Applied Intelligence (2023) 53:1605–1618
https://doi.org/10.1007/s10489-022-03218-8

Semi-supervised pedestrian re-identification via a teacher–student model with similarity-preserving generative adversarial networks

Botong Zhao[1,2] · **Yanjie Wang**[1,2] · **Keke Su**[1,2] · **Hong Ren**[1] · **Xiyu Han**[1]

Accepted: 6 January 2022 / Published online: 30 April 2022

Abstract

This paper describes a pedestrian re-identification algorithm, which was developed by integrating semi-supervised learning and similarity-preserving generative adversarial networks (SPGAN). The pedestrian re-identification task aimed to rapidly capture the same target using different cameras. Importantly, this process can be applied in the field of security. Because real-life environments are complex, the number of detected identities is uncertain, and the cost of manual labeling is high; therefore, it is difficult to apply the re-identification model based on supervised learning in real-life scenarios. To use the existing labeled dataset and a large amount of unlabeled data in the application environment, this report proposes a semi-supervised pedestrian re-identification model, which combines a teacher–student model with SPGAN. SPGAN was used to reduce the difference between the target domain and the source domain by transferring the style of the labeled dataset from the source domain. Additionally, the dataset from the source domain was used after the style transfer to pre-train the model; this enabled the model to adapt more rapidly to the target domain. The teacher–student model and the transformer model were then employed to generate soft pseudo-labels and hard pseudo-labels (via iterative training) and to update the parameters through distillation learning. Thus, it retained the learned features while adapting to the target domain. Experimental results indicated that the maps of the applied method on the Market-to-Duke, Duke-to-Market, Market-to-MSMT, and Duke-to-MSMT domains were 70.2, 79.3, 30.2, and 33.4, respectively.

Keywords Deep learning · Pedestrian re-identification · Semi-supervised learning

1 Introduction

The task of pedestrian re-recognition involves evaluating a pedestrian image taken by a camera, and then re-identifying that pedestrian from a large number of images captured by different cameras. This process can be widely applied in the field of security and has recently emerged as a research hotspot in the field of computer vision. Pedestrian re-recognition tasks can be deconvoluted into two processes: feature extraction and feature matching. Because the images captured by different cameras have large differences in terms of the background, brightness, camera resolution, and other parameters, feature extraction and feature matching processes face significant challenges. The key to attaining pedestrian re-identification (re-ID) lies in extracting robust feature representation.

Conventional pedestrian re-recognition based on supervised models can achieve suitable performance in each dataset; however, this approach is not robust, and it has difficulty adapting to the application environment after training. In general, models based on supervised learning are difficult to apply in real environments because of the uncertain number of identities during real-life applications and the high cost of manual labeling. Moreover, in the field of pedestrian re-identification, a large amount of unmarked data can be obtained; therefore,

✉ Yanjie Wang
wangyj@ciomp.ac.cn

Botong Zhao
zhaobotong19@mails.ucas.ac.cn

Keke Su
sukeke@mails.ucas.ac.cn

Hong Ren
renv587@126.com

Xiyu Han
hanxiyusdu@163.com

[1] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

semi-supervised pedestrian re-identification technology is required.

In recent years, four main strategies have been employed for target re-identification based on unsupervised learning: (1) method based on pseudo-labeling; (2) method based on image generation; (3) method based on instance classification; (4) method based on domain adaptation.

Considering that there are already many available labeled datasets, and a significant amount of unlabeled data can be obtained during pedestrian re-identification tasks, this study used unsupervised domain-adaptive methods for modeling. First, similarity-preserving generative adversarial networking (SPGAN) was used to adapt the style of the source domain image to make it closer to the target domain style. Then, ResNet-50 was used to extract the discriminative features shared by the target domain and the source domain.

Next, a clustering algorithm was used to generate pseudo-labels for the unlabeled target domain images. Considering that the number of identities is uncertain in real application scenarios, we used density-based spatial clustering of applications with noise (DBSCAN) to generate pseudo-labels. To minimize the influence of the noise contained in each pseudo-label and to reduce the impact of the hard pseudo-label, this study applied the predicted value of the network as a soft pseudo-label instead of the output in a one-hot format. Meanwhile, to avoid the network's predicted value being used directly under the network's own supervision, this study implemented a teacher–student model, which involved constructing two networks for collaborative training and ensuring the relative independence of the two networks. This principle is illustrated in Fig. 1.

The ENC [1] described three characteristics of pedestrian re-identification tasks, i.e., exemplar invariance, camera invariance, and neighborhood invariance, which are presented in Fig. 2.

The task of this article is to build a semi-supervised pedestrian re-recognition system based on the teacher-student model and SPGAN. The main challenges we face are:

1. The noise of the pseudo-label will interfere with the training of the neural network.

2. The loss function needs to consider instance invariance, camera invariance and neighborhood invariance.

3. How to improve the decoupling ability of the network framework and the robustness in application scenarios.

4. How to provide reliable pre-training network for teacher-student model.

Based on the above issues, the main contributions of this article are divided into three aspects:

1. In response to problems 1 and 2, we propose a new compound loss function, which makes the teacher-student model pay attention to the three characteristics of the pedestrian re-recognition task during the training process, and reduces the noise of pseudo-label.

2. In order to solve problem 3, we introduced Transformer to adjust the structure of ResNet to improve the decoupling ability of ResNet in the task of pedestrian re-recognition.

3. By introducing SPGAN into the teacher-student model, it provides the pre-training network with a labeled

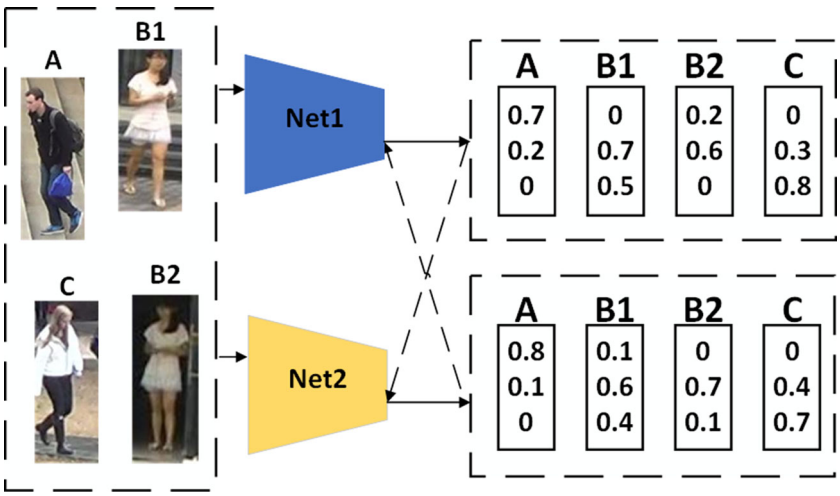

Fig. 1 The principle of network collaborative training via the teacher–student model. In the training process, the probability of each identity obtained after classification is used as a soft pseudo- label. Compared with the labels generated directly by the clustering algorithm, the use of soft pseudo-labels can find elucidate the relative relationships between different identities and reduce the noise caused by the defects of the clustering algorithm itself. By using two networks for collaborative training, it is possible to prevent the label generated by the network from directly supervising itself, and to ensure the independence of the two networks

