Supplementary Material of Disentangling based Environment-Robust Feature Learning for Person ReID

BMVC 2022 Submission # 428

1 Details of ME-ReID Dataset

Fig. 1 shows statistics of ME-ReID. 6∼71 pedestrians are found under each camera, with an average of 29.0. The number of images under each camera ranges from 30 to 509, on average of 196.9. There are 6∼43 images belonging to each person, with an average of 15.97, and most of the pedestrians have 10∼20 images. ME-ReID contains totally 5908 boundingboxes.

We divide our dataset into train, query and gallery parts. We randomly sampled 150 identities to form the training set, while other 220 identities are used to form testing set. Among each identity in the testing set, about 30% of the images are chosen into the query set, while others into the gallery set. There are totally 2296 images in training set, 1265 in query set and 2347 in gallery set.

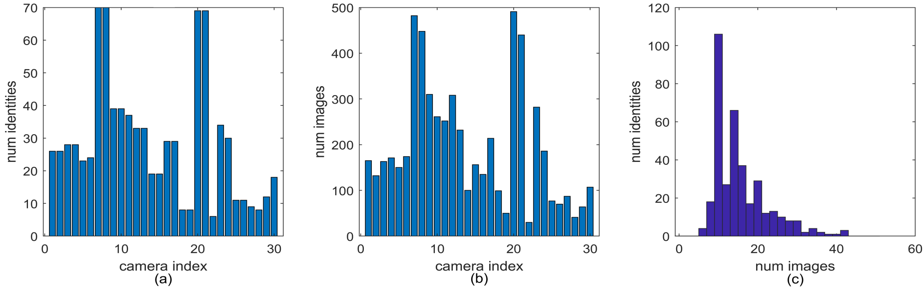

Figure 1: Statistics of ME-ReID. (a) number of identities under each camera. (b) number of images under each camera. (c) distribution of the number of images belonging to the pedestrians.

Data Source. We collected raw data from 30 real urban surveillance cameras, including 3 indoor cam-eras and 27 outdoor cameras. These cameras are distributed in several streets and neighborhoods. The clips are distributed over a 15-day span in winter, covering day and night, sunny and snowy.

Procedures. Yolo-v3 [2] is adopted as pedestrians detector to acquire person bounding-boxes. To get identity labels, we adopt a hierarchical clustering method to get raw ID labels, and then sort out wrong labeled samples.

Privacy Protection. We adopt DSFD [1] for face detection, and add Gaussian blur to the detected face areas, in order to protect privacy of pedestrians.

2 Camera Pairwise ReID Setting

Our method mainly target to eliminate environment related factors from identity features, then make more positive samples rank ahead of negative samples from the same camera with query. We design a camera pairwise ReID setting for direct evaluation. Different from the traditional ReID process that retrieve each query in the whole gallery set, we retrieve in a mini gallery sets with positive samples of same camera and negative sample of different camera. We calculate retrieval scores (mAP and CMC for ReID) between each camera pairs. The result of camera pairwise ReID scores serve as better evaluations on how a method eliminates environment factors and extract robust identity features. The detailed calculation is shown in the following algorithm.

Algorithm 1: Camera Pairwise Retrieving

Initialize: define $q_{ij} = 0$ as query times between camera pair (i, j), $S_{ij} = 0$ as the camera pairwise ReID scores, including camera pairwise mAP and CMC.

```
for each query do
    denote the camera label of this query as u ;
    find a set of cameras C containing positive samples ;
    for each camera v in C do
        form a mini gallery set with positive samples in v and negative samples in u. ;
        retrieve the query image in the mini gallery set, calculate retrieval scores S_temp. ;
        S_uv = S_uv + S_temp, q_uv = q_uv + 1;
    end
    regularize pairwise scores: S_ij = S_ij / q_ij for i, j in 1, 2, ..., N_c;
end
```

Fig. 2 shows results of camera pairwise scores on MSMT17 dataset of baseline and our methods with ResNet50 as backbone. In both figures, X axis and Y axis denote camera pairwise scores of baseline and our method respectively, and each node denotes a camera pair. Camera pairs with lower mAP and R1, have larger retrieval difficulties, indicating greater environment differences between the two cameras. Both of the figures shows that, our method exceeds baseline on most of the camera pairs, on both mAP and Rank1 scores. Besides, our EFL method acquire statistically larger improvements on camera pairs with larger environment difference, which shows the effect of our method to eliminate the interfere of environment features.

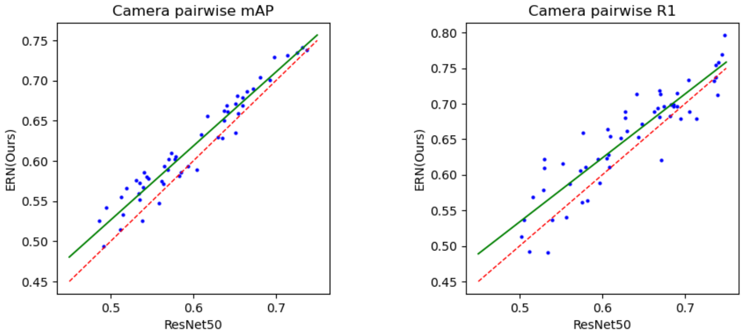

Figure 2: Visualization of distribution of camera pairwise mAP (a) and camera pairwise Rank1 (b). X axis and Y axis denote camera pairwise scores of baseline and our method respectively. Each node denotes a camera pair.

References

[1] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. Dsfd: dual shot face detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5060–5069, 2019.

[2] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv e-prints*, 2018.