Efficient and Deep Person Re-Identification using Multi-Level Similarity Supplementary Material

Yiluan Guo, Ngai Man Cheung
Singapore University of Technology and Design
yiluan_guo@mymail.sutd.edu.sg, ngaiman_cheung@sutd.edu.sg

1. trade-off between classification and ranking

For the model trained with combined loss in Eq.6, the tradeoff factor λ in Eq.7 between $s_{softmax}$ and d is studied, shown in Figure 1 with different value from 0 to 1. It can be observed that even when $\lambda = 0$, the top-1 accuracy, 77.2%, is higher than ours-cls, which proves that the ranking loss helps the binary classification converge to a better optimal than itself alone. The top-1 accuracy reaches the highest value when $\lambda = 0.2$. When λ keeps increasing, the top 1 accuracy decreases apparently, which implies that the information provided by ranking loss is likely to be less important than the binary classification. Therefore, we can conclude that the combination of two losses is necessary.

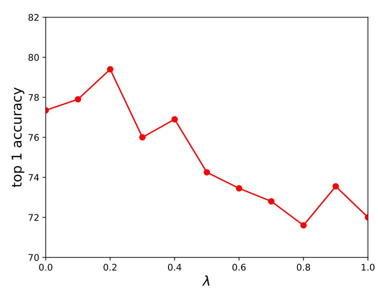

Figure 1. Top 1 accuracy on CUHK03 detected dataset with different λ.

2. Effect of regularization on transform parameters

The effect of regularization on the 6 transform parameters is studied. We consider three cases: 1) removing all the constraints; 2) ignoring the rotation situations by keep the r_w and r_h 0; 3) adopting regularization the same as [1]. The results on CUHK03 detected datasets are show in Table 1. Our method outperforms all the three cases. In (1), we found that the extracted parts fell out of the original images easily, which is definitely what we want to avoid. Serious rotation situation is rarely seen in real world application but rotating to a small extent is likely to exist due to the cameras, which is proven by the 2% improvement between our method and (2). If we adopt the same regularization as BD-LatPart, the performance drops drastically in (3), which justifies the superiority of our dividing-and-regularizing strategy.

Table 1. CMC results on CUHK03 detected datasets for different regularization on transform parameters.

Method	top-1	top-5	top-10
Case 1	77.25	94.35	96.65
Case 2	77.80	94.55	97.20
Case 3	72.10	93.00	97.10
Ours	**79.45**	**94.70**	**97.90**

3. Visualization

Figure 2 depicts some learned semantic parts. Here we show the learned parts of the upper region. For query and positive images, our model puts emphasis on the black coat with white logos. For the negative, our model puts emphasis on the more distinguishable white shirt.

Figure 2. Learned parts for query, positive, negative images using STN-based attention.

References

[1] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.