IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON ADVANCED OPTICAL IMAGING FOR EXTREME ENVIRONMENTS

Received June 25, 2019, accepted July 17, 2019, date of publication July 24, 2019, date of current version August 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930865

Double-Domain Imaging and Adaption for Person Re-Identification

SHUREN ZHOU[1,2], PENG LUO[1,2], DEEPAK KUMAR JAIN[3], XIANGYUAN LAN[4], AND YUDONG ZHANG[5]

[1]Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology, Changsha 410114, China
[2]School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China
[3]Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[4]Department of Computer Science, Hong Kong Bapist University, Hong Kong
[5]Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K.

Corresponding author: Shuren Zhou (zsr@csust.edu.cn)

This work was supported in part by the Scientific Research Fund of Hunan Provincial Education Department of China under Project 17A007, in part by the Teaching Reform and Research Project of Hunan Province of China under Project JG1615, in part by the Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing and the Key Laboratory of Industrial IoT and Networked Control, Ministry of Education, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China, and in part by Hong Kong Baptist University Tier 1 Grant.

ABSTRACT Person re-identification (re-ID) performance has been significantly boosted in recent works, but re-ID model trained on one dataset usually cannot work effectively on another one. To address this problem, we proposed a novel framework—double-domain translation generative adversarial network (DTGAN) that can train images between two domains and generalize the model trained on one domain well to another domain. We divide this paper into two steps. First, the source images are translated in an unsupervised manner, and the translated images retain the style of target images and the ID labels in the source domain. Second, the translated images are used as the training data for supervised feature learning. Besides, with the purpose of moderating the influence of noise, we employ the strategy of label smoothing regularization (LSR). In our experiments, we observe that the images generated by the DTGAN are of high quality and more appropriate for domain adaption. In addition, the re-ID accuracy of the DTGAN is competitive to the state-of-the-art methods on Market-1501 and DukeMTMC-reID.

INDEX TERMS Generative adversarial network, domain adaption, person re-identification, deep learning.

I. INTRODUCTION

Person re-identification (re-ID) [2] recently gains its popularity in the field of computer vision. It is a significant task that aims at retrieving images of a person from a large-scale gallery collected by multiple cameras, given a query of person-of-interest. Re-ID has been a very challenging and demanding task because the appearance of a person can be dramatically changed across views and cameras and the lighting, viewpoint, occlusion, and body configuration [28], [39] may change every time. We train a re-ID model on one dataset and it works effectively when we test it on this dataset, but when we apply the model to other datasets, the accuracy of re-ID often falls drastically because of a domain bias. As a consequence, the expensive labels of the source dataset may be wasted. Moreover, the problem may bring an issue that presently fully supervised trained re-ID model cannot be made full use in real world scenes.

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

Unsupervised domain adaption (UDA) is a typical strategy to address the problem. However, this series of methods has a main drawback when applied to re-ID. These methods suppose that the classes of source and target domain are exactly same, but this supposition is not applicable to re-ID, as individual re-ID dataset usually contains absolutely different classes. In the field of domain adaption, domain translation [10], [13], [23] is gaining increasingly popularity. Reference [10], [16], [17] and [36] use style transfer methods.

In our work, we employ CycleGAN [10] referring to the settings in [9]. However, in the task of re-ID, if we only use CycleGAN to translate images between two domains, the ID label information of the persons in source domain will be lost. Therefore, we should preserve the ID information when we perform image-to-image translation for re-ID task. Compared to CycleGAN, we add identity mapping loss and identity preserving loss to CycleGAN to ensure that an image should contain the same ID information before and after translation. Besides, because the source dataset contains entirely different ID information from the target dataset, the translated image should contain distinct ID information from any target images. Compared to SPGAN [9], we adopt different loss functions to preserve ID information and add LSR strategy to eliminate the influence of noise.

In this paper, we propose the double-domain translation generative adversarial network (DTGAN), an unsupervised domain adaption method to translate images from source to target domain and apply the translated images to supervised feature learning [45]. The images generated by DTGAN are supposed to maintain the ID information from source domain. Meanwhile, the translated images possess similar style of target domain. Our work is mainly divided into two steps: first, the source images are translated in an unsupervised manner [44], and the translated images retain the style of target images and the ID labels in source domain; second, the translated images are used as the training data for supervised feature learning. DTGAN is composed of two parts, one part is a Siamese network [2], and the other part is a CycleGAN. In Fig. 2, the CycleGAN learns the generative mappings G and F between source domain and target domain in the process of image-to-image translation. On the basis of CycleGAN, we add ID-discriminative embedding (IDE) [2] as an auxiliary for image-to-image translation so as to improve re-ID accuracy and image quality. With the identity preserving loss, the Siamese network pulls close translated image and its counterpart in source domain. On the contrary, because the translated images should maintain different ID information from any target images, the Siamese network pushes them further away. This satisfies the particularity of re-ID that translation between re-ID datasets needs to preserve the ID information from source domain. In training process, the images are first used to train generator and discriminator of CycleGAN, and then sent to train the Siamese network. With the cooperation of CycleGAN and Siamese network, the translated images not only obtain the style of target domain, but maintain the ID information during translation as well.

With the purpose of moderating the influence of noise [22], [43], we further adopt label smoothing regularization (LSR) in feature learning. LSR assigns less weights to ground truth classes and small value to non-ground truth classes instead of 0. Therefore, LSR can softly distribute data labels and avoid over-fitting risk. On the basis of DTGAN, LSR further improves the re-ID accuracy.

The proposed domain adaption approach, DTGAN, has three main advantages. First, the images generated by DTGAN possess the style of target domain and maintain the latent ID information [34] from source domain. Second, it can be considered as an augmentation scheme smoothing the domain gap. Third, it is an unsupervised re-ID method with good scalability.

Overall, we summarize our contributions as follows,

(1) We introduce a novel framework DTGAN for re-ID (Section III.C). It generates images possessing the style of target domain and preserving the underlying ID information during image-to-image translation.

(2) We apply an LSR method (Section III.D) to translated images to regularize the feature learning process and improve re-ID accuracy.

