



## Multi-domain Cross-dataset and Camera Style Transfer for Person Re-Identification

Rabia Tahir<sup>1</sup>, Keyang Cheng<sup>2</sup>, Lubamba Kasangu Eric<sup>3</sup>, Muhammad Saddam Khokhar<sup>4</sup>

<sup>1,2,3,4</sup> School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China

<sup>1</sup> rabiatahir074@gmail.com

<sup>2</sup> kychem@ujs.edu.cn

<sup>3</sup> 5103171302@stmail.ujs.edu.cn

<sup>4</sup> saddam\_khokhar@hotmail.com

### ABSTRACT

The major reason of performance drop in person re-identification (re-id) models is domain gap between various datasets. Each person re-id dataset is different from other dataset because of change in lightning conditions, camera angle, pose, illumination, background and resolution. This results in domain gap between these datasets and leads to performance drop in person re-identification models. The domain gap not only exists between distinct datasets but it is also present in a dataset which contains images taken from multiple cameras. Generative Adversarial Networks (GANs) have achieved successful results in various fields, especially for image-to-image translation. In this paper, we propose an unsupervised domain adaptation method based on ComboGAN and Xception to perform cross-dataset and camera style translation. Our framework generates images in different styles of a single person according to various domains. Experimental results show the effectiveness of the proposed model on two popular re-id datasets Market-1501 and PETA.

**Key words:** Person re-identification, Camera Style Transfer, GAN, Xception

### 1. INTRODUCTION

Person re-identification (re-id) is a very famous field because of its research significance and applications. It is a task of identifying a person who has previously been located over a camera network [1]. The ultimate purpose of any surveillance system is to understand and analyze the scene and decide either the given person has already been observed by some camera network or not. Although performance of person re-id has been increased a lot in past years but still this field is facing many issues such as viewpoint and pose changes, complex scene, lightening variations and large number of identities in a camera network. The appearance of a person in one camera is different as compared to appearance captured by some other camera which is called domain gap [2], [3], [4]. Traditionally, handcrafted algorithms and small-scale techniques were used to perform person re-id. Recently, many deep learning techniques have been used in person re-id tasks because of emergence of very large and complex datasets [5]. Deep learning methods have obtained high accuracy as compared to handcrafted methods on datasets iLIDS, CUHK01, CUHK03, Market-1501 and PRID450S.

Especially, Generative Adversarial Net (GAN) variants helped to solve problems of domain gap, pose estimation for person re-id [6]. The main disadvantage of deep learning models in person re-ID is limited training data. For example, VIPeR contains only two images for a person. In real applications, the existing datasets are not enough to deal with large number of identities [5]. Moreover, there is another major problem for person re-id datasets i.e., domain gap. It means performance of person re-id model which is trained on one re-id dataset does not gives satisfactory results on testing dataset. Annotating new datasets is difficult, time consuming and costly. Therefore, it is important to reduce the domain gap between person re-id datasets [2]. To address these issues, we perform multi-domain style adaptation (or style translation) on person re-id datasets with an unsupervised domain adaptation method based on ComboGAN [7] and Xception. Our main contribution in this paper are:

- We perform cross-dataset and camera style translation to bridge domain gap between different re-id datasets with an unsupervised domain adaptation method which is based on ComboGAN.
- We use Xception to quantitatively evaluate the translated images and perform comparative analysis.

The rest of paper is organized as follows: Section 2 explains related work about person re-id and style translation with GAN. Section 3 describes our methodology and models to perform style translation. Experiments and results are explained in section 4. Conclusion is given in section 5. References are delineated at the end.

### 2. RELATED WORK

#### 2.1 Image-to-Image Translation with Generative Adversarial Networks

In past few years, impressive results have been obtained in the field of Image-to-Image (I2I) translation with different supervised and unsupervised models of Generative Adversarial Network (GAN). There are many examples of I2I such as super resolution, colorization, image synthesis, style transfer and domain adaptation [8], [2]. The aim of I2I is to learn a mapping function among various visual domains. BicycleGAN [9] and Pix2Pix [6] are supervised image-to-image translation methods proposed by Zhu and Isola, respectively. Unsupervised image translation task is very challenging to perform because it is difficult to find correspondence among images from two different domains

when there is no paired images as an example [10]. In 2017, CycleGAN [11], DiscoGAN [12], and DualGAN [13] were proposed to perform unsupervised image-to-image translation tasks. They used two mapping function (forward and inverse) and cycle consistency loss to perform image translation without paired data [14]. Other unsupervised models for image-to-image translation are XGAN [15], StarGAN [16], UNIT [17], DA-GAN [10], and PairedCycleGAN [18]. Anoosheh *et al.* [7] proposed ComboGAN which is an extension of CycleGAN model for multiple domain style transfer. Our work is based on ComboGAN for multi-domain style transfer on person re-id datasets.

