k_2^{intra} and k_2^{inter} are the k number of k-nearest neighbors we select from intra-camera and inter-camera ranking lists, V_j and V_l are the weighted CKRNNs vectors of x_j and x_l respectively. CLQE regards camera variation as a strong constraint to mine reliable samples in neighbors and enlarge their weights. Specifically, CLQE averages the weighted CKRNNs vectors of samples from multiple cameras. Due to the existence of camera variation, the reliability of a sample increases with its frequency of occurrence in the CKRNNs of samples from multiple cameras, indicating that it is more likely to be a positive sample. In this way, CLQE assigns reliable samples that have high frequency in the CKRNNs of intra-camera and inter-camera k-nearest neighbors higher weight.

3.5. Camera-aware Jaccard Distance

We name the proposed distance metric camera-aware Jaccard (CA-Jaccard) distance, which improves the reliability of Jaccard distance by replacing the robust KRNNs and

LQE with CKRNNs and CLQE. We utilize CKRNNs to increase the proportion and total weight of inter-camera samples which are informative and exclude the intra-camera samples beyond relevant neighbors, enhancing the reliability of neighbors. Meanwhile, we utilize CLQE to assign high weights to reliable samples, which further improves the reliability of relevant neighbors. Our CA-Jaccard distance is a simple but effective distance metric, offering lower computational complexity and higher reliability than Jaccard distance. The detailed steps of CA-Jaccard distance computation are presented in Fig. 2(a).

4. Experiments

4.1. Datasets and Evaluation Protocols

We evaluate the proposed method on two person re-ID datasets (Market1501 [48], MSMT17 [35]) and one vehicle re-ID dataset (VeRi-776 [22]). We adopt mean Average Precision (mAP) [2] and Cumulative Matching Characteristic (CMC) [16] to evaluate performance.

4.2. Implement Details

Our proposed CA-Jaccard distance can be applied in the clustering and re-ranking scenes of person re-ID. Therefore, to fully verify the effectiveness of our CA-Jaccard distance, we conduct experiments in both scenes. CA-Jaccard distance can be applied with marginal modification. Specifically, only the Jaccard distance needs to be replaced with CA-Jaccard distance, while all other settings remain unchanged. In CA-Jaccard distance, we set the k_1^{intra} and k_1^{inter} to 5 and 20 in Eq. (11). The k_2^{intra} and k_2^{inter} are set to 2 and 4 respectively in Eq. (12).

4.3. Performance Improvement in Clustering Scene

In Tab. 1, We verify the effectiveness of our CA-Jaccard distance by applying it in state-of-the-art unsupervised person re-ID methods (*e.g.* CAP [34], CC [10], ICE[4] and PPLR [8]). We can observe that when the CA-Jaccard distance is applied for clustering, the performance of these methods gains significant improvement. Especially when applying our CA-Jaccard distance to a more powerful method PPLR [8], we achieve 86.1%/94.4% mAP/Rank-1 on Market1501, 44.3%/75.1% mAP/Rank-1 on MSMT17, and 45.3%/90.4% mAP/Rank-1 on VeRi-776, which surpasses all unsupervised person re-ID methods by a large margin. Moreover, we can find that CA-Jaccard distance can bring greater performance improvement on the MSMT17 and VeRi-776 datasets with larger camera variation compared to Market1501, demonstrating that CA-Jaccard distance effectively solves the problem of unreliable Jaccard distance caused by camera variation. The results show the effectiveness and generalization of our method.

Table 1. Comparison with the state-of-the-art unsupervised re-ID methods on Market1501, MSMT17 and VeRi-776. The best results are in **bold** and the second-best results are in underline. CC* denotes our results with the official CC code without hard instance memory updating mechanism and generalized mean (GeM) pooling [27]. "CAJ" represents CA-Jaccard distance.