II. RELATED WORKS

A. GENERATIVE ADVERSARIAL NETWORKS(GAN)

GAN has achieved iconic results in the field of computer vision in recent years, especially in image generation [6], [12]–[15]. GAN was originally introduced by Goodfellow *et al.* [1] and GAN mainly learns two parts: a generator and a discriminator. The discriminator intends to discriminate if the generated images are real or fake whereas the generator learns to generate fake images as true as possible to fool the discriminator. Zheng *et al.* [6] apply the GAN method to the field of re-ID for the first time. This work uses DCGAN [12] as a data augmentation method [30] to generate unlabeled images and adds them to the training in a semi-supervised manner. Recently, GANs have been widely used in image-to-image translation [9], [10], [16], [17], [21], style transfer [7], [19], [20] and domain image generation [13], [18].

B. IMAGE-TO-IMAGE TRANSLATION

Recently, the image translation research has made impressive results. Conditional GAN [21] translates an image and learns the mapping during the translation. For each input image, it requires a relevant image as training data, however, pairs of training images are often difficult to obtain. To solve the hot spots problem [41], a couple generative adversarial network [23] (CoGAN) is proposed to generate pairs of corresponding images and learn joint distribution in two different domains. CycleGAN [10] is proposed to learn the cycle consistency for image-to-image translation. The advantage of this method is that it no longer requires paired images as training data, but uses unpaired samples as input. The drawback is that the color of images generated by CycleGAN may be strange. Our proposed DTGAN is similar to CycleGAN because both of them are image-to-image translation, but the difference is that our work adds some additional restrictions on preserving persons' ID information. It is more suitable than direct applying CycleGAN to re-ID model learning, and we adopt LSR which further improves the re-ID accuracy. Choi *et al.* [25] propose starGAN to solve the image-to-image translation problem on multi-domain. Bousmalis *et al.* [13] use an unsupervised GAN network to translate images to a simulation in target domain. Deng *et al.* [9] propose the SPGAN adding contrastive loss based on CycleGAN, which can identify and process [37] the ID information in the translation process. It has the same effect as our work, but our work adopts different loss functions and adds LSR on the translated samples to prevent over-fitting problems in training, so that our work achieves better results in re-ID.

C. UNSUPERVISED PERSON RE-IDENTIFICATION

Unsupervised re-ID is aimed at addressing the problem of extensibility of existing re-ID methods. Most re-ID

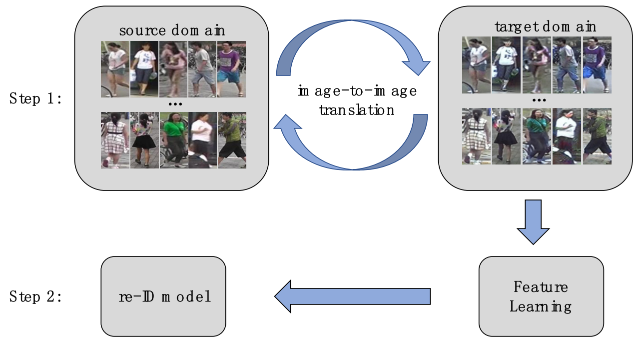

FIGURE 1. **The pipeline of DTGAN consists of two steps. In step 1, the images with labels from source domain are translated to target domain using an unsupervised method. In step 2, the re-ID model is trained in a supervised manner with the translated images.**

approaches require a great amount of paired, cross-view label samples as training data, which needs a lot of time and labor costs. Besides, because of the challenge of re-ID problem itself (cross-view change, high similarity between different persons, etc.), there may be some errors in the label itself, which leads to poor scalability of these methods. Lv *et al.* [4] propose an unsupervised incremental learning algorithm, TFusion that learns the persons' spatial-temporal patterns. Reference [31] and [40] directly apply hand-craft features to unsupervised re-ID. These methods focus on feature extraction [24] whereas ignoring the large amount of information contained in the data distribution of the samples. Pumarola *et al.* [8] uses an unsupervised method to generate person images in arbitrary poses. Zhong *et al.* [7] use CycleGAN to transfer the styles between different cameras in the re-ID dataset in an unsupervised manner.

Unlike upper approaches which primarily focus on improving image quality, our work can not only generate images of high quality, but save the latent ID information during the image-to-image translation as well.

III. METHODOLOGY

We first briefly review the baseline and CycleGAN in Section III.A and Section III.B respectively. In Section III.C, we will introduce the DTGAN method in detail, and in Section III.D, we will introduce the training strategy and LSR.

A. BASELINE OVERVIEW

Given two different datasets from source domain and target domain, respectively. We serve the purpose of training a re-ID model using the labeled source images that works well in target dataset. In this paper, we divide the baseline into two steps. The first step is the image translation from source to target domain. The second step is to extract features from translated images for re-ID model training. The pipeline of DTGAN is shown as Fig. 1.

Step1: Image-to-image translation. In this step, we use the generative mapping function G:A→B to translate the annotated dataset from source domain A to target domain B, so we generate a labeled dataset G(A) with the style of target domain. Similarly, we apply F:B→A to translate images in a opposite direction and generate F(B). We adopt the translation strategy of [36] and [38].

Step2: Feature learning. After we translate the annotated dataset in the first stage, we can use the translated images G(A) to train the re-ID model. We refer to the practice as [2] and add LSR [11] to softly distribute the labels.

In this paper, our main work focuses on improving the first step with the purpose of improving the entire re-ID accuracy. In the first step we used CycleGAN, but we added IDE and identity preserving loss on it. In the second step we add LSR and further improved re-ID accuracy.

