



DCSG: data complement pseudo-label refinement and self-guided pre-training for unsupervised person re-identification

Qing Han^{1,2,3} · Jiongjin Chen¹ · Weidong Min^{1,2,3} · Jiahao Li¹ · Lixin Zhan¹ · Longfei Li¹

Accepted: 8 June 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Existing unsupervised person re-identification (Re-ID) methods use clustering to generate pseudo-labels that are generally noisy, and initializing the model with ImageNet pre-training weights introduces a large domain gap that severely impacts the model's performance. To address the aforementioned issues, we propose the data complement pseudo-label refinement and self-guided pre-training framework, referred to as DCSG. Firstly, our method utilizes image information from multiple augmentation views to complement the source image data, resulting in aggregated information. We employ this aggregated information to design a correlation score that serves as a reliability evaluation for the source features and cluster centroids. By optimizing the pseudo-labels for each sample, we enhance their robustness. Secondly, we propose a pre-training strategy that leverages the potential information within the training process. This strategy involves mining classes with high similarity in the training set to guide model training and facilitate smooth pre-training. Consequently, the model acquires preliminary capabilities to distinguish pedestrian-related features at an early stage of training, thereby reducing the impact of domain gaps arising from ImageNet pre-training weights. Our method demonstrates superior performance on multiple person Re-ID datasets, validating the effectiveness of our proposed approach. Notably, it achieves an mAP metric of 84.3% on the Market1501 dataset, representing a 2.8% improvement compared to the state-of-the-art method. The code is available at <https://github.com/duolaJohn/DCSG.git>.

Keywords Person re-identification · Unsupervised learning · Frequent itemset mining · Label smoothing

1 Introduction

Person Re-ID aims to identify individuals in images across different camera views and timestamps. It finds applications in various domains, including public safety, smart cities, and autonomous driving. The rapid advancement of deep neural networks [1, 2], coupled with extensive research on attention mechanisms [3–5], has led to notable developments. For instance, EAPT [3] introduces a variable attention mechanism capable of encompassing non-fixed attention information from diverse visual elements. Additionally, BaGFN [4] proposes a novel generalized attention mechanism that dynamically learns importance weights for cross-features. Thus, supervised methods [6–11] have achieved remarkable performance. However, these methods necessitate a substantial amount of costly annotated data, which restricts their usage in large-scale real-world re-labeling scenarios. To tackle this issue, considerable attention has been given to unsupervised learning-based person Re-ID methods.

✉ Weidong Min
minweidong@ncu.edu.cn

Qing Han
hanqing@ncu.edu.cn

Jiongjin Chen
416100210351@email.ncu.edu.cn

Jiahao Li
359100230012@email.ncu.edu.cn

Lixin Zhan
zhanlixin@email.ncu.edu.cn

Longfei Li
lilongfei@email.ncu.edu.cn

¹ School of Mathematics and Computer Science, Nanchang University, Nanchang 330031, China

² Institute of Metaverse, Nanchang University, Nanchang 330031, China

³ Jiangxi Provincial Key Laboratory of Virtual Reality, Nanchang 330031, China

Previous unsupervised methods [12, 13] for person Re-ID employ a two-stage iterative process, which includes clustering to generate pseudo-labels and training the model using these labels. Nevertheless, the presence of noise in pseudo-labels presents an inevitable challenge in these methods, as errors resulting from incorrect pseudo-labels accumulate during training, exerting a significant impact on the performance of unsupervised methods. Furthermore, the majority of approaches [14, 15] directly utilize pre-trained weights from ImageNet for model initialization. However, given that the person Re-ID task involves fine-grained classification, utilizing pre-trained weights introduces a substantial domain gap, which may not be suitable for the Re-ID task and ultimately results in diminished fine-tuning performance [16]. Therefore, it is imperative to investigate the reduction of pseudo-label noise and the mitigation of the domain gap caused by pre-training weights in person Re-ID methods.

To address these two problems, many existing methods [17–20] employ label softening techniques to mitigate the noise impact of the labels. For instance, PPLR [20] leverages the local information of pedestrians to optimize the labels, while MMT [18] utilizes feature information from two networks with different initialization parameters to optimize the labels in a mutually beneficial manner. However, these methods fail to consider the integration of multiple information sources from augmented viewpoints to supplement the information obtained from the original viewpoints and optimize the pseudo-labels during model training. One of the recent advancements in person Re-ID involves exploring unsupervised pre-training methods [16, 21, 22] by introducing a large-scale unlabeled dataset specifically designed for person Re-ID. This approach aims to enhance the model's generalization capabilities and mitigate the noise impact of ImageNet pre-training weights on pseudo-labels. Nevertheless, this method necessitates the creation of a new person dataset for unsupervised pre-training, which incurs significant manpower and material resource expenses.

This paper addresses the two aforementioned key issues. To mitigate the impact of pseudo-label noise during the training process, we propose the data complement pseudo-label refinement (DCLR) framework. This framework leverages aggregated features, obtained by complementing the source features with information from multiple augmentation views, to optimize the pseudo-labels. In comparison with previous methods, our approach takes into account the limited availability of pedestrian feature information and utilizes information from multiple augmentation views to enhance its comprehensiveness. To address the domain gap caused by pre-training weights, we propose the self-guided pre-training (SGP) strategy. This strategy harnesses potential information in the training process, conducts data mining on the training set and employs Easy-To-Discriminate (ETD) class (more details in Sect. 3.3.1) samples to guide the model's pre-

training. As a result, the model gains the ability to initially discern relevant pedestrian features and further explore its potential. In comparison to previous pre-training methods, our approach utilizes the original training set as a dataset for pre-training, enhancing its generalizability without requiring additional labor-intensive dataset collection. The method only incurs a modest increase in training time and offers new research directions in the field of unsupervised person Re-ID. Notably, our method achieves significant performance improvements across multiple datasets. Our contributions can be summarized as follows:

- We propose the DCLR framework, which enhances the source features by incorporating information from multiple augmentation views to obtain aggregated features. These aggregated features are then used to design a correlation score, evaluating the reliability between the source features and cluster centroids, in order to optimize the pseudo-labels for each sample.
- We introduce the SGP strategy that exploits the coupling information between pairs of images and mining ETD classes in the training set. This strategy enables the model to develop an initial ability to discriminate pedestrian-related features during the early stage of training and reduces the impact of the domain gap caused by ImageNet pre-training weights.
- We conduct extensive experiments on three datasets [23–25] to demonstrate the effectiveness of our method. Moreover, our method surpasses the current state of the art in purely unsupervised approaches without requiring camera labeling.

2 Related work

2.1 Unsupervised person re-ID

Unsupervised domain adaptation person Re-ID (UDA) [12, 13, 19, 26, 27] utilizes source domain labels for pre-training and pure unsupervised person Re-ID (USL) [17, 20, 28]. Recent research in UDA methods [29] has primarily focused on aligning feature distributions between the source and target domains to address inter-domain offsets. JVTC [30] addresses this challenge by jointly considering visual and temporal consistency. It combines both local single-class classification and global multi-class classification. In USL methods, clustering approaches are commonly employed to generate pseudo-labels for guiding model training. While person Re-ID methods [31–33] based on contrastive learning have achieved impressive results, CACL [34] proposes cluster-guided asymmetric contrast learning, which allows the clustering results to be used to guide feature learning the presence of noise in pseudo-labeling significantly impacts

model performance. Several recent works have explored solutions for this issue. For example, SECRET [35] improves label consistency by mutually optimizing pseudo-labels generated from different feature spaces in a multibranching model. X.Han [36] proposes a group sampling approach for unsupervised person re-identification, which improves performance by addressing over-fitting and enhancing statistical stability. HCM [37] propose a novel Hybrid Contrastive Model for unsupervised person re-identification, which combines identity-level and image-level contrastive learning to effectively leverage feature similarities among hard sample pairs. CACHE [38] reports a complementary attention-driven contrast learning and hard sample exploration algorithm to improve feature discriminability and make clustering more compact. In contrast to these works, our approach incorporates multiple augmentation views information, allowing it to complement the source features and obtain aggregated features with more comprehensive information about pedestrians. These aggregated features are then utilized to optimize the labels.

