Supplementary Material for Domain Adaptive Person Search

I Additional Results

This is the supplementary material for the paper entitled "Domain Adaptive Person Search". Although the main paper stands on its own, it is still worthwhile providing more experimental results and performance analysis.

I.1 Analysis on Balancing Factor λ

To investigate the effectiveness of our designed instance-level task-sensitive balancing term λ, we compare it with some manually assigned values. When CUHK-SYSU serves as the target dataset, λ is set as 0.98. As illustrated in Table I, in comparison with a wide range of values, our task-sensitive design achieves the best performance. Moreover, it can be observed that the performance of detection is positively associated with λ, which validates our hypothesis in paper that the detection performance mainly depends more on the first standard Faster R-CNN head.

I.2 Comparison with Two-Step Person Search Models

To evaluate the performance and efficiency of our proposed end-to-end framework, we compare it with the baseline two-step model, where the detector is first trained to localize pedestrians from raw images, and ReID model trained with unified contrastive loss is subsequently employed for person retrieval from the former detected crops. Specifically, to make a fair comparison, we use the detection branch of SeqNet as the detector. From Table II, we can make the observations that DAPS not only outperforms the two-step method, but also takes a lead in efficiency by a large margin. Specifically, DAPS displays shorter inference time (117 ms $\rightarrow$ 46 ms) and lower FLOPs (557G $\rightarrow$ 423G). It is also noteworthy that the ReID training prototype adopted by two-step method is the same with DAPS, and our proposed target data utilization strategies help boost the performance of the end-to-end DAPS framework.

I.3 Comparison with Joint-Domain Fully-Supervised Settings

We further train the state-of-the-art fully supervised person search model with both of the CUHK-SYSU and PRW to measure the theoretical upper bound of DAPS, and the results are reported in Table III. When trained with both datasets, the model trained with joint-domain achieves better performance on

Table I: Comparative results of task-sensitive instance-level alignment. When CUHK-SYSU serves as the target domain, the task-sensitive balancing term is 0.98.

λ	Target: CUHK-SYSU			
	mAP	top-1	recall	AP
0.2	59.0	60.6	66.9	56.3
0.4	57.8	59.0	70.2	59.0
0.5	58.2	60.5	66.3	56.3
0.8	59.7	61.4	70.3	61.3
1.0	61.9	62.7	71.7	65.6
task-sensitive	62.2	63.6	70.8	63.1

Table II: Comparison with two-step methods on CUHK-SYSU. Runtime is measured by the average inference time, and the unit is milliseconds (ms).

Methods	Target: CUHK-SYSU					
	GFLOPs	Runtime	mAP	top-1	recall	AP
SeqNet+SPCL [1]	557	117	65.2	67.5	70.8	63.1
DAPS(ours)	423	46	77.6	79.6	77.7	69.9

PRW, but evidently underperforms the normal training strategy on CUHK-SYSU, especially for the detection sub-task. The result indicates that simply adding more training data is not beneficial for the discriminability and generalization ability of model, and proper domain adaptation operation is necessary. Moreover, there still exists a significant margin between DAPS and its theoretical upper limit, and we hope this work will encourage future works to explore solutions for bridging this gap.

I.4 Analysis on Image-Level Alignment Positions

We compare the cases of adding the image-level alignment module to the output of different convolutional blocks, and to multi-scale (all three blocks) of the backbone ResNet-50. As shown in Table IV, in comparison with baseline, adding domain alignment to any intermediate layer improves both of detection and ReID performance, and the best performance is achieved with adding alignment to res4. Following some recent domain adaptation methods in object detection [4,3], we also perform multi-scale alignment by conducting image-level alignment on all three blocks. However, the results indicate that the multi-scale strategy has a negative effect on ReID sub-task, and the scale misalignment issue is not the top priority for person search task.

Table III: Comparison with joint-domain fully-supervised person search models. * denotes training on both of CUHK-SYSU and PRW under the fully supervised setting.

Methods	Target: PRW				Target: CUHK-SYSU			
	mAP	top-1	recall	AP	mAP	top-1	recall	AP
SeqNet [2]	46.7	83.4	96.4	94.0	93.8	94.6	92.1	89.2
SeqNet*	49.4	85.2	97.7	94.7	92.2	93.2	86.6	83.9
DAPS(ours)	34.7	80.6	97.2	90.9	77.6	79.6	77.7	69.9

Table IV: Comparative results of different positions to apply image-level alignment. Baseline denotes not applying any alignment operation. "res" denotes the output of corresponding residual stage in ResNet-50 backbone. "multi-scale" refers to applying image-level alignment on the output of all the three residual stages.

Methods	Target: CUHK-SYSU			
	mAP	top-1	recall	AP
baseline	52.5	54.8	55.2	55.1
res2	59.0	60.5	70.4	58.2
res3	57.9	59.1	71.4	62.7
res4	62.2	63.6	70.8	63.1
multi-scale	57.3	59.0	71.2	64.1

References

1. Ge, Y., Zhu, F., Chen, D., Zhao, R., Li, H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In: NeurIPS (2020)
2. Li, Z., Miao, D.: Sequential end-to-end network for efficient person search. In: AAAI. pp. 2011–2019 (2021)
3. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. In: CVPR. pp. 6956–6965 (2019)
4. Zhuang, C., Han, X., Huang, W., Scott, M.R.: ifan: Image-instance full alignment networks for adaptive object detection. In: AAAI. pp. 13122–13129 (2020)