

Supplementary Material on

Context-Aware Unsupervised Clustering for Person Search

Byeong-Ju Han^{*,1}

bjhan@unist.ac.kr

Kuhyeun Ko^{*,1}

khko@unist.ac.kr

Jae-Young Sim^{†,1,2}

jysim@unist.ac.kr

¹ Department of Electrical Engineering,
UNIST, Ulsan, Korea² Graduate School of Artificial
Intelligence, UNIST, Ulsan, Korea

*Equal contribution

†Corresponding author

1 Training Strategy Details

We warm up the learning rate to $3e-3$ at the first epoch, and decrease the learning rate to $3e-4$ at the 20th and 14th epoch on CUHK-SYSU and PRW datasets, respectively, to enhance the training stability. We only update the feature vectors in the feature memory during the first 4 epochs of training regarding each person instance has a unique positive sample, and then we apply the proposed clustering methods of HNM and HPM to train the network for the remaining epochs. In every training iteration, we generate a mini-batch of size 4 and update the network parameters to minimize the training loss by using an SGD optimizer with momentum 0.9. The network is totally trained for 26 epochs on CUHK-SYSU dataset and 18 epochs on PRW dataset, respectively.

2 Quantitative Comparison

We evaluate the person search performance according to different gallery sizes in CUHK-SYSU, and compare the proposed method with more existing supervised person search methods.

Effect of gallery size. CUHK-SYSU dataset provides not only the standard gallery including 100 test images but also the other galleries containing different numbers of test images. In the submitted paper, the person search performances in CUHK-SYSU dataset are evaluated on the standard gallery. In Figure S-1, we additionally investigate the performance variation of the proposed method and the extended unsupervised person re-ID methods, DET+BUC [13] and DET+MLC [19], according to different gallery sizes, where we see that the proposed method always achieves outstanding performance than that of DET+BUC [13] and DET+MLC [19].

Person search results. Table 1 summarizes quantitative performances of more supervised person search methods including the methods compared in the submitted paper. The scores of the existing supervised person search methods listed in the top block were taken from literature, and that in the rest blocks were practically evaluated by the same evaluation code. In Table 1, we can see that the performance scores of the proposed method trained without person identity labels are even better than that of several supervised methods in both datasets.

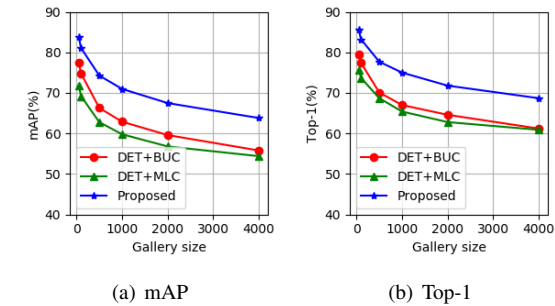


Figure S-1: Comparative quantitative performance on CUHK-SYSU dataset according to different gallery sizes. (a) mAP and (b) Top-1 score.

