Supplementary Material: Interpretable and Generalizable Deep Image Matching with Query-adaptive Convolutions

Anonymous Author(s)
Affiliation
Address
email

1 Random Block Data Augmentation

A Random Block (RB) module is implemented for data augmentation of the QAConv training, similar to the Random Erasing (RE) method [5]. In Zhong et al.'s implementation of RE, the probability of random erasing is 0.5, the target erasing area is randomly sampled from 0.02 to 0.2 of the image area, the target aspect ratio is randomly sampled from 0.3 to 3, and this is tried at most 100 times to generate a reasonable region for erasing. In contrast, in our implementation of the random block module, we always block a random portion of the image. We use a square block, with the size randomly sampled from 0.2 to 1.0 of the width of the image. Then the square block is filled with white pixels with RGB values (255, 255, 255). In the experiments, we also set the probability of random erasing to 1.0, and observed a better performance. Note that with a simple square block, there is no need to sample multiple times of areas and aspect ratios and check the validity, and hence the generation process is more efficient.

Comparison of the random erasing method of [5] and the random block method implemented here is shown in Table 1 for cross-dataset evaluation and Table 2 for within-dataset evaluation. From the results it is clear that the implementation of the random block module in this work is better than the random erasing method, and hence we use the new implementation in the training of the proposed QAConv algorithm.

2 Results on the CUHK03 dataset

The CUHK03 dataset [2] includes 13,164 images of 1,360 pedestrians. It is captured with six surveillance cameras. Each person is observed by two disjoint camera views and has an average of 4.8 images in each view. Apart from manually cropped pedestrian images, samples detected with a state-of-the-art pedestrian detector is also provided. This is a more realistic setting and poses problems like misalignment, occlusions and body part missing. Images were obtained from a series of videos recorded over months. Illumination changes were caused by weather, sun directions, and shadow distributions even within a single camera view.

In the experiments, we adopted the new evaluation protocol provided in [4], denoted as CUHK03-NP. That is, images of 767 identities are used for training, and the remaining images of 700 identities are used for testing. We directly applied the learned models on the DukeMTMC-reID and the Market-1501 datasets for the cross-dataset evaluation on the CHUK03 dataset. The results for the detected subset are reported in Table 3. As can be observed, the proposed QAConv method without transfer learning performs better than a recent transfer learning method PUL [1]. QAConv is not as good as another unpublished method UDARTP [3]. However, with the help of re-ranking, QAConv+RR performs comparable to UDARTP under Market→CUHK03, and much better than UDARTP under Duke→CUHK03. Note that re-ranking computed on the fly is much more efficient than transfer learning which requires training on the target dataset. Also note that since the CUHK03 dataset does

Table 1: Comparison of cross-dataset evaluation results (%).

Method	Data Augmentation	Duke→Market		Market→Duke	
		Rank-1	mAP	Rank-1	mAP
QAConv	Random Erasing [5]	61.9	30.8	49.4	29.3
QAConv+RR	Random Erasing [5]	66.5	47.9	55.5	45.3
QAConv+RR+TLift	Random Erasing [5]	78.8	54.4	77.8	58.9
QAConv	Random Block	61.2	30.5	54.2	33.3
QAConv+RR	Random Block	66.6	50.3	61.4	52.5
QAConv+RR+TLift	Random Block	**79.6**	**57.6**	**82.6**	**66.1**

Table 2: Comparison of within-dataset evaluation results (%).

Method	Data Augmentation	Market-1501		DukeMTMC-reID	
		Rank-1	mAP	Rank-1	mAP
QAConv	Random Erasing [5]	93.6	81.7	83.8	71.4
QAConv+RR	Random Erasing [5]	94.8	92.8	87.7	86.3
QAConv+RR+TLift	Random Erasing [5]	97.6	94.0	95.0	91.1
QAConv	Random Block	93.7	83.3	88.3	76.7
QAConv+RR	Random Block	95.4	94.1	91.5	90.5
QAConv+RR+TLift	Random Block	**97.7**	**95.0**	**96.6**	**93.8**

not provide temporal information of the person images, the proposed TLift method cannot be applied.

