Student ID: 17251311

Student Name: Zheng Chuanyang

Supervisor Name: Yuen Pong Chi

Report Author: Zheng, C.Y., Ye, M. & Yuen, P.C.

Submit Time: 2019-09-04

Contact Mail: 17251311@life.hkbu.edu.hk

We Need More Data: Cross-Modality Person Re-identification via MSECycleGAN Data Augmentation

Abstract

Nowadays, though single modality person re-identification has a great breakthrough and improve the mAP to 90% on market1501 dataset, cross-modality person re-identification still has several serious challenges in terms of the modality variance of the same person. In this paper, we mainly try to train a cycleGAN which can generate visible images and thermal images in order to help us translate an image in one domain to another domain. The generated images should be the visible image or thermal image of the original persons. Using our methods, we can get more data, which is three times more than the original data.

1. introduction

Person re-identification, which is usually called re-id, tries to build an identity correspondence across different cameras [1]. Similarly, the purpose of cross-modality person re-identification is building an identity correspondence across different cameras and different image style (visible image style, thermal image style, gray image style and so on). Simply, cross-modality re-id can be regarded as an image retrieve. Given a query dataset which contains modality A (such as visible modality) images of different identities', the aim is to retrieve the corresponding images in modality B (such as thermal) images of the identities.

Compared to the performance of single modality re-id, the performance cross-modality re-id is usually poorer. The main reason is the variance of the two different modalities, which increases the difficulty of convolutional neural networks (CNNs) to extract features of images. To solve this problem, the main method is using GAN to translate two modalities to one modality [2]. For example, translating all visible images to thermal images. However, the limitation is that we do not fully use our data because we only use the generated thermal images and the real thermal images.

We now propose a new method for cross-modality person re-id: use GAN to generate more data. For example, generate visible images by thermal images and generate thermal images by visible images. What is more, we can also generate visible images by visible images and generate thermal images by thermal images. By this method, we have more data.

After doing extensive experiments, we believe that the second method is much better than the first method. We get three times images than before to train the re-id model, and the performance improves a lot. The main contribution of this paper is::

1. We can get three times data than before by our model

2. Extensive experiments show that our model is state-of-art

3. Discuss several possible methods to improve our model

2. Related Work

Single Modality Person Re-identification.

The main challenge of single modality person re-id is appearance changes [2], such as camera variance, viewpoint, and environment. To solve the single modality problems, researchers try to generate more images [1,3], such as change a person's pose [1] and dress [3]. In 2019 CVPR paper, these unsupervised learning methods become the main ways to solve single modality person re-id problems.

image transfer.

The main method is using GAN. And there are several famous and powerful GAN, such as Conditional GAN [4], CycleGAN [5] and StarGAN [6]. Using the GANs, we can change the camera style of images, image style (such as visible, thermal) of images and so on. In this paper, we use CycleGAN as the image translation tool

Random Erasing

According to [8], Random Erasing is that "we randomly choose a rectangle region of an arbitrary size, and assign the pixels within the selected region with random values".

Figure 1 Example in [8]

3. Proposed Methods

We divide our training period into two parts: Training Period I and Training Period II.

After finishing Training Period I, we start Training Period II. For example, we want to MSECycleGAN for 200 epochs at Training Period I. Therefore, we train MSECycleGAN for 200 epochs. After that, we begin Training Period II.

In the first training period, we just train a MSECycleGAN. MSECycleGAN is a CycleGAN which has additional MSE loss.

In the second training period, we use MSECycleGAN to triple our data. We will use the generated data and the original data to train the re-id model.

3.1 Training Period I

During Training Period I, we only update MSECycleGAN, and we do not update the Re-id model.

3.1.1 CycleGAN Review

Given two domains: domain A and domain B. $real_A \in \mathrm{A}, real_B \in B$. The aim of cycleGAN is learning a mapping function $\mathrm{G_A}$ than can translate images in domain A to image domain B. There are two adversarial discriminators $\mathrm{D_A}$ and $\mathrm{D_B}$ which are proposed to distinguish whether images are translated from another domain. For example, $\mathrm{D_B}$ must distinguish images translated from domain A. The following are important loss functions of CycleGAN.

