Machine Vision and Applications

Semi-supervised learning for person re-identification based on style-transfer-generated data by CycleGANs

--Manuscript Draft--

Manuscript Number:	MVAP-D-19-00424
Full Title:	Semi-supervised learning for person re-identification based on style-transfer-generated data by CycleGANs
Article Type:	Full Paper
Corresponding Author:	Yunzhou Zhang Northeastern University CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Northeastern University
Corresponding Author's Secondary Institution:	
First Author:	Shangdong Zhu
First Author Secondary Information:	
Order of Authors:	Shangdong Zhu
	Yunzhou Zhang
	Sonya Coleman
	Song Wang
	Ruilong Li
	Shuangwei Liu
Order of Authors Secondary Information:	
Funding Information:	National Natural Science Foundation of China (61973066) — Prof. Yunzhou Zhang
	National Natural Science Foundation of China (61471110) — Prof. Yunzhou Zhang
	National Natural Science Foundation of China (61733003) — Prof. Yunzhou Zhang
	Fundamental Research Funds for the Central Universities (N172608005) — Prof. Yunzhou Zhang
Abstract:	Person re-identification (re-ID) is an exceedingly significant branch in the field of computer vision, especially for video surveillance. It is still a challenge to obtain more labeled training data and use them reasonably for more precise matching, though the person re-ID performance has been improved significantly. In order to solve this challenge, this study proposes a semi-supervised learning algorithm for data augmentation, the style-transfer-generated data as an extra class (STGDEC), which is aided by the Cycle-Consistent Adversarial Networks (CycleGANs) in generating extra unlabeled training data. Specifically, the algorithm firstly trains the CycleGANs and Deep Convolutional Generative Adversarial Networks (DCGANs) so as to generate large amounts of unlabeled data. Secondly, we propose a recognition model that depends on what kind of label distribution the generated samples are. Finally, this paper proposes a deformed loss function, the label smoothing and only extra class regularization (LSOECR), which is the combination of the Label Smoothing Regularization for Outliers (LSRO) and the additional added class. Comprehensive experiments based on the STGDEC are conducted, and these results show that the

	proposed algorithm gains a significant improvement over the ID-discriminative embedding (IDE) baseline, the Basel.+LSRO and state-of-the-art approaches of person re-ID in many cases.
Suggested Reviewers:	Weishi Zheng wszheng@ieee.org Prof. Weishi Zheng's main research interest is machine vision and intelligent learning, mainly for the human-centered image recognition, including person re-identification.
	Jianhuang Lai stsljh@mail.sysu.edu.cn Prof. Jianhuang Lai's main research areas are computer vision, biometrics, digital image processing, pattern recognition and machine learning.
	Liang Zheng liang.zheng@anu.edu.au Dr. Liang Zheng is working on interesting topics in computer vision: applications including person re-identification and medical image understanding; fundamental problems including deep learning and synthetic data analysis.

Noname manuscript No.
(will be inserted by the editor)

Semi-supervised learning for person re-identification based on style-transfer-generated data by CycleGANs

Shangdong Zhu[1] · Yunzhou Zhang[2] · Sonya Coleman[3] · Song Wang[1] · Ruilong Li[2] · Shuangwei Liu[2]

Abstract Person re-identification (re-ID) is an exceedingly significant branch in the field of computer vision, especially for video surveillance. It is still a challenge to obtain more labeled training data and use them reasonably for more precise matching, though the person re-ID performance has been improved significantly. In order to solve this challenge, this study proposes a semi-supervised learning algorithm for data augmentation, the style-transfer-generated data as an extra class (STGDEC), which is aided by the Cycle-Consistent Adversarial Networks (CycleGANs) in generating extra unlabeled training data. Specifically, the algorithm firstly trains the CycleGANs and Deep Convolutional Generative Adversarial Networks (DCGANs) so as to generate large amounts of unlabeled data. Secondly, we propose a recognition model that depends on what kind of label distribution the generated samples are. Finally, this paper proposes a deformed loss function, the label smoothing and only extra class regularization (LSOECR), which is the combination of the Label Smoothing Regularization for Outliers (LSRO) and the additional added class. Comprehensive experiments based on the STGDEC are conducted, and these results show that the proposed algorithm gains a significant improvement over the ID-discriminative embedding (IDE) baseline, the Basel.+LSRO and state-of-the-art approaches of person re-ID in many cases.

Keywords Person re-identification · Generative adversarial network · Data augmentation · Semi-supervised learning · Convolutional neural network

✉ Yunzhou Zhang
zhangyunzhou@ise.neu.edu.cn

Shangdong Zhu
zhushangdong@gmail.com

Sonya Coleman
sa.coleman@ulster.ac.uk

Song Wang, Ruilong Li and Shuangwei Liu
{1801946, 1700912, 1700915}@stu.neu.edu.cn

[1] Faculty of Robot Science and Engineering, Northeastern University, Shenyang, 110819, China

[2] College of Information Science and Engineering, Northeastern University, Shenyang, 110819, China

[3] Intelligent Systems Research Centre, University of Ulster, Derry, BT52 1SA, United Kingdom

1 Introduction

The intelligent video analysis method based on person re-identification (re-ID) [39] recently has become a research hotspot in the field of computer vision. In this task, pedestrian images undergo intensive changes in appearance and background frequently. Due to important applications of person re-ID in security and monitoring, it has received extensive attention from academia and industry. Person re-ID, a cross-camera retrieval task [41], which aims to verify a probe person's identity in image sequences captured by cameras that are orientated in different directions at different times. These algorithms essentially study the similarity measurements and the pedestrian features between the learned features, which are robust and view-invariant to the monitoring camera changes.

