AIMS Mathematics, 6(5): 4715–4733.
DOI: 10.3934/math.2021277
Received: 09 November 2020
Accepted: 18 February 2021
Published: 24 February 2021

http://www.aimspress.com/journal/Math

Research article

Random style transfer for person re-identification with one example

Yang Li, Tianshi Wang and Li Liu*

School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

* **Correspondence:** Email: liuli_790209@163.com.

Abstract: Person re-identification with only one labeled image for each identify can not eliminate the style variations of different cameras in the same dataset. In this paper, we propose a random style transfer strategy that randomly transforms the labeled images on the one-example person re-identification task. In this strategy, we focus on twofolds: 1) Randomly transform the camera style of labeled images and unlabeled images during the training stage and 2) use the average feature of labeled data and its camera style transform data to estimate pseudo label on unlabeled data. Notably, our strategy exhibits state-of-the-art performance on large-scale image datasets and its Rank-1 accuracy outperforms the state-of-the-art method by 10.3% points on Market-1501, and 8.4% points on DukeMTMC-reID.

Keywords: one-example; person re-identification; semi-supervised learning; style transfer
Mathematics Subject Classification: 68T07, 68U10

1. Introduction

Person re-identification (re-ID) aims to find the person-of-interest across different cameras from the gallery when the person-of-interest is given. With the development of surveillance equipment and the increasing demand for public safety, many camera networks have been installed in public places such as theme parks, airports, streets and university campuses. Therefore, there is an urgent need to develop an intelligent technology for monitoring image analysis. At present, re-ID has been widely used in the security field and has become the focus of academic research. However, person re-ID faces more challenges such as lighting, pose, viewpoint, camera variation, and so on. At the early work of person re-ID, there is only a few small dataset, and the texture feature, color feature, and hand-crafted feature are used in the field of person re-ID. In recent years, with the development of deep Convolutional Neural Networks (CNN) and the emergence of large-scale datasets, CNN-based deep learning models have made a great success in the field of re-ID. Most person re-ID methods focus on supervised learning [1–7]. These methods rely on labeled datasets, that is, each pedestrian in the

training data has an identity label. However, due to the high cost of labeling large-scale datasets, the semi-supervised person re-ID has been proposed.

The semi-supervised person re-ID methods use part of the labeled data and part of the unlabeled data for training. We focus on one-example setting, i.e., each identity has only one labeled example. The setting can use the labeled data to estimate the pseudo-label of unlabeled data by calculating the feature distance between the labeled data and the unlabeled data. Then the reliable pseudo-label data are selected and added to the label data to train the model together with unlabeled data. However, due to the difference in the camera frames, the lighting, and the location of the camera, the training dataset under one-example setting will be affected by the style variation between the cameras and reduce the model performance. This problem can be solved by style transfer.

The style transfer method is to use the Generative Adversarial Network (GAN) to transfer the labeled pedestrian image style to the unlabeled test data and use it to train the model. It can better solve the problem of performance degradation when training and testing on different datasets or between different cameras in the same dataset. At the same time, this method also avoids the tedious work of labeling data on the unlabeled test data and reduces the expensive cost of labeling data on the unlabeled test data and it is widely used in supervised person re-ID or semi-supervised person re-ID. Figure 1 shows the style transfer from random cameras to other cameras in the same dataset and the label of the generated image is the same as the original image. Therefore, we propose a random style transfer strategy that effectively solves the problem of camera style variation in the one-example setting.

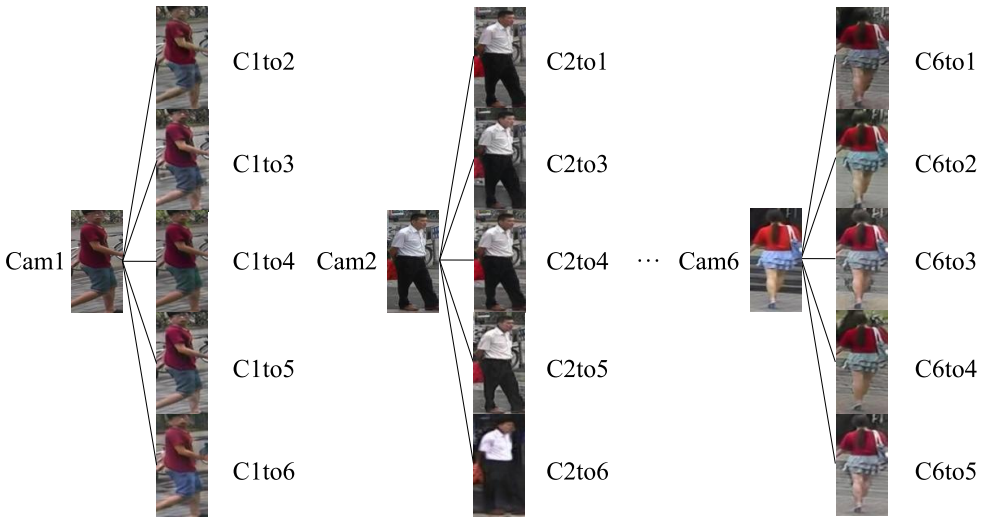


Figure 1. An image can be generated with the help of CycleGAN to generate images with different camera styles.

In the one-example person re-ID task, Wu et al. [8] use part of the labeled person images and part of the unlabeled person images for training. During training, an image is randomly selected as the labeled data for each identity under the camera with the smallest number, and the rest is used as unlabeled data. Then these two kinds of data are used for training. However, the method of [8] cannot eliminate

the domain difference between cameras in the training phase and the label evaluation phase. The random style transfer strategy we propose is to randomly transform the image style of the labeled data and unlabeled data to other camera styles without changing the person identity label during training. This can solve the problem of camera style variation in the same training dataset, thereby improving performance. Besides, when evaluating pseudo-labels, Wu et al. [8] use the features of the labeled data as the evaluation features. However, we adopt an average feature method to average the features of the labeled data and the features after the camera style transformation, as the final evaluation feature. In the same way, there will be a better improvement. Our method can obtain a more robust initialized CNN model, and in subsequent iterations, it will obtain better performance.

Our contributions are summarized as follow:

- We propose a random style transfer strategy to transform the camera style of the training data, which eliminates the style variation between different cameras in the same datasets and obtains a better initialized CNN model.
- We adopt an average feature strategy to estimate the pseudo-labels of unlabeled data, and the estimation results are improved.
- Our method has achieved good results on commonly used datasets.

2. Related work

2.1. Supervised person re-ID

In recent years, with the advent of the CNN model [9] and the emergence of large datasets, methods based on deep learning have been widely used in computer vision tasks [10–12], including person re-ID. More methods [13–19] are used for person re-ID tasks and achieve good performance results. Deng et al. [20] use triplets as input on the siamese model. Zheng et al. [21] use a pair of images as the input of the network, and indicates whether the two images are a person according to the similarity value of the output. Ahmed et al. [22] propose a joint learning framework that combines end-to-end re-ID learning and data generation to make better use of the generated data. Zheng et al. [18] use a conventional fine-tuning approach called the Identity Discriminative Embedding (IDE) on the Market-1501 dataset and finally obtain competitive results. Zheng et al. [21] combine the verification model and recognition model to learn more differentiated pedestrian descriptors with a pair of training images. Huynh-The et al. [23] learn the full gait information of an individual by comprehensively studying gait information from 3D human skeleton data with a deep learning-based identifier.