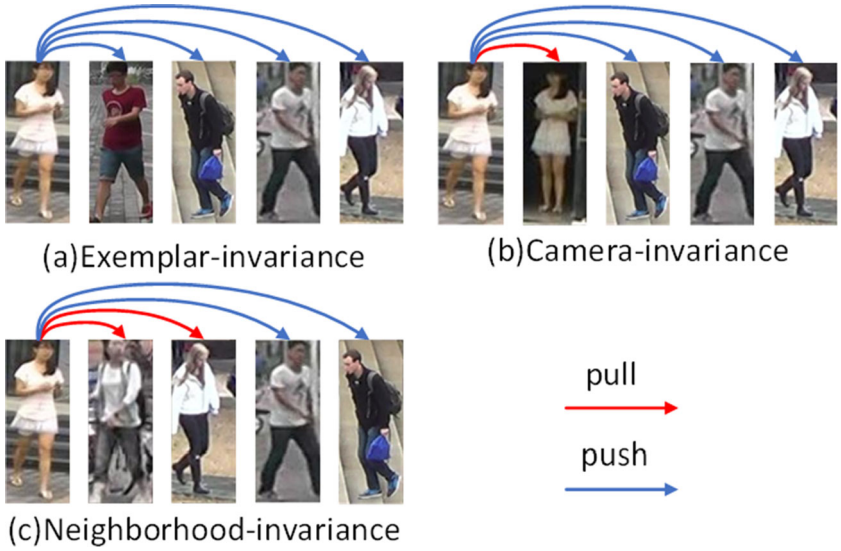

Fig. 2 Three characteristics of pedestrian re-identification tasks: (a) the distance between individuals should be increased; (b) the distance between different cameras capturing the same individual should be shortened; (c) the distance between similar individuals should be shortened

data set similar to the image style of the target domain, and then provides a better pre-training model for the teacher-student model.

2 Related work

2.1 Re-ID via generating pseudo-labels

The re-ID method based on generating pseudo-labels involved generating high-quality pseudo-labels for unlabeled data to train and update the network. Yu et al. [2] proposed a soft label-based learning method to overcome the challenge of unsupervised pedestrian re-identification. This method generated pseudo-labels by supplementing the dataset with labeled data. Specifically, a cluster center was generated for each class in the target domain, and then, a vector was generated according to the similarity between each unlabeled sample and each class. Then, the similarity between these datasets was calculated. Yang et al. [3] proposed a discriminative feature learning method based on blocks. This method first brought similar images closer together, and pushed dissimilar images away. Then, the original image after style transfer was considered as a positive sample, and the most difficult negative samples were identified. Finally, the system was optimized based on the triplet loss. Fu et al. [4] used a DBSCAN clustering algorithm to cluster the unlabeled data based on the features extracted from the source domain, and then applied the triple loss technique for training. Ding et al. [5] proposed a clustering method based on the degree of dispersion to cluster the target domain samples. This clustering method not only considered the differences between individuals, but also the compactness of similar individuals. Compared with alternative clustering methods, this approach more widely captured the relationship between multiple samples and effectively dealt with the problems caused by unbalanced data distribution.

Currently, the generation of pseudo-labels has become a mainstream technical route. This method involves clear steps and achieves good performance (similar to that of supervised learning methods). However, as their name suggests, pseudo-labels are not real, and they contain noise. Therefore, improving the quality of pseudo-labels and the effective use of tags is required for such methods, e.g., by improving the extraction and analysis of features so that the clustering algorithm can generate more accurate labels, or using the extracted features as soft labels to reduce the influence of pseudo-label noise.

2.2 Re-ID via generating images

Recently, with the rapid development of generative adversarial networks (GAN), researchers have tried to solve the problem of pedestrian re-identification from the perspective of style transfer. Huang et al. proposed SBSGAN [6], which removes the background area of an image by generating a soft mask. This method can effectively suppress the errors of the image segmentation method. Zhong et al. [7] developed StarGAN [8] to transform images captured with different camera styles in the target domain. The positive samples obtained in the training process adopted the style of the same camera, which combined with the original target domain image, the source domain image, and the transformed image to generate a triplet to train the neural network. Wei et al. proposed PTGAN [9] to transfer images from the source domain to the target domain. This method introduced the pedestrian background segmentation image on the basis of CycleGAN [10] to verify the consistency of the pedestrian area before and after the style transfer. The

SPGAN [11] proposed by Deng et al. (based on CycleGAN) increased similarity preservation for the characteristics of an unchanged identity before and after conversion; this approach made the generated image more reasonable.

This type of approach relies on the quality of the images generated by the GAN, but images from surveillance videos generally exhibit poor quality and noise, which causes the quality of the image after the style transfer to be unstable. However, this method makes full use of the image in the source domain, so it is essentially complementary to the method based on pseudo-labels. Therefore, this type of method must use images in the application scene to further improve the model after the style transfer. This enables the model to be transferred to the application environment, and the robustness of the model in the application scene is improved.

2.3 Re-ID via exemplar classification

This type of re-ID method focuses obtaining and utilizing better relationships between samples. Zhong et al. [12] proposed a prediction method based on graph neural networks to determine whether two samples were real neighboring samples. Ding et al. [13] selected nearby samples by setting a distance threshold. Considering that the imbalance of adjacent samples for each instance can lead to bias in learning, they suppressed this phenomenon by applying a loss function.

Although this method demonstrates superior performance, the relationship between samples requires further research. For example, it is necessary to design an effective loss function so the model can learn more nuanced feature relationships among samples in order to not be limited to whether or not they are the same sample.

2.4 Re-ID via unsupervised domain adaptation

This method based on unsupervised domain adaptation (UDA) follows the traditional domain-adaptive framework, aims to eliminate or reduce the differences between domains, and transfers discriminative information in the source domain to the target domain.

Both Delorme et al. [14] and Qi et al. [15] proposed camera-based GAN methods to solve the problem of data distribution differences in cross-domain pedestrian re-ID tasks. Ge et al. [16] used network joint training to ensure independence between the two networks and employed soft labels to alleviate the noise problem of the clustering algorithm. However, the integrated loss function considered only the differences between different samples but not the neighbor invariance.

Compared with methods based on pseudo-labels or instance-based classification, this approach achieves lower performance. However, in pedestrian re-recognition tasks, it shows that the effect of transferred learning is better at the feature level than at the image level.

According to above works, the main challenge for pedestrian re-identification is how to effectively use labeled source domain data sets and a large number of unlabeled images in application scenarios. In the existing methods, Re-ID via generating pseudo-labels directly ignores the noise of pseudo-label generated by the clustering algorithm. And Re-ID via generating images can only generate images similar in style to the target domain image. It is not the image in the application scenario, so it also contains noise. This also leads to the performance of the network after training cannot meet the requirements of the application. Although the method based on domain adaptation focuses on distinguishing different samples, we need to make the model notice that there are some similar discriminative features between similar samples, which is the neighborhood invariance. Based on the above challenges, we propose our model to alleviate these problems, which will be introduced in Section 3.

3 Proposed method

The method proposed herein has three stages. First, SPGAN is used to transfer the style of the source domain dataset, which makes the sample style of the source domain dataset similar to that of the target domain sample. Additionally, the samples from the source domain dataset after the style transfer are independent of any samples in the target domain dataset. Then, supervised training is conducted on the source domain data to search for discriminative features shared between the target domain and source domain datasets. Finally, a clustering algorithm is used to generate labels for the unlabeled data, and the teacher–student model is applied to update the parameters of the network. In this way, the model gradually adapts to the sample style of the target domain dataset.

