Learning Deep Context-aware Features over Body and Latent Parts for Person Re-identification
Supplementary Materials

Dangwei Li[1,2], Xiaotang Chen[1,2], Zhang Zhang[1,2], Kaiqi Huang[1,2,3]
[1]CRIPAC & NLPR, CASIA [2]University of Chinese Academy of Sciences
[3]CAS Center for Excellence in Brain Science and Intelligence Technology
{dangwei.li, xtchen, zzhang, kaiqi.huang}@nlpr.ia.ac.cn

1. Market1501 dataset

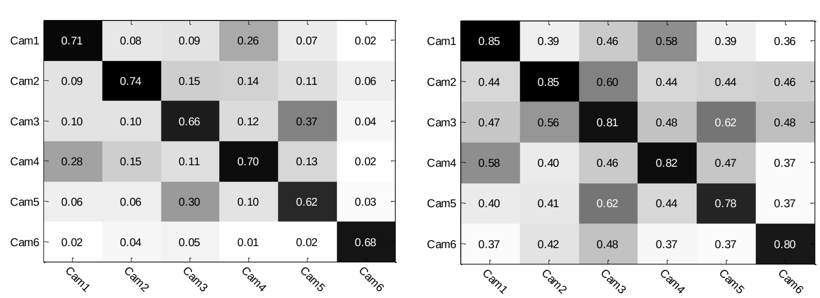
To further understand the results on Market1501 [8], we show mean Average Precision (mAP) and Rank-1 identification rate between camera pairs in Figure 1 and Figure 2. Compared to the BOW methods, the proposed method improves mean mAP and Rank-1 identification rate between camera pairs by 35.09% and 40.01% respectively. In addition, we show some searching results with different query images in Figure 3. The dataset is challenging and the returned images have very similar appearances and some pedestrians have large backgrounds and occlusions. For the query image in first row of Figure 3, even though the query person has large occlusions and some groundtruth images have large backgrounds, our proposed method can still return the right results. This shows the effectiveness of our proposed method.

2. CUHK03 dataset

CUHK03 [3] is one of the largest person re-identification datasets. It provides two types of pedestrian bounding boxes, including detected and manually annotated. In this paragraph, we show the overall Cumulated Matching Characteristics (CMC) on both detected and labeled datasets in Figure 4. For the GateSCNN [5] in Figure 4(a), we use the singe-query results to approximate the single-shot results. The DGD [6] is trained using multiple datasets. In this paper, we use the results trained on single CUHK03 dataset for fair comparison. Compared with the state-of-the-art approaches, our proposed method improves the results by a large margin.

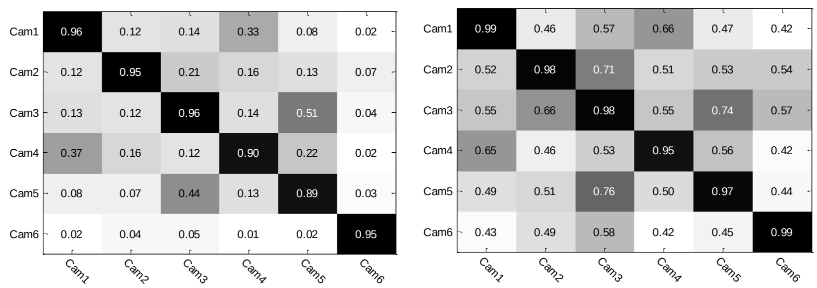
3. MARS dataset

MARS [7] is currently largest sequence-based person re-identification dataset. In this paragraph, experimental results of CMC curves under single query and multiple query are shown in Figure 5. We show the results of our proposed method using