Due to the rapid development of the usability of a large amount of datasets and deep learning, the performance of person re-ID has been improved significantly. The majority of the proposed person re-ID algorithms [1,4,16,21,28,33] conduct supervised learning on small set of labeled training datasets. Since the target domain might differ from the small and practical training

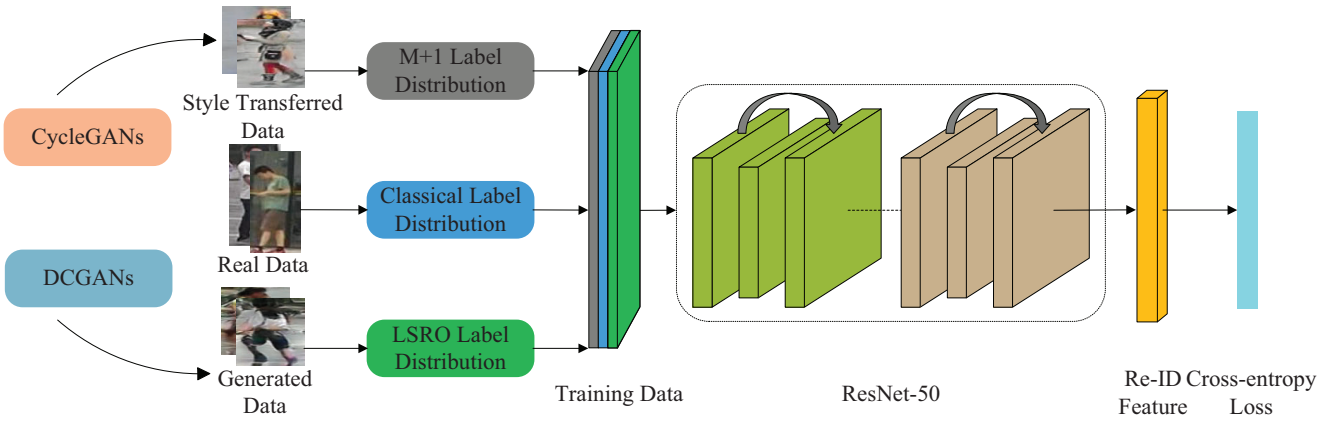

Fig. 1 Overview of the proposed semi-supervised learning framework. The training data are divided into three classes: 1) The original real training data is assigned the classical label distribution, i.e., the ground-truth label distribution; 2) The data generated by DCGANs is assigned LSRO label distribution; 3) The data generated by CycleGANs is assigned M+1 label distribution, where M represents the total number of pedestrian identities in the real training data domain.

dataset significantly, applying these well-trained models directly to a real-world scene with a massive network of cameras can result in performance degradation. Therefore, incremental optimization in practical applications is crucial to obtaining the excellent performance of person re-ID algorithms. Nevertheless, marking large amounts of online monitoring videos to support supervised learning is often expensive and impractical. How to obtain and utilize the rich unmarked data reasonably is a challenging and practical issue.

In order to solve this issue, some semi-supervised algorithms [40,10] are proposed to generate artificial sample data through Generative Adversarial Networks (GANs) [9] and assign reasonable label distribution to the unlabeled data for regularization. Furthermore, there is also a proposed semi-supervised person re-ID framework [34] that uses some extra unlabeled images and only a small part of labeled samples. In general, the performance of these algorithms are not satisfactory. On the other hand, some unsupervised algorithms [41, 15] to measure the similarity between different samples and extract view-invariant features in unmarked datasets are proposed. These unsupervised algorithms usually perform poorly without strong supervised adjustment and optimization. An unsupervised algorithm which is cross-dataset transfer learning [22] has been proposed recently in addition to these unsupervised algorithms used in an individual dataset, which uses a dictionary learning mechanism from a marked source dataset to another unmarked target dataset to transfer the view-invariance representation of one person's appearance, and achieves much better performance than some of the unsupervised algorithms used in an individual dataset. Nevertheless, the performance of the methods mentioned above is still far lower than that of the supervised learning approaches.

In this paper, we propose a semi-supervised learning algorithm, named the style-transfer-generated data as an extra class (STGDEC), to enable high person re-ID performance with unmarked datasets generated by Cycle-Consistent Adversarial Networks (CycleGANs) [42]. Unlike the above semi-supervised methods, which generate unlabeled datasets only based on the GANs or Deep Convolutional Generative Adversarial Networks (DCGANs) [23], this paper tries to learn and integrate with the samples generated by DCGANs and CycleGANs in the steps shown in the leftmost column of Fig. 1. Firstly, we apply the DCGANs and CycleGANs models to generate unlabeled datasets to extend the original training data. Secondly, a recognition model is proposed that depends on what type of label distribution the generated samples are. Finally, this paper proposes a novel deformed loss function, the label smoothing and only extra class regularization (LSOECR), which is the combination of the Label Smoothing Regularization for Outliers (LSRO) and the additional added class.

Comparative experiments on the Market-1501 [38] dataset and the DukeMTMC-reID [40] dataset indicate that the proposed STGDEC performs much better than the Basel.+LSRO [40] and some other semi-supervised algorithms. Meanwhile, the STGDEC performs considerable or even better than, to some extent, some of the state-of-the-art supervised methods using the exact same datasets. The main contributions can be summarized as follows:

1. This paper proposes a semi-supervised learning algorithm for data augmentation, the style-transfer-generated data as an extra class (STGDEC), which

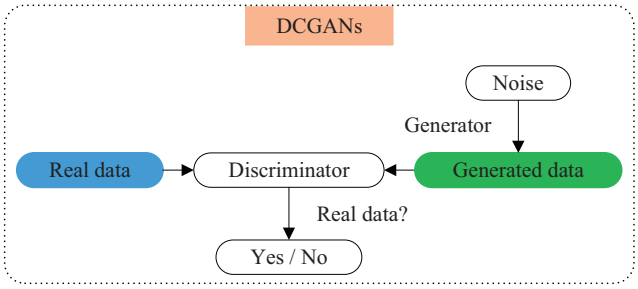

Fig. 2 Illustration of a DCGANs framework which contains a generator and a discriminator. Both of the generator and discriminator are consist of convolutional neural networks.

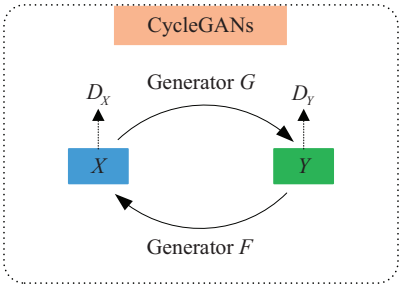

Fig. 3 Illustration of a CycleGANs framework which contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_X and D_Y.

is aided by CycleGANs in generating extra unmarked training data.

2. This study proposes a recognition model that depends on the label distribution of the generated samples.
3. This research proposes a deformed loss function, the label smoothing and only extra class regularization (LSOECR), which is the combination of the LSRO and the additional added class.