For $\mathrm{fake_A}$ and $\mathrm{fake_B}$

$$\mathrm{fake_A} = G_B(real_B)$$

$$\mathrm{fake_B} = \mathrm{G_A}(\mathrm{real_A})$$

Calculate GAN Loss

Discriminator loss:

$$-(logD_{A(real\_A)} + \log(1 - D_A(fake_A)) + logD_{B(real_B)} + \log(1 - D_{B(fake_B)}))$$

Generator loss:

$$-(logD_A((fake_A) + logD_{B(fake_B)})$$

Cycle loss

$$Cycle_A = G_{B(fake_B)}$$

$$Cycle_B = G_{A(fake_A)}$$

$$Cycle_{Consistency_{loss}} = |real_A - Cycle_A|_1 + |real_B - Cycle_B|_1$$

Identity mapping loss

Identity loss:

$$Idt_A = G_{B(real_A)}$$

$$idt_B = G_A(real_B)$$

$$|real_A - idt_A|_1 + |real_B - idt_B|_1$$

3.1.2 MSE loss for CycleGAN

$$Mean: (|real_A - fake_A|_2 + |real_B - fake_B)|_2)$$

Compared with original cycleGAN, we add another constraint: MSE loss for fake images and real images. By this method, we try to reduce the gap between fake images and real images in pixel level, which will help CNN to extract features.

3.2 Training Period II

During Training Period II, we try to use the MSECycleGAN trained in Training Period I to generate more data. **During Training Period II, we do not update MSECycleGAN, and we only update Re-id model.**

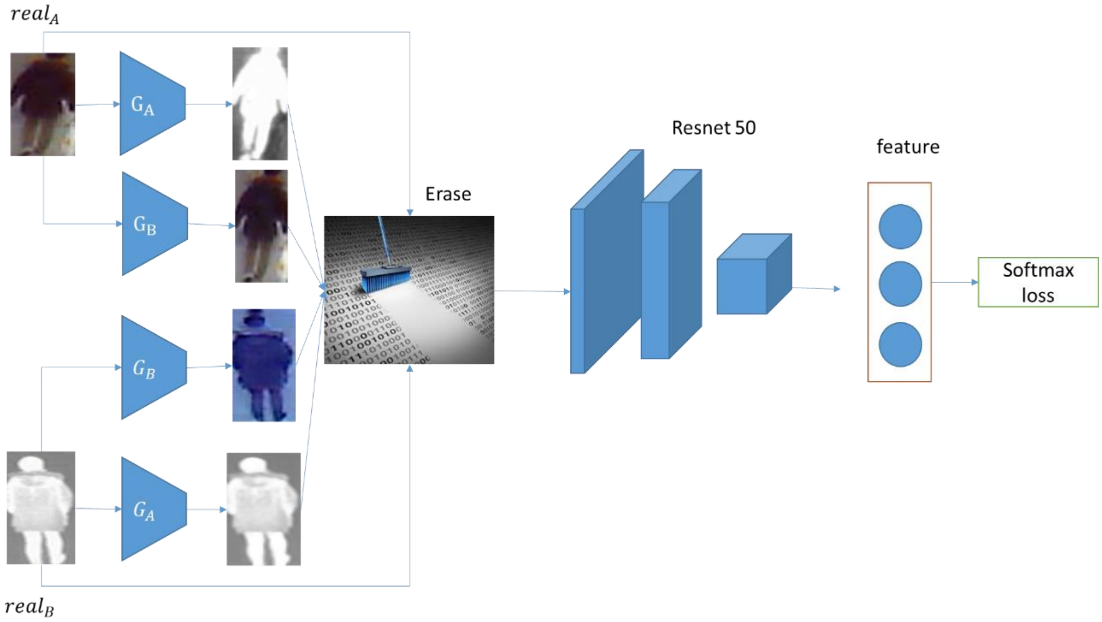

Figure 2 The Re-ID model. During this part, we do not update MSECycleGAN, we only update the Re-id model.