### 2.2 Unsupervised Domain Adaptation for Person Re-identification

Unsupervised Domain adaptation (UDA) has got attention in past few years that learns cross-domain mapping i.e., from source domain to target domain. In UDA, we have provided labeled source domain with unlabeled target domain. For target domain, it learns a discriminative representation. Domain adaptation is sometimes referred as transfer learning [19], [20], [21]. Different models were proposed which perform style transfer for domain adaptation [8], [22]. Our work performs multi-domain style transfer for person re-id models. In case of person re-id, UDA assumes that the source and target domains contain different categories or classes (person as different classes) which makes it more complex and challenging [23]. Recent work of UDA focused on decreasing gap between the source and target domains. Like, Deng *et al.* [24] proposed SPGAN (Similarity Persevering Image Generation) to learn domain adaption for person re-identification tasks. Zhong *et al.* [4] proposed camera style adaptation method which used LSR

(Label Smooth regularization) loss along with CycleGAN to remove noise and over-fitting in translated images. They considered camera styles as different domains in Market-151 and DukeMTMC-reID datasets. As domain gap also exists in a dataset because of illumination, camera angle and pose etc. Therefore, they aimed to reduce domain gap in a dataset. Another model PTGAN [2] was proposed to bridge domain gap between distinct re-id datasets. Other works include CTGAN [25], SimPGAN [26], HHL [27], IPGAN [23]. The aim of all this research is to increase training dataset for person re-id by performing domain adaptation on re-id datasets.

### 3. THE PROPOSED MODEL

In this paper, we have a single source domain and multiple target domains. The source domain contains set of images  $X_s$  with their labels  $L_s$  and person identities  $N_s$ . The multiple target domains contain images  $[X_{t_1}, X_{t_2}, \dots, X_{t_n}]$  and person identities  $N_t$  without labels. Our model learns a mapping function  $G: X_s \rightarrow X_{t_1}$  from source domain to target domain in an unsupervised way to perform domain style adaptation. We use ComboGAN [7] as a base model with some modifications to perform cross-dataset and camera style translation. Our work is different from other works because we do not use traditional re-id models to quantitatively evaluate the images. Instead, we use very recently proposed Xception semantic classifier to inspect the quality of style transferred images. Figure 1 shows overall working of our method.

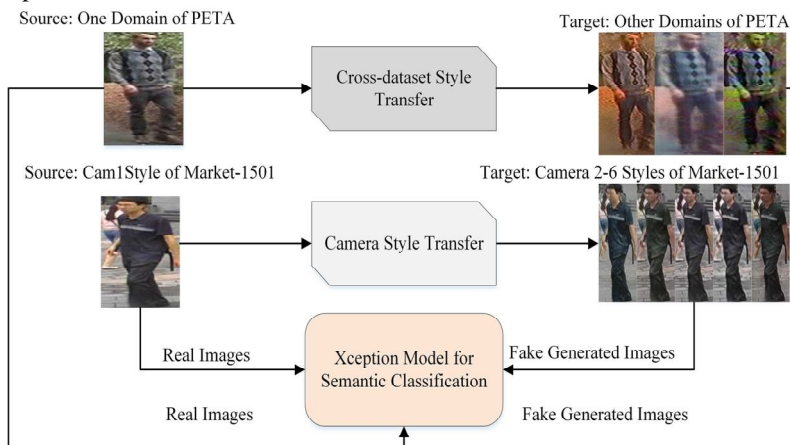


Figure 1: The proposed framework: Style transfer on PETA and Market-1501 with ComboGAN and Xception

