ORIGINAL ARTICLE

Hybrid feature constraint with clustering for unsupervised person re-identification

Tongzhen Si[1] · Fazhi He[1] · Penglei Li[2]

Abstract
Unsupervised person re-identification (Re-ID) has better scalability and usability in real-world deployments due to the lack of annotations, which is more challenging than supervised methods. State-of-the-art approaches mainly employ clustering algorithms to generate pseudo-labels for transferring the process into a supervised operation. However, the clustering algorithm depends on discriminative pedestrian features. Only using the clustering algorithm produces low-quality labels and hinders the performance of the Re-ID model. In the paper, we propose the hybrid feature constraint network (HFCN) to adequately restrict the pedestrian feature distribution for unsupervised person Re-ID. Specifically, we first define a feature constraint loss to restrict the feature distribution so that different pedestrians can be clearly distinguished at the first step. And then, we design a multi-task operation with the iterative update for clustering algorithm to further implement the feature constraint. This can adequately utilize predicted label information and identify complex samples. Finally, we integrate the feature constraint loss and multi-task operation to optimize the Re-ID model, which could promote the clustering to generate high-quality labels and learn valuable information. Extensive experiments prove that the proposed HFCN is effective and outperforms the state-of-the-art.

Keywords Person re-identification · Unsupervised learning · Feature constraint · Clustering

1 Introduction

Person re-identification (Re-ID) is to search for target pedestrians according to given query images across non-overlapping cameras. It has attracted widespread attention due to its wide applications in cross-camera tracking, video surveillance, etc [1–3]. Encouraged by deep learning, supervised Re-ID has obtained great success [4–7]. However, these methods will have a significant drop when they face a new database. In addition, supervised Re-ID requires abundant labeled pedestrian data, and its cost is expensive. Hence, supervised Re-ID is difficult to satisfy the requirement of real-world scenarios.

To alleviate the scalability challenge, researchers attempt to study unsupervised methods. Most of them utilize the labeled source domain and unlabeled target domain to optimize deep models, treated as unsupervised domain adaptation (UDA) methods. Some works leverage GAN to transform the style from the source domain to the target domain for bridging the domain gap [8,9]. Some other methods treat each image as an individual identity and learn the pedestrian invariance [10,11]. In addition, several state-of-the-art methods employ clustering algorithms to generate pseudo-labels and translate them into a supervised learning problem [12,13]. But, UDA requires leveraging the prior knowledge to optimize Re-ID models and lacks flexibility. It is closely related to the domain gap and needs to select a suitable source domain.

Fully unsupervised person Re-ID aims to mine robust features by no labeled image data. A few existing methods learn the softened label distribution or design the multi-label classification loss to optimize deep models [14,15], which obtains the significant improvement. But, these methods are difficult to identify pedestrians with significant variation due to lack of effective supervision information. Inspired by UDA, some approaches employ clustering algorithms to produce pseudo-labels. However, the clustering algorithm depends on

✉ Fazhi He
  fzhe@whu.edu.cn

[1] School of Computer Science, Wuhan University, Bayi Road, Wuhan 430072, China

[2] Institute of Geophysics and Geomatics, China University of Geosciences, Lumo Road, Wuhan 430074, China

(a) Original Feature Distribution **(b)** Feature Constraint **(c)** Optimized Result
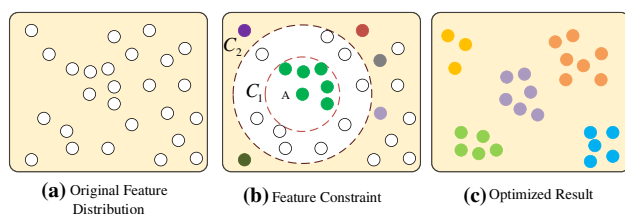
Fig. 1 **a** Original feature distribution is represented as different circles. **b** The optimization method for example A is presented. The near-neighbor similar features are obtained inside circle C_1, and the top-n dissimilar features are extracted outside circle C_2. The similarity of these near-neighbor features is further enhanced, and the similarity of the top-n features is further reduced. **c** The optimization result is visualized by our optimization strategy

discriminative pedestrian features. As shown in Fig. 1a, the initial feature distribution has poor pedestrian information. Simply utilizing the clustering generates low-quality labels and cannot learn effective features. They still have a significant performance gap with supervised learning or UDA methods.

In the paper, we propose a novel Hybrid Feature Constraint Network (HFCN) to adequately implement the feature constraint and provide meaningful supervision information without any labeled identities for the Re-ID model. To this end, we define the feature constraint loss to force near-neighbor features to have high similarity and decrease the similarity among different pedestrian features as shown in Fig. 1b. As a result, pedestrian features with the same identity have the similar feature distribution and different pedestrians possess large distribution gaps as shown in Fig. 1c. Afterward, we design the multi-task operation with the iterative update for clustering algorithm to adequately employ predicted labels and identify complex samples. Finally, we integrate the feature constraint loss and multi-task operation to jointly optimize the Re-ID model. This could stimulate the clustering algorithm to produce high-quality labels, and in turn, effective labels provide effective supervision signals for the deep model. The paper has the following contributions:

(1) We define the feature constraint loss to restrict the feature distribution so that pedestrian features with the same identity have the similar distribution and different pedestrians possess large distribution gaps.

(2) The proposed HFCN successfully integrates the feature constraint loss and multi-task operation process. This facilitates the clustering algorithm to generate high-quality labels and provides effective supervision information for the Re-ID model.

(3) Extensive experiments prove that our method is valuable for fully unsupervised Re-ID and exceeds the state-of-the-art methods on mainstream Re-ID databases.

2 Related work

2.1 Supervised person Re-ID

With the renaissance of neural networks [16–21], deep learning has become the mainstream technology and dominated the Re-ID community to optimize objective features [22,23]. Most Re-ID methods mainly focus on supervised learning by obtaining abundant annotated data [24–29]. For extracting discriminative features, some methods explore the similarity relationship among pedestrian images and constrain the spatial distribution of features. For example, Chen et al. [30] take into account the local similarity and dependencies of global features to optimize the deep model. Luo et al. [25] construct sample pairs according to pedestrian identities and employ the triplet loss and center loss to constrain feature distributions.

The above methods suffer significant performance degradation while applied to a new domain. In our paper, we abandon annotated pedestrian labels to focus on unsupervised person Re-ID. We explore the feature distribution among different pedestrians to mine meaningful pedestrian information.

2.2 Unsupervised domain adaptation

Unsupervised domain adaptation (UDA) methods leverage the labeled source domain data and unlabeled target domain data to optimize deep models [31–33]. Some works implement the style transformation operation from one scenario to the other for aligning the feature distribution. Wei et al. [8] design an adversarial network to transfer the pedestrian style from one database to the other database so as to decrease the domain gap. Liu et al. [31] propose the Adaptive Transfer Network (ATNet) to learn three critical styles (illumination, resolution, and camera) for addressing the domain discrepancy. Some approaches study the intra-domain invariance to mine robust features. For example, Zhong et al. [10] consider each image as a class and require the nearest neighbor of each image to have the same label by assigning weights. In [34], GAN is utilized to generate positive training pairs by image-image translation and then employ the triplet loss to constrain the intra-domain variation. In addition, several clustering-based methods assign pseudo-labels to pedestrian images and achieve impressive performance. For instance, Ge et al. [35] mitigate the effect of predicted false labels by designing the mutual mean-teaching framework. Zhai et al. [12] pre-trains several different networks based on clustering algorithms and integrate these models to implement a mutual learning process.