Methods	Reference	Market1501				MSMT17				VeRi-776			
		mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
MMCL [32]	CVPR'20	45.5	80.3	89.4	92.3	11.2	35.4	44.8	49.8	-	-	-	-
HCT [41]	CVPR'20	56.4	80.0	91.6	95.2	-	-	-	-	-	-	-	-
GCL [5]	CVPR'21	66.8	87.3	93.5	95.5	21.3	45.7	58.6	64.5	-	-	-	-
IICS [37]	CVPR'21	72.9	89.5	95.2	97.0	26.9	56.4	68.8	73.4	-	-	-	-
SpCL [15]	NeurIPS'20	73.1	88.1	95.1	97.0	19.1	42.3	55.6	61.2	36.9	79.9	86.8	89.9
RLCC [45]	CVPR'21	77.7	90.8	96.3	97.5	27.9	56.5	68.4	73.1	39.6	83.4	88.8	90.9
OPLG-HCD [49]	ICCV'21	78.1	91.1	96.4	97.7	26.9	53.7	65.3	70.2	-	-	-	-
MCRN [36]	AAAI'22	80.8	92.5	–	-	31.2	63.6	-	-	-	-	-	-
Secret [17]	AAAI'22	81.0	92.6	-	-	31.3	60.4	-	-	-	-	-	-
CC [10]	ACCV'22	82.6	93.0	97.0	98.1	33.3	63.3	73.7	77.8	42.5	87.7	91.4	93.1
RESL [20]	AAAI'22	83.1	93.2	96.8	98.0	33.6	64.8	74.6	79.6	-	-	-	-
RIDE [7]	SCIS'23	84.0	93.0	97.3	-	39.5	68.4	79.6	-	-	-	-	-
ISE [46]	CVPR'22	84.7	94.0	97.8	98.8	35.0	64.7	75.5	79.4	-	-	-	-
CAP [34]	AAAI'21	79.2	91.4	96.3	97.7	36.9	67.4	78.0	81.4	40.4	86.8	90.8	92.7
CAP [34]+CAJ	-	80.4	91.7	96.4	97.7	39.9	70.0	80.5	83.7	43.4	90.4	93.4	95.1
CC* [10]	ACCV'22	81.0	91.1	96.2	97.4	31.1	60.2	71.3	75.7	38.1	80.3	85.1	87.5
CC* [10]+CAJ	-	84.8	93.6	97.6	98.4	42.8	72.3	82.2	85.6	43.1	90.1	92.8	95.0
ICE [4]	ICCV'21	82.3	93.8	97.6	98.4	38.9	70.2	80.5	84.4	42.5	87.5	91.5	93.2
ICE [4]+CAJ	-	82.7	93.8	97.7	98.4	43.0	74.1	83.8	86.9	44.5	**91.0**	93.6	95.0
PPLR [8]	CVPR'22	84.4	94.3	97.8	98.6	42.2	73.3	83.5	86.5	43.5	88.3	92.7	94.4
PPLR+CAJ	-	**86.1**	**94.4**	**97.9**	**98.7**	**44.3**	**75.1**	**84.3**	**87.3**	**45.3**	90.4	**93.9**	**95.2**

4.4. Peformance Improvement in Re-ranking Scene

Table 2. Comparison with the state-of-the-art re-ranking methods for person re-ID on Market1501, MSMT17 and VeRi-776. The best results are in **bold**.

Methods	Market1501			MSMT17			VeRi-776		
	mAP	R1	R5	mAP	R1	R5	mAP	R1	R5
BoT [25]	85.9	94.5	**98.2**	50.7	74.0	85.6	76.2	95.5	98.1
+KR [51]	94.2	95.4	97.9	66.9	79.4	86.6	78.7	95.8	97.2
+ECN [28]	94.4	95.9	97.8	69.0	80.5	86.3	79.5	96.8	97.3
+CAJ	**94.5**	**96.2**	98.1	**74.1**	**86.2**	**90.5**	**81.4**	**97.6**	**98.3**
CC* [10]	81.0	91.1	96.2	31.1	60.2	71.3	38.1	80.3	85.1
+KR [51]	89.7	93.2	95.5	42.6	65.3	73.4	39.0	80.0	81.6
+ECN [28]	90.0	93.4	95.0	43.9	65.3	71.8	40.0	81.1	82.4
+CAJ	**90.2**	**93.7**	**95.9**	**45.3**	**68.9**	**75.3**	**42.5**	**88.6**	**91.1**