2 Related work

For the person re-ID task, we review the existing works related to the GANs, supervised learning and semi-supervised learning in this section.

2.1 Generative adversarial network

In [9], Goodfellow et al. first propose the GANs so as to obtain the generated samples and understand the neural networks in depth. After that, DCGANs (shown in Fig. 2) is proposed by Radford et al. [23] to introduce techniques for case of improving the stability of training. GANs have been applied widely in many fields recently, especially for image generation [23]. In recent years, GANs also have been used in the fields of style transfer [8], image-to-image translation [11,42,17] and cross-domain image generation [27]. Among these areas, image-to-image translation has received a lot of attention. Isola et al. [11] learn to map from an input image to an output image by using conditional GANs to perform image-to-image translation. Nevertheless, paired training samples are difficult to obtain required in quite a few tasks [42]. In order to overcome this problem, Liu and Tuzel [17] introduce the coupled generative adversarial network (CoGAN), which is used to learn the joint distribution of multi-domain images. Also, the work [42] proposes a similar framework, the CycleGANs (shown in Fig. 3), which uses cycle consistency loss and aims to overcome the unpaired image-to-image translation task. Unlike most of previous algorithms that primarily consider the quality of generated images, this paper directly uses the combination of the style-transfer-generated samples and the images generated by DCGANs so as to improve the person re-ID performance.

2.2 Supervised person re-ID

Most of the proposed person re-ID algorithms are supervised algorithms based on either deep learning [1,4, 33], metric learning [21,28] or invariant feature learning [16]. Among them, the deep learning based algorithm of person re-ID tasks has been particularly popular recently. Meanwhile, the performance reported is generally superior to conventional methods [36,14] which are not using deep learning. Nevertheless, labeling lots of online monitoring videos to perform supervised learning is often expensive and unpractical in the practical application of person re-ID methods in a massive network of cameras, as described in [22].

2.3 Semi-supervised person re-ID

The semi-supervised learning uses both marked and unmarked samples to strengthen the performance of a learning task which is driven by the practical significance for learning cheaper, better and faster feature representations. It is reasonably easy to obtain plenty of unmarked samples in many practical applications, such as pedestrian detection and image segmentation. The goal of semi-supervised learning is to train models that can predict future unseen test data more accurately than models that only learned from labeled training data.

Some semi-supervised methods of person re-ID have been proposed. Figueira et al. [7] propose the multi-

feature semi-supervised learning approach to adderss the learning-based and appearance-based person re-ID issue jointly. In order to learn two paired dictionaries from both marked and unmarked data jointly in the training stage, Liu et al. [18] introduce the semi-supervised paired dictionary learning algorithm. In a semi-supervised setting, Yang et al. [35] propose the technology of multi-kernel embedding within a learning framework of self-training subspace for the person re-ID tasks, which effectually handles the nonlinearity in the appearance of a person and explores the supplementary information shared between multi-kernels. Nevertheless, these algorithms which are not based on deep learning simply cope with small-scale datasets.

In the past few years, some deep learning based algorithms of semi-supervised person re-ID have been proposed. In [19], Liu et al. design a simple but efficient learning algorithm, which only replaces the last fully connected layer of the CNN with the presented Transductive Centroid Projection (TCP) module. Meanwhile, they introduce a new large-scale dataset called Person Tracker Re-Identification dataset (PT-ReID) as unmarked data. Xin et al. [34] introduce a semi-supervised algorithm for a person re-ID task that enable to use multiple CNN models and both marked and unmarked training data to improve the person re-ID performance. The works of semi-supervised person re-ID mentioned above primarily concentrate on how to use enough unmarked samples with available marked samples to improve the person re-ID performance as obtaining training labels is expensive. After the emergence of GANs [9], the research branch of semi-supervised learning has turned to exploring GANs-generated images [25]. Also, exploring how to apply the generated images to person re-ID has become a prominent research topic. Huang et al. [10] propose a virtual label named Multi-pseudo Regularized Label (MpRL) and assign this label to the generated samples, which are able to used to supplement real training samples to train a deep neural network in the semi-supervised mannner. Ding et al. first [5] propose the semi-supervised method, which employs pseudo-marking by considering the complicated relationships between marked and unmarked training data in feature space. Zheng et al. [40] introduce the LSRO method that uses the uniform label distribution for the unmarked data to regularize the neural network training in the person re-ID task. All of the literature mentioned above use GANs to generate artificial labeled training samples in the feature space. Therefore, we use the existing training samples directly in this paper for generating unmarked samples by GANs, and incorporate unmarked data generated by GANs alongside with the real data available in the marked datasets. In addition, we demonstrate that the generated samples are able to help improve the performance of the discriminative model by assigning the pre-determined label distribution.

3 The proposed algorithm

In this section, this paper introduces the proposed semi-supervised framework in detail for the person re-ID task. Firstly, we briefly look back at the virtual label LSRO [40] which is the state of the art. Secondly, the STGDEC algorithm is introduced in the model, meanwhile, the deformed LSOECR is illustrated. Finally, we present three training strategies based on the proposed STGDEC model. The framework of semi-supervised learning for the person re-ID task is demonstrated in Fig. 1.

3.1 The review of the LSRO algorithm for the person re-ID task

The design of the LSRO method is inspired from the Label Smoothing Regularization (LSR) [26] method, which assigns relatively low-value weights to the non-ground-truth categories instead of 0 and assigns small confidence to the ground-truth label. LSRO assumes that all of the generated data do not fall into any existing classes and assigns uniform label distribution to each of them to avoid model over-fitting. Specifically, for a real training image X, the cross-entropy loss $\mathcal{L}_R$ can be defined as follows:

$$\mathcal{L}_R = -\sum_{m=1}^{M} q(m)\log(p(Y_m)), \tag{1}$$

where M represents the total number of the predefined training classes, that is, the number of the pedestrian identities in the real training data domain in the person re-ID task. m belongs to $\{1, 2, ..., M\}$ is the m-th training class and Y_m represents the output of the m-th predefined training category. $p(Y_m) \in (0, 1)$ denotes the softmax predictive probability of X which belongs to the predefined training category m and $q(m)$ represents the ground-truth label distribution. Assuming that n is the ground-truth category label of X, $q(m)$ and $p(Y_m)$ can be respectively given by the two following equations:

$$q(m) = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}, \tag{2}$$

$$p(Y_m) = \frac{e^{Y_m}}{\sum_{i=1}^{M} e^{Y_i}}. \tag{3}$$

Fig. 4 The label distributions of a real sample and a generated sample. The real data (a) is assigned the ground-truth label, i.e., the classical label distribution. For a generated sample, one-hot (b) assigns each unmarked sample a dynamic class label in each training epoch. LSRO (c) uses a uniform label distribution. All-in-one (d) treats all unmarked images belong to a new class. (e) the label distribution proposed in this work.