B. CYCLEGAN OVERVIEW

CycleGAN serves the purpose of learning two pairs of generator and discriminator, $\{G, D_B\}$ and $\{F, D_A\}$. The generator $G: A \rightarrow B$ maps the source images to target domain and $F: B \rightarrow A$ maps images in the opposite direction. The discriminator D_A and D_B are respectively used to discriminate if the generated images come from another domain. The adversarial loss for discriminator D_B and its related generator G can be expresses as,

$$\mathcal{L}_{B_{adv}}(G, D_B, p_x, p_y) = \mathbb{E}_{y\sim p_y}[(D_B(y)-1)^2] + \mathbb{E}_{x\sim p_x}[(D_B(G(x)))^2], \tag{1}$$

where p_x denote the distribution in source domain and p_y denote the distribution in target domain. The adversarial loss for discriminator D_A and its corresponding generator F can be written as,

$$\mathcal{L}_{Aadv}(G, D_A, p_y, p_x) = \mathbb{E}_{x\sim p_x}[(D_A(x)-1)^2] + \mathbb{E}_{y\sim p_y}[(D_A(G(y)))^2]. \tag{2}$$

However, only two separate pairs of generator and discriminator loss functions are not enough for image-to-image translation. CycleGAN proposes a cycle consistent loss that combines two pairs of generator and discriminator. Forcing $F(G(x)) \approx x$ and $G(F(y)) \approx y$ so that every picture passes a cycle mapping for reconstruction. Cycle-consistent loss is,

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x\sim p_x}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_y}[\|G(F(y)) - y\|_1]. \tag{3}$$

In addition to adversarial loss, cycle-consistent loss, we also added target domain identity constraint [10] which helps regularizing generator to learn the identity mapping when the real samples from target dataset are set to be input images. Therefore, the identity mapping loss can be written as,

$$\mathcal{L}_{ide}\left(G, F, p_x, p_y\right) = \mathbb{E}_{x\sim p_x}\left[\|F(x) - x\|_1\right] + \mathbb{E}_{y\sim p_y}\left[\|G(y) - y\|_1\right]. \tag{4}$$

As mentioned in [10], the color of the images generated from source domain may change without L_{ide} (Fig. 3(b)). This is confirmed by our later experiments.

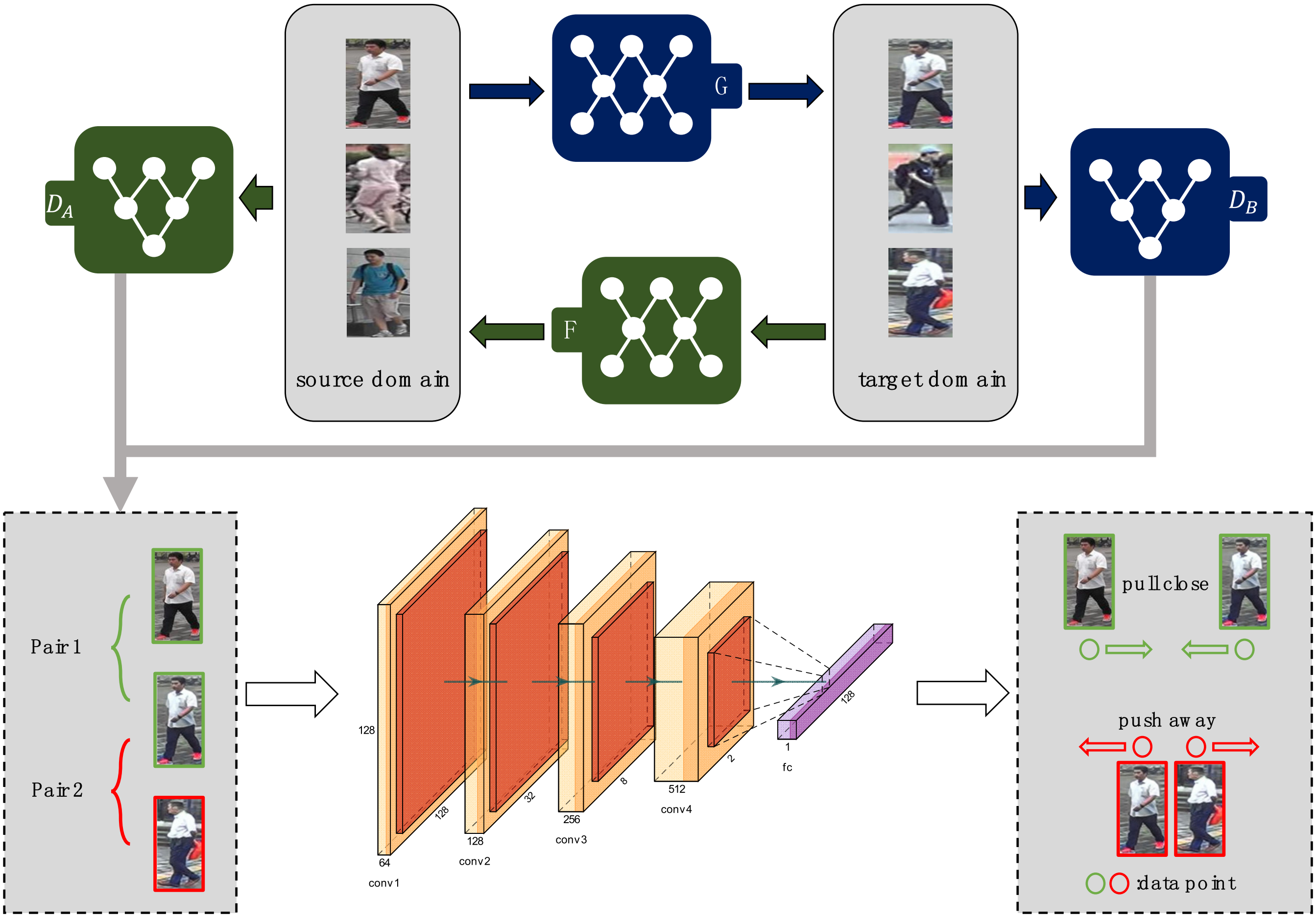

FIGURE 2. The proposed DTGAN is composed of two parts: a CycleGAN and a Siamese network. The CycleGAN learns the generative mappings G and F in the process of image-to-image translation. The Siamese network pulls close the images of the same ID and pushes away the images of different IDs. The blocks in yellow are convolution layers, the blocks in red are max pooling layers and the block in purple is fully connected layer. During training, the training images are first used to train the generators, then the discriminators, and finally sent to train the Siamese network.

C. DTGAN

After the translation between source domain and target domain using CycleGAN and IDE, we found that using only CycleGAN and IDE is not enough for re-ID tasks. The ID information should be utilized, so we proposed DTGAN. In DTGAN, we can not only translate the images, but also preserve the relevant ID information in the process. In DTGAN, we combine Siamese network and CycleGAN together, as shown in Fig. 2. CycleGAN learns the mapping between source domain and target domain, whereas Siamese network pulls close pairs of images containing the same person ID information and pushes away the distance of different persons' image pairs. During training process, the images are first used to update the generator and discriminator of CycleGAN, and then sent to train the Siamese network.