3.2.1 Generate Data

We can use MSECycleGAN to triple our data. Just like generated images in Training Period I, we have idt_A,idt_B,$fake_A$,$fake_B$

$$Idt_A = G_{B(real_A)}$$

$$idt_B = G_A(real_B)$$

$$\mathrm{fake}_A = G_B(real_B)$$

$$\mathrm{fake}_B = \mathrm{G}_A(\mathrm{real}_A)$$

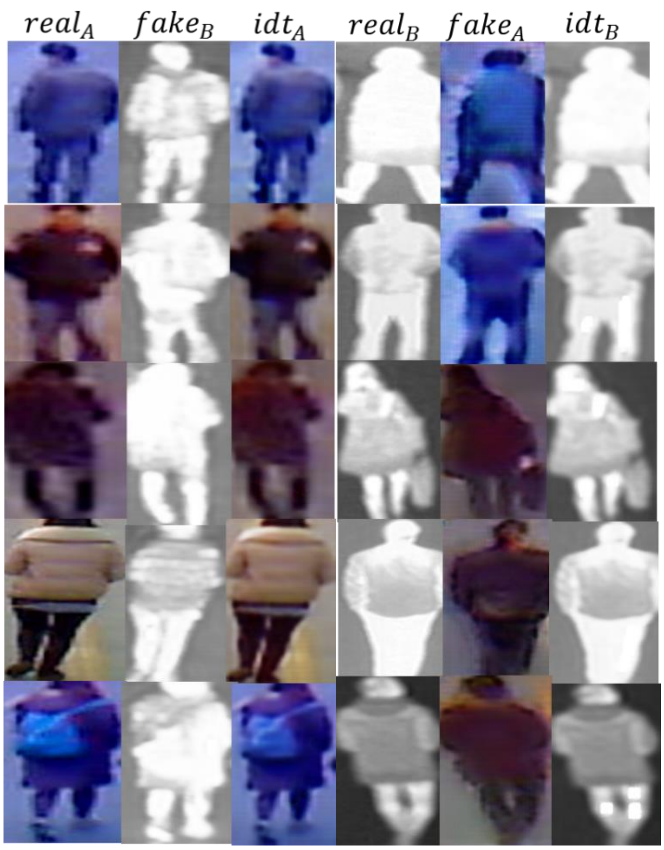

Figure 3. Examples for generated images

3.2.2 Erase Image.

For the real images and generated images, we use the Random Erasing method to erase them. And we give the value $\frac{1}{2}$ to the area.

3.2.3 Softmax loss

Just like normal softmax loss, our loss function is:

$$\text{loss} = -\log(\mathrm{P}(\mathrm{i}|x_i)$$

$\mathrm{x_i}$ is a person' image, and the person's id is i. Given an image, we first get the $\mathrm{P}(\mathrm{i}|\mathrm{x_i})$ by CNN. Then, we calculate $-\log(\mathrm{P}(\mathrm{i}|\mathrm{x_i})$ as a loss. The higher the $\mathrm{P}(\mathrm{i}|\mathrm{x_i})$ is, the better our re-id model is.

4. Experiments

Our baseline comes from [9], but we **remove the Random Erasing method and Padding method in transform part.** Because of the limit of time, we only do experiments on RegDB dataset.

RegDB dataset is “collected by dual camera systems, it contains 412 persons. For each person, 10 different visible light images are captured by a visible camera, and 10 different thermal images are obtained by a thermal camera” [15].