3.1 SPGAN

SPGAN is a learning framework based on style transfer and composed of a Siamese network (SiaNET) and a CycleGAN. Through the coordination between these two networks, SPGAN can generate samples that adopt the style of the target domain while maintaining their identity information, so that the network can converge faster in the teacher–student model.

3.1.1 CycleGAN

The main principle of CycleGAN is that when an image is transferred from the style of dataset A to the style of dataset B, the new image should form an image similar to the original via the second style transfer, and the main information is retained. This concept is shown schematically in Fig. 3.

We proposed that the discriminator D_Y could effectively determine whether the object is in the style of the target domain Y. At the same time, the generator G(X → Y) was used to effectively generate images in the style of the target domain Y. The loss function involving the generator G(X → Y) and the discriminator D_Y is shown in (1),

$$L_{YGAN}(G, D_Y, X, Y) = E_{y\sim py}[(D_Y(Y) - 1)^2] + E_{x\sim px}[D_Y(G(X))^2] \quad (1)$$

where *px* and *py* represent the sample distribution of the X and Y datasets, respectively.

Similarly, the loss function involving the generator F and the discriminator D_X is shown in (2):

$$L_{XGAN}(G, D_X, Y, X) = E_{x\sim px}[(D_X(X) - 1)^2] + E_{y\sim py}[D_X(F(Y))^2] \quad (2)$$

In particular, the discriminator *D*, is used to determine whether the style of the image after the style transfer is similar to the style of the image in the target domain.

We also proposed that an image could be restored to one that was similar to the original image by using another generator after the style transfer. The loss function for this case is shown in (3):

$$L_{cyc}(G, F) = E_{x\sim px}[\|F(G(X)) - X\|_1] + E_{y\sim py}[\|G(F(Y)) - Y\|_1] \quad (3)$$

According to the ablation experiments (Section 4.3, vide infra), the source domain and the target domain share common discriminative features, which supports the effectiveness of the UDA method. Therefore, while satisfying the above conditions, the original image should retain the information of the original image as much as possible after the style transfer, giving the loss function shown in (4):

$$L_{id}(G, F, X, Y) = E_{x\sim px}[\|F(X) - X\|_1] + E_{y\sim py}[\|G(Y) - Y\|_1] \quad (4)$$

3.1.2 SiaNet and the Loss Function of SPGAN

For the SPGAN, the sample should retain the information contained in the original sample after the style transfer is complete, and it should be independent of any sample in the target domain (in terms of the exemplar invariance and camera invariance in the re-ID task). Therefore, adding SiaNet on the basis of CycleGAN can constrain the learning process of the mapping function (Fig. 4).

The similarity preservation loss function employed to train SiaNet is shown in (5),

$$L_{con}(i, x_1, x_2) = (1 - i)\{max(0, m - d)\}^2 + i * d^2 \quad m \in [0, 2] \quad (5)$$

where *x1* and *x2* are a pair of input vectors, d is the Euclidean distance between those two vectors, *i* indicates

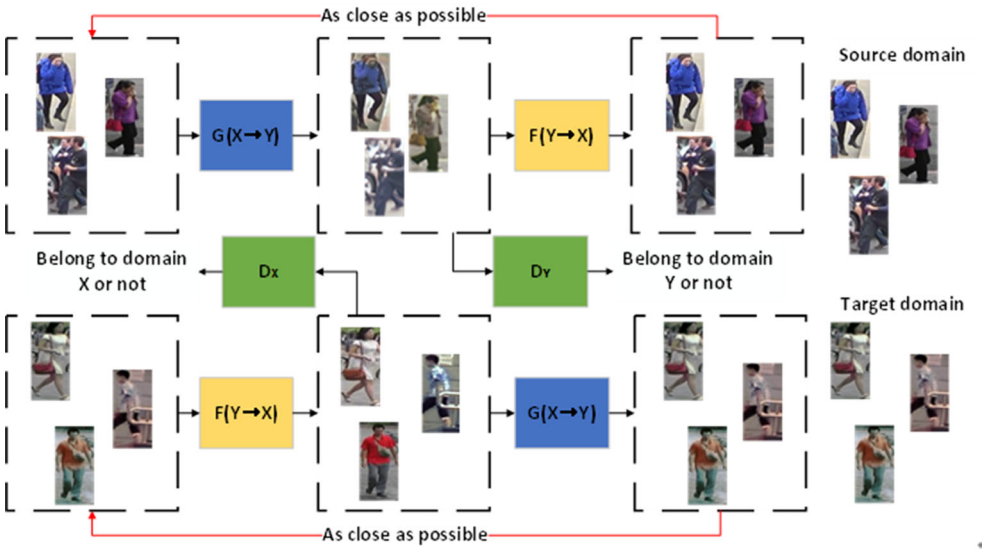

Fig. 3 The defining principles of CycleGAN.When the style transfer is performed on the source domain dataset, the style of the resulting image should be the same as the style of the target domain; after the second style transfer, the original image should still retain the information of the source domain. The same should be true for the style transfer from the target domain to the source domain

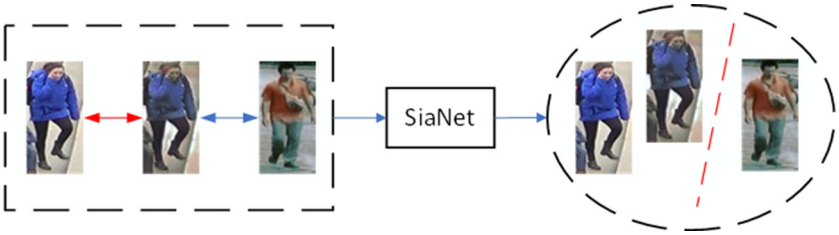

Fig. 4 The principle of SPGAN: First, the style of the source domain image is transferred based on the image style of the target domain. Note that the image after the style transfer does not belong to any category in the target domain. Then, SiaNet is used to constrain the learning process of the mapping function

whether *x1* and *x2* are a pair of positive samples (i = 1 indicates positive samples; $i = 0$ indicates negative samples), m represents the discrimination boundary between positive and negative samples in the feature space (when m = 0, negative samples are ignored by the loss function and cannot be introduced into back propagation; when m ¿ 0, both positive and negative samples will be introduced by the loss function) and determines the proportion of positive and negative samples in the loss function.

3.1.3 Loss Function of SPGAN

Through the loss function of SiaNet, it is possible to reduce the distance between positive sample pairs and increase the distance between negative sample pairs. Therefore, the overall loss function of SPGAN is shown in (6),

$$L_{SPGAN} = L_{XGAN} + L_{YGAN} + \lambda_1 L_{cyc} + \lambda_2 L_{id} + \lambda_3 L_{con} \tag{6}$$

where λ_1, λ_2, and λ_3 are the weights that control the relationship between the four loss functions (Fig. 5).