In addition, Ge et al. [36] propose a self-paced contrastive learning strategy by designing a hybrid memory to encode available information from both source and tar-

get domains. Differently, the hybrid memory method solves the inconsistency problem between the source and target domains and effectively utilizes outliers in the clustering process. To this end, they design a hybrid memory to encode available information from training samples for feature learning. Our method resolves the feature distribution problem, which stimulates the clustering algorithm to generate effective labels. We define the feature constraint loss to restrict the feature distribution and then integrate the multi-task operation with clustering to jointly optimize the Re-ID model.

In the optimization process, the hybrid memory method builds the memory to store clustering features and employs the contrastive loss to learn the identity information, while our method directly employs the classifier to identify clustering-based labels and perform the multi-task operation to mine pedestrian features. In addition, they design a self-paced contrastive learning strategy to prevent training error amplification, while our method defines the feature constraint loss to constrain the feature distribution for promoting the clustering algorithm. Furthermore, as for all instance features, the hybrid memory method minimizes the negative log-likelihood over feature similarities, while we utilize the specificity of cosine similarity to maximize the feature similarity. In conclusion, our method is completely different from the existing hybrid memory [36]. Our method and the existing hybrid memory possess different motivations and implementation processes.

These methods mentioned above obtain impressive performance in the UDA field. However, they require sufficient labeled source domain data. Otherwise, these methods will have significant performance degradation. In contrast, we explore the fully unsupervised Re-ID method by effectively constraining the feature distribution, which achieves superior performance and exceeds many UDA methods.

2.3 Unsupervised person Re-ID

The conventional hand-craft features do not employ any external information [37–39]. But, they are difficult to extract robust features, and their performance suffers a significant degradation when dealing with large databases. Recently, several deeply learned approaches have been proposed to stimulate the performance of unsupervised person Re-ID. Lin et al. [40] utilize a clustering algorithm to predict pedestrian labels and iteratively train the Re-ID model. Nevertheless, the clustering may generate many wrong labels and affect the model's generalization capability. On the other hand, since the camera-ID can be directly employed, some works utilize camera-ID to finish unsupervised Re-ID tasks and achieve impressive performance. For example, Lin et al. [15] propose the cross-camera encouragement to learn the potential feature similarity, which has a significant improvement. Wang et al. [14] apply GAN to generate pedestrian images with

different styles and design multiple labels for each image. However, the two methods easily cause performance degradation when pedestrian images have large variations due to lack of effective supervision signals.

In the paper, we define the feature constraint loss to restrict the feature distribution and then integrate the multi-task operation with clustering to adequately employ generated pseudo-labels. The strategy could promote the clustering algorithm to generate high-quality labels, which provides effective supervision information for the Re-ID model.

3 Proposed method

The study aims to restrict the feature distribution of different pedestrians for solving the unsupervised person Re-ID challenge. Given an unlabeled pedestrian database $X = \{x_1, x_2, x_3, \ldots, x_N\}$, where N is the total number of samples, our method could extract robust features to match the same pedestrian. We first introduce the whole structure in Sect. 3.1 and then explain the feature constraint loss in Sect. 3.2. Afterward, we illustrate the multi-task operation process in Sect. 3.3. Finally, we present the final optimization objective in Sect. 3.4.

3.1 The framework of HFCN

In the paper, we employ the ResNet-50 [41] as the backbone to extract image features, and the whole framework is shown in Fig. 2. After the final pooling operation, we remove the original fully connected layer and extract corresponding pedestrian features. Afterward, we add two branches to optimize the deep model. Specifically, we define the feature constraint loss to restrict the feature distribution for distinguishing different pedestrians in the first stage in the upper branch. After m epochs, the multi-task operation is implemented to further optimize pedestrian features with clustering algorithm for identifying complex samples in the below branch. We integrate the two branches to jointly train the Re-ID model and update pedestrian labels at each epoch.

For producing pedestrian labels, the density-based DBSCAN [42,43] is utilized to assign pseudo-labels for all samples. According to predicted labels, we randomly select P identities, and each identity has K pedestrian images as the input. So, each batch contains PK pedestrian images. As for the input, we implement the random horizontal flipping, random cropping, and random erasing [44] data enhancement methods. Since the camera-ID can be utilized directly, Camstyle [45] is leveraged to increase the variety of camera styles.

In the test stage, we employ the trained Re-ID model to process pedestrian images and obtain corresponding features behind the pooling layer. We compute the Euclidean distance
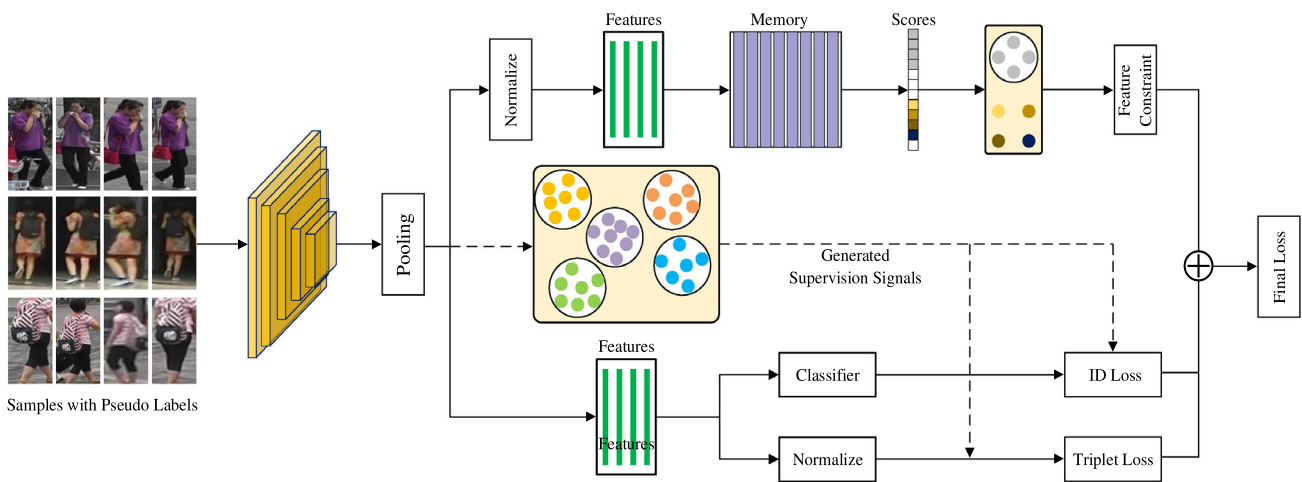
Fig. 2 The whole framework of the proposed HFCN. We extract image features and calculate the similarity between them. The feature constraint loss is defined to restrict the feature distribution. Meanwhile, we design the multi-task operation with the iterative update for the clustering algorithm to learn robust features. We integrate the feature constraint loss and the multi-task operation to jointly optimize the Re-ID model and update pedestrian labels. The solid line represents the operation in each iteration, and the dotted line denotes the process on the entire database

to match the same pedestrian, which obtains a remarkable improvement.