4.1 Transform

In the transform part, we resize the image to 144X144. Then we random horizontal flip image. After that, we random crop the image to 128X128 and change it to Tensor. Finally, we normalize the Tensor with mean[0.5,0.5,0.5] and stand division[0.5,0.5,0.5]

4.2 Training Period I

During Training Period I, the batch size is 16. That is, each time, we have 16 visible images and 16 thermal images. We totally train MSECycleGAN for 70 epochs

4.3 Training Period II

During Training Period II, the batch size is totally 24. This is, we may have 12 visible images and 12 thermal images. We may also have 8 visible images, 8 $\mathrm{idt_A}$ images, 8 thermal images and 8 $\mathrm{idt_B}$ images.

4.3.1Compare with SOTA

Method	mAP	Cmc-1	Cmc-10
LOMO [10]	2.28	0.85	2.47
MLBP[11]	6.77	2.02	7.33
HOG[12]	10.31	13.49	33.22
GSM[13]	15.06	17.28	34.47
One-stream[14]	14.02	13.11	32.98
Two-stream[14]	13.42	12.43	30.36
Zero-Padding[14]	18.90	17.75	34.21
Tone[15]	14.92	16.87	34.03
HCML[15]	20.80	24.44	47.53
BDTR[16]	31.83	33.47	58.42
Proposed $\mathrm{D^2RL}$[2]	44.1	43.4	66.1
Ours	**65.1**	**74.9**	**88.3**

Table 1. Compare with Other state-of-the-are methods

4.3.2 Compare with ourselves

Method	mAP	Cmc-1	Cmc-5	Cmc-10
baseline	48.7	49.1	63.1	69.8
Baseline+Random Erasing	51.0	54.6	68.3	77.2
Baseline+MSECycleGAN Data Augmentation	58.6	68.0	78.3	83.2
Baseline+MSECycleGAN Data Augmentation+Random Erasing	65.1	74.9	84.4	88.3

Table 2 Compare with ourselves

4.4 Discussion

Impact of CycleGAN Data Augmentation

Compared to baseline, the mAP increase 9.9% after adding CycleGAN Data Augmentation. Compared to baseline+Random Erasing, the mAP increases 14.1% after adding CycleGAN Data Augmentation. These two results show that our method is excellently powerful

Compared to baseline, the mAP increases 2.3% after adding Random Erasing method. Compared to baseline+MSECycleGAN Data Augmentation, the mAP increases 6.5% after adding Random Erasing method. 6.5% is about three times more than 2.3% where 58.6% is just 1.2 times more than 48.7%. That means MSECycleGAN Data augmentation can help increase Random Erasing method performance.

5. Future Work

Though our model performance is state of the art, we still have some method to improve our model.

5.1 Triplet Loss

We should extend our loss function. Now, we only have softmax loss. It is necessary that we add triplet loss to our loss function

5.2 Cycle-consistency Image

We discuss cycle-consistency images in Proposed Method part. If we add cycle-consistency images to our model, our data is four times than before.

5.3 Just three times?

During training re-id model, we choose MSECycleGAA which is trained for 70 epochs. However, we can choose another two MSECycleGAN which is trained for 60 epochs and 80 epochs separately. The weights of these three MSECycleGAN should not be the same. Therefore, we can get data that is 7 times than before. What is more, if we choose N MSECycleGAN that are trained for different epochs, we can get (2*N+1) times images than before.

5.4 We Need Infinite-CycleGAN

Well, can we improve CycleGAN? The Answer is Yes! Look at the flow.

$$\mathrm{fake_A} = G_B(real_B)$$

$$\mathrm{fake_B} = \mathrm{G_A}(\mathrm{real_A})$$

$$Cycle_A = G_{B(fake_B)}$$

$$Cycle_B = G_{A(fake_A)}$$

$$Idt_A = G_{B(real_A)}$$

$$idt_B = G_A(real_B)$$

The above is the original CycleGAN flow. Now, we regard rename $\mathrm{real_A}$ as $\mathrm{idt_{A-0}}$ and we rename $\mathrm{real_B}$ as $\mathrm{idt_{B-0}}$ The following is Infinite-CycleGAN

