



Unsupervised person re-identification based on high-quality pseudo labels

Yanfeng Li¹ · Xiaodi Zhu¹ · Jia Sun¹ · Houjin Chen¹ · Zhiyuan Li¹

Accepted: 15 October 2022 / Published online: 10 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The unsupervised domain adaptive (UDA) person re-identification (re-ID) method is of great significance to promote the practical application of person re-ID. However, the noisy pseudo labels in the target domain hinder its performance. In this paper, a novel high-quality pseudo labels (HQP) method for UDA person re-ID is proposed, which improves the performance from the perspectives of sample feature expression and similarity measurement in the clustering. In order to obtain better feature representation for target domain samples, a source domain generalization method based on contrastive learning (SCL) is designed. SCL learns the inherently consistent information within a sample, thereby improving the expression ability of the source domain pre-trained model. In order to provide a more reasonable similarity measurement for the clustering method, a soft label similarity based on neighborhood information integration (NII) is designed, which aids the clustering method to generate reliable pseudo labels. Market-1501, DukeMTMC-ReID and MSMT17 datasets are employed to evaluate the performance of the proposed HQP method. It achieves the results of 80.3%/92.3%, 68.0%/82.6% and 25.4/53.3 mAP/Rank-1 on DukeMTMC-ReID-to-Market-1501, Market-1501-to-DukeMTMC-ReID and DukeMTMC-ReID-to-MSMT17 tasks. Experimental results demonstrate that our HQP method performs favorably against the state-of-the-art UDA person re-ID methods.

Keywords Person re-identification · Unsupervised domain adaptive · Contrastive learning · Clustering method

1 Introduction

The task of person re-identification (re-ID) is to match the target persons in the multi-camera surveillance networks with the computer vision technology, so as to realize the trajectory tracking of a specific person. With the development of deep learning methods and the emergence of large-scale datasets, supervised person re-ID methods [1, 2] have achieved great breakthroughs on public datasets. However, when a large amount of monitoring data is collected in the surveillance network, manually labelling person identity takes a lot of manpower. Unsupervised person re-ID [3–28] can use the unlabeled data, which is of great significance to promote the practical application. In the field of unsupervised person re-ID, the main focus is currently on the unsupervised domain adaptive (UDA) person re-ID method, which adapts the model pre-trained with the labeled source domain to the unlabeled target domain. Among the UDA methods, pseudo

label-based method has attracted widespread attention due to its high performance and stability. This type of method extracts the features of the target domain images using the source domain pre-trained model, and then performs the clustering method to generate pseudo labels for model retraining in the target domain. However, due to the domain difference and the pseudo label noise, the performance of unsupervised person re-ID method is far from practical application.

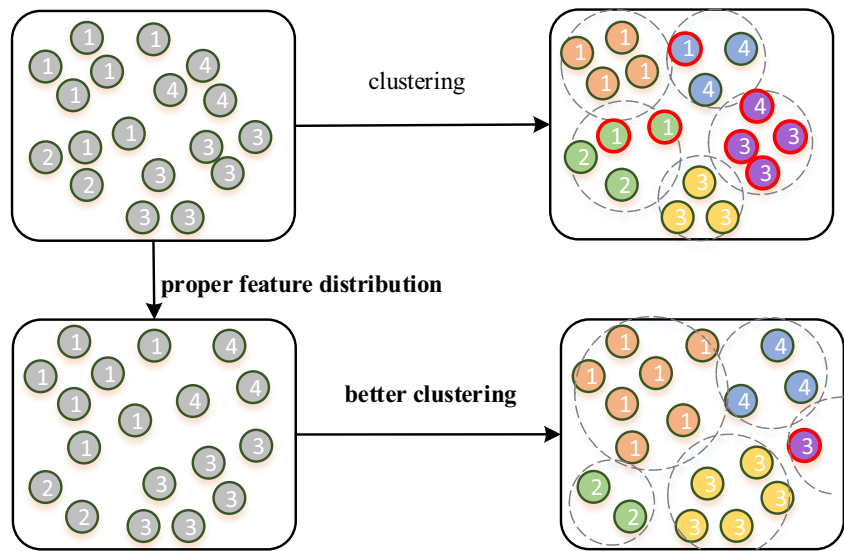
Existing UDA person re-ID methods usually design effective training strategy to suppress the impact of pseudo label noise, thus improving the recognition performance. However, distinguishing the correct pseudo label from the incorrect pseudo label is difficult, and the pseudo label noise will interfere with the model learning for feature expression. Therefore, improving the quality of pseudo labels before model training is more advantageous than suppressing the impact of pseudo label noise. According to the generation process, the accuracy of features for target domain extracted by the pre-trained model and the validity of feature similarity measurement are two key factors to improve the clustering quality. These two factors can be interpreted as proper feature distribution and better clustering method, shown in Fig. 1.

To address the limitation of existing pseudo label-based methods, a novel UDA person re-ID method based on high-

✉ Yanfeng Li
yf.li@bjtu.edu.cn

¹ School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Fig. 1 Pseudo label generation. The numbers in the circles represent the correct pedestrian IDs. The same color represents the same pseudo label, and the circles marked in red are with the wrong pseudo labels



quality pseudo labels (HQP) is proposed, which improves the re-ID performance from the perspective of boosting the quality of pseudo labels in the clustering procedure. The source domain pre-trained model is employed to extract features of the target domain, which are the inputs for the clustering method. Due to the difference between the source domain and the target domain, the extracted feature distribution is not close to its true distribution. Aiming at this problem, a source domain generalization method based on contrastive learning (SCL) is designed, which enhances the feature representation ability of the source domain pre-trained model. The feature similarity measurement is also important to the clustering quality. Just using the distance between samples to measure their relationships may lead to large error due to the complex background, different poses, and occlusion in the target domain. To provide a more reasonable similarity measurement for the clustering method, a soft label similarity based on the neighborhood information integration (NII) is designed. NII combines the similarity between samples, the shared neighbors between samples, and the similarity between neighbors.

The contributions of this paper can be summarized as follows: (1) From the perspective of improving the quality of pseudo labels, a high-quality pseudo label (HQP) method is proposed for UDA person re-ID. (2) A source domain generalization method based on contrastive learning is designed. Different augmentations are applied for each sample in the classification loss and the triplet loss, which improves the feature invariance of the source domain pre-trained model to the visual changes of the same person. (3) A soft label similarity based on the neighborhood information integration (NII) is designed to guide the clustering method. Existing neighborhood methods normally consider the similarity between the image and its neighbors. Our NII considers the shared neighbors between samples and the similarity between these neighbors.

2 Related work

To improve the re-ID performance, the pseudo label-based method studies two main problems: difference between the source domain and the target domain and pseudo label noise in the target domain. One type is sample screening method. In order to reduce the interference of pseudo label noise, the progressive unsupervised learning (PUL) method [18] selected the reliable samples for model training according to the distance to the cluster center. Similar to PUL, the progressive unsupervised co-learning (PUCL) method [19] employed two source domain pre-trained models and performed sample screening. However, simply discarding the suspected noisy samples may lead to information deficiency of the training samples, which may in reverse hinder the model training. Different from this idea, the asymmetric co-teaching (ACT) method [20] designed an asymmetric cooperative network. One model was employed to receive the possibly cleanest samples and the other model was trained on outlier samples to preserve the sample diversity. Ge et al. [21] proposed a self-paced contrastive learning framework (SpCL), including an image feature encoder and a hybrid memory model. Based on the idea of self-paced learning, this method started with the most reliable samples, and then the training samples were gradually increased. As SpCL screened noisy samples and employed outliers for training, the impact of pseudo label noise was reduced and the diversity of training samples can also be reserved. Similarly, Li et al. [22] introduced a multi-label learning guided self-paced clustering (MLC) method, which learned the discriminative features with three crucial modules, and removed some noisy samples through self-paced clustering.

