# Beyond Human Parts: Dual Part-Aligned Representations for Person Re-Identification Supplementary Material

Jianyuan Guo[1†], Yuhui Yuan[2,3,4†], Lang Huang[1], Chao Zhang[1*], Jin-Ge Yao[2], Kai Han[5]
[1]Key Laboratory of Machine Perception (MOE), Peking University, [2]Microsoft Research Asia
[3]Institute of Computing Technology [4]University of Chinese Academy of Sciences [5]Noah's Ark Lab
{jyguo, laynehuang}@pku.edu.cn, chzhang@cis.pku.edu.cn
{yuhui.yuan, jinge.yao}@microsoft.com, kai.han@huawei.com

In this supplementary material, we provide the complexity analysis of our method, details about triplet loss, and show more typical experiments and cases on DukeMTMC-ReID and CUHK03.

## 1. Appendix

### 1.1. Triplet loss

As mentioned in Sec 3.2, we use triplet loss to improve the performance in final results. The details are as folows: (1) We prepare each mini-batch by randomly sampling 16 classes (identities) and 4 images for each class. (2) We set the weight rate as 1:1 on all three datasets. (3) Given a mini-batch of 64 samples, we construct a triplet for each sample by choosing the hardest positive sample and the hardest negative sample measured by their Euclidean distances.

### 1.2. Strategies for inserting DPB.

We do the ablation study to find the results of adding DPB after different Res-$k$ residual blocks. As shown in the Table 1 of main paper, we can find that all types of blocks (DPB / Human Part Branch / Latent Part Branch) achieve better performances when they are inserted after the Res-2 and Res-3 stages, compared to Res-1 and Res-4 stages. Specifically, the DPB improves the Rank-1 accuracy and mAP by 4.4% and 9.5% when inserted after res-2 stage, 3.4% and 7.3% when inserted after res-3 stage, respectively. One possible explanation is that the feature map from Res-1 has more precise localization information but less semantic information, and the deeper feature map from Res-4 is insufficient to provide precise spatial information. In conclusion, Res-2 and Res-3 can benefit more from the proposed DPB. So the 5×DPB in all experiments means that we add 2 DPB blocks to Res-2 and 3 DPB blocks to Res-3, if not specified.

*Corresponding author. †Equal contribution.

Table 1: Complexity comparison of DPB/Baseline on CUHK03.

| Method | 5×DPB | Params | FLOPs | Time | R-1 | mAP |
|---|---|---|---|---|---|---|
| R-50 | × | 24.2M | 14.9G | 19ms | 60.29 | 54.79 |
| R-101 | × | 43.2M | 22.1G | 32ms | 68.14 | 63.45 |
| R-50 | ✓ | 31.6M | 18.6G | 27ms | **71.55** | **64.23** |

### 1.3. Complexity analysis

We compare the proposed model with ResNet-50 and ResNet-101 in model size and computation complexity, measured by the number of parameers and FLOPs during inference on CUHK03. And we test the inference time of each forward pass on a single GTX 1080Ti GPU with CUDA8.0 given an input image of size $3 \times 384 \times 128$. Table 1 shows that our method outperforms ResNet-101 with smaller model size, less computation amount and faster inference speed, the improvement of $P^2$-Net is not just because the added depth to the baseline model.

### 1.4. Experiments on DukeMTMC-reID

To further verify that the latent part branch and the human part branch are complementary, we also conduct the controlled experiments on both DukeMTMC-ReID and CUHK03.

We present the results on DukeMTMC-reID in Table 2. It can be seen that DPB achieves better performance than either only employing the latnet part branch or only employing the human part branch. e.g., "1 × DPB" improves the mAP of "1 × DPB (HP-5)" from 66.99 to 67.93. "5 × DPB" improves the mAP of "5 × HPP (HP-5)" from 68.64 to 70.84.