2.2. Semi-supervised Person re-ID

Semi-supervised learning methods [24–27] use partially labeled data and partially unlabeled data to solve a given task. In recent years, some semi-supervised person re-ID [28] tasks use the Progressive Cross-camera Soft Label Learning (PCSL) framework. Kipf et al. [29] use a scalable method of semi-supervised learning on graph structure data. Yu et al. [30] propose an asymmetric metric clustering to discover potential labels in unlabeled target data. Liu et al. [31] use K-Nearest Neighbor (KNN) to update the classifier. Ye et al. [32] adopt a Dynamic Graph Matching (DGM) method to iteratively update graph matching and label estimation. Liu et al. [33] propose a newemantics-Guided Clustering with Deep Progressive Learning (SGC-DPL) to gradually enhance the labeled training data.

In this paper, we follow the one-example setting person re-ID as in [8]. It assumes that only one image of each person in the training set is labeled, while the rest of the data in the training set is not labeled. In [8], a progressive sampling strategy is proposed to increase the number of the selected pseudo-labeled candidates step by step. However, it lacks consideration of the cameras variation when estimating the pseudo-label. This can be solved with style transfer.

2.3. Unsupervised domain adapation for person re-ID

The unsupervised domain adaptation re-ID uses an auxiliary source dataset to label an unlabeled target dataset. Recently, some cross-domain learning methods [34–38] have emerged. Peng et al. [34] propose an asymmetric multi-task dictionary learning model to learn a discriminative represtation for target data. Fan et al. [35] propose a learning via translation framework to reduce the performance deviation after dataset conversion and the unsupervised self-similarity and domain-dissimilarity to ensure potential ID information. Zhong et al. [38] propose to learn camera invariance and domain connectedness simultaneously to improve the generalization ability of re-ID models on the target testing set. Xiang et al. [39] propose a new unsupervised Re-ID method through domain adaptation to use synthetic data to get rid of heavy data annotation and improve the performance of re-identification in a completely unsupervised way. Ge et al. [40] propose an unsupervised framework, namely Mutual Mean-Teaching (MMT), to reduce the inevitable label noise caused by the clustering procedure.

Style transfer refers to the transfer of the style of image A to image B to obtain a new image. The new image contains both the content of image B and the style of image A. The conversion process is achieved through GAN. Since Goodfellow et al. propose GAN [41], many variants of GAN [42–46] have been proposed to handle different tasks, such as natural style conversion, super-resolution, image conversion etc. Zheng et al. [47] use GAN to generate new samples for data augmentation of person re-ID, which is an early work on personnel migration done by GAN for person re-ID. Isola et al. [48] propose a conditional adversarial network to learn the mapping function from input to output image. However, this method requires paired training data, which is difficult to obtain in many tasks. Therefore, for the task of unpaired image-to-image conversion, Zhu et al. [49] propose a loop consistency loss training unpaired data. Later, Person Transfer GAN (PTGAN) [50] proposed by Wei et al. is similar to Cycle-GAN [49], and it can also perform image-to-image conversion. The difference is that, in order to ensure that the transmitted images can be used for model training, additional constraints are imposed on the identity of the person. In Camstyle [6], CycleGAN is used to transfer the labeled training image style to each camera, and form an enhanced training set with the original training samples. Zhang et al. [51] propose a novel semi-supervised re-ID by Similarity-Embedded Cycle GANs (SECGAN), which can learn cross-view features with limited labeled data by using cycle GAN. Our work focuses on using style transfer on one-example setting to eliminate the cameras variation. Liu et al. [52] propose a UnityStyle adaption method to solve the problem of more image artifacts when the difference between the images taken by different cameras is large. Chong et al. [53] propose an unsupervised domain-adaptive person re-identification method based on style transfer (STReID) to solve the potential image distinctions between different domains.

3. Method

This section consists of five parts: We first describe the process of using CycleGAN [49] to generate camera conversion data in subsection 3.1. We then introduce the preliminary work in subsection 3.2. The random style transfer (RST) strategy and average feature estimation (Avg) strategy are described in subsection 3.3 and subsection 3.4 respectively. The last subsection shows the overall progressive iteration strategy. The overall framework is shown in Figure 2. In each iteration, (1) In the training phase, we use the style transfer dataset(Cam set) to perform random style transfer on labeled data, pseudo-labeled data and unlabeled data to train the CNN model. We train label data and pseudo-label data through the Cross-Entropy loss, and train unlabeled data through the Exclusive Loss. (2) In the label estimation stage, we first use style transfer data to average the features of the labeled data. Then, according to the distance in the feature space, some reliable pseudo-label candidates are selected from the unlabeled data U. Nodes with different colors in the feature space frame represent different identification samples.

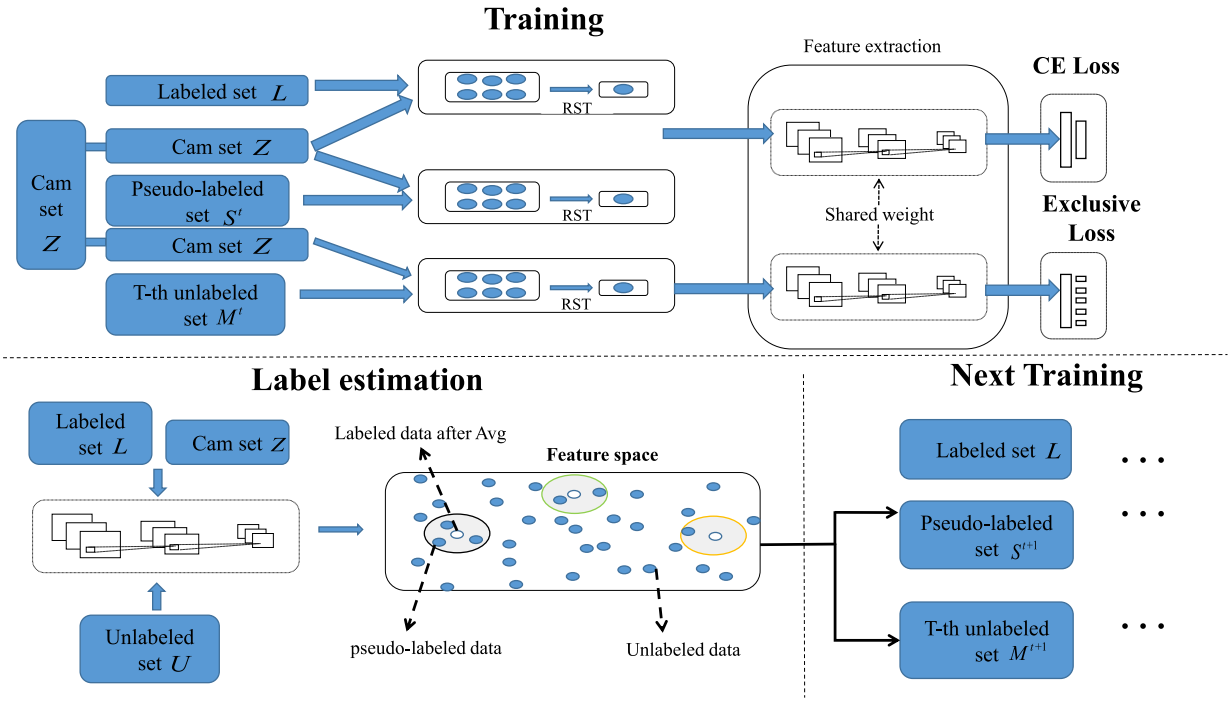


Figure 2. An overview of the framework.