By removing the 0 terms in Eq. (1), i.e., merely considering the ground-truth term, the loss $\mathcal{L}_R$ can be redefined as follows:

$$\mathcal{L}_R = -\log(p(Y_n)). \tag{4}$$

Given the generated image g, its uniform label distribution $q^g_{LSRO}(m)$ can be formulated as:

$$q^g_{LSRO}(m) = \frac{1}{M}. \tag{5}$$

Therefore, the cross-entropy loss of the LSRO method for the generated image g can be defined as:

$$\mathcal{L}^g_{LSRO} = -\frac{1}{M}\sum_{m=1}^{M}\log(p(Y_m)). \tag{6}$$

3.2 The proposed STGDEC algorithm for the person re-ID task

The proposed STGDEC algorithm is to be applied to setting visual labels on the generated samples in order to merge the unlabeled samples into the network. Nevertheless, unlike the LSRO method, we do not assign uniform label distribution to all the generated data, instead, we set uniform label distribution and all-in-one label distribution (shown in Fig. 4) to the generated data by DCGANs (images shown in Fig. 5) and CycleGANs (images from DukeMTMC-reID to Market-1501 or from Market-1501 to DukeTMTC-reID are shown in Fig. 6 and Fig. 7 respectively) respectively. For the sake of convenience, we simply abbreviate the data generated by DCGANs as D-images and the samples generated by CycleGANs as C-images.

For a real training image X, which is assigned a ground-truth label distribution, the cross-entropy loss $\mathcal{L}_R$ is equal to Eq. (1). For a D-image, which is assigned a uniform label distribution, the cross-entropy loss $\mathcal{L}^g_{LSRO}$ is the same as Eq. (6). For C-images, we apply the all-in-one label distribution to all of them, i.e., let all the corresponding samples belong to the additional $M+1$ class. The loss $\mathcal{L}^g_C$ of a C-image can thus

(a) Market-1501. (b) DukeMTMC-reID.

Fig. 5 Examples of images generated by DCGANs and real images on the Market-1501 dataset and the DukeMTMC-reID dataset. The top two rows of (a) and (b) present the pedestrian images generated by DCGANs trained on the Market-1501 and DukeMTMC-reID. The bottom row presents the real images in the training set.

be defined as follows:

$$\mathcal{L}_C^g = -\sum_{m=1}^{M+1} q(m)\log(p(Y_m)), \tag{7}$$

where m belongs to $\{1, 2, ..., M, M+1\}$.

Similar to Eq. (4), Eq. (7) can also be simplified as follows:

$$\mathcal{L}_C^g = -\log(p(Y_n)). \tag{8}$$

If we train using the real training data and the C-images together, the cross-entropy loss $\mathcal{L}_{R+C}^g$ for each of the images in the dataset is equal to Eq. (7). On the other hand, if we train with the real training data alongside with the C-images and D-images, the cross-entropy loss $\mathcal{L}_{R+C+LSRO}^g$ (i.e., $\mathcal{L}_{STGDEC}^g$) can be formulated as follows:

$$\begin{aligned}\mathcal{L}_{STGDEC}^g = &-(1-K)\log(p(Y_n)) \\ &-\frac{K}{M+1}\sum_{m=1}^{M+1}\log(p(Y_m)),\end{aligned} \tag{9}$$

where this system has two types of losses: 1) for each of the samples in the combination of the real training data and the C-images, the value of K is 0; 2) for each of the D-images, the value of K is 1. Additionally, we need to further explain this training method. For the training mechanism using mixed data, since the C-images participate in the training, the number of training data classes is $M+1$ in the first form of loss. Meanwhile, in the second form of loss, though the number of training data classes is M as there are only D-images, we still set $M+1$ to accommodate the last classification layer in the training network.

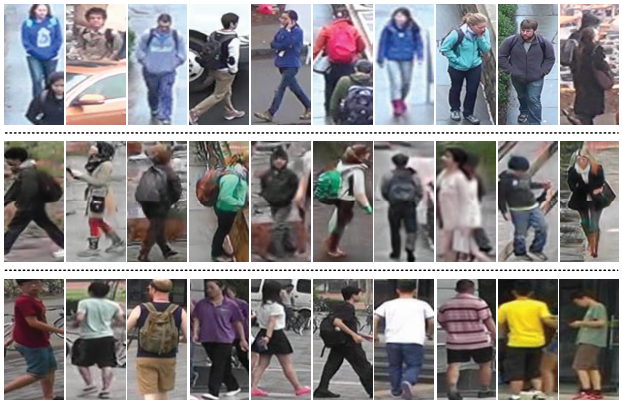

Fig. 6 Style-transfer-generated samples by CycleGANs from the DukeMTMC-reID to Market-1501and real images on the DukeMTMC-reID dataset and the Market-1501 dataset. Transferred images to the Market-1501 are shown in the second row. Images in the first and third rows are real samples from the DukeMTMC-reID and Market-1501 respectively.

On the basis of the proposed STGDEC algorithm and the deformed loss LSOECR, we briefly introduce three training strategies in the next subsection.

3.3 Training strategies

There are three training strategies for applying the proposed STGDEC approach to the experiment. The first strategy is the training mechanism of combining the real training data with the C-images. In this strategy, the real data are assigned the ground-truth label distribution, and the C-images are assigned all-in-one label distribution. In the second and the third strategies, we add the D-images and assign uniform label distribution to all of them based on the training mechanism of combining the real training data with the C-images and the D-images. Since in [40], DCGANs uses the Market-1501 (or DukeMTMC-reID) dataset as the training set to generate samples, that is, the target domain is used as the training set of DCGANs. Subsequently, since the person re-ID network is tested on the target domain, the generated images have a certain degree of similarity with the target domain. Thus, it is reasonable to assign the LSRO label to the generated samples. However, the CycleGANs performs style conversion on the source domain in order to augment the data size of the target domain. Though the samples generated in this way have the same style as the target domain, the data source is from a different dataset. Objectively, the images generated in this way are almost not similar to that in the target domain. Therefore, this paper considers assigning all-in-one label to the C-images. Descriptions are as follows:

Fig. 7 Style-transfer-generated samples by CycleGANs from the Market-1501 dataset to the DukeMTMC-reID dataset and real images on the Market-1501 dataset and the DukeMTMC-reID dataset. Transferred images to the DukeMTMC-reID are shown in the second row. Images in the first and third rows are real samples from the Market-1501 and DukeMTMC-reID respectively.