After the introduction of SPGAN to process the source domain data set, the pre-training model can be provided with pictures that are more similar in style to the target domain, thereby improving the reliability of the pre-training model. From Fig. 6 of the 4.3 ablation experiment, it can be found that when the model is pre-trained using the data set generated by SPGAN, although the accuracy of the pre-training model cannot meet the needs of the application, the pre-training model can achieve higher performance. Furthermore, we can conclude that compared to directly using the source domain data set, using the SPGAN-processed data set can provide a small amount of information unique to the target domain for the pre-training model, and improve the reliability of the pre-training model.

Algorithm 1: The style transfer model based on SPGAN.

Input: Unlabeled target domain data set D_t; labeled source domain data set D_s; The weight between the four loss functions, which are λ_1, λ_2 and λ_3.

Output: A labeled source domain data set D_{st} that has the image style of the target domain and retains the main information of the source domain.

```
for n in [i,num-epochs] do
    for each mini-batch B ⊂ D_s, iteration T do
        1.Take CycleGAN and SiaNet as the basic framework to get the source domain image after style transfer;
        2.Perform gradient descent based on equation 6 and update network parameters;
    end
end
```

3.2 Teacher–student model

The teacher–student model comprises two steps. First, a clustering algorithm is used to generate pseudo-labels, and then, those pseudo-labels are used for collaborative training of the network. This process is illustrated in Fig. 5.

3.2.1 Pre-training of teacher–student model

The most recent UDA methods focus on pre-training the source domain dataset to identify common discriminative features and gradually adapting to the environment of the target domain via transfer learning. Although it is difficult for the network to achieve satisfactory performance after pre-training (i.e., because of distinct camera parameters, brightness, environments, and other factors), the ablation experiments described in Section 4.3 (vide infra) indicated that while the network is trained in the source domain, the accuracy of the target domain gradually increases. SPGAN

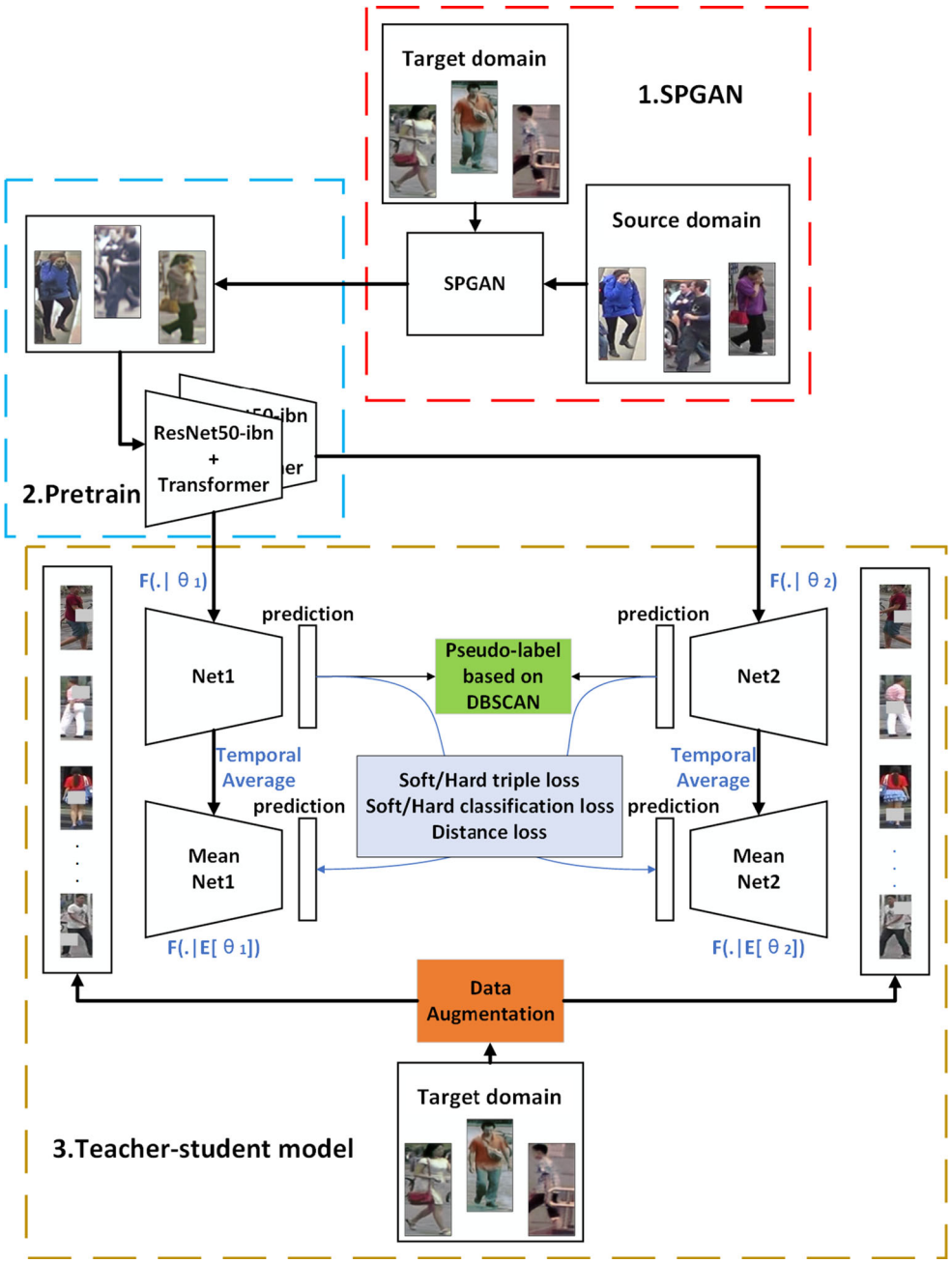

Fig. 5 Principles of the teacher–student model. Stage 1: Use SPGAN to transfer the image style of the source domain dataset to the target domain dataset. Stage 2: Pre-train the model with the source domain data set after the style transfer. Stage 3: Use the teacher–student model to train the network to adapt to the unlabeled target domain dataset

is more suitable for determining the common discriminative features while retaining the source domain information, and this approach can also help the teacher–student model learn faster.

The loss function of the traditional UDA-based re-ID task consists of a classification loss function and a triplet loss function, as shown in (7) and (8), respectively:

$$L_{id}^{s}(\theta) = \frac{1}{N_s}\sum_{i=1}^{N_s} L_{ce}(C^s(F(x_i^s|\theta)), y_i^s) \tag{7}$$

In Eq. 7, $L_{id}^{s}(\theta)$ represents the classification loss function of the source domain; Ns represents the sample size of the source domain; L_{ce} is the cross entropy loss function; $F(x_i^s|\theta)$ is the feature extracted by the pre-training network; $C^s(F(x_i^s|\theta))$ is the source domain classifier, which determines whether the output result of the pre-training model is the label of the corresponding sample; and y_i^s represents the label of the i-th sample of the source domain.

$$\begin{aligned} L_{tri}^{s}(\theta) = \tfrac{1}{N_s}\sum_{i=1}^{N_s} \max(0, ||F(x_i^s|\theta) - F(x_{i,p}^s|\theta)|| \\ +m - ||F(x_i^s|\theta) - F(x_{i,n}^s|\theta)||) \end{aligned} \tag{8}$$

In (8), $L_{tri}^{s}(\theta)$ represents the triplet loss function of the source domain; $x_{i,p}^s$ represents the positive i-th sample;