3.1. Camera style conversion

3.1.1. CycleGAN review

Given two datasets $\{x\}_{i=1}^{M}$ and $\{y\}_{j=1}^{N}$ collected from domain A and domain B, $\{x\}_{i=1}^{M}$ belongs to A and $\{y\}_{j=1}^{N}$ belongs to B. The goal of GAN is to learn a generator G and a discriminator D, where $G : A \rightarrow B$ is to realize the conversion of the image from domain A to domain B, that is, $G(A) \approx B$, D is to identify whether the image is from another domain conversion. CycleGAN contains two mapping functions

$G_{A2B} : A \to B$ and $G_{B2A} : A \to B$. Two adversarial discriminators D_A and D_B. The overall loss function of CycleGAN is as follows:

$$\begin{aligned} L(G_{A2B}, G_{B2A}, D_A, D_B) =& L_G(G_{A2B}, D_B, A, B) \\ &+ L_G(G_{B2A}, D_A, B, A) \\ &+ \lambda L_{cyc}(G_{A2B}, G_{B2A}, A, B) \end{aligned} \tag{3.1}$$

where $L_G(G_{A2B}, D_B, A, B)$ and $L_G(G_{B2A}, D_A, B, A)$ are the loss functions for the mapping functions G_{A2B} and G_{B2A} and for the discriminators D_B and D_A. $L_{cyc}(G_{A2B}, G_{B2A}, A, B)$ is the cycle consistency loss, in which each image can be reconstructed after a cycle mapping. λ weighs the importance of L_G and L_{cyc}. More details about CycleGAN can be accessed in [49].

3.1.2. Image-to-Image conversion

CycleGAN is employed to transform the style of one domain into the style of another domain to realize the generation of different camera data. Given a re-ID dataset collected from C cameras, the styles between different cameras are regarded as different domains and CycleGAN is used to learn the image-image conversion model of each camera pair to realize the generation of data of different camera styles.

In this paper, we need to generate camera style transfer data. There are many deep learning models that can realize style transfer data generation. This paper only uses CycleGAN as an example.

Using the learned CycleGAN model, for the training images collected from a specific camera, we generate $C-1$ new training samples. Figure 3 shows pictures with other camera styles generated by the Market-1501 dataset with the help of CycleGAN. For example, the picture in cam1 uses CycleGAN to generate a picture of cam2 style. The cam1to2 retains the pedestrian of cam1 but changes the camera style of cam1 to the camera style of cam2. There are 6 cameras in the Market-1501 dataset. Under the action of CycleGAN, each image generates 5 other camera-style images. Compared with the original image, the generated image has the same identity labeled, only the camera style changes. In this work, we will generate each style conversion image to retain the content of the original image and have the same identity as the original image. This kind of data is called CameraStyle data.

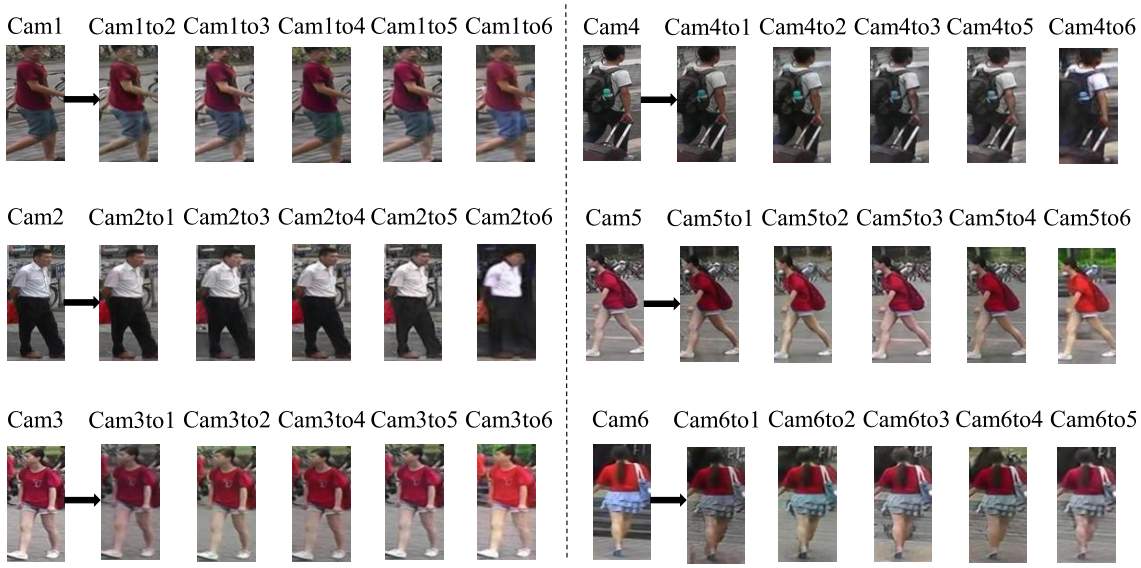


Figure 3. Examples of style transfer on Market-1501 [18].

3.2. Preliminaries

We first introduce the necessary symbols for the one-example re-ID task. Let x and y represent the person image and the identity label, respectively. For the training of one-example re-ID task, we have a labeled dataset $L = \{(x_1, y_1), \ldots, (x_{n_l}, y_{n_l})\}$, an unlabeled dataset $U = \{x_{n_l+1}, \ldots, x_{n_l+n_u}\}$, and a CamStyle dataset Z. In the training phase, these data are used to train the re-ID model $\phi(\theta, \cdot)$ in the form of identity classification. In the evaluation phase, the trained CNN model ϕ is used to embed query data and gallery data in the feature space. The query result is a ranking list of all gallery data based on the Euclidean distance between the query data and each gallery data, *i.e.*, $\left\|\phi(\theta; x_q) - \phi(\theta; x_g)\right\|$, where x_q and x_g represent the query image and gallery image, respectively. And we denote S^t and M^t as the pseudo-labeled dataset and unlabeled dataset of the tth step, respectively.