3.2 Feature constraint loss

Fully unsupervised person Re-ID is a challenging technology because there are no annotated identities. Several state-of-the-art methods apply clustering algorithms to produce pseudo-labels and obtain impressive performance [35,46]. Due to large variances in cameras, poses, backgrounds, etc., pedestrian images have a large distribution discrepancy as shown in Fig. 1a. Since effective clustering depends on discriminative pedestrian features, these existing methods are easy to produce low-quality labels by directly employing clustering algorithms and affect the performance of person Re-ID.

Pedestrian images in the person Re-ID field generally have large background changes and relatively small foreground variations. It means that the foreground region contains the identifiable semantic feature and the foreground feature has a strong correlation, which is an essential factor for solving unsupervised Re-ID. In order to explore the potential correlation, we define the effective feature constraint loss to restrict the feature distribution without any external information. The strategy forces image features with the same identity to possess the similar distribution and different pedestrians to have large distribution gaps. More importantly, the strategy could stimulate clustering algorithms to generate high-quality labels by distinguishing various pedestrians and extract valuable information for unsupervised person Re-ID.

Specifically, we compute the feature similarity and sort these similarity scores for each pedestrian image. We set the threshold v to filter high similarity features, treated as near-neighbor features. To this end, we construct an exemplar memory M to store pedestrian features, where $M[i]$ represents the feature of the i-th pedestrian image. In the initialization, we assign all values to be 0 in M. During the iteration, we forward each batch through the Re-ID model and extract the L2-normalized feature. In the backpropagation stage, M is updated:

$$M[i] \leftarrow \theta f_i + (1 - \theta) \cdot M[i], \tag{1}$$

where $\theta \epsilon [0, 1]$ is the updating rate, f_i denotes the current normalized feature, and $M[i]$ is then L2-normalized via $M[i] \leftarrow ||M[i]||_2$. So, the similarity between f_i and $M[j]$ is defined as:

$$s_{ij} = f[i] \times M[j]^T, \tag{2}$$

We compute the similarity between the feature f_i and all other features. Afterward, we rank the similarity to get the corresponding sort score:

$$S_i = \underset{j}{\arg \operatorname{sort}} (s_{ij}), \; j \epsilon [1, N], \tag{3}$$

where N denotes the total number of pedestrian images.

In the beginning, the initial model has poor identification capability, and initial features stored in M have poor discriminability. These features have a large distribution gap, so it

is difficult to search for the same pedestrian properly. Nevertheless, it is easy to search for different pedestrians because there are many negative samples in M. Therefore, we first treat each image as an individual identity by constraining the feature similarity between f_i and $M[i]$. Simultaneously, we force dissimilar features to have low similarity. As for dissimilar samples, pedestrian features with high similarity scores are considered hard negative sample features, while pedestrian features with low similarity scores are treated as easy negative sample features. Obviously, when hard negative sample pairs possess large distribution gaps, easy negative sample pairs naturally produce a larger difference. Thus, we select n difficult scores S_i for the feature f_i and obtain the similarity score set, i.e., V_i^n. As for any two features, their cosine similarity is the matrix inner product of two features after L2 normalization. Since pedestrian features stored in M are L_2 normalized, the feature similarity s_{ij} is restricted between $[-1, 1]$. Therefore, the same pedestrian features should have high similarity with the highest similarity close to 1, while different pedestrian features have low similarity with the lowest similarity reaching -1. So, the feature constraint loss is defined as:

$$L_{sc}^i = \eta(1 - s_{ii})^2 + \frac{1}{n} \sum_{s_{iv} \in V_i^n} (1 + s_{iv})^2, \qquad (4)$$

where η controls the importance of reliable features, s_{ii} indicates the similarity of current epoch feature and last updated feature $M[i]$, and s_{iv} denotes the similarity score between the feature f_i and f_v in V_i^n.

After m epochs, pedestrian features achieve a certain discriminant ability. We search for reliable samples and force them to have the similar feature distribution. Intuitively, reliable samples have high similarities. Hence, we set the threshold v to filter low similarity score in S_i, and then obtain a reliable score set U_i^v. Hence, the feature constraint loss is escalated to:

$$L_{sc}^i = \frac{\eta}{T} \sum_{s_{iu} \in U_i^v} (1 - s_{iu})^2 + \frac{1}{n} \sum_{s_{iv} \in V_i^n} (1 + s_{iv})^2, \qquad (5)$$

where T is the total number of selected reliable samples and s_{iu} denotes the similarity score between the feature f_i and f_u in U_i^v. So, the loss value within one batch could be written as:

$$L_{sc} = \frac{1}{PK} \sum_{i=1}^{PK} L_{sc}^i. \qquad (6)$$

By implementing the feature constraint, near-neighbor features generate high similarity; otherwise, they have a low similarity. Approximatively, image features with the same

identity are forced to have the similar feature distribution, and different pedestrians possess larger distribution gaps. This promotes the Re-ID model to distinguish different pedestrians without any external labels and learn meaningful features. Furthermore, the strategy could effectively facilitate clustering algorithms to predict pedestrian identities.

3.3 Multi-task operation

The feature constraint loss encourages pedestrian features with the same identity to have high similarity and forces different pedestrians to possess low similarity. The effect could stimulate the deep model to identify different pedestrians. With the promotion of the feature distribution, the clustering algorithm can generate valuable label information. To this end, we design the multi-task operation to adequately employ generated labels for distinguishing different pedestrians. Specifically, the label smoothing regularization loss is utilized to perform the classification task, while the hard triplet loss is employed to accomplish the ranking task, which can improve the generalization ability of the Re-ID model.

Since the number of pedestrian labels is uncertain, we employ DBSCAN [42,43] to predict pedestrian identities. In each epoch, the clustering algorithm generates a variable number of labels. Thus, we create a new classifier depending on the number of generated identities in each epoch to perform the classification task. We calculate the mean feature of each cluster and implement the normalization operation to initialize the classifier.