### 3.1 Revisit ComboGAN

Anoosheh *et al.* [7] proposed Image-to-Image translation model ComboGAN which solves  $\theta(n^2)$  scaling problem in CycleGAN. It uses encoders and decoders by decoupling generators. During inference process, any image  $y$  from domain  $Y$  can translate into style of domain  $Z$  with the help of these encoders and decoders. It is mathematically expressed in equation 1. With the increase of domains, the number of generators linearly scales with a pair of encoder and decoder. The model performs training by covering all domains uniformly. The discriminator in ComboGAN works

same as in CycleGAN. Also, ComboGAN has same architecture like CycleGAN except the extension of domains. The network specifications of ComboGAN is explained in Fig. 2. For multi-domain case, the encoder output is suitable for any other decoder. It uses same losses as of CycleGAN i.e. adversarial loss and cycle consistency loss. But it replaces the negative log-likelihood loss function by mean-squared loss. During training, ComboGAN randomly selects any two domains  $Y$  and  $Z$  from  $i = \{1, 2, 3 \dots n\}$  domains. Then, uses two mapping functions i.e.,  $G: Y \rightarrow Z$  and inverse function  $I: Z \rightarrow Y$ . The adversarial loss for  $G$  and  $I$  is expressed in equation 2 and 3.

$$z = G_{ZY}(y) = Decoder_Z Encoder_Y(y) \tag{1}$$

$$L_{adv}(G, D_z, Y, Z) = \mathbb{E}_z [(D_z(z)-1)^2] + \mathbb{E}_y [(D_z(G(y)))^2] \quad (2)$$

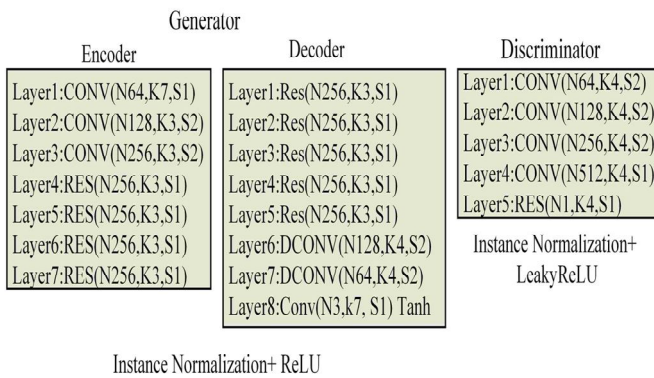
$$L_{adv}(I, D_y, Z, Y) = \mathbb{E}_y [(D_y(y)-1)^2] + \mathbb{E}_z [(D_y(G(z)))^2] \quad (3)$$

Where  $D_z$  and  $D_y$  are two different discriminators which differentiate generated images from source domain. Further, they used cycle consistency loss that makes sure after style translation, image should retain its original state i.e.  $G(I(z)) \approx z$  and  $I(G(y)) \approx y$  and it is expressed in equation 4 [6].

$$L_{cycle}(G, I, Y, Z) = \mathbb{E}_y [|| I(G(y))-y ||_1] + \mathbb{E}_z [|| G(I(z))-z ||_1] \quad (4)$$

The overall loss function for ComboGAN is described in equation 5. Where  $\lambda$  is a hyper-parameter for adjusting weight and usually its value is 10.

$$Loss_{full} = \lambda L_{cycle}(G, I, Y, Z) + \lambda L_{adv}(G, D_z, Y, Z) + \lambda L_{adv}(I, D_y, Z, Y) \quad (5)$$



**Figure 2:** Network specifications of ComboGAN. N=Neurons, S=Stride, K=kernel, RES= Residual Block, Conv=Convolutional layer, Dconv= transposed convolutional layer.

### 3.2 Cross-dataset and Camera Style Transfer with ComboGAN

In original ComboGAN, 14-painter dataset and Alps images were used to perform multi-domain image style translation. This is related to our work because we want to perform multi-domain style translation on person re-id datasets. Therefore, we keep the underlying foundation of ComboGAN same and adopt different weight initializing method for ComboGAN. Instead of using simple weight initialization, we use Xavier weight initialization [28], [29]. By following [28], we use Instance normalization in both generator and discriminator because Instance normalization performs better as compared to batch normalization for style transform. Also, we set  $\lambda_{id}=1$  because it gives better results as compared to  $\lambda_{id}=0$  and use the PatchGAN discriminator [25]. All images are scaled to  $128 \times 128 \times 3$  for PETA dataset and  $256 \times 256 \times 3$  for Market-1501 dataset.