3.3.1 STGDEC-I

In this strategy, we extend the original training dataset by adding C-images which are assigned an all-in-one label distribution. For fair comparison and approximate proportion of samples, 12,000 C-images are randomly selected. We demonstrate that the result of this strategy in the experiments can be improved compared with that Basel.+LSRO in Section 4.3.

3.3.2 STGDEC-II

The original training dataset is expanded by adding 12,000 C-images and 12,000 D-images to this strategy. Different from the C-images, the D-images are all assigned a uniform label distribution. The later experimental result shows that the STGDEC-II strategy also outperforms the Basel.+LSRO.

3.3.3 STGDEC-III

Similar to the second training strategy, this strategy is also a combination of C-images and D-images, which differs from the last strategy only in the number of D-images, i.e., 24,000. The experimental result of this strategy demonstrates that the STGDEC algorithm gains a significant improvement over the ID-discriminative embedding (IDE) baseline [39], Basel.+LSRO and state-of-the-art approaches of person re-ID in most cases. Algorithm 1 shows the detailed training process of this strategy.

Algorithm 1: Detailed training process of the STGDEC-III strategy: The original training dataset is expanded by adding 12,000 C-images and 24,000 D-images to this strategy. All of the C-images are assigned an all-in-one label distribution. Different from the C-images, the D-images are all assigned a uniform label distribution.

Input: Real data set: R;
C-images: C;
D-images: D;
Incorporated dataset: $M = R \cup C \cup D$;
Loss of the real dataset and C-images: $\mathcal{L}^g_{R+C}$;
Loss of the D-images: $\mathcal{L}^g_{LSRO}$.

for *the number of the training epochs* **do**
  Shuffle M;
  for *the number of the training iterations in every epoch* **do**
    Set $\mathcal{L}^g_{R+C} = 0$, $\mathcal{L}^g_{LSRO} = 0$;
    Sample minibatch from $M \rightarrow M'$;
    if $M'_i \in R \cup C$ **then**
      Set $K = 0$ in the Eq. (9);
      Calculate loss $\mathcal{L}^g_{R+C}$;
    else
      Set $K = 1$ in the Eq. (9);
      Calculate loss $\mathcal{L}^g_{LSRO}$;
    Calculate the final loss $\mathcal{L}^g_{STGDEC}$;
    Backward propagation;
    Update parameters;
final;

4 Experiment

In this section, we use two re-ID datasets (the Market-1501 dataset and the DukeMTMC-reID dataset) to validate the availability of the proposed STGDEC algorithm, since they are both large-scale datasets and have fixed test/training split.

4.1 Person re-ID datasets and evaluation metrics

4.1.1 The Market-1501 dataset

The Market-1501 is collected with six cameras in the open environment of Tsinghua University. The training set contains 12,936 images and the testing set contains 19,732 images. The training set and the testing set contain 751 identities and 750 identities respectively. A total of 1,501 identities are annotated. There are 17.2 images on average for each identity in the training set and 26.3 images on average for each identity in the testing set. Both of multi-query settings and single-query settings are applied to this dataset. All of the images are automatically detected through the deformable part

model (DPM) [6] algorithm. In this paper, all of the 12,936 automatically detected images in the training set are used for training the DCGANs to get abundant unlabeled samples. Furthermore, we randomly select 1,645 training images to train the CycleGANs.

4.1.2 The DukeMTMC-reID dataset

The DukeMTMC-reID dataset is collected with eight cameras in Duke University, which is in the same format as Market-1501 dataset. It is a subset of the original DukeMTMC dataset [24] for the purpose of image-based re-recognition. In this paper, we employ a subset of [24], i.e., the re-ID version benchmark [40], for the image-based person re-ID task. It contains 1,404 identities, randomly selecting 702 of them for the training set (a total of 16,522 images), and then using the remaining 702 identities for the testing set. All of the 1,404 identities do not appear in just one camera. In this testing set, a given query image for every identity is selected in every camera and the rest of the images are placed in the gallery. Therefore, for the 702 testing identities, there are 2,228 query images and 17,661 gallery images. In this study, we use this training set which contains 16,522 images to train the DCGANs to generate unlabeled images. Moreover, 1,472 training images are randomly picked to train the CycleGANs model.

4.1.3 Evaluation protocol

We employ the mean average precision (mAP) and rank-1 accuracy for evaluating the presented STGDEC algorithm on the DukeMTMC-reID dataset and the Market-1501 dataset. The rank-i accuracy represents the rate at which one or more correctly matched images appear in the top-i ranked images. The mAP value reflects the overall accuracy and recall rate, thus providing a more comprehensive evaluation metric.

4.2 Implementation details

4.2.1 CNN baseline model for person re-ID

In this paper, the ResNet-50 model is employed as the primary structure of the proposed method. The network structure has been applied to evaluating highly correlated pseudo-labeling methods, such as the one-hot pseudo label [13], all-in-one [25] and LSRO. The Matconvnet [32] package is used to implement the network. There are no other changes to the training architecture other than replacing the number of neurons in the final classification layer with the number of the target identities (703 and 752 for the DukeMTMC-reID dataset and the Market-1501 dataset, respectively). Before randomly cropping all images to 224×224 pixels by random horizontal flip, all of the images will be resized to 256×256 pixels firstly. In order to prevent over-fitting, we need to insert one dropout layer before the last convolutional layer and set the dropout rate for both of the two datasets to 0.75. We use the Stochastic Gradient Descent (SGD) algorithm with a momentum of 0.9 to optimize the entire model. The training is performed for a total of 50 epochs. We set the learning rate for each convolutional layer to 0.002, which decays to 0.0002 after 40 epochs. During testing, the last fully-connected layer with 2048-dim activations is extracted for a 224×224 pixels input image so as to calculate the similarity between two images for ranking.