$$\mathrm{fake_{A-1}} = G_B(real_B) = G_B(idt_{B-0})$$

$$\mathrm{fake_{B-1}} = \mathrm{G_A}(\mathrm{real_A}) = \mathrm{G_A}(\mathrm{idt_{A-0}})$$

$$Cycle_{A-1} = G_{B(fake_{B-1})}$$

$$Cycle_{B-1} = G_{A(fake_{A-1})}$$

$$Idt_{A-1} = G_{B(real_A)} = G_B(\mathrm{idt_{A-0}})$$

$$idt_{B-1} = G_A(real_B) = G_A(idt_{B-0})$$

Regard idt_A as $\mathrm{real_A}$, regard $\mathrm{idt_B}$ as new $\mathrm{real_B}$. Then, we go again.

$$\mathrm{fake_{A-2}} = G_B(Idt_{B-1})$$

$$\mathrm{fake_{B-2}} = \mathrm{G_A}(Idt_{A-1})$$

$$Cycle_{A-2} = G_{B(fake_{B-2})}$$

$$Cycle_{B-2} = G_{A(fake_{A-2})}$$

$$Idt_{A-2} = G_{B(real_{A-1})}$$

$$idt_{B-2} = G_A(real_{B-1})$$

Similarly, we can do it for infinite times. And A Question: Why choose idt? Well, you can choose cycle-images or fake-images. But, I have two reasons to choose idt-images.

The first one: we have identity mapping constraint.

The second one. I do an experiment. Baseline+idt-images+real-images, the performance only increases 1.2% (batch size is 16, baseline is mAP is 50.6%). The new real images must belong to a visible domain or thermal domain The performance increase but only increase 1.2%, it means idt-images are not totally the same as the original visible or thermal images but still in a visible domain or thermal domain.

By this method, we can get $\mathrm{fake}_{\mathrm{A-1}}$ … $\mathrm{fake}_{\mathrm{A-N}}$, $\mathrm{fake}_{\mathrm{B-1}}$…$\mathrm{fake}_{\mathrm{B-N}}$, Idt_{A-1}…Idt_{A-N}, Idt_{B-1}…Idt_{B-N}. The data generated is infinite

Acknowledgment: We thank Simon Wang for mang helpful comments. And our work is supported by Prof Sunney Chan. Thanks for his sponsorship.

Reference:

[1]. Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., & Kautz, J. (2019). Joint discriminative and generative learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2138-2147).

[2]. Wang, Z., Wang, Z., Zheng, Y., Chuang, Y. Y., & Satoh, S. I. (2019). Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 618-626).

[3]. Zhong, Z., Zheng, L., Luo, Z., Li, S., & Yang, Y. (2019). Invariance matters: Exemplar memory for domain adaptive person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 598-607).

[4]. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[5]. Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.

[6]. Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8789-8797).

[7]. Zhong, Z., Zheng, L., Zheng, Z., Li, S., & Yang, Y. (2018). Camera style adaptation for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5157-5166).

[8]. Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. arXiv preprint arXiv:1708.04896, 2017.

[9]. https://github.com/L1aoXingyu/reid_baseline

[10]. Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). 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).

[11]. Liao, S., & Li, S. Z. (2015). Efficient psd constrained asymmetric metric learning for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3685-3693).

[12]. Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection.

[13]. Lin, L., Wang, G., Zuo, W., Feng, X., & Zhang, L. (2016). Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1089-1102.

[14]. Wu, A., Zheng, W. S., Yu, H. X., Gong, S., & Lai, J. (2017). Rgb-infrared cross-modality person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5380-5389).

[15]. Ye, M., Lan, X., Li, J., & Yuen, P. C. (2018, April). Hierarchical discriminative learning for visible thermal person re-identification. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[16]. Ye, M., Wang, Z., Lan, X., & Yuen, P. C. (2018, July). Visible Thermal Person Re-Identification via Dual-Constrained Top-Ranking. In *IJCAI* (pp. 1092-1099).