



Cross-Dataset Person Re-identification Using Similarity Preserved Generative Adversarial Networks

Jianming Lv^(✉) and Xintong Wang

South China University of Technology, Guangzhou, China
jmlv@scut.edu.cn, w.xintong@mail.scut.edu.cn

Abstract. Person re-identification (Re-ID) aims to match the image frames which contain the same person in the surveillance videos. Most of the Re-ID algorithms conduct supervised training in some small labeled datasets, so directly deploying these trained models to the real-world large camera networks may lead to a poor performance due to underfitting. The significant difference between the source training dataset and the target testing dataset makes it challenging to incrementally optimize the model. To address this challenge, we propose a novel solution by transforming the unlabeled images in the target domain to fit the original classifier by using our proposed similarity preserved generative adversarial networks model, SimPGAN. Specifically, SimPGAN adopts the generative adversarial networks with the cycle consistency constraint to transform the unlabeled images in the target domain to the style of the source domain. Meanwhile, SimPGAN uses the similarity consistency loss, which is measured by a siamese deep convolutional neural network, to preserve the similarity of the transformed images of the same person. Comprehensive experiments based on multiple real surveillance datasets are conducted, and the results show that our algorithm is better than the state-of-the-art cross-dataset unsupervised person Re-ID algorithms.

Keywords: Cross-dataset · Person re-identification
Similarity preserved

1 Introduction

As one of the most important and challenging problems in the field of surveillance video analysis, person re-identification (Re-ID) aims to match the image frames which contain the same person in the surveillance videos. How to extract the view invariant features from the images and design a robust visual classifier to identify the persons is the core challenge of the Re-ID algorithms.

The work described in this paper was supported by the grants from NSFC (No. U1611461), Science and Technology Program of Guangdong Province, China (No. 2016A010101012), and CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, 100190, Beijing, China (No. CASNDST201703).



Fig. 1. Samples in different datasets.

Due to the privacy problem regarding the collection of surveillance videos and the expensive cost of data labeling, most of the proposed Re-ID algorithms [12, 16] conduct supervised learning on small labeled datasets. Directly deploying these trained models to the real-world large-scale camera networks may lead to a poor performance, because the images captured from different camera networks usually have totally different backgrounds, noise distributions, brightness and resolution as shown in Fig. 1. How to incrementally optimize the Re-ID algorithms based on the abundant unlabeled data collected from the target domain is a practical and extremely challenging problem.

To address this problem, some unsupervised algorithms [7, 10] are proposed to extract view invariant features and measure the similarity of pedestrians based on the unlabeled dataset. Without the powerful supervision based on labels, this kind of pure unsupervised learning based algorithms working on one single dataset have a poor performance in most cases. Recently, a cross-dataset unsupervised transfer learning algorithm, named UMDL [11], is proposed to make use of data samples from both labeled source datasets and the unlabeled target dataset to learn the view-invariant feature representation and similarity measurement by the dictionary learning mechanism. UMDL gains much better performance than purely unsupervised algorithms, but is still much weaker than the state-of-the-art supervised algorithms based on the labeled dataset. Most of above algorithms try to incrementally optimize the visual classifier, which is pre-trained in the source dataset, to fit the new data in the target domain. However, without the labels in the target domain, it is hard to fine-tune the classifier to suit the source and target datasets simultaneously, which have diverse feature distributions.

We address this cross-dataset Re-ID challenge in a totally new direction in this paper. Instead of optimizing the classifier to fit the new data, we transform the unlabeled images in the target domain to fit the classifier by using our proposed similarity preserved generative adversarial networks model, SimP-GAN. As shown in Fig. 5 regarding an example from the real datasets, after the

transformation, the features of the images in the target datasets are projected into the features close to the distribution in the source dataset, and are easier to be processed by the visual classifier trained in the source dataset.