4.2.2 GANs models

In the experiments, we use the CycleGANs model and the DCGAN-tensorflow package to generate the unmarked images from the given training samples in the initial training set without preprocessing. We adopt Tensorflow to train the above two models. For fair comparison, we also adopt the same setup of the model architecture in the above two models. For the DCGANs model: 1) all of the images are resized to 128×128 pixels and randomly flip before training; 2) in order to obtain stochastic optimization, we use the Adam [12] algorithm with parameter β_1 and β_2 (the values of β_1 is 0.5 and the value of β_2 is 0.99); 3) after 30 epochs, the training is stopped; 4) the generator takes a 100-dimensional vector randomly as input; 5) for the later semi-supervised CNN training, we resize all of the unlabeled images to 256×256 pixels. For the CycleGANs model, the setup is the same as in [42]. The generated samples are still able to improve performance and regularize this model, while some of them are far from the real data distribution. In Section 4.3, we present the results of experiments. Some of the data generated by DCGANs (the D-images) are shown in Fig. 5 and CycleGANs (the C-images) are shown in Fig. 6 and Fig. 7.

4.3 Experimental results

4.3.1 The C-images improve the baseline and the result outperforms the Basel.+LSRO

As can be seen in Table 1, when 12,000 C-images are added to the CNN baseline model for training, the proposed STGDEC-I improves the person re-ID baseline on both two datasets in the single-query way. Compared

Table 1 Comparison of our results between the Basel., Basel.+LSRO and STGDEC-I on the Market-1501 dataset and the DukeMTMC-reID dataset in the single-query way.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
Basel. [39]	73.69	51.48	65.22	44.99
Basel.+LSRO [40]	78.06	56.23	67.68	47.13
STGDEC-I (Ours)	**78.39**	**57.02**	**67.87**	**48.03**

Table 2 Comparison of our results between the Basel., Basel.+LSRO and STGDEC-II on the Market-1501 dataset and the DukeMTMC-reID dataset in the single-query way.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
Basel. [39]	73.69	51.48	65.22	44.99
Basel.+LSRO [40]	78.06	56.23	67.68	47.13
STGDEC-II (Ours)	**78.27**	**56.83**	**67.77**	**47.87**

Table 3 Comparison of our results with the published state-of-the-art algorithms on the Market-1501 dataset. All our results of three training strategies are both in single-query way and multi-query way and shown in bold. Rank-1 and mAP are listed.

Method	Single-query		Multi-query		Reference
	Rank-1	mAP	Rank-1	mAP	
BoW+kissme [38]	44.42	20.76	–	–	2015 ICCV
MR CNN [29]	45.58	26.11	56.59	32.26	2015 arXiv
FisherNet [33]	48.15	29.94	–	–	2017 PR
SL [3]	51.90	26.35	–	–	2016 CVPR
S-LSTM [31]	–	–	61.6	35.3	2016 ECCV
DNS [37]	61.02	35.68	71.56	46.03	2016 CVPR
Gate Reid [30]	65.88	39.55	76.04	48.45	2016 ECCV
SOMAnet [2]	73.87	47.89	81.29	56.98	2018 CVIU
Basel. [39]	73.69	51.48	81.47	63.95	2016 arXiv
MVC Re-ID [34]	75.23	52.63	80.08	59.54	2019 PR
SpindleNet [43]	76.90	–	–	–	2017 CVPR
Basel.+LSRO [40]	78.06	56.23	85.12	68.52	2017 ICCV
STGDEC-II	78.27	56.83	85.52	69.05	Proposed
STGDEC-I	78.39	57.02	85.67	69.13	Proposed
STGDEC-III	**78.57**	**57.07**	**86.25**	**69.49**	Proposed

with the Basel. in Table 1, we can see the improvements of 4.70% (78.39 - 73.69) in rank-1 and 5.54% (57.02 - 51.48) in mAP on the Market-1501 dataset. Meanwhile, it can be seen that the improvements of 2.65% (67.87 - 65.22) in rank-1 and 3.04% (48.03 - 44.99) in mAP on the DukeMTMC-reID dataset. These results show that the added C-images improve the baseline based on the STGDEC-I strategy. Compared with the 2^{nd} best Basel.+LSRO, there are also improvements of 0.33% (78.36 - 78.06) in rank-1, 0.79% (57.02 - 56.23) in mAP on the Market-1501 dataset and 0.09% (67.87 - 67.68) in rank-1, 0.9% (48.03 - 47.13) in mAP on the DukeMTMC-reID dataset. Therefore, these results indicate that the STGDEC-I strategy outperforms the Basel.+LSRO method, even though the number of the GANs images used in STGDEC-I is 12,000 less than the Basel.+LSRO, i.e., only uses half of the GANs images.

4.3.2 The combination of the C-images and D-images improve the baseline and the result outperforms the Basel.+LSRO

As shown in Table 2, we use the STGDEC-II implementation to demonstrate that the combination of the C-images and D-images assigned different label distribution improves the person re-ID baseline on both two datasets in the single-query way. Compared with the Basel. in Table 2, we can see that by adding 12,000 C-images and 12,000 D-images to both of the above training sets, the STGDEC-II strategy (improvements of 4.58% (78.27 - 73.69) in rank-1 and 5.35% (56.83 - 51.48) in mAP on the Market-1501 dataset, 2.55%

Table 4 Comparison of our results with the published state-of-the-art algorithms on the DukeMTMC-reID dataset. All our results of three training strategies are both in single-query way and multi-query way and shown in bold. Rank-1 and mAP are listed.

Method	Single-query		Reference
	Rank-1	mAP	
BoW+kissme [38]	25.13	12.17	2015 ICCV
LOMO+XQDA [16]	30.75	17.04	2015 CVPR
MVC Re-ID [34]	55.70	37.80	2019 PR
Basel. [39]	65.22	44.99	2016 arXiv
Basel.+LSRO [40]	67.68	47.13	2017 ICCV
STGDEC-II	67.77	47.87	Proposed
STGDEC-I	67.87	48.03	Proposed
STGDEC-III	**69.28**	**48.42**	Proposed

(67.77 - 65.22) in rank-1 and 2.88% (47.87 - 44.99) in mAP on the DukeMTMC-reID dataset) achieves improvements over the baseline. In comparison with the Basel.+LSRO method, we achieve 0.21% (78.27 - 78.06) in rank-1 and 0.60% (56.83 - 56.23) improvements in mAP on the Market-1501 dataset. Meanwhile, we observe improvements of 0.09% (67.77 - 67.68) in rank-1 and 0.74% (47.87 - 47.13) in mAP on the DukeMTMC-reID dataset. Although result of this strategy slows only a small improvement over the Basel.+LSRO, this implementation indicates that this kind of combination of C-images and D-images helps to enhance the overall person re-ID performance.