We apply CA-Jaccard to re-ranking the retrieval results of pre-trained models of supervised and unsupervised commonly used baselines (BoT [25] and CC [10]). For a fair comparison, we also apply the state-of-the-art re-ranking methods *i.e.* KR [51], and ECN [28]. The experiment results are reported in Tab. 2. We can observe that our method improves the performance of BoT and CC by a large margin. Meanwhile, our method consistently brings greater performance improvement than the state-of-the-art re-ranking methods. These results demonstrate the effectiveness and

superiority of our method.

4.5. Ablation Studies

In this section, we conduct extensive experiments on Maret1501 and MSMT17 in clustering and re-ranking scenes to validate the effectiveness of each component in our method. We select CC with instance memory updating mechanism and average pooling as the baseline for clustering scene and BoT as the baseline for re-ranking scene. We present the results of the baselines and three variants of our CA-Jaccard distance in Tab. 3. Then we analyze each component of our method respectively.

Effect of CKRNNs. To verify the effectiveness of CK-RNNs, we replace the robust KRNNs in Jaccard distance with CKRNNs. The results in Tab. 3 demonstrate a significant performance improvement compared to the baselines. Specifically, in the clustering scene, applying CK-RNNs brings 1.4% mAP and 1.8 % Rank-1 improvement on Market1501, and 4.0% mAP and 5.1% Rank-1 improvement on MSMT17. In the re-ranking scene, CKRNNs consistently improve the performance of BoT+KR. Especially on the challenging MSMT17 datasets, it brings 6.2%/4.6% mAP/Rank-1 improvement. These results validate the effectiveness of the CKRNNs.

Effect of CLQE. To validate the necessity of CLQE, we incorporate CLQE into the Jaccard distance. The experimental results, presented in Tab. 3, show that CLQE pro-
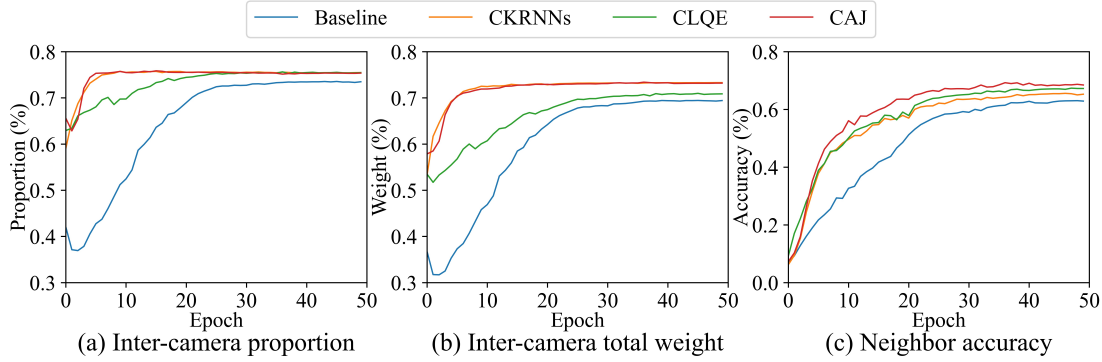
Figure 3. (a) average inter-camera proportion, (b) average inter-camera total weight and (c) average neighbor accuracy of all training samples' weighted expanded neighbors vectors over different epochs from baseline, CKRNNs, CLQE and CAJ in clustering scene.

Table 3. Ablation study on individual components in the clustering and re-ranking scenes. "CAJ" represents CA-Jaccard distance.