Method	Supervised	CUHK-SYSU		PRW	
		mAP	Top-1	mAP	Top-1
OIM [14]	Yes	75.5	78.7	21.3	49.9
DPM+IDE [15]	Yes	-	-	20.5	48.3
IAN [16]	Yes	76.3	80.1	23.0	61.9
NPSM [17]	Yes	77.9	81.2	24.2	53.1
RCAA [18]	Yes	79.3	81.3	-	-
CNN+CLSA [19]	Yes	87.2	88.5	38.7	65.0
CTX [20]	Yes	84.1	86.5	33.4	73.6
QEEPS [21]	Yes	88.9	89.1	37.1	76.7
FPN+RDLR [8]	Yes	93.0	94.2	42.9	70.2
BiNet [9]	Yes	90.0	90.7	45.3	81.7
PBNet [22]	Yes	90.5	88.4	48.5	87.9
DIOIM [9]	Yes	88.7	89.6	36.0	76.1
APNet [23]	Yes	88.9	89.3	41.2	81.4
TCTS [24]	Yes	93.9	95.1	46.8	87.5
IGPN+PCB [9]	Yes	90.3	91.4	47.2	87.0
MGTS [9]	Yes	83.3	84.2	32.8	72.1
DMRNet [9]	Yes	93.2	94.2	46.9	83.3
SeqNet [25]	Yes	94.8	95.7	47.6	87.6
AlignPS [26]	Yes	94.0	94.5	46.1	82.1
PGA [10]	Yes	92.3	94.7	44.2	85.2
OR [27]	Yes	93.2	93.8	52.3	71.5
HOIM [9]	Yes	89.7	90.8	44.2	84.5
NAE [9]	Yes	92.1	92.9	45.7	83.3
DET+BUC [10]	No	74.8	77.4	26.0	83.6
DET+MLC [10]	No	69.2	73.7	25.4	84.7
Proposed	No	81.1	83.2	41.7	86.0
*HOIM [9]	Yes	-	-	36.5	64.9
*NAE [9]	Yes	-	-	40.0	67.5
*DET+BUC [10]	No	-	-	18.6	53.0
*DET+MLC [10]	No	-	-	17.1	50.8
*Proposed	No	-	-	36.6	64.9

Table 1: Further comparison of the quantitative performances.

3 Qualitative Comparison

We provide the qualitative results of the proposed method compared with that of DET+BUC [13] and DET+MLC [19].

Feature Distribution. We first visualize the features of detected persons in 2-D space by using t-SNE [17] to see the qualitative performance of clustering. Figures S-2 and S-3 show the feature distribution of detected persons for 40 identities in CUHK [20] and PRW [25] testsets, respectively, where the points are marked with the same color whose bounding boxes yield IoU scores higher than 0.5 to a certain ground truth of the same identity. We see that the existing methods produce the feature points mixed with each other and scattered in relatively narrow areas in the feature space, making it difficult to clearly separate them according to their identities. On the other hand, the proposed method distributes the feature points over a larger area while locally grouping the points of the same color. We also select four sample points of different identities, and then show the detected persons whose features lie within a certain distance from that of the selected samples in the feature space, respectively. Figures S-2 and S-3 show eight samples uniformly selected for each identity. Whereas DET+BUC and DET+MLC tend to over-estimate the similarities between two persons of different identities due to partially similar human poses and/or cloth colors, the proposed method achieves a better clustering performance by removing such false positive samples reliably.

Qualitative person search results. As demonstrated in the submitted paper, the proposed method quantitatively outperforms both DET+BUC and DET+MLC. In this section, Figures S-4~S-7 and Figures S-8~S-11 show the person search results on CUHK-SYSU and PRW testsets, respectively. We see that the proposed method more frequently provides reliable results of person search even in the challenging cases, where the persons are under-exposed due to low-lighting as shown in Figures S-4~S-7, and the persons locally exhibit distinct features, e.g., the hats, umbrellas, and bags, as shown in Figures S-8~S-11. However, in such cases, DET+BUC and DET+MLC usually fail to find correct persons matching to the queries.