### 3.3 Xception Classifier

There are different ways to quantitatively evaluate the generated images. Mostly recent works used pre-trained classifiers which semantically measure the generated images. In this paper, we use classification accuracy as a metric to evaluate the performance of the model [28]. Our work is different from other works because we do not use common re-id models to quantitatively evaluate the translated images of re-id datasets. Instead, we use Xception classifier to inspect the quality of generated images. There are various deep learning image classifiers such as VGG [30], ResNet50 [31], Inception V3 [32] and Xception [33]. Among them, Xception is the latest one proposed in the year 2017. Basically, it is an extension of Inception network and uses depth wise separable modules instead of inception modules. We train Xception with new training dataset  $D_{trans}$  which is a combination of original and translated images. We set 0.5 % images in training and validation sets for both PETA and Market-1501 datasets.

## 4. EXPERIMENTS AND RESULTS

In this section, we perform experiments for cross-dataset style transfer and camera style transfer on PETA and Market-1501 datasets respectively.

### 4.1 Datasets

We used two state-of-the-art datasets PETA and Market-1501 for experiments.

PETA dataset [34] is one of the most dynamic and largest person re-id dataset. It contains 19000 images with different resolution ranges from  $17 \times 39$  to  $169 \times 365$  pixels. This dataset is divided into ten different subsets according to various attributes such as scene, illumination, resolution, and camera angle and view point. We choose GRID, VIPeR, 3DPeS and CAVIAR4REID among them. Other detail of these datasets is given in Table 1.

Market-1501 [35] dataset has been used for most of person re-id tasks. It has 32,668 images of 1,501 persons gathered from six different cameras. The dataset contains five folders `bounding_box_train`, `bounding_box_test`, `gt_bbox`, `gt_query` and `query`. As an average, there are 17.2 images per person in `bounding_box_train` folder for training.

**Table 1:** Detail of four-person re-id datasets from PETA

Dataset	GRID	VIPeR	3DPeS	CAVIAR 4REID
No of images	1275	1264	1012	1220
Illumination	Low	Varying	Varying	Low
Resolution	From 29x67 to 169x365	48x128	From 31x100 to 236 x 178	From 17x39 to 72x141
Camera Angle	Varying	Ground	High	Ground
Scene type	Indoor	Outdoor	Outdoor	Outdoor
View point	Frontal and back	Varying	Varying	Varying

**Table 2:** Hyper-parameters of the proposed model.

Parameters	Values
Epochs	100
Image load size and fine size	128×128×3 (PETA)
Image load size and fine size	256×256×3 (Market-1501)
Normalization type	Instance Normalization
Batch-size	1
Decay rate	30
$\beta_1$ and $\beta_2$	0.5 and 0.999
$\lambda_{\text{cycle}}$ , $\lambda_{\text{identity}}$ , $\lambda_{\text{latent}}$ and $\lambda_{\text{forward}}$	10, 1, 1 and 0.2
Learning rate	0.0001