2.2 Frequent itemset mining

A significant volume of data conceals valuable and useful information. Frequent itemset mining (FIM) [39, 40] is a crucial task in data analysis aimed at extracting frequently occurring events or patterns from a large dataset [41]. Since the introduction of the frequent itemset mining problem in the early 1990s, numerous solutions have been proposed, with the most notable ones being the Apriori algorithm, FP-Growth algorithm and Eclat algorithm [42]. To mitigate the issue of extensive memory consumption, certain studies have proposed algorithmic variants within a distributed architecture [43–45]. The data can be decomposed into smaller units for parallel processing [46, 47]. However, it is important to note that these enhancements primarily pertain to innovations in distributed architecture, rather than breakthroughs in the FIM algorithm itself [48]. Moreover, to tackle the complexity problem, the non-exhaustive search model has been extended, ensuring a fixed-time solution for this problem [41, 48, 49]. However, the computational complexity arising from a large number of data items grows exponentially, posing a persistent challenge. Diverging from previous approaches, our study addresses this challenge by transforming the task of mining frequent multiple data items into a similarity search problem between pairs of data items. Subsequently, we employ a clustering algorithm to extract highly correlated multiple data items, effectively reducing memory consumption and computational complexity. Finally, we utilize these identified multiple data items for pre-training the class guidance model.

3 Methodology

To address the problem of label noise in the model learning process, we present the data complement pseudo-label refinement (DCLR) framework. DCLR utilizes aggregated features obtained by complementing the source features with multiple augmentation views information from the same sample. These aggregated features are then employed to design a correlation score, which serves as a reliability evaluation between the source features and the cluster centroids. Subsequently, the reliability evaluation score is utilized to refine the pseudo-labels during the training stage. The framework diagram is illustrated in Fig. 1d. In order to mitigate the domain gap issue resulting from the initialization of ImageNet pre-training weights applied to the model, we introduce the self-guided pre-training strategy (SGP). SGP leverages the coupling information between pairs of images during the training process, which is converted into a distance metric matrix. Subsequently, pseudo-labels are generated through clustering to guide the model during pre-training. The framework is depicted in Fig. 2.

3.1 Overview of the baseline network

In this study, we introduce a clustering-based USL framework as the baseline for our model. The training process consists of two stages. The first stage, depicted in Fig. 1a and referred to as the Clustering Stage, involves extracting features from all images in the training set and generating pseudo-labels through clustering using the DBSCAN [50] algorithm. The second stage, also depicted in Fig. 1b, serves as the Baseline. In this stage, the images are randomly augmented with additional data, and the model is trained using the pseudo-labels. The training process alternates between the cross-entropy loss function and the soft triplet loss until the network converges.

In this study, we consider the unlabeled training dataset denoted as $X = \{x_i\}_{i=1}^{N_D}$, where x_i represents an image and N_D denotes the total number of images. Our model extracts the features of an image through the process $F_\theta(x_i) \in \mathbb{R}^C$, where C represents the channel size of the feature. These extracted features are denoted as f_i . We utilize the DBSCAN clustering algorithm to cluster the set of features $F = \{f_i\}_{i=1}^{N_D}$ and assign pseudo-labels $Y = \{y_i\}_{i=1}^{N_D}$. During training, we only consider images that have been assigned class labels and disregard unclustered outliers. The pseudo-labels y_i that have been assigned classes are represented using one-hot encoding with M clusters.

Pseudo-labels serve as the identities of pedestrians and are incorporated into the training process. The loss function employed in our baseline model using the following formula:

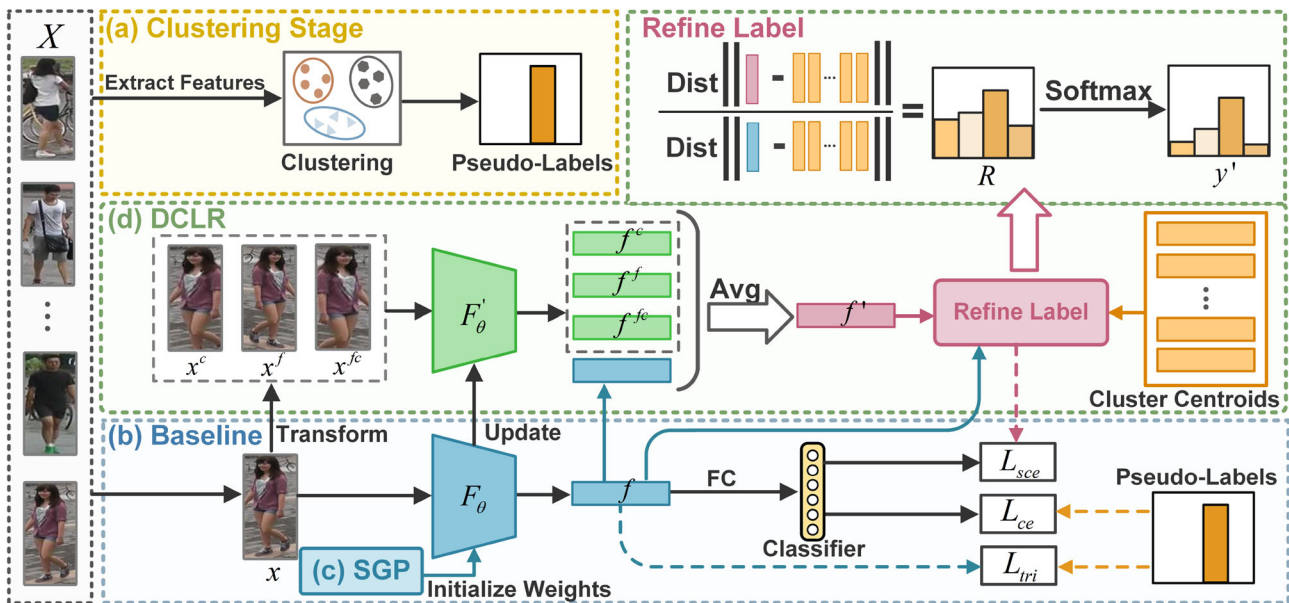


Fig. 1 DCSG method architecture and components. **a** Clustering Stage: Pseudo-labels are generated for each sample by clustering all features extracted from the training set. **b** Baseline: During training, the pseudo-labels generated in the Clustering Stage are used to guide model training. After training, only F_θ is retained for test, while the source features f are used for testing. **c** SGP Strategy: Pre-trained weights obtained from the SGP strategy are used to initialize the network F_θ . Further details

of the SGP strategy can be found in Fig. 2. **d** DCLR Module: Network parameters are updated prior to each epoch of training for F'_θ . Aggregate features are obtained by combining features from multiple augmented views with the source features. These aggregate features, along with the source features, are used to calculate the reliability evaluation score R with respect to all cluster centroids. The score R is then transformed into a soft label y' using the softmax function