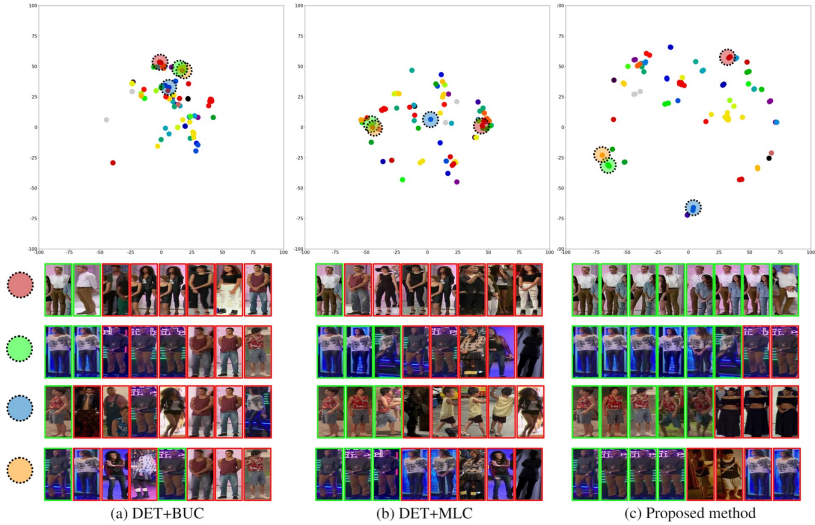


Figure S-2: The feature distributions of detected persons for 40 identities in CUHK-SYSU testset, visualized in 2-D space by using t-SNE [17]. (a) DET+BUC. (b) DET+MLC. (c) The proposed method. The points are marked with the same color whose bounding boxes yield IoU scores higher than 0.5 to a certain ground truth of the same identity. The detected persons whose features lie within a certain distance from that of four selected samples in the feature space are displayed below.

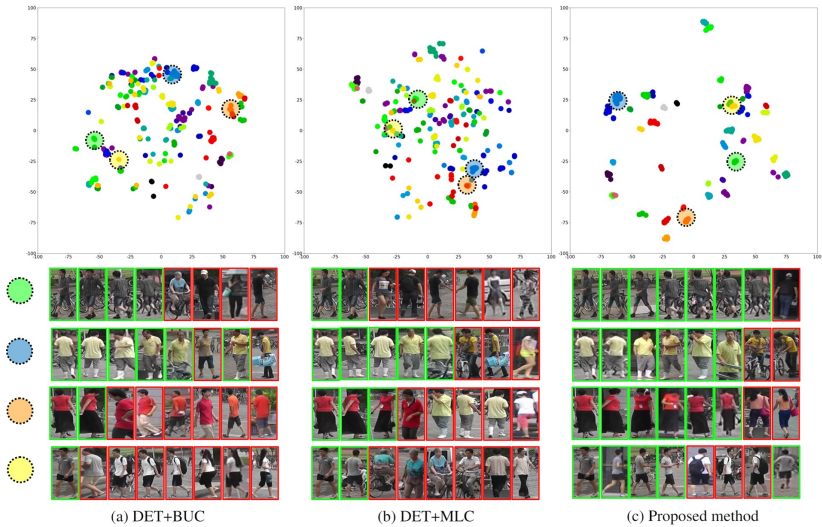


Figure S-3: The feature distributions of detected persons for 40 identities in PRW testset, visualized in 2-D space by using t-SNE [17]. (a) DET+BUC. (b) DET+MLC. (c) The proposed method. The points are marked with the same color whose bounding boxes yield IoU scores higher than 0.5 to a certain ground truth of the same identity. The detected persons whose features lie within a certain distance from that of four selected samples in the feature space are displayed below.

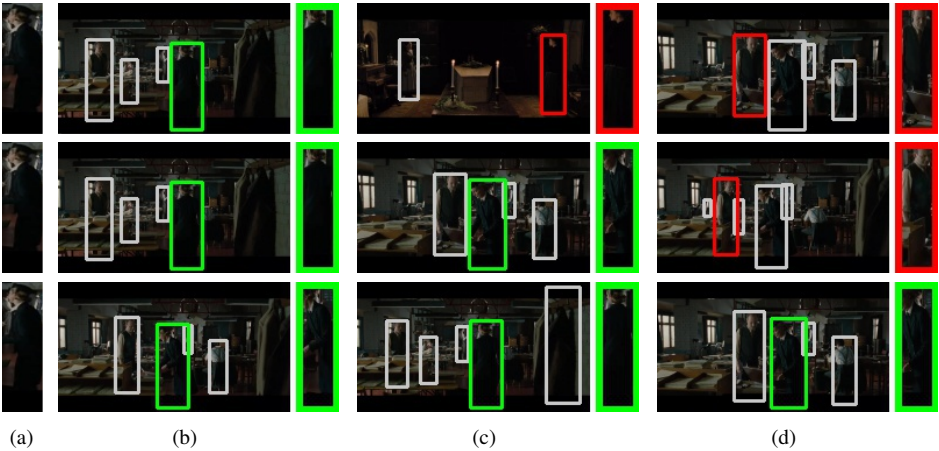


Figure S-4: Qualitative comparison of the person search results on CUHK-SYSU testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