The main contributions of this paper are summarized as follows:

- We propose a novel efficient solution, named SimPGAN, to improve the performance of cross-dataset person Re-ID by transforming the unlabeled images in the target domain into the style of the source domain to fit the visual classifier transferred from the source domain.
- SimPGAN adopts the generative adversarial networks with cycle consistency constraint to avoid sharp change of the images after transformation. Meanwhile, SimPGAN uses the similarity consistency loss, which is measured by a siamese deep convolutional neural network, to preserve the similarity of the transformed images of the same person.
- We conduct comprehensive experiments based on real datasets (Market1501 [17], CUHK01 [14], GRID [1]), which show that SimPGAN is better than the state-of-the-art cross-dataset unsupervised transfer learning algorithm [11].

The rest of this paper is organized as follows. Section 2 reviews the related work of Re-ID. Section 3 offers clear definitions of the problem about Re-ID in unlabeled dataset. Section 4 presents our proposed methods. Section 5 evaluates the performance of this system by conducting experiments on real datasets. We conclude the work in Sect. 6.

2 Related Work

Supervised Learning: Most existing person Re-ID models are supervised, and based on either invariant feature learning [8], metric learning [12] or deep learning [16]. However, in the practical deployment of Re-ID algorithms in large-scale camera networks, it is usually costly and unpractical to label the massive online surveillance videos to support supervised learning as mentioned in [11].

Unsupervised Learning: In order to improve the effectiveness of the Re-ID algorithms towards large-scale unlabeled datasets, some unsupervised Re-ID methods [7] are proposed to learn cross-view identity-specific information from unlabeled datasets. However, due to the lack of the knowledge about identity labels, these unsupervised approaches usually yield much weaker performance compared to supervised learning approaches.

Transfer Learning: Recently, some cross-dataset transfer learning algorithms [6, 9–11] are proposed to leverage the Re-ID models pre-trained in other labeled datasets to improve the performance on target dataset. This type of Re-ID algorithms can be classified further into two categories: supervised transfer learning and unsupervised transfer learning according to whether the label information of target dataset is given or not. Specifically, in the **supervised transfer learning** algorithms [6, 9, 10], both of the source and the target datasets are labeled

or have weak labels. [6] is based on a SVM multi-kernel learning transfer strategy, and [10] is based on cross-domain ranking SVMs. [9] adopts multi-task metric learning models. On the other hand, the recently proposed cross-dataset **unsupervised transfer learning** algorithm for Re-ID, UMDL [11], is totally different from above algorithms, and closer to real-world deployment environment where the target dataset is totally unlabeled. UMDL [11] transfers the view-invariant representation of a person's appearance from the source labeled dataset to the unlabeled target dataset by dictionary learning mechanisms, and gains much better performance. Although this kind of cross-dataset transferring algorithms are proved to outperform the purely unsupervised algorithms, they still have a long way to catch up the performance of the supervised algorithms, e.g. in the CUHK01 [14] dataset, UMDL [11] can achieve 27.1% rank-1 accuracy, while the accuracy of the state-of-the-art supervised algorithms [13] can reach 67%.

Generative Adversarial Networks: Generative adversarial networks (GAN) [2] provide a way to learn deep representations without extensively annotated training data. Within the adversarial network, the generative model takes random inputs and tries to generate data samples, while the discriminative model learns to determine whether a sample is drawn from the model distribution or the data distribution. GAN is broadly used for image synthesis, image classification, image-to-image translation and super-resolution and achieves impressive results. Specifically, the image-to-image translation aims to learn the mapping between an input image and an output image using a set of aligned image pairs. More recently, the Pix2Pix framework [5] uses a conditional generative adversarial network to learn a mapping from input to output images by using paired training data. Unlike these prior works, Zhu [19] learns the mapping without paired training examples using the cycle-consistent adversarial networks.

3 Preliminaries

3.1 Problem Definition of Re-ID

Given a surveillance image which contains a target pedestrian, the goal of a person Re-ID algorithm is to retrieve the surveillance videos for the image frames which contain the same person. For clarity of the problem definition, some notations describing Re-ID are introduced in this section.

