

Article

A GAN-Based Self-Training Framework for Unsupervised Domain Adaptive Person Re-Identification

Yuanyuan Li ¹, Sixin Chen ¹, Guanqiu Qi ^{2,*}, Zhiqin Zhu ³, Matthew Haner ⁴ and Ruihua Cai ⁵

¹ College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; liyy@cqupt.edu.cn (Y.L.); chensixin9868@gmail.com (S.C.)

² Computer Information Systems Department, State University of New York at Buffalo State, Buffalo, NY 14222, USA

³ College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; zhuzq@cqupt.edu.cn

⁴ Department of Mathematics & Computer and Information Science, Mansfield University of Pennsylvania, Mansfield, PA 16933, USA; mhaner@mansfield.edu

⁵ Computer Engineering Department, San Jose State University, San Jose, CA 95192, USA; ruihua.cai@sjsu.edu

* Correspondence: qig@buffalostate.edu

Abstract: As a crucial task in surveillance and security, person re-identification (re-ID) aims to identify the targeted pedestrians across multiple images captured by non-overlapping cameras. However, existing person re-ID solutions have two main challenges: the lack of pedestrian identification labels in the captured images, and domain shift issue between different domains. A generative adversarial networks (GAN)-based self-training framework with progressive augmentation (SPA) is proposed to obtain the robust features of the unlabeled data from the target domain, according to the preknowledge of the labeled data from the source domain. Specifically, the proposed framework consists of two stages: the style transfer stage (STrans), and self-training stage (STrain). First, the targeted data is complemented by a camera style transfer algorithm in the STrans stage, in which CycleGAN and Siamese Network are integrated to preserve the unsupervised self-similarity (the similarity of the same image between before and after transformation) and domain dissimilarity (the dissimilarity between a transferred source image and the targeted image). Second, clustering and classification are alternately applied to enhance the model performance progressively in the STrain stage, in which both global and local features of the target-domain images are obtained. Compared with the state-of-the-art methods, the proposed method achieves the competitive accuracy on two existing datasets.

Keywords: person re-ID; domain shift; style transfer; self-training



Citation: Li, Y.; Chen, S.; Qi, G.; Zhu, Z.; Haner, M.; Cai, R. A GAN-Based Self-Training Framework for Unsupervised Domain Adaptive Person Re-identification. *J. Imaging* **2021**, *7*, 62. <https://doi.org/10.3390/jimaging7040062>

Academic Editor: Cosimo Distanto

Received: 22 February 2021

Accepted: 22 March 2021

Published: 25 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Person re-identification (re-ID) as a crucial task in surveillance and security strives to retrieve the same people across multiple images captured by non-overlapping cameras or across multi-scene images captured by the same camera. Despite the great success in person re-ID, some limitations still exist in practical applications, such as the acquisition of high-quality feature representation, the domain shift between training and testing data, and the difficulty of model migration from source domain to target domain.

Although existing person re-ID methods achieve high recognition rates on different types of single dataset, the great disparity exists between these person re-ID methods and practical applications, which is usually caused by the difference between the training and testing datasets [1–9]. As shown in Figure 1, different camera parameters, shooting conditions, and other factors cause the differences in exposure, image size, clarity, and other aspects. Therefore, if a model is trained on a single dataset according to the manually labeled data, the trained model often has poor recognition performance on real-world datasets.

The domain adaptive person re-ID was proposed to alleviate the domain shift issue between the labeled source domain and target domain. Some existing domain adaptive solutions achieve good performance in person re-ID. Existing domain adaptive person re-ID solutions can be roughly categorized into cross domain person re-ID [10] and shared domain learning person re-ID [11,12]. Specifically, cross domain solutions usually first conduct supervised learning on the labeled source domain, and then apply the learned model to the unlabeled target domain by migration learning. However, cross domain solutions ignore the global and local feature distribution of target-domain data, which is crucial for high-quality prediction. To compensate this deficiency, shared domain learning person re-ID solutions attempt to transfer the features in both source and target domains to a shared feature domain. Actually, due to the lack of supervisory signals during the transformation, it is difficult to guarantee the quality of visual feature information without enough constraints.

This paper proposes a GAN-based self-training framework with progressive augmentation (SPA) to solve the aforementioned two main challenges: the lack of labeled data and domain shift. This proposal expands the target-domain dataset by analyzing the consistency of each person identity in the labeled data and gains the pedestrian scale features in different learning strategies. Specifically, clustering and classification are applied to obtain the global and local features of each person respectively. The proposed SPA consists of style transfer stage and self-training stage, which act on the expanding and training processes respectively.

Style transfer stage (STrans): As a widely used style transfer solution in image processing, the proposed solution uses CycleGAN to achieve the transformation of the labeled image from source domain to target domain. Therefore, the generated training samples retain the style of target domain, such as resolution and light conditions. Specifically, Siamese network is applied to preserve the identities of pedestrians in the transformed images by using adversarial loss and contrastive loss. Besides, circle loss is used to mitigate the inflexibility and sensitivity of the proposed solution to image quality.

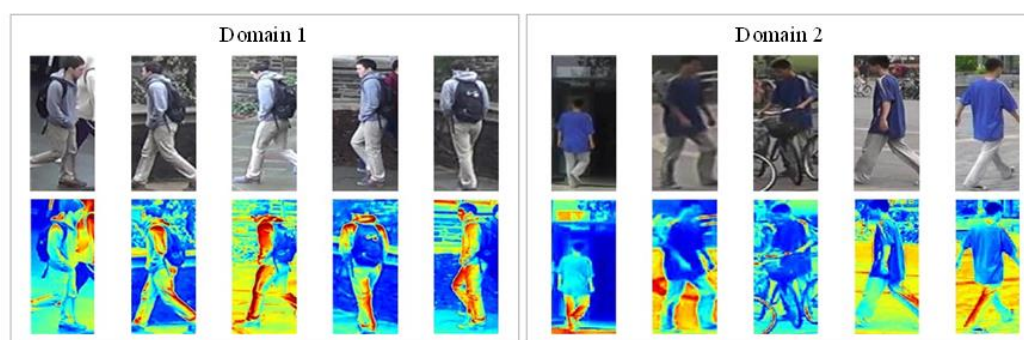


Figure 1. The sample images of person re-ID. The corresponding activation maps (AM) of the same pedestrians from two different domains as shown in the left- and right-hand sides, respectively.

Self-training stage (STrain): Clustering and classification are integrated to learn the robust features of the unlabeled target domain. Therefore, the learned global and local features are semantically complementary. As the progressive augmentation learning, both global and local features of the target-domain data are gradually enhanced by alternate clustering and classification. Since clustering and classification can promote and supervise each other, the self-training process can be completed without external intervention.