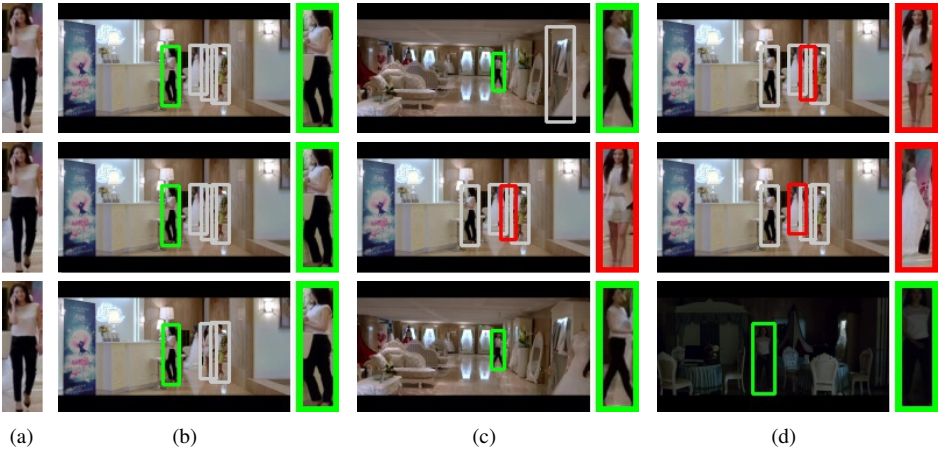


Figure S-5: Qualitative comparison of the person search results on CUHK-SYSU testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

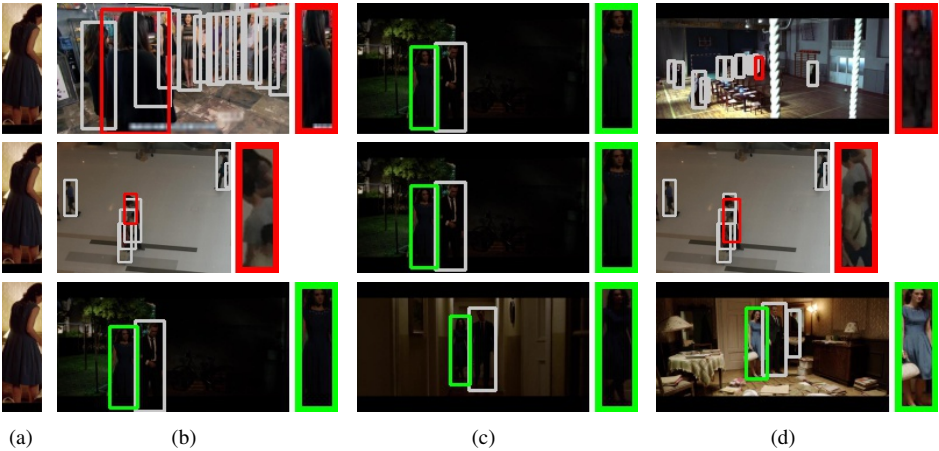


Figure S-6: Qualitative comparison of the person search results on CUHK-SYSU testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

