Supplementary Materials of “Occluded Person Re-Identification with Single-scale Global Representations”

Cheng Yan[1]*, Guansong Pang[2*†], Jile Jiao[3] Xiao Bai[1†], Xuetao Feng[3], Chunhua Shen[4]
[1]Beihang University [2]The University of Adelaide [3]Alibaba Group [4]Monash University

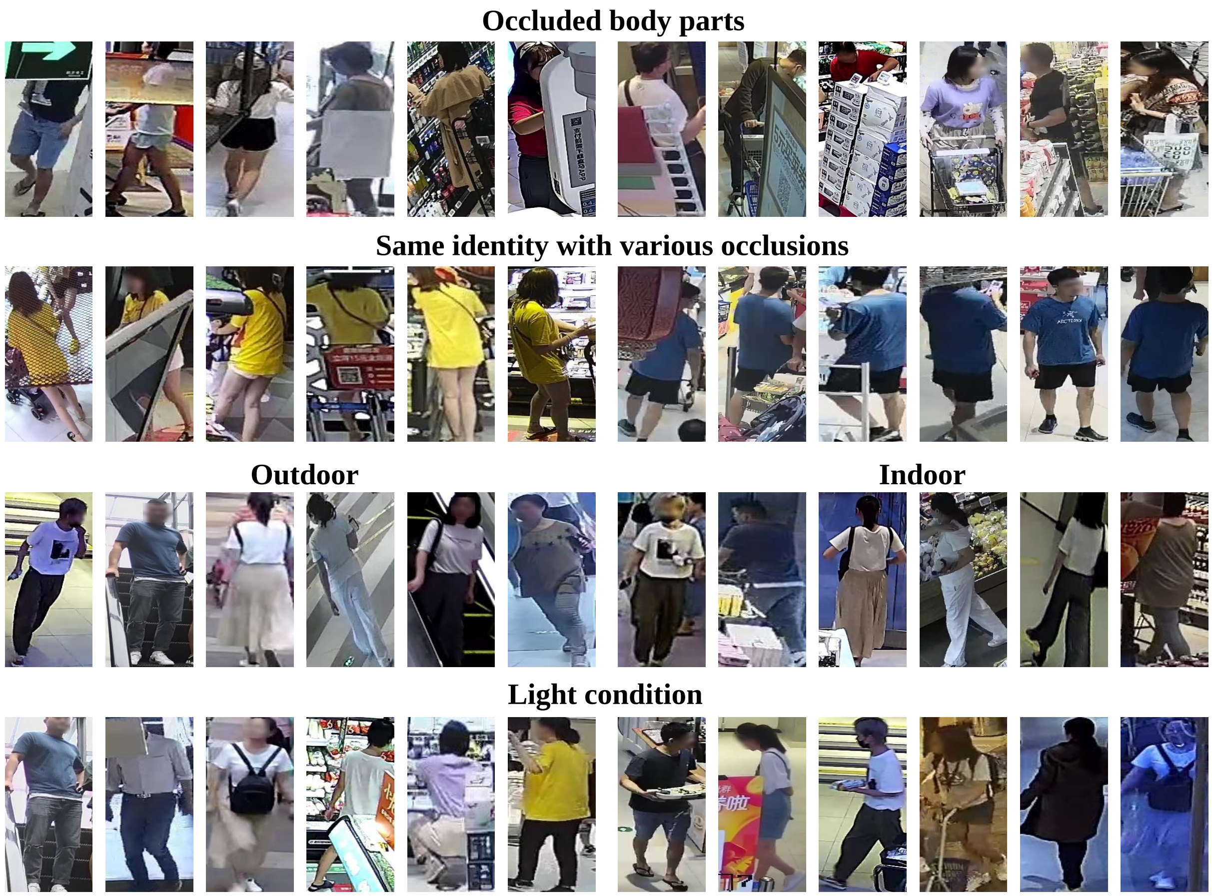

Figure 1 – Image examples from the proposed OPReID dataset for occluded person ReID.