Method	Market1501			MSMT17		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
Clustering scene						
CC* [10]	81.0	91.1	96.2	31.1	60.2	71.3
+CKRNNs	82.4	92.9	97.1	35.1	65.3	75.8
+CLQE	83.5	92.6	97.0	40.4	70.1	81.2
+CAJ	**84.8**	**93.6**	**97.6**	**42.8**	**72.3**	**82.2**
Re-ranking scene						
BoT [25]	85.9	94.5	**98.2**	50.7	74.0	85.6
BoT+KR[51]	94.2	95.4	97.9	66.9	79.4	86.6
+CKRNNs	94.4	95.7	97.9	73.1	84.0	90.3
+CLQE	94.3	95.7	98.0	72.0	85.5	90.2
+CAJ	**94.5**	**96.2**	98.1	**74.1**	**86.2**	**90.5**

vides a significant performance improvement in both scenarios. CLQE improves the mAP and Rank-1 by 2.5% and 1.5% on Market1501, and 9.3% and 9.9% on MSMT17 respectively in the clustering scenario. Meanwhile, in the re-ranking scene, mAP/Rank-1 are improved when CLQE is applied. These results underscore the importance of CLQE in effectively mining reliable samples and increasing the weights of reliable samples.

Neighbors analysis. To further investigate the effectiveness of CKRNNs and CLQE, we plot three line charts in Fig. 3, which represent the average inter-camera proportion, average inter-camera total weight, and average neighbor accuracy of all training samples' weighted expanded neighbors vectors over different epochs from baseline, CK-RNNs, CLQE and CA-Jaccard distance in clustering scene. As shown in Fig. 3 (a) and (b), we can observe that CK-RNNs and CLQE improve the average proportion and total weight of inter-camera samples in the weighted expanded neighbors vectors. However, the combination of CKRNNs and CLQE results in a subtle difference in proportion and weight compared to using CKRNNs alone. This can be at-

tributed to the fact that most of the inter-camera samples brought by CLQE are already included in CKRNNs. Therefore, when CKRNNs and CLQE are used together, CK-RNNs mainly focus on improving the proportion and total weight of inter-camera samples in relevant neighbors, while CLQE focuses more on improving the weights of reliable samples. Moreover, Fig. 3 (c) demonstrates that the simultaneous use of CKRNNs and CLQE leads to better average neighbor accuracy compared to using either one alone. This suggests that CA-Jaccard distance maximizes the reliability of relevant neighbors and distance, resulting in performance improvement.

4.6. Parameter Analysis

In CA-Jaccard distance, four parameters are introduced, including k_1^{intra}, k_1^{inter} for CKRNNs and k_2^{intra}, k_2^{inter} for CLQE. We conduct experiments to analyze the impact of each parameter on Market1501 and MSMT17 datasets in both clustering and re-ranking scene. CC and BoT are the baselines for clustering and re-ranking scene. The mAP results are presented in Fig. 4.

Impact of k_1^{intra} and k_1^{inter} in CKRNNs. In Fig. 4 (a) and (b), we investigate the impact of k_1^{intra} and k_1^{inter}. We observe that the performance remains stable when k_1^{intra} is within the range from 1 to 20 and k_1^{inter} is within the range from 15 to 30. This is because CLQE decreases the impact of k_1^{intra} and k_1^{inter} in CKRNNs by emphasizing reliable samples in weighted expanded neighbors vectors. However, when k_1^{inter} is set to 5, there is a significant decrease in performance. Conversely, setting k_1^{intra} to 1, meaning that the intra-camera neighbors of samples only include themselves, still achieves high performance. This finding validates that inter-camera samples have more information and reliability than intra-camera samples. Moreover, we find that setting k_1^{intra} or k_1^{inter} with too large values will bring too many noise samples and hinder the performance. Therefore, considering the performance on two datasets, we set k_1^{intra} to
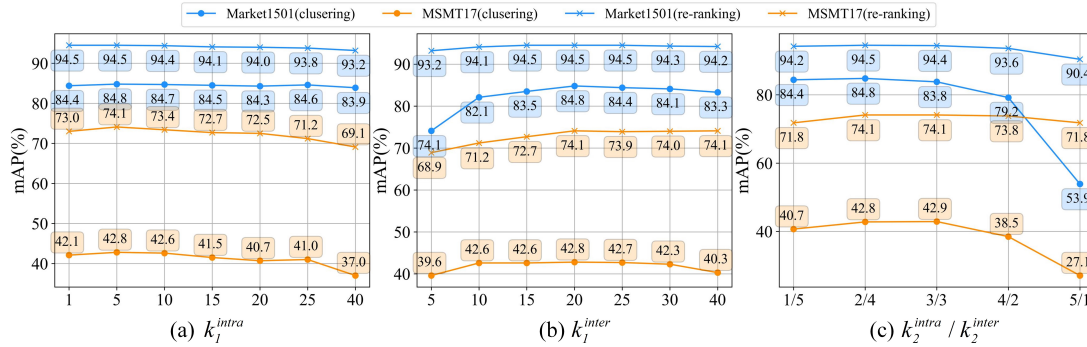
Figure 4. Parameter analysis of k_1^{intra}, k_1^{inter} and k_2^{intra}/k_2^{inter} on Market1501 and MSMT17.
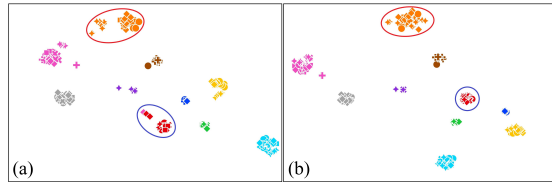