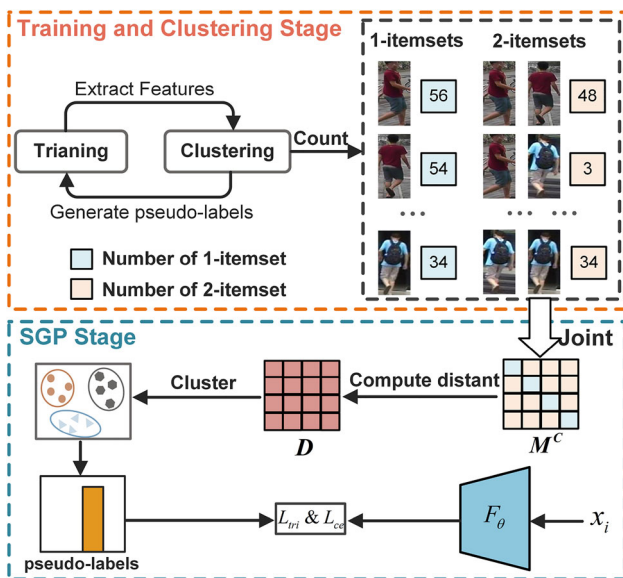


Fig. 2 Illustration of the proposed SGP strategy. In the Training and Clustering Stage, we calculate the frequency of occurrence for 1-itemsets and 2-itemsets. In the SGP Stage, we combine the 1-itemsets and 2-itemsets to create a statistical matrix M^C , which is then transformed into a distance matrix D . Clustering is performed using D , and the resulting pseudo-labels are used in model pre-training

$$L_{base} = L_{ce} + L_{tri} \tag{1}$$

However, the performance of this baseline approach is constrained by the accuracy of the pseudo-labels, which introduces considerable noise during training. To mitigate this noise impact, we propose the DCLR method, which optimizes the labels.

3.2 Data complement pseudo-label refinement framework

The DCLR method leverages features extracted from multiple augmented views to augment the source features and generate aggregated features. It utilizes aggregated feature computations and reliability evaluation scores for all categories to optimize label information, effectively mitigating the noise impact of labels. The DCLR method is illustrated in Fig. 1d.

To preserve the original color information and pedestrian-specific details in the images, our experiments incorporate various data augmentation techniques. Specifically, we apply random cropping, flipping and combinations of random cropping and flipping. By employing these strategies, we generate three sets of augmented images denoted as $X^a = \{x_i^c, x_i^f, x_i^{fc}\}$. These images are then fed into the feature extraction network F'_θ to obtain corresponding features $F^a =$

$\{f_i^c, f_i^f, f_i^{fc}\}$. To enhance the feature representation, we average these augmented features with the original feature, resulting in aggregated features denoted as f_i' using the following formula:

$$f_i' = \text{Avg}(\{f_i^c, f_i^f, f_i^{fc}\} \cup f_i) \quad (2)$$

where the averaging operation denoted as $\text{Avg}(\cdot)$ is employed to calculate the average of multiple features. Notably, f_i' contains more informative characteristics than f_i . Subsequently, a cluster centroids vector $w_i^c \in \mathbb{R}^{M \times C}$ is constructed to represent the cluster centroids. The Euclidean distance $O(f_i, w_i^c)$ measures the distance between the original feature f_i and the M cluster centroids, while the Euclidean distance $O(f_i', w_i^c)$ quantifies the distance between the aggregated feature f_i' and the M cluster centroids. The expression uses the following formula:

$$\begin{aligned} O(f_i, w_i^c) &= \|\mathbf{u} \cdot f_i^T - w_i^c\| \\ O(f_i', w_i^c) &= \|\mathbf{u} \cdot f_i'^T - w_i^c\| \end{aligned} \quad (3)$$

where $\|\cdot\|$ denotes the L2-norm of the row vectors of a matrix, and $\mathbf{u} \in \mathbb{R}^C$ is an $\mathbb{1}$ vector. We employ the similarity distance defined by Eq. 3 to compute a reliability evaluation score, which assesses the reliability of the source feature f_i for all classes of cluster centroids using the following formula:

$$R_i = \frac{O(f_i', w_i^c)}{O(f_i, w_i^c)} \quad (4)$$

According to Eq. 4, we define $R_{i,j}$, $j \in \{1, 2, \dots, M\}$ as the reliability evaluation value of the i -th image for the j -th cluster centroid. If $R_{i,j} > 1$, it indicates that the feature is further away from the cluster centroid and has a weaker association with it. Conversely, if $R_{i,j} < 1$, the feature is closer to the cluster centroid, indicating a stronger relationship. In this case, the feature point is assigned a higher labeling value, and the control model's learning is more aligned with this cluster centroid. Finally, R_i is normalized to obtain the smoothed labels using the following formula:

$$y'_{i,k} = \frac{\exp(R_{i,k})}{\sum_{j=1}^M \exp(R_{i,j})} \quad (5)$$

where the $y'_{i,k}$ represents the label value of node k in the i -th image. Given other relevant class label values, the model is directed to focus on other associated classes as a way to better learn the similarities between pedestrian images. We add y'_i to the following soft cross-entropy loss function to guide the model through training using the following formula:

$$L_{sce} = - \sum_{i=1}^{N_D} y'_i \cdot \log(q_i) \quad (6)$$

The overall loss function of DCLR is defined by the following formula:

$$L_{dclr} = L_{sce} + L_{ce} + L_{tri} \quad (7)$$

The DCLR method efficiently leverages information from multi-incremental generalized views to smooth pseudo-labels, directs the model's attention toward other classes using the loss function L_{sce} , mitigates the noise impact of pseudo-labels and ensures sample concentration in the model category through the loss function L_{ce} . Moreover, when the model converges and the reliability evaluation scores of all categories reach 1, all categories are assigned the same probability value, and the smoothed label y'_i is ultimately compressed into a uniform vector, thus making the training signal lack meaningful information. During this stage, the cross-entropy loss function assumes a primary role by utilizing the label y_i to guide the model toward further convergence of the sample to its corresponding cluster centroid.

3.3 Self-guided pre-training strategy

To address the domain gap caused by ImageNet pre-training weights, we propose the SGP strategy. This strategy explores the temporal dimension of the model training process, identifies ETD samples in the training set and guides the model's pre-training. Consequently, the model acquires an initial ability to discriminate pedestrian features and realizes its full potential.

3.3.1 Cluster dataset definition

During the model's training process, there is consistently a group of samples that appear frequently within the same class of clustering generated by each epoch. These samples exhibit a high degree of correlation among themselves, making it highly likely to correctly identify them as samples of the same identity or as extremely similar samples with different identities. We refer to this group of samples as "Easy-To-Discriminate (ETD)" class. By identifying and utilizing ETD classes, we can guide the model's pre-training, enabling it to develop the initial ability to recognize pedestrian features.

Drawing inspiration from the field of frequent itemset mining, we consider each sample in the training set as an item. Furthermore, each cluster generated through clustering in the training process of all epochs is treated as a record (or transaction), and these records are combined to form a cluster dataset. We aim to extract frequent patterns (or itemsets) from the cluster dataset, which represents the set of items that occur frequently in the dataset and reflect the interdependence and association among them. These frequent patterns are utilized as class labels to guide the pre-training of the model.

In a complete training process, we begin by defining the set of clusters in the t -th epoch phase as $C_t^s = \{A_{t,i}\}_{i=1}^{N_t^s}$, $A_{t,i} \subseteq \{1, 2, \dots, N_D\}$. Here, N_t^s represents the number of clusters generated by clustering in the t -th epoch phase, and $A_{t,i}$ denotes the set of index values of images within a cluster. Next, we define the cluster dataset as $X^c = \{C_0^s \cup C_1^s \cup \dots \cup C_T^s\}$, where T denotes the total number of training epochs. Additionally, we define the number of records in X^c as $K = \sum_t^T N_t^s$.