4.3.3 The effects when employing different quantities of the GANs images

Our goal is to explore how different quantities of GANs samples affects the person re-ID performance. Since unlabeled images are readily available, we hope that as the growing number of unmarked images, the model is able to learn more common sense. Table 3 shows the experimental results on the Market-1501 dataset while using different quantities of GANs images. What we need to point out is that the quantity of the real training samples in the Market-1501 dataset is 12,936. By observing and comparing the results, we draw two conclusions. Firstly, adding different quantities of C-images and D-images can continuously improve the baseline. Secondly, the optimal performance can be achieved when 12,000 C-images and 24,000 D-images are added into the STGDEC-III. The reason why the STGDEC-III strategy achieves the best performance is that this strategy has the best combination of different quantities of C-images and D-images.

4.3.4 The comparison between three different modes of the STGDEC

We use three different training strategies in the experiments to present the validity of the proposed STGDEC, the experimental results on the Market-1501 dataset and the DukeMTMC-reID dataset are shown in Table 3 and Table 4 respectively. It can be seen that by adding double the D-images to the training dataset, the STGDEC-III strategy achieves better improvements compared with the STGDEC-II on both of the two datasets. Though the experiments show that the STGDEC-I achieves better performance than the STGDEC-II, STGDEC-III achieves the optimal performance. This is because there is a more reasonable training data ratio.

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

We perform a number of experiments and the results show that the proposed algorithm achieves better than many of the existing methods.

Comparison on the Market-1501 dataset: we can see in Table 3 that the proposed STGDEC-III is better than all the listed state-of-the-art methods, surpassing the second-best Basel.+LSRO by improvements of 0.54% (78.36 - 78.06) in rank-1 and 0.84% (57.07 - 56.23) in mAP for the single-query mode, 1.13% (86.25 - 85.12) in rank-1 and 0.97% (69.49 - 68.52) in mAP for the multi-query mode.

Comparison on the DukeMTMC-reID dataset: as can be seen in Table 4, the performance of our proposed STGDEC-III is higher than all the compared existing algorithm, surpassing the 2^{nd} best Basel.+LSRO by 1.6% (69.28 - 67.68) in rank-1 and 1.29% (48.42 - 47.13) in mAP for the single-query mode.

Compare to Basel.+LSRO, our model contains two types of GANs samples and the combination of the GANs images, not just the semi-supervised algorithm as the Basel.+LSRO does. It demonstrates the regularization of diverse kinds of GANs data.

4.3.6 Ablation analysis

For the ablation analysis, we assign the opposite label distribution to the C-images and D-images, which is different from the label distribution of the real data and the generated samples in the three training strategies proposed in this paper. Specifically, we assign the label type LSRO or all-in-one to the C-images and D-images in the form of mathematical combination, removing the assigned label distribution in our model and [40], and the rest is all cases with the reverse label distribution. As Shown in Table 5 and 6, all the results of the ablation

Table 5 Experimental results of STGDEC-II after assigning the opposite label distribution to the C-images and D-images. Rank-1 and mAP are listed. D represents D-images, C represents C-images, 1.2 represents 12,000, 2.4 represents 24,000.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
D_2.4_All-in-one [40]	76.63	55.12	–	–
D_2.4_Pseudo label [40]	75.80	53.03	–	–
D_1.2_LSRO+C_1.2_LSRO	77.35	55.92	66.83	46.73
D_1.2_All-in-one+C_1.2_LSRO	76.75	55.23	67.30	46.89
D_1.2_All-in-one+C_1.2_All-in-one	77.51	55.72	67.40	47.05
Basel.+LSRO [40]	78.06	56.23	67.68	47.13
STGDEC-II (ours)	78.27	56.83	67.77	47.87

Table 6 Experimental results of STGDEC-I after assigning the opposite label distribution to the C-images and D-images. Rank-1 and mAP are listed. D represents D-images, C represents C-images, 1.2 represents 12,000.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
D_1.2_All-in-one [40]	75.33	52.82	–	–
D_1.2_Pseudo label [40]	76.07	53.56	–	–
D_1.2_LSRO [40]	76.81	55.32	–	–
C_1.2_LSRO	76.91	55.73	66.07	46.25
Basel.+LSRO [40]	78.06	56.23	67.68	47.13
STGDEC-I (ours)	78.39	57.02	67.87	48.03

Table 7 The performance of different models is evaluated on cross-domain datasets. M→D means that the model is first trained on the Market-1501 dataset and then tested on the DukeMTMC-reID dataset. D→M means that the model is first trained on the DukeMTMC-reID dataset and then tested on the Market-1501 dataset.

Method	M→D		D→M	
	Rank-1	mAP	Rank-1	mAP
Basel. [39]	25.8	12.8	34.4	13.8
Basel.+LSRO [40]	25.7	12.8	35.3	14.1
Baseline [20]	24.4	12.9	34.2	**14.5**
STGDEC-II (Ours)	**26.5**	**13.4**	**35.9**	14.0

study are not higher than the Basel.+LSRO, and thus not higher than the experimental results of our method. This ablation analysis indicates the reasonability and effectiveness of the label distribution proposed in our algorithm.

4.3.7 Cross-domain experiments

In order to validate the availability of the proposed STGDEC algorithm, we also perform comparative cross-domain experiments on both the Market-1501 dataset and the DukeMTMC-reID dataset. The cross-domain experiments on both the above two datasets are that the model is first trained on the Market-1501 dataset and then tested on the DukeMTMC-reID dataset, or first trained on the DukeMTMC-reID dataset and then tested on the Market-1501 dataset. Compared with the 2^{nd} best person re-ID baseline [20] in Table 7, we can see the improvements of 0.5% (13.4 - 12.9) in rank-1 and 2.1% (26.5 - 24.4) in mAP on the Market-1501 dataset, 1.7% (35.9 - 34.2) in mAP on the DukeMTMC-reID dataset. These cross experiments present the availability of the proposed STGDEC algorithm.