Figure 5. The t-SNE visualization of 10 persons' features extracted by the models of (a) CC and (b) CC+CAJ. Different colors and shapes indicate different identities and camera labels.

Figure 6. Ranking results of a probe produced by BoT, BoT+KR and BoT+CAJ respectively.

5 and k_1^{inter} to 20.

Impact of k_2^{intra} and k_2^{inter} in CLQE. Due to intra-camera and inter-camera k-nearest neighbors being noisy, we follow [51] and limit the sum of k_2^{intra} and k_2^{inter} to 6. In Fig. 4 (c), we vary k_2^{intra}/k_2^{inter} from 1/5 to 5/1. A smaller k_2^{intra} and a larger k_2^{inter} lead to the disregard of intra-camera information, thereby limiting the performance. Meanwhile, too large k_2^{intra} and too small k_2^{inter} weaken the mining ability of CLQE for reliable samples and lead to a decrease in performance. These experimental results lead us to set $k_2^{intra} = 2$ and $k_2^{inter} = 4$.

4.7. Visualizations

To better understand the effect of our CA-Jaccard distance, we conduct visualizations to qualitatively analyze the impact of CA-Jaccard distance.

Clustering scene. We make t-SNE visualization [31] on Market1501. As illustrated in Fig. 5, our method compacts the samples of same person from different cameras (*e.g.* red circle and blue circle in Fig. 5), indicating that our CA-Jaccard distance helps generate more accurate pseudo labels that guide the model learning camera-invariant features.

Re-ranking scene. Ranking results of BoT, BoT+KR, BoT+CAJ on Market1501 are represented in Fig. 6. Compared to KR re-ranking that uses Jaccard distance for re-ranking, CA-Jaccard distance achieves better ranking results, which indicates the superiority of our method.
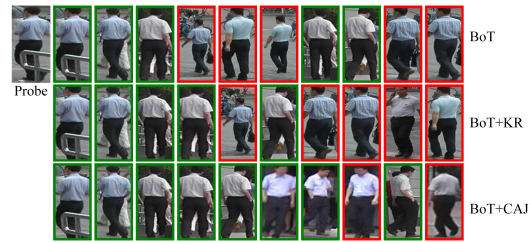
4.8. Computational Complexity Analysis

We replace the robust KRNNs and LQE in the Jaccard distance with CKRNNs and CLQE, while keeping other parts consistent. The computation of CKRNNs includes sorting and applying k-reciprocal nearest constraint. Thus the computational complexity of CKRNNs is $O(N^2 log N)$, which is comparable to that of KRNNs. However, robust KRNNs still have a recall operation which is a time-consuming set operation. The computational complexity of CLQE stays the same as LQE. In summary, CA-Jaccard distance has lower computational complexity and more reliable distance than the Jaccard distance.