In addition to sample screening method, pseudo label noise suppression method based on cross-camera problem is also

effective. The augmented discriminative clustering (AD-Cluster) method [23] expanded the sample data for each camera style, increasing the sample diversity while learning camera invariant features. To suppress feature changes caused by pseudo label noise and camera offset, Yang et al. [24] proposed a dynamic and symmetric cross-entropy loss and a camera-aware meta-learning algorithm. Dynamic and symmetric cross-entropy loss mitigated the negative effect of noisy samples and adapted to the changes in clusters after each clustering step. The camera-aware meta-learning algorithm split the training data into meta-training and meta-testing based on the camera ID to simulate cross-camera constraints, and forced the model to learn camera-invariant features through the interactive gradients of meta-training and meta-testing. The camera penalty learning (CPL) method [25] improved the UDA re-ID performance from the camera-ID penalty strategy. A camera penalty-based triplet loss (PTL) was designed, which reduced the sample distance imbalance caused by cross-camera problem. A camera-penalty-neighborhood loss (PNL) was combined with the push loss (PL), which could reduce the dependence on pseudo labels.

The cross-camera problem is just one factor to cause pseudo label noise. Using multiple networks co-training can suppress pseudo label noise more comprehensively. The simultaneous mean teaching framework (MMT) [26] combined the hard pseudo labels with the soft pseudo labels for joint training. Hard pseudo labels were generated by the clustering algorithm and updated before each training epoch. Soft pseudo labels were generated by the co-trained networks and optimized online. MMT exploited the outputs of the two networks to mitigate pseudo labels noise. Similar to MMT, Zhao et al. [27] put forward a noise-resistant reciprocal training method (NRMT), which maintained two networks simultaneously during training and allowed them to share aggregation through cooperative clustering in each iteration. Zhai et al. proposed the multiple expert brainstorming network (MEB-Net) method [28], which utilized multiple networks with different structures for model pre-train in the source domain. The feature of the target domain sample was obtained as the average feature of the multiple networks, which was employed to produce the pseudo labels. Zhu et al. [29] proposed a learning with noisy labels (LNL) method, which promoted the model training from noise correction and noise resistance. Through the closed-loop learning mechanism, the triplet ensemble student-teacher (TEST) model [30] relaxed the constraints between the teacher network and the student network, and enhanced the expression ability of the student network. Furthermore, knowledge exchange between student networks could better handle noisy labels and avoid coupling.

Different from suppressing the impact of pseudo label noise, other methods tend to improve the pseudo label quality. High-quality pseudo labels can enable the model to learn a better representation for the target domain, thereby improving the

recognition performance. The Dual-Refinement method [31] performed the K-means clustering algorithm to re-cluster each class of the initial cluster, and then the cluster centers of the subclasses were used to refine the pseudo labels. Li et al. [32] proposed an iterative intra-domain consistency enhancement (ICE) method based on the mean teacher framework to fully mine the two underlying consistency constraints on multi-granularity features. The impact of noisy pseudo labels was reduced through the joint action of the instance-ensembling consistency constraint and the cross-granularity consistency constraint. Chen et al. [33] incorporated a generative adversarial network (GAN) and a contrastive learning module into one joint training framework. The newly generated views could provide more reference for the network and improve the quality of generated pseudo labels.

3 Proposed method

In this paper, a novel high-quality pseudo labels (HQP) method is proposed for UDA person re-ID task. Different from suppressing the impact of pseudo label noise in model training, we directly improve the quality of pseudo labels from the perspectives of proper feature expression and reliable similarity measurement in clustering.

3.1 Framework overview

The overall framework of our HQP method is described in Fig. 2. IBN-ResNet50 [34] pre-trained on ImageNet [35] is selected as the backbone. We first train the model with a supervised manner using the labeled training data in the source domain. In order to enhance the feature representation ability of the pre-trained model, a source domain generalization method based on contrastive learning (SCL) is designed. In the fine-tuning stage of the target domain, the pseudo labels are first generated by the clustering method. In order to improve the clustering quality, a soft label similarity based on the neighborhood information integration (NII) is designed for the clustering method.

3.2 Source domain generalization method based on contrastive learning

The source domain pre-trained model is employed to extract the feature representation of the target domain, which will be input to the clustering method for pseudo label generation. The consistency between the extracted feature distribution and its true distribution is directly related to the clustering quality. Therefore, enhancing the feature representation generalization of the pre-trained model is important. Contrastive learning can learn the invariance of the image by applying different augmentations on the sample. Accordingly, a source domain generalization method based on contrastive learning (SCL) is designed.

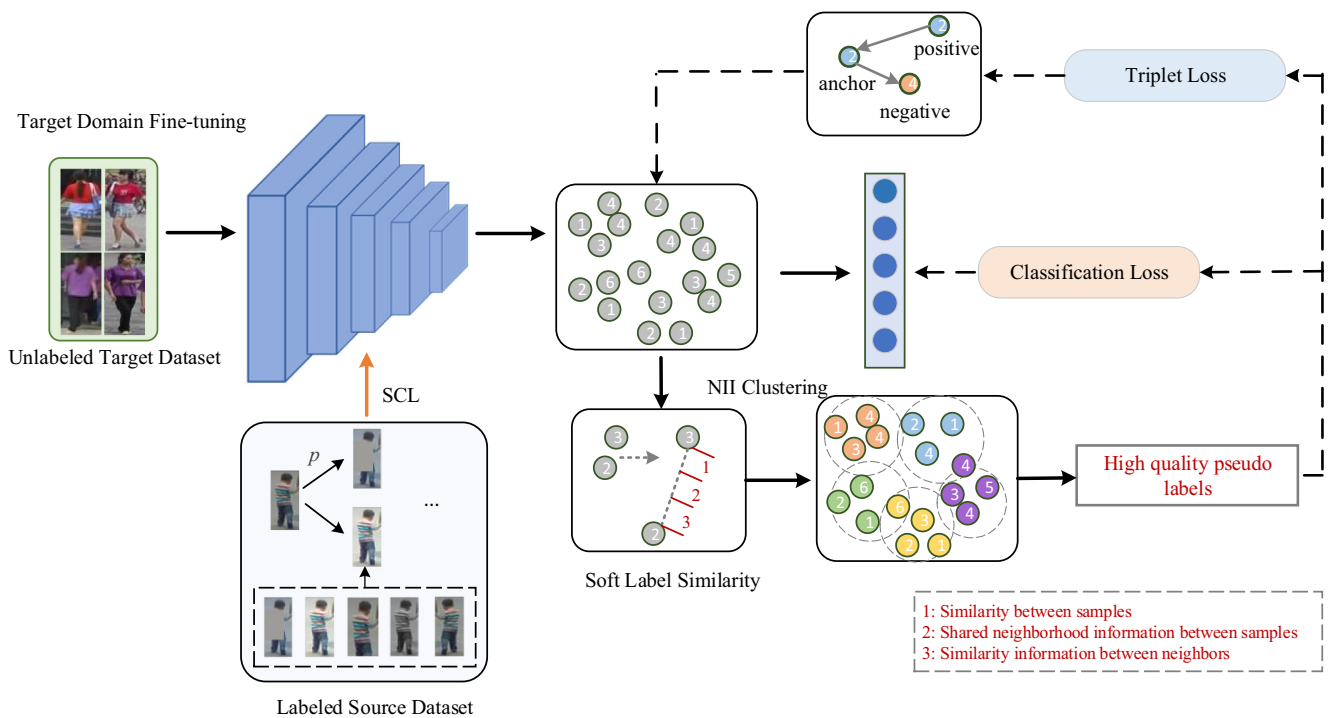


Fig. 2 The overall framework of the HQP method. The number in the circle represents the camera ID, and the circle of the same color represents the same pseudo label

In SCL, two different image augmentation operations are applied for each sample. In order to preserve the original characteristic of the image, image occlusion or flip operation with a probability of p is performed as the first augmentation. In order to learn the feature invariance of the pedestrian, the second augmentation includes background interference, posture change, color change, occlusion and scale change. To increase the randomness, one transformation form is randomly selected as the second augmentation for each sample. The pre-training process of the source domain is shown in Fig. 3. The training loss is the combination of the classification loss and the triplet loss.