Each surveillance image containing a pedestrian is denoted as I_j , which is cropped from an image frame of a surveillance video. The ID of the pedestrian in I_j is denoted as $\mathcal{Y}(I_j)$. Given any surveillance image I_j , the person Re-ID problem is to retrieve the images $\{I_k | \mathcal{Y}(I_j) = \mathcal{Y}(I_k)\}$, which contain the same person $\mathcal{Y}(I_j)$.

The traditional strategy of person Re-ID is to train a classifier \mathcal{C} based on visual features to judge whether two given images contain the same person. The output of the classifier is usually the similarity score, which measures the likelihood that the two images contain the same person. The similarity score can be used to rank the image frames to retrieve the Re-ID results.

3.2 Cross-Dataset Person Re-ID

Like most of the traditional person Re-ID algorithms [8], we can conduct supervised learning on some public labeled dataset (denoted as Ω_s below), which is usually of small size, to train a classifier \mathcal{C} . While directly deploying the trained \mathcal{C} to a real-world unlabeled target dataset Ω_t collected from a large-scale camera network, it tends to have a poor performance, due to the significant difference between Ω_s and Ω_t as shown in the Fig. 1.

How to effectively transfer the classifier trained in a small labeled source dataset to another unlabeled target dataset is the fundamental challenging problem addressed in this paper.

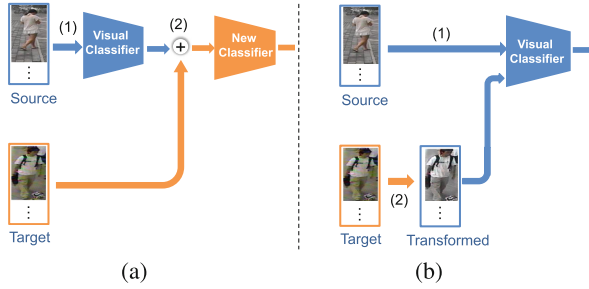


Fig. 2. Two different kind of solutions to perform cross-dataset person Re-ID. (a) Incrementally optimizing the classifier to fit the data in the target domain. Here (1) indicates supervised learning, and (2) indicates incrementally optimization. (b) Transforming the data in the target domain to fit the original classifier. Here (1) indicates supervised learning, and (2) indicates transforming the image into the style of the source domain.

4 Model

4.1 Model Overview

As shown in Fig. 1, the images taken in different camera networks usually have significantly different feature distributions. Thus, direct transferring of a visual classifier, which is pre-trained in a small labeled source dataset, to another unlabeled target dataset may cause a poor performance in most cases. As shown in Fig. 2(a), most of the proposed cross-dataset transfer learning algorithms try to incrementally optimize the classifier based on the unlabeled data from the target dataset. However, without the powerful supervised tuning, the performance of the visual classifier over the target domain usually does not gain significant improvement.

We propose a novel generative model based solution to address the challenge in the opposite way as shown in Fig. 2(b). Instead of incrementally optimizing the classifier to fit the new data, we transform the data in the target dataset to

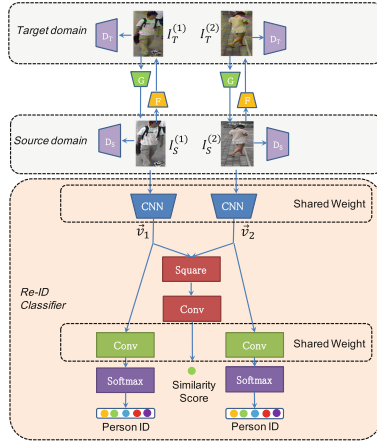


Fig. 3. The SimPGAN model for cross-dataset person Re-ID.

fit the classifier. In order to make the transformation benefit the improvement of the cross-dataset person Re-ID, the image transformation should satisfy the following constraints:

- **Data fitness.** The visual features of the transformed images in the target dataset should fit the feature distribution in the source dataset, and should be more suitable for the visual classifier, which is pre-trained in the source dataset.
- **Similarity preservation.** The similarity of the transformed images, which contain the same person, should be as high as possible.