Similarly, we extract pedestrian features behind the pooling operation and then feed them into the classifier. Afterward, we employ the label smoothing regularization loss to calculate the loss of feature f_i:

$$L_{id}^i = -(1 - \varepsilon) \log(q(r \mid f_i)) - \frac{\varepsilon}{Z} \sum_{\substack{z=1 \\ z \neq r}}^{Z} \log(q(z \mid f_i)), \qquad (7)$$

where ε is set to 0.1 that denotes the hyperparameter, $q(z \mid f_i)$ is the predicted probability that the feature f_i belongs to the z-th label, Z is the total number of predicted labels, and r is the corresponding prediction label. So, the loss value within one batch is formulated as:

$$L_{id} = \sum_{i=1}^{PK} L_{id}^i. \qquad (8)$$

In the optimization process, hard samples are difficult to be effectively distinguished. The hard triplet loss could decrease

the distance among the same pedestrian features and increase the distance among different pedestrians. It is widely used in supervised methods to perform the similarity measurement learning operation and obtains the stimulative effect [47–49]. In the paper, we intuitively employ the hard triplet loss to unsupervised person Re-ID based on generated labels to implement the ranking task. The loss could further constrict the feature distribution to identify hard samples and in turn, it facilitates the clustering to produce effective labels. The loss value within one batch is written as:

$$L_t = \sum_{i=1}^{PK} [\psi + D_{ip} - D_{ib}]_+. \tag{9}$$

where ψ denotes the margin, D_{ip} represents the maximum distance between the i-th feature and the hardest positive sample feature within one batch, D_{ib} is the minimum distance between the i-th feature and the hardest negative sample feature within one batch, and $[w]_+ = \max(w, 0)$.

In a word, the label smoothing regularization loss could make full use of predicted label information, while the hard triplet loss increases the inter-class variation and decreases the intra-class variation. The two loss functions enable the deep model to easily identify hard samples, which effectively improves the performance of the Re-ID model.

3.4 Final objective

On the one hand, we define the feature constraint loss to learn the potential feature correlation, which enforces pedestrian features with the same identity to have high similarity and different pedestrians to possess low similarity. This not only promotes the model to learn robust features but also facilitates the clustering algorithm to generate high-quality labels. On the other hand, we design the multi-task operation with the iterative update to make full use of predicted labels and distinguish hard pedestrian samples. The final objective function is formulated as:

$$L = L_{sc} + \alpha L_{id} + \beta L_t. \tag{10}$$

where α and β denote the importance of the label smoothing regularization loss and the hard triplet loss, respectively.

We integrate the feature constraint loss and multi-task operation to optimize the Re-ID model with iteratively updated pedestrian labels. The iterative optimization strategy in each epoch promotes the clustering algorithm to generate high-quality labels, and in turn, these labels stimulate the Re-ID model to learn potential valuable information. The proposed HFCN substantially improves the performance of the Re-ID model and exceeds the state-of-the-art methods.

4 Experiments

4.1 Databases and implementation details

Our method is evaluated on three mainstream person Re-ID databases: **Market1501** [39], **DukeMTMC-reID** [2] and **MSMT17** [8]. Two typical assessment criteria, i.e., rank-1 and mAP, are utilized to evaluate the performance of the algorithm.

In the paper, we employ the ResNet-50 [41] as the backbone due to its competitive advantages. We set P to 32 and K to 4, resulting in 128 pedestrian images within one batch. We crop each pedestrian image into 256×128. The total number of epoch is 60. We set the learning rate to 0.1 and change it to 0.01 after 40 epochs. We employ SGD to optimize the Re-ID model. The parameter θ in Eq. 1 starts from 0 increases linearly to 0.5. The threshold ψ in Eq. 9 is set to 0.5. Several other important parameters are evaluated in Sect. 4.4.

4.2 Ablation study

We systematically assess the effectiveness of each component in HFCN by performing the ablation study, and the results are reported in Table 1.

Direct transfer denotes the deep model pre-trained on ImageNet [66] is directly applied to the person Re-ID field. This has a poor performance and illustrates the pre-trained model

Table 1 Ablation study on three person Re-ID databases

Methods	Market1501		DukeMTMC-reID		MSMT17	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Direct transfer	7.8	2.6	6.1	1.9	3.3	0.7
$Invariance$	63.5	30.5	48.5	23.8	10.2	3.1
$Invariance + L_{id} + L_t$	74.5	44.1	61.1	35.9	20.6	6.2
$Triplet + L_{id} + L_t$	83.4	61.3	72.4	53.1	50.2	18.6
Ours w/L_{sc}	61.0	30.8	47.9	26.5	8.2	2.0
Ours w/$L_{id} + L_t$	74.2	48.5	64.7	42.0	42.1	12.9
Ours w/$L_{sc} + L_{id} + L_t$	**88.7**	**69.3**	**75.5**	**56.6**	**52.3**	**21.9**

Bold values indicate the best results

contains very little pedestrian information, which can not provide meaningful features for the clustering algorithm.

L_{sc} indicates the feature constraint loss is utilized to explore the potential pedestrian correlation. The strategy surpasses direct transfer by 53.2% rank-1, 28.2% mAP on Market1501 and 41.8% rank-1, 24.6% mAP on DukeMTMC-reID. It also surpasses direct transfer by 4.9% rank-1, 1.3% mAP on MSMT17. This proves the feature constraint loss is valuable for restricting the feature distribution. More importantly, the method could promote the pedestrian features with the same identity to have the similar distribution and different pedestrians to possess large distribution gaps.

$L_{id} + L_t$ represents that the clustering algorithm is directly applied to predict pedestrian labels. Afterward, the classification and ranking tasks are performed, which reveals a certain performance. Nevertheless, the effect is not satisfying for us. This is because that simple clustering cannot produce high-quality label information due to lack of effective constraint, limiting the effect of unsupervised person Re-ID.

Next, we add L_{sc} to $L_{id} + L_t$ to validate the effect of our final HFCN, i.e, $L_{sc} + L_{id} + L_t$. We can see that rank-1 and mAP are improved by 14.5% and 20.8% on Market1501 compared with $L_{id} + L_t$. Meanwhile, the improvement on DukeMTMC-reID is 10.8% and 14.6% on rank-1 and mAP, respectively. The significant advantage proves that predicted labels possess the high quality. Our feature constraint loss could provide effective constraint to restrict the feature distribution. In addition, this also verifies we effectively fuse the feature constraint loss and the multi-task operation to mine robust pedestrian features.

To further analyze HFCN, we compare the invariance loss presented in ECN [10], which is expressed as *Invariance*. ECN constructs a memory to save all pedestrian features and assigns near-neighbor features to have the same identity. The invariance loss exhibits a certain performance for the Re-ID model. However, we replace our feature constraint loss with the invariance loss to integrate the multi-task operation, i.e., $Invariance + L_{id} + L_t$, which reflects worse performance than that of the proposed HFCN. Even the invariance loss makes the performance of multi-task operation degrade, especially in the two large-scale person Re-ID databases, i.e., DukeMTMC-reID and MSMT17. This is because the invariant loss sets hard labels for each pedestrian, which is inconsistent with the multi-task operation. Our method is superior $Invariance + L_{id} + L_t$ by **14.2% rank-1**, **25.2% mAP** on Market1501 and **14.4% rank-1**, **20.7% mAP** on DukeMTMC-reID. In addition, HFCN surpasses $Invariance + L_{id} + L_t$ over **31.7% rank-1**, **15.7% mAP** on MSMT17. The significant advantage proves the proposed feature constraint loss can stimulate the clustering algorithm to generate high-quality labels. The proposed HFCN could effectively integrate the feature constraint loss and the multi-task operation to mine potential pedestrian features.