In the training procedure, two augmented images are generated for each sample. Thus, each batch contains $2N_b$ images. The labels of the augmented images are their corresponding original labels, shown in Fig. 4a. The cross entropy loss is used as the classification loss, shown below:

$$L_{cls-g}^s = -\frac{1}{2N_b} \sum_{i=1}^{2N_b} \log p(y_i^s | x_i^s) \tag{1}$$

where N_b and i represent the number of images and image index in a batch, respectively, and $p(y_i^s | x_i^s)$ is the probability that image x_i^s belongs to y_i^s . s is the abbreviation for source domain.

For triplet loss, the augmented samples in each batch participate in the selection of the hardest positive sample and the hardest negative sample, shown in Fig. 4b. It can be seen that with SCL, the hardest positive samples and the hardest negative samples selected in the triplet loss are more effective than before. The expression for triplet loss is as follows:

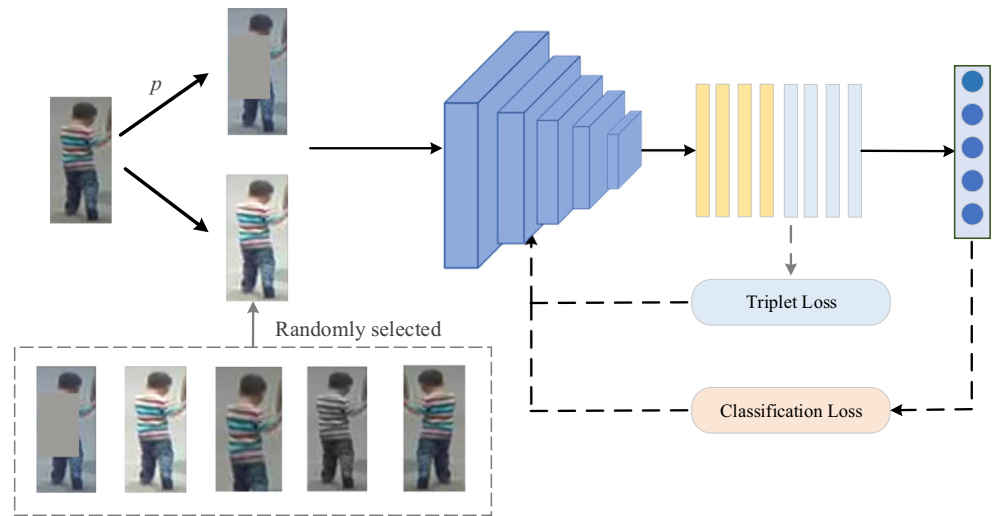
$$L_{tri-g}^s = \sum_{i=1}^{2N_b} [m + \|f(x_i^s) - f(x_{i+}^s)\|_2 - \|f(x_i^s) - f(x_{i-}^s)\|_2] \tag{2}$$

where x_{i+}^s represents the farthest positive sample and x_{i-}^s represents the nearest negative sample for anchor x_i^s . $f(\cdot)$ is the feature extracted by the model, and m represents the margin parameter.

3.3 Clustering method based on soft label similarity

We use the DBSCAN clustering method to generate pseudo labels for the target domain. Due to domain difference between source domain and target domain, the complex background, pedestrian pose, and occlusion in the target domain, measuring the relationship between samples only with the samples' distance may be not reliable. As shown in Fig. 5 (the red rectangle represents the wrong sample), if the cosine similarity between samples is employed, sample B is similar with sample A, leading to pseudo label noise. From Fig. 5, we can also find that the neighbors of sample A and that of sample B are different. Therefore, measuring sample similarity by integrating the relationship between sample itself and the relationship between their neighbors is more reasonable. According to the above analysis, in order to provide a more reasonable similarity measurement for the clustering method, a soft label similarity based on neighborhood information integration (NII) is developed. NII combines the similarity between samples, the shared neighbors between samples, and the similarity information between neighbors. The calculation process for NII is shown in Fig. 6.

Fig. 3 Source domain generalization method based on contrastive learning



First, the cosine similarity matrix of the samples in the target domain is calculated. Let $F \in \mathbb{R}^{N \times l}$ denote the L2 normalized feature matrix of all samples in the target domain, where N is the data size of the target domain, l indicates that the data comes from the target domain, and l indicates the feature length. Then the cosine similarity matrix of the target domain data M can be expressed as:

$$M = F \cdot F^T \tag{3}$$

where $M(i, j)$ represents the cosine similarity between the i -th sample and the j -th sample, and ‘ \cdot ’ represents matrix multiplication.

Then, each row of matrix M is sorted from small to large, generating the sorted similarity matrix M_s . For each normal

(non-outlier) sample, we select k reference neighbors according to the sorting result. Compared with normal samples, outlier samples may be less similar to other samples and require more neighborhood information. Therefore, in order to assist the outlier samples in finding the correct category, $k + n$ neighbors are selected for outlier samples. For each sample in M , the similarities of the selected neighbors are kept, and the similarities of other samples are set to 0, resulting in the neighborhood representation matrix M_{nei} . Thus, the neighborhood representation matrix M_{nei} can be expressed as:

$$M_{nei}(i, j) = \begin{cases} M(i, j), & j \in R_i^k \text{ and } i \in C \\ M(i, j), & j \in R_i^{k+n} \text{ and } i \in O \\ 0, & \text{others} \end{cases} \tag{4}$$

Fig. 4 Loss function illustration. Image with the dotted box represents the augmented sample, and image with the solid box represents the original sample. x_i^s and $x_i^{s'}$ represent two augmented images of each sample. The green arrow represents the hardest positive sample, and the blue arrow represents the hardest negative sample. The dashed line represents the selection result before adding the augmented samples. The solid line represents the selection result after

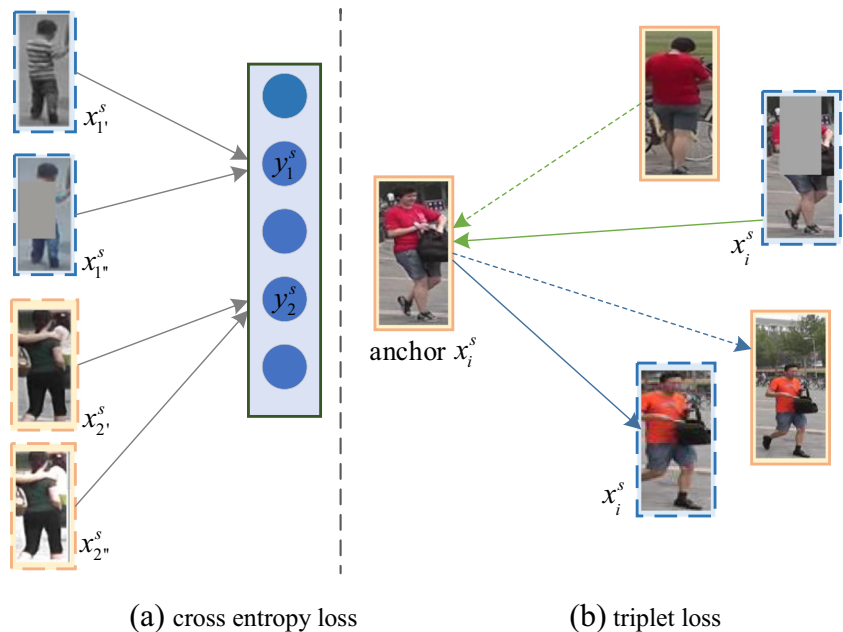


Fig. 5 Sample relationship with cosine similarity. The red rectangle represents the wrong sample



where $j \in R_i^k$ indicates that j is in the k neighborhood for sample i , and $i \in C$ indicates that i is a non-outlier sample in the cluster. Similarly, $j \in R_i^{k+n}$ indicates that j is in $k + n$ neighborhood for sample i , and $i \in O$ indicates that i is an outlier in the cluster.

By multiplying M_{nei} with M_{nei}^T , the soft label similarity R based on neighborhood information integration can be obtained:

$$R = M_{nei} \cdot M_{nei}^T \tag{5}$$

The soft label similarity R considers the similarity between two samples and the similarity between their neighbors. In order to verify its effectiveness, the neighborhood samples calculated by the cosine similarity and that calculated by the proposed NII (R) are compared. As shown in Fig. 7, there are two incorrect neighborhood samples for sample A with the cosine similarity, and just one incorrect neighborhood sample with NII. Since NII is more reliable, DBSCAN clustering method using NII can improve the accuracy of pseudo labels.