To address the above constraints, we propose the similarity preserved GAN model, SimPGAN, to conduct the transformation. As shown in Fig. 3, we adopt a siamese deep neural network as the Re-ID classifier, which measures the similarity of two input images and determines the identity of the person in the image. We firstly pre-train this Re-ID classifier in the labeled source dataset, and then we transfer it to the target dataset. Given a pair of images in the target dataset, they are simultaneously transformed into the style of the source dataset using the generative model, and input into the Re-ID classifier to measure their similarity.

To address the ‘data fitness’ constraint, we integrate the cycle consistency loss [19] in the GAN model, which is proved to be very powerful to improve the quality and steadiness of unpaired images translation. Meanwhile, to address the ‘similarity preservation’ constraint, we use the similarity loss from the Re-ID classifier as the high-level semantic signal to fine-tune the generative model to preserve the similarity of the transformed images of the same person. We will introduce the detail of each component in the following sections.

4.2 Siamese CNN Based Re-ID Classifier

As shown in Fig. 3, we select the recently proposed siamese convolutional neural network [18] as the Re-ID classifier, which makes better use of the label information and has a good performance in the large-scale datasets such as Market1501 [17]. The network adopts a siamese scheme including two ImageNet pre-trained CNN modules, which share the same weight parameters and extract visual features from the input images $I_S^{(1)}$ and $I_S^{(2)}$. The CNN module is achieved from the ResNet-50 network [3] by removing its final fully-connected (FC) layer. The outputs of the two CNN modules are flattened into two one-dimensional vectors: \mathbf{v}_1 and \mathbf{v}_2 , which act as the embedding visual feature vectors of the input images.

To measure the matching degree of the input images, their feature vectors \mathbf{v}_1 and \mathbf{v}_2 are fed into the following square layer to conduct subtracting and squaring element-wisely: $\mathbf{v}_s = (\mathbf{v}_1 - \mathbf{v}_2)^2$. Finally, a convolutional layer is used to transform \mathbf{v}_s into the similarity score as:

$$\hat{q} = \text{sigmoid}(\theta_s \circ \mathbf{v}_s) \quad (1)$$

Here θ_s denotes the parameters in the convolutional layer, \circ denotes the convolutional operation, and *sigmoid* indicates the *sigmoid* activation function. By comparing the predicted similarity score with the ground-truth matching result of $I_S^{(1)}$ and $I_S^{(2)}$, we can achieve the **variation loss** as a cross entropy form:

$$L_v(I_S^{(1)}, I_S^{(2)}) = -q \cdot \log(\hat{q}) - (1 - q) \cdot \log(1 - \hat{q}) \quad (2)$$

Here $q = 1$ when $I_S^{(1)}$ and $I_S^{(2)}$ contain the same person. Otherwise, $q = 0$.

Besides predicting the similarity score, the model also predicts the identity of each image in the following steps. Each visual feature vector ($\mathbf{v}_x (x = 0, 1)$) is fed into one convolutional layer to be mapped into an one-dimensional vector with the size K , where K is equal to the total number of the pedestrians in the dataset. Then the following softmax unit is applied to normalize the output as follows:

$$\hat{P}^{(x)} = \text{softmax}(\theta_x \circ \mathbf{v}_x) (x = 1, 2) \quad (3)$$

Here θ_x is the parameter in the convolutional layer and \circ denotes the convolutional operation. The output $\hat{P}^{(x)}$ is used to predict the identity of the person contained in the input image $I_S^{(x)} (x = 1, 2)$. By comparing $\hat{P}^{(x)}$ with the ground-truth identify label, we can achieve the **identification loss** as the cross-entropy form:

$$\begin{aligned} & L_{id}(I_S^{(1)}, I_S^{(2)}) \\ &= \sum_{k=1}^K (-\log \hat{P}_k^{(i)} \cdot P_k^{(i)}) + \sum_{k=1}^K (-\log \hat{P}_k^{(j)} \cdot P_k^{(j)}) \end{aligned} \quad (4)$$

Here $P^{(x)}(x = 1, 2)$ is the identity vector of the input image $I_S^{(x)}$. $P_k^{(x)} = 0$ for all k except $P_t^{(x)} = 1$, where t is ID of the person in the image $I_S^{(x)}$.