1. Dataset

Our OPReID dataset contains highly diverse occlusions at different body parts with various light conditions and background inside and outside the supermarkets and shopping malls. As shown in Figure 1, the first row demonstrates the diverse occlusions and occluded body parts of OPReID. The occlusions include exit signs, nets, air orbits, commodities, carts, etc. The occluded body parts are from head to foot. Moreover, as illustrated in the second row, each identity has several images with diverse occlusions at different body parts. The last two rows show the different scenes inside and outside the supermarkets and shopping malls, i.e., squares, parking lots, and crossroads, with various light condition from natural light to shimmering light. These are some unique and important characteristics that existing occluded person ReID datasets do not have, overcoming issues of monotonous occlusions and scenes.

Table 1 – R-1, R-5, R-10 and mAP on three occluded ReID datasets and runtime results on OPReID.

Year	Method	Backbone	OPReID				O-Duke				O-ReID				Runtime (seconds)	
			R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	Training	Testing
2019	FPR	ResNet-50	48.7	51.1	53.0	46.6	52.4	67.9	73.1	41.7	78.3	-	-	68.0	4,013	8,433
2019	PGFA		48.1	53.2	55.3	48.3	51.4	-	-	37.3	57.1	-	-	56.2	11,435	750
2020	HOReID		47.6	48.5	50.7	46.0	55.1	-	-	43.8	80.3	-	-	70.2	39,126	199
2020	Baseline		61.1	70.0	73.1	62.1	62.8	75.8	81.0	51.9	68.2	80.3	84.5	63.4	8,460	139
	Ours	ResNet-50	65.8	74.1	77.5	67.2	69.0	81.9	85.1	57.2	78.5	87.6	91.3	72.9	10,870	140
	Ours w\o RE	ResNet-50	65.1	73.8	77.0	66.5	68.2	81.0	84.6	56.6	77.9	87.0	90.8	72.3	10,230	140
	Ours	ResNet-101	66.2	74.6	78.3	67.9	69.5	82.7	86.6	57.8	78.5	88.0	91.8	73.1	12,961	177

2. Additional Empirical Results

2.1. Using Other Backbones

Our model can also work well with other CNN structures as the backbone, such as ResNet-101. For example, as shown in Table 1, our model is able to obtain better performance by using a deeper network backbone, ResNet-101. The main reason for using ResNet-50 in our experiments is that almost all existing deep ReID methods use this network as the backbone in their feature learning. We use the same backbone setting as these existing methods to exclude the variation of the backbone among the methods, so that we can have a genuine effectiveness evaluation of our proposed modules.

2.2. R-5 and R-10 Performance

The R-5 and R-10 results on occluded ReID datasets are presented in Table 1. Compared to state-of-the-art occluded ReID models, our model obtains similar superiority in the R-5 and R-10 performance to that in R-1.

2.3. Occlusion-based Data Augmentation

The proposed occlusion-based data augmentation, i.e., the synthesis of RE and BcE, is to simulate occlusions in real images and augment occluded images. Particularly, RE that erases random small image patches well complements the proposed BcE augmentation that erase a striping part of the image; their combination helps produce images with occlusions at diverse body parts of different size. This enables better feature learning for occluded ReID, as shown by the results of 'Ours' and 'Ours w\o RE' in Table 1 above.

2.4. Runtime

Our model can run considerably faster than SOTA occluded ReID models in the training and/or inference stage, such as FPR, PGFA and HOReID, as shown by the runtime on the OPReID dataset in Table 1 (all experiments are done under the same environment: Xeon E5@2.5GHz and 2 GTX-1080 GPUs). Our model uses the same backbone as the light-weight general ReID model – Baseline – but it takes more training time due to the inclusion of the proposed data augmentation (CBE) and several other modules (i.e., DNL, RP and ME) that enable our model to achieve substantial improvement over Baseline (see Table 5 in p.8). FPR is a feature pyramid matching-based method that has efficient training but is very computationally costly during inference.