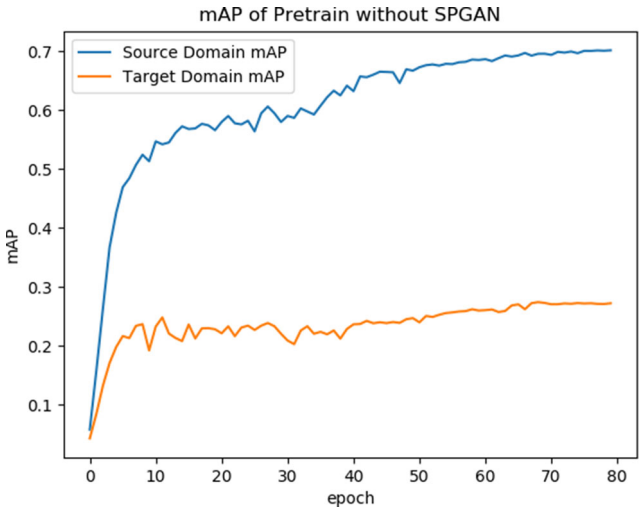

Fig. 6 Comparison of the mAP curves of the source domain dataset and the target domain dataset in the pretraining phase

$x_{i,n}^s$ represents the negative i-th sample; m represents the threshold between the positive and negative samples; and ||.|| designates the normal form distance.

The pre-trained loss function is shown in (9),

$$L_{pre}^s(\theta) = (1-\lambda)L_{tri}^s(\theta) + \lambda L_{id}^s(\theta) \tag{9}$$

where λ represents the weight relationship between the classification function and the triplet loss function.

In the pre-training model, the basic framework comprised a ResNet [17] based on an IBN-Net [18] proposed by Pan et al. for improving the performance of cross-domain transfer learning. Considering that each extracted feature is determined considering global features, a transformer [19] was used to replace the final convolutional layer.

The advantage of the Transformer module is that it can process features in parallel. As in the case of CNN , the same knowledge should be used for all image locations. The transformer also uses the Seq2seq [20] concept to ensure that the new features of each output are obtained after summarizing and analyzing the global features.

The calculation process of Transformer involved (10–14):

$$A = Sigmoid(X + Station) \tag{10}$$

$$Q = W_q A + b_q \tag{11}$$

$$K = W_k A + b_k \tag{12}$$

$$V = W_v A + b_v \tag{13}$$

$$Output = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{14}$$

In (10), A represents the activation of input X after adding the position weight matrix Station. The transformer function maps a query and a set of key-value pairs to an output, wherein the query, keys, values, and outputs are all vectors.

In the transformer, the matrix Q contains a set of queries packed together. Keys and values are also packed together into matrices K and V, respectively, where dk denotes the dimension of the keys.

Algorithm 2: Pre-training model.

Input: The source domain image D_{st}) after the style transfer (following Algorithm 1); The weight λ of the classification loss function in the loss function.

Output: Pre-trained model based on ResNet and Transformer.

```
for n in [i,num-epochs] do
    for each mini-batch B ⊂ D_st, iteration T do
        1, Using ResNet50 as the baseline, the transformer is introduced locally in the last two layers, and the prediction result of the source domain sample X_s is obtained for gradient descent;
        2, Perform gradient descent based on equation 9, and update the parameters θ1, θ2.
    end
end
```

3.2.2 Updating parameters via distillation learning

When constructing the teacher–student model, it is important to pay attention to the following five issues:

1. Because pseudo-labels are not real labels, directly using the labels generated by the clustering algorithm will impact the accuracy of the results. Moreover, the imperfection of the clustering algorithm causes the label itself to be noisy.

2. The model cannot supervise itself using the pseudo-labels that it generates. This does not achieve a learning effect, but will rather diverge the learning results.

3. While updating the parameters, it is important to avoid forgetting the learned knowledge while adapting to the target domain.

4. Re-ID is an open class problem, so the number of identities in the task is unknown.

5. During the training process, the three characteristics of the re-ID task (i.e., exemplar invariance, camera invariance, and neighbor invariance) must be considered.

From the ablation experiments (Section 4.3, vide infra), it was clear that the pseudo-labels generated by K-means and DBSCAN led the model to achieve similar levels of accuracy; however, DBSCAN could cluster dense datasets with any pore size distribution and find abnormal points while clustering, while involving no bias in the clustering results. Therefore, this approach is not affected by the position of the initial cluster center (as with K-means). Therefore, to control the model to pay attention to the open

class characteristics during the learning process, DBSCAN was selected to generate pseudo-labels.

To prevent the model from using clustered pseudo-labels for self-supervision, a collaborative training network was built. The same input from each batch underwent two different data enhancements, and the output results supervised one another. This method guaranteed the independence between the two networks. In particular, after each training, the network will only retain the better performance parameters in the target domain; therefore, in essence, only one network was trained.

To retain the discriminative features learned in the pre-training and subsequent training steps of the learning process, this study applied the idea of distillation learning to update the parameters. The parameter update formulas are presented as (15),

$$\begin{aligned} E^{(T)}[\theta_1] &= \alpha E^{(T-1)}[\theta_1] + (1-\alpha)\theta_1 \\ E^{(T)}[\theta_2] &= \alpha E^{(T-1)}[\theta_2] + (1-\alpha)\theta_2 \end{aligned} \tag{15}$$

where $E^{(T)}[\theta_1]$ and $E^{(T)}[\theta_2]$ represent the parameters of the previous epoch, i.e., $E^{(0)}[\theta_1] = \theta_1$, $E^{(0)}[\theta_2] = \theta_2$; and α, with the range (0,1], represents the proportion of distillation learning required to retain old knowledge each iteration.

To reduce the noise caused by the pseudo-label itself during the learning process, the classification loss function and triple loss function based on the soft pseudo-label were introduced. The soft pseudo-label represents the probability that the output result corresponds to each identity. The classification losses based on soft pseudo-labels were determined using (16),

$$\begin{aligned} L^t_{sid}(\theta_1|\theta_2) &= -\tfrac{1}{Nt}\sum_{i=1}^{Nt}(C^t_2(F(x'^t_i|E^{(T)}[\theta_2])).logC^t_1(F(x^t_i|E^{(T)}[\theta_1]))) \\ L^t_{sid}(\theta_2|\theta_1) &= -\tfrac{1}{Nt}\sum_{i=1}^{Nt}(C^t_1(F(x^t_i|E^{(T)}[\theta_1])).logC^t_2(F(x'^t_i|E^{(T)}[\theta_2]))) \end{aligned} \tag{16}$$

where $C^t_j(F(x^t_i|E[\theta_1]))$ represents the target domain classifier based on the $E^{(T)}[\theta_j]$ parameter in the j-th network; and x'^t_i and x^t_i represent the two data enhancements of the i-th sample of the target domain.