To utilize the abundant unlabeled data, we use the joint training method [8] in the training phase to perform joint training of labeled data, pseudo-labeled data, and unlabeled data. The objective function is as follows:

$$\begin{aligned} &\min_{\theta,\omega} \lambda \sum_{i=1}^{n_l} \ell_{CE}(f(\omega; \phi(\theta; x_i)), y_i) \\ &+ \lambda \sum_{i=n_l+1}^{n_l+n_u} s_i^{t-1} \ell_{CE}(f(\omega; \phi(\theta; x_i)), \hat{y}_i) \\ &+ (1-\lambda) \sum_{i=n_l+1}^{n_l+n_u} (1 - s_i^{t-1}) \ell_e(V; \tilde{\phi}(\theta; x_i)) \end{aligned} \tag{3.2}$$

where $\min\limits_{\theta,\omega} \sum\limits_{i=1}^{n_l} \ell_{CE}(f(\omega; \phi(\theta; x_i)), y_i)$ represents the optimized part of the labeled dataset L,

$\min\limits_{\theta,\omega} \sum\limits_{i=n_l+1}^{n_l+n_u} s_i^{t-1}\ell_{CE}(f(\omega;\phi(\theta;x_i)),\hat{y}_i)$ represents the optimized part of the pseudo-labeled dataset S^t, and $\min\limits_{\theta,\omega} \sum\limits_{i=n_l+1}^{n_l+n_u} (1-s_i^{t-1})\ell_e(V;\tilde{\phi}(\theta;x_i))$ represents the optimized part of the unlabeled dataset M^t. $f(w;\cdot)$ is an identity classifier, parameterized by w, to classify the embedded feature $\phi(\theta;x_i)$ into a k-dimension confidence estimation. $s_i \in \{0,1\}$ is the selection indicator for the unlabeled sample x_i. ℓ_{CE} and ℓ_e represent identity classification loss and exclusive loss, respectively, and λ is a hyper-parameter used to adjust their contribution. More details about joint training method can be accessed in [8].

3.3. The random style transfer strategy

We first introduce the types of data used in training. In the tth iteration, we will use four kinds of data: label data L, pseudo label data S^t, unlabeled data M^t, and CameraStyle data Z. Before the training starts, we need to generate the CamStyle dataset. We use CycleGAN mentioned in subsection 3.1 to generate one corresponding picture of other camera styles for each picture in the training dataset, that is, we keep the person label, and only change the camera style to other camera styles.

The random style transfer strategy is to perform a random conversion of the camera style with the help of CamStyle data for each piece of labeled data, pseudo-labeled data, and unlabeled data during the training process:

$$x_k \to \tilde{x}_k, \tilde{x}_k \in \{x_k\} \cup \{x|x \in Z_k\} \tag{3.3}$$

where x_k represents a picture in the dataset, and Z_k represents pictures of other cameras corresponding to A in the CamStyle dataset, so $\tilde{x}_k$ represents a camera-style image of A randomly converted, including itself.

After adopting a random style conversion strategy for training, (3.2) will be optimized as:

$$\begin{aligned}&\min_{\theta,\omega} \lambda \sum_{i=1}^{n_l} \ell_{CE}(f(\omega;\phi(\theta;\tilde{x}_i)),y_i)\\&+\lambda \sum_{i=n_l+1}^{n_l+n_u} s_i^{t-1}\ell_{CE}(f(\omega;\phi(\theta;\tilde{x}_i)),\hat{y}_i)\\&+(1-\lambda)\sum_{i=n_l+1}^{n_l+n_u}(1-s_i^{t-1})\ell_e(V;\tilde{\phi}(\theta;\tilde{x}_i))\end{aligned} \tag{3.4}$$

where $\min\limits_{\theta,\omega} \sum\limits_{i=1}^{n_l} \ell_{CE}(f(\omega;\phi(\theta;\tilde{x}_i)),y_i)$ represents the optimized part of the labeled dataset L after random style transfer, $\min\limits_{\theta,\omega} \sum\limits_{i=n_l+1}^{n_l+n_u} s_i^{t-1}\ell_{CE}(f(\omega;\phi(\theta;\tilde{x}_i)),\hat{y}_i)$ represents the optimized part of the pseudo-labeled dataset S^t after random style transfer, and $\min\limits_{\theta,\omega} \sum\limits_{i=n_l+1}^{n_l+n_u} (1-s_i^{t-1})\ell_e(V;\tilde{\phi}(\theta;\tilde{x}_i))$ represents the optimized part of the unlabeled dataset M^t after random style transfer.

Therefore, the training data will reduce the domain difference caused by different cameras, thereby making the initial training model more robust and making the subsequent training process more effective.

3.4. The average feature estimation strategy

Previous works use the distance between labeled data and unlabeled data in the feature space as a measure of the reliability of pseudo-labels. For label estimation of unlabeled data, the nearest neighbor (NN) classifier is used to assign pseudo labels to each unlabeled data through the nearest labeled neighbor in the feature space [8]. However, it is difficult to eliminate the domain difference caused by different cameras by just using a labeled picture feature to calculate the distance from the unlabeled data. To solve this problem, we propose the average feature estimation strategy. For label estimation of unlabeled data, we find other camera-style pictures corresponding to the labeled picture in CamStyle data, and then take the average feature of all the pictures. Finally, the distance between the average feature and the unlabeled data is used as a measure of the reliability of the pseudo-label.

We evaluate pseudo-labels for each unlabeled data $x_i \in U$ by:

$$x^*, y^* = \arg\min_{(x_l, y_l)} ||\phi(\theta; x_i) - \phi(\theta; \tilde{x}_l)|| \tag{3.5}$$

$$d(\theta; x_i) = ||\phi(\theta; x_i) - \phi(\theta; x^*)|| \tag{3.6}$$

$$\hat{y}_i = y^* \tag{3.7}$$

where $\tilde{x}_l$ is the average of label data $x_l \in L$ and other camera pictures in the corresponding CamStyle data. Where $d(\theta; x_i)$ is the dissimilarity cost of label estimation. In order to select candidates, in iterative step t, we sample pseudo-labeled candidates into training by setting the selection indicators as follows:

$$s^t = \arg\min_{\|s^t\|_0 = m_t} \sum_{i=n_l+1}^{n_l+n_u} s_i d(\theta; x_i) \tag{3.8}$$

where m_t denotes the size of selected pseudo-labeled set. s^t is the vertical concatenation of all s_i. (3.8) selects the top m_t nearest unlabeled data for all the labeled data at the iteration step t.

3.5. The overall progressive iteration strategy

We train the CNN model iteratively. In the initial iteration, we only use the labeled data with a random style transfer strategy to initialize the model. Then in each subsequent iteration, we first optimize the model through (3.4). Then we use (3.7) to estimate the pseudo label of the unlabeled data, and select some reliable pseudo-label data by applying the trained model on (3.8).