3.3.2 Compute the similarity measure matrix

The commonly used algorithms for FIM are the Apriori algorithm and the FP-Growth algorithm. However, the cluster dataset X^c obtained from the USL task is characterized by a large number of items, resulting in a potentially exponential number of feasible candidate sets, reaching $2^{N_D} - 1$ [48]. The computational complexity of calculating the frequency for each candidate set can even reach $O((2^{N_D} - 1) \times K \times N_D)$ [48]. Consequently, the computational and spatial complexity associated with mining this dataset using the commonly used FIM algorithms is impractical. To address this challenge and achieve correlation analysis between multiple items with low computational and memory consumption, we simplify the problem by reducing the mining of frequent itemsets between multiple items to exploring the correlation between pairs of items. This involves investigating the similarity distance between each pair of items. Subsequently, we utilize clustering algorithms to mine relationships among multiple items.

Drawing inspiration from the work of the re-ranking [51], we propose a novel similarity distance matrix. This matrix is constructed by tallying the occurrences of 1-itemset and 2-itemset rules derived from the clustering results. These rules are integrated into a frequency matrix, which is subsequently transformed into a similarity measure matrix representing the relationships between pairs of images.

Specifically, we define the 1-itemsets of statistical occurrences N'_i in X^c and the 2-itemsets of statistical occurrences $N''_{i,j}$ in X^c , where $\{i\}, \{i, j\} \subseteq \{1, 2, \dots, N_D\}$. Furthermore, we introduce the statistical matrix $M^C \in \mathbb{R}^{N_D \times N_D}$, which tallies the frequency of simultaneous occurrences of one or two items in all transactions. The expression of this matrix uses the following formula:

$$M^C_{i,j} = \begin{cases} N'_i, & \text{if } i = j \\ N''_{i,j}, & \text{if } i \neq j \end{cases} \quad (8)$$

The diagonal of matrix M^C represents the frequency of N'_i , while the off-diagonal elements represent the frequency of $N''_{i,j}$. To quantify the level of association between two images, we opted to convert the matrix M^C into a Jaccard distance matrix using the following formula:

$$D^J_{i,j} = \begin{cases} 0, & \text{if } i = j \\ 1 - \frac{M^C_{i,j}}{M^C_{i,i} + M^C_{j,j} - M^C_{i,j}}, & \text{if } i \neq j \end{cases} \quad (9)$$

The distance matrix obtained from Eq. 9 effectively captures the degree of association between pairs of images. However, it fails to address situations where the number of 1-itemsets in both images is low and approaches the number of 2-itemsets. This can lead to unexpectedly high similarity distances. To mitigate the occurrence of such anomalies, we introduce a novel distance called the frequency distance using the following formula:

$$D^F_{i,j} = \begin{cases} 0, & \text{if } i = j \\ 1 - \frac{M^C_{i,j}}{\text{Max}(M^C)}, & \text{if } i \neq j \end{cases} \quad (10)$$

where $\text{Max}(\cdot)$ represents the maximum value in the computation matrix. To obtain the Final Similarity Distance $D^F_{i,j}$, we combine the Jaccard distance $D^J_{i,j}$ with the frequency distance $D^F_{i,j}$. This distance matrix quantifies the similarity distance between pairs of images and is calculated using the following formula:

$$D_{i,j} = D^J_{i,j} + D^F_{i,j} \quad (11)$$

3.3.3 Labels refinement and pre-training

The Final Similarity Distance matrix D is inputted into the DBSCAN clustering algorithm to extract multiple sets of ETD classes and generate pseudo-labels for the corresponding samples.

During the SGP-enabled phase, these ETD class pseudo-labels are incorporated into the pre-training process while disregarding outliers. These pseudo-labels contribute to the smooth training of the model, enabling it to initially discriminate pedestrian feature information. The SGP method effectively mitigates the domain gap caused by the ImageNet pre-training parameters. Finally, the ImageNet pre-training parameters are replaced with the pre-trained model weights obtained in the SGP strategy for subsequent model training.

4 Experiments

4.1 Datasets and evaluation protocols

4.1.1 Datasets

We evaluated the effectiveness of our method on three datasets: Market1501, DukeMTMC-reID and MSMT17. Market1501 is widely used in person Re-ID research. The Market1501 dataset consists of a training set with 12,936

Table 1 Comparison with the state-of-the-art unsupervised person Re-ID methods on Market1501, DukeMTMC-reID and MSMT17 datasets

Methods	Market1501			DukeMTMC-reID			MSMT17		
	mAP (%)	R1 (%)	R5 (%)	mAP (%)	R1 (%)	R5 (%)	mAP (%)	R1 (%)	R5 (%)
JVTC [30]	47.5	79.5	89.2	50.7	74.6	82.9	17.3	43.1	53.8
SpCL [32]	73.1	88.1	95.1	65.3	81.2	90.3	19.1	42.3	55.6
HCD [29]	78.1	91.1	96.4	65.6	79.8	88.6	26.9	53.7	65.3
ICE [17]	79.5	92.0	97.0	67.2	81.3	90.1	29.8	59.0	71.7
GCL [33]	66.8	87.3	93.5	62.8	82.9	87.1	21.3	45.7	58.6
CACL [34]	80.9	92.7	97.4	69.6	82.6	91.2	23.0	48.9	61.2
SECRET [35]	80.8	92.1	–	63.1	77.4	–	30.5	60.3	–
PPLR [20]	81.5	92.8	97.1	69.1	81.1	90.3	31.4	61.1	73.4
X.Han [36]	79.2	92.3	96.6	69.1	82.7	91.1	24.6	56.2	67.3
HCM [37]	79.0	91.8	96.7	67.9	82.3	90.2	26.9	59.1	70.1
CACHE [38]	81.0	92.0	96.9	70.6	83.8	91.6	31.8	58.2	69.9
Ours	84.3	93.5	97.7	72.6	84.1	92.7	35.8	65.0	76.7

images, including 751 pedestrians. The query set contains 3,368 images of 750 pedestrians, and the gallery set contains 19,732 images. The DukeMTMC-reID training set comprises 16,522 images of 702 pedestrians. The query set contains 2,228 images of 702 pedestrians, while the gallery set consists of 702 pedestrians and 408 interfering pedestrians, totaling 17,661 images. MSMT17 is a larger and more challenging dataset that closely represents real-world scenarios. The training set contains 32,621 images of 1,041 pedestrians. The query set includes 11,659 images of 3,060 pedestrians, and the gallery set contains 82,161 test images corresponding to 3,060 identities.

4.1.2 Evaluation protocols

In our experiments, we employed mean average precision (mAP) and cumulative match characteristic (CMC) as evaluation metrics for assessing the performance of our model. We did not incorporate post-processing operations such as re-ranking or multi-query fusion during the testing process.

4.2 Implementation details

Our model is based on a ResNet50 [52] network with ImageNet [53] pre-trained weights. We used a mini-batch size of 32 for the Market1501 and DukeMTMC-reID datasets and a mini-batch size of 64 for the MSMT17 dataset. Additionally, we resized the pedestrian images to 328×128 and applied data augmentation techniques such as random flipping, random cropping and random erasure. During the training process, we initialized the learning rate to 3.5×10^{-4} and conducted a total of 80 epochs, with each epoch consisting of 400 iterations. The learning rate was reduced by a factor of 10 at the 40th and 70th epochs. When employing

the SGP strategy, we fixed all class labels and utilized a baseline model for 5 epochs of training. Prior to each training session, we employed the DBSCAN [50] clustering algorithm to cluster the feature vectors of all images and generate pseudo-labels. In the testing phase, we evaluated the model's performance using the f .

