Cross-dataset Person Re-Identification Using Similarity Preserved Generative Adversarial Networks.

Jianming Lv, Xintong Wang

South China University of Technology

jmlv@scut.edu.cn, w.xintong@mail.scut.edu.cn

Person re-identification (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.

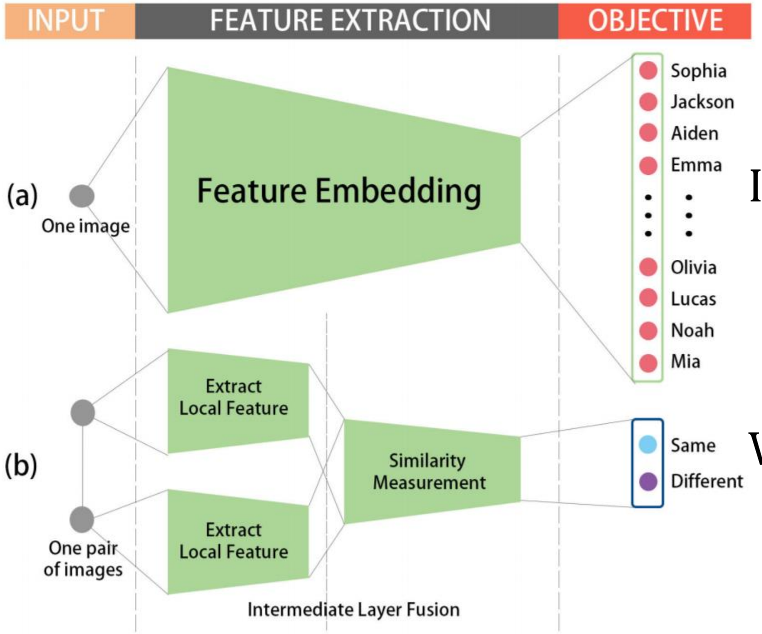
Each surveillance image containing a pedestrian is denoted as I_j.

The ID of the pedestrian in I_j is denoted as $\gamma(I_j)$.

Given any surveillance image I_j, the person Re-ID problem is to retrieve the images $\{I_k | \gamma(I_j) = \gamma(I_k)\}$, which contain the same person $\gamma(I_j)$.

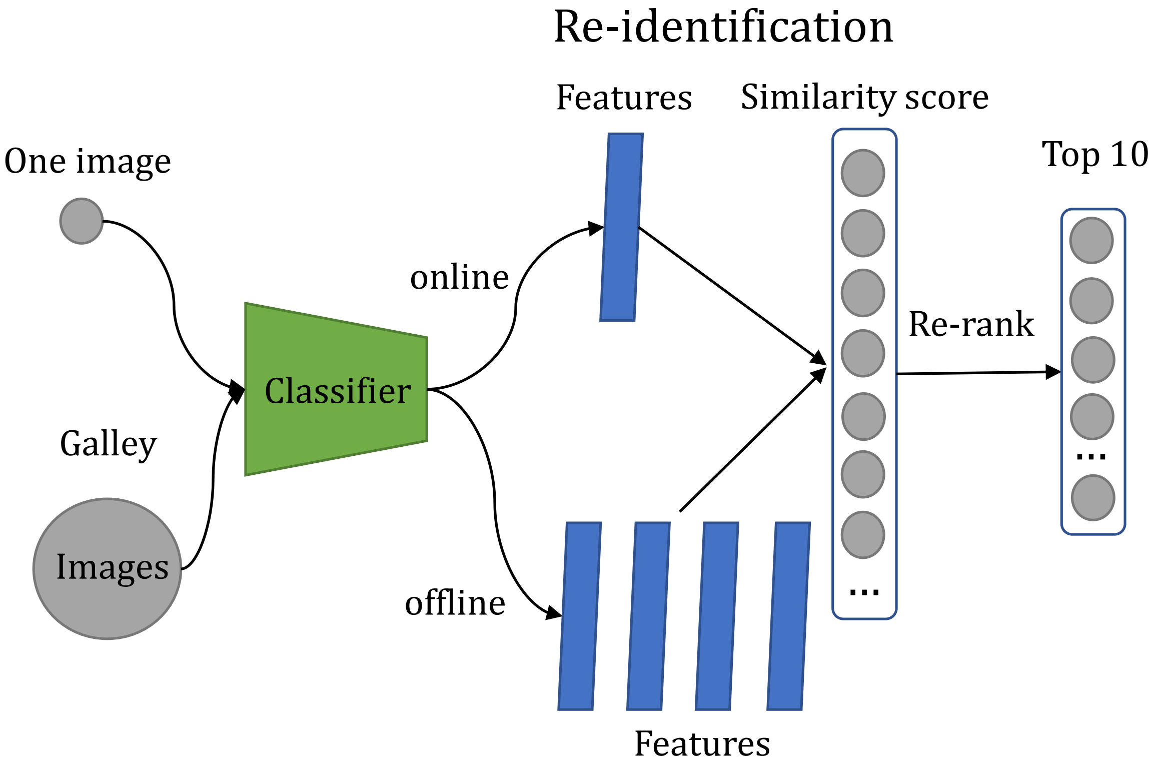
Person re-identification (Re-ID)