5. Conclusion

In this paper, we propose a novel CA-Jaccard distance for person re-ID that overcomes camera variation and enhances the reliability of Jaccard distance through the use of CK-RNNs and CLQE. CKRNNs improve reliability by incorporating informative inter-camera positive samples while excluding intra-camera negative samples in neighbors. CLQE mines reliable samples in CKRNNs and assigns higher weights to them to further enhance the reliability. Extensive ablation studies and experiment results validate the effectiveness and robustness of our method. The low computational complexity and effectiveness of our CA-Jaccard distance make it a general distance metric for person re-ID.

References

[1] Song Bai and Xiang Bai. Sparse contextual activation for efficient visual re-ranking. *IEEE Transactions on Image Processing*, 25(3):1056–1069, 2016. 3

[2] Song Bai, Xiang Bai, and Qi Tian. Scalable person re-identification on supervised smoothed manifold. In *CVPR*, pages 2530–2539, 2017. 5

[3] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *PKDD*, pages 160–172, 2013. 1

[4] Hao Chen, Benoit Lagadec, and Francois Bremond. Ice: Inter-instance contrastive encoding for unsupervised person re-identification. In *ICCV*, pages 14960–14969, 2021. 5, 6

[5] Hao Chen, Yaohui Wang, Benoit Lagadec, Antitza Dantcheva, and Francois Bremond. Joint generative and contrastive learning for unsupervised person re-identification. In *CVPR*, pages 2004–2013, 2021. 6

[6] Tianlong Chen, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren, and Zhangyang Wang. Abd-net: Attentive but diverse person re-identification. In *ICCV*, pages 8351–8361, 2019. 1

[7] Yiyu Chen, Zheyi Fan, Shuni Chen, and Yixuan Zhu. Improving pseudo-labeling with reliable inter-camera distance encouragement for unsupervised person re-identification. *Science China Information Sciences*, 66(5):1–12, 2023. 2, 6

[8] Yoonki Cho, Woo Jae Kim, Seunghoon Hong, and Sung-Eui Yoon. Part-based pseudo label refinement for unsupervised person re-identification. In *CVPR*, pages 7308–7318, 2022. 1, 2, 5, 6

[9] Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, pages 1–8, 2007. 3

[10] Zuozhuo Dai, Guangyuan Wang, Weihao Yuan, Siyu Zhu, and Ping Tan. Cluster contrast for unsupervised person re-identification. In *ACCV*, pages 1142–1160, 2022. 1, 2, 5, 6, 7

[11] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996. 1, 2

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

[13] Yang Fu, Yunchao Wei, Guanshuo Wang, Yuqian Zhou, Honghui Shi, and Thomas S. Huang. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In *ICCV*, 2019. 1, 2

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

[15] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In *NeurIPS*, 2020. 1, 2, 6

[16] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *PETS*, pages 1–7, 2007. 5

[17] Tao He, Leqi Shen, Yuchen Guo, Guiguang Ding, and Zhenhua Guo. Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification. In *AAAI*, pages 879–887, 2022. 1, 2, 6

[18] Herve Jegou, Hedi Harzallah, and Cordelia Schmid. A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*, pages 1–8, 2007. 3

[19] Zilong Ji, Xiaolong Zou, Xiaohan Lin, Xiao Liu, Tiejun Huang, and Si Wu. An attention-driven two-stage clustering method for unsupervised person re-identification. In *ECCV*, pages 20–36, 2020. 2

[20] Zongyi Li, Yuxuan Shi, Hefei Ling, Jiazhong Chen, Qian Wang, and Fengfan Zhou. Reliability exploration with self-ensemble learning for domain adaptive person re-identification. In *AAAI*, pages 1527–1535, 2022. 6

[21] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, pages 8738–8745, 2019. 2

[22] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *ECCV*, pages 869–884, 2016. 5

[23] Stuart Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, pages 129–137, 1982. 1

[24] Chuanchen Luo, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Spectral feature transformation for person re-identification. In *ICCV*, pages 4976–4985, 2019. 1

[25] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *CVPR Workshops*, pages 0–0, 2019. 1, 6, 7