The proposed solution applies a two-stage (STrans and STrain) method to data expanding and training. Source images are first transformed without distorting semantic contents, and then credible pseudo labels are generated. Therefore, the proposed solution can achieve good prediction performance. According to the comparative results, the proposed solution outperforms other state-of-the-art unsupervised domain adaptive person

re-ID solutions on two benchmark datasets Market-1501 [13] and Duke-MTMC [14]. This paper has two main contributions as follows.

- A two-stage (STrans and STrain) framework is proposed for unsupervised domain adaptive person re-ID, which can achieve good performance on both image style transformation and self-training.
- A progressive augmentation learning strategy integrates clustering and classification to obtain both global and local features of the target-domain data, and generates credible pseudo labels without any interventions.

The rest of this paper is organized as follows. Section 2 introduces the related work; Section 3 presents the proposed image dehazing framework in detail; Section 4 discusses and compares the comparative experimental results; and Section 5 concludes this paper.

2. Related Work

As a critical task in intelligent monitoring, person re-ID that was first proposed by Gandhi in 2006 [15] has attracted considerable attention. Gray et al. [16] published a standard dataset called person re-ID VIPeR to test the performance of person re-ID solutions. Subsequently, the related person re-ID research boomed after the solutions of Zheng [17] and Farenzena [18] were published. Following the development of deep learning, person re-ID has achieved a significant breakthrough in both theories and applications. Existing solutions can achieve high recognition performance. Some recently published supervised person re-ID solutions have achieved more than 90% recognition rate on the relevant testing datasets, which greatly promote the development of the related applications.

Supervised person re-ID: Supervised person re-ID methods conduct supervised training and testing on the same datasets [1–9,13]. As shown in Figure 2, the identity labels of pedestrian images are required. The features of the whole dataset are extracted by training the feature extraction network with the guidance of labels, which can be used to calculate the similarity between different images. According to the obtained similarity, the pedestrian images are sorted. A top ranked image contains highly similar features. Zheng et al. [19] explored how to use the generated data in training. Pedestrians are encoded as appearance and structure codes. Therefore, both self-identity and cross-identity people are generated, which make the dataset expansion become realistic. Considering both posture changes and unconstrained detection errors, a new joint learning method proposed by Li [20] integrates multi-scale attention selection and feature representation to maximize the relevant supplementary information of pedestrians.

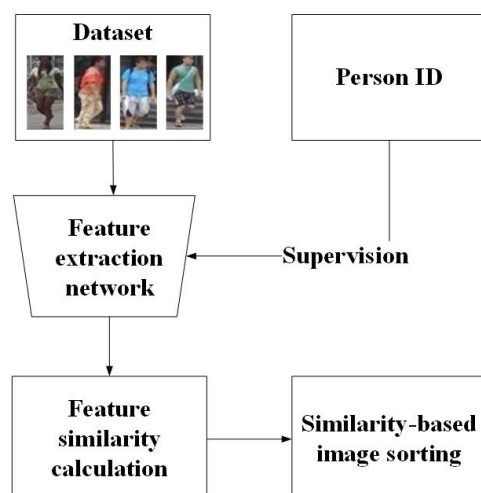


Figure 2. The general process of supervised person re-ID.

Although existing supervised person re-ID methods can achieve good performance in the source domain, the lack of labeled samples and domain shift as two main issues

still exist. Due to the difference of feature distribution between source and target domains, their recognition performance is often unsatisfactory in the target domain. Therefore, unsupervised domain adaptive (UDA) learning was proposed and applied to person re-ID to address the domain shift issue, which can be roughly categorized into cross domain learning and shared domain learning.

Cross domain person re-ID: Cross domain models can improve the object recognition accuracy, which are usually based on the supervised learning in the labeled source domain and applied to the unlabeled target domain by migration learning [21–25]. Peng et al. [24] proposed an unsupervised multi-task dictionary learning model, which represented the transferred visual features in unchanged visual angles from the source domain to the target domain. With the emergence and improvement of autoencoder, Potapov et al. [25] decomposed the interference variables of pedestrian images by potential coding, and a triple loss was used in the person feature extraction network. In addition, McLaughlin et al. [26] proposed a new data augmentation scheme based on the change of image background to alleviate the difference of data distribution caused by domain shift, which improved the cross-domain recognition ability.

Shared domain person re-ID: Shared domain-based person re-ID methods mainly focus on migrating the images in both source and target domains to a shared feature space [27–30]. In the shared domain, the consistency of visual feature information is preserved to solve the domain shift issue. To alleviate the dependence of existing methods on the labeled data, Li et al. [29] constructed a depth structure to project the features of different domains into the shared feature space by considering the labeled auxiliary dataset and the dataset of interest (without any label). In the process of shared domain person re-ID, the features from different domains are migrated to the shared feature space, and the similarity measurement of different images is realized in the shared feature space [30].

GAN-based person re-ID: The acquisition and learning of valid datasets are two main steps of recognition. GAN [31] adopts the adversarial learning. Generator and discriminator can interact with each other in the process of adversarial learning, which are conducive to improving the recognition performance of person re-ID. Therefore, GAN-based person re-ID methods are booming. A similarity preserving generative adversarial network (SPGAN) proposed by Deng et al. [10] maintains the self-similarity and inter-domain differences to eliminate the domain shift by transforming the labeled samples from the source domain (called cycle-consistent generative adversarial networks (CycleGAN) [32]) to target domain. Inspired by CycleGAN, the Camstyle network proposed Zhong et al. [12] achieves the data augmentation by transferring the camera style of each image to different ones. Wei et al. [33] introduced the semantic segmentation of images to person re-ID and proposed the person transfer generative adversarial networks (PTGAN) to alleviate the domain shift issue between different domains. Figure 3 shows the transformed images obtained by different GAN-based person re-ID methods.

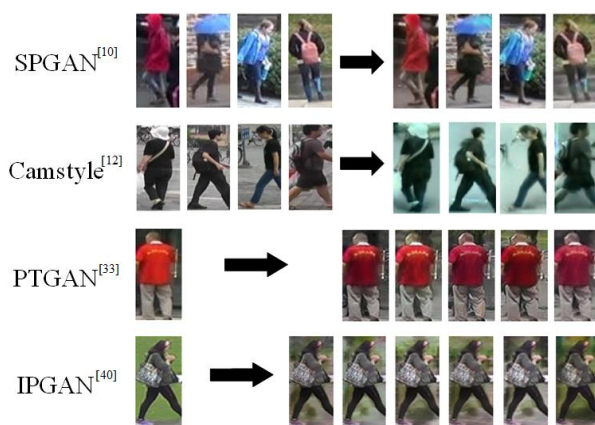


Figure 3. Transformed images obtained by different GAN-based person re-ID solutions.

The robustness of UDA person re-ID methods is determined by the differentiated information from different domains. Due to the varying degrees of domain shift, the overall recognition performance of cross- and shared-domain person re-ID methods is not stable. Therefore, this paper explores how the labels, feature representation, and metric learning affect the performance of person re-ID and proposes an effective GAN-based self-training framework.

3. The Proposed Solution

As shown in Figure 4, the proposed SPA consists of STrans and STrain.

