Two-side Network for Person Detection and Person Re-identification

胡 力 博[1] 王濤[2] 渡 辺 裕[3]
Libo HU[1] Tao WANG[2] Hiroshi WATANABE[3]

早稲田大学大学院† 基幹理工学研究科2‡
Waseda University† School of Fundamental Science and Engineering 2‡

Abstract: Person Re-IDentification (ReID) aims to retrieve the images of a person recorded by different surveillance cameras, has been playing a significant role in Intelligent security. However, when it comes to real-time field, ReID still meet difficulties in latency and accuracy. In this paper we proposed a two-side network that meets the demand of ReID request in both off-time and real-time. For edge-side, we propose to set the front-end network to MobileNet and the back-end algorithm to SSD for target detection and then compare with other lightweight networks. For the Cloud-side, we chose the DG-NET as a data augmentation network for ReID tasks. By switching the appearance or structure codes, the generative module can generate high-quality cross-id composed images to let the ReID learning better leverage the generated data.

1. Introduction

Person Re-IDentification (ReID) aims to retrieve the images of a person recorded by different surveillance cameras. Most recent ReID models are based on deep CNNs. Although classification backbones can be used for retrieval as the inputs of both tasks are images, there are still some differences between the ReID task and the classification task. For example, the inputs are all person images which meet those difficulties such as visual angle and post change, illumination changes, objects cover, etc. On the other hand, edge computing allows more computing tasks to take place on the decentralized nodes at the edge of networks. Today many delay sensitive, mission-critical applications can leverage edge devices to reduce the time delay or even to enable real-time, online decision making thanks to their onsite presence.

As a following research to meet the demand of both cloud-side and edge-side ReID task, we propose a two-side network for Person Detection and Person Re-identification

2. Approach

2.1 Cloud-side

With recent progress in the generative adversarial networks (GANs), generative models have become appealing choices to improve ReID network for free. the re-id method combined with GAN learns complementary features from both original images and pose-normalized synthetic images, which can make huge data augmentation to overcome the difficulties in posture differences.

In light of the above observation, DG-NET come as a new network for ReID tasks, which can generate high-quality cross-id composed images, and online fed back to the appearance encoder and used to improve the discriminative module. The framework renders significant improvement over the baseline without using generated data, leading to the state of-the-art performance on several benchmark datasets.

DG-NET introduce a generative module, of which encoders decompose each pedestrian image into two latent spaces: an appearance space that mostly encodes appearance and other identity related semantics; and a structure space that encloses geometry and position related structural information as well as other additional variations. The appearance space encoder is also shared with the discriminative module, serving as a ReID learning backbone.

Table 1: Description of the information encoded in the latent appearance and structure spaces

Appearance Space	Structure Space
Clothing/shoes color,	Body size, hair, carrying,
Texture and style	Pose, background,
Other id-related cues, etc.	Position, viewpoint, etc.

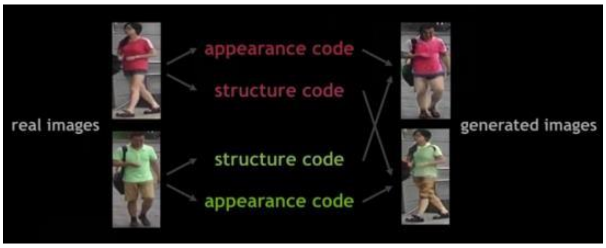

Figure 1: swap learned code between every two images

2.2 Edge-side

The MobileNet model is based on depthwise separable convolutions which is a form of factorized convolutions which factorize a standard convolution into a depthwise convolution and a 1×1 convolution called a pointwise convolution.

The output feature map for standard convolution assuming stride one and padding is computed as:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \qquad (1)$$

Standard convolutions have the computational cost of:

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \qquad (2)$$

The combination of depthwise convolution and 1 × 1 (pointwise) convolution is called depthwise separable convolution which was originally introduced in Depthwise separable convolutions cost:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \qquad (3)$$

which is the sum of the depthwise and 1×1 pointwise convolution. By expressing convolution as a two steps process of filtering and combining we get a reduction in computation of:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \qquad (4)$$

3. Experiments

3.1 Datasets

In the cloud-side experiments, we use Market-1501, The Market-1501 dataset is a dataset for person re-identification, In the edge-side experiments, we use the dataset named PASCAL VOC2007

3.2 Cloud-side experiment

Two experiments are performed in this project. The first one is to test the generating performance of DG-Net, the generated result is showed as below.

Figure 2: success and failed results generated by DG-Net

The second one is to make ReID experiment and compare the result between Res50 and DG-Net.

Table 2: Comparison between Res50, Dense121, and DG-Net

Methods	Rank@1	mAP
[ResNet-50]	88.84%	74.02%
[DG-Net]	94.10%	85.60%

3.3 Edge-side experiment

In the experiment, there are two main scene: inside and outside. And inside scene is mainly detecting people and indoor objects like tableware and furniture, outside scene is aiming at vehicle, people and animal.

Figure 3: Indoor scene and outdoor scene

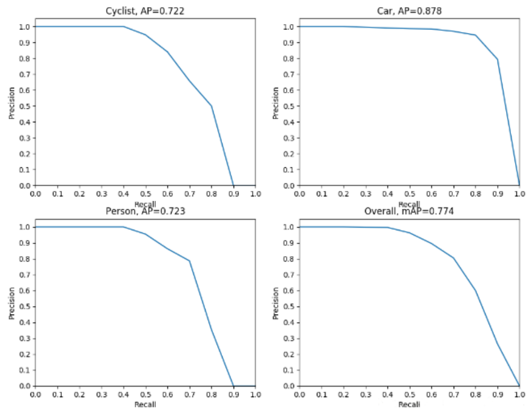

Figure 4: P-R Chart

4. Conclusion

As a two-side network, on the cloud-side, DG-NET does have significant performance in ReID tasks. According the result of generated images by DG-NET, it shows that DG-NET still has room to improve its performance.

MobileNet performance well in detecting both simple and complex scene. So it can be applied in some real tasks on edge devices. But the performance of SSD on small objects is very bad. We will try to find method dealing with the small object detection.

5. Reference

[1] Yutao Huang, When deep learning meets edge computing IEEE10.1109 ICNP.2017.8117585

[2] Longhui Wei, Shiliang Zhang, Wen Gao, Qi Tian Person Transfer GAN to Bridge Domain Gap for Person Re-Identification10.1016/j.patcog.2019.06.006

[3] Zheng Z, Yang X, Yu Z, et al. Joint discriminative and generative learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 2138-2147.

[4] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

WATANABE Lab, Communications and Computer Engineering department, Waseda University.
Shillman Hall 401, 3-14-9 Okubo, Shinjuku, Tokyo, 169-0072
Phone: +81-3-5286-2509, Fax: +81-3-5286-3488
E-mail: foo@example.jp