OIMNet++: Prototypical Normalization and Localization-aware Learning for Person Search

Sanghoon Lee[1], Youngmin Oh[1], Donghyeon Baek[1],
Junghyup Lee[1], and Bumsub Ham[1,2⋆]

[1]Yonsei University [2]Korea Institute of Science and Technology (KIST)

In this supplementary material, we present qualitative analysis on head/tail classes for person search (Sec. S1) and visualizations of positive proposals that sufficiently overlap with ground-truth bounding boxes during training (Sec. S2).

S1 Quantitative analysis

ProtoNorm calibrates feature distributions while considering minority IDs during training. To demonstrate the effectiveness, we categorize training IDs into head and tail classes by splitting at the median occurrence value. For each ID, we compute cosine similarity scores with all features within the LUT. We then average the scores w.r.t the head/tail classes separately and visualize in Fig. S1 the results. Lower similarity scores indicate stronger inter-class separability, indicating higher discriminative power of person representations. We compare the results between three variants; vanilla OIMNet [2], OIMNets adopting either BatchNorm [1] or ProtoNorm. We can observe that vanilla OIMNet shows the worst inter-class separability as it does not consider feature distribution prior to L2 normalization. Employing BatchNorm alleviates this problem. This, however, calibrates feature distribution without considering the long-tail distribution across person IDs, restraining the discriminative power of person representations. ProtoNorm computes feature statistics that are less biased to head IDs for the calibration. This allows to encourage inter-class separateness regardless of the sample distribution across person IDs. As a result, ProtoNorm strengthens the inter-class separability offers lowest similarity scores for both cases.

S2 Visualizations of positive proposals

During training, we have observed that features extracted from positive proposals contain