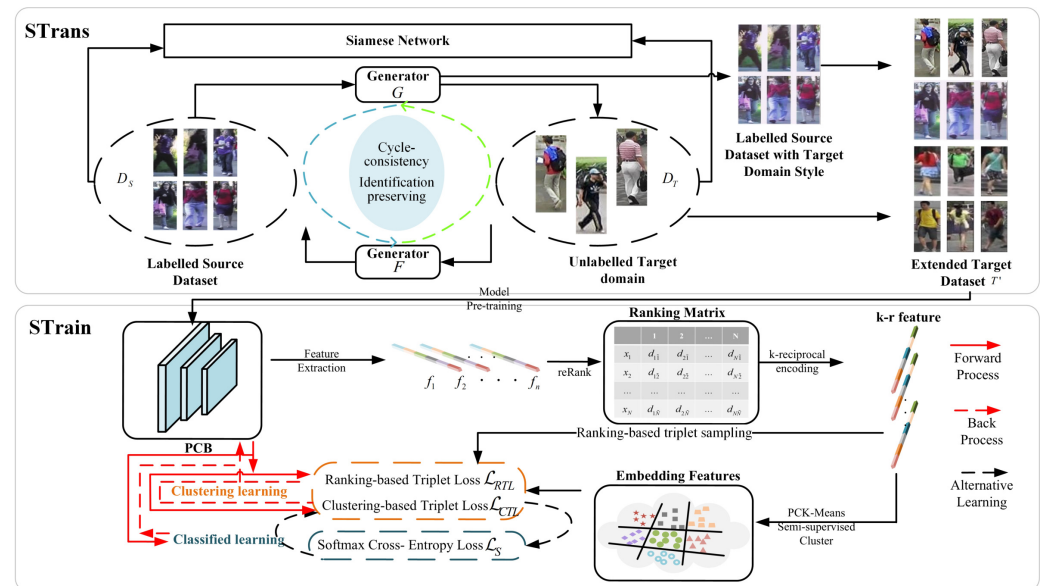


Figure 4. The framework of the proposed SPA.

In STrans, both CycleGAN and Siamese Network are integrated to ensure the self-similarity (the same identity in an image is remained) and inter-domain difference (the original style is remained across different domains) before and after transformation. When any part of STrans is changed, the corresponding parameters of CycleGAN and Siamese Network are updated accordingly.

In STrans, the global and local structures of target-domain data are obtained in the two-stage self-training process of the progressive augmentation framework. In particular, the global and local features of each person are obtained by clustering and classification, respectively. Two stages process alternately in the self-training process until reaching the goal. Similarly, the corresponding parameters are updated according to any change of STrain.

3.1. Style Transfer Stage

Similar to SPGAN [10], CycleGAN [32] is used to realize the basic style transformation, and Siamese Network [34] is applied to maintain the consistency of pedestrian identity. Figure 5 illustrates the structure of STrans (Style transfer stage). As shown in the upper part of Figure 5, Euclidean distance is used to measure the similarity between two different images. The images with high similarity are clustered.

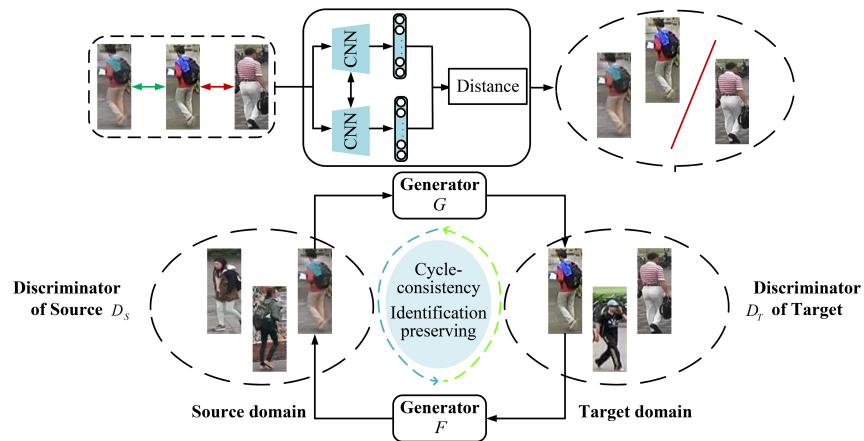


Figure 5. The structure of Strans (Style transfer stage).

As shown in the lower part of Figure 5, CycleGAN learns generators G and F by capturing the fine information of the labeled source-domain dataset $S=\{x_i\}_{i=1}^M$ and unlabeled target-domain dataset $T=\{y_j\}_{j=1}^N$, respectively, which are used in the image style transformation from source domain to target domain. Adversarial loss and cycle-consistent loss are used to ensure the antagonism and consistency of image contents between G and D . CycleGAN is formalized as follows.

$$\mathcal{L}_{cyc}(G, F, D_T, D_S) = \mathcal{L}_{adv}(G, D_T) + \mathcal{L}_{adv}(F, D_S) + \alpha \mathcal{L}_{rec}(G, F) \quad (1)$$

where $D_T(D_S)$ represents the discriminator corresponding to the generator $G(F)$, $\mathcal{L}_{adv}$ and $\mathcal{L}_{rec}$ denote adversarial loss and cycle-consistency loss respectively, and α controls the relative importance of the cycle-consistent loss.

In addition to adversarial losses and cycle-consistency loss, style retain function is designed to ensure that the color composition between the input and output is preserved and the generator is prevented from outputting unreal results. In particular, when the generator transfers an image, it needs to preserve the identity information of source images. Therefore, a unit matrix is formed to ensure the identity mapping as follows.

$$\mathcal{L}_{ide}(G, F) = E_{x^s \sim p_{data}(S)} [\|F(x^s) - x^s\|_1] + E_{x^t \sim p_{data}(T)} [\|G(x^t) - x^t\|_1] \quad (2)$$

It is necessary to ensure the identity consistency and domain dissimilarity of pedestrian after transformation. During the training process, Siamese network is optimized by minimizing the sum of contrastive loss and circle loss [35] on the designed input pair.

$$\mathcal{L}_{con}(w, x_1, x_2) = \frac{1}{2N} \sum_{n=1}^N (1-w) [\max(0, m-d)]^2 + w \cdot d^2 \quad (3)$$

where (x_1, x_2) is an input matching pair, d is the Euclidean distance between the pair, $w = 0$ ($w = 1$) denotes the input pair is negative (positive), and the parameter m controls the margin of decision boundary.

$$\mathcal{L}_{circle} = \log[1 + \sum_{i=1}^L \exp(\eta \alpha_p^i (d_p^i - \Delta p))] \times \sum_{j=1}^K \exp(-\eta \alpha_n^j (d_n^j - \Delta n)) LK \quad (4)$$

where L and K represent the number of Euclidean distances corresponding to positive and negative input pairs respectively, $L + K = N$, and d_p^i and d_n^j denote the Euclidean distance between each matching pair. Due to the asymmetry of positive and negative pairs, Δp and Δn are the margin corresponding to them, respectively. η is used as an extended factor to realize the gradient control. To realize the self-paced weighting, α_p^i and α_n^j can be defined as follows.