The traditional strategy of person Re-ID is to train a classifier 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 likehood 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. *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.*



Identification Model

Verification Model

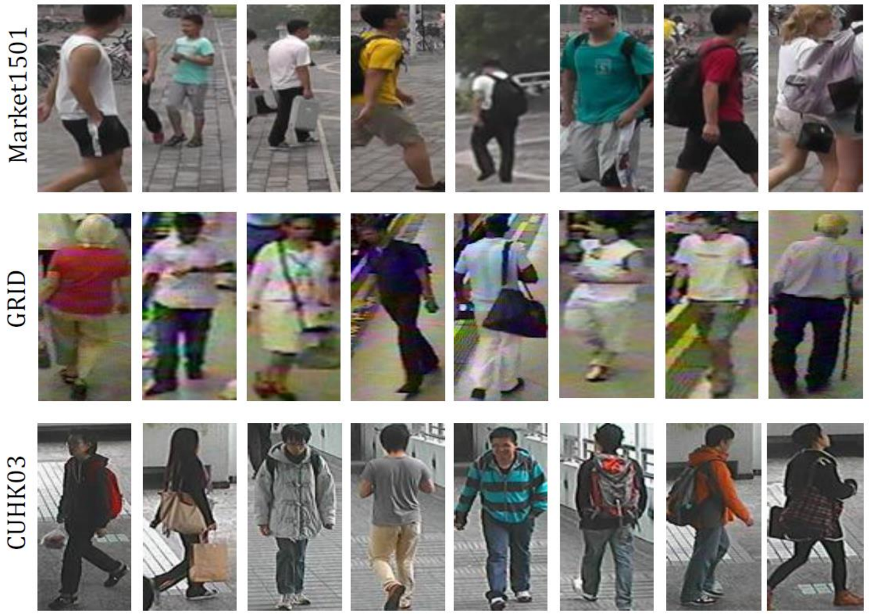


Cross-Dataset Person Re-ID

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 conduct supervised learning on small labeled datasets.

Like most of the traditional person Re-ID algorithms, we can conduct supervised learning on some public labeled dataset Ω_s, which is usually of small size, to train a classifier C. While directly deploying the trained 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.

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.



Related work

Supervised Learning: Most existing person Re-ID models are supervised.

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.

Unsupervised Learning: Some unsupervised Re-ID methods 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: Transfer learning algorithms leverage the Re-ID models pre-trained in other labeled datasets to improve the performance on target datasets.

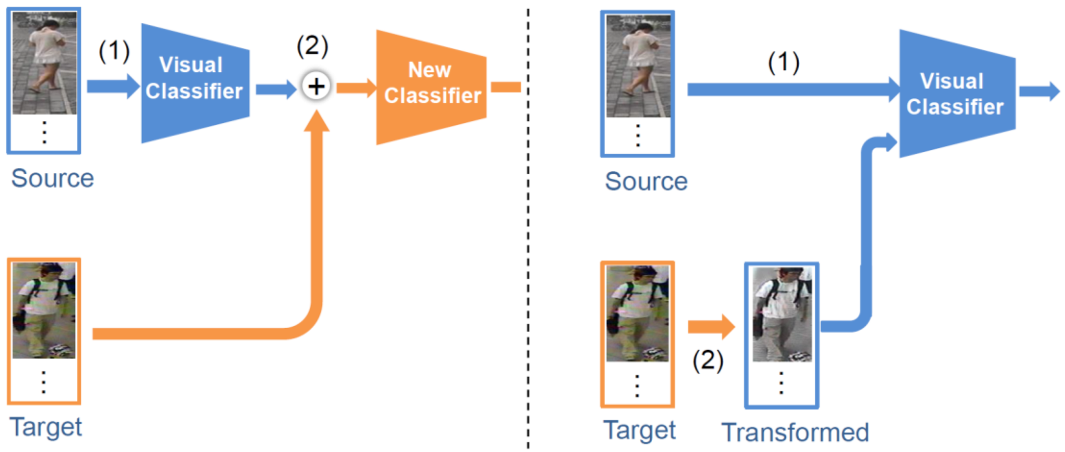
Supervised transfer learning: Both of the source and the target datasets are labeled or have weak labels.