In addition, we also replace our feature constraint loss with the triplet loss at the first stage to prove its effectiveness, i.e., $Triplet + L_{id} + L_t$. Obviously, the performance is inferior to our final results. This is because the triplet loss mainly employs pseudo-labels to constrain the feature space. The feature constraint loss only utilizes the feature similarity to constrain the feature distribution, which stimulates the clustering algorithm to produce high-quality labels. The feature constraint loss is complementary to the multi-task operation. This further demonstrates that our feature constraint loss is valuable for unsupervised person Re-ID.

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

We compare the proposed HFCN with other competitive unsupervised person Re-ID methods on Market1501 and DukeMTMC-reID, and the compared results are reported in Table 2.

We divide these competing approaches into three categories. The first category employs hand-craft feature methods. The second category adopts deep learning methods for unsupervised person Re-ID, and the third type utilizes the labeled source domain to implement UDA tasks. From Table 2, deep learning methods obtain higher accuracy than the two hand-crated features. This is because neural networks can learn discriminative features by adapting to various situations.

More importantly, our method obtains the best performance and outperforms all other unsupervised methods by a large margin. The proposed HFCN achieves **rank-1 = 88.7%**, **mAP = 69.3%** on Market1501. Meanwhile, it also obtains **rank-1 = 75.5%** and **mAP = 56.6%** on DukeMTMC-reID. The impressive performance proves that the proposed HFCN successfully integrates the feature constraint loss and multi-task operation, which could promote the clustering algorithm to produce high-quality labels. The method could adequately utilize predicted labels and identify different pedestrians. All results demonstrate that the proposed HFCN is effective and could extract discriminative features without any annotated labels.

BUC [40] and HCT [52] apply the clustering method to fully unsupervised person Re-ID. From Table 2, the proposed HFCN outperforms them by a large margin, which proves our method is valuable. HCT also utilizes the hard triplet loss to optimize the Re-ID model. Our method is still superior HCT by **8.7% rank-1**, **12.9% mAP** on Market1501 and **5.9% rank-1**, **5.9% mAP** on DukeMTMC-reID. The distinct advantage demonstrates the proposed HFCN is a meaningful algorithm.

Furthermore, even though UDA methods leverage the external source domain data, our method still exceeds them. This further proves HFCN is an innovative approach for unsupervised Re-ID, which could extract discriminative features

Table 2 Comparison with other state-of-the-art methods on Market1501 and DukeMTMC-reID. In column "Setting", "U" denotes fully unsupervised methods and "UDA" indicates utilizing external labeled data

Methods	Setting	Market1501				DukeMTMC-reID			
		Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
LOMO [38]	U	27.2	41.6	49.1	8.0	12.3	21.3	26.6	4.8
BOW [39]	U	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
FedUReID [50]	U	65.2	77.8	82.2	34.2	51.0	62.4	67.6	29.5
BUC [40]	U	66.2	79.6	84.5	38.3	47.4	62.6	68.4	27.5
JVTC+ [51]	U	79.5	89.2	91.9	47.5	74.6	82.9	85.3	50.7
HCT [52]	U	80.0	91.6	95.2	56.4	69.6	83.4	87.4	50.7
MMCL [14]	U	80.3	89.4	92.3	45.5	65.2	75.9	80.0	40.2
STAR-DAC [53]	U	67.0	80.6	84.9	33.9	56.4	72.1	76.5	31.6
SCL [54]	U	82.2	89.9	92.6	54.4	69.9	79.7	82.2	47.2
KNNL [55]	U	82.6	91.3	94.5	54.1	68.5	79.0	82.8	45.1
MPRD [56]	U	83.0	91.3	93.6	51.1	67.4	78.7	81.8	43.7
DSCE [57]	U	83.9	84.2	–	53.8	73.8	84.2	–	53.8
MLC [58]	U	86.7	93.5	95.6	66.2	73.6	82.3	85.5	52.3
HFCN (Ours)	**U**	**88.7**	**94.7**	**96.4**	**69.3**	**75.5**	**85.6**	**89.0**	**56.6**
ECN [10]	UDA	75.1	87.6	91.6	43.0	63.3	75.8	80.4	40.4
D-MMD [59]	UDA	70.6	87.0	91.5	48.8	63.4	78.8	83.9	46.0
ADTC [60]	UDA	79.3	90.8	94.1	59.7	71.9	84.1	87.5	52.5
ACT [61]	UDA	80.5	–	–	60.6	72.4	–	–	54.5
GDS-H [62]	UDA	81.1	–	–	61.2	73.1	–	–	55.1
pMR-SADA [63]	UDA	83.0	91.8	94.1	59.8	74.5	85.3	88.7	55.8
MMCL [14]	UDA	84.4	92.8	95.0	60.4	72.4	82.9	85.0	51.4
AD-Cluster [46]	UDA	86.7	94.4	96.5	68.3	72.6	82.5	85.5	54.1
TAL-MIRN [64]	UDA	73.1	86.3	–	40.0	63.5	76.6	–	41.3
PREST [65]	UDA	81.1	–	–	60.5	72.7	–	–	54.1
MLC [58]	UDA	85.6	93.9	96.0	65.9	74.1	83.8	86.3	55.0

Bold values indicate the best results

and markedly boost the generalization ability of the Re-ID model.

MSMT17 is the most challenging database in the person Re-ID field. We also verify our method on MSMT17, and the results are listed in Table 3. From the Table, the proposed HFCN achieves **rank-1 = 52.3%** and **mAP = 21.9%**, which outstands others competitive approaches. Although UDA methods employ different external data, our method still has a competitive advantage. This proves HFCN can generate high-quality labels to learn discriminative features once again.

4.4 Parameter analysis

We investigate six hyperparameters and analyze their effect on Market1501, i.e., the weight η in Eq. 4, the candidate number n in Eq. 5, the threshold v in Eq. 5, the parameter α, β in Eq. 10, and m.

The weight η. Figure 3a investigates the effect of the weight η in Eq. 4. The parameter η controls the importance

of reliable samples. We verify the parameter value from 2 to 7, and the performance is best when η is 5. When η gradually increases, the accuracy is not significantly changed, which means that appropriate similar feature constraints can promote the deep model to identify different pedestrians.

The candidate number n. We employ parameter n in Eq. 4 to control the candidate number of pedestrian features that are different from feature f_i. The evaluation results are shown in Fig. 3b. We rank the feature similarity and select n similarity scores after one hundred. From the figure, the proposed HFCN achieves the best performance when n is equal to 30. When n is too small, the feature constraint loss is difficult to stimulate the Re-ID model to distinguish different pedestrians, resulting in performance degradation. So, $n = 30$ is utilized in our experiments.