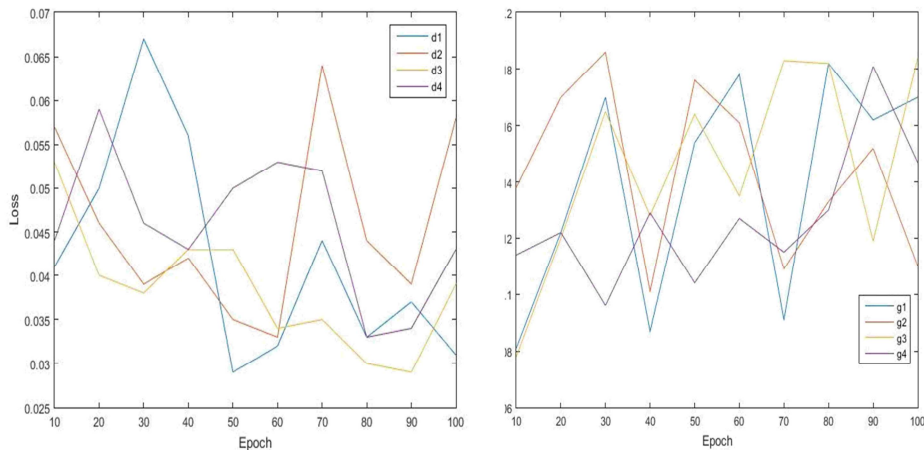
**4.2 Implementation Details**

We used Pytorch to train ComboGAN with some modifications in the network. The Adam optimization algorithm is used with  $b_1=0.5$  and  $b_2=0.999$ . All images are scaled to 128×128×3 for PETA and 256×256×3 for Market-1501 dataset. One image is used as a batch and learning rate is set to 0.0001 which slowly decays to zero at the end. We enabled random crops, random flipping, and random erasing for data augmentation and set dropout value as 0.5 to avoid over-fitting in the model. Our model has various hyper-parameters, whose values are given in Table 2. In cross-dataset style transfer and camera style transfer, we generate K-1 (3 for PETA and 5 for Market-1501) fake images while preserving their original ID. For cross-dataset style transfer, we run 100 epochs on four different domains i.e. GRID, VIPeR, 3DPeS and CAVIAR4REID. It took about 2 days on NVIDIA Tesla P100 GPU. For camera style transfer, we run 100 epochs on six camera style domains i.e. cam1, cam2, cam3, cam4, cam5 and cam 6. It took about 4 days on NVIDIA Tesla P100 GPU.

**4.3 Results of Cross-dataset Style Transfer on PETA Dataset**

We performed cross-dataset style transfer to bridge the domain gap which usually exists between different person re-id datasets. For this purpose, we used PETA dataset as it contains images from various popular re-id datasets. We choose GRID, VIPeR, 3DPeS and CAVIAR4REID from PETA dataset and set four different domains  $K=4$  accordingly. The purpose was to obtain K-1 domain style images for a single image. The generated images are called style transfer images, it means they preserve the content of original image but adopt the style (texture) of target image. Therefore, the person in the original image remains same and we kept same labels for both original and translated images. We used four generators and discriminators in ComboGAN and Fig. 3 shows fluctuations of loss values in all generators and discriminators.

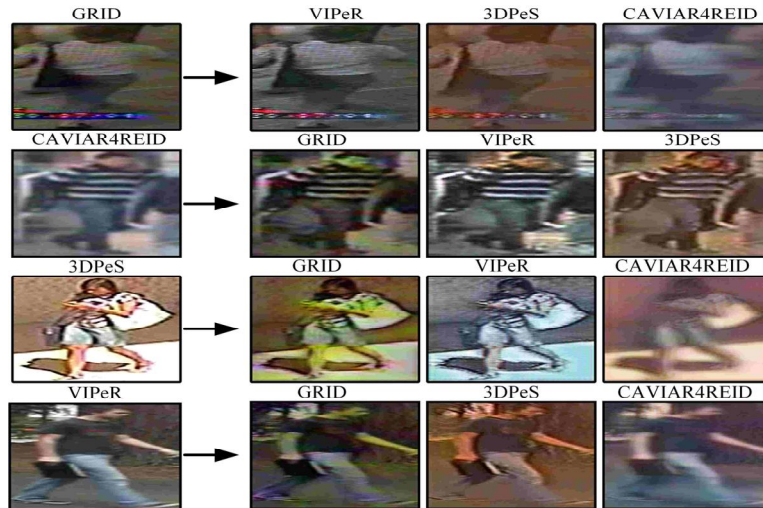
It can be seen clearly that there is great fluctuation among losses of generator and discriminators in Fig. 3. The reason is that training of GAN is not very stable. Decrease in generator loss means increase in discriminator loss and vice versa. Therefore, it is always crucial to select parameters of GAN for stabilize training. At last epochs, all generators and discriminators losses are decreased as compared to initial epochs. Fig. 4 shows the style transferred images of PETA dataset. It can be seen that the source image is successfully translated into style of three target domains. For quantitative evaluation, we combined original dataset images with translated images to train Xception [33]. Total size of images became 13,849 after integration. The obtained accuracy and loss values are reported in Table 3 and Fig. 5 for this experiment. It is clearly seen that at 10 epoch, both losses are decreased as compared to initial epochs which shows the better quality of generated images. We compared our results with baseline ComboGAN [7] and SingleGAN [28]. We used same dataset with same number of epochs and hyper-parameters and applied original ComboGAN and SingleGAN on PETA dataset. Table 3 shows comparison of our results with these models. Our results are better than baseline ComboGAN but slightly less than SingleGAN.