1) IDENTITY PRESERVING LOSS FUNCTION

The Siamese network is trained with the following loss from [29],

$$\mathcal{L}_{ip}\left(C, x_1, x_2\right) = (C)\,\frac{2}{Q}\left(d\left(x_1, x_2\right)\right)^2 + (1 - C)\,2Qe^{-\frac{2.77}{Q}d\left(x_1, x_2\right)}, \quad (5)$$

where x_1 and x_2 represent a pair of input vectors, and $d(x_1, x_2)$ denotes the Euclidean distance of the two input person images: $d\left(x_1, x_2\right) = \|x_1 - x_2\|_2$. C denotes the ground-truth label, where $C = 1$ if x_1 and x_2 form a positive pair; $C = 0$ if x_1 and x_2 form a negative pair. Q denotes a constant.

2) POSITIVE AND NEGATIVE IMAGE PAIR DEFINITION

The proposed DTGAN requires positive and negative pairs [27] as shown in (5). Assuming sample $\mathcal{X}_A$ comes from source domain and $\mathcal{X}_B$ comes from target domain, for the generator G and F, we define 1)$\mathcal{X}_A$ and $G(\mathcal{X}_A)$, 2) $\mathcal{X}_B$ and $G(\mathcal{X}_B)$ as the positive image pairs. In a positive pair, two samples have the same ID information. The difference is that one contains the style of source domain, and the other contains the style of target domain. When training the Siamese network, we pull close the distance of the positive pairs. At the same time, we define two negative pairs: 1) $\mathcal{X}_B$ and $G(\mathcal{X}_A)$, 2)$\mathcal{X}_A$ and $G(\mathcal{X}_B)$. We introduced the two datasets in re-ID with different ID information, the images translated from source domain should have different ID information than any images in target domain. So, when training the Siamese network, we push the distance of the negative

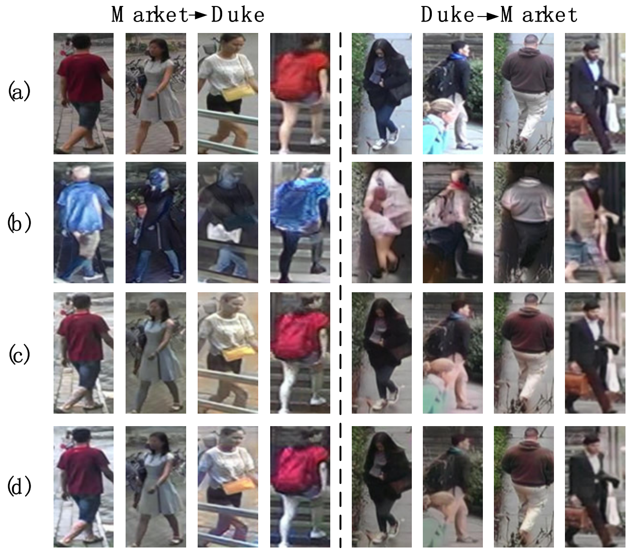

FIGURE 3. Sample images of image-to-image translation between Market-1501 and DukeMTMC-reID. The left side is translating images from Market-1501 to DukeMTMC-reID, and the right side is translating images from DukeMTMC-reID to Market-1501. (a) original dataset images. (b) translated images by CycleGAN. (c) translated images by CycleGAN +L_{ide}. (d) translated images by DTGAN.

pairs away. Some positive image pairs can be shown as (a) and (d) in Fig. 3.

3) THE OVERALL TRAINING OBJECTIVE

The mentioned adversarial loss, cycle-consistent loss, identity preserving loss above jointly work for learning a Siamese network and preserving ID information. The overall DTGAN loss function can be written as,

$$\mathcal{L}_{\rm dt} = \mathcal{L}_{\rm Aadv} + \mathcal{L}_{\rm Badv} + \lambda_1\mathcal{L}_{\rm cyc} + \lambda_2\mathcal{L}_{\rm ide} + \lambda_3\mathcal{L}_{\rm ip}, \tag{6}$$

where λ_1, λ_2, and λ_3 regulate the relevant proportion of the four objectives. We obtain the optimal value of λ_3 in the following experiment (Fig. 6). $\mathcal{L}_{Aadv}$, $\mathcal{L}_{Badv}$, and $\mathcal{L}_{cyc}$ come from CycleGAN formulation [10]. The identity preserving loss $\mathcal{L}_{ip}$ is used to train the Siamese network, which is a new constraint on the system.

D. FEATURE LEARNING WITH LSR

Once we obtained the translated images and associated labels, we can start using these translated samples for feature learning and training the re-ID model in a supervised manner. We employ the baseline ID-discriminative Embedding (IDE) in [2] for network training. We adopt ResNet50 [26] as backbone and refer to the practice in [2]. Besides, we modify the last 1000-dimensional fully-connected layer to the number of the dataset identities.

1) TRAINING WITH LSR

Although the translated images can effectively transform to the style of target images and retain the ID information to promote our re-ID effect, the translation process still produces noise, which may affect the accuracy of re-ID. The generation of noise can be attributed to two aspects. On the one hand, CycleGAN cannot fully learn the mapping relationship between two domains, so that errors may occur during the translation process; On the other hand, because of occlusion and detection errors, there is already some noise in the sample. Translating these noisy samples may generate more noise samples. For the sake of addressing this problem, the proposed DTGAN utilizes label smoothing regularization applying to images to softly distribute these labels and avoid over-fitting risk. LSR assigns less weights to ground truth classes and small value to non-ground truth classes instead of 0. This strategy avoids our excessive tendency to ground truth, which decreases the possibility of over-fitting. LSR is commonly used on cross-entropy loss, and it is written as,

$$\mathrm{L} = -\sum\nolimits_{k=1}^{K} \log\left(p\left(k\right)\right) q\left(k\right), \tag{7}$$

where k in {1, 2, ..., K} corresponds to a class in the dataset and K is the number of classes. $p\left(k\right) \epsilon \left[0, 1\right]$ denotes the predicted probability of the samples belonging to class k and $q\left(k\right)$ denotes the ground truth distribution. Let y be the ground truth label, $q\left(k\right)$ could be expressed as,

$$\mathrm{q}\left(\mathrm{k}\right) = \begin{cases} 0 & k \neq y \\ 1 & k = y. \end{cases} \tag{8}$$