The threshold v. We evaluate the threshold v in Eq. 5, and the results are listed in Fig. 3c. The parameter is utilized to select reliable pedestrian features. A higher threshold selects fewer and more similar features, and the lower thresh-

Table 3 Performance comparison with the state-of-the-art methods on MSMT17. The column "Source" denotes external data utilized, where "None" stands for fully unsupervised methods

Methods	Source	MSMT17			
		Rank-1	Rank-5	Rank-10	mAP
MMCT [14]	None	35.4	44.8	49.8	11.2
JVTC+ [51]	None	43.1	53.8	59.4	17.3
STAR-DAC [53]	None	41.5	–	64.9	17.7
DSCE [57]	None	35.2	48.3	–	15.5
MLC [58]	None	37.1	47.5	52.4	12.0
KNNL [55]	None	37.5	48.5	52.0	13.4
MPRD [56]	None	37.7	51.3	57.1	146
SCL [54]	None	41.4	53.6	58.7	13.3
HFCN (Ours)	**None**	**52.3**	**63.0**	**68.0**	**21.9**
D-MMD [59]	Market	29.1	46.3	54.1	13.5
SILC [67]	Market	27.8	38.1	45.8	10.9
TAL-MIRN [64]	Market	30.9	43.5	–	11.2
PREST [65]	Market	37.8	51.8	57.8	15.9
MLC [58]	Market	43.9	55.3	60.4	16.5
MLOL [68]	Market	46.9	59.4	64.7	21.7
D-MMD [59]	Duke	34.4	51.1	58.5	15.3
SILC [67]	Duke	33.1	45.2	48.0	12.6
TAL-MIRN [64]	Duke	39.0	51.5	–	14.2
PREST [65]	Duke	43.8	57.5	63.6	18.5
MLC [58]	Duke	46.4	57.9	62.7	18.0
MLOL [68]	Duke	48.3	60.7	66.1	22.4

Bold values indicate the best results

Table 4 Evaluation of parameter α in Eq. 10

α	Market1501			
	Rank-1	Rank-5	Rank-10	mAP
0.8	87.6	94.0	96.1	68.3
1.0	88.4	94.2	96.1	68.5
1.2	**88.7**	**94.7**	**96.4**	**69.3**
1.4	88.6	94.5	96.2	68.8
1.6	84.5	93.8	95.4	61.9

Bold values indicate the best results

Table 5 Evaluation of parameter β in Eq. 10

β	Market1501			
	Rank-1	Rank-5	Rank-10	mAP
0.8	87.8	94.5	96.3	67.4
1.0	88.0	94.5	96.2	68.1
1.2	88.2	94.6	96.3	68.9
1.4	**88.7**	**94.7**	**96.4**	**69.3**
1.6	88.5	94.5	96.2	69.2

Bold values indicate the best results

Table 6 Evaluation of parameter m

m	Market1501			
	Rank-1	Rank-5	Rank-10	mAP
1	74.6	87.7	91.3	47.9
2	87.0	94.4	96.1	67.4
3	87.9	94.2	96.0	68.0
4	87.6	94.2	96.3	67.9
5	**88.7**	**94.7**	**96.4**	**69.3**
6	88.3	94.5	96.1	68.1
7	88.7	94.6	96.3	68.6
8	88.2	94.6	96.4	68.5

Bold values indicate the best results

old selects more features for optimization. The results show that the threshold of 0.8 drives this model to achieve the best accuracy.

The parameter α. Table 4 shows the effect of parameter α in Eq. 10. The performance achieves the best result when $\alpha = 1.2$. This proves the generated label information is important. Since the accuracy is relatively stable, the parameter is not sensitive to the optimization process. So, we set α to be 1.2 in our experiments.

The parameter β. The parameter β denotes the importance of the hard triplet loss, and the assessment is shown in

Table 5. The proposed HFCN obtains the highest accuracy when $\beta = 1.4$. This illustrates that the hard triplet loss is vital for identifying hard samples and further stimulates the deep model to distinguish different pedestrians.
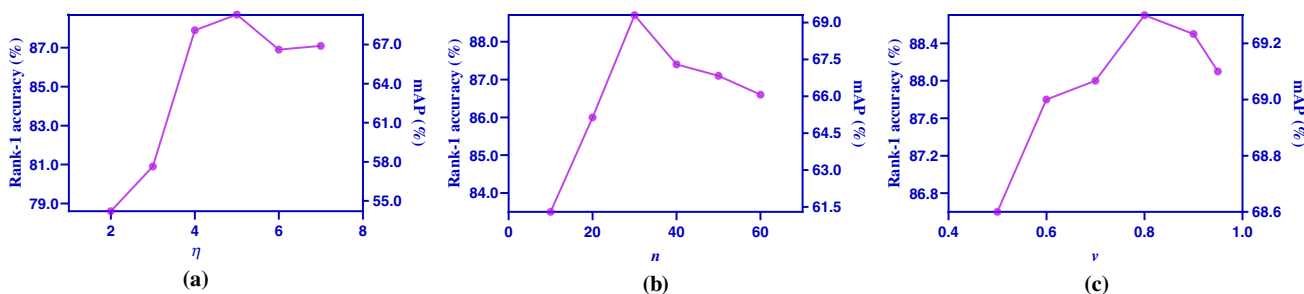

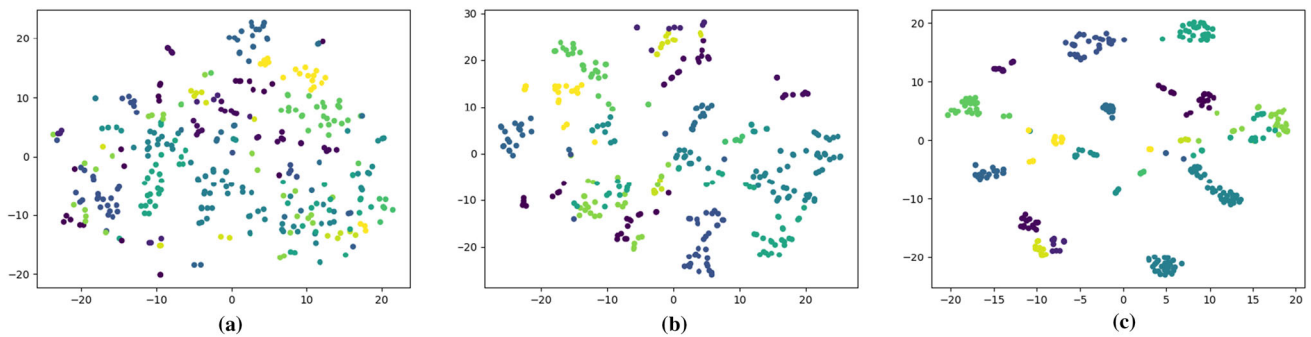
Fig. 3 Evaluation of three parameters

Fig. 4 Visualization of twenty randomly selected identities from the test set on Market1501. **a** Denotes the initial distribution, **b** represents the optimized result by performing the multi-task operation, and **c** indicates the final feature distribution by HFCN

The parameter m. Table 6 reveals the effect of the parameter m. The parameter expresses we start to search for reliable features after m epochs. The results illustrate that we obtain the best performance when m is equal to 5. This reason may be pedestrian features have certain discrimination after proper constraints, which is more conducive to the clustering algorithm to produce effective labels.