**Figure 3:** Fluctuation of all four generator and discriminator loss values.

**Table 3:** Comparison of our method with baseline ComboGAN and SingleGAN on PETA dataset

Model	Accuracy (%)	Loss (%)	Validation Accuracy (%)	Validation Loss (%)
ComboGAN	98.46	5.62	99.65	1.30
SingleGAN	99.74	0.7	99.96	0.32
Our model	99.31	2.46	99.68	1.41

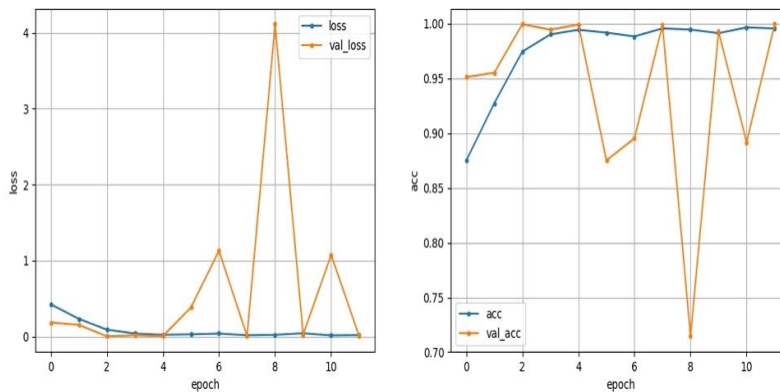


**Figure 4:** Cross-dataset style transfer on PETA dataset: The first column shows source domain and next three columns show translated image in the target domains. Labels on the top of images show names of datasets

#### 4.4 Results of Camera Style Adaptation on Market-1501 Dataset

Large person re-id datasets such Market-1501 and DukeMTMC-reID possess domain gap because of camera

style variations in a dataset. Market-1501 dataset contains images taken from six different cameras and DukeMTMC-reID contains images taken from eight cameras. Therefore, domain gap also exists in a dataset because of variation in camera’s setting [4]. We used modified ComboGAN



**Figure 5:** Loss and accuracy values of Xception for PETA dataset (original images + translated images)



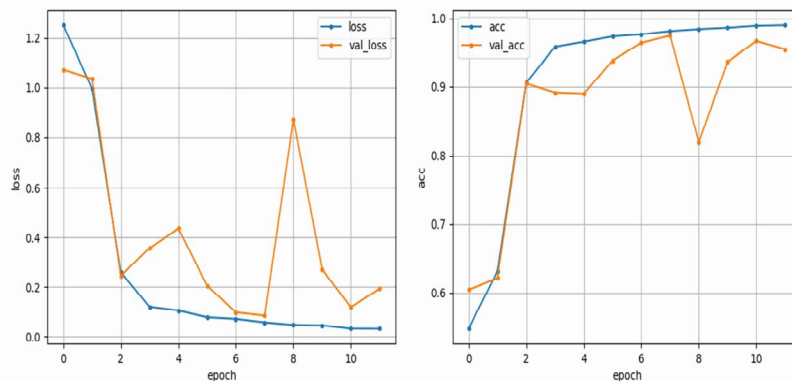
**Figure6:** Qualitative results of Market-1501 dataset. Source image is on left side while translated images are on right sides

to produce new training images by considering each camera style as a single domain. Market-1501 re-id dataset contains images taken from six cameras, therefore we set six domains and produce K-1 images for each source image (where K is number of total cameras). We used 12,936 images from the training subset of Market-1501 dataset. Fig. 6 shows result of camera style transfer on Market-1501. For Market-1501 dataset, we combined 12,936 original training images with 35,064 fake generated images. A total of 48,000 images were used to train Xception with 8,000 images for each camera style. The obtained accuracy and loss values for Market-1501 are reported in Table 4 and Fig. 7. We can see that obtained accuracy value for cross-data style transfer on PETA is high as compared to Market-1501. The reason is smaller number of images and domains in this experiment. However, Market-1501 dataset contains more images with 6

different camera style images. Therefore, obtained classification accuracy is relatively low. Also, it takes more time to complete experiment of style transfer on GPU as compared to PETA dataset.