$$\begin{aligned}\alpha_p^i &= [O_p - d_p^i]_+ \\ \alpha_n^j &= [d_n^j - O_n]_+\end{aligned}\quad (5)$$

In Equations (3) and (4), loss functions use the binary labels of input image pairs. As shown in Figure 6, positive input pair $(x_s^{id,i}, G(x^{S,id,j}))$ and negative pair $(G(x^S), x_t)$ are designed to ensure the identity consistency and domain dissimilarity of pedestrians. Specifically, the i -th sample in source domain can be directly used to form a positive pair with any transformed image which has the same identity but not necessarily converted from the same sample. As the a priori knowledge that pedestrian images from two datasets do not cross and contain the same person, the pedestrians in the transformed images must be different from anyone from target domain. A negative pair is constructed as $(G(x^S), x_t)$ or $(F(x^T), x_s)$.

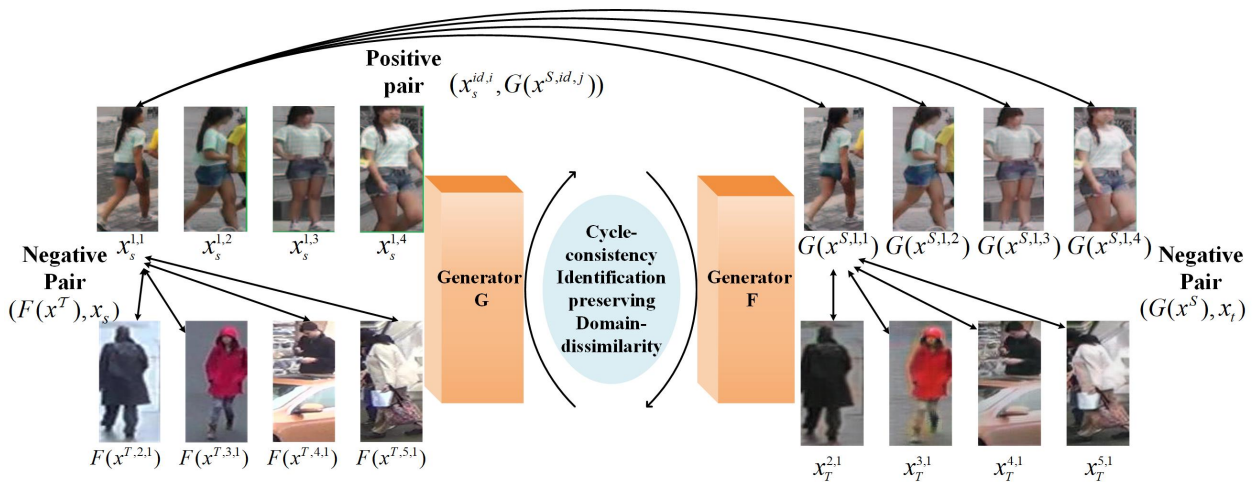


Figure 6. The image pair construction.

The overall objective function of style transfer stage can be formalized as follows.

$$\begin{aligned}\mathcal{L}(G, F, D_T, D_S, M) &= \mathcal{L}_{cyc}(G, F, D_T, D_S) + \lambda_1 \mathcal{L}_{ide}(G, F) \\ &+ \lambda_2 \mathcal{L}_{con} + \lambda_3 \mathcal{L}_{circle}\end{aligned}\quad (6)$$

The extended target domain dataset T' is obtained for further learning.

3.2. Self-Training Stage

Due to the dramatic appearance changes and identity dissimilarity between different domains, it is expensive and impractical to label data in the unsupervised and domain adaptation settings. To alleviate the above limitations, a two-step self-training process is proposed, which takes advantage of classification and clustering.

3.2.1. Semi-Supervised Learning

Since the extended target domain dataset $T' : \{t_1, t_2, \dots, t_{N'}\}$ contains both true unlabeled samples (original target-domain images) and untrue labeled samples (converted

from labeled source-domain images) after style transformation, semi-supervised learning is used to extract the embedding features from the pre-training part-based convolutional baseline (PCB) [36]. Subsequently, the pairwise constrained K-Means [37,38] (PCK-Means) is applied to semi-supervise sample clustering to obtain the reliable pseudo labels of untrue labeled samples. The semi-supervised learning structure is shown in Figure 7.

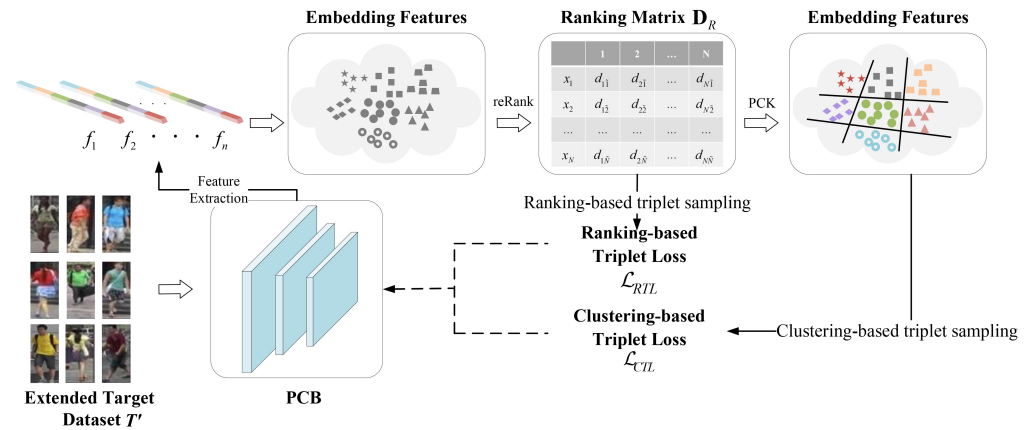


Figure 7. The semi-supervised learning structure. In semi-supervised learning, PCB first extracts the related features. Then k-reciprocal encoding is used to describe the fine difference between any two images, which yields a ranking score matrix $\mathbf{D}_R$ for the next clustering operation. With the help of partially labeled images in the extended dataset T' , semi-supervised clustering PCK-Means is used to mark pseudo-labels for the related data. The whole stage is trained by RTL and CTL.