At this point, we have to reduce the loss of (7), which is equivalent to increasing the proportion of ground truth, so if the 0 term is not taken into consideration in (7), the formulation can be rewritten as,

$$\mathrm{L} = -\log\left(\mathrm{p}\left(\mathrm{y}\right)\right). \tag{9}$$

However, only considering the ground truth may cause the problem of over-fitting. Therefore, the LSR takes non-ground-truth into consideration. The ground truth label distribution can be re-defined as,

$$\mathrm{q}_{LSR}\left(k\right) = \begin{cases} \dfrac{\varepsilon}{K} & k \neq y \\ 1-\varepsilon+\dfrac{\varepsilon}{K} & k = y, \end{cases} \tag{10}$$

where $\varepsilon \epsilon \left[0, 1\right]$, if $\varepsilon = 0$, (10) will be reduced to (8). If ε is set as an overly large value, then the model may be difficult to predict the ground truth label. Therefore, ε is usually set to 0.1. At this time, loss function can be rewritten as,

$$\mathcal{L}_{LSR} = -\left(1-\varepsilon\right)\log\left(p\left(y\right)\right) - \frac{\varepsilon}{K}\sum\nolimits_{k=1}^{K} \log\left(p\left(k\right)\right) \tag{11}$$

2) DISCUSSION

Label Smoothing Regularization for Outliers (LSRO) [6] was proposed to train a re-ID model with the unlabeled images generated by DCGAN [12].It assigns a uniform distribution over the unlabeled images. There are two differences between LSRO and our proposed work. First, LSRO is applied when the samples do not contain labels, which is a semi-supervised method. We train the re-ID model with labels, so we apply LSR to our work. Second, the images generated by DTGAN

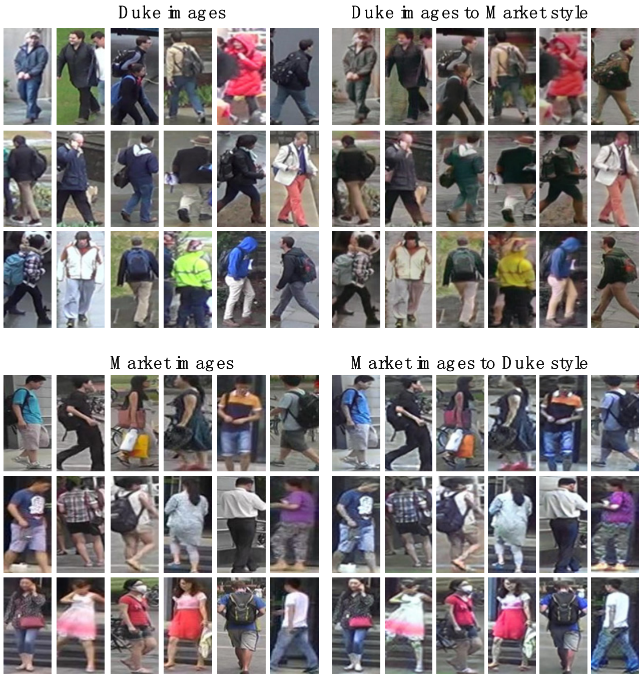

FIGURE 4. Examples of the style transfer. The upper left is images from DukeMTMC-reID, and the upper right is DukeMTMC-reID images translated to Market-1501 style. The lower left is images from Market-1501, and the lower right is Market-1501 translated to DukeMTMC-reID style.

maintain the main characteristics of source domain and contain the style of the target domain.

IV. EXPERIMENT

A. DATASET

In this paper, we test and verify our proposed DTGAN method on Market-1501 [31] and DukeMTMC-reID [6], [32]. Market-1501 collects 12,936 images from 751 identities for training and 19,732 images from 750 identities for testing. DukeMTMC-reID is a re-ID version of the DukeMTMC dataset, which includes 1,402 identities, 16,522 images from 702 identities for training, 2,228 query images from the other 702 identities for testing, and 17,661 database images. We employ mean average precision (mAP) and rank-1 accuracy to the two large-scale datasets to evaluate the performance of our method for re-ID. Fig. 4 shows some image samples of Market-1501 and DukeMTMC-reID.

B. EXPERIMENT SETTINGS

1) TARGET-SOURCE DOMAIN TRANSLATION

We employ the deep learning framework TensorFlow [33] to train DTGAN. Noted that the image translation is an unsupervised process, therefore, we do not employ any annotation of identity information during the training process. In all experiment, some of our parameters are set as follows, for $\lambda_1 = 10$, $\lambda_2 = 5$, $\lambda_3 = 2$ and $Q = 2$ in (6). We set the learning rate as 0.0002, and the number of epochs is 5. ε in LSR is set as 0.1. In the test phase, we use G for mapping images from Market-1501 to DukeMTMC-reID and F for mapping images from DukeMTMC-reID to Market-1501. The translated images are used to train re-ID models.

To train CycleGAN, we refer to the training strategy in [10], and when training Siamese network, we refer to the training strategy in [9], which consists of 4 convolution layers, 4 maximum pooling layers, and 1 fully connected layer.

2) FEATURE LEARNING

We refer to the training strategy in [2] and train a classification network (IDE). ResNet-50 [26] is utilized as the backbone to fine-tune the training images. Specially, we modify the output of the last fully connected layer to the number of identities of the dataset. For example, we changed the output of the fully connected layer to 702 on DukeMTMC-reID and changed to 751 on Market-1501. We utilize the mini-batch stochastic gradient descent to train re-ID model on a TITAN X GPU.

C. EVALUATION

1) SUPERVISED LEARNING COMPARED TO DIRECT TRANSFER

In Table 1, it can be apparently observed that when we use a model trained on source domain to classify target domain images, the accuracy of the model drops dramatically. For example, as the experiment reported in [35], the ResNet-50 model trained on the DukeMTMC-reID achieves 63.4% in rank-1 accuracy, however, when we test the model on Market-1501, the accuracy falls sharply to 33.1%. It can be also observed that the model trained on Market-1501 attains 76.2% in rank-1 accuracy and when we test it on DukeMTMC-reID, the accuracy drops drastically to 43.1%.