While training the classifier in the labeled source dataset, the parameters are optimized by using the final loss function:

$$L_{all} = L_v + L_{id} \quad (5)$$

According to [15], this kind of composite loss makes the classifier more efficient to extract the view-invariant visual features for Re-ID than the single loss function.

4.3 Similarity Preserved GAN

While transferring the Re-ID classifier to the target domain, we propose a siamese GAN architecture to transform each pair of the images in the target domain to the style of the source domain and feed them into the Re-ID classifier as shown in Fig. 3. Similar with the implementation of CycleGAN [19], the model includes two generators performing the cross-dataset image mapping: $G : I_S \rightarrow I_T$ and $F : I_T \rightarrow I_S$, where I_S indicates the source dataset and I_T indicates the target dataset. In addition, the model integrates the adversarial discriminator classifier D_S to discriminate the real images I_S and generated images $F(I_T)$, and D_T to discriminate the real images I_T and generated images $G(I_S)$.

The optimization of G , F , D_S and D_T is based on the combination of the cycle consistency adversarial loss [19] and our proposed similarity consistency loss, which will be introduced in the following sections. As shown in Fig. 3, in each iteration of optimization, the model takes four images as input to calculate the loss: one pair of labeled images $(I_S^{(1)}, I_S^{(2)})$ from the source dataset, and another pair of unlabeled images $(I_T^{(1)}, I_T^{(2)})$ from the target dataset.

Cycle Consistency Adversarial Loss. We apply the recently proposed least square adversarial loss [11] to optimize each pair of generator and discriminator. Specifically, the loss for the generator $G : I_S \rightarrow I_T$ is defined as:

$$L_G = \sum_{x=1,2} (D_T(G(I_S^{(x)})) - 1)^2 \quad (6)$$

Minimizing the loss means to transform the image $I_S^{(x)}(x = 1, 2)$ in the source domain to make the discriminator D_T believe it is an image from the target domain. Similarly, we can define the loss for the discriminator D_T as:

$$L_{D_T} = \sum_{x=1,2} [(D_T(G(I_S^{(x)})))^2 + (D_T(I_S^{(x)}) - 1)^2] \quad (7)$$

Here the discriminator D_T tries to distinguish the real image $I_S^{(x)}$ and the transformed image $G(I_S^{(x)})$. Obviously, L_G and L_{D_T} are a pair of adversarial loss. In a similar way, we can define the loss for the generator $F : I_T \rightarrow I_S$ as:

$$L_F = \sum_{x=1,2} (D_S(F(I_T^{(x)})) - 1)^2 \quad (8)$$

and the loss of the discriminator D_S as:

$$L_{D_S} = \sum_{x=1,2} [(D_S(F(I_T^{(x)})))^2 + (D_S(I_T^{(x)}) - 1)^2] \quad (9)$$

Similar to the CycleGAN [19], we also integrate the cycle consistency loss into the model to avoid sharp change the images:

$$L_{cycle} = \sum_{x=1,2} [\|G(F(I_T^{(x)})) - I_T^{(x)}\|_1 + \|F(G(I_S^{(x)})) - I_S^{(x)}\|_1]$$

Similarity Consistency Loss. We propose the similarity consistency loss to address the ‘similarity preservation’ constraint as:

$$L_{sim} = L_v(F(G(I_S^{(1)})), F(G(I_S^{(2)}))) \quad (10)$$

Here L_v denotes the variation loss of the two images in the source domain as defined in Eq. (2), which indicates the loss of the similarity measurement of the images based on their ground-truth labels. $F(G(I_S^{(x)}))$ ($x = 0, 1$) indicates transforming the image $I_S^{(x)}$ in the source domain to the style of the target domain, and then further transforming back into the style of the source domain. Minimizing this similarity consistency loss (Eq. (10)) means making the similarity of the transformed images containing the same person as high as possible, and thus preserving the precision of the similarity measurement during the transformation.