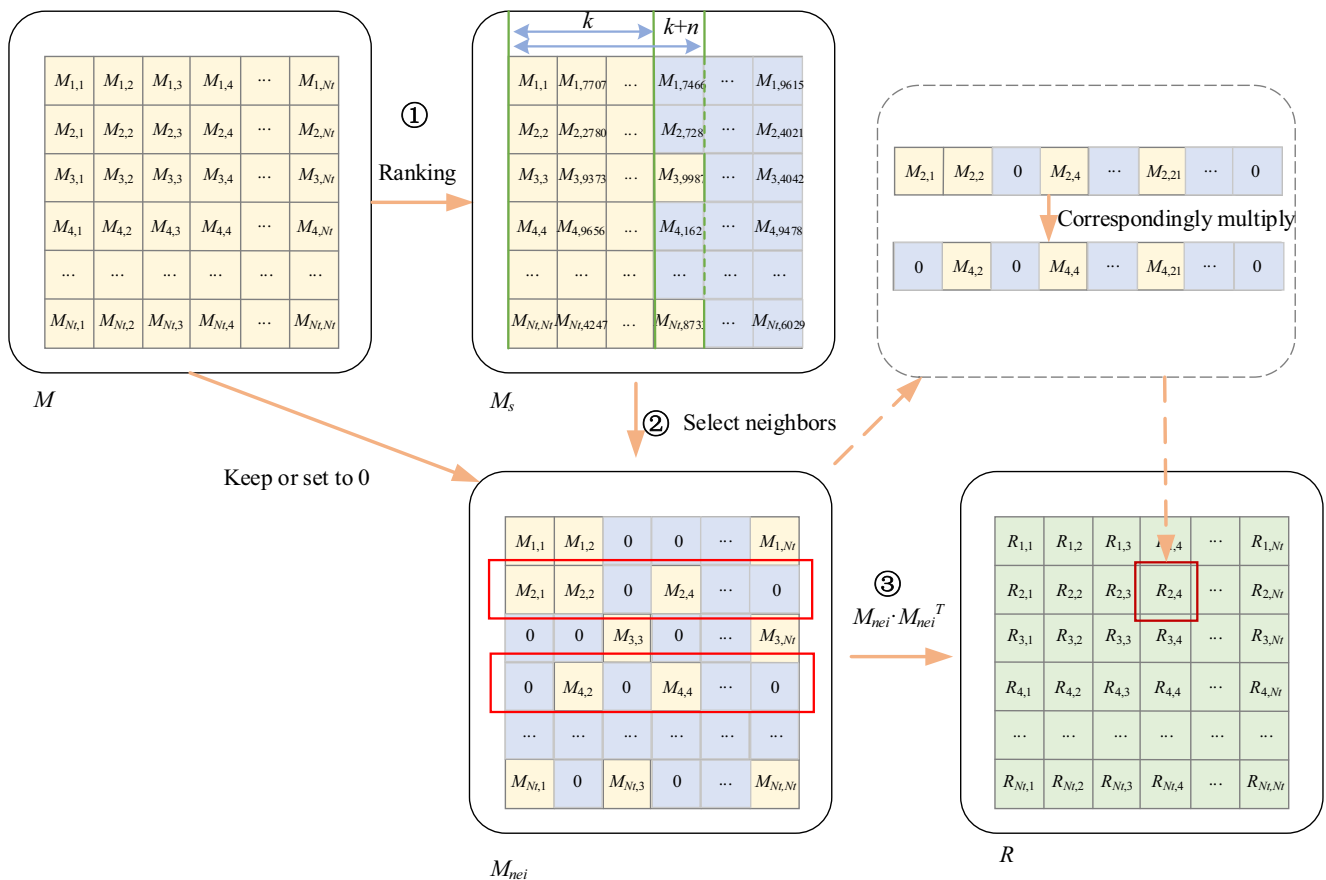


Fig. 6 Soft label similarity based on neighborhood information integration

3.3.1 Overall algorithm

First, the source domain generalization method based on contrastive learning (SCL) is performed to obtain the pre-trained model. The model is trained using a combination of the cross entropy loss L_{cls-g}^s and the triplet loss L_{tri-g}^s computed among the augmented images:

$$L^s = L_{cls-g}^s + L_{tri-g}^s \quad (6)$$

Then, features of the target domain images are extracted using the pre-trained model. Afterwards, the DBSCAN clus-

tering method with NII is performed on the extracted features, generating the pseudo labels for the target domain. Finally the model is fine-tuned through the images and the pseudo labels in the target domain. We alternate the pseudo labels generation and the model fine-tuning. The loss for fine-tuning is the combination the cross entropy loss L_{cls}^t and the triplet loss L_{tri}^t in the target domain:

$$L^t = L_{cls}^t + L_{tri}^t \quad (7)$$

The detailed optimization procedure is summarized in Algorithm 1.

Algorithm 1: Training strategy of high-quality pseudo labels method (HQP)

Input: Source domain data D^s and target domain data D^t .

Iteration number T and epoch number $(num_epoch^s, num_epoch^t)$. The parameters of k and n in NII in Section 3.3, and the sample number threshold $MinPts$ in the DBSCAN clustering algorithm.

Output: Best model θ .

```

1  for  $n$  in  $[1, num\_epoch^s]$  do
2      for each mini-batch, iteration  $T$  do
3          Perform the first image augmentation with probability  $p$  on the mini-batch samples in  $D^s$ 
4          Randomly select a form for the second image augmentation
5          Calculate  $L_{cls-g}^s$  by Eq. (1).
6          Calculate  $L_{tri-g}^s$  by Eq. (2).
7          Update the model  $\theta$  through the total loss  $L^s$  in Eq. (6).
8      end for
9  end for
10 for  $n$  in  $[1, num\_epoch^t]$  do
11     The soft label similarity matrix  $R$  is calculated by Eq. (5), where the number of neighborhood information for all
        samples is  $k$ .
12     The DBSCAN clustering algorithm based on  $R$  is performed to find the outlier samples.
13     The soft label similarity matrix  $R$  is calculated again by Eq. (5), where the number of neighborhood information for
        non-outlier samples is  $k$ , and the number for outlier samples is  $k+n$ .
14     The DBSCAN clustering algorithm based on  $R$  is performed, generating the pseudo labels  $y_i^t$  for sample  $x_i^t$  in  $D^t$ .
15     for each mini-batch, iteration  $T$  do
16         Update the model  $\theta$  through the total loss  $L^t$  in Eq. (7).
17     end for
18 end for

```

Fig. 7 Comparison of cosine similarity and NII in finding neighbors. The red border represents the incorrect sample



4 Experiments

4.1 Datasets and evaluation protocol

We evaluate the proposed HQP method on three widely-used person re-ID datasets, Market-1501 [36], DukeMTMC-ReID [37] and MSMT17 [6]. Mean average precision (mAP) and cumulative matching characteristic (CMC) [38] are adopted as the evaluation metrics.

Market-1501: This dataset was released in 2015, captured with five high-resolution cameras and one low-resolution camera. In this dataset, each pedestrian is captured by at least two cameras, and each camera is allowed to capture multiple images for one pedestrian. The dataset contains 32,217 fixed-size pedestrian images for 1,501 pedestrian IDs, with an average of 21.46 pedestrian images per ID. The training set includes 12,936 images with 751 pedestrian IDs. For the test set, 3,368 images with 750 pedestrian IDs constitute the query set, and 19,936 images with 750 pedestrian IDs constitute the gallery set. The query image is obtained using the human-annotated boxes, while the gallery images are obtained using the boxes generated by the deformable part model (DPM) [39]. All the set is the same with existing methods. Image examples for this dataset are shown in Fig. 8a.

