Ph.D Forum: Generating Domain and Pose Variations between Pair of Cameras for Person Re-Identification

Asad Munir
University of Udine
asad.munir@uniud.it

Gian Luca Foresti
University of Udine
gianluca.foresti@uniud.it

Christian Micheloni
University of Udine
christian.micheloni@uniud.it

ABSTRACT

Person re-identification (re-id) remains an important task that aims to retrieve a person's images from an image dataset, given a probe image. The lack of cross-view (pose variations) training data and significant intra-class (domain) variations across different cameras make re-id more challenging. To solve these issues, this work proposes a Domain and Pose Invariant Generative Adversarial Network (DPI-GAN) to generate images for both domain and pose variations capture. It is based on a CycleGAN structure in which the generator networks are conditioned on a new pose. Identity and pose discriminators networks are used to monitor the image generation process. These generated images are used for learning domain and pose invariant features to improve the performance of person re-identification.

KEYWORDS

person re-identification, image generation, domain variations , pose variations, camera to camera translation

1 INTRODUCTION

One of the most important and challenging problem in the field of video surveillance is person re-identification (re-id) which aims to match person images with the same identity across non overlapping camera views. In this task, person image encounters many changes independent of the person's identity. These changes include appearance, background (domain variations), viewpoint, pose variations, lightning and occlusions. Wide range of deep learning methods have been proposed to improve the performance of re-id. Generative adversarial network (GAN) [2] is gaining popularity in image generation to increase the re-id performance. Existing GAN based methods consider either domain variations [6] or pose variations [4] to generate new images.

In this work, we propose Domain and Pose Invariant Generative Adversarial Network (DPI-GAN) to generate images by changing both domain and pose in a pair of cameras. The proposed DPI-GAN uses CycleGAN [7] approach to translate images from one domain to another. The generators are conditioned on a new pose to generate image in new domain with given pose. Identity and pose discriminators are used with each generator to preserve the identity and conversion to new pose in the generated images. For each of the two training cycles, the proposed framework trains the two generators and two discriminators. The images are generated from one camera's domain to other camera's domain with a new pose and returning back to original domain with a new pose.

The next section explains the proposed DPI-GAN framework. Section 3 describes experimental results and parameters. Acknowledgement and conclusion are included in last two sections.

ICDSC 2019, September 9–11, 2019, Trento, Italy

ACM ISBN 978-1-4503-7189-6/19/09…$15.00
https://doi.org/10.1145/3349801.3357135

2 DOMAIN AND POSE INVARIANT GENERATIVE ADVERSARIAL NETWORK

Our proposed DPI-GAN aims to generate an image with new pose and domain from an input image and a skeleton pose image. Skeleton pose images are calculated using human pose estimator [1]. Input image and skeleton pose image are concatenated and fed into the generator network to generate the image into given pose. We used CycleGAN [7] approach to train the network for capturing the domain changes between pair of cameras. First Cycle of our framework is shown in Figure 1 which is transferring an image from camera A to camera B and then reconstruct that image back to camera A with a different pose. Second Cycle is the same with starting from camera B to A and the reconstruct to Camera B. With this, we are training the two generators which are generating images from domain A to B and vice versa. Identity and pose discriminators are used to identify real and fake generated images.

2.1 Training

Assume we have $\{a_i^m, x_m\}_{i=1\ldots Mi}^{m=1\ldots M}$ and $\{b_j^n, y_n\}_{i=1\ldots Nj}^{n=1\ldots N}$ persons images in domain A and B respectively, where M and N are the number of images in both domains. i and j are the pose indexes from the total poses M_i and N_j of a person. The skeleton pose images are denoted as P_i and P_j for ith and jth pose respectively. x_m and y_n are the persons identities and for every training sample $x_m = y_n$. The full loss function is denoted as:

$$L = \arg\min_G \max_D \lambda_1 L_{GAN} + \lambda_2 L_{cycle} + \lambda_3 L_{identity} \tag{1}$$

where

$$\begin{aligned} L_{GAN} &= \mathbb{E}_{a, a_i \in \rho, p_i \in \rho_p} \log(D_p(P_i, a_i).D_i(a, a_i)) \\ &\quad + \mathbb{E}_{a \in \rho, p_i \in \rho_P, \hat{a}_i \in \hat{\rho}} \log[(1 - D_p(P_i, \hat{a}_i)).(1 - D_i(a, \hat{a}_i))] \end{aligned} \tag{2}$$

$$L_{cycle} = ||\hat{a}_i - a_i||_1 + ||\hat{b}_j - b_j||_1 \tag{3}$$

$$L_{identity} = ||G_{AB}(b, P_j) - b_j||_1 + ||G_{BA}(a, P_i) - a_i||_1 \tag{4}$$