When selecting pseudo-labels, we adopt a dynamic sampling strategy to ensure the reliability of the selected pseudo-labeled samples. It starts with a small amount of pseudo-labeled data in the initial stage, and then merges more samples in the following stages. We set the sampled pseudo-labeled data $m_0 = 0$ and unlabeled data $M^0 = U$ at the beginning. In subsequent iterations, we gradually increase the size of the selected pseudo-labeled candidate set S^t. In iterative step t, we expand the size of the sampled pseudo-labeled data by setting $m_t = m_{t-1} + p \cdot n_u$, where $p \in (0, 1)$ is the selection factor, which represents the speed of magnifying the candidate set during the iteration. As described in Algorithm 1,

Algorithm 1 The proposed method

Require: Labeled data L, unlabeled data U, selection factorr $p \in (0, 1)$ initialized CNN model θ_0.
Ensure: The best CNN model θ^*.

Initialize the selected pseudo-labeled data $S_0 \leftarrow \emptyset$, sampling size $m_1 \leftarrow p \cdot n_u$, iteration step $t \leftarrow 0$, best validation performance $V^* \leftarrow 0$.
while $m_{t+1} \leq \|U\|$ **do**
    $t \leftarrow t + 1$
    Update the model (θ_t, w_t) on L, S^t and M^t after random style transfer via (3.4).
    Estimate pseudo labels for U via (3.7)
    Generate the selection indicators s_t via (3.8)
    Update the sampling size: $m_{t+1} \leftarrow m_t + p \cdot n_u$
end while
for $i \leftarrow 1$ to T **do**
    Evaluate θ_i on the validation set $\rightarrow$ performance V_i
    if $V_i > V^*$ **then**
        $V^*, \theta^* \leftarrow V_i, \theta_i$
    end if
end for

4. Experiment

We first introduce the datasets and settings show the results compared with advanced methods, and finally introduce ablation experiments and result analysis.

4.1. Datasets and settings

4.1.1. Datasets

Market-1501 [18] contains 32,668 pieces of labeled data with 1,501 identities taken under 6 cameras. Among them, 12,936 labeled images with 751 identities are used as the training set, and 19,732 labeled images with 750 identities are used as the test set.

DukeMTMC-reID [21] is a re-ID dataset derived from the DukeMTMC dataset [54]. It contains 36,411 pieces of labeled data with 1,404 identities taken under 8 cameras. Among them, there are 16,522 training set images with 702 identities, 2,228 query images with 702 identities, and 17,661 gallery images.

Camarket is a dataset of the other 5 camera styles of the Market-1501 dataset generated by CycleGAN [49], with a total of 64,680 images. It's just a collection of images in the training set under other 5 camera styles. The generation of Camarket dataset can refer to [6].

Camduke is a dataset of the other 7 camera styles of the duke dataset generated CycleGAN [49], with a total of 115,654 images. It's just a collection of pictures in the training set under other 7 camera styles. The generation of Camduke dataset can refer to [6].

4.1.2. Evaluation metrics

We employ the Cumulative Matching Characteristic (CMC) curve and the mean average precision (mAP) for re-ID evaluation. The CMC scores reflect the accuracy of response retrieval and we use Rank-1, Rank-5, Rank-10 scores to represent the CMC curve. Rank-1 recognition rate means that after matching according to a certain similarity matching rule, the ratio of the number of tests with the correct label to the total number of test samples can be judged for the first time. Rank-5 means that there are five opportunities to judge. The mAP is the average of the Average Precision (AP) for each query, and the AP refers to the average of precision.

4.1.3. Experiment setting

For one-example experiments, we use the same protocol as [8]. In all datasets, we randomly select an image from camera 1 as the labeled data for each identity. If camera 1 does not record any data for an identity, we will randomly select a sample from the next camera to ensure that each identity has a sample as labeled data. Other data are used as unlabeled data. Before training, we first form a set, which contains each picture and its corresponding other camera style pictures on the camarket/camduke data. During training, labeled data and unlabeled data will be randomly converted into one picture in the corresponding set. In this paper, we only use an image (single-shot) as input, not a video (multi-shot).

4.1.4. Parameter setting

We use ResNet-50 (with the last classification layer removed) as the feature embedding model for all experiments. We initialize it with ImageNet [57] pre-trained model. We set the temperature scalar τ to 0.1. The setting of λ will be discussed in subsection 4.1. In each model update step, Stochastic Gradient Descent (SGD) with a momentum of 0.5 and a weight decay of 0.0005 is used to optimize the parameters of 70 epochs with a batch size of 16. The total learning rate is initialized to 0.1. In the last 15 epochs, in order to stabilize the model training, we change the learning rate to 0.01 and set $\lambda = 1$.

4.2. Comparison with the state-of-the-art methods

The recent work on one-example person re-ID did not eliminate the domain variation between different cameras. Our method solves this problem well. The re-ID performance of our method on the two large-scale re-ID datasets are summarized in Table 1. We use a selection factor of $p = 0.05$. The baseline($\lambda = 0.8$) is a model training based on one-example label data. Ours(RST, 0.8) represents the result after using the random style transfer strategy when $\lambda = 0.8$. Ours(RST+Avg, 0.8) represents the result of adopting random style transfer strategy and average feature estimation strategy when $\lambda = 0.8$. Ours(RST+Avg, 0.9) represents the best result after adjusting the hyperparameters $\lambda = 0.9$. Even if there is only one labeled example for each identity, our method achieves amazing performance on image-based re-ID tasks, *i.e.*, we achieve 10.3% and 8.4% points of Rank-1 accuracy improvement over the Baseline (one-example) on Market-1501 and DukeMTMC-reID, respectively. The performance on two large-scale datasets proves the effectiveness of our method.

We compare our method with the state-of-the-art image-based person re-ID approaches. Among them, there are four hand-crafted feature representation methods (LOMO [7], BoW [18], UDML [34], ISR [55]), and multiple deep-learning-based methods. The latter includes three recent pseudo-label-learning-based methods(CAMEL [30], PUL [35], TJ-AIDL [36]), three

domain-adaptation-based methods(PTGAN [50], SPGAN [37], HHL [38]) and two recent one-example-based methods (Rank [56], Baseline [8]). The results show that our method is the best compared with all methods on the two datasets. The main reason for the gap with the hand-crafted feature method is that most of the early works are based on heuristic design, so they cannot learn the best distinguishing features. Our method is superior to the unsupervised re-ID method based on pseudo-label learning. The main reason is that our average feature strategy on KNN can better assign pseudo-labels to unlabeled data. On the contrary, the pseudo-label learning of [30, 35] directly compares the visual features, ignoring potential label information and camera variation. Compared with domain-adaptation-based approaches, our approach achieves superior performance. A key reason is that we have adopted style transfer and information mining for unlabeled data. After adopting the random style transfer strategy (Ours(RST, 0.8)), on the market-1501 dataset, Rank-1 and mAP increase by 7.5% and 4.2% than Baseline [8], respectively; on the DukeMTMC-reID dataset, Rank-1 and mAP increase by 4.8% and 1.5% than Baseline [8], respectively. On this basis, after using the average feature estimation strategy (Ours(RST+Avg, 0.8)), the Rank-1 and mAP of the market-1501 dataset and DukeMTMC-reID dataset can be improved again. After adjusting the hyperparameter $\lambda = 0.9$, the best performance can be obtained: on the market-1501 dataset, Rank-1 and mAP increase by 10.3% and 6.5% than Baseline [8], respectively; on the DukeMTMC-reID dataset, Rank-1 and mAP increase by 8.4% and 4.0% than Baseline [8], respectively. Compared with the state-of-the-art image-based person re-ID methods, it can be seen that our method performs better than existing methods, and will get higher accuracy in many practical fields. The specific process of adjusting hyperparameters is in the next section.