The triplet loss functions based on soft labels are represented by (17):

$$\begin{aligned} L^t_{stri}(\theta_1|\theta_2) &= \tfrac{1}{N_t}\sum_{i=1}^{N_t} L_{bce}(\tau_i(\theta_1), \tau_i(E^{(T)}[\theta_2])) \\ L^t_{stri}(\theta_2|\theta_1) &= \tfrac{1}{N_t}\sum_{i=1}^{N_t} L_{bce}(\tau_i(\theta_2), \tau_i(E^{(T)}[\theta_1])) \end{aligned} \tag{17}$$

$$\tau_i(\theta) = \frac{\exp(||F(x^t_i|\theta) - F(x^t_{i,n}|\theta)||)}{\exp(||F(x^t_i|\theta) - F(x^t_{i,p}|\theta)|| + ||F(x^t_i|\theta) - F(x^t_{i,n}|\theta)||)} \tag{18}$$

The soft triplet loss function can be used to shorten the distance between the sample and the positive sample, and extend the distance between the sample and the negative sample. However, the triplet loss function and the classification loss function do not consider bringing similar samples closer together. That is, only the exemplar invariance and camera invariance are considered, but the neighbor invariance is not considered. Therefore, it is necessary to introduce a loss function that narrows the distance between similar samples, as shown in (19),

$$\begin{aligned} L_{push}(\theta_1|\theta_2) &= \sum_{i=1}^{Nt}(F(x^t_i|E^{(T)}[\theta_1]) - C(y^t_i)) \\ L_{push}(\theta_2|\theta_1) &= \sum_{i=1}^{Nt}(F(x^t_i|E^{(T)}[\theta_2]) - C(y^t_i)) \end{aligned} \tag{19}$$

where $C(y^t_i)$ is the feature vector of the cluster center of a certain class. Considering that this method employs the DBSCAN algorithm, the cluster center is the center of the nearest n points of the same type in the current sample (i.e., a mini-K-means-based clustering space is established for each point in each cluster of DBSCAN). Then, the loss function can be summarized as shown in (20),

$$\begin{aligned} L(\theta_1,\theta_2) = {} & \beta_1((1-\lambda^t_{id})(L^t_{id}(\theta_1) + L^t_{id}(\theta_2)) + \lambda^t_{id}(L^t_{sid}(\theta_1|\theta_2) \\ & + L^t_{sid}(\theta_2|\theta_1))) + \beta_2((1-\lambda^t_{tri})(L^t_{tri}(\theta_1) + L^t_{tri}(\theta_2)) \\ & + \lambda^t_{tri}(L^t_{stri}(\theta_1|\theta_2) + L^t_{stri}(\theta_2|\theta_1))) + (1-\beta_1-\beta_2) \\ & (L^t_{pull}(\theta_1|\theta_2) + L^t_{pull}(\theta_2|\theta_1)) \end{aligned} \tag{20}$$

Where, $\beta_1, \beta_2, \beta_1 + \beta_2 \in [0,1]$, and $\beta_1, \beta_2, 1-\beta_1-\beta_2$ represent the classification loss, triple loss, and the weight of attraction loss in the loss function; λ^t_{id} is the weight of soft pseudo-labels and hard pseudo-labels in the classification loss function; and λ^t_{tri} is the weight of soft pseudo-labels and hard pseudo-labels in the triplet loss function.

Algorithm 3: Teacher–student model.

Input: The target domain dataset D_t; the pre-training model from Algorithm 2.Weighting factorsβ_1, β_2, λ^t_{id}, λ^t_{tri} for (20), with momentum α from (15).

Output: The network model can be applied to the target domain dataset environment.

for *n in [i,num-epochs]* **do**
- Generate hard pseudo-labels $\widetilde{y^t_i}$ for each sample x^t_i in the target domain dataset D_t by DBSCAN clustering algorithms.
- **for** *each mini-batch* $B \subset D_t$, *iteration* T **do**
  - 1, Generate soft pseudo-labels from the collaborative networks by predicting $\tau_{i\in B}(E^{(T)}[\theta_1])$, $\tau_{i\in B}(E^{(T)}[\theta_2])$, $C^t_1F(x^t_{i\in B}|E^{(T)}[\theta_1])$, $C^t_2F(x'^t_{i\in B}|E^{(T)}[\theta_2])$;
  - 2, Update parameters θ_1, θ_2 by gradient descent based on (20);
  - 3, Update the model weights $E^{(T+1)}[\theta_1]$ and $E^{(T+1)}[\theta_2]$ following (15).
- **end**
- Compare the performance of the two networks; reserve a better performance network.

end

4 Experiments

We conducted experiments on three datasets (i.e., Market-1501 [21], MSMT17 [22], and dukeMTMC [23]) and compare the performance of the developed method with others reported in the literature based on cumulative matching characteristics (CMC) and mean average precision (mAP). An ablation experiment was also carried out to illustrate the importance of each component for improving the performance.

4.1 Preparation

4.1.1 Dataset

The DukeMTMC dataset is a large-scale, labeled, multi-target, multi-camera pedestrian tracking dataset. It provides new large-scale high-definition video data recorded by eight synchronized cameras, with more than 7000 single-camera tracks and more than 2700 independent characters. It samples an image from every 120 frames in the video and obtains 36411 images. A total of 1404 people appeared in images captured by more than two cameras, and 408 people (distractor ID) only appeared in one camera.

The MSMT dataset was proposed on CVPR2018. Specifically, MSMT17 contains 126441 bounding boxes and 4101 identities, including 12 outdoor cameras and three indoor cameras. Four days with different weather conditions were selected in a given month, and three hours of video were collected every day, covering three time periods (i.e., morning, noon, and afternoon), which allowed it to better simulate real scenes and consider more complex lighting changes.

The Market-1501 dataset was collected on the campus of Tsinghua University in the summer; it was built and made public in 2015. This dataset includes five high-definition cameras and one low-definition camera, which together collected images of 1501 pedestrians and 32668 detected pedestrian rectangular frames . Each pedestrian was captured by at least two cameras, and there may be multiple images containing a given pedestrian in each camera. The training set identified 751 people in 12936 images, and each person corresponded to an average of 17.2 training data points. The test set contained 750 people in 19732 images, and each person corresponded to an average of 26.3 test data points. The pedestrian detection rectangles of the 3368 query images were drawn manually, whereas the pedestrian detection rectangles in the gallery were detected using the DPM(Deformable Parts Model) detector (Table 1).

4.1.2 Evaluation Indicators: mAP and CMC

Cumulative match characteristic curves and mean average precision (mAP) are commonly used to evaluate the performance of pedestrian re-identification algorithms.

The CMC curve comprehensively reflects the performance of the classifier, and it can indicate the probability that the matching target appears in a candidate list of size k. Intuitively, the CMC curve can be given in the form of a *Rank-k* accuracy rate, i.e., the probability that the correct match of the target appears in the top k positions of the match list. In the pedestrian re-identification problem, the algorithm performance is typically evaluated when k = 1,5,10. For example, the accuracy of *Rank-1* indicates the probability that the correct match appears at the first place in the matching list, i.e., the probability that the system can return the correct result only by looking it up once. Generally, the final *Rank-k* accuracy rate refers to the average of the results obtained after querying all retrieval targets.

However, when there are multiple correct matches in the test set, the CMC accuracy rate cannot fully evaluate the algorithm, considering that the pedestrian re-identification algorithm should retrieve all of the correct targets. Essentially, while the algorithm considers precision, it should also consider recall. Therefore, mAP is used to account for the retrieval and recall capabilities of the algorithm. Specifically, the mAP calculation process needs to traverse all retrieval targets and calculate the average precision (AP) for each retrieval target to obtain the average. The AP calculation process involves computing the integral of the precision-recall (PR) curve. As a result, mAP considers the precision and recall of the target under certain thresholds. Therefore, mAP and Rank-k accuracy rates are typically used together as an evaluation index for pedestrian re-identification problems to ensure comprehensive evaluations of the algorithm performance.