4.5 Visualization analysis

We visualize the feature distribution as shown in Fig. 4, where the same color indicates the same identity. Figure 4a is the initialized feature extracted by a pre-trained model on ImageNet. It hardly distinguishes any pedestrians, reflecting the poor performance. By performing the clustering algorithm and implementing the multi-task operation, Fig. 4b reveals certain performance with its ability to distinguish different pedestrians. The proposed HFCN achieves superior performance as shown in Fig. 4c, which significantly identifies pedestrian images. This further proves that our feature constraint loss effectively encourages the same pedestrian to have the similar distribution and different pedestrians to generate the large discrepancy. More importantly, we successfully integrate the feature constraint loss and the multi-task operation, which stimulates the clustering algorithm to generate high-quality labels for learning discriminative features.

5 Conclusion

In the paper, we propose HFCN to restrict potential pedestrian features and mine effective information with no labeled data for fully unsupervised person Re-ID. Specifically, we define the feature constraint loss to restrict the feature distribution. As a result, pedestrian features with the same identity are facilitated to have the similar distribution and different pedestrians possess large distribution gaps. Hence, the operation could promote the deep model to distinguish different pedestrians, which further stimulates the clustering algorithm

to generate high-quality labels. Afterward, we design the multi-task operation to sufficiently utilize predicted labels and identify hard samples. Finally, the feature constraint loss and multi-task operation are integrated to jointly optimize the Re-ID model, improving the Re-ID model's generalization ability. Numerous experiments demonstrate the effectiveness of the proposed HFCN that surpasses the state-of-the-art for fully unsupervised Re-ID.

The proposed HFCN possesses six hyperparameters that have been analyzed in Sect. 4.4. Although the six parameters are easy to determine, their choice still requires some experience when expanding to other fields. In the future, we will extend the idea to unsupervised cross-modality person ReID to solve the cross-modality problem and further verify its effectiveness and stability [69].

Data availability The database used is: Market1501, DukeMTMC-reID and MSMT17. They are already public.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wang, X.: Intelligent multi-camera video surveillance: a review. Pattern Recognit. Lett. **34**(1), 3–19 (2013)
2. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: European Conference on Computer Vision, pp. 17–35 (2016)
3. Chen, Z., Lv, X., Sun, T., Zhao, C., Chen, W.: Flag: feature learning with additional guidance for person search. Vis. Comput. **37**(4), 685–693 (2021)

4. Si, T., He, F., Zhang, Z., Duan, Y.: Hybrid contrastive learning for unsupervised person re-identification. IEEE Trans. Multimed. (2022). https://doi.org/10.1109/TMM.2022.3174414

5. Fan, X., Jiang, W., Luo, H., Mao, W.: Modality-transfer generative adversarial network and dual-level unified latent representation for visible thermal person re-identification. Vis. Comput. 38, 279–294 (2022)

6. Wei, D., Wang, Z., Luo, Y.: Video person re-identification based on RGB triple pyramid model. Vis. Comput. (2022). https://doi.org/10.1007/s00371-021-02344-7

7. Pervaiz, N., Fraz, M., Shahzad, M.: Per-former: rethinking person re-identification using transformer augmented with self-attention and contextual mapping. Vis. Comput. (2022). https://doi.org/10.1007/s00371-022-02577-0

8. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 79–88 (2018)

9. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camstyle: a novel data augmentation method for person re-identification. IEEE Trans. Image Process. 28(3), 1176–1190 (2018)

10. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Invariance matters: exemplar memory for domain adaptive person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 598–607 (2019)

11. Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y.: Learning to adapt invariance in memory for person re-identification. IEEE Trans. Pattern Anal. Mach. Intell. 43(8), 2723–2738 (2021)

12. Zhai, Y., Ye, Q., Lu, S., Jia, M., Ji, R., Tian, Y.: Multiple expert brainstorming for domain adaptive person re-identification. In: European Conference on Computer Vision, pp. 594–611 (2020)

13. Chen, H., Lagadec, B., Bremond, F.: Enhancing diversity in teacher–student networks via asymmetric branches for unsupervised person re-identification. In: IEEE Winter Conference on Applications of Computer Vision, pp. 1–10 (2021)

14. Wang, D., Zhang, S.: Unsupervised person re-identification via multi-label classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 10981–10990 (2020)

15. Lin, Y., Xie, L., Wu, Y., Yan, C., Tian, Q.: Unsupervised person re-identification via softened similarity learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3390–3399 (2020)

16. Zhang, S., He, F.: DRCDN: learning deep residual convolutional dehazing networks. Vis. Comput. 36(9), 1797–1808 (2020)

17. Pan, Y., He, F., Yu, H.: Learning social representations with deep autoencoder for recommender system. World Wide Web 23(4), 2259–2279 (2020)

18. Liu, T., Cai, Y., Zheng, J., Thalmann, N.M.: Beacon: a boundary embedded attentional convolution network for point cloud instance segmentation. Vis. Comput. (2021). https://doi.org/10.1007/s00371-021-02112-7

19. Tulsulkar, G., Mishra, N., Thalmann, N.M., Lim, H.E., Lee, M.P., Cheng, S.K.: Can a humanoid social robot stimulate the interactivity of cognitively impaired elderly? a thorough study based on computer vision methods. Vis. Comput. 37(12), 3019–3038 (2021)

20. Arora, S., Bhatia, M., Mittal, V.: A robust framework for spoofing detection in faces using deep learning. Vis. Comput. 38(7), 2461–2472 (2022)

21. Wei, T., He, F., Liu, Y.: YDTR: infrared and visible image fusion via y-shape dynamic transformer. IEEE Trans. Multimed. (2022). https://doi.org/10.1109/TMM.2022.3192661

22. Liang, Y., He, F., Zeng, X., Luo, J.: An improved loop subdivision to coordinate the smoothness and the number of faces via multi-objective optimization. Integr. Comput. Aided Eng. 29(1), 23–41 (2021)

23. Li, H., He, F., Chen, Y., Pan, Y.: MLFS-CCDE: multi-objective large-scale feature selection by cooperative coevolutionary differential evolution. Memet. Comput. 13(1), 1–18 (2021)

24. Zhang, Z., Si, T., Liu, S.: Integration convolutional neural network for person re-identification in camera networks. IEEE Access 6, 36887–36896 (2018)

25. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 4321–4329 (2019)

26. Liu, S., Huang, W., Zhang, Z.: Learning hybrid relationships for person re-identification. In: Association for the Advance of Artificial Intelligence, pp. 2172–2179 (2021)

27. Xie, J., Ge, Y., Zhang, J., Huang, S., Chen, F., Wang, H.: Low-resolution assisted three-stream network for person re-identification. Vis. Comput. (2021). https://doi.org/10.1007/s00371-021-02127-0

28. Ding, Y., Duan, Z., Li, S.: Source-free unsupervised multi-source domain adaptation via proxy task for person re-identification. Vis. Comput. 38(6), 1871–1882 (2022)

29. Si, T., He, F., Wu, H., Duan, Y.: Spatial-driven features based on image dependencies for person re-identification. Pattern Recognit. 124, 108462 (2022)

30. Chen, D., Xu, D., Li, H., Sebe, N., Wang, X.: Group consistent similarity learning via deep CRF for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 8649–8658 (2018)

31. Liu, J., Zha, Z., Chen, D., Hong, R., Wang, M.: Adaptive transfer network for cross-domain person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 7202–7211 (2019)

