Do Not Disturb Me: Person Re-identification Under the Interference of Other Pedestrians (Supplementary Material)

Fig. 1. Examples in PI-PRW (a) and PI-CUHK-SYSU (b).

1 PI-PRW and PI-CUHK-SYSU

Figure 1 shows the examples of our constructed dataset (*i.e.* PI-PRW and PI-CUHK-SYSU), which are derived from the PRW and CUHK-SYSU datasets, respectively. PRW is a multi-camera dataset recorded outdoors at the Tsinghua University campus with 6 synchronized cameras. It is annotated every 25 frames (1 second in duration) with manually drawn bounding boxes and person ID information on each frame. PRW is divided into a training set with 5, 704 frames and a test set with 6,112 frames. As for CUHK-SYSU, it is collected from street snaps and movies, which is densely annotated with all the 96, 143 pedestrians bounding boxes and 8,432 labeled identities. It is divided into a training set with 11,206 frames and a test set with 6,978 frames. We use the off-the-shelf detector (*i.e.* Faster R-CNN) to perform pedestrian detection on both datasets. Then we select the bounding boxes with multiple pedestrians, following the selection criterion as mentioned in Subsection 4.1.

2 Retrieval results

Figure 2 shows the retrieval results, using PISNet, under the setting of PI Re-ID. The correct results in Figure 2(a) show the targets, who are interfered by other pedestrians. In Figure 2(b), the targets are even occluded by other pedestrians.

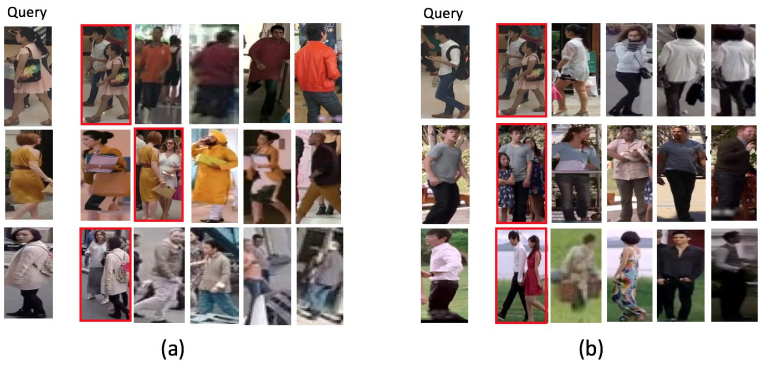

Fig. 2. Person retrieval results of PISNet under our proposed setting. The red rectangles indicate correct retrieval results. The correct results in (a) show the targets interfered by the other pedestrians, while in (b), the targets are even occluded by other pedestrians.

The results indicate that PISNet is able to re-identify people in both cases. Moreover, in the first row of Figure 2(a) and Figure 2(b), the same multi-person gallery image is retrieved using different query images. This further proves the effectiveness of our query-guided attention mechanism.

3 Additional Analysis

Table 1. Experiment results (%) of the strong baseline combined with the two multi-label learning losses on PI-PRW and PI-CUHK-SYSU dataset. BCE and SLF the denote Binary Cross Entropy and the Square Loss Function, respectively

Method	PI-PRW				PI-CUHK-SYSU			
	rank1	*rank 5*	*rank 10*	*mAP*	*rank1*	*rank 5*	*rank 10*	*mAP*
Strong Baseline	34.7	59.4	70.3	36.0	72.5	83.9	88.2	70.1
Strong Baseline + BCE	29.8	55.0	63.8	30.2	67.5	75.6	80.3	62.2
Strong Baseline + SLF	27.9	54.1	62.3	28.9	64.5	72.2	76.1	60.4

Is the Multi-Label Loss harmful for PI Re-ID? A straightforward solution to PI Re-ID is utilizing the multi-label loss to supervise the traditional Re-ID model. Two popular multi-label losses are considered: Binary Cross Entropy (BCE) and Square Loss Function (SLF). Table 1 shows the results of the strong baseline combining with the above multi-label loss functions on the PI-PRW and PI-CUHK-SYSU datasets. The results in Table 1 show that, when combining with the multi-label losses, the performance degrades significantly. This is because the multi-label learning losses try to activate feature representation of

all related objects, which causes the more severe interference, in order to tag all the label in images.

Table 2. Comparison with the pose-guided methods on PI-PRW and PI-CUHK-SYSU dataset (%)

Method	PI-PRW				PI-CUHK-SYSU			
	rank1	*rank 5*	*rank 10*	*mAP*	*rank1*	*rank 5*	*rank 10*	*mAP*
PGFA	25.3	50.9	60.3	25.4	60.1	72.4	79.2	59.3
PGFA*	37.0	60.2	71.2	38.1	74.1	80.3	88.4	71.4
Ours	42.7	67.4	76.2	43.2	79.1	88.4	91.9	76.5

Is the pose estimation algorithm helpful for PI Re-ID? The Pose-Guided Feature Alignment (PGFA) uses a pose estimation algorithm to calculate the landmarks of the pedestrians in images, and assume that the person with the largest number of landmarks is the target person. Then PGFA enhances the features of the target through the locations of the landmarks. Therefore, we modify the PGFA into PGFA* which can separately enhance the features of containing people. For example, if there are two people in the images, PGFA* outputs two features, which are enhanced for two people, respectively. As shown in Table 2, our method surpasses PGFA* by +9.5% Rank-1 accuracy and +8.5% mAP on PI-PRW and +6.6% Rank-1 accuracy and +6.4% mAP on PI-CUHK-SYSU. This result further demonstrates the advantages of our approach to the PI Re-ID problem, even compared with the method with the pose estimation algorithm.