In practice, features $F : \{f(t_1), f(t_2), \dots, f(t_{N'})\}$ are extracted from the current PCB to construct the feature embedding space, and k-reciprocal encoding [39] is adopted to describe the fine difference between any two images. By calculating the Euclidean distance $d_m(t_i, t_j)$ between the features of each pair, the neighbour set N corresponding to the k-closest distances of the probe is calculated. N that contains both positive and negative samples is defined as: $N(\text{probe}, k) = \{t_1^0, t_2^0, \dots, t_k^0\}$, where t_1^0, t_2^0, t_k^0 represent the 1st, 2nd, and k-th closest samples to the probe, respectively. At the same time, each t_i^0 of N also has its own neighbour set N' . If a probe is included, probe and t_i^0 are adjacent to each other. Otherwise, they are not adjacent to each other. Thus, the k-reciprocal set R of the probe can be obtained, and all the elements in R are close to the probe. A ranking score matrix $\mathbf{D}_R$ is obtained to describe the distance as follows.

$$\begin{aligned} \mathbf{D}_R &= [\mathbf{D}_R(t_1) \mathbf{D}_R(t_2) \dots \mathbf{D}_R(t_N)]^T, \\ \mathbf{D}_R(t_i) &= [d_J(t_i, \tilde{t}_1) d_J(t_i, \tilde{t}_2) \dots d_J(t_i, \tilde{t}_N)] \\ &\quad \forall i \in \{1, 2, \dots, N\} \end{aligned} \quad (7)$$

where $\mathbf{D}_R(t_i)$ represents the ascending order of the distance between the probe t_i and other samples in the gallery.

Given a large gallery, it is difficult to distinguish the samples with high similarity. PCK-Means is applied to mark pseudo labels for the extended target-domain dataset T' , and the associated relationship in a mini-batch is explored to improve the operation speed of the proposed model in practical applications. In the end, P clusters and K instances are sampled randomly, and the cluster-based triplet Loss (CTL) is formulated as Equation (9).

$$\begin{aligned}
\mathcal{L}_{CTL} &= \sum_{a=1}^{PK} [m + \|f(t_a) - f(t_p)\|_2 - \|f(t_a) - f(t_n)\|_2]_+ \\
&= \sum_{i=1}^P \sum_{a=1}^K [m + \max_{p=1 \dots K} \|f(t_a) - f(t_p)\|_2 \\
&\quad - \min_{\substack{n=1 \dots K \\ j=1 \dots P \\ i \neq j}} \|f(t_a) - f(t_n)\|_2]_+
\end{aligned} \tag{8}$$

where (t_a, t_p, t_n) is a triplet, m is the margin between positive and negative pairs as same as Equation (3), and for the anchor t_a , i represents a certain class in P clusters, and j represents an instance under this class. Subsequently, benefiting from the PCK-Means, some samples could be added into T_U , which is the new image training set with pseudo labels to optimize PCB. However, it is clear that the effectiveness of CTL is highly subjected to the correctness of the clustering result. Hence, ranking-based triplet loss (RTL) is proposed as follows, which does not depend on any pseudo labels, but is only related to the sorting matrix $\mathbf{D}_R$.

$$\begin{aligned}
\mathcal{L}_{RTL} &= \sum_{a=1}^{PK} [\frac{P_p - P_n}{\eta} m + \|f(t_a) - f(t_p)\|_2 \\
&\quad - \|f(t_a) - f(t_n)\|_2]_+
\end{aligned} \tag{9}$$

where the triplet and parameter m are constructed in the same way as CTL, and for each anchor t_a , P_p and P_n represent the number of positive and negative pairs respectively. The combination of CTL and RTL can optimize the feature extraction network and capture the local information of data distribution effectively. Therefore, the final triplet loss function in the semi-supervised learning can be defined as follows.

$$\mathcal{L}_C = \mathcal{L}_{RTL} + \lambda \mathcal{L}_{CTL} \tag{10}$$

where the parameter λ controls the relative importance of feature learning constraints.

3.2.2. Classification Learning

Conventionally, according to the difference of objective loss function, person re-ID consists of representation learning and metric learning corresponding to classification and clustering respectively. Most existing methods use one way to train the network and the two learning methods are applied to further improve the network performance. Theoretically, due to PCK-Means clustering, the network focuses on the local structure of data distribution and may ignore the global information in semi-supervised learning. Therefore, the model is easy to fall into a sub-optimal local minimum.

As an optimization way, clustering and classification are used alternately. In this way, a fully connected layer is added to the end of the model as a classification layer, which is initialized by the current T_U . The objective function can be calculated by Softmax cross-entropy loss as follows.

$$\mathcal{L}_s = \sum_{i=1}^{PK} \log \frac{e^{\hat{y}_i^T x_i}}{\sum_{c=1}^C e^{W_c^T x_i}} \tag{11}$$

where $\hat{y}_i$ is the pseudo label of x_i , C denotes the cluster number of the updated training set T_U after PCK-Means clustering, and W is the initialized classifier weight.

4. Comparative Experiments

4.1. Datasets and Objective Evaluation Indicators

Two large-scale person re-ID datasets as shown in Table 1, Market-1501 and Duke, are used to test the performance of the proposed model.

Table 1. The details of person Re-ID datasets (Market-1501 and DukeMTMC-ReID).

Dataset	Publication Time	Number of IDs	Number of Cameras	Number of Images	Labeling Method	Image Size
Market-1501 [13]	2015	1501	6	32,688	Manual + DPM	128 × 64
DukeMTMC-ReID [14]	2017	1812	8	36,411	Manual	Random

Market-1501 [13] is a dataset collected and published by Tsinghua University in 2015. 32,688 images were captured by six cameras including a low-definition camera, which involved 1501 pedestrians. Each pedestrian appeared in at least two camera views. Market-1501 is divided into training and testing sets, which have 12,936 images with 751 pedestrians and 19,732 images (including 3,368 manually drawn images) with 750 pedestrians, respectively.

DukeMTMC-ReID [14] is a large-scale multi-pedestrian dataset collected by Duke University, which contains a large number of labels. Eight high-definition cameras collected 85-min video, involving 36,411 images and 1812 pedestrians. 1404 pedestrians appeared in at least two camera views. Zheng et al. [14] divided the dataset into the training set containing 1622 images with 702 people, testing query set containing 2228 images, and testing gallery set with 17,661 images. For convenience, Duke is short for DukeMTMC-ReID in the following paragraphs.

Cumulative matching feature (CMC) is the most widely used in person re-ID, which can be regarded as the accuracy rate in the related papers. For each pedestrian in the query set, it calculates the distance to n gallery samples in turn, and then sorts the obtained distances to check whether the same identity samples are located in the top- k , and finally the CMC curve is obtained by statistics. Specifically, it is a floating-point number in an interval. In convenience, it usually takes the form of percentage and only compares three-digits accuracy rates.

As an index widely used in reflecting the recall rate of the model, mean average precision (mAP) is the mean value of average accuracy (AP) of all query samples [40–43]. For the query sample probe, the calculation of its AP is mainly determined by the accuracy of recall rate. Specifically, AP of a query sample can be calculated as the area of precision-recall (PR) curve and horizontal axis.

4.2. Implementation

CycleGAN and Siamese network are adopted in the style transfer stage. Adam optimizer [44] is also used. The batch size is 1 and the initial learning rate is set to 0.0002. The training stops after the network has passes 6 epochs. Siamese network contains 3 convolutional layers (Con.), 3 maximum pooling layers (Max pooling), and 2 fully connected layers (FC). The specific network structure is shown in Table 2.