Table 1 Information from some image-based person re-identification datasets

Dataset	ID	Training set's ID	Training set's images	Test set's ID	Test set's images	Camera
Market1501	1501	751	12936	750	16384	6
DukeMTMC	1404	702	16522	702	17661	8
MSMT17	4101	1041	32621	3060	82161	15

Table 2 Experimental results of the proposed approach and state-of-the-art methods for Market1501, DukeMTMC-ReID, and MSMT17 datasets

Comparison:	Market-to-duke				Duke-to-market			
Method	Map	Top-1	Top-5	Top-10	Map	Top-1	Top-5	Top-10
ECN [1]	40.4	63.3	75.8	80.4	43	75.1	86.3	90.2
SPGAN [11]	22.3	41.1	56.6	63	22.8	51.5	70.1	76.8
MMT [14]	68.7	81.8	91.2	93.4	76.5	90.9	96.4	**97.9**
PUL [24]	16.4	30	43.4	48.5	20.5	45.5	60.7	66.7
TJ-AIDL [25]	23	44.3	59.6	65	26.5	58.2	74.8	81.1
HHL [26]	27.2	46.9	61	66.7	31.4	62.2	78.8	84
CFSM [27]	27.3	49.8	-	-	28.3	61.2	-	-
BUC [28]	27.5	47.4	62.6	68.4	38.3	66.2	79.6	84.5
ARN [29]	33.4	60.2	73.9	79.5	39.4	70.3	80.4	86.3
UDAP [30]	49	68.4	80.1	83.5	53.7	75.8	89.5	93.2
UCDA-CCE [31]	31	47.7	-	-	39	60.4	-	-
PDA-Net [32]	45.1	63.2	77	82.5	47.6	75.2	86.3	90.2
SSG [33]	53.4	73	80.6	83.2	58.3	80	90	92.4
PAUL [34]	35.7	56.1	-	-	36.8	66.7	-	-
UNRN [35]	69.1	82	**90.7**	**93.5**	78.1	91.9	96.1	97.8
GLT [36]	**69.2**	72	90.2	92.8	**79.5**	**92.2**	**96.5**	97.8
SpCL [37]	68.8	**82.9**	90.1	92.5	76.7	90.3	96.2	97.7
Dual-Refinement [38]	67.7	82.1	90.1	92.5	78	90.9	96.4	97.7
This work	***70.2***	***83.4***	***93.2***	***95.1***	***79.5***	***93.2***	***97.4***	***98.8***
Comparison:	Market-to-MSMT				Duke-to-MSMT			
Method	Map	Top-1	Top-5	Top-10	Map	Top-1	Top-5	Top-10
ECN [1]	8.5	25.3	36.3	42.1	10.2	30.2	41.5	46.8
SSG [4]	13.2	31.6	-	49.6	13.3	32.2	-	51.2
PTGAN [5]	2.9	10.2	-	24.4	3.3	11.8	-	27.4
LAIM [12]	15.2	40.4	53.1	58.7	16	42.5	55.9	61.5
AE [13]	9.2	25.5	37.3	42.6	11.7	32.3	44.4	50.1
MMT [14]	26.3	52.5	66.3	71.7	**29.7**	58.8	**71**	**76.1**
UNRN [35]	25.3	52.4	64.7	69.7	26.2	54.9	67.3	70.6
GLT [36]	26.5	**56.6**	**67.5**	**72**	27.7	**59.5**	70.1	74.2
SpCL [37]	**26.8**	53.7	65	69.8	26.5	53.1	65.8	70.5
Dual-Refinement [38]	25.1	53.3	66.1	71.5	26.9	55	68.4	73.2
This work	***30.2***	***58.2***	***71.3***	***77.4***	***33.4***	***62.1***	***75.5***	***79.8***

Bold entries indicate the best performance of other methods and the performance of our method

Table 3 Comparison of the impacts of pseudo-labels generated based on K-means-500, -600, and -700, and the DBSCAN clustering algorithm on model performance

	Duke-to-market			
Ablation experimental details	Map	Top-1	Top-5	Top-10
ResNet-K-means-500 without transformer	**73.2**	88.3	95	97.1
ResNet-K-means-600 without transformer	70.6	86.3	95.1	96.7
ResNet-K-means-700 without transformer	66.5	85.4	94.5	96.9
ResNet-DBSCAN without transformer	73	**88.6**	**96**	**97.8**

Bold entries mean the best performance in this experiment

Table 4 The influence of SPGAN on the extraction of discriminative features shared by the source domain and target domain datasets during the pre-training stage of the model

	Duke-to-market			
Ablation experimental details	Map	Top-1	Top-5	Top-10
ResNet-50 pre-training without SPGAN	27.2	54.3	71	77.3
ResNet-50 pre-training with SPGAN	**35.4**	**63.7**	**78.3**	**82.4**

Bold entries mean the best performance in this experiment

4.2 Comparative experiments

Four sets of comparative experiments were carried out: market-to-duke, duke-to-market, market-to-MSMT, and duke-to-MSMT (Table 2). The significance of duke-to-market is that dukeMTMC is used as a labeled source domain dataset to pre-train the model, and the teacher–student model is used to make the network applicable to the unlabeled target domain dataset, market1501, and vice versa.

All hyper-parameters were selected based on the verification machine of the duke-to-market task, the number of pseudo-labels was 500, and IBN-ResNet50 comprised the basic framework. The same hyper-parameters were also applied to the other three areas.

These comparison experiments revealed that the method developed in this study demonstrated excellent performance in terms of mAP and CMC.

4.3 Ablation experiments

To confirm the role of each module in the developed model, ablation experiments were designed using the duke-to-market results for a self-comparison of the performance (Table 2).

When the number of K-means settings was 500, the model achieved the same performance as with DBSCAN. However, it is important to recall that re-ID is an open class problem, meaning that there is an unknown number of identities in the environment in practical applications. Therefore, the samples in the source domain dataset must be independent from the samples in the target domain dataset, i.e., exemplar invariance and camera invariance must be taken into account (Table 3).

By observing the changes in the mAP of the source and target domains that were not processed by SPGAN during the pre-training process (Fig. 6), we determined that although the model had poor performance in the target domain, its performance gradually increased as the training improved. This indicated that although the two pictures had different styles, they shared some discriminative features. It was therefore confirmed that SPGAN can accelerate the convergence of the teacher–student model.(Table 4)

Finally, to improve the decoupling ability of the model while saving computational space, we replaced the last two layers of the IBN-based ResNet with Transformer and compared their performance. From Table 5, it is clear that Transformer can improve the accuracy of the algorithm.