4.3 Comparison with state of the art

We compare our method with state-of-the-art USL methods on three datasets: Market1501 [23], DukeMTMC-reID [24] and MSMT17 [25]. The results are presented in Table 1. We include a list of USL methods, namely JVTC [30], SpCL [32], HCD [29], ICE [17], GCL [33], CACL [34] SECRET [35], PPLR [20], X.Han [36], HCM [37] and CACHE [38], which do not rely on any labeled data sources. To ensure fairness, we remove the startup camera labels. From Table 1, it is evident that our proposed method outperforms previous methods, achieving 84.3% mAP and 93.5% Rank-1 on the Market1501 dataset, and 72.6% mAP and 84.1% Rank-1 on the DukeMTMC-reID dataset. Moreover, our method demonstrates state-of-the-art performance on the challenging and difficult MSMT17 dataset, achieving 35.8% mAP and 65.0% Rank-1.

Our method surpasses previous state-of-the-art methods on all three datasets: Market1501, DukeMTMC-reID and MSMT17. Our method maximizes the utilization of multi-view information to optimize labels and mitigate label noise. In comparison with other state-of-the-art label smoothing methods such as PPLR and SECRET, our method achieves a significant increase of 2.8% and 3.5% in the mAP value on the Market1501 dataset, as well as an increase of 3.5% and 9.5% on the DukeMTMC-reID dataset. Our method also demonstrates superior performance on large datasets. The

Table 2 Comparative analysis of DCLR and constant label smoothing (CLS), SGP and LUPerson-NL Pre-trained Models, and the Combined DCLR Method with SGP Strategy for Person Re-Identification on Market1501, DukeMTMC-reID and MSMT17 Datasets

Different Individual Modules	Market1501			DukeMTMC-reID			MSMT17		
	mAP (%)	R1 (%)	R5 (%)	mAP (%)	R1 (%)	R5 (%)	mAP (%)	R1 (%)	R5 (%)
Baseline	78.8	90.4	96.7	68.1	80.7	90.2	21.5	47.0	60.6
+CLS	79.0	91.0	96.8	69.4	81.8	90.9	26.1	52.9	66.1
+DCLR	81.4	91.9	96.7	70.0	83.0	91.4	32.5	61.8	73.7
+LUPerson-NL	78.9	90.4	96.7	68.6	81.9	90.9	24.2	50.7	64.2
+SGP	80.1	91.4	97.0	70.0	82.0	91.7	26.6	53.2	66.5
+DCLR+SGP	83.0	93.3	97.5	71.6	83.8	92.2	34.3	64.0	75.7

Table 3 Ablation study on different combinations of data augmentation effects as aggregate features during model training in DCLR on the Market1501 and DukeMTMC-reID datasets

Aggregate selections					Market1501		DukeMTMC-reID	
	f	f^c	f^f	$f^f c$	mAP(%)	R1 (%)	mAP (%)	R1 (%)
✓	✓				79.9	90.3	69.6	81.8
✓		✓			80.9	91.8	69.3	82.4
✓			✓		80.5	91.6	69.6	82.5
✓	✓	✓	✓		81.4	91.9	70.0	83.0

The bolded parts indicate the optimal results among different ablation parameter configurations

MSMT17 dataset significantly affects the performance gap due to its large number of pedestrian categories and complex scenes. Our method employs the SGP pre-training strategy to mining the ETD classes in the dataset and guide model pre-training. This approach provides an advantage over methods that rely on ImageNet pre-training weights, effectively mitigating the impact of the performance gap. For instance, when compared to other state-of-the-art methods HCM and CACHE, our method exhibits significant improvements of 8.9% and 4.0% in the mAP value on the MSMT17 dataset.

4.4 Ablation study

In this subsection, we conducted ablation experiments to validate the effectiveness of the proposed DCLR method and the SGP strategy. When combining these two approaches, the model achieved significant performance gains. The results of these experiments are presented in Table 2. Furthermore, Table 3 displays the experimental results of different data augmentation methods. To assess the independence between the label distribution generated by clustering and the real labels, we analyzed the variation of the NMI index during the training process on the Market1501 dataset, as depicted in Fig. 3. Additionally, Fig. 4 illustrates the change in mAP value during the training process on the Market1501 dataset. Figure 5 presents the graph of ablation results for distance thresholds in the domain of the DBSCAN clustering algorithm with the SGP strategy. In order to evaluate the compactness of clustering, Fig. 6 showcases t-SNE [54] visualizations before and after applying the SGP strategy. Lastly,

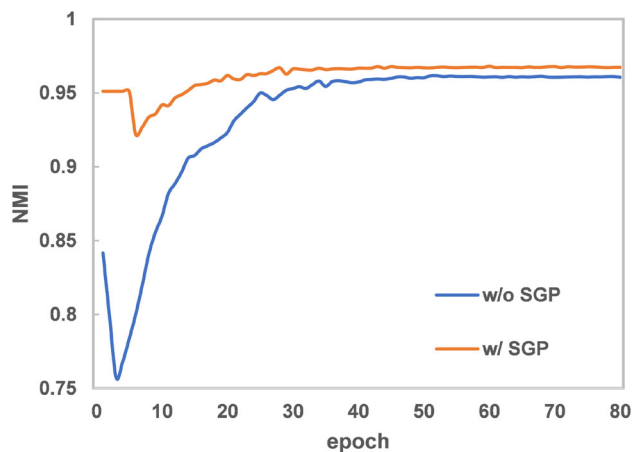


Fig. 3 Clustering quality over time during training on the Market1501 dataset, where larger numbers indicate a closer alignment between the distribution of clustered generated labels and the distribution of real labels

Fig. 7 provides visualizations of the results obtained by the Baseline and our method under the Rank-5 evaluation metric.

4.4.1 Effectiveness of DCLR

The performance of feature combinations obtained from different data augmentation effects on various datasets is presented in Table 3. During the training process, we observe that data augmentation techniques such as random cropping, flipping and combinations of flip and random crop are employed to enrich the source image with additional information. The

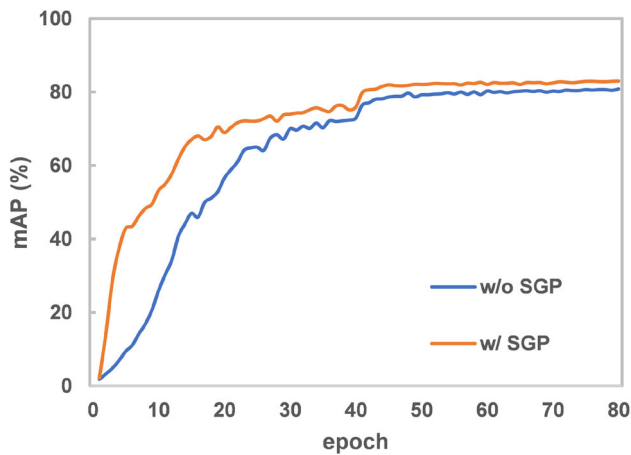


Fig. 4 Change in the mAP index with and without the SGP strategy during training on the Market1501 dataset

