Supplementary Materials of "BV-Person: A Large-scale Dataset for Bird-view Person Re-identification"

Cheng Yan[1]*, Guansong Pang[2]*, Lei Wang[4], Jile Jiao[3]†, Xuetao Feng[3], Chunhua Shen[5], Jingjing Li[6]
[1]Tianjin University [2]The University of Adelaide [3]Alibaba Group
[4]University of Wollongong [5]Monash University [6]University of Alberta

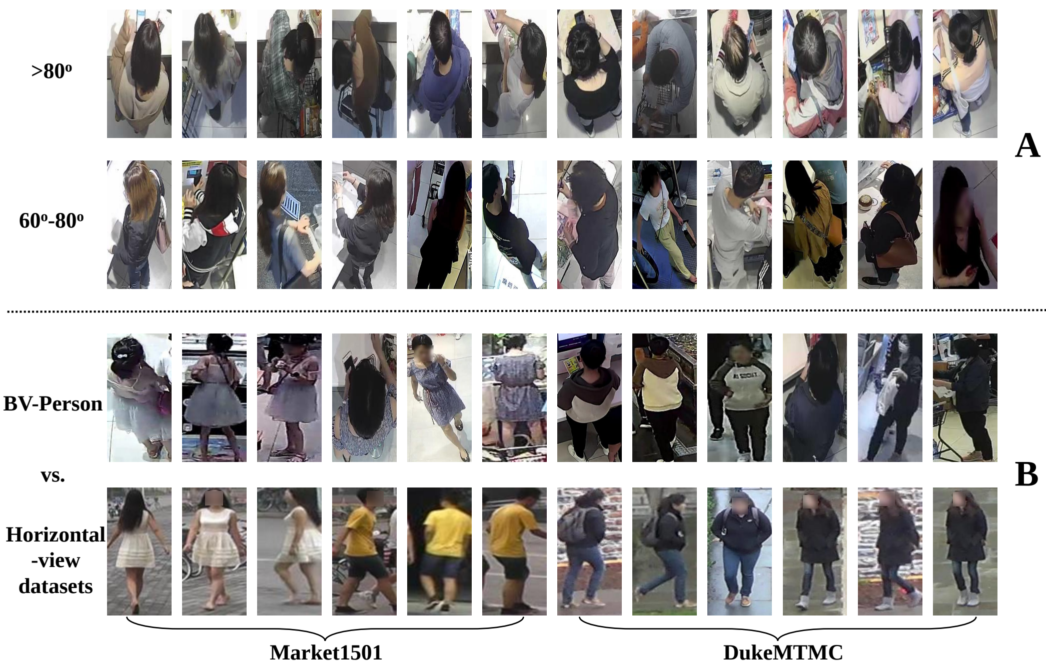

Figure 1: (A) Bird-view image examples from the proposed BV-Person dataset OPReID dataset and (B) Illustrative comparison of BV-Person and two widely-used person ReID benchmarks, Market1501 and DukeMTMC.

1. The Proposed Dataset: BV-Person

This section provides more image examples from our proposed BV-Person dataset for bird-view person ReID. In Figure 1(A), the first two rows demonstrate the images of different identities with diverse appearances captured by a bird-view angle greater than 80^o and between 60^o and 80^o, respectively. Figure 1(B) shows an illustrative comparison of BV-Person and two horizontal-view person ReID benchmarks – Market1501 [5] and DukeMTMC [6] – that are widely used in general ReID tasks. BV-Person provides identity images with diverse dressing styles from summer shirts to winter coats in both of the bird's-eye view and the horizontal view, while Market1501 and DukeMTMC contain images of identities with exclusively either summer or winter dressing captured from the horizontal view only.

2. Implementation Details of Our Model

The implementation of our model is built upon the publicly available source code of the widely-used state-of-the-art model, FastReID [2], by plugging in the three proposed network layers. Specifically, similar to most of the competing methods, we adopt the original Resnet-50 as the network backbone. We replace the self-attention layers in [2], i.e., the 2/3 SA layers in stage 2/3 of Resnet-50, with our proposed Cross Attention (CA) layers, and then add the Multi-scale Attention (MA) layer and Feature Reconstruc-

Table 1: Updated results. '-M': the dataset with all faces masked; '-B': bird-view images in both query and gallery sets.

Method	80H		80H-M		80E		80E-M		60H		60H-B		60E		60E-B		Time
	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	train/test
ABD	11.8	7.3	11.0	7.1	41.5	29.6	41.2	29.6	20.1	15.6	21.8	16.5	51.8	35.5	52.1	35.6	3,360m/112ms
AGW	26.3	19.0	26.3	19.1	54.6	40.3	54.7	40.2	44.1	28.3	48.0	30.3	67.8	52.9	69.6	54.0	235m/21ms
HOReID	23.3	19.3	23.1	19.2	35.7	25.0	35.7	24.9	23.3	19.3	22.5	18.6	53.6	33.5	55.1	35.1	602m/¿1000ms
Baseline*	27.6	19.9	28.0	20.0	55.1	40.6	55.1	40.7	45.0	29.7	48.1	30.5	69.9	54.6	72.5	55.7	245m/21ms
Ours	**31.8**	**22.7**	**31.9**	**22.8**	**64.4**	**50.2**	**64.5**	**50.2**	**50.1**	**33.9**	**52.3**	**35.1**	**75.4**	**60.0**	**78.8**	**62.3**	250m/22ms

tion (FR) layer into the end of the backbone. The GeM pooling [4, 2] and batch normalization are used to obtain the final global feature. The learning rate, training epochs, and batch-size are, by default, set to 0.00035, 60 and 64, respectively. Adam [3] is used as our optimizer. All the experiments are done using a single machine with 2 GTX-1080 GPUs.

3. Additional Experiment Results

3.1. Bird-view Images Presented in Both Query and Gallery Sets

We create two more datasets 60H-B (and 60E-B) that contain 60-degree bird-view images in both of the query and gallery sets. Particularly, we keep one image of each person in the query set and put the rest into the gallery set in 60H (60E) to create 60H-B (60E-B). This guarantees bird-view images in both of the query and gallery sets in 60H-B (60E-B), resulting in datasets with less challenges than that as in 60H (60E). This is demonstrated by the results in Table 1 where the ReID models perform generally much better on 60H (60E) than on 60H-B (60E-B).

3.2. Effects of Using Masked Faces

As shown by some recent studies [1], the faces are not distinguishable features in ReID tasks due to the use of low-resolution cameras in real applications. They showed that some methods can even perform better using images with masked faces than using original images. We had similar empirical findings in our datasets, as shown in Table 1, where 80H-M/80E-M denotes the results of using 80H/80E with all faces masked. The results show clearly that masking faces can hardly affect the performance.

3.3. Computational Cost

The runtimes of our model and the best contenders from four different approaches are shown on Table 1. Our method runs comparably fast to the most efficient counterparts in both training and inference.

References

[1] Julia Dietlmeier, Joseph Antony, Kevin McGuinness, and Noel E O'Connor. How important are faces for person re-identification? In *ICIP*, pages 6912–6919. IEEE, 2021. 2

[2] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. Fastreid: A pytorch toolbox for general instance re-identification. *arXiv preprint arXiv:2006.02631*, 2020. 1, 2

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[4] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. 2

[5] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1116–1124, 2015. 1

[6] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Int. Conf. Comput. Vis.*, pages 3754–3762, 2017. 1