3 Training and Evaluation Time

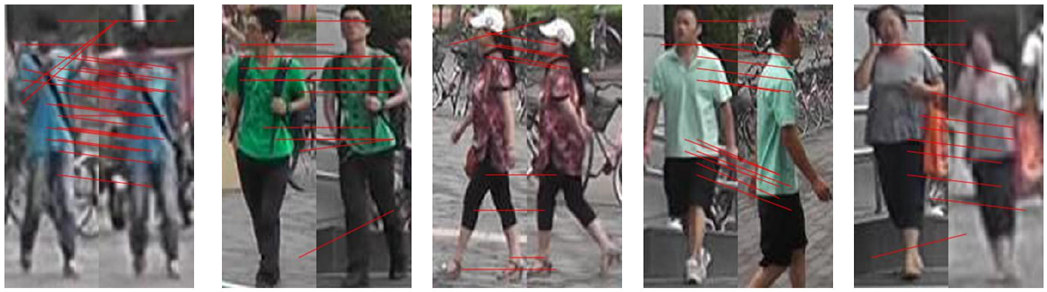
We train the QAConv network on a NVIDIA DGX-1 server, with 4 V100 GPU cards. With the backbone network Resnet152 and input image size 384×128, the training time on the Market-1501 dataset is about 3.76 hours, and 4.95 hours on the DukeMTMC-reID dataset. The evaluation on the Market-1501 dataset requires about 395 seconds, and it is about 376 seconds on the DukeMTMC-reID dataset.

4 Qualitative Analysis

The unique characteristic of the proposed QAConv method is its interpretation ability of the matching. Therefore, we show some qualitative matching results in Fig. 1 for a better understanding of the proposed method. The model shown here is learned on the Market-1501 training data, and the evaluations are done on the query subsets of the Market-1501 and DukeMTMC-reID datasets. Results of both positive pairs and hard negative pairs are shown. It can be observed that the proposed method is able to find correct local correspondences of positive image pairs, even if there are notable misalignments or pose/viewpoint changes. Besides, for hard negative pairs, the maching of QAConv still appears to be mostly reasonable, by linking visually similar parts or even the same person (may be ambiguously labeled).

Table 3: Comparison of state-of-the-art cross-dataset evaluation results (%) on the detected subset of the CUHK03 dataset with the CUHK03-NP protocol. Transfer learning methods used the training set of the CUHK03 dataset.

Method	Publication	Transfer learning	Duke→CUHK03		Market→CUHK03	
			Rank-1	mAP	Rank-1	mAP
PUL [1]	TOMM 2018	✓	5.6	5.2	7.6	7.3
UDARTP [3]	arXiv 2018	✓	11.1	12.4	21.6	**23.8**
QAConv			10.9	8.9	15.8	13.0
QAConv+RR			**15.7**	**16.4**	**21.9**	23.5



score=0.66 score=0.66 score=0.65 score=0.65 score=0.62

(a) Positive pairs of samples on the Market-1501 dataset

score=0.58 score=0.53 score=0.53 score=0.53 score=0.49

(b) Negative pairs of samples on the Market-1501 dataset

score=0.69 score=0.65 score=0.64 score=0.64 score=0.56

(c) Positive pairs of samples on the DukeMTMC-reID dataset

score=0.70 score=0.64 score=0.61 score=0.57 score=0.55

(d) Positive pairs of samples on the DukeMTMC-reID dataset

Figure 1: Examples of qualitative matching results by the proposed QAConv method.

References

[1] H. Fan, L. Zheng, C. Yan, and Y. Yang. Unsupervised person re-identification: Clustering and fine-tuning. *TOMM*, 14(4):83, 2018.

[2] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[3] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *arXiv preprint arXiv:1807.11334*, 2018.

[4] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1318–1327, 2017.

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