32. Zhang, Z., Wang, Y., Liu, S., Xiao, B., Durrani, T.: Cross-domain person re-identification using heterogeneous convolutional network. IEEE Trans. Circuits Syst. Video Technol. 32(3), 1160–1171 (2022)

33. Jia, Z., Li, Y., Tan, Z., Wang, W., Wang, Z., Yin, G.: Domain-invariant feature extraction and fusion for cross-domain person re-identification. Vis. Comput. (2022). https://doi.org/10.1007/s00371-022-02398-1

34. Zhong, Z., Zheng, L., Li, S., Yang, Y.: Generalizing a person retrieval model hetero-and homogeneously. In: European Conference on Computer Vision, pp. 172–188 (2018)

35. 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)

36. 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)

37. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European Conference on Computer Vision, pp. 262–275 (2008)

38. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)

39. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)

40. Lin, Y., Dong, X., Zheng, L., Yan, Y., Yang, Y.: A bottom-up clustering approach to unsupervised person re-identification. In: Association for the Advance of Artificial Intelligence, pp. 8738–8745 (2019)

41. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

42. Wang, W., Wu, Y., Tang, C., Hor, M.: Adaptive density-based spatial clustering of applications with noise (DBSCAN) according to data. In: International Conference on Machine Learning and Cybernetics, pp. 445–451 (2015)

43. Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., Wang, X.: Unsupervised domain adaptive re-identification: theory and practice. Pattern Recognit. **102**, 107173 (2020)

44. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Association for the Advancement of Artificial Intelligence, pp. 13001–13008 (2020)

45. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5157–5166 (2018)

46. Zhai, Y., Lu, S., Ye, Q., Shan, X., Chen, J., Ji, R., Tian, Y.: Adcluster: augmented discriminative clustering for domain adaptive person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 9021–9030 (2020)

47. Dai, Z., Chen, M., Gu, X., Zhu, S., Tan, P.: Batch dropblock network for person re-identification and beyond. In: IEEE International Conference on Computer Vision, pp. 3691–3701 (2019)

48. Si, T., Zhang, Z., Liu, S.: Compact triplet loss for person re-identification in camera sensor networks. Ad Hoc Netw. **95**, 101984 (2019)

49. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: a survey and outlook. IEEE Trans. Pattern Anal. Mach. Intell. **44**(6), 2872–2893 (2021)

50. Zhuang, W., Wen, Y., Zhang, S.: Joint optimization in edge-cloud continuum for federated unsupervised person re-identification. In: ACM International Conference on Multimedia, pp. 433–441 (2021)

51. Li, J., Zhang, S.: Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In: European Conference on Computer Vision, pp. 483–499 (2020)

52. Zeng, K., Ning, M., Wang, Y., Guo, Y.: Hierarchical clustering with hard-batch triplet loss for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 13657–13665 (2020)

53. Prasad, M.V., Balakrishnan, R., Ramadoss, B.: Spatio-temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification. Pattern Recognit. **122**, 108287 (2022)

54. Pang, B., Zhai, D., Jiang, J., Liu, X.: Fully unsupervised person re-identification via selective contrastive learning. ACM Trans. Multimed. Comput. Commun. Appl. **18**(2), 1–15 (2022)

55. Xie, K., Wu, Y., Xiao, J., Li, J., Xiao, G., Cao, Y.: Unsupervised person re-identification via k-reciprocal encoding and style transfer. Int. J. Mach. Learn. Cybern. **12**, 1–18 (2021)

56. Ji, H., Wang, L., Zhou, S., Tang, W., Zheng, N., Hua, G.: Meta pairwise relationship distillation for unsupervised person re-identification. In: IEEE International Conference on Computer Vision, pp. 3661–3670 (2021)

57. Yang, F., Zhong, Z., Luo, Z., Cai, Y., Lin, Y., Li, S., Sebe, N.: Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4855–4864 (2021)

58. Li, Q., Peng, X., Qiao, Y., Hao, Q.: Unsupervised person re-identification with multi-label learning guided self-paced clustering. Pattern Recognit. **125**, 108521 (2022)

59. Djebril, M., Amran, B., George, E., Eric, G.: Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: European Conference on Computer Vision, pp. 159–174 (2020)

60. Ji, Z., Zou, X., Lin, X., Liu, X., Huang, T., Wu, S.: An attention-driven two-stage clustering method for unsupervised person re-identification. In: European Conference on Computer Vision, pp. 20–36 (2020)

61. Yang, F., Li, K., Zhong, Z., Luo, Z., Sun, X., Cheng, H., Guo, X., Huang, F., Ji, R., Li, S.: Asymmetric co-teaching for unsupervised cross-domain person re-identification. In: Association for the Advance of Artificial Intelligence, pp. 12597–12604 (2020)

62. Jin, X., Lan, C., Zeng, W., Chen, Z.: Global distance-distributions separation for unsupervised person re-identification. In: European Conference on Computer Vision, pp. 735–751 (2020)

63. Wang, G., Lai, J., Liang, W., Wang, G.: Smoothing adversarial domain attack and p-memory reconsolidation for cross-domain person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 10568–10577 (2020)

64. Li, H., Dong, N., Yu, Z., Tao, D., Qi, G.: Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification. IEEE Trans. Circuits Syst. Video Technol. **32**(5), 2814–2830 (2021)

65. Zhang, H., Cao, H., Yang, X., Deng, C., Tao, D.: Self-training with progressive representation enhancement for unsupervised cross-domain person re-identification. IEEE Trans. Image Process. **30**, 5287–5298 (2021)

66. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Li, F.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)

67. Ainam, J.P., Qin, K., Owusu, J.W., Lu, G.: Unsupervised domain adaptation for person re-identification with iterative soft clustering. Knowl. Based Syst. **212**, 106644 (2021)

68. Sun, J., Li, Y., Chen, H., Peng, Y., Zhu, J.: Unsupervised cross domain person re-identification by multi-loss optimization learning. IEEE Trans. Image Process. **30**, 2935–2946 (2021)

69. Liang, W., Wang, G., Lai, J., Xie, X.: Homogeneous-to-heterogeneous: unsupervised learning for RGB-infrared person re-identification. IEEE Trans. Image Process. **30**, 6392–6407 (2021)

Tongzhen Si is currently pursuing Ph.D. degree with the School of Computer Science, Wuhan University, China. His research interests include computer vision, deep learning, and person re-identification.

Fazhi He received the bachelor's, master's, and Ph.D. degrees from the Wuhan University of Technology. He was a Postdoctoral Researcher at The State Key Laboratory of CAD&CG, Zhejiang University, a Visiting Researcher with the Korea Advanced Institute of Science and Technology, and a Visiting Faculty Member at the University of North Carolina at Chapel Hill. He is currently a Professor with the School of Computer Science, Wuhan University. His research interests are artificial intelligence, intelligent computing, computer graphics, image processing, computer-aided design, computer-supported cooperative work, and co-design of software/hardware.

Penglei Li is currently pursuing Ph.D. degree with the Institute of Geophysics and Geomatics, China University of Geosciences, China. Her research interests include image processing, deep learning, and computer vision.