TABLE 1. Comparison of different methods applied to target domain.

Methods	DukeMTMC-reID				Market-1501			
	rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
Supervised Learning	63.4	78.9	83.3	42.6	76.2	89.5	93.3	53.1
Direct Transfer	33.1	49.3	55.6	16.7	43.1	60.8	68.1	17.0
CycleGAN	38.1	54.4	60.5	19.6	45.6	68.8	71.3	19.1
CycleGAN + L_{ide}	38.5	54.6	60.8	19.9	48.1	66.2	72.7	20.7
DTGAN	41.1	58.9	64.9	22.6	54.0	72.5	79.3	23.6
DTGAN+LSR	44.9	61.0	66.2	25.0	56.7	73.3	79.2	25.8

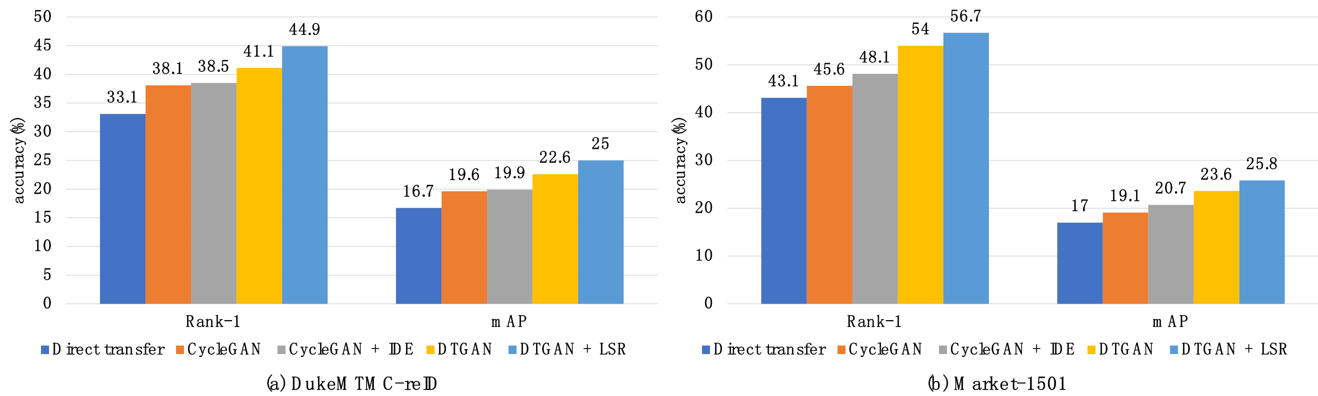

FIGURE 5. Visualization of the comparison of different settings testing on target domain. (a) Comparison of different methods testing on DukeMTMC-reID. (b) Comparison of different methods testing on Market-1501.

The main reason is the domain bias and we propose DTGAN to address this problem.

2) THE DOMAIN ADAPTION BASELINE USING CYCLEGAN

In the baseline method (Section III.A), we perform the image-to-image translation following two steps. As displayed in Table 1, there is an obvious improvement between direct transfer and the baseline framework. Compared to direct transfer, the performance of CycleGAN baseline obtains +2.9% and +5.0% improvement in mAP and rank-1 accuracy on DukeMTMC-reID. The performance gains +2.1% in mAP and +2.5% in rank-1 accuracy on Market-1501. The result shows the effectiveness of the baseline than direct transfer method.

3) THE INFLUENCE OF TARGET DOMAIN IDENTITY CONSTRAINT

As shown in Table 1, on DukeMTMC-reID, CycleGAN $+\mathcal{L}_{ide}$ attains analogous mAP and rank-1 accuracy compared to CycleGAN. However, CycleGAN $+\mathcal{L}_{ide}$ gains +1.6% in mAP and +2.5% in rank-1 accuracy on Market-1501. The reason why testing on Market-1501 achieves better results is that Market-1501 dataset is more inclusive of the inter-camera variance. The target identity constraint does not largely improve the re-ID accuracy, but as shown in Fig. 3, this loss can significantly improve color distortion of the images. Therefore, we adopt the identity constraint as an auxiliary tool. The experiment shows the effectiveness of the identity loss for re-ID.

4) THE EFFECTIVENESS OF THE PROPOSED DTGAN

As shown in Table 1, the proposed DTGAN achieves higher re-ID performance than previous methods. Compared with CycleGAN $+\mathcal{L}_{ide}$, DTGAN obtains +2.7% and +2.6% in mAP and rank-1 accuracy on DukeMTMC-reID, respectively. Similarly, DTGAN gains +2.9% in mAP and +5.9% in rank-1 accuracy on Market-1501. The proposed DTGAN can preserve the latent ID information of the source images and thus improves the re-ID performance. The consistent growth of the re-ID accuracy proves that the proposed DTGAN generates images suitable for the model training in the target domain. Translated images by DTGAN are displayed in Fig. 4.

5) STUDY OF THE SENSITIVITY OF THE PARAMETER –

We evaluate the influence of the key parameter λ_3 in (6). λ_3 controls the weight of the identity preserving loss $\mathcal{L}_{ip}$. As shown in Fig. 6, the identity preserving loss achieves improvement when compared to $\lambda_3 = 0$. However, a larger λ_3 does not bring better results. Therefore, $\lambda_3 = 2$ achieves the best result.

6) LSR FURTHER IMPROVES THE RE-ID PERFORMANCE

In our experiments, we apply LSR so as to softly distribute labels and avoid over-fitting risk. We show in the Table 1, using LSR achieves higher performance. On DukeMTMC-reID, the performance of adding LSR gains +2.4% and +3.8% improvement in mAP and rank-1 accuracy than DTGAN. The performance improvement is +2.2% and +2.7% in mAP and rank-1 accuracy when testing on Market-1501. The LSR method reduces the focus on labels and prevents the network from over-fitting, which explains why applying LSR obtains a superior performance. The visualization of the effect of these settings is shown as Fig. 5.