5 Conclusion

In this study, we build a semi-supervised pedestrian re-identification system based on the teacher-student model and SPGAN. It enables the pedestrian re-identification system to be trained with a small amount of labeled data and a large number of unlabeled data in application scenarios, and the performance of the system meets the requirements of the application. Our main work is: 1. Proposed a new loss function so that the teacher-student model can satisfy the instance invariance, camera invariance and neighborhood invariance during the training process, and reduce the impact of pseudo-label noise on the system. 2. By locally introducing Transformer into ResNet, the decoupling ability

Table 5 The impact of Transformer on model performance

	Duke-to-market			
Ablation experimental details	Map	Top-1	Top-5	Top-10
IBN-ResNet-DBSCAN without transformer	77.9	91.5	97.1	98.2
IBN-ResNet-DBSCAN with transformer	**79.5**	**93.2**	**97.4**	**98.8**

Bold entries mean the best performance in this experiment

of the model is improved and the speed of the system is guaranteed. 3. By introducing SPGAN to process the labeled source domain data set, the pre-training model is provided with more labeled data sets with the same discriminative features as the target domain.

References

1. Zhong Z, Zheng L, Luo ZM, Li SZ, Yang Y (2019) Invariance matters: Exemplar memory for domain adaptive person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 598–607
2. Yu HX, Zheng WS, Wu AC, Guo XW, Gong SG, Lai JH (2019) Unsupervised person re-identification by soft multilabel learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 2148–2157
3. Yang QZ, Yu HX, Wu AC, Zheng WS (2019) Patch-Based discriminative feature learning for unsupervised person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 3633–3642
4. Fu Y, Wei YC, Wang GS, Zhou YQ, Shi HH, Huang TS (2019) Self-Similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 6112–6121
5. Ding GD, Khan S, Tang ZM, Zhang J, Porikli F (2019) Towards better validity:, Dispersion based clustering for unsupervised person re-identification. arXiv:1906.01308
6. Huang Y, Wu Q, Xu JS, Zhong Y (2019) SBSGAN: Suppression of inter-domain background shift for person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 9527–9536
7. Zhong Z, Zheng L, Li SZ, Yang Y (2018) Generalizing a person retrieval model hetero-and homogeneously. In: Proc. of the European Conf. on Computer Vision (ECCV). pp 172–188
8. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J (2018) Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 8789–8797
9. Wei LH, Zhang SL, Gao W, Tian Q (2018) Person transfer gan to bridge domain GAP for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 79–88
10. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 2223–2232
11. Deng WJ, Zheng L, Ye QX, Kang GL, Yang Y, Jiao JB (2018) Image-Image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 994–1003
12. Zhong Z, Zheng L, Luo ZM, Li SZ, Yang Y (2019) Learning to adapt invariance in memory for person re-identification. arXiv:1908.00485
13. Ding YH, Fan HH, Xu ML, Yang Y. (2019) Adaptive exploration for unsupervised person re-identification. arXiv:1907.04194
14. Delorme G, Xu YH, Lathuilière S, Horaud R (2019) Alameda-pineda X. CANU-reID: A conditional adversarial network for unsupervised person re-identification. arXiv:1904.01308
15. Qi L, Wang L, Huo J, Zhou LP, Shi YH, Gao Y. (2019) A novel unsupervised camera-aware domain adaptation framework for person re-identification. In: Proc. of the int'l conf. on computer vision
16. Ge Y, Chen D, Li H. (2020) Mutual mean-teaching:, Pseudo label refinery for unsupervised domain adaptation on person re-identification[J]. arXiv:2001.01526
17. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778
18. Pan X, Luo P, Shi J et al (2018) Two at once:, Enhancing learning and generalization capacities via ibn-net[C]//Proceedings of the European Conference on Computer Vision (ECCV). 464–479
19. Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need[C]//Advances in neural information processing systems. 5998–6008
20. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 3104–3112
21. Zheng L, Shen LY, Tian L, Wang SJ, Wang JD, Tian Q (2015) Scalable person re-identification: A benchmark. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 1116–1124
22. Wei LH, Zhang SL, Gao W, Tian Q (2018) Person transfer gan to bridge domain GAP for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 79–88
23. Zheng ZD, Zheng L, Yang Y (2017) Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 3754–3762
24. Fan H, Zheng L, Yan C et al (4) Unsupervised person re-identification: Clustering and fine-tuning[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14:1–18
25. Wang J, Zhu X, Gong S, Li W (2018) Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR
26. Zhong Z, Zheng L, Li S, Yang Y (2018) Generalizing a person retrieval model heteroand homogeneously. In: ECCV
27. Chang X, Yang Y, Xiang T, Hospedales TM (2019) Disjoint label space transfer learning with common factorised space AAAI
28. Lin Y, Dong X, Zheng L, Yan Y, Yang Y (2019) A bottom-up clustering approach to unsupervised person re-identification. In: AAAI
29. Li Y-J, Yang F-E, Liu Y-C, Yeh Y-Y, Du X, Wang Y-CF (2018) Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In: CVPRW
30. Song L, Wang C, Zhang L, Du B, Zhang Q, Huang C, Wang X (2018) Unsupervised domain adaptive re-identification:, Theory and practice. arXiv:1807.11334
31. Qi L, Wang L, Huo J, Zhou L, Shi Y, Gao Y (2019) A novel unsupervised camera-aware domain adaptation framework for person re-identification. ICCV
32. Li Y-J, Lin C-S, Lin Y-B, Wang Y-CF (2019) Cross-dataset person reidentification via unsupervised pose disentanglement and adaptation. ICCV

33. Fu Y, Wei YC, Wang GS, Zhou YQ, Shi HH, Huang TS (2019) Self-Similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision. pp 6112–6121
34. Yang QZ, Yu HX, Wu AC, Zheng WS (2019) Patch-Based discriminative feature learning for unsupervised person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp 3633–3642
35. Zheng K, Lan C, Zeng W, Zhan Z, Zha Z-J (2021) Exploiting sample uncertainty for domain adaptive person re-identification. In: AAAI
36. Zheng K, Liu W, He L, Mei T, Luo J, Zha Z-J (2021) Group-aware label transfer for domain adaptive person re-identification. In: CVPR, pp 5310–5319
37. Ge Y, Chen D, Zhu F, Zhao R, Li H (2020) Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: NeurIPS
38. Dai Y, Liu J, Bai Y, Tong Z, Duan L-Y (2021) Dual-refinement: Joint label and feature refinement for unsupervised domain adaptive person re-identification. IEEE TIP

Botong Zhao was born in Meihekou, Jilin, China, in 1999. He received the B.E. degree from Heilongjiang University, Harbin, China, in 2019. He is currently pursuing the Master degree with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. His research interests include computer vision, object detection, and Machine Learning.

Yanjie Wang received the M.S degree in mechatronic engineering from Changchun Institution of Optics, Fine Mechanics and Physics (CIOMP), University of China Academy of Science, in 2020. Her research interests include image processing.

Keke Su was born in Luoyang, Henan, China, in 1987. He received the B.E. degree from Henan Institute of Science and Technology, Xinxiang, China, in 2011. He is currently pursuing the Master degree with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. His research interests include computer vision, object detection, and optoelectronic imaging.

Hong Ren received the M.S. degree in optical engineering from the Nanjing University of Science & Technology (NJUST), Nanjing, China, in 2015. Currently, he is pursuing the Ph.D. degree at Changchun Institute of Optics, Fine Mechanics and physics, Chinese Academy of Sciences, Changchun, China. His research interests include aerospace image processing and pose estimation.

Xiyu Han received the PhD degree in Mechatronic Engineering from Changchun Institution of Optics, Fine Mechanics and Physics (CIOMP), University of China Academy of Science, in 2020. Her research interests include image processing.