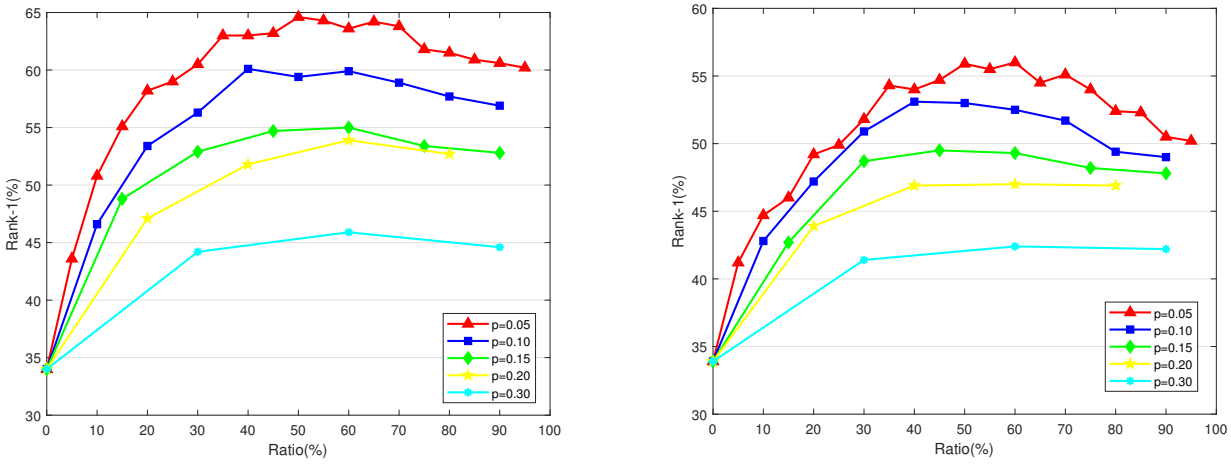
Table 1. Performance comparison.

Methods	Marekt-1501				DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
LOMO [7]	27.2	41.6	49.1	8.0	12.3	21.3	26.6	4.8
BoW [18]	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
UDML [34]	34.5	-	-	12.4	18.5	-	-	7.3
ISR [55]	40.3	62.2	-	14.3	-	-	-	-
CAMEL [30]	54.5	73.1	-	26.3	40.3	57.6	-	19.8
PUL [35]	45.5	60.7	66.7	20.5	30.0	43.4	48.5	16.4
TJ-AIDL [36]	58.2	74.8	81.1	26.5	44.3	59.6	65.0	23.0
PTGAN [50]	38.6	57.3	-	15.7	27.4	43.6	-	13.5
SPGAN [37]	51.5	70.1	76.8	22.8	41.1	56.6	63.0	22.3
SPGAN+LMP [37]	57.7	75.8	82.4	26.7	46.4	62.3	68.0	26.2
HHL [38]	62.2	78.8	84.0	31.4	46.9	61.0	66.7	27.2
Rank [56]	26.0	41.4	49.2	9.0	16.4	27.9	32.8	6.8
Baseline [8]	55.8	72.3	78.4	26.2	48.8	63.4	68.4	28.5
Ours(RST, 0.8)	63.3	78.8	83.5	30.4	54.6	66.4	71.2	30.0
Ours(RST+Avg, 0.8)	64.6	78.5	83.0	31.6	55.5	68.1	72.9	31.0
Ours(RST+Avg, 0.9)	**66.1**	**80.0**	**84.2**	**32.7**	**57.2**	**69.7**	**74.4**	**32.5**

4.3. Hyperparameter analysis

4.3.1. Hyperparameter p

p is a key parameter in our framework , and it controls the speed at which pseudo-marked data are selected in the iterative process. A smaller selection factor indicates a lower selection speed, therefore, more iteration steps and training time are required. The results of different selection factors can be found in Figure 4. Figure 4a is the result of Rank-1 with different selection factors on Marekt-1501. Figure 4b is the result of Rank-1 with different selection factors on DukeMTMC-reID. The x-axis represents the ratio of selected data from the entire unlabeled dataset. Here we set $\lambda = 0.8$. In experiments, a smaller selection factor can produce better performance. An important reason is that each selection step is more cautious, so the label estimation is more accurate. We can also find that the gap between the five curves in the first few iterations is relatively small, and the gap gradually increases in the subsequent iterations, which indicates that the estimation error continues to accumulate during the iteration.



(a) Marekt-1501 Dataset

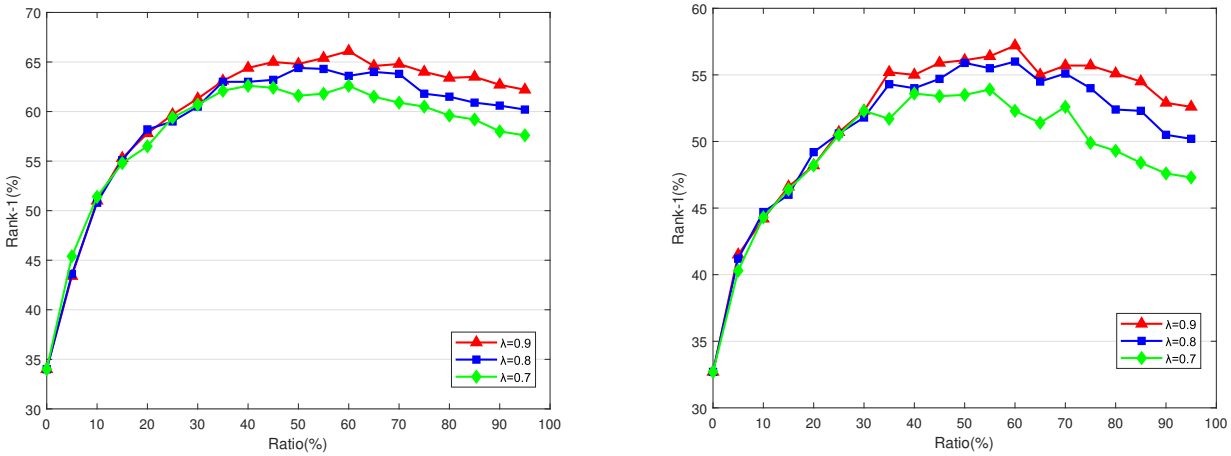
(b) DukeMTMC-reID Dataset

Figure 4. Rank-1 comparison of different selection factor p on Marekt-1501 and DukeMTMC-reID, respectively.