**Table 4:** Quantitative results of both experiments on two datasets.

Experiment	Cross-dataset Style Transfer	Camera Style Transfer
Dataset	PETA	Market-1501
Accuracy	<b>99.31</b>	98.96
Loss	<b>2.46</b>	3.19
Validation Accuracy	<b>99.68</b>	95.37
Validation loss	<b>1.41</b>	19.33
No of Domains	4	6
Time	<b>~2 days</b>	~4 days



**Figure7:** Loss and accuracy values of Xception for Market-1501 (original images + translated images).

## 5 CONCLUSION AND FUTURE WORK

Person re-id datasets possess different characteristics because of change in lightening conditions, illumination, and background and camera angle. Therefore, training on

one dataset does not gives satisfactory results on other dataset. To achieve high accuracy for person re-id models, it is necessary to train that model on different datasets and fill the domain gap. In this paper, we proposed an unsupervised domain adaptation method based on ComboGAN and

Xception. We performed multiple experiments on Market-1501 and PETA re-id datasets. The purpose of these experiments was to increase the training images of person re-id datasets and bridge the domain gap between these datasets. Due to source limitations, we compared our method with only two image-to-image translation models. In future, we aim to perform comparative analysis with recent and new style translation models and improve accuracy on complex and large re-id datasets.

#### ACKNOWLEDGMENTS

This research is supported by Natural Science Foundation of China (No.61602215, No.61672268), and the science foundation of Jiangsu province (No.BK20150527).

#### REFERENCES

- [1] X. Qian, Y. Fu, T. Xiang, W. Wang, and J. Qiu, "Pose-Normalized Image Generation for Person," in *European Conference on Computer Vision (ECCV)*, 2018, p. 650–667.  
[https://doi.org/10.1007/978-3-030-01240-3\\_40](https://doi.org/10.1007/978-3-030-01240-3_40)
- [2] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person Transfer GAN to Bridge Domain Gap for Person Re-Identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 79–88.  
<https://doi.org/10.1109/CVPR.2018.00016>
- [3] S. Bak *et al.*, "Person Re-identification Using Spatial Covariance Regions of Human Body Parts," *Adv. video signal based Surveill.*, pp. 435–440, 2010.
- [4] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camera Style Adaptation for Person Re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, p. 5157–5166.  
<https://doi.org/10.1109/CVPR.2018.00541>
- [5] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person Re-identification : Past , Present and Future," *arXiv Prepr. arXiv1610.02984*, pp. 1–20, 2016.
- [6] P. Isola, A. A. Efros, B. Ai, and U. C. Berkeley, "Image-to-Image Translation with Conditional Adversarial Networks," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.  
<https://doi.org/10.1109/CVPR.2017.632>
- [7] A. Anoosheh, E. Agustsson, R. Timofte, and L. Van Gool, "ComboGAN: Unrestrained Scalability for Image Domain Translation," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, vol. 1, no. 1, pp. 783–790.
- [8] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. K. Singh, and M.-H. Yang, "Diverse Image-to-Image Translation via Disentangled Representations," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 35–51.  
[https://doi.org/10.1007/978-3-030-01246-5\\_3](https://doi.org/10.1007/978-3-030-01246-5_3)
- [9] J. Zhu, T. Darrell, R. Zhang, and A. A. Efros, "Toward Multimodal Image-to-Image Translation," *Adv. Neural Inf. Process. Syst.*, no. 1, p. 465–476, 2017.
- [10] T. Ma, S., Fu, J., Wen Chen, C., and Mei, "DA-GAN: Instance-level Image Translation by Deep Attention Generative Adversarial Networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, p. 5657–5666.  
<https://doi.org/10.1109/CVPR.2018.00593>
- [11] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017–Octob, pp. 2242–2251, 2017.
- [12] J. Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, "Learning to discover cross-domain relations with generative adversarial networks.," in *34th International Conference on Machine Learning*, 2017, p. 1857–1865.
- [13] M. Yi, Zili and Zhang, Hao and Tan, Ping and Gong, "Dualgan: Unsupervised dual learning for image-to-image translation.," in *IEEE international conference on computer vision*, 2017, pp. 2849–2857.  
<https://doi.org/10.1109/ICCV.2017.310>
- [14] D. Bang, S. Korea, and H. Shim, "Resembled Generative Adversarial Networks: Two Domains with Similar Attributes," in *british machine vision conference*, 2018, pp. 1–12.
- [15] K. Royer, Amelie and Bousmalis, Konstantinos and Gouws, Stephan and Bertsch, Fred and Mosseri, Inbar and Cole, Forrester and Murphy, "XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings," *arXiv Prepr. arXiv1711.05139*, pp. 1–19, 2017.
- [16] M. I. Translation, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.
- [17] M. Liu, T. Breuel, and J. Kautz, "Unsupervised Image-to-Image Translation Networks," *Adv. Neural Inf. Process. Syst.*, no. Nips, p. 700–708., 2017.
- [18] H. Chang, J. Lu, A. Research, F. Yu, and A. Finkelstein, "PairedCycleGAN: Asymmetric Style Transfer for Applying and Removing Makeup," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 40–48, 2018.  
<https://doi.org/10.1109/CVPR.2018.00012>
- [19] Z. Li, W. Wang, and Y. Zhao, "Image Translation by Domain-Adversarial Training," *Comput. Intell. Neurosci.*, 2018.
- [20] Q. Pan, Sinno Jialin and Yang, "A survey on transfer learning.," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. (10), pp. 1345–135, 2009.  
<https://doi.org/10.1109/TKDE.2009.191>
- [21] Y. Zhong, Z., Zheng, L., Luo, Z., Li, S., and Yang, "Invariance matters: Exemplar memory for domain domain adaptive person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, p. 598–607.
- [22] X. Mao and Q. Li, "Unpaired Multi-Domain Image Generation via Regularized Conditional GANs," in *International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018, pp. 2553–2559.
- [23] J. Liu, W. Li, and H. Pei, "Identity Preserving Generative Adversarial Network for Cross-Domain Person Re-identification," *arXiv Prepr. arXiv1811.11510*, 2018.