4.4 Optimization Procedure

We adopt the Stochastic Gradient Descent method to optimize the generators (G , F) and the discriminators (D_S , D_T). Specifically, we integrate all of the generators related loss as:

$$L_{gen} = L_G + L_F + \lambda_1 L_{cycle} + \lambda_2 L_{sim} \quad (11)$$

where λ_1 and λ_2 control the relative importance of each loss. We also combine all of the discriminators related loss as:

$$L_{dis} = L_{D_S} + L_{D_T} \quad (12)$$

In each training iteration, we use the loss L_{gen} to optimize the parameters of G and F , and use L_{dis} to optimize the parameters of D_S and D_T by gradient descent.

5 Experiment

5.1 Dataset and Setting

Three widely used benchmark datasets are chosen in our experiments, including GRID [1], Market1501 [17], CUHK01 [14]. As shown in the Table 1, we select one of the above datasets as the source dataset and another one as the target dataset to test the performance of cross-dataset person Re-ID. In each source dataset, all the labeled images are used for the training of the visual classifier. On the other hand, the configurations of the target dataset ‘GRID’ follow the instructions of [1] to divide the training and testing set, and a 10-fold cross validation is conducted.

We compare the performance of the following cross-dataset person Re-ID algorithms in the experiments:

- **UMDL** [11]: It is the state-of-the-art unsupervised cross-dataset person Re-ID algorithm. UMDL is based on the cross-dataset dictionary learning method and outperforms other unsupervised learning methods as reported in [11].
- **Direct Transfer**: It is the baseline model, which directly transfers the siamese neural network based Re-ID classifier (as shown in Fig. 3), which is pre-trained in the source dataset, to the target dataset.
- **GAN**: It is the original version of GAN, which only uses the adversarial loss (L_G , L_F , L_{D_S} and L_{D_T}) to train the model.
- **SimPGAN**: It is the SimPGAN model, which integrates the adversarial loss, the cycle consistency loss and the similarity consistency loss L_{sim} . λ_1 and λ_2 are two tunable parameters in Eq. (11). By default, we set $\lambda_1 = 10$ and $\lambda_2 = 1$.

Table 1. Cross-dataset re-id precision.

Method	Source	Target	Performance		
			rank-1	rank-5	rank-10
UMDL	Market1501	GRID	3.77	7.76	9.71
	CUHK01	GRID	3.58	7.56	9.50
Direct transfer	Market1501	GRID	9.60	21.20	28.40
	CUHK01	GRID	3.60	7.20	9.20
GAN	Market1501	GRID	4.40	10.80	15.20
	CUHK01	GRID	2.40	6.40	11.60
SimPGAN	Market1501	GRID	18.00	34.40	43.20
	CUHK01	GRID	13.20	24.40	32.40



Fig. 4. Transform the images in the source dataset to the style of the target dataset.

5.2 Re-ID Results

We compare the performance of different cross-dataset transfer learning algorithms in Table 1. The performance of ‘Direct Transfer’ baseline model is quite poor, which is due to the different styles of the source and target datasets as shown in Fig. 1. The ‘SimPGAN’ model gains a lot of improvement compared with the ‘Direct Transfer’ model, and also outperforms UMDL [11] with a large margin. This proves that the image transformation greatly improves the fitness of the target data to the classifier transferred from the source domain. To verify the effectiveness of each component in SimPGAN, we also test the original version of GAN. Without the constraint of the cycle consistency loss and similarity consistency loss, ‘GAN’ achieves a much worse performance. This confirms the importance of the cycle consistency loss and the similarity consistency loss to preserve the identity information of the transformed images.