Unsupervised transfer learning: UMDL 27.1% rank-1 accuracy vs 67% (Supervised)

A long way to go!

Model overview

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.



(a) (b)

We propose a novel generative model based solution. Instead of incrementally optimizing the classifier to fit the new data, we transform the data in the target dataset to fit the classifier.

Model overview

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.

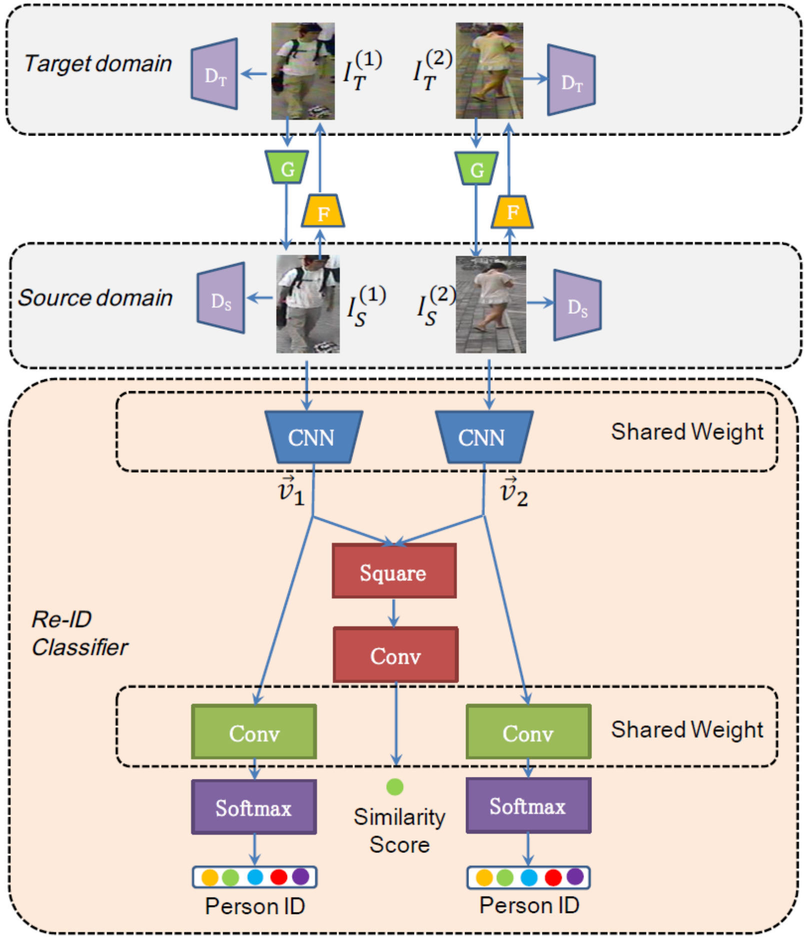
Similarity preservation: The similarity of the transformed images, which contain the same person, should be as high as possible.

To address the **'data fitness' constraint**, we integrate the cycle consistency loss in the GAN model, which is proved to be very powerful to improve the quality and steadiness of unpaired images translation.

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.

Model: Siamese CNN based Re-ID Classifier

We select the recently proposed siamese convolutional neural network as the Re-ID classifier, which makes better use of the label information and has a good performance in the large-scale datasets.



Variation loss:

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

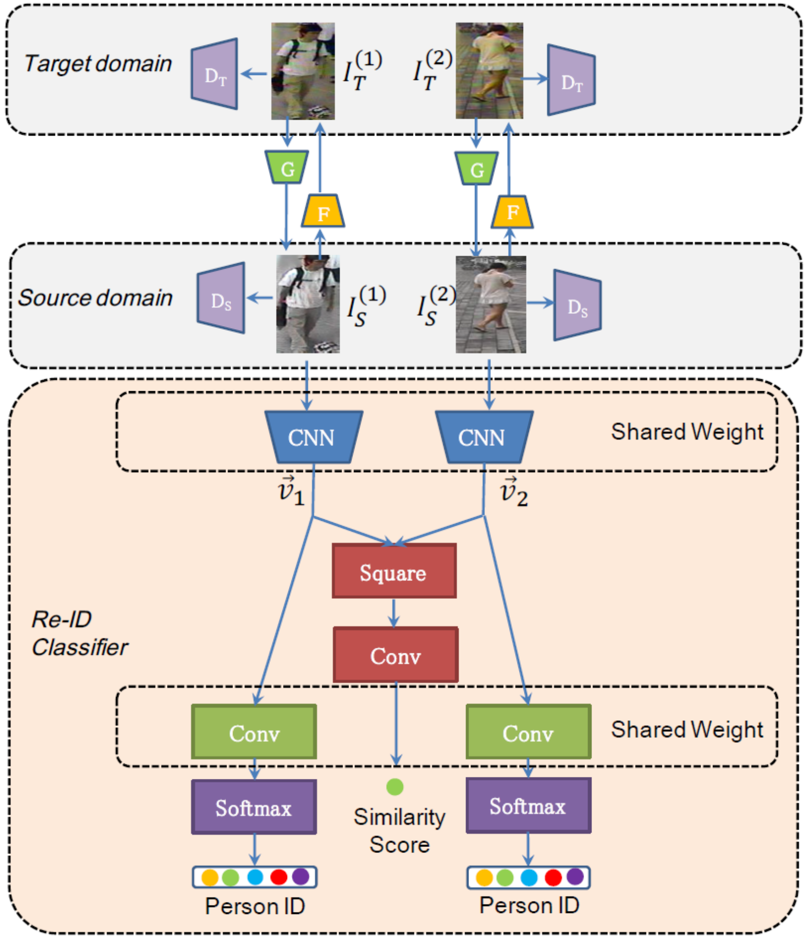
Identification loss:

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

Final loss function: $L_{all} = L_v + L_{id}$

Model: 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



G: $I_S \rightarrow I_T$ and F: $I_T \rightarrow I_S$

D_S: Discriminate real images I_S and generated images $F(I_T)$

D_T: Discriminate 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 and our proposed similarity consistency loss.

Model: Cycle Consistency Adversarial Loss

G: $I_S \rightarrow I_T$

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

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

F: $I_T \rightarrow I_S$

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

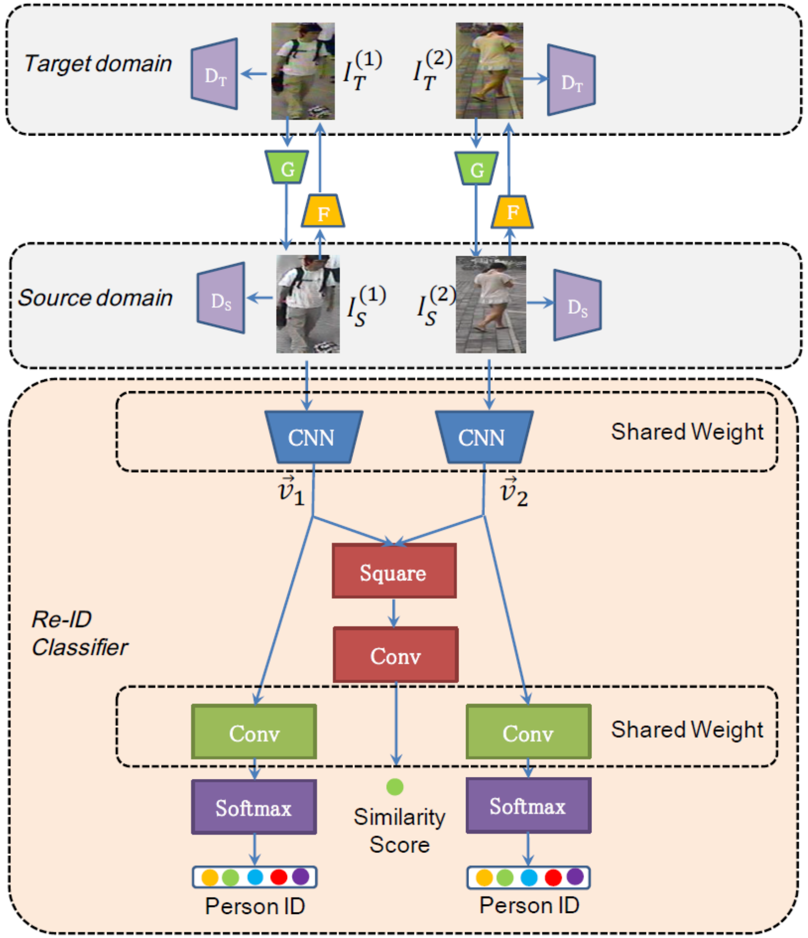
$$L_{D_S} = \sum_{x=1,2} [(D_S(F(I_T^{(x)})))^2 + (D_S(I_S^{(x)}) - 1)^2]$$

Cycle consistency loss:

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

Similarity Consistency Loss: similarity preservation

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



Model: Optimization procedure

We adopt the Stochastic Gradient Descent method to optimize the generators (G,F) and the discriminators (D_S, D_T).

$$L_{gen} = L_G + L_F + \lambda_1 L_{cycle} + \lambda_2 L_{sim}$$

$$L_{dis} = L_{D_S} + L_{D_T}$$

$\lambda_1\ and \lambda_2$ control the relative importance of each loss.

