Supplementary Material: Anchor-Free Person Search

This is the supplementary material for the paper entitled "Anchor-Free Person Search". Although the main paper stands on its own, it is still worthwhile providing more experimental results and performance analysis. In this supplementary document, we provide

- More visualization of the person search results on CUHK-SYSU [2].

- More visualization of the learned offsets of the deformable convolution filters on CUHK-SYSU [2].

1. Qualitative Results of Person Search

Fig. 1 visualizes more person search results, which are complementary to Fig. 7 in the main paper. This figure covers various challenging cases, including low resolution, scale/viewpoint variations, and occlusions. We observe that OIM [2] and NAE [1] fail to retrieve the correct targets due to either detection failure or matching to the wrong identity, while in most of these challenging cases, the proposed AlignPS and AlignPS+ models successfully detect and retrieve correct target persons. It is also interesting to notice that, in the first three extremely difficult cases, where the low-resolution targets are in crowded scenes, our models can still obtain correct search results. These results further demonstrate the robustness and the discriminative capability of the proposed framework.

2. Visualization of Learned Offsets

We also provide more qualitative results on the learned offsets of the deformable convolution filters in Fig. 2, as a complement to Fig. 5 in the main paper. Similar to the observations in our main manuscript, we can see that the proposed framework is capable of learning adaptive receptive fields in accordance with the human body layout. It is also noteworthy that the receptive fields can well adapt to the regions of the overlapping persons in crowded scenes, thus yielding more reliable results under these difficult circumstances.

References

[1] Di Chen, Shanshan Zhang, Jian Yang, and Bernt Schiele. Norm-aware embedding for efficient person search. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12612–12621, 2020. 1, 2

[2] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3376–3385, 2017. 1, 2

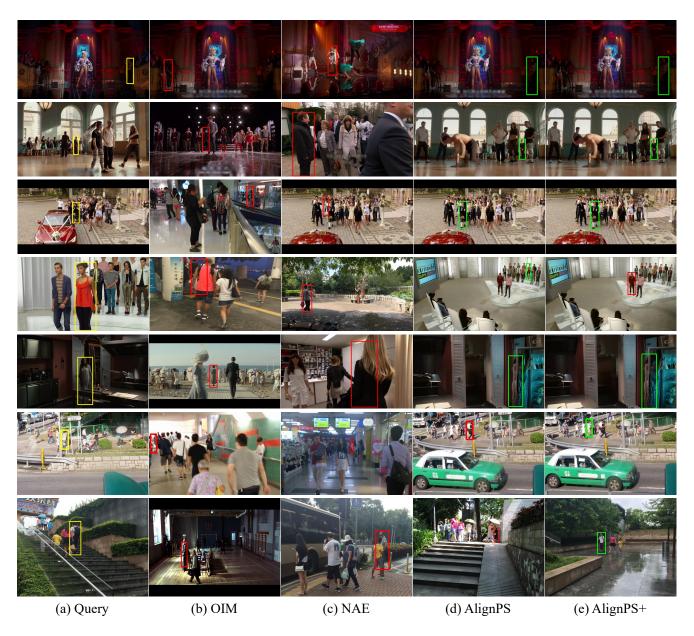| (a) Query | (b) OIM | (c) NAE | (d) AlignPS | (e) AlignPS+ |

Figure 1: Difficult cases that can be successfully retrieved by AlignPS/AlignPS+ but not OIM [2] and NAE [1]. The yellow bounding boxes denote the queries, while the green and red bounding boxes denote correct and incorrect top-1 matches, respectively.

(a) Deformable conv at lateral C_3 layer in AFA

(b) Deformable conv at lateral C_4 layer in AFA

Figure 2: Each image shows the sampling locations of two levels of 3×3 ($9^2 = 81$ points at each location) deformable filters: (a) Lateral deformable conv C_3 + Output deformable conv; (b) Lateral deformable conv C_4 + Output deformable conv. We illustrate different locations with different colors, while center locations of people are marked in green.