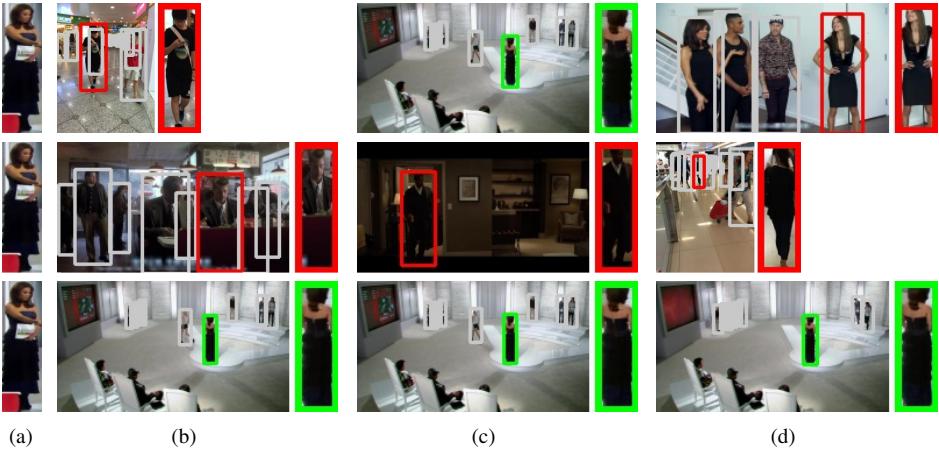


Figure S-7: Qualitative comparison of the person search results on CUHK-SYSU testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

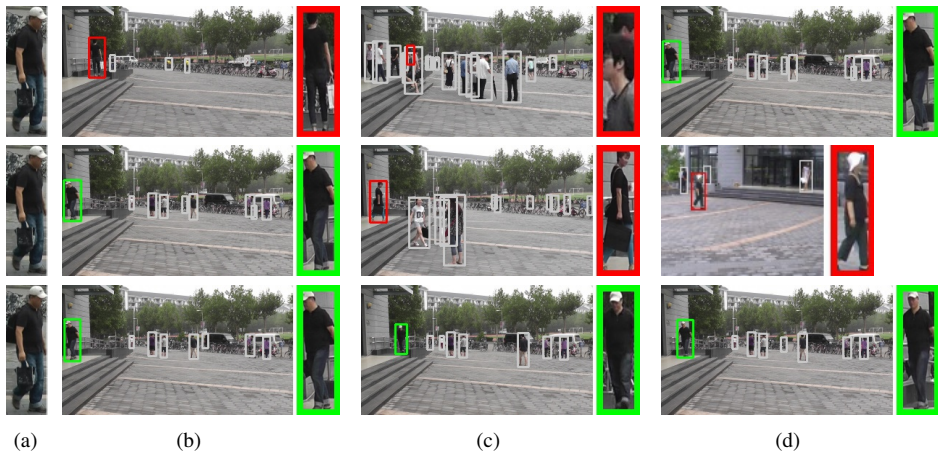


Figure S-8: Qualitative comparison of the person search results on PRW testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

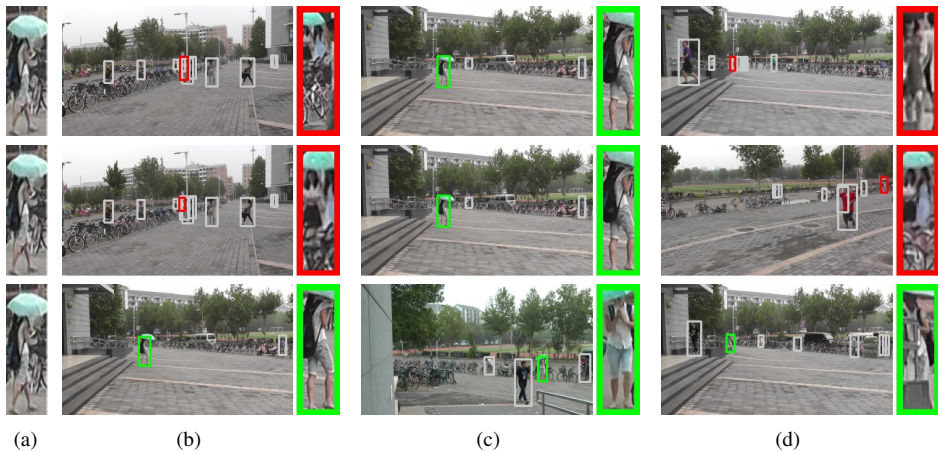


Figure S-9: Qualitative comparison of the person search results on PRW testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