[26] Danfeng Qin, Stephan Gammeter, Lukas Bossard, Till Quack, and Luc Van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR*, pages 777–784, 2011. 3

[27] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1655–1668, 2018. 6

[28] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *CVPR*, pages 420–429, 2018. 1, 2, 3, 6

[29] Xiaohui Shen, Zhe Lin, Jonathan Brandt, Shai Avidan, and Ying Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *CVPR*, pages 3013–3020, 2012. 2

[30] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, pages 480–496, 2018. 1

[31] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9 (11), 2008. 8

[32] Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *CVPR*, pages 10981–10990, 2020. 2, 6

[33] Haochen Wang, Jiayi Shen, Yongtuo Liu, Yan Gao, and Efstratios Gavves. Nformer: Robust person re-identification with neighbor transformer. In *CVPR*, pages 7297–7307, 2022. 1

[34] Menglin Wang, Baisheng Lai, Jianqiang Huang, Xiaojin Gong, and Xian-Sheng Hua. Camera-aware proxies for unsupervised person re-identification. In *AAAI*, pages 2764–2772, 2021. 5, 6

[35] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *CVPR*, pages 79–88, 2018. 5

[36] Yuhang Wu, Tengteng Huang, Haotian Yao, Chi Zhang, Yuanjie Shao, Chuchu Han, Changxin Gao, and Nong Sang. Multi-centroid representation network for domain adaptive person re-id. In *AAAI*, pages 2750–2758, 2022. 6

[37] Shiyu Xuan and Shiliang Zhang. Intra-inter camera similarity for unsupervised person re-identification. In *CVPR*, pages 11926–11935, 2021. 6

[38] Fengxiang Yang, Ke Li, Zhun Zhong, Zhiming Luo, Xing Sun, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Shaozi Li. Asymmetric co-teaching for unsupervised cross-domain person re-identification. In *AAAI*, pages 12597–12604, 2020. 1, 2

[39] Hong-Xing Yu, Wei-Shi Zheng, Ancong Wu, Xiaowei Guo, Shaogang Gong, and Jian-Huang Lai. Unsupervised person re-identification by soft multilabel learning. In *CVPR*, pages 2148–2157, 2019. 2

[40] Rui Yu, Zhichao Zhou, Song Bai, and Xiang Bai. Divide and fuse: A re-ranking approach for person re-identification. In *BMVC*, 2017. 2

[41] Kaiwei Zeng, Munan Ning, Yaohua Wang, and Yang Guo. Hierarchical clustering with hard-batch triplet loss for person re-identification. In *CVPR*, pages 13657–13665, 2020. 2, 6

[42] Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *CVPR*, pages 9021–9030, 2020. 2

[43] Yunpeng Zhai, Qixiang Ye, Shijian Lu, Mengxi Jia, Rongrong Ji, and Yonghong Tian. Multiple expert brainstorming for domain adaptive person re-identification. In *ECCV*, pages 594–611, 2020. 1, 2

[44] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *ICCV*, pages 8222–8231, 2019. 1, 2

[45] Xiao Zhang, Yixiao Ge, Yu Qiao, and Hongsheng Li. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *CVPR*, pages 3436–3445, 2021. 1, 2, 6

[46] Xinyu Zhang, Dongdong Li, Zhigang Wang, Jian Wang, Errui Ding, Javen Qinfeng Shi, Zhaoxiang Zhang, and Jing-dong Wang. Implicit sample extension for unsupervised person re-identification. In *CVPR*, pages 7369–7378, 2022. 1, 6

[47] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *ECCV*, pages 526–544, 2020. 2

[48] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015. 5

[49] Yi Zheng, Shixiang Tang, Guolong Teng, Yixiao Ge, Kaijian Liu, Jing Qin, Donglian Qi, and Dapeng Chen. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In *ICCV*, pages 8371–8381, 2021. 6

[50] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *CVPR*, pages 2138–2147, 2019. 1

[51] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, pages 1318–1327, 2017. 1, 2, 3, 6, 7, 8

[52] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *CVPR*, pages 598–607, 2019. 2