7) DISCUSSION

As shown in Fig. 4, the style of the translated images has changed close to the target domain. It is important that the translated images are visually closer to the style of target images. However, whether the model trained with the translated images achieves better performance in target domain than direct transfer is of greater importance. As shown in Table 1, on DukeMTMC-reID, DTGAN obtains +5.9% and +8% than direct transfer in mAP and rank-1 accuracy, respectively. On Market-1501, DTGAN obtains +6.6% and +8.9% than direct transfer in mAP and rank-1 accuracy, respectively. The result proves that the model brings the source images closer to target domain and the model trained

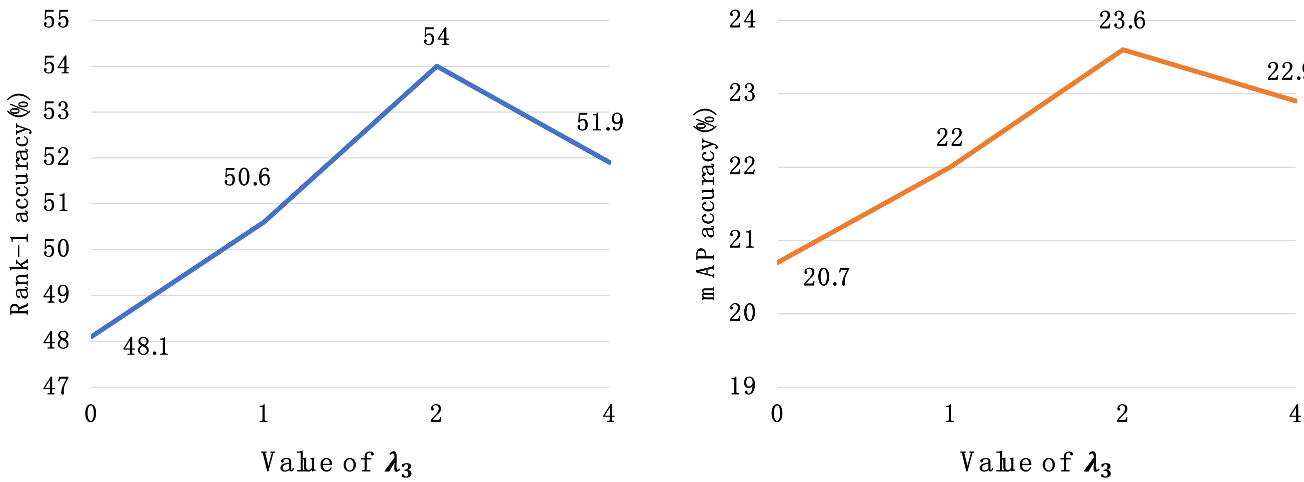

FIGURE 6. Influence of λ_3 on re-ID accuracy. A larger λ_3 means larger weight of identity preserving loss.

TABLE 2. Comparison with the state-of-the-art methods on market-1501.

Methods	Market-1501			
	Rank-1	Rank-5	Rank-10	mAP
LOMO [40]	27.2	41.6	49.1	8.0
Bow [31]	35.8	52.4	60.3	14.8
UMDL [42]	34.5	52.6	59.6	12.4
PUL [35]	45.5	60.7	66.7	20.5
Direct transfer	43.1	60.8	68.1	17.0
SPGAN [9]	51.5	70.1	76.8	22.8
DTGAN	54.0	72.5	79.3	23.6
DTGAN+LSR	56.7	73.3	79.2	25.8

TABLE 3. Comparison with the state-of-the-art methods on dukemtmc-reid.

Methods	DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP
LOMO [40]	12.3	21.3	26.6	4.8
Bow [31]	17.1	28.8	34.9	8.3
UMDL [42]	18.5	31.4	37.6	7.3
PUL [35]	30.0	43.4	48.5	16.4
Direct transfer	33.1	49.3	55.6	16.7
SPGAN [9]	41.1	56.6	63.0	22.3
DTGAN	41.1	58.9	64.9	22.6
DTGAN+LSR	44.9	61.0	66.2	25.0

with translated images is more appropriate than the model trained with original source images.

D. COMPARISON WITH THE STATE-OF-THE-ART METHODS

As shown in Table 2, on Market-1501, we compare our DTGAN approach with local maximum occurrence (LOMO) [40] and Bag-of-Words (BoW) [31] at first. Unlike other domain adaption methods, they are not trained and directly tested on dataset, and obviously obtain poor performance. We also compare our methods to other unsupervised methods, including UMDL [42], the Progressive Unsupervised Learning (PUL) [35] and Similarity Preserving Generative Adversarial Net (SPGAN) [9]. We achieve better performance than these existing methods, where mAP = 23.6% and rank-1 accuracy = 54.0%. It is +0.8% in mAP and +2.5% in rank-1 accuracy beyond the second-best method. After applying LSR, we arrived at mAP = 25.8% and rank-1 = 56.7%, which is +3% in mAP and +5.2% in rank-1 accuracy than SPGAN, respectively. These comparisons show that our DTGAN method is competitive to the state-of-the-art unsupervised methods.

As shown in Table 3, on DukeMTMC-reID, we make comparisons between DTGAN and LOMO [40], BoW [31], UMDL [42], PUL [35], SPGAN [9] methods. The proposed method obtains the excellent performance of mAP = 22.6% and rank-1 accuracy = 41.1%. After applying LSR, we obtain the result of mAP = 25.0% and rank-1 accuracy = 44.9%, which gains the improvement of +3.8% and +2.7% in rank-1 accuracy and mAP than SPGAN. These comparisons prove the superiority of DTGAN on DukeMTMC-reID.

Discussion: As shown in our experiments, we prove the superiority and effectiveness of the proposed DTGAN to direct transfer and CycleGAN method. Moreover, we show the effectiveness of added identity preserving loss and LSR. Both our method and SPGAN adopt CycleGAN to translate images between two domains. But compared to SPGAN, our method adopts different loss functions to preserve the identity information, and we add LSR strategy to moderate the noise in feature learning process. Therefore, our method achieves better results. Besides, this method achieves better results compared with the state-of-the-art methods on Market-1501 and DukeMTMC-reID.