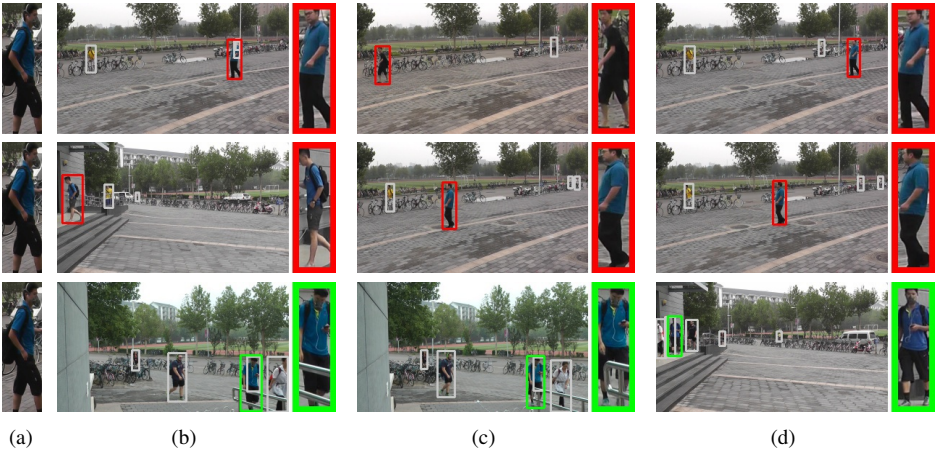


Figure S-10: Qualitative comparison of the person search results on PRW testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

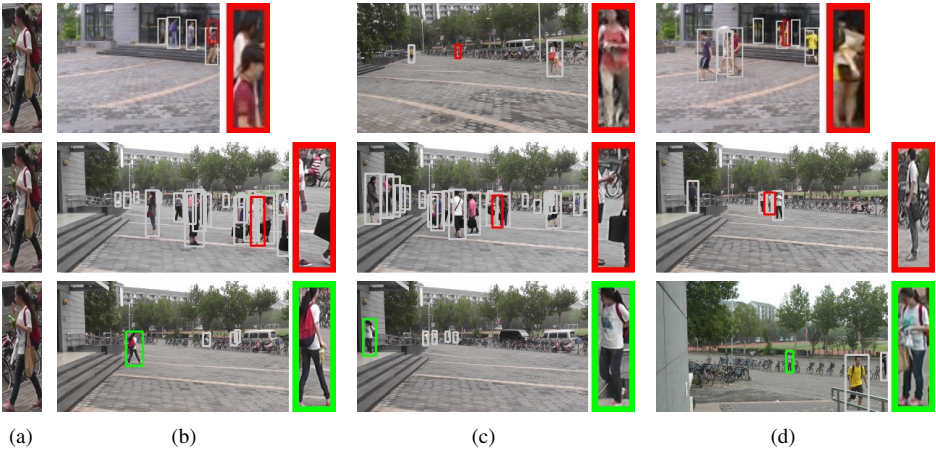


Figure S-11: Qualitative comparison of the person search results on PRW testset obtained by using DEC+BUC, DEC+MLC, and the proposed method from top to bottom, respectively. (a) A test query. The person search results of (b) the most, (c) the second most, and (d) the third most similar features to the given query. Detected persons are localized by the predicted bounding boxes, and the true and false positive samples are highlighted in green and red, respectively.