4.3.2. Hyperparameter λ

The hyperparameter λ is the only parameter that adjusts the proportion of labeled data, pseudo-labeled data and unlabeled data in the optimization process. It represents the contribution degree of labeled data and pseudo-labeled data in training. We have selected three values, λ=0.7, λ=0.8 (λ=0.8 in the baseline), λ=0.9, The results with λ=0.7, λ=0.8, λ=0.9, are shown in Figure 5. Figure 5a is the Rank-1 result of different hyperparameter λ on Marekt-1501. Figure 5b is the Rank-1 result of different hyperparameters λ on DukeMTMC-reID. The x-axis represents the ratio of data selected from the entire unlabeled dataset. Here we set $p = 0.05$. As seen, λ=0.9 can achieve the best performance under different selection factors. An important reason is that the random style transfer strategy eliminates camera variation and the average feature estimation strategy makes the label estimation more accurate,

so the labeled data and pseudo-labeled data become more reliable.



(a) Marekt-1501 Dataset (b) DukeMTMC-reID Dataset

Figure 5. Rank-1 comparison of different λ value on Marekt-1501 and DukeMTMC-reID respectively.

4.4. Ablation experiment

4.4.1. The effectiveness of the average feature estimation strategy

The effectiveness of the average feature estimation strategy. The average feature estimation strategy uses the average feature of the labeled data and its corresponding style transfer data to evaluate the pseudo-label. The results of it under different selection factors are shown in Table 2. Here we only take values of 0.05, 0.10, 0.15 for p. Baseline(0.8) and Avg(0.8) represent the result of baseline when λ=0.8 and the result after we only adopt the average feature estimation strategy when λ=0.8, respectively. As seen, using the average feature estimation strategy in the experiment can produce better performance. An important reason is that the average feature is better than the original feature to reduce the impact of camera style variation, so the pseudo label estimation is more accurate.

Table 2. The results of the average feature estimation strategy under different selection factors.

Selection factor	Methods	Marekt-1501				DukeMTMC-reID			
		Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
p=0.05	Baseline(0.8)	55.8	72.3	78.4	26.2	48.8	63.4	68.4	28.5
	Avg(0.8)	**59.5**	**74.9**	**80.7**	**30.1**	**51.7**	**66.2**	**70.7**	**29.1**
p=0.10	Baseline(0.8)	51.5	66.8	73.6	23.2	40.5	53.9	60.2	21.8
	Avg(0.8)	**60.1**	**75.0**	**81.0**	**27.4**	**53.1**	**64.6**	**69.2**	**28.3**
p=0.15	Baseline(0.8)	44.8	61.8	69.1	19.2	35.1	49.1	54.3	18.2
	Avg(0.8)	**51.0**	**68.1**	**74.1**	**23.6**	**51.3**	**63.1**	**67.8**	**26.4**

4.4.2. The effectiveness of the random style transfer strategy

The random style transfer strategy is to randomly change the camera style of labeled data and unlabeled data in the training phase. As shown in Table 3, here p is set to 0.05, Baseline-start(0.8) represents the model initialized in the first iteration when the random style conversion strategy is not used; Ours-start(RST, 0.8) represents the model initialized in the first iteration when the random style conversion strategy is used. As seen, using a random style transfer strategy can achieve better performance. This is because the random style transfer strategy eliminates the style variation between cameras, so the obtained initialization model will be better.

Table 3. The result of model initialization when we adopt the random style transfer strategy.

Methods	Marekt-1501				DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
Baseline-start(0.8)	28.3	43.6	51.0	10.1	17.2	28.1	34.1	7.1
Ours-start(RST, 0.8)	**34.0**	**53.0**	**61.6**	**12.4**	**32.9**	**48.1**	**54.4**	**13.7**

5. Conclusions

In this paper, we propose a random style transfer strategy and the average feature estimation strategy. In the training process, we adopt a random style transfer strategy to randomly change the styles of labeled data and unlabeled data. In the pseudo-label evaluation process, we adopt an average feature estimation strategy to more accurately evaluate pseudo-labels for unlabeled data. These two strategies eliminate the camera style variation during the training process and the pseudo-label evaluation process. The obvious performance improvement proves the effectiveness of our method.

In the future, we hope to use these two strategies to improve the accuracy of video-based person re-ID tasks.

Acknowledgments

The work is partially supported by the National Natural Science Foundation of China (Nos. U1836216, 61772322, 62076153), the major fundamental research project of Shandong, China (No. ZR2019ZD03), and the Taishan Scholar Project of Shandong, China (No. ts20190924).

Conflict of interest

The authors declare there is no conflict of interests.

References

1. X. Zhu, X. Zhu, M. Li, P. Morerio, S. Gong, Intra-Camera Supervised Person Re-Identification, *ArXiv*, 2002.05046, 2020. Available from: https://arxiv.org/abs/2002.05046.

2. Y. Li, Y. Chen, Y. Lin, Y. F. Wang, Cross-Resolution Adversarial Dual Network for Person Re-Identification and Beyond, *ArXiv*, 2002.09274, 2020. Available from: https://arxiv.org/abs/2002.09274.

3. S. M. Saquib, A. Schumann, A. Eberle, R. Stiefelhagen, A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 420–429, .

4. C. Song, Y. Huang, W. Ouyang, L. Wang, Mask-guided contrastive attention model for person re-identification, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 1179–1188.

5. Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang, Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline), J. Serrin, In: *The European Conference on Computer Vision*, 2018, 480–496.

6. Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 5157–5166.

7. S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2015, 2197–2206.

8. Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, Y. Yang, Progressive learning for person re-identification with one example, *IEEE Transactions on Image Processing,* **28** (2019), 2872–2881.

9. T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2016, 1249–1258.

10. L. Zhu, Z. Xu, Y. Yang, A. G. Hauptmann, Uncovering the temporal context for video question answering, *Int. J. Comput. Vision,* **124** (2017), 409–421.

11. C. Deng, Z. Chen, X. Liu, X. Gao, and D. Tao, Triplet-based deep hashing network for cross-modal retrieval, *IEEE Trans. Image Proc.,* **27** (2018), 3893–3903.

12. X. Dong, Y. Yan, M. Tan, Y. Yang, I. W. Tsang, Late fusion via subspace search with consistency preservation, *IEEE Trans. Image Proc.,* **28** (2018), 518–528.

13. W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: Deep filter pairing neural network for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2014, 152–159.

14. Y. Lee, S. Chen, J. Hwang, Y. Hung, An ensemble of invariant features for person reidentification, *IEEE Trans. Circuits Syst. Video Technol.,* **27** (2016), 470–483.

15. Z. Feng, J. Lai, X. Xie, Learning view-specific deep networks for person re-identification, *IEEE Trans. Image Proc.,* **27** (2018), 3472–3483.

16. Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, A bottom-up clustering approach to unsupervised person re-identification, J. Serrin, In: *The AAAI Conference on Artificial Intelligence*, 2019, 8738–8745.

17. M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, Large scale metric learning from equivalence constraints, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2012, 2288–2295.

18. L. Zheng, L. Shen, Scalable person re-identification: A benchmark, J. Serrin, In: *The IEEE international conference on computer vision*, 2015, 1116–1124.

19. Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, J. Kautz, Joint discriminative and generative learning for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2019, 2138–2147.

20. C. Deng, Z. Chen, X. Liu, X. Gao, D. Tao, Triplet-based deep hashing network for cross-modal retrieval, *IEEE Trans. Image Proc.,* **27** (2018), 3893–3903.

21. Z. Zheng, L. Zheng, Y. Yang, A discriminatively learned cnn embedding for person reidentification, *ACM Trans Multimedia Comput., Commun., Appl.,* **14** (2017), 1–20.

22. E. Ahmed, M. Jones, T. K. Marks, An improved deep learning architecture for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition,*, 2015, 3908–3916.

23. T. Huynh-The, CH. Hua, NA. Tu, DS. Kim, Learning 3D spatiotemporal gait feature by convolutional network for person identification, *Neurocomputing,* **397** (2020), 192–202.

24. D. P. Kingma, S. Mohamed, D. J. Rezende, M. Welling, Semi-supervised learning with deep generative models, J. Serrin, In: *The Advances in neural information processing systems*, 2014, 3581–3589.

25. A. Rasmus, M. Berglund, M. Honkala, H. Valpola, T. Raiko, Semi-supervised learning with ladder networks, J. Serrin, In: *The Advances in neural information processing systems*, 2015, 3546–3554.

26. H. Ma, W. Liu, A progressive search paradigm for the internet of things, *IEEE MultiMedia,* **25** (2017), 76–86.

27. X. Dong, L. Zheng, F. Ma, Y. Yang, D. Meng, Few-example object detection with model communication, *IEEE Trans. Pattern Anal. Mach. Intell.,* **41** (2018), 1641–1654.

28. L. Qi, L. Wang, J. Huo, Y. Shi, Y. Gao, Progressive Cross-camera Soft-label Learning for Semi-supervised Person Re-identification, *IEEE Trans. Circuits Syst. Video Technol.,* **30** (2020), 2815–2829.

29. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, *ArXiv*, 1609.02907, 2016. Available from: `http://arxiv.org/abs/1609.02907`.

30. H. Yu, A. Wu, W. Zheng, Cross-view asymmetric metric learning for unsupervised person re-identification, J. Serrin, In: *The IEEE international conference on computer vision*, 2017, 994–1002.

31. Z. Liu, D. Wang, H. Lu, Stepwise metric promotion for unsupervised video person re-identification, J. Serrin, In: *The IEEE international conference on computer vision*, 2017, 2429–2438.

32. M. Ye, A. J. Ma, L. Zheng, J. Li, P. C. Yuen, Dynamic label graph matching for unsupervised video re-identification, J. Serrin, In: *The IEEE international conference on computer vision*, 2017, 5142–5150.

33. C. Liu, A. Y. Li, S. Chien, J. Li, Y. Wang, Semantics-Guided Clustering with Deep Progressive Learning for Semi-Supervised Person Re-identification, *ArXiv*, 2010.01148, 2020. Available from: https://arxiv.org/abs/2010.01148.

34. P. Peng, X. Tao, Y. Wang, M. Pontil, Y. Tian, Unsupervised cross-dataset transfer learning for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2016, 1306–1315.

35. H. Fan, L. Zheng, C. Yan, Y. Yang, Unsupervised person re-identification: Clustering and fine-tuning, *ACM Trans. Multimedia Comput., Commun. Appl.,* **14** (2018), 1–18.

36. J. Wang, X. Zhu, S. Gong, W. Li, Transferable joint attribute-identity deep learning for unsupervised person re-identification, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 2275–2284.

37. W. Deng, L. Zheng, Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2018, 994–1003.

38. Z. Zhong, L. Zheng, S. Li, Y. Yang, Generalizing a person retrieval model hetero-and homogeneously, J. Serrin, In: *The European Conference on Computer Vision*, 2018, 172–188.

39. S. Xiang, Y. Fu, G. You, T. Liu, Unsupervised domain adaptation through synthesis for person re-identification, J. Serrin, In: *The IEEE International Conference on Multimedia and Expo*, 2020, 1–6.

40. Y. Ge, D. Chen, H. Li, Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification, *ArXiv*, 2001.01526, 2020. Available from: https://arxiv.org/abs/2001.01526.

41. L. A. Gatys, A. S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2016, 2414–2423.

42. C. Ledig, L. Theis, Photo-realistic single image super-resolution using a generative adversarial network, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2017, 4681–4690.

43. W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: Deep filter pairing neural network for person re-identification, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2014, 152–159.

44. L. Ma, Q. Sun, S. Georgoulis, L. Van Gool, B. Schiele, M. Fritz, Disentangled person image generation, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 99–108.

45. S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, Generative Adversarial Text to Image Synthesis, J. Serrin, In: *The International Conference on Machine Learning*, 2016, 1060–1069.

46. K. P. Dirgantoro, J. M. Lee, D. S. Kim, Generative adversarial networks based on edge computing with blockchain architecture for security system, J. Serrin, In: *The International Conference on Artificial Intelligence in Information and Communication*, 2020, 039–042.

47. Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, J. Serrin, In: *The IEEE International Conference on Computer Vision*, 2017, 3754–3762.

48. P. Isola, J. Zhu, T. Zhou, A. Efros, Image-to-image translation with conditional adversarial networks, J. Serrin, In: *The IEEE conference on computer vision and pattern recognition*, 2017, 1125–1134.

49. J. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, J. Serrin, In: *The IEEE international conference on computer vision*, 2017, 2223–2232.

50. L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer gan to bridge domain gap for person re-identification, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 79–88.

51. X. Zhang, X. Jing, X. Zhu, F. Ma, Semi-supervised person re-identification by similarity-embedded cycle GANs, *Neural Comput. Appl.,* **32** (2020), 14143–14152.

52. C. Liu, X. Chang, Y. Shen, Unity style transfer for person re-identification, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2020, 6887–6896.

53. Y. Chong, C. Peng, J. Zhang, S. Pan, Style transfer for unsupervised domain-adaptive person re-identification, *Neural Comput. Appl.,* **422** (2021), 314–321.

54. E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, J. Serrin, In: *The European Conference on Computer Vision*, 2016, 17–35.

55. G. Lisanti, I. Masi, A. D. Bagdanov, B. A. Del, Person re-identification by iterative re-weighted sparse ranking, *IEEE transactions on pattern analysis and machine intelligence,* **37** (2014), 1629–1642.

56. Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, J. Serrin, In: *The IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 1318–1327.

57. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, J. Serrin, In: *The Advances in neural information processing systems*, 2012, 1097–1105.

AIMS Press