We present the advantages of human part branch in Figure 1 . The results with human part branch perform more robust compared with the results of baseline and the results with the latent part branch. For example, the query image on the $1^{st}$ line carries the misleading information caused by

Table 2: Comparison experiments on **DukeMTMC-ReID**. DPB (HP-5) only uses the human part branch and sets $K = 5$. DPB (Latent) only uses the latent part branch. DPB uses both the human part branch and the latent part branch.

| Method | R-1 | R-5 | R-10 | mAP |
|---|---|---|---|---|
| Baseline | 79.85 | 89.81 | 92.19 | 62.57 |
| $1 \times$ DPB (HP-5) | 83.04 | 91.18 | 93.22 | 66.99 |
| $1 \times$ DPB (Latent) | 82.20 | 90.33 | 92.69 | 65.09 |
| $1 \times$ DPB | **83.80** | **91.38** | **93.58** | **67.93** |
| $5 \times$ DPB (HP-5) | 84.08 | 91.82 | 94.10 | 68.64 |
| $5 \times$ DPB (Latent) | 84.45 | 91.97 | 94.25 | 69.07 |
| $5 \times$ DPB | **84.91** | **92.08** | **94.45** | **70.84** |

Table 3: Comparison experiments on **CUHK03**. DPB (HP-5) only uses the human part branch and sets $K = 5$. DPB (Latent) only uses the latent part branch. DPB uses both the human part branch and the latent part branch.

| Method | R-1 | R-5 | R-10 | mAP |
|---|---|---|---|---|
| Baseline | 60.29 | 78.21 | 84.86 | 54.79 |
| $1 \times$ DPB (HP-5) | 67.57 | 81.32 | 87.36 | 60.02 |
| $1 \times$ DPB (Latent) | 68.59 | 83.14 | 87.96 | 61.75 |
| $1 \times$ DPB | **70.43** | **84.50** | **89.64** | **63.93** |
| $5 \times$ DPB (HP-5) | 69.93 | 83.86 | 88.90 | 63.34 |
| $5 \times$ DPB (Latent) | 69.84 | 83.50 | 89.83 | 63.25 |
| $5 \times$ DPB | **71.55** | **85.71** | **90.80** | **64.23** |

the part of a car. Both the baseline method and the method with latent part branch return the images carrying parts of the car, and the method with human part branch returns the correct result by removing the influence of the car.

We also present the benefits of latent part branch in Figure 2. The failure cases in both the baseline and the method with human part branch are solved by using the latent part masks generated by the latent part branch. It can be seen that these latent part masks capture some non-human but important part information that fail to be captured by both the baseline method and the method with only human part branch. We can conclude that the latent part branch and the human part branch are complementary accordingly.

### 1.5. Experiments on CUHK03

We report the results on CUHK03 (detected) in Table 3. And we can find similar performance improvements by combining the human part branch and the latent part branch. e.g., "$1 \times$ DPB" improves the mAP of "$1 \times$ DPB (HP-5)" from 60.02 to 63.93. "$5 \times$ DPB" improves the mAP of "$5 \times$ DPB (HP-5)" from 63.34 to 64.23.

Our approach boosts the performance of baseline model by a large margin, especially on CUHK03 dataset, the probable reasons are (i) the quality is better (less blurring effects, higher image resolutions: $266 \times 90$ in CUHK03, $128 \times 64$ in Market-1501), thus the DBP can estimate more accurate human parsing and latent attention results. (ii) the background across different images is more noisy, DPB can remove the influence of the background.