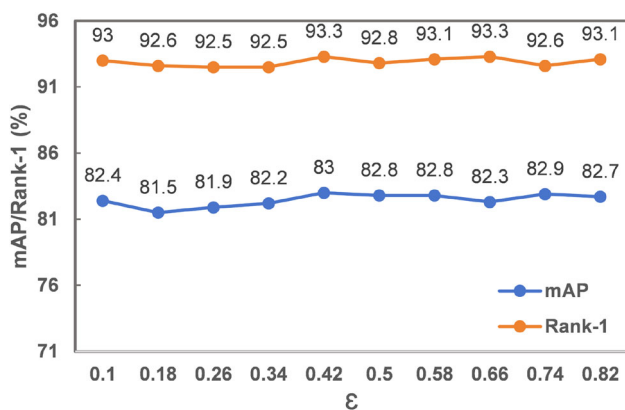


Fig. 5 Parameter analysis of ϵ in the DBSCAN clustering algorithm for the SGP strategy on the Market1501 dataset

aggregated features derived from these techniques are then utilized to optimize the label information, thereby effectively reducing labeling errors and improving noise resistance to a significant extent. Notably, employing all three data augmentation techniques simultaneously yields more stable results. To account for performance across multiple datasets, we select flip, crop and a combination of flip and random crop as the final settings for our model, utilizing these three data augmentations. Table 2 displays the results of ablation experiments using different label smoothing methods. From table 2, it is apparent that the DCLR method outperforms CLS in terms of performance. While CLS uniformly distributes label values and demonstrates limited effectiveness, the DCLR method proves to be more efficient in mitigating label noise. In comparison with the baseline, the DCLR method achieves 2.6% and 1.9% improvement in mAP on the Market1501 and DukeMTMC-reID datasets. This improvement surpasses the 0.2% and 1.3% mAP achieved by CLS.

To demonstrate the comprehensiveness and reliability of the superimposed aggregated features, we investigated the

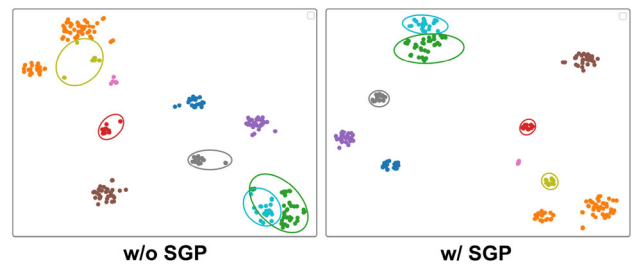


Fig. 6 The distribution of features extracted by the model in the embedding space under two different training strategies is visualized using t-SNE for data visualization. The data points are derived from a randomly selected sample of pedestrians encompassing 10 identities from the Market1501 dataset. Dots of the same color represent images of the identical identity. As illustrated in the figure, the utilization of the SGP strategy (right) enhances intra-class compactness and effectively discriminates misclassified samples from other classes (yellow rounded box) compared to the unused case (left)

effectiveness of various data augmentation methods, including flipping features f^f , cropping features f^c , combining flipping and cropping features f^{fc} , grayscale features f^g and erasing features f^e . Experimental verification was conducted to assess the effects resulting from the combination of multiple data augmentation schemes. During the model testing phase, we employed the aggregated features f' , and the results are presented in Table 4. It is evident that incorporating flipping, cropping and the combination of flipping and cropping with any data enhancement method significantly improves the model's performance. Notably, when all three data enhancement methods are combined, the performance of the model further enhances the mAP by 1.3% compared to the original test using only feature f . This improvement emphasizes the effective complementarity of our method in capturing source feature information and obtaining more comprehensive characteristic information. However, we also observed that data enhancement operations involving erasure and grayscale processing significantly diminish the information abundance. Erasure operations remove crucial local key information of pedestrians, while grayscale operations erase essential color information of pedestrians.

4.4.2 Effectiveness of SGP

Here we have used four evaluation dimensions to analyze the effectiveness of the method.

- (1) *Comparison with alternative pre-training weights* During training, we replaced the ImageNet pre-training weights with those provided by the latest person ReID pre-training method, LUPerson-NL [22]. The results are presented in Table 2, revealing that the LUPerson-NL pre-training weights yield only modest improvements in the model's performance compared to the training results



Fig. 7 Visualization of Baseline, DCLR Method, SGP Strategy and DCSG Framework under Rank-5 Evaluation. The images enclosed in green boxes represent correctly matched images, while those in red boxes indicate incorrectly matched images

Table 4 Performance of aggregated features f in model testing, obtained using different combinations of data augmentation on the Market1501 and DukeMTMC-reID datasets

Test feature selections (Avg(\cdot))	Market1501		DukeMTMC-reID	
	mAP (%)	R1 (%)	mAP (%)	R1 (%)
$[f^g]$	21.3	48.1	14.8	31.6
$[f^e]$	56.8	78.3	49.1	72.3
$[f^f]$	80.5	91.5	65.4	81.6
$[f^c]$	82.6	92.9	71.5	83.8
$[f^{fc}]$	82.9	92.9	71.3	84.1
$[f]$	83.0	93.3	71.6	83.8
$[f, f^g]$	77.8	91.5	65.0	80.5
$[f, f^e]$	79.3	91.6	67.5	82.4
$[f, f^f]$	84.0	93.6	71.5	83.8
$[f, f^c]$	83.5	93.2	72.1	84.1
$[f, f^{fc}]$	84.1	93.4	72.7	83.9
$[f, f^g, f^e]$	79.1	91.3	65.9	80.4
$[f, f^f, f^c, f^{fc}]$	84.3	93.5	72.6	84.1

The bolded parts indicate the optimal results among different ablation parameter configurations

obtained using ImageNet pre-training weights. In comparison with the aforementioned pre-training weights, the SGP method achieves substantial performance gains by mining ETD classes for training on the training set. For instance, on the Market1501 dataset, it enhances the model's mAP by 1.2% compared to using the LUPersonNL pre-training weights and by 1.3% compared to using the ImageNet pre-training weights.

- (2) *Normalized mutual information (NMI)* The NMI score measures the independence between the label assignments generated by clustering and the real labels, providing an indication of the reliability of the generated pseudo-labels. An NMI value of 1 signifies a strong correspondence between the clustering-generated labels and the real labels, while an NMI value of 0 indicates

their independence. Figure 3 illustrates the impact of the SGP training strategy. We enable the SGP strategy for the first 5 epochs, resulting in a slight drop in performance during the 6th epoch. This decline can be attributed to the exclusion of the majority of outliers during the SGP training strategy implementation, causing a temporary disturbance in the model's ability to distinguish features of the new samples. As a result, there is a transient decrease in the NMI value when merging clusters with the new samples. Additionally, we observe a consistent and continuous improvement in NMI and mAP throughout the subsequent performance, as depicted in Figs. 3 and 4. Notably, these performance metrics consistently outperform those obtained without utilizing the SGP method. By employing the SGP strategy, the

pre-training phase enables the acquisition of initial differentiation ability for pedestrian features, facilitating rapid performance growth and achieving superior results in subsequent training. This demonstrates the effectiveness of our pre-training strategy in maximizing the model's performance potential.

- (3) *Analysis of DBSCAN hyperparameters in the SGP stage*
We conduct an experimental evaluation of the distance threshold ε within the parameter domain of the DBSCAN clustering algorithm during the SGP strategy. The results are depicted in Fig. 5. The value of ε does not exert a significant influence on the overall model performance. However, to consider mAP and Rank-1, we set $\varepsilon = 0.42$.
- (4) *Visualization Analysis of Features*: We randomly selected identity samples from 10 individuals and compared the training process with and without the SGP strategy. The distribution of the model's extracted features in the embedding space under both training strategies is visualized in Fig. 6. Notably, the SGP strategy promotes cluster compactness within the model and effectively separates individuals with different and overlapping IDs. This observation confirms the effectiveness of our proposed method, which ultimately enhances the model's classification performance on pedestrian features.