Table 2. The design details of the STrans network structure.

Layer	Property	Parameter
1	Conv.1	4×4 , stride = 2, fmap = 64
2	Max pooling 1	2×2 , stride = 2,
3	Conv.2	4×4 , stride = 2, fmap = 128
4	Max pooling 2	2×2 , stride = 2,
5	Conv.3	4×4 , stride = 2, fmap = 256
6	Max pooling 3	2×2 , stride = 2,
7	Max pooling 4	2×2 , stride = 2,
8	FC	Out: 256

Similar to the EANet, PCB [36] is used as the feature extractor in the self-training stage. The feature tensor is horizontally divided into six parts to ensure the retention of local information. Deriving from numerous experiments and previous experiences, m in

Equations (3) and (8), and α in Equation (1) are empirically set to 2 and 10, respectively. All input images are resized to $384 \times 128 \times 3$. The dimension of each embedding layer is set to 256, the batch size is set to 64, and the number of iterations is set to 4.

The two-step learning rate can improve the learning performance of the progressive self-monitoring learning framework, rather than using the same learning rate directly in both self-monitoring stage and classification stage. Therefore, the false label guidance can be avoided. Specifically, in the semi-supervised learning, the learning rate of the backbone network is initialized to 1×10^{-4} , and the learning rate of the embedded layer is 2×10^{-4} . In the classification learning, the classification layer is 1×10^{-3} , while all other layers are set to 5×10^{-5} . After three iterations, all learning rates are multiplied by 0.1. The super parameter m is set to 2 which is consistent with Equations (3) and (8).

4.3. Comparisons with the State-of-the-Art Solutions

SPA proposed in this paper is compared with the state-of-the-art style transfer learning and UDA learning solutions on Market1501 [13] and DukeMTMC-reID [14]. Tables 3 and 4 show the comparisons, in which M and D represent Market-1501 and Duke respectively. In each column, the highest result is marked in bold.

Table 3. Person re-ID performance comparison with the state-of-the-art solutions of style transfer on Market-1501 and DukeMTMC-reID.

Method	D $\rightarrow$ M				M $\rightarrow$ D			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
Camstyle [12]								
PTGAN [33]	38.6	-	66.1	-	27.4	-	50.7	-
SPGAN [10]	51.5	70.1	76.8	22.8	41.1	56.6	63	22.3
IPGAN [45]	56.4	75.6	82.5	25.6	46.8	62	67.9	25.7
MMFA [46]	56.7	75	81.8	27.4	45.3	59.6	65	23
UCDA [47]	64.3	-	-	34.5	55.4	-	-	36.7
Ours	65.4	80.6	85.7	35.2	59.2	72.8	76.7	37.6

Table 4. Person re-ID performance comparison with the state-of-the-art solutions of unsupervised on Market-1501 and DukeMTMC-reID.

Method	Market-1501				Duke			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
LOMO [48]	27.2	41.6	49.1	8	12.3	21.3	26.6	4.8
BOW [21]	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
PUL [49]	45.5	60.7	66.7	20.5	30	43.4	48.5	16.4
PCB [36]	57.7	-	-	29.01	42.73	-	-	25.7
BUC [50]	66.2	76.6	84.5	38.3	47.4	62.6	68.4	27.5
DBC [51]	69.2	83	87.8	41.3	51.5	64.6	70.1	30
MAR [52]	67.7	81.9	-	40	67.1	79.8	-	48
Ours	73.93	82.8	91.5	53.35	65.18	81.3	87.2	52.43

As shown in Table 3, transfer learning-based methods include Camstyle [12], PTGAN [33], SPGAN [10], IPGAN [45], MMFA [46], and UCDA [47]. PTGAN uses the semantic segmentation to constrain local images and retain the pedestrian information, but the direct conversion causes the loss of identity information easily. Camstyle, SPGAN, and IPGAN are all based on CycleGAN, which realize the unity of image styles between source and target domains. SPGAN and IPGAN use the identity retention to eliminate domain offsets, but they are limited by the matching pair construction methods. UCDA uses the transfer learning to minimize the invariance in target domain. STTrans obtains 65.4 and 59.3 on Market and Duke of Rank-1, respectively, which benefit from the novel effective method to construct the matching pairs and optimize the model through circle loss with the target convergence.

As shown in Table 4, unsupervised methods include LOMO [48], BOW [21], PUL [49], BUC [50], DBC [51], PCB [36], and MAR [52]. LOMO and BOW use the hand-crafted

features, which show low performance. MAR adopts the idea of multi-soft labeling. PCB is a baseline commonly used in recent research, which uses the horizontal division of high-dimensional tensors to retain the detailed information. TFSuion uses the spatio-temporal information to estimate the matching probability through Bayesian inference. However, the framework proposed in this paper is much more concisely and effective than existing methods. As shown in Table 4, mAP reaches 53.35% and 52.43% and rank-1 reaches 73.93% and 65.18% in $D \rightarrow M$ and $M \rightarrow D$, respectively.

It is useful to use the expanded labeled data to train the model in the last two rows of Table 4. Specifically, compared with the style transfer stage alone, the incremental self-monitoring learning framework in rank-1 and mAP can improve by 3.86% and 3.07% in $D \rightarrow M$, respectively.

4.4. Ablation Study

The impact of the each component of the proposed algorithm. As mentioned in introduction, the accuracy of person re-ID in UDA setting relies on the generation quality and identity recognition accuracy. Four components of the proposed GAN-based self-training network are evaluated as follows.

- GAN-based transformation network: According to the SPGAN model, the performance of the proposed GAN-based transformation network is significantly improved, which benefits from the adoption of both novel training data construction and circle loss methods.
- Progressive self-training framework: The semi-supervised clustering and classification learning are combined to learn the robust features of the unlabeled target domain effectively.
- Semi-supervised learning: k-reciprocal encoding and PCK-Means are used when a ranking score matrix is constructed and the initial images are clustered.
- Classification: It is identical to general softmax classification but needs to initialize the classification layer.

As shown in Table 5, when the network only contains STrans, the rank-1 accuracy on $M \rightarrow D$ and $D \rightarrow M$ increases by 18.43% and 12.48%, respectively. The rank-1 and map of $M \rightarrow D$ increase by 20.53% and 12.47% respectively, while the rank-1 and map of $D \rightarrow M$ increase by 12.75% and 5.50% respectively. The improvement of efficiency shows that both triple losses can be used to enhance the performance of the proposed model, but the performance of STrans is slightly lower than that of STrain. STrans and STrain are combined to jointly optimize the model at the self-monitoring stage, and they achieve good results in $M \rightarrow D$ and $D \rightarrow M$. Compared with STrans only, 1.71% and 3.40% improvements on rank-1 and map are achieved on $D \rightarrow M$. Therefore, it confirms that a powerful target-domain feature extraction model is learned by the proposed SPA.