Experiment

Dataset and Baselines:

Three widely used benchmark datasets are chosen in our experiments, including **GRID, Market1501, CUHK01**. 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.

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

UMDL: It is the state-of-the-art unsupervised cross-dataset person Re-ID algorithm.

Direct Transfer: It is the baseline model, which directly transfers the siamese neural network based Re-ID classifier, 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 to train the model.

SimPGAN: It is the SimPGAN model, which integrates the adversarial loss, the cycle consistency loss and the similarity consistency loss. $\lambda_1 = 10 \; and \; \lambda_2 = 1$

Experiment

Re-ID Results:

The performance of 'Direct Transfer' baseline model is quite poor, which is due to the different styles of the source and target datasets.

The 'SimPGAN' model gains a lot of improvement compared with the 'Direct Transfer' model, and also outperforms UMDL 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.

Without the constraint of the cycle consistency loss and similarity consistency loss, 'GAN' achieves a much worse performance.

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
SimPGANS	Market1501	GRID	17.40	32.60	40.40
	CUHK01	GRID	12.80	23.80	31.20
SimPGAN	Market1501	GRID	**18.00**	**34.40**	**43.20**
	CUHK01	GRID	**13.20**	**24.40**	**32.40**

Cycle consistency loss avoids sharp change the images. With the similarity consistency loss can filter out more noise and color shift while highlighting the important portions which can facilitate identification.

Experiment

Re-ID Results:

The performance of 'Direct Transfer' baseline model is quite poor, which is due to the different styles of the source and target datasets.

The 'SimPGAN' model gains a lot of improvement compared with the 'Direct Transfer' model, and also outperforms UMDL 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.

Without the constraint of the cycle consistency loss and similarity consistency loss, 'GAN' achieves a much worse performance.

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**

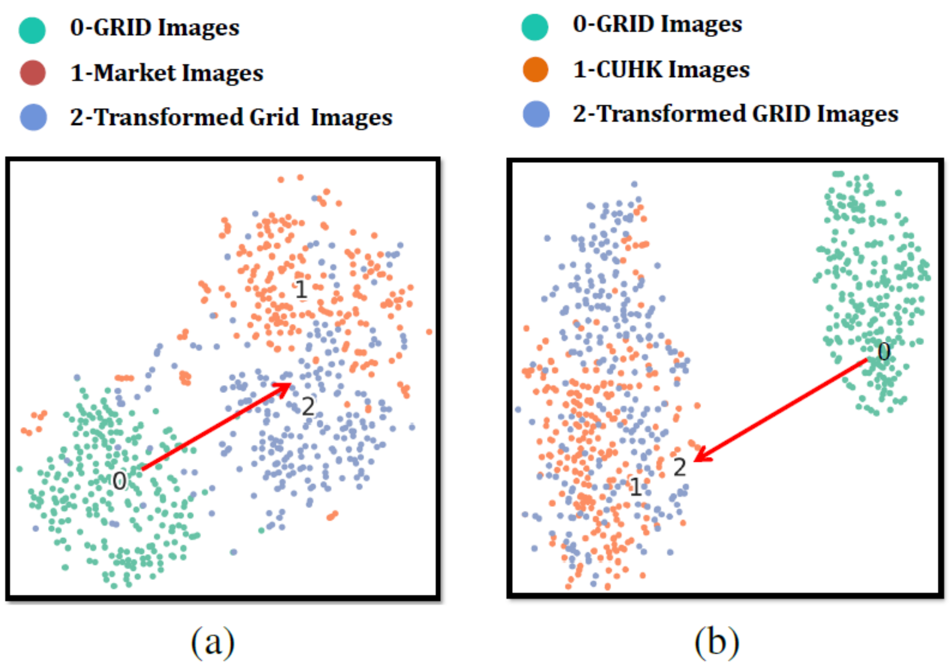
Cycle consistency loss avoids sharp change the images. With the similarity consistency loss can filter out more noise and color shift while highlighting the important portions which can facilitate identification.

Experiment

Re-ID Results:

We compare the transformed images with the original ones. It shows clearly that the images in the target domain are transformed into the style of the source domain.

Furthermore, the right figure shows how the distribution of the visual features of the images changes after transformation.



That's all!

Intelligent Information Fusion Research Group
School of Computer Science and Engineering
South China University of Technology