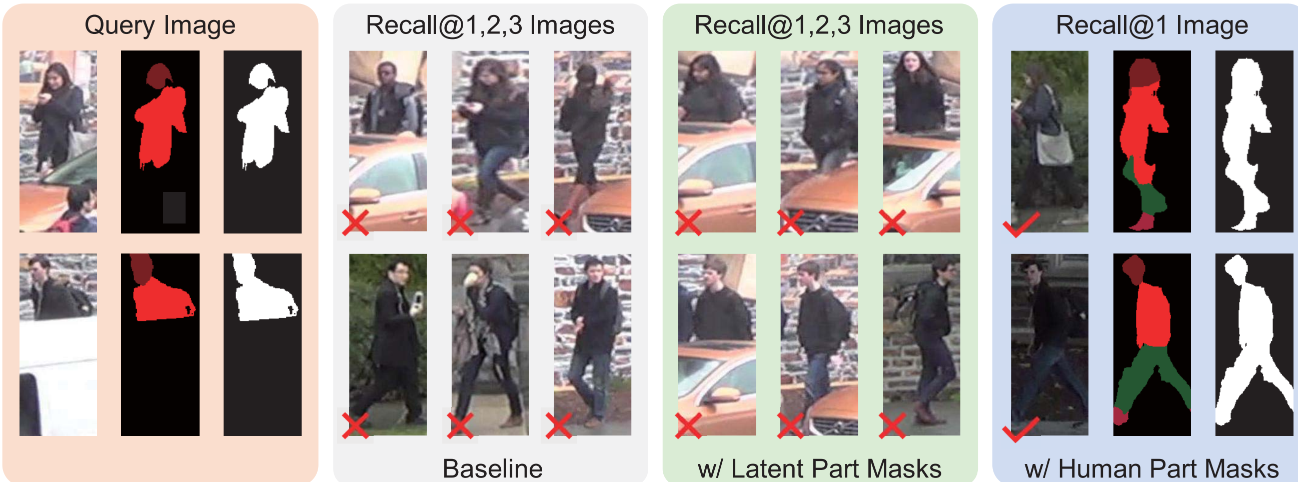


Figure 1: Comparison of Baseline, DPB (w/ Latent Part Masks) and DPB (w/ Human Part Masks) on **DukeMTMC-ReID**. We denote $P^2$-Net that only employs human part branch as the method w/ Human Part Masks. Both these two query images suffer from the problem of occlusions and contain useless or misleading background information. Both the baseline and DPB (w/ Latent Part Masks) fail to return the correct results within the top 3 positions while DPB (w/ Human Part Masks) returns the correct result at top 1 position.

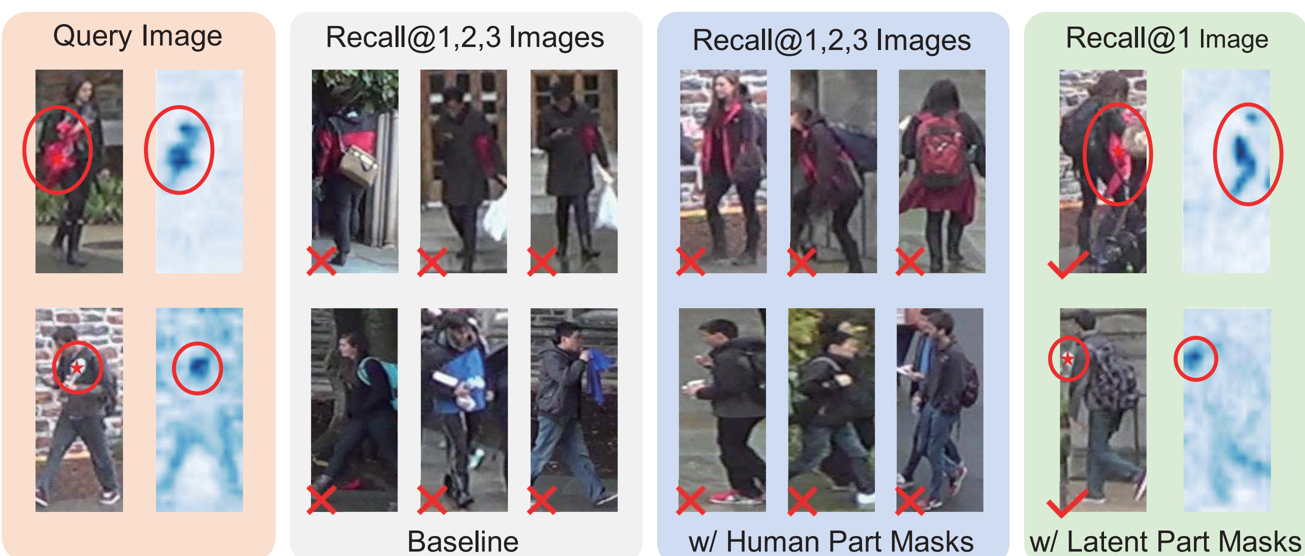


Figure 2: Comparison of Baseline, DPB (w/ Human Part Masks) and DPB (w/ Latent Part Masks) on **DukeMTMC-ReID**. There exist some important non-human parts within all these two query images. The DPB (w/ Human Part Masks) categorizes these important parts to background and fails to return the correct image. The DPB (w/ Latent Part Masks) predicts the latent part mask associated with these parts, which helps to find the correct image.