To further investigate the effectiveness of our proposed method, we selected a subset of test samples to visualize the Rank-5 matching results from the image library, as depicted in Fig. 7. In these results, images with a green box represent correct matches, while those with a red box indicate incorrect matches. Notably, we observed a significant enhancement in the model's accuracy in identifying pedestrians by employing the DCLR method and the SGP strategy, particularly in matching the first few pedestrians. Furthermore, the combination of these two methods further improves the matching performance, enabling the model to effectively discriminate the fine-grained differences in pedestrian characteristics, as demonstrated in the DCSG results shown in Fig. 7.

5 Conclusions

In this study, we introduce a label smoothing scheme and a pre-training strategy. Our label smoothing scheme leverages the valid information from the multiple augmentation views and optimizes the labels accordingly to guide the model during training. Additionally, we propose a novel pre-training strategy. This strategy calculates the similarity distance between pairs of images by mining the frequency information of image occurrences within clustered classes during

the training process. It further utilizes a clustering algorithm to identify ETD classes that guide the pre-training of the model, mitigating the domain disparity caused by the pre-training weights of ImageNet. Furthermore, we conducted a variety of ablation experiments, including the visualization of experimental data, to validate the effectiveness of the proposed method. Extensive experiments conducted on benchmark datasets demonstrate that our method surpasses the performance of current state-of-the-art methods.

For future research, we intend to investigate diverse data supplementation approaches. We will explore the integration of a vision-language model to generate text information relevant to pedestrian images. This text information will be utilized to supplement features, thereby enhancing the robustness of the model by mitigating the noise impact of pseudo-labels. Additionally, we intend to introduce an adaptive distance scheme in our future work. This scheme will enable us to adjust to different model parameter ratios and enhance the accuracy of calibrating the distance between multiple samples.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant No. 62166026 and 62076117 and the Jiangxi Provincial Key Laboratory of Virtual Reality under Grant No.2024SSY03151.

Author Contributions Q.H. and J.C. conceived and designed the experimental ideas, with J.C. leading the writing of the source manuscript, implementing the experiments and creating Figures 1 and 2. Q.H. and W.M. managed and coordinated the overall experimental process, provided financial support, supervised the project and revised the manuscript. J.L. implemented specific experiments, visualized the experimental results and created Figures 3–7. L.Z. collected and organized the experimental data and created Tables 1–4. L.L. visualized certain experimental results and contributed to the manuscript revisions. All authors reviewed the manuscript.

Data availability The datasets utilized in this article are publicly available and free of charge, such as Market1501 [23], DukeMTMC-reID [24] and MSMT17 [25].

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

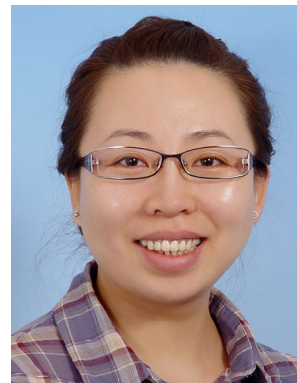
- Li, J., Chen, J., Sheng, B., Li, P., Yang, P., Feng, D.D., Qi, J.: Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network. *IEEE Trans. Industr. Inf.* **18**(1), 163–173 (2022). <https://doi.org/10.1109/TII.2021.3085669>
- Zhan, L.X., Li, W., Min, W.D.: Fa-ResNet: feature affine residual network for large-scale point cloud segmentation. *Int. J. Appl. Earth Obs. Geoinf.* (2023). <https://doi.org/10.1016/j.jag.2023.103259>
- Lin, X., Sun, S., Huang, W., Sheng, B., Li, P., Feng, D.D.: EAPT: efficient attention pyramid transformer for image processing. *IEEE*

- Trans. Multimed. **25**, 50–61 (2023). <https://doi.org/10.1109/TMM.2021.3120873>
4. Xie, Z., Zhang, W., Sheng, B., Li, P., Chen, C.L.P.: BaGFN: broad attentive graph fusion network for high-order feature interactions. *IEEE Trans. Neural Netw. Learn. Syst.* **34**(8), 4499–4513 (2023). <https://doi.org/10.1109/TNNLS.2021.3116209>
 5. Qin, Y., Chi, X., Sheng, B., Lau, R.W.H.: Guiderender: large-scale scene navigation based on multi-modal view frustum movement prediction. *Vis. Comput.* **39**(8), 3597–3607 (2023). <https://doi.org/10.1007/s00371-023-02922-x>
 6. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)
 7. Wei, L., Zhang, S., Yao, H., Gao, W., Tian, Q.: Glad: global-local-alignment descriptor for pedestrian retrieval. In: Proceedings of the 25th ACM International Conference on Multimedia, pp. 420–428 (2017)
 8. Sheng, B., Li, P., Ali, R., Chen, C.L.P.: Improving video temporal consistency via broad learning system. *IEEE Trans. Cybern.* **52**(7), 6662–6675 (2022). <https://doi.org/10.1109/TCYB.2021.3079311>
 9. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
 10. Li, J., Zhang, S., Tian, Q., Wang, M., Gao, W.: Pose-guided representation learning for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(2), 622–635 (2022). <https://doi.org/10.1109/TPAMI.2019.2929036>
 11. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
 12. Fu, Y., Wei, Y., Wang, G., Zhou, Y., Shi, H., Uiu, U., Huang, T.: Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6111–6120 (2019). <https://doi.org/10.1109/ICCV.2019.00621>
 13. Zhai, Y., Lu, S., Ye, Q., Shan, X., Chen, J., Ji, R., Tian, Y.: Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
 14. Jiang, N., Sheng, B., Li, P., Lee, T.-Y.: Photohelper: portrait photographing guidance via deep feature retrieval and fusion. *IEEE Trans. Multimed.* **25**, 2226–2238 (2023). <https://doi.org/10.1109/TMM.2022.3144890>
 15. Chen, Z., Qiu, G., Li, P., Zhu, L., Yang, X., Sheng, B.: MNGNAS: distilling adaptive combination of multiple searched networks for one-shot neural architecture search. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(11), 13489–13508 (2023). <https://doi.org/10.1109/TPAMI.2023.3293885>
 16. Yang, Z., Jin, X., Zheng, K., Zhao, F.: Unleashing potential of unsupervised pre-training with intra-identity regularization for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14298–14307 (2022)
 17. Chen, H., Lagade, B., Bremond, F.: Ice: inter-instance contrastive encoding for unsupervised person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14960–14969 (2021)
 18. Ge, Y., Chen, D., Li, H.: Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In: International Conference on Learning Representations (2020). <https://openreview.net/forum?id=rJlnOhVYPS>
 19. Han, J., Li, Y.-L., Wang, S.: Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 790–798 (2022)
 20. Cho, Y., Kim, W.J., Hong, S., Yoon, S.-E.: Part-based pseudo label refinement for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7308–7318 (2022)
 21. Fu, D., Chen, D., Bao, J., Yang, H., Yuan, L., Zhang, L., Li, H., Chen, D.: Unsupervised pre-training for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14750–14759 (2021)
 22. Fu, D., Chen, D., Yang, H., Bao, J., Yuan, L., Zhang, L., Li, H., Wen, F., Chen, D.: Large-scale pre-training for person re-identification with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2476–2486 (2022)
 23. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
 24. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Hua, G., Jégou, H. (eds.) *Computer Vision—ECCV 2016 Workshops*, pp. 17–35. Springer, Cham (2016)
 25. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
 26. Zhang, P.Y., Dou, H.Z., Yu, Y.L., Li, X.: Adaptive cross-domain learning for generalizable person re-identification. In: 17th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 13674, pp. 215–232 (2022). https://doi.org/10.1007/978-3-031-19781-9_13. <GotoISI>://WOS:000904096200013
 27. Zhang, X., Cao, J., Shen, C., You, M.: Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8221–8230 (2019). <https://doi.org/10.1109/ICCV.2019.00831>
 28. Peng, W., Chen, H., Li, Y., Sun, J.: MCFR: multi-confidence contrastive learning with feature refined for unsupervised person re-identification. *The Visual Computer* (2023). <https://doi.org/10.1007/s00371-023-02890-2>
 29. Zheng, Y., Tang, S., Teng, G., Ge, Y., Liu, K., Qin, J., Qi, D., Chen, D.: Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8371–8381 (2021)
 30. Li, J., Zhang, S.: Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *Computer Vision—ECCV 2020*, pp. 483–499. Springer, Cham (2020)
 31. Song, X., Jin, Z.: Robust label rectifying with consistent contrastive-learning for domain adaptive person re-identification. *IEEE Trans. Multimed.* **24**, 3229–3239 (2022). <https://doi.org/10.1109/TMM.2021.3096014>
 32. Ge, Y., Zhu, F., Chen, D., Zhao, R., Li, H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: *Advances in Neural Information Processing Systems* (2020)
 33. Chen, H., Wang, Y., Lagade, B., Dantcheva, A., Bremond, F.: Joint generative and contrastive learning for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2004–2013 (2021)
 34. Li, M., Li, C.-G., Guo, J.: Cluster-guided asymmetric contrastive learning for unsupervised person re-identification. *IEEE Trans. Image Process.* **31**, 3606–3617 (2022). <https://doi.org/10.1109/TIP.2022.3173163>
 35. He, T., Shen, L., Guo, Y., Ding, G., Guo, Z.: Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person