DukeMTMC-ReID: This dataset was released in 2017. It is a multi-target and multi-camera pedestrian tracking dataset, which is a subset of the DukeMTMC dataset. It is captured with 8 cameras, which contains 36,411 images of 1,812 pedestrian IDs. The training set contains 16,522 images with 702 pedestrian IDs. For the test set, 2,228 images with 702 pedestrian IDs constitute the query set, and 17,661 images with 1,110 pedestrian IDs constitute the gallery set. Image examples for this dataset are shown in Fig. 8b.

MSMT17: It is captured by 15 cameras, which has 4,101 identities with 126,441 images. 32,621 images of 1041 identities are in the training set, and 93,820 images of 3060 identities are in the test set. The query set includes 11,659 images which are randomly sampled, and the rest 82,161 images form the gallery set.

4.2 Implementation details

All the images are resized to 256×128 , and IBN-ResNet50 pre-trained on ImageNet is selected as the backbone. For pre-training on the source domain and fine-tuning on the target domain, the batch size N_b is set to 64. That is 16 pedestrian identities are selected in each batch using ground-truth labels for the source domain and pseudo labels for the target domain.

Source domain pre-training 80 epochs are trained with an initial learning rate of 0.00035. At the 40th and 70th epochs, the learning rate is reduced to 1/10 of the previous epoch.

Target domain fine-tuning 40 epochs are trained with the learning rate of 0.00035. We use the DBSCAN clustering method to generate pseudo labels. The sample number threshold $MinPts$ controls the number of minimum samples in the range of neighborhood distance threshold in DBSCAN clustering method. As DukeMTMC-ReID dataset is more complex than the Market dataset, a larger value for DukeMTMC-ReID is suggested. When Market-1501 is used as the target domain, the sample number threshold $MinPts$ is set to 9. When DukeMTMC-ReID is used as the target domain, the sample number threshold $MinPts$ is set to 12. The neighborhood distance threshold ϵ is calculated according to the similarity matrix.



Fig. 8 Examples of person re-ID dataset

4.3 Ablation studies

In this section, ablation experiments are conducted on Market-1501 and DukeMTMC-ReID datasets to evaluate the effectiveness of each module in the proposed HQP method. The experimental results are shown in Table 1. ‘DukeMTMC-ReID->Market-1501’ indicates that the DukeMTMC-ReID dataset is the source domain, and the Market-1501 dataset is the target domain, and vice versa. ‘Direct Transfer’ represents the result of directly applying the source domain pre-trained model to the target domain. ‘Direct Transfer + SCL’ represents adding the proposed source domain generalization method-SCL to pre-train the source domain, and then directly applying the pre-trained model to the target domain. ‘Baseline’ represents the traditional source domain pre-trained model and fine-tuning with the DBSCAN clustering method using cosine similarity. ‘Baseline + SCL’ represents adding SCL to pre-train the source domain and then fine-tuning using the cosine similarity. ‘Baseline + NII’ represents performing the DBSCAN clustering method using the proposed similarity-NII. ‘HQP’ denotes the proposed method, which uses SCL for model pre-training on the source domain, and DBSCAN clustering with NII.

The effectiveness of source domain generalization method based on contrastive learning (SCL) In order to enhance the feature expression ability of the pre-trained model, SCL is designed to learn the invariance of image by applying different augmentations on the sample. We compare the results of directly applying the source domain pre-trained model to the

target domain without and with SCL. Compared with ‘Direct Transfer’, the mAP/Rank-1 performance of ‘Direct Transfer + SCL’ is improved by 3.6%/3.5% and 1.8%/3.7% on Market-1501 and DukeMTMC-ReID datasets. This may show that when SCL is added, the ability of the pre-trained model adapting to unseen samples becomes stronger. Besides, the results of fine-tuned model on the target domain without and with SCL are also compared. Compared with ‘Baseline’, the mAP/Rank-1 performance of ‘Baseline + SCL’ on the Market-1501 dataset is improved by 2.9%/0.7%. On DukeMTMC-ReID dataset, the mAP/Rank-1 performance is improved by 1.2%/0.7%. These experimental results show that SCL can effectively improve the model’s feature expression ability, thereby improving the UCD re-ID performance.

The effectiveness of soft label similarity based on neighborhood information integration (NII) In order to provide a more accurate similarity measurement for the clustering method, NII is designed. The results of fine-tuned model on the target domain with cosine similarity and that with NII are compared. Compared with ‘Baseline’, ‘Baseline + NII’ improves mAP/Rank-1 by 2.4%/0.3% on Market-1501 dataset. On DukeMTMC-ReID dataset, mAP/Rank-1 is improved by 4.3%/2.7%. This result show that, compared with just considering the similarity between samples, combining the similarity between samples, the shared neighborhood information between samples, and the similarity information between neighbors is more credible in providing measurement for DBSCAN clustering, and improves the quality of pseudo labels.

Table 1 Ablation study of the proposed high quality pseudo label (HQP) method

Methods	DukeMTMC-ReID→Market-1501				Market-1501→DukeMTMC-ReID			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
Direct Transfer	32.8	62.4	77.1	82.4	33.2	51.3	65.4	70.6
Direct Transfer+SCL	36.4	65.9	80.6	85.3	35.0	55.0	68.1	73.8
Baseline	75.6	89.8	95.4	97.3	62.8	79.1	88.2	90.4
Baseline+SCL	78.5	90.5	96.3	97.5	64.0	79.8	88.8	91.4
Baseline+NII	78.0	90.1	96.2	97.6	66.9	82.0	89.7	92.3
HQP	80.3	92.3	96.9	97.9	68.0	82.6	90.2	92.4

Table 2 Comparison of the proposed HQP method with the state-of-the-art unsupervised person re-ID methods

Methods		DukeMTMC-ReID→Market-1501				Market-1501→DukeMTMC-ReID			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
SSM	PUL [18]	20.1	44.7	59.1	65.6	16.4	30.4	44.5	50.7
	PUCL [19]	22.9	48.5	63.8	69.9	19.5	33.1	46.9	53.0
	ACT [20]	60.6	80.5	-	-	54.5	72.4	-	-
	SpCL [21]	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5
	MLC [22]	65.9	85.6	93.9	96.0	55.0	74.1	83.8	86.3
NSCC	DSCE [24]	61.7	83.9	92.3	-	53.8	73.8	84.2	-
	AD-Cluster [23]	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
	CPL [25]	70.7	87.4	94.6	96.7	59.0	75.2	85.1	88.1
MNC	MMT [26]	76.5	90.9	96.4	97.9	65.7	79.3	89.1	92.4
	NRMT [27]	72.2	88.0	94.5	96.4	62.0	77.5	86.8	89.2
	MEB-Net [28]	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
	LNL [29]	75.2	88.9	95.7	97.3	62.5	77.4	88.1	90.6
	TEST [30]	75.2	90.1	96.0	97.4	66.3	80.1	88.7	91.7
OTHERS	Dual [31]	78.0	90.9	96.4	97.7	67.7	82.1	90.1	92.5
	ICE [32]	73.8	90.8	95.8	97.2	66.4	80.2	88.5	91.6
	GCL [33]	75.4	90.5	96.2	97.1	67.6	81.9	88.9	90.6
	HQP(ours)	80.3	92.3	96.9	97.9	68.0	82.6	90.2	92.4

The effectiveness of high-quality pseudo labels method (HQP)

When both SCL and NII are applied, the proposed HQP method shows the best re-ID performance. On Market-1501 dataset, the performance of mAP/Rank-1 reaches 80.3%/92.3%. On DukeMTMC-ReID dataset, the performance of mAP/Rank-1 reaches 68.0%/82.6%.

4.4 Comparison with the state-of-the-art methods

In this section, the proposed HQP method is compared with existing unsupervised person re-ID methods. Experimental

results of different methods on Market-1501 and DukeMTMC-ReID datasets are shown in Table 2.

The methods used for comparison include sample screening methods (SSM) [18–22], pseudo label noise suppression methods based on cross-camera problem (NSCC) [23–25], co-training methods of multiple networks (MNC) [26–30], and other methods [31–33]. The HQP method improves the quality of pseudo labels, which shows better re-ID performance than the sample screening methods [18–22] and noise suppression methods based on cross-camera problem [23–25] on the whole. Compared with the best sample screening

Table 3 Comparison of our HPQ model with unsupervised person re-ID methods on MSMT17.