Table 5. Ablation study on different components of the proposed method.

Method	Training	Market-1501		DUKEMTMC-reID	
		Rank-1	mAP	Rank-1	mAP
PCB	-	59.74	41.93	39.38	49.69
PCB+Style transform	Unsupervised	70.09	50.28	62.09	49.59
PCB+Self-training	Semi-supervised	71.22	51.85	63.49	48.43
PCB+SPA	Semi-supervised	73.93	53.35	65.18	52.43

The impact of the hyperparameters. The generalization properties of a loss function are governed by a small number of hyper-parameters. The hyperparameter values are determined in the process of model selection. In Equation (10), λ is used to control the weight between RTL and CTL. Values are selected from the set of 0.1, 0.2, 0.5, 1.0, and 2.0 to test the impact on the $D \rightarrow M$ task. When λ is low, RLT plays a major role, which tends to cause the overall network relying on the ranking score matrix D_R . Particularly,

when the feature representations are in poor quality, the network performance is really low. When λ is high, the pseudo-label quality is low in the early stage of clustering process, and the network emphasizes the clustering results. As shown in the top left of Figure 8, the best result is obtained, when λ set to 0.5. However, the change in the size limits the performance improvement.

Subsequently, three hyperparameters in Equation (6) of STrans are tested, and the results are shown in the top right, bottom left, and bottom right of Figure 8. λ_1 , λ_2 , and λ_3 control the degree of style retain loss, contrastive loss, and circle loss respectively, which balance the impact of the losses and change from 0.2 to 1.0. When λ_1 , λ_2 , and λ_3 are set to 0.4, 0.6, 0.4, the best performance is achieved. When λ_1 , λ_2 , and λ_3 continually increase, a obvious drop occurs.

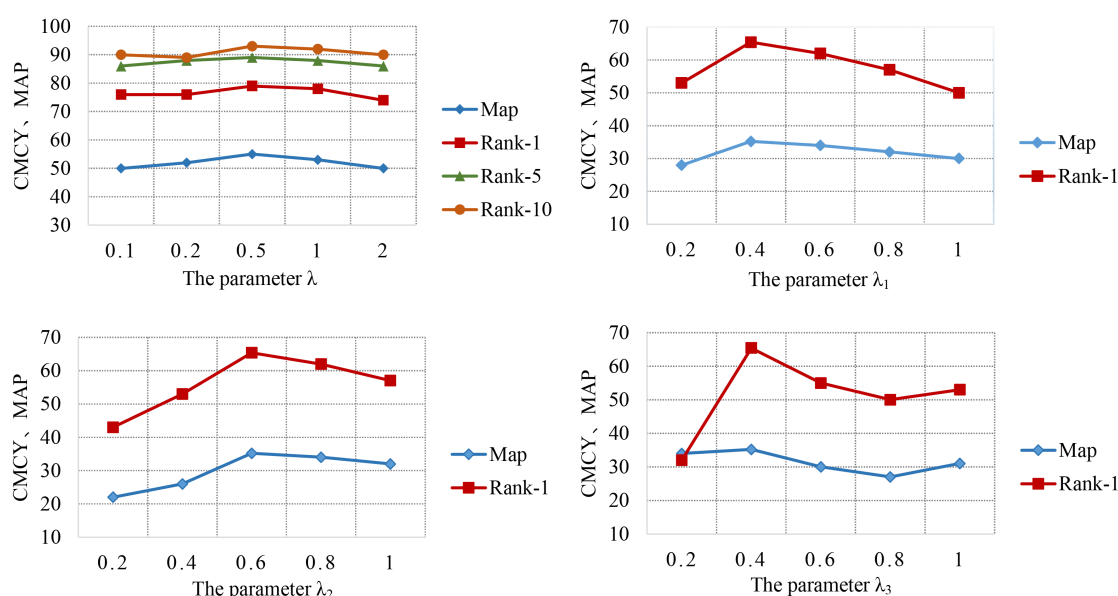


Figure 8. Analysis of parameters λ in Equation (10) and λ_1 , λ_2 , λ_3 in Equation (6).

5. Conclusions

This paper proposes a GAN-based self-training framework for UDA person Re-ID, which focuses on solving the lack of pedestrian identification labels in the captured images and domain shift issue between different domains.

In the proposed SPA, the a priori knowledge from the labeled source domain is used to obtain the robust features of target domain. In style transfer stage, CycleGAN and Siamese Network are combined to ensure the self-similarity and inter-domain difference of person identification. Besides the widely used adversarial loss and contrastive loss, which are inflexible and sensitive to the quality of pair, circle loss is used to optimize the model with a targeted convergence. The self-training stage captures the global and local structure of target-domain data in the progressive augmentation framework, which takes advantage of clustering and classification on person re-ID. The comparative experimental results confirm the proposed solution achieves better performance than the state-of-the-art unsupervised cross-domain re-ID solutions in person re-ID. In future, the proposed method will be extended to other unsupervised cross-domain applications.

Author Contributions: Conceptualization, Y.L. and G.Q.; methodology, S.C. and Z.Z.; software, S.C.; validation, G.Q., Z.Z., M.H., and R.C.; formal analysis, Y.L., S.C., G.Q., and Z.Z.; investigation, Z.Z.; resources, Y.L. and S.C.; data curation, S.C.; writing—original draft preparation, S.C. and G.Q.; writing—review and editing, S.C., G.Q., Z.Z. and M.H.; visualization, R.C.; supervision, G.Q. and M.H.; project administration, Z.Z.; funding acquisition: Y.L. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61803061, Grant 61906026, in part by the Chongqing Natural Science Foundation under Grant cstc2020jcyj-msxmX0577.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chen, D.; Zhang, S.; Ouyang, W.; Yang, J.; Tai, Y. Person search via a mask-guided two-stream cnn model. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
- Geng, M.; Wang, Y.; Xiang, T.; Tian, Y. Deep transfer learning for person re-identification. *arXiv* **2016**, arXiv:1611.05244.
- Xu, J.; Zhao, R.; Zhu, F.; Wang, H.; Ouyang, W. Attention-aware compositional network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2119–2128.
- Chang, X.; Hospedales, T.M.; Xiang, T. Multi-level factorisation net for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2109–2118.
- Chen, W.; Chen, X.; Zhang, J.; Huang, K. A multi-task deep network for person re-identification. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
- Li, W.; Zhu, X.; Gong, S. Harmonious attention network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2285–2294.
- Sun, Y.; Zheng, L.; Deng, W.; Wang, S. Svdnet for pedestrian retrieval. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3800–3808.
- Chen, Y.; Zhu, X.; Gong, S. Person re-identification by deep learning multi-scale representations. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2590–2600.
- Paisitkriangkrai, S.; Shen, C.; Van Den Hengel, A. Learning to rank in person re-identification with metric ensembles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1846–1855.
- Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; Jiao, J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 994–1003.
- Peng, P.; Tian, Y.; Xiang, T.; Wang, Y.; Pontil, M.; Huang, T. Joint semantic and latent attribute modelling for cross-class transfer learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 1625–1638. [[CrossRef](#)] [[PubMed](#)]
- Zhong, Z.; Zheng, L.; Zheng, Z.; Li, S.; Yang, Y. Camstyle: A novel data augmentation method for person re-identification. *IEEE Trans. Image Process.* **2018**, *28*, 1176–1190. [[CrossRef](#)] [[PubMed](#)]
- Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Las Condes, Chile, 11–18 December 2015; pp. 1116–1124.
- Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; Tomasi, C. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 17–35.
- Gandhi, T.; Trivedi, M.M. Panoramic appearance map (pam) for multi-camera based person re-identification. In Proceedings of the 2006 IEEE International Conference on Video and Signal Based Surveillance, Sydney, Australia, 22–24 November 2006; p. 78.
- Gray, D.; Brennan, S.; Tao, H. Evaluating appearance models for recognition, reacquisition, and tracking. In Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), Honolulu, HI, USA, 14–18 May 2007; Volume 3, pp. 1–7.
- Zheng, W.S.; Gong, S.; Xiang, T. Person re-identification by probabilistic relative distance comparison. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 22–25 June 2011; pp. 649–656.
- Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR); San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
- Zheng, Z.; Yang, X.; Yu, Z.; Zheng, L.; Yang, Y.; Kautz, J. Joint discriminative and generative learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2138–2147.
- Li, W.; Zhu, X.; Gong, S. Person re-identification by deep joint learning of multi-loss classification. *arXiv* **2017**, arXiv:1705.04724.
- Wang, J.; Zhu, X.; Gong, S.; Li, W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2275–2284.
- Zhong, Z.; Zheng, L.; Li, S.; Yang, Y. Generalizing a person retrieval model hetero-and homogeneously. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 172–188.