- re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 879–887 (2022)
36. Han, X.M., Yu, X.H., Li, G.R., Zhao, J., Pan, G., Ye, Q.X., Jiao, J.B., Han, Z.J.: Rethinking sampling strategies for unsupervised person re-identification. *IEEE Trans. Image Process.* **32**, 29–42 (2023). <https://doi.org/10.1109/tip.2022.3224325>
 37. Si, T., He, F., Zhang, Z., Duan, Y.: Hybrid contrastive learning for unsupervised person re-identification. *IEEE Trans. Multimed.* **25**, 4323–4334 (2023). <https://doi.org/10.1109/TMM.2022.3174414>
 38. Liu, Y., Ge, H., Sun, L., Hou, Y.: Complementary attention-driven contrastive learning with hard-sample exploring for unsupervised domain adaptive person re-id. *IEEE Trans. Circuits Syst. Video Technol.* **33**(1), 326–341 (2023). <https://doi.org/10.1109/TCSVT.2022.3200671>
 39. Hikmawati, E., Maulidevi, N.U., Surendro, K.: Minimum threshold determination method based on dataset characteristics in association rule mining. *J. Big Data* (2021). <https://doi.org/10.1186/s40537-021-00538-3>
 40. Aryabarzan, N., Minaei-Bidgoli, B.: Neclatclosed: a vertical algorithm for mining frequent closed itemsets. *Expert Syst. Appl.* **174**, 114738 (2021). <https://doi.org/10.1016/j.eswa.2021.114738>
 41. Li, H., Sheu, P.C.-Y.: A scalable association rule learning heuristic for large datasets. *J. Big Data* (2021). <https://doi.org/10.1186/s40537-021-00473-3>
 42. Gan, W., Lin, J.C.-W., Fournier-Viger, P., Chao, H.-C., Tseng, V.S., Yu, P.S.: A survey of utility-oriented pattern mining. *IEEE Trans. Knowl. Data Eng.* **33**(4), 1306–1327 (2021). <https://doi.org/10.1109/TKDE.2019.2942594>
 43. Luna, J.M., Padillo, F., Pechenizkiy, M., Ventura, S.: Apriori versions based on mapreduce for mining frequent patterns on big data. *IEEE Trans. Cybern.* **48**(10), 2851–2865 (2018). <https://doi.org/10.1109/tycb.2017.2751081>
 44. Kumar, S., Mohbey, K.K.: A review on big data based parallel and distributed approaches of pattern mining. *J. King Saud Univ. Comput. Inf. Sci.* **34**(5), 1639–1662 (2022). <https://doi.org/10.1016/j.jksuci.2019.09.006>
 45. Wu, J.M.-T., Srivastava, G., Wei, M., Yun, U., Lin, J.C.-W.: Fuzzy high-utility pattern mining in parallel and distributed Hadoop framework. *Inf. Sci.* **553**, 31–48 (2021). <https://doi.org/10.1016/j.ins.2020.12.004>
 46. Sumalatha, S., Subramanyam, R.B.V.: Distributed mining of high utility time interval sequential patterns using mapreduce approach. *Expert Syst. Appl.* **141**, 112967 (2020). <https://doi.org/10.1016/j.eswa.2019.112967>
 47. Lin, J.C.-W., Djenouri, Y., Srivastava, G.: Efficient closed high-utility pattern fusion model in large-scale databases. *Inf. Fusion* **76**, 122–132 (2021). <https://doi.org/10.1016/j.inffus.2021.05.011>
 48. Luna, J.M., Fournier-Viger, P., Ventura, S.: Frequent itemset mining: a 25 years review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **9**(6), 1329 (2019)
 49. Bashir, S.: An efficient pattern growth approach for mining fault tolerant frequent itemsets. *Expert Syst. Appl.* **143**, 113046 (2020). <https://doi.org/10.1016/j.eswa.2019.113046>
 50. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*, vol. 96, pp. 226–231 (1996)
 51. Zhong, Z., Zheng, L., Cao, D.L., Li, S.Z., Ieee: Re-ranking person re-identification with k-reciprocal encoding. In: 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3652–3661 (2017). <https://doi.org/10.1109/cvpr.2017.389>. <GotoISIpsdoiGtsps://WOS:000418371403078
 52. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
 53. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
 54. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(11) (2008)

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.



Qing Han received the B.E. and M.E. degrees in computer application from Tianjin Polytechnic University, China in 1997 and 2006, respectively. She is currently an Associate Professor of School of Mathematics and Computer Science, Nanchang University, China. Her current research interests include image and video processing, network management, virtual reality and metaverse.



Jiongjin Chen received his B.E. degree in network engineering from Huizhou University, China in 2021. He is currently pursuing the M.E. degree at Nanchang University, China. His current research interests include computer vision.



Weidong Min received the B.E., M.E. and Ph.D. degrees in computer application from Tsinghua University, China in 1989, 1991 and 1995, respectively. He is currently a Full Professor of School of Mathematics and Computer Science, Nanchang University, China, the Dean of Institute of Metaverse, Nanchang University, China, and an Executive Director of China Society of Image and Graphics. His current research interests include image and video processing, artificial intelligence, big data,

virtual reality and metaverse.



Jiahao Li received his B.E. degree in information and computing science from Shandong University of Technology, China, in 2020 and M.E. degree in computer science and technology from Nanchang University, China, in 2023. He is currently pursuing the Ph. D. degree at Nanchang University, China. His current research interests include computer vision.



Lixin Zhan obtained the B.E. degree in Computer Science and Technology from Hunan First Normal University, China, in 2021. He is currently pursuing a master's degree in computer technology at Nanchang University, China. His research interests include 3D computer vision and deep learning.



Longfei Li received his B.E. degree in software engineering from Jiangxi Normal University, China in 2021. He is currently pursuing the M.E. degree at Nanchang University, China. His current research interests include computer vision.