L_{GAN} is the adversarial loss for first cycle and is calculated in the same way for the other cycle. As we are using two discriminators with each generator so the final output of these discriminators are multiplied to get the final score. ρ,$\hat{\rho}$ and ρ_P denote the distributions

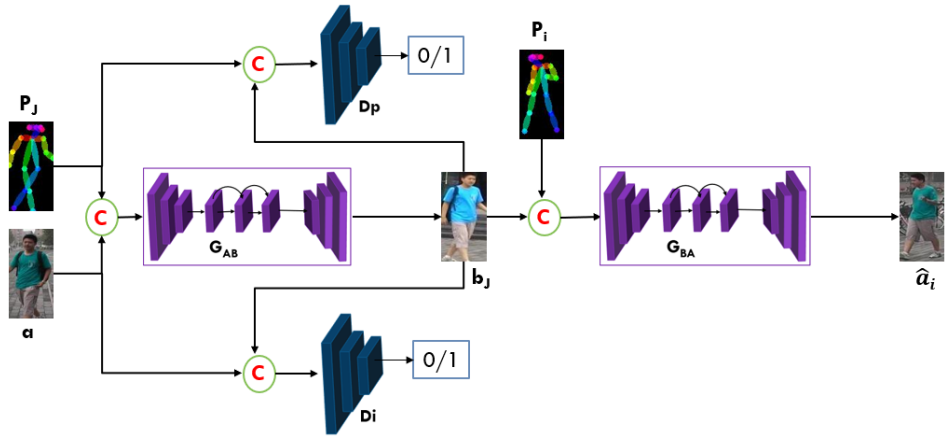

Figure 1: Overview of our framework. Gen A-B and Gen B-A are the generators to transfer from domain A to B and from B to A respectively. Dp and Di are the pose and identity discriminators. C is symbol for concatenation

for real,fake and skeleton pose images. We use least square loss which is more stable [7]. In L_{cycle}, $\hat{a}_i$ and $\hat{b}_j$ are the reconstructed images as shown in figure 1. a_i and b_j are the ground truth images for skeleton poses P_i and P_j. Identity mapping loss is used to preserve the color composition between input and output [7].

(a) **(b)** **(c)** **(d)**

Figure 2: Results generated by the two generators. (a) and (c) are the ground truths from camera 1 (A domain) and camera 6 (B domain) respectively. (b) shows the output of generator $B - A$. The output of generator $A - B$ is shown in (d).

3 EXPERIMENTAL RESULTS

We select camera 1 and camera 6 images from Market-1501 [5] dataset and all the images are resized into 256 x 128. Adam optimizer is used with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Learning rates for generator and discriminator are 0.0002 and 0.0001 respectively. Generator consists of encoder decoder network with 9 ResNet basic blocks and PatchGAN [3] structure is used for all discriminators

3.1 Quantitative Results

The qualitative results of the proposed method are shown in figure 2. Generated images between two camera domains and their ground truths are shown. The inputs are from opposite domain and having different poses in each case.

4 CONCLUSION

We have proposed an image generation method which captures both domain and pose changes for re-id. In contrast with the previous approaches the proposed method merge both these variations in a single network. Generated images with the proposed approach provide domain and pose invariant features for person re-identification. Experimental results prove the image generation with above mentioned variations.

ACKNOWLEDGMENTS

This work was supported by EU H2020 MSCA through Project ACHIEVE-ITN (Grant No 765866)

REFERENCES

[1] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7291–7299.
[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
[4] Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, and Xiangyang Xue. 2018. Pose-normalized image generation for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 650–667.
[5] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*. 1116–1124.
[6] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. 2018. Camera style adaptation for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5157–5166.
[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.