5 Conclusion

We propose the novel virtual label STGDEC for the data generated by the DCGANs and the CycleGANs. Aiming at training with a CNN-based model, we use the STGDEC to assign virtual label to the C-images and D-images, which are used to perform semi-supervised learning. Using a baseline DCGANs model and a baseline CycleGANs, we present that the imperfect GANs samples effectually indicate their regularization ability in the process of training a ResNet-based model. The ResNet-50 baseline model is is employed as the primary structure of our model to demonstrate the availability of the STGDEC algorithm. Experimental results demonstrate that diverse kinds of generated samples effectually improve the person re-ID performance based on the proposed STGDEC. Compare to the second-best method LSRO, the STGDEC achieves better improvements. Considering the strong ability of GANs and their derivative models to generate high-quality samples, we will continuously study how to assign reasonable virtual

labels to diverse kinds of generated samples for the future semi-supervised person re-ID task and apply the proposed algorithm to other fields.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (No. 61973066, 61471110, 61733003), National Key R&D Program of China (No. 2017YFC0805000/5005, 2017YFB1301103), Fundamental Research Funds for the Central Universities (N172608005, N182608004), Advanced Technology Project (No. 41412050202), Natural Science Foundation of Liaoning (No. 20180520040) and the Distinguished Creative Talent Program of Shenyang (RC170490).

References

1. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3908–3916 (2015)
2. Barbosa, I.B., Cristani, M., Caputo, B., Rognhaugen, A., Theoharis, T.: Looking beyond appearances: Synthetic training data for deep cnns in re-identification. Computer Vision and Image Understanding (CVIU) **167**, 50–62 (2018)
3. Chen, D., Yuan, Z., Chen, B., Zheng, N.: Similarity learning with spatial constraints for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1268–1277 (2016)
4. Chen, S., Guo, C., Lai, J.: Deep ranking for person re-identification via joint representation learning. IEEE Transactions on Image Processing (TIP) **25**(5), 2353–2367 (2016)
5. Ding, G., Zhang, S., Khan, S., Tang, Z., Zhang, J., Porikli, F.: Feature affinity based pseudo labeling for semi-supervised person re-identification. arXiv:1805.06118 (2018)
6. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **32**(9), 1627–1645 (2010)
7. Figueira, D., Bazzani, L., Minh, H.Q., Cristani, M., Bernardino, A., Murino, V.: Semi-supervised multi-feature learning for person re-identification. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 111–116 (2013)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414–2423 (2016)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
10. Huang, Y., Xu, J., Wu, Q., Zheng, Z., Zhang, Z., Zhang, J.: Multi-pseudo regularized label for generated data in person re-identification. IEEE Transactions on Image Processing (TIP) **28**(3), 1391–1403 (2019)
11. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1125–1134 (2017)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980 (2014)
13. Lee, D.-H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on Challenges in Representation Learning, ICML, pp. 2 (2013)
14. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 152–159 (2014)
15. Liang, C., Huang, B., Hu, R., Zhang, C., Jing, X., Xiao, J.: A unsupervised person re-identification method using model based representation and ranking. In: Proceedings of the 23rd ACM international conference on Multimedia, pp. 771–774 (2015)
16. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2197–2206 (2015)
17. Liu, M.-Y., Tuzel, O.: Coupled generative adversarial networks. In: Neural Information Processing Systems (NIPS), pp. 469–477 (2016)
18. Liu, X., Song, M., Tao, D., Zhou. X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3550–3557 (2014)
19. Liu, Y., Song, G., Shao, J., Jin, X., Wang, X.: Transductive centroid projection for semi-supervised large-scale recognition. In: European Conference on Computer Vision (ECCV), pp. 70–86 (2018)
20. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bags of tricks and a strong baseline for deep person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 0–0 (2019)
21. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Learning to rank in person re-identification with metric ensembles. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1846–1855 (2015)
22. Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., Tian, Y.: Unsupervised cross-dataset transfer learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1306–1315 (2016)
23. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arxiv:1511.06434 (2016)
24. 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 (ECCV), pp. 17–35 (2016)
25. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Neural Information Processing Systems (NIPS), pp. 2234–2242 (2016)
26. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)
27. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. arxiv:1611.02200 (2016)
28. Tao, D., Guo, Y., Song, M., Li, Y., Yu, Z., Tang, Y.Y.: Person re-identification by dual-regularized KISS metric learning. IEEE Transactions on Image Processing (TIP) **25**(6), 2726–2738 (2016)

29. Ustinova, E., Ganin, Y., Lempitsky, V.: Multi bilinear convolutional neural networks for person re-identification. arXiv:1512.05300 (2015)
30. Varior, R.R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification. In: The European Conference on Computer Vision (ECCV), pp. 791–808 (2016)
31. Varior, R.R., Shuai, B., Lu, J., Xu, D., Wang, G.: A siamese long short-term memory architecture for human re-identification. In: The European Conference on Computer Vision (ECCV), pp. 135–153 (2016)
32. Vedaldi, A., Lenc, K.: Matconvnet: Convolutional neural networks for matlab. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 689–692 (2015)
33. Wu, L., Shen, C., van den Hengel, A.: Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. Pattern Recognition (PR) **65**, 238–250 (2017)
34. Xin, X., Wang, J., Xie, R., Zhou, S., Huang, W., Zheng N.: Semi-supervised person re-identification using multi-view clustering. Pattern Recognition (PR) **88**, 285–297 (2019)
35. Yang, X., Wang, M., Hong, R., Tian, Q., Rui, Y.: Enhancing person re-identification in a self-trained subspace. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) **13**(3), 27:1–27:23 (2017)
36. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 34–39 (2014)
37. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1239–1248 (2016)
38. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: IEEE International Conference on Computer Vision (ICCV), pp. 1116–1124 (2015)
39. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: Past, present and future. arXiv:1610.02984 (2016)
40. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: IEEE International Conference on Computer Vision (ICCV), pp. 3774–3782 (2017)
41. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5157–5166 (2018)
42. Zhu. J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251 (2017)
43. Zhao, H., Tian, M., Sun, S., Shao, J., Yan, J., Yi, S., Wang, X., Tang, X.: Spindle net: person re-identification with human body region guided feature decomposition and fusion. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 907–915 (2017)

References