(a) BOW [8] (b) Our-Fusion

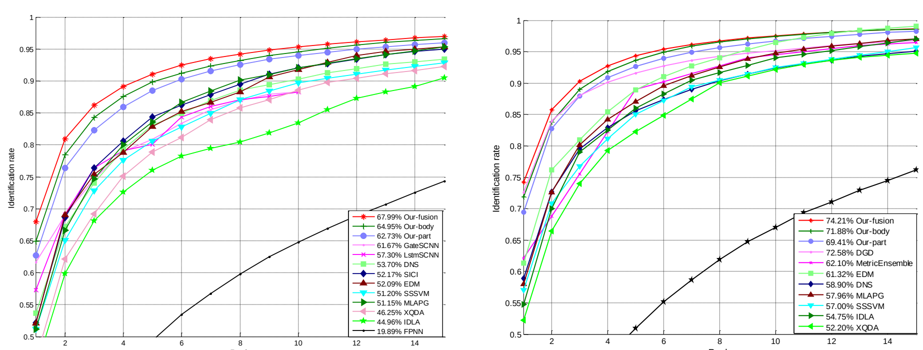
Figure 1. mAP between camera pairs using BOW and our fusion model respectively. The average cross camera mAP of BOW and our fusion are 10.51% and 45.60% respectively.



(a) BOW [8] (b) Our-Fusion

Figure 2. Rank-1 identification rate between camera pairs using BOW and our fusion model respectively. The average cross camera Rank-1 identification rate of BOW and our fusion are 13.72% and 53.73% respectively.

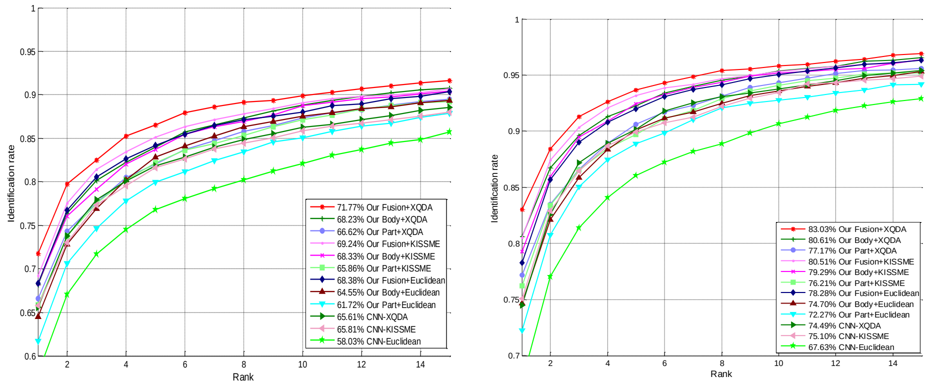
Figure 3. Query results using our fusion model on Market1501 dataset. The first column are query images, and the rest images in each row are corresponding searching results. Images with green boxes are the same person with the probe image. These results are generated using single query protocol. Best viewed in color.



(a) CUHK03 detected dataset (b) CUHK03 labeled dataset

Figure 4. Experimental results on CUHK03 datasets, including (a) detected dataset and (b) labeled dataset. Best viewed in color.

three metric learning algorithms, including Euclidean metric, Keep It as Simple and straightforward Metric (KISSME) [1], and Cross-view Quadratic Discriminant Analysis (XQDA) [4]. Compared with the popular AlexNet [2], which is denoted as CNN in Figure 5, the proposed method has shown much better performance in both single query and multiple query cases.



(a) Single query results

(b) Multiple query results

Figure 5. Experimental results on MARS datasets, including (a) single query and (b) multiple query. Best viewed in color.

References

[1] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Proc. CVPR*, pages 2288–2295, 2012.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. NIPS*, pages 1097–1105, 2012.

[3] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proc. CVPR*, pages 152–159, 2014.

[4] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *Proc. CVPR*, pages 2197–2206, 2015.

[5] R. R. Varior, M. Haloi, and G. Wang. Gated siamese convolutional neural network architecture for human re-identification. In *Proc. ECCV*, pages 791–808, 2016.

[6] T. Xiao, H. Li, W. Ouyang, and X. Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proc. CVPR*, 2016.

[7] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *Proc. ECCV*, pages 868–884, 2016.

[8] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *Proc. ICCV*, pages 1116–1124, 2015.