Clear ReID

Tackling Occlusion in Person Re-identification

Authors: Boxiao Pan, Soutik Chakraborty, Pragya Mishra
{bxpan, soutikc, pragmi}@stanford.edu

Introduction

Person Re-identification (Person Re-ID) is a challenging task to retrieve a given person among all the gallery pedestrian images captured across different security cameras. The main challenges for person Re-ID come from large variations on persons such as pose, occlusion, clothes, etc.[1] The task can be even harder when we take occlusion into consideration, which means some part of the person to be identified might be occluded (by objects, other people, etc), which is often the case in reality but wasn't paid much attention by researchers. In this project, we improved on the state-of-the-art Person Re-ID model (i.e., Aligned Re-ID[2]) by incorporating a multi-task loss[3], and trained the new model jointly with non-occluded images and occluded images, which greatly enhanced its performance on occluded person images without losing too much accuracy on non-occluded images.

Baseline Model

We used the current state-of-the-art Aligned Re-ID model as our baseline, whose architecture is shown below in Figure 1.

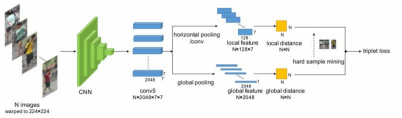

Figure 1: Aligned Re-ID Model Architecture

Feature Extraction:

- Global feature: global pooling on feature map
- Local feature: global pooling horizontally (e.g., reduces a $C \times H \times W$ image to $C \times H) + 1 \times 1$ convolution
- Each image is represented by 1 global feature and H local features

Loss Function:

- Global distance: L_2 distance of the global features
- Local distance: total distance of the shortest path from $(1,1)$ to (H, H) in the distance matrix, where

$$d_{i,j} = \frac{e^{\|f_i - g_j\|_2} - 1}{e^{\|f_i - g_j\|_2} + 1}$$

is the distance between the i-th horizontal part of the first image and the j-th part of the second image.

- Triplet loss: triplet loss is applied by hard sample mining using global distance

Multi-Task Loss

In order to grant our model the ability of discriminating an image to be whether occluded or not, we add another occluded / non-occluded binary classification (OBC) loss, which is defined as follows:

$$L^O(\hat{y}_i', y_i') = \sum_{c=0}^{C} \{y_i' = c\} log \frac{e^{\hat{y}_i'}}{\sum_{c=0}^{C} e^{\hat{y}_c}}$$

Where y_i' is the occlusion label which equals to 0 if the i-th image is occluded and 1 otherwise. We add another fully-connected layer to predict this label.
Combined with the original triplet loss, we obtain the final multi-task loss as:

$$L = \alpha L^T + (1 - \alpha) L^O$$

Where L^T denotes the triplet loss and α is a hyperparameter which balances the two losses.

Occlusion Simulator

Since we need to train our model using occluded images but currently there is not a publicly available occluded person image dataset, we created an occlusion simulator which first averaged all the pixel values of an input image and created a 50×50 occlusion patch with that value, then applied it on the 256×128 image at a random location. In this way we get an occluded version of the input image.

Data Description

Market1501:
We used Market1501 as our base dataset, which is a widely used Person Re-ID benchmark dataset and contains 32,668 images of 1,501 labeled persons of six camera views. There are 751 identities in the training set and 750 identities in the testing set.

Occluded Market1501:
We applied our occlusion simulator on the original Market1501 dataset to get the occluded version of it. And we used it together with the original non-occluded version to train our model. Note that only the query images get occluded while all gallery images are kept non-occluded. We give some examples of the non-occluded and occluded images in Figure 2.

Experimental Results

In this section we present our results in both quantitative (accuracy) and qualitative (heatmap) forms.

Model	Market1501	Occluded Market1501
Aligned Re-ID	**95.61%**	18.2%
Aligned Re-ID + OBC	77.40%	**30.94%**

Table 1: Quantitative experimental results. OBC stands for the new model with OBC loss. The reported numbers are all CMC10 scores.

Figure 2: Examples of occluded and non-occluded image. The left column are non-occluded versions while the right ones are occluded.

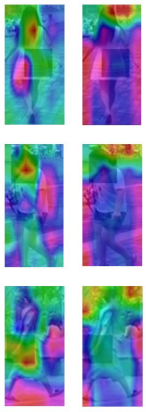

Figure 3: Heatmaps from the last conv layer using the occluded images from Figure 2. Left side is from the original Aligned ReID network while the right side is from our new network.

Result Analysis:

- From Table 1, our new model achieves a big performance boost on occluded images, while still achieving good accuracy on non-occluded images.
- From the heatmaps, we can clearly see that by incorporating the OBC loss and training jointly on both occluded and non-occluded images, our new model puts less attention on occluded areas, thus making it more robust to occlusions.

Conclusion and Future Work

As seen in the experimental result section, after incorporating OBC loss, our model performed much better on occluded images than the original Aligned Re-ID model without dropping too much accuracy on non-occluded images. Also from the heatmaps, we can see that our model has learnt to put less attention on the occluded areas, which is exactly what we expected. The occlusions are applied randomly and are of the average value of all pixels. This confuses the model so much as to where to pay attention to and where not to. Given this, the results are indeed very promising. In the future, we will carry out the following ideas:

- Adopt a GAN instead of a single fully-connected layer to predict the occlusion label, which should be better at helping network learn features that are more robust to occlusion.
- Test on more datasets (still using our occlusion simulator).
- Finally, gather occluded images in a realistic setting rather than automatically generating the occlusions.

References

[1] Learning Discriminative Features with Multiple Granularity for Person Re-Identification. Wang et al. https://arxiv.org/pdf/1804.01438.pdf
[2] AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. Zhang et al. https://arxiv.org/pdf/1711.08184.pdf
[3] Occluded Person Re-identification. Zhuo et al. https://arxiv.org/pdf/1804.02792.pdf