References

- S1. Xiaojun Chang, Po-Yao Huang, Yi-Dong Shen, Xiaodan Liang, Yi Yang, and Alexander G. Hauptmann. RCAA: Relational context-aware agents for person search. In *Proceedings of the European Conference on Computer Vision*, 2018.
- S2. Di Chen, Shanshan Zhang, Wanli Ouyang, Jian Yang, and Bernt Schiele. Hierarchical online instance matching for person search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.
- S3. Di Chen, Shanshan Zhang, Wanli Ouyang, Jian Yang, and Ying Tai. Person search by separated modeling and a mask-guided two-stream cnn model. *IEEE Transactions on Image Processing*, 2020.
- S4. Di Chen, Shanshan Zhang, Jian Yang, and Bernt Schiele. Norm-aware embedding for efficient person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- S5. Ju Dai, Pingping Zhang, Huchuan Lu, and Hongyu Wang. Dynamic imposter based online instance matching for person search. *Pattern Recognition*, 100:107120, 2020.
- S6. Wenkai Dong, Zhaoxiang Zhang, Chunfeng Song, and Tieniu Tan. Instance guided proposal network for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- S7. Wenkai Dong, Zhaoxiang Zhang, Chunfeng Song, and Tieniu Tan. Bi-directional interaction network for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- S8. Chuchu Han, Jiacheng Ye, Yunshan Zhong, Xin Tan, Chi Zhang, Changxin Gao, and Nong Sang. Re-id driven localization refinement for person search. In *Proceedings of the IEEE Conference on International Conference on Computer Vision*, 2019.
- S9. Chuchu Han, Zhedong Zheng, Changxin Gao, Nong Sang, and Yi Yang. Decoupled and memory-reinforced networks: Towards effective feature learning for one-step person search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
- S10. Hanjae Kim, Sunghun Joung, Ig-Jae Kim, and Kwanghoon Sohn. Prototype-guided saliency feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- S11. Xu Lan, Xiatian Zhu, and Shaogang Gong. Person search by multi-scale matching. In *Proceedings of the European Conference on Computer Vision*, 2018.
- S12. Zhengjia Li and Duoqian Miao. Sequential end-to-end network for efficient person search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
- S13. Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

- S14. Hao Liu, Jiashi Feng, Zequn Jie, Karlekar Jayashree, Bo Zhao, Meibin Qi, Jianguo Jiang, and Shuicheng Yan. Neural person search machines. In *Proceedings of the IEEE Conference on International Conference on Computer Vision*, 2017.
- S15. Bharti Munjal, Sikandar Amin, Federico Tombari, and Fabio Galasso. Query-guided end-to-end person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- S16. Kun Tian, Houjing Huang, Yun Ye, Shiyu Li, Jinbin Lin, and Guan Huang. End-to-end thorough body perception for person search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.
- S17. Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9:2579–2605, 2008.
- S18. Cheng Wang, Bingpeng Ma, Hong Chang, Shiguang Shan, and Xilin Chen. TCTS: A task-consistent two-stage framework for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- S19. Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- S20. Jimin Xiao, Yanchun Xie, Tamмам Tillo, Kaizhu Huang, Yunchao Wei, and Jiashi Feng. IAN: the individual aggregation network for person search. *Pattern Recognition*, 87:332–340, 2019.
- S21. Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- S22. Yichao Yan, Qiang Zhang, Bingbing Ni, Wendong Zhang, Minghao Xu, and Xiaokang Yang. Learning context graph for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- S23. Yichao Yan, Jinpeng Li, Jie Qin, Song Bai, Shengcai Liao, Li Liu, Fan Zhu, and Ling Shao. Anchor-free person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- S24. Hantao Yao and Changsheng Xu. Joint person objectness and repulsion for person search. *IEEE Transactions on Image Processing*, 2021.
- S25. Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, Yi Yang, and Qi Tian. Person re-identification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- S26. Yingji Zhong, Xiaoyu Wang, and Shiliang Zhang. Robust partial matching for person search in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.