Methods		DukeMTMC-ReID→MSMT17				Market-1501→MSMT17			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
SSM	SpCL [21]	26.5	53.1	65.8	70.5	25.4	51.6	64.3	69.7
	MLC [22]	18.0	46.4	57.9	62.7	16.5	43.9	55.3	60.4
MNC	MMT [26]	23.3	50.0	62.8	68.4	19.6	43.3	56.1	61.6
	NRMT [27]	20.6	45.2	57.8	63.3	19.8	43.7	56.5	62.2
	TEST [30]	24.7	51.4	-	-	22.4	51.2	-	-
OTHERS	Dual [31]	26.9	55.0	68.4	73.2	25.1	53.3	66.1	71.5
	ICE [32]	24.3	51.6	-	-	24.0	50.9	-	-
	GCL [33]	29.7	54.4	68.2	74.2	27.0	51.1	63.9	69.9
	HQP(ours)	25.4	53.5	66.0	70.5	23.6	49.8	62.4	68.0

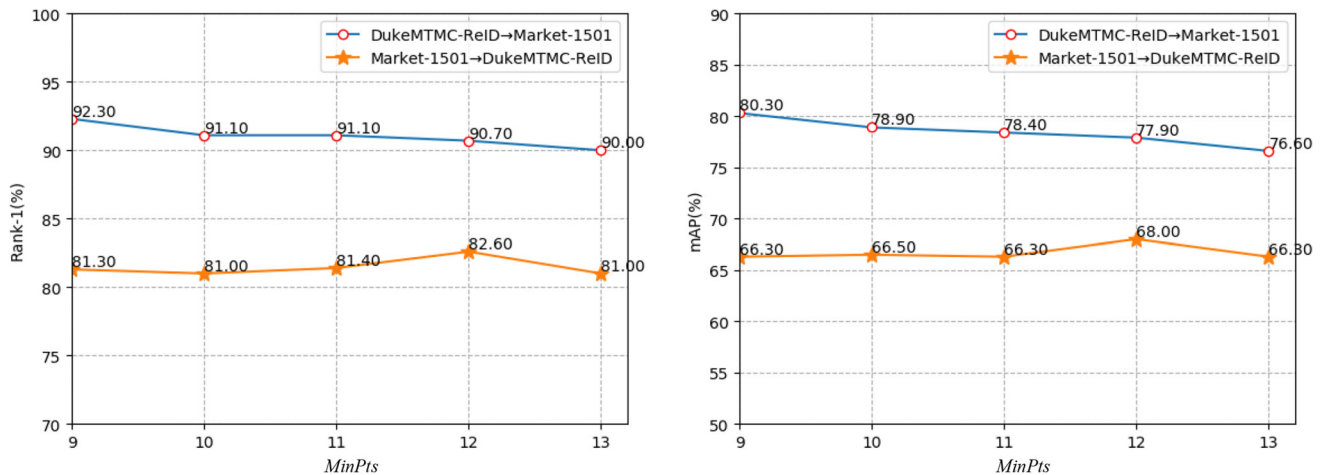


Fig. 9 Analysis of hyper-parameter *MinPts*.

method SpCL [21], our HQP method achieves 3.6%/2% improvement in mAP/Rank-1 on the Market-1501 dataset. On the DukeMTMC-ReID dataset, our HQP method shows slightly lower re-ID performance in mAP/Rank-1 compared with SpCL [21]. Compared with the best cross-camera based method CPL [25], our HQP method improves mAP/Rank-1 by 9.6%/4.9% on Market-1501 dataset, and 9%/7.4% on DukeMTMC-ReID dataset respectively. Compared with the co-training methods of multiple networks [26–30], our HQP method also shows advantages. Compared with MMT [26], the performance of mAP/ Rank-1 on the Market-1501 dataset is increased by 3.8%/1.4%. On the DukeMTMC-ReID dataset, the performance of mAP/Rank-1 is increased by 2.3%/3.3%. From Table 2, it can be found that the pseudo label improving methods ([31] and our HQP method) show better performance than other UCD methods. These experimental results indicate that improving the quality of pseudo label is more effective than suppressing the influence of

pseudo label noise in model training. Compared with the pseudo label improving method Dual [31], our HQP method improves mAP/Rank-1 by 2.3%/1.4% on Market-1501 dataset, and 0.3%/0.5% on DukeMTMC-ReID dataset respectively.

To further verify the effectiveness of our method, comparisons on the largest dataset MSMT17 are conducted. We use Market-1501 and DukeMTMC-Re-ID as the source domains respectively, and the comparison results are shown in Table 3. It should be noted that just the methods tested on MSMT17 are shown in Table 3. It can be seen that the performance of Dual [31] or GCL [33] is better. As indicated by [31], Dual introduces extra GPU memory cost and time cost because of its instant memory bank. GCL requires interval and camera ID information for each image in the target domain, where more priori information needs to be provided. In contrast, the proposed method does not require additional information and is also competitive on MSMT17.

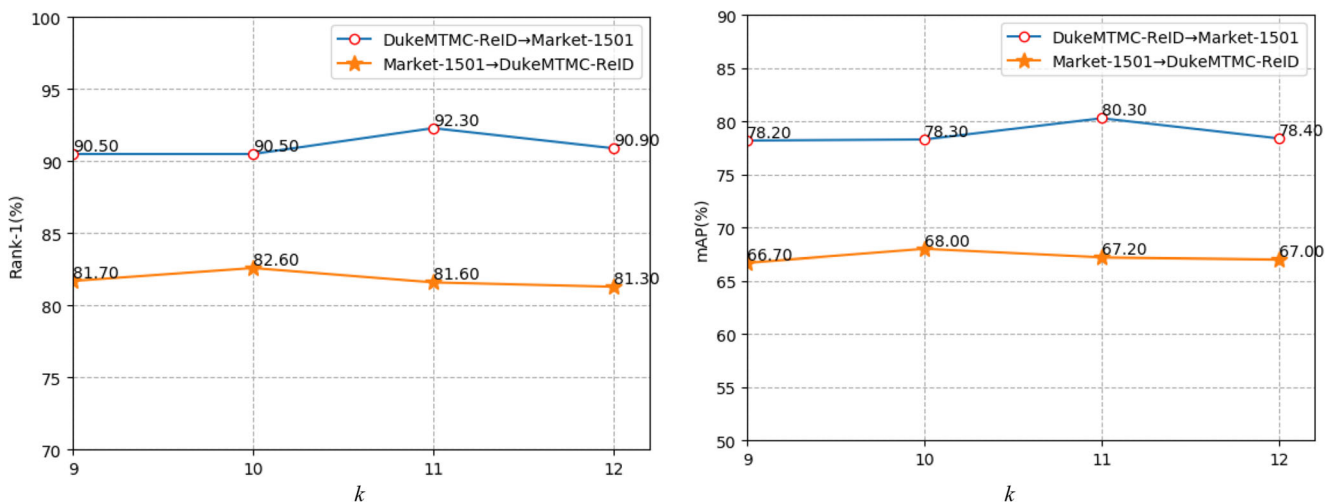


Fig. 10 Analysis of hyper-parameters *k*

4.5 Parameter analysis

This section analyzes the sensitivity of important hyper-parameters in our HQP method. These hyper-parameters are the sample number threshold *MinPts* in DBSCAN clustering method, the number of reference neighbors *k* in NII, and the number of extra neighbor support *n* for outliers in NII. We change the value of one parameter and keep the others unchanged.

Analysis of hyper-parameter *MinPts* This parameter controls the number of minimum samples in the range of neighborhood distance threshold ϵ for a core point in DBSCAN clustering. The sample number threshold *MinPts* is critical to the clustering results. Figure 9 shows the re-ID performance under different *MinPts* values. It can be seen that when Market-1501 is used as the target domain, HQP method has the best performance when *MinPts* = 9. When DukeMTMC-ReID is used as the target domain, the performance of HQP is the best when *MinPts* = 12. We may conclude that when the dataset is more complex, a larger value for *MinPts* is suggested.