- [24] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 994–1003.  
<https://doi.org/10.1109/CVPR.2018.00110>
- [25] J. V. C. I. R, S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification q," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 393–400, 2019.
- [26] J. Lv and X. Wang, "Cross-dataset Person Re-Identification Using Similarity Preserved Generative Adversarial Networks," in *International Conference on Knowledge Science, Engineering and Management*, 2018, pp. 171–183.  
[https://doi.org/10.1007/978-3-319-99247-1\\_15](https://doi.org/10.1007/978-3-319-99247-1_15)
- [27] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing A Person Retrieval Model Hetero- and Homogeneously," in *European Conference on Computer Vision (ECCV)*, 2018, vol. 1, pp. 172–188.
- [28] X. Yu, X. Cai, Z. Ying, T. Li, and G. Li, "SingleGAN: Image-to-Image Translation by a," *Prepr. arXiv1810.04991.*, 2018.
- [29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, vol. 9, pp. 249–256.
- [30] K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR," in *ICLR*, 2015, pp. 1–14.
- [31] J. He, K., Zhang, X., Ren, S., and Sun, "Deep residual learning for image recognition.," in *IEEE conference on computer vision and pattern recognition*, 2016, p. 770–778.  
<https://doi.org/10.1109/CVPR.2016.90>
- [32] A. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and A. Rabinovich, "Going deeper with convolutions," in *IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [33] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *IEEE conference on computer vision and pattern recognition*, 2017, p. 1251–1258.  
<https://doi.org/10.1109/CVPR.2017.195>
- [34] X. Deng, Y., Luo, P., Loy, C. C., and Tang, "Pedestrian attribute recognition at far distance.," in *22nd ACM international conference on Multimedia, ACM.*, 2014, p. 789–792.  
<https://doi.org/10.1145/2647868.2654966>
- [35] Q. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, "Scalable person re-identification: A benchmark," in *IEEE International Conference on Computer Vision*, 2015, pp. 116–1124.  
<https://doi.org/10.1109/ICCV.2015.133>