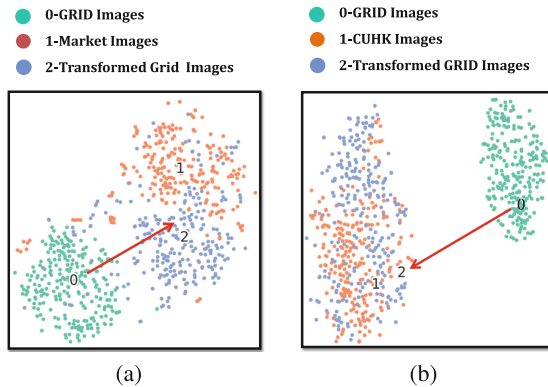


Fig. 5. Feature distribution of the images. (a) Transforming GRID to Market1501. (b) Transforming GRID to CUHK01.

To further observe how SimPGAN transforms images, we compare the transformed images with the original ones in Fig. 4. It shows clearly that the images in the target domain are transformed into the style of the source domain. Furthermore, Fig. 5 shows how the distribution of the visual features of the images changes after transformation. The feature vector of each image is extracted by the siamese convolutional neural network, which is trained on the source dataset, and is corresponding to $\mathbf{v}_i (i = 1, 2)$ in Fig. 3. Each feature vector is transformed into a 2-dimensional embedding vector by the t-SNE [4] algorithm and plotted in Fig. 5. It shows that the feature vectors of the transformed images in the target dataset are much closer to those in the source dataset after the transformation.

6 Conclusions

In this paper, we have presented SimPGAN as an efficient unsupervised cross-dataset person Re-ID algorithm based on image transformation. In particular, SimPGAN transforms the unlabeled images in the target dataset into the style of the source dataset to fit the siamese convolutional classifier, which is pre-trained on the labeled source dataset. By integrating with the similarity consistency loss and cycle consistency loss, SimPGAN preserves the important identity information of the images during the transformation. Experiments show that SimPGAN can greatly improve the fitness of unlabeled target dataset to the Re-ID classifier transferred from the source dataset, and outperforms the state-of-the-art unsupervised cross-dataset transferring algorithm by a big margin.

Furthermore, SimPGAN is easy to use as an image enhancement tool to improve the performance of other cross-dataset person Re-ID algorithms. In our future work, we will integrate the image transformation of SimPGAN with other cross-dataset person Re-ID algorithms to enhance their performance.

References

1. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1988–1995. IEEE (2009)
2. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
4. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(Nov), 2579–2605 (2008)
5. Isola, P., Jun-Yan, Z., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. *arXiv preprint* (2017)
6. Layne, R., Hospedales, T.M., Gong, S.: Domain transfer for person re-identification. In: Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, pp. 25–32. ACM (2013)

7. 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. ACM (2015)
8. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
9. Ma, L., Yang, X., Tao, D.: Person re-identification over camera networks using multi-task distance metric learning. *IEEE Trans. Image Process.* **23**(8), 3656–3670 (2014)
10. Ma, A.J., Li, J., Yuen, P.C., Li, P.: Cross-domain person reidentification using domain adaptation ranking svms. *IEEE Trans. Image Process.* **24**(5), 1599–1613 (2015)
11. Peng, P., et al.: Unsupervised cross-dataset transfer learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1306–1315 (2016)
12. Tao, D., Guo, Y., Song, M., Li, Y., Yu, Z., Tang, Y.Y.: Person re-identification by dual-regularized kiss metric learning. *IEEE Trans. Image Process.* **25**(6), 2726–2738 (2016)
13. Matsukawa, T., Okabe, T., Suzuki, E., Sato, Y.: Hierarchical Gaussian descriptor for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1363–1372 (2016)
14. Li, W., Wang, X.: Locally aligned feature transforms across views. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3594–3601. IEEE (2013)
15. Li, W., Zhu, X., Gong, S.: Person re-identification by deep joint learning of multi-loss classification. arXiv preprint [arXiv:1705.04724](https://arxiv.org/abs/1705.04724) (2017)
16. Wu, L., Shen, C., van den Hengel, A.: Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification. *Pattern Recogn.* **65**, 238–250 (2017)
17. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
18. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned CNN embedding for person reidentification. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **14**(1), 13 (2017)
19. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint [arXiv:1703.10593](https://arxiv.org/abs/1703.10593) (2017)