V. CONCLUSION

This paper mainly focuses on domain adaption in re-ID. When the model trained on one dataset is directly tested on another dataset, re-ID accuracy usually drops drastically owing to the dataset bias. To address this problem, we introduce the DTGAN. The proposed DTGAN is composed of two parts: a CycleGAN and a Siamese network. The DTGAN baseline is divided into two steps, 1) unsupervised image-to-image translation, 2) supervised re-ID model learning with translated images. The proposed method can not only translate images between different double domains but preserve the latent ID information as well, which satisfies the requirement of re-ID. Besides, we adopt the LSR strategy in order to moderate the influence of noise, which further improves the re-ID accuracy. The experiments demonstrate the effectiveness and superiority of our proposed DTGAN method on Market-1501 and DukeMTMC-reID.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[2] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," 2016, *arXiv:1610.02984*. [Online]. Available: https://arxiv.org/abs/1610.02984

[3] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person reidentification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, p. 13, 2017.

[4] J. Lv, W. Chen, Q. Li, and C. Yang, "Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7948–7956.

[5] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 79–88.

[6] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3754–3762.

[7] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camera style adaptation for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5157–5166.

[8] A. Pumarola, A. Agudo, A. Sanfeliu, and F. Moreno-Noguer, "Unsupervised person image synthesis in arbitrary poses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8620–8628.

[9] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 994–1003.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1511.06434

[13] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3722–3731.

[14] T. Kaneko, K. Hiramatsu, and K. Kashino, "Generative attribute controller with conditional filtered generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6089–6098.

[15] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 5907–5915.

[16] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2849–2857.

[17] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1857–1865.

[18] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," 2016, *arXiv:1611.02200*. [Online]. Available: https://arxiv.org/abs/1611.02200

[19] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2414–2423.

[20] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 9906, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016.

[21] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[22] D. Zeng, Y. Dai, F. Li, R. Sherratt, and J. Wang, "Adversarial learning for distant supervised relation extraction," *Comput., Mater. Continua*, vol. 55, no. 1, pp. 121–136, 2018.

[23] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.

[24] Y. Gui and G. Zeng, "Joint learning of visual and spatial features for edit propagation from a single image," *Vis. Comput.*, 2019, pp. 1–14.

[25] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[27] W. Li, H. Liu, J. Wang, L. Xiang, and Y. Yang, "An improved linear kernel for complementary maximal strip recovery: Simpler and smaller," *Theoretical Computer Science*, to be published.

[28] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 34–39.

[29] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. CVPR*, Jun. 2005, pp. 539–546.

[30] Y. Chen, J. Xiong, W. Xu, and J. Zuo, "A novel online incremental and decremental learning algorithm based on variable support vector machine," *Cluster Comput.*, vol. 22, pp. 1–11, Jan. 2018. doi: 10.1007/s10586-018-1772-4.

[31] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.

[32] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Computer Vision—ECCV 2016 Workshops* (Lecture Notes in Computer Science), vol. 9914, G. Hua and H. Jégou, Eds. Cham, Switzerland: Springer, 2016.

[33] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.

[34] M. Long, F. Peng, and H.-Y. Li, "Separable reversible data hiding and encryption for HEVC video," *J. Real-Time Image Process.*, vol. 14, no. 1, pp. 171–182, 2018.

[35] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multi Media Comput., Commun., Appl.*, vol. 14, no. 4, p. 83, Nov. 2018.

[36] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 700–708.

[37] Q. Tang, K. Yang, D. Zhou, Y. Luo, and F. Yu, "A real-time dynamic pricing algorithm for smart grid with unstable energy providers and malicious users," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 554–562, Aug. 2016.

[38] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. A. Efros, and T. Darrell, "CyCADA: Cycle-consistent adversarial domain adaptation," 2017, *arXiv:1711.03213*. [Online]. Available: https://arxiv.org/abs/1711.03213

[39] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools Appl.*, vol. 77, pp. 1–21, Aug. 2018. doi: 10.1007/s11042-018-6562-81.

[40] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2197–2206.

[41] J. Wang, Y. Gao, X. Yin, F. Li, and H.-J. Kim, "An enhanced PEGASIS algorithm with mobile sink support for wireless sensor networks," *Wireless Commun. Mobile Comput.*, to be published. doi: 10.1155/2018/9472075.

[42] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, " Unsupervised cross-dataset transfer learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1306–1315.

[43] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[44] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[45] Y. Wu, B. Jiang, and N. Lu, "A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

SHUREN ZHOU received the B.S. degree from the Changsha University of Science and Technology, Changsha, China, in 1999, and the M.S. and Ph.D. degrees from Central South University, China, in 2004 and 2009, respectively. He was a member of the Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology. His research interests include person re-identification, action recognition, image processing, pattern recognition, human pose estimation, and computer vision.

PENG LUO received the B.E. degree in computer science and technology from the Changsha University of Science and Technology, Changsha, China, in 2017, where he is currently pursuing the M.E. degree in computer science. His research interests include person re-identification, action recognition, and object tracking.

DEEPAK KUMAR JAIN received the B.Eng. degree from Rajiv Gandhi Proudyogiki Vishwavidyalaya, India, in 2010, the M.Tech. degree from the Jaypee University of Engineering and Technology, India, in 2012, and the Ph.D. degree from the Institute of Automation, University of Chinese Academy of Sciences, Beijing, China. He was an Assistant Professor with the Institute of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China. He has presented several papers in peer-reviewed conferences and has published numerous studies in science cited journals. His research interests include deep learning, machine learning, pattern recognition, and computer vision. He was an Awardee of the CAS-TWAS Presidential Fellowship, from 2014 to 2018. He was invited as "Foreign Experts" by the Shandong Taian Administration of foreign Expert Affairs.

XIANGYUAN LAN received the B.Eng. degree in computer science and technology from the South China University of Technology, China, in 2012, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2016, where he is currently a Research Assistant Professor. His current research interests include intelligent video surveillance and biometric security.

YUDONG ZHANG received the Ph.D. degree from Southeast University, China, in 2010. He held a postdoctoral position, from 2010 to 2012, and a Research Scientist, from 2012 to 2013, at Research Foundation in Mental Hygiene (RFMH) affiliated with Columbia University, USA. From 2013 to 2017, he was a Professor with Nanjing Normal University, where he was the Director and the Founder of the Advanced Medical Image Processing Group. Since 2017, he has been a Professor (permanent) with the Department of Informatics, University of Leicester, U.K.