23. Fan, H.; Zheng, L.; Yan, C.; Yang, Y. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2018**, *14*, 1–18. [\[CrossRef\]](#)
24. Peng, P.; Xiang, T.; Wang, Y.; Pontil, M.; Gong, S.; Huang, T.; Tian, Y. Unsupervised cross-dataset transfer learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1306–1315.
25. Potapov, A.; Rodionov, S.; Latapie, H.; Fenoglio, E. Metric Embedding Autoencoders for Unsupervised Cross-Dataset Transfer Learning. In *International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 289–299.
26. McLaughlin, N.; Del Rincon, J.M.; Miller, P. Data-augmentation for reducing dataset bias in person re-identification. In Proceedings of the 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, Germany, 25–28 August 2015; pp. 1–6.
27. Lv, J.; Chen, W.; Li, Q.; Yang, C. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7948–7956.
28. Song, L.; Wang, C.; Zhang, L.; Du, B.; Zhang, Q.; Huang, C.; Wang, X. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognit.* **2020**, *102*, 107173. [\[CrossRef\]](#)
29. Li, Y.J.; Yang, F.E.; Liu, Y.C.; Yeh, Y.Y.; Du, X.; Frank Wang, Y.C. Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–178.
30. Huang, H.; Yang, W.; Chen, X.; Zhao, X.; Huang, K.; Lin, J.; Huang, G.; Du, D. EANet: Enhancing alignment for cross-domain person re-identification. *arXiv* **2018**, arXiv:1812.11369.
31. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, UK, 2014; pp. 2672–2680.
32. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
33. Wei, L.; Zhang, S.; Gao, W.; Tian, Q. Person transfer gan to bridge domain gap for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 79–88.
34. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; pp. 539–546.
35. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle loss: A unified perspective of pair similarity optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual conference, Seattle, WA, USA, 14–19 June 2020; pp. 6398–6407.
36. Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 480–496.
37. Grira, N.; Crucianu, M.; Boujemaa, N. Semi-Supervised Fuzzy Clustering with Pairwise-Constrained Competitive Agglomeration. In Proceedings of the 14th IEEE International Conference on Fuzzy Systems, Reno, NV, USA, 22–25 May 2005; pp. 867–872.
38. Qi, G.; Zhu, Z.; Erqinhu, K.; Chen, Y.; Chai, Y.; Sun, J. Fault-diagnosis for reciprocating compressors using big data and machine learning. *Simul. Model. Pract. Theory* **2018**, *80*, 104–127. [\[CrossRef\]](#)
39. Zhong, Z.; Zheng, L.; Cao, D.; Li, S. Re-ranking person re-identification with k-reciprocal encoding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1318–1327.
40. Qi, G.; Hu, G.; Wang, X.; Mazur, N.; Zhu, Z.; Haner, M. EXAM: A Framework of Learning Extreme and Moderate Embeddings for Person Re-ID. *J. Imaging* **2021**, *7*, 6. [\[CrossRef\]](#)
41. Li, H.; Chen, Y.; Tao, D.; Yu, Z.; Qi, G. Attribute-Aligned Domain-Invariant Feature Learning for Unsupervised Domain Adaptation Person Re-Identification. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 1480–1494. [\[CrossRef\]](#)
42. Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf. Sci.* **2018**, *432*, 516–529. [\[CrossRef\]](#)
43. Zhu, Z.; Wei, H.; Hu, G.; Li, Y.; Qi, G.; Mazur, N. A Novel Fast Single Image Dehazing Algorithm Based on Artificial Multiexposure Image Fusion. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–23. [\[CrossRef\]](#)
44. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
45. Liu, J.; Li, W.; Pei, H.; Wang, Y.; Qu, F.; Qu, Y.; Chen, Y. Identity preserving generative adversarial network for cross-domain person re-identification. *IEEE Access* **2019**, *7*, 114021–114032. [\[CrossRef\]](#)
46. Lin, S.; Li, H.; Li, C.T.; Kot, A.C. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. *arXiv* **2018**, arXiv:1807.01440.
47. Qi, L.; Wang, L.; Huo, J.; Zhou, L.; Shi, Y.; Gao, Y. A novel unsupervised camera-aware domain adaptation framework for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8080–8089.

48. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2197–2206.
49. Li, J.; Zhang, S.; Tian, Q.; Wang, M.; Gao, W. Pose-guided representation learning for person re-identification. In *IEEE Transactions on PATTERN analysis and Machine Intelligence*; IEEE: New York, NY, USA, 2019.
50. Lin, Y.; Dong, X.; Zheng, L.; Yan, Y.; Yang, Y. A bottom-up clustering approach to unsupervised person re-identification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 8738–8745.
51. Ding, G.; Khan, S.H.; Tang, Z. Dispersion based Clustering for Unsupervised Person Re-identification. *arXiv* **2019**, arXiv:1906.01308.
52. Yu, H.X.; Zheng, W.S.; Wu, A.; Guo, X.; Gong, S.; Lai, J.H. Unsupervised person re-identification by soft multilabel learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 2148–2157.