Analysis of hyper-parameter *k* This parameter controls the number of reference neighbors in NII. The re-ID performance under different *k* values is shown in Fig. 10. When *k* is set to 11, the HQP method achieves better performance on the Market-1501 dataset. When *k* is set to 10, the HQP method achieves better performance on DukeMTMC-ReID dataset. These results reveal that fewer reference neighbors may lead to insufficient sample similarity information, while more reference neighbors may introduce noise.

Analysis of hyper-parameter *n* This parameter controls the number of extra neighbors for outlier samples in NII. The re-

ID performance under different *n* values is shown in Fig. 11. When *n* is set to 1, the model achieves better performance on the Market-1501 dataset. When *n* is set to 2, the model achieves better performance on DukeMTMC-ReID dataset. These experimental results show that more neighbors are suggested for outlier samples when the dataset is more complex. When *n* is set to a large number, the re-ID performance is reduced on both datasets, indicating that adding too much neighborhood support has the risk of introducing noise.

4.6 Discussion

This section compares and analyzes our method with other related methods. Ding et al. [40] also used neighborhood information to solve the problem of unsupervised person re-ID, and proposed the adaptive exploration (AE) method. According to a threshold, AE adaptively selected neighbors for each image in the feature space. By treating these neighbors as the same class, the non-parametric classifier forced them to stay closer. However, treating the neighbors as the same class is not that reasonable. As shown in the top row of Fig. 7, there are two incorrect samples in the neighbors. Unlike the AE method, we do not assume that the neighbors have the same person ID. Our NII considers the shared neighbors between samples and the similarity between these neighbors to provide a more reliable similarity. When calculating the similarity between sample A and sample B, the neighbors of sample A (NS_A) are found and the neighbors of sample B (NS_B) are also found. Then the similarities between sample A and NS_A and that between sample B and NS_B are both calculated. The final value is the combination of these two, which can aid the clustering method to generate high quality pseudo labels.

Our NII employs the neighborhood information as an auxiliary clue to guide the clustering method. This idea is related to multiple knowledge representation (MKR) [41], which

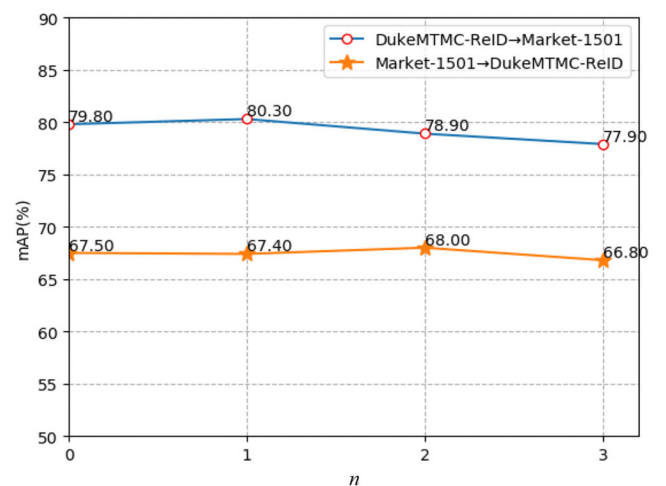
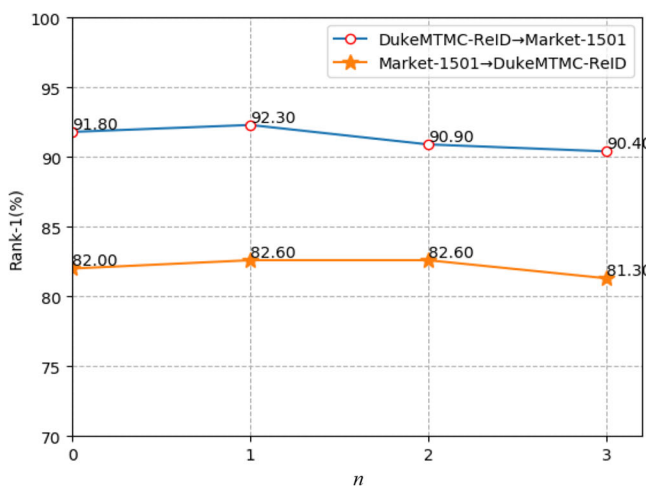


Fig. 11 Analysis of hyper-parameters *n*

introduced a general framework to enhance feature representation through multi-source feature aggregation. MKR aimed at integrating many different forms of input features to enhance the feature representation. Our method just integrates the features of the neighborhood images, and the source of the input features is the same (extracted from IBN-ResNet50), which is more concise.

5 Conclusion

In this paper, from the perspective of improving the quality of pseudo labels, a high-quality pseudo labels (HQP) method is proposed for UDA person re-ID. In order to obtain better feature representation for the target samples, a source domain generalization method based on contrastive learning (SCL) is designed. SCL aids the model to learn the invariance of image, thus improving the feature expression ability of the source domain pre-trained model. In order to provide a more reasonable similarity measurement for the clustering method, a soft label similarity based on the neighborhood information integration (NII) is designed. Using NII to guide the clustering method, the generated pseudo labels are more reliable. The effectiveness of each module in the proposed HQP method is verified by detailed ablation experiments. Compared with existing unsupervised person re-ID methods, the proposed method shows strong competitiveness.

Acknowledgements This work was supported by the [National Natural Science Foundation of China] (Grant numbers [62172029] and [61872030]).

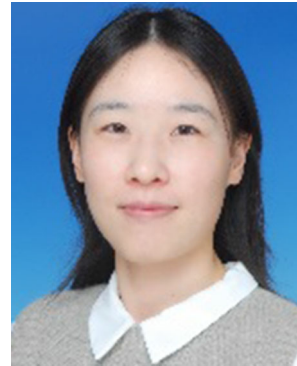
References

1. Zhao D, Chen C, Li D (2022) Multi-stage attention and center triplet loss for person re-identification. *Appl Intell* 52:3077–3089. <https://doi.org/10.1007/s10489-021-02511-2>
2. Lyu C, Ning W, Wang C, Wang K (2022) A multi-branch attention and alignment network for person re-identification. *Appl Intell Online*. <https://doi.org/10.1007/s10489-021-02885-3>
3. Li M, Zhu X, Gong S (2020) Unsupervised tracklet Person re-identification. *IEEE Trans Pattern Anal Mach Intell* 42(7):1770–1782. <https://doi.org/10.1109/TPAMI.2019.2903058>
4. Jiang K, Zhang T, Zhang Y, Wu F, Rui Y (2020) Self-supervised agent learning for unsupervised cross-domain person re-identification. *IEEE Trans Image Processing* 29:8549–8560. <https://doi.org/10.1109/TIP.2020.3016869>
5. Cheng D, Li J, Kou Q, Zhao K, Liu R (2022) H-net: unsupervised domain adaptation person re-identification network based on hierarchy Image. *Vis Comput* 124:104493. <https://doi.org/10.1016/j.imavis.2022.104493>
6. Wei L, Zhang S, Gao W, Tian Q (2018) Person transfer GAN to bridge domain gap for person re-identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 79–88. <https://doi.org/10.1109/CVPR.2018.00016>
7. Li Y, Lin C, Lin Y, Wang Y (2021) Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp 7919–7929. <https://doi.org/10.1109/ICCV.2019.00801>
8. Zheng D, Xiao J, Chen K, Huang X, Chen L, Zhao Y (2022) Soft pseudo-Label shrinkage for unsupervised domain adaptive person re-identification. *Pattern Recogn* 127:108615. <https://doi.org/10.1016/j.patcog.2022.108615>
9. Zhang C, Tang Y, Zhang Z, Li D, Yang X, Zhang W (2021) Improving domain-adaptive person re-identification by dual-alignment learning with camera-aware image generation. *IEEE Trans Circuits Syst Video Technol* 31(11):4334–4346. <https://doi.org/10.1109/TCSVT.2020.3047095>
10. Zhong Z, Zheng L, Luo Z, Li S, Yang Y (2019) Invariance matters: exemplar memory for domain adaptive person re-identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 598–607. <https://doi.org/10.1109/CVPR.2019.00069>
11. Wu A, Zheng W, Lai J (2019) Unsupervised person re-identification by camera-aware similarity consistency learning. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp 6922–6931. <https://doi.org/10.1109/ICCV.2019.00702>
12. Liu G, Wu J (2021) Unsupervised person re-identification by intra-inter camera affinity domain adaptation. *J Vis Commun Image Represent* 80:103310. <https://doi.org/10.1016/j.jvcir.2021.103310>
13. Xiang W, Yong H, Huang J, Hua X, Zhang L (2021) Second-order camera-aware color transformation for cross-domain person re-identification. *Asian Conference on Computer Vision (ACCV)*, pp 36–53. https://doi.org/10.1007/978-3-030-69532-3_3
14. Wu J, Liu H, Yang Y, Lei Z, Liao S, Li S (2019) Unsupervised graph association for person re-identification. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp 8320–8329. <https://doi.org/10.1109/ICCV.2019.00841>
15. Luo C, Song C, Zhang Z (2020) Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup. *European Conference on Computer Vision (ECCV)*, pp 224–241. https://doi.org/10.1007/978-3-030-58555-6_14
16. Qi L, Wang L, Huo J, Zhou L, Shi Y, Gao Y (2019) A novel unsupervised camera-aware domain adaptation framework for person re-identification. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp 8080–8089. <https://doi.org/10.1109/ICCV.2019.00817>
17. Mekhazni D, Bhuiyan A, Ekladios G, Granger E (2020) Unsupervised domain adaptation in the dissimilarity space for person re-identification. *European Conference on Computer Vision (ECCV)*, pp 159–174. https://doi.org/10.1007/978-3-030-58583-9_10
18. Fan H, Zheng L, Yan C, Yang Y (2018) Unsupervised person re-identification: clustering and fine-tuning. *ACM Trans Multimed Comput* 14(4):1–18. <https://doi.org/10.1145/3243316>
19. Xian Y, Hu H (2018) Enhanced multi-dataset transfer learning method for unsupervised person re-identification using co-training strategy. *IET Comput Vis* 12(8):1219–1227. <https://doi.org/10.1049/iet-cvi.2018.5103>
20. Yang F, Li K, Zhong Z, Luo Z et al (2020) Asymmetric co-teaching for unsupervised cross-domain person re-identification. *Proc AAAI Conf Artif Intell* 34:12597–12604
21. Ge Y, Zhu F, Chen D, Zhao R, Li H (2020) Self-paced contrastive learning with hybrid memory for domain adaptive object Re-ID. *Conference on Neural Information Processing Systems (NeurIPS)*, pp 1–14
22. Li Q, Peng X, Qiao Y, Hao Q (2022) Unsupervised person re-identification with multi-label learning guided self-paced clustering. *Pattern Recogn* 125:108521. <https://doi.org/10.1016/j.patcog.2022.108521>

23. Zhai Y, Lu S, Ye Q, Shan X, Chen J, Ji R, Tian Y (2020) AD-cluster: augmented discriminative clustering for domain adaptive person re-identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 9018–9027. <https://doi.org/10.1109/CVPR42600.2020.00904>
24. Yang F, Zhong Z, Luo Z et al (2021) Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 4853–4862. <https://doi.org/10.1109/CVPR46437.2021.00482>
25. Zhu X, Li Y, Sun J, Chen H, Zhu J (2021) Unsupervised domain adaptive person re-identification via camera penalty learning. *Multimed Tools Appl* 80:15215–15232. <https://doi.org/10.1007/s11042-021-10589-6>
26. Ge Y, Chen D, Li H (2020) Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. *International Conference on Learning Representations (ICLR)*, pp 1–15
27. Zhao F, Liao S, Xie G, Zhao J, Zhang K, Shao L (2020) Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. *European Conference on Computer Vision (ECCV)*, pp 526–544. https://doi.org/10.1007/978-3-030-58621-8_31
28. Zhai Y, Ye Q, Lu S, Jia M, Ji R, Tian Y (2020) Multiple expert brainstorming for domain adaptive person re-identification. *European Conference on Computer Vision (ECCV)*, pp 594–611. https://doi.org/10.1007/978-3-030-58571-6_35
29. Zhu X, Li Y, Sun J, Chen H, Zhu J (2021) Learning with noisy labels method for unsupervised domain adaptive person re-identification. *Neurocomputing* 452:78–88. <https://doi.org/10.1016/j.neucom.2021.04.120>
30. Li Y, Yao H, Xu C (2021) TEST: Triplet ensemble student-teacher model for unsupervised person re-identification. *IEEE Trans Image Process* 30:7952–7963. <https://doi.org/10.1109/TIP.2021.3112039>
31. Dai Y, Liu J, Bai Y, Tong Z, Duan L (2021) Dual-refinement: joint label and feature refinement for unsupervised domain adaptive person re-identification. *IEEE Trans Image Process* 30:7815–7829. <https://doi.org/10.1109/TIP.2021.3104169>
32. Li Y, Yao H, Xu C (2022) Intra-domain consistency enhancement for unsupervised person re-identification. *IEEE Trans Multimedia* 24:415–425. <https://doi.org/10.1109/TMM.2021.3052354>
33. Chen H, Wang Y, Lagadec B, Dantcheva A, Francois B (2021) Joint generative and contrastive learning for unsupervised person re-identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 2004–2013. <https://doi.org/10.1109/CVPR46437.2021.00204>
34. Pan X, Luo P, Shi J, Tang X (2018) Two at once: enhancing learning and generalization capacities via IBN-Net[C]. *European Conference on Computer Vision (ECCV)* 484–500. https://doi.org/10.1007/978-3-030-01225-0_29
35. Deng J, Dong W, Socher R, Li L, Li K, Li F (2019) ImageNet: a large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
36. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1116–1124. <https://doi.org/10.1109/ICCV.2015.133>
37. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. *IEEE International Conference on Computer Vision (ICCV)*, pp 3754–3762. <https://doi.org/10.1109/ICCV.2017.405>
38. Bolle R, Connell J, Pankanti S, Ratha N, Senior A (2005) The relation between the ROC curve and the CMC. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pp 15–20. <https://doi.org/10.1109/AUTOID.2005.48>
39. Felzenszwalb P, Mcallester D, Ramanan D (2018) A discriminatively trained, multiscale, deformable part model. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1–8. <https://doi.org/10.1109/CVPR.2008.4587597>
40. Ding Y, Fan H, Xu M (2020) Adaptive exploration for unsupervised person re-identification. *ACM Trans Multimed Comput Commun Appl* 16(1):1551–6857. <https://doi.org/10.1145/3369393>
41. Yang Y, Zhuang Y, Pan Y (2021) Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies. *Front Inform Technol Electron Eng* 22(12):1551–1558. <https://doi.org/10.1631/FITEE.2100463>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yanfeng Li is a professor at the School of Electronic and Information Engineering, Beijing Jiaotong University. She received the B.S. degree in communication engineering from Beijing Jiaotong University in 2010 and the Ph.D. degree in circuit and system from Beijing Jiaotong University in 2015. Her research interests include signal processing and artificial intelligence.



Xiaodi Zhu received the B.S. degree at Qingdao University in 2019 and M.S. at Beijing Jiaotong University in 2022. Her current research interests include image processing, pattern recognition and deep learning.



Jia Sun received the B.S. degree in electronic science and technology from Beijing Jiaotong University in 2017 and the Ph.D. degree in information and communication engineering from Beijing Jiaotong University in 2022. Her current research interests include image processing, pattern recognition and deep learning.



Houjin Chen is a professor at the School of Electronic and Information Engineering, Beijing Jiaotong University. He received his B.S. degree in traffic signal and control from Lanzhou Jiaotong University in 1986. He received his M.S. and Ph.D. degrees in signal and information processing from Northern Jiaotong University in 1989 and 2003, respectively. His research interests include signal processing and artificial intelligence.



Zhiyuan Li is currently studying for the M.S. degree at the School of Electronic and Information Engineering, Beijing Jiaotong University. He received the B.S. degree in communication engineering from Tianjin University of Technology in 2020. His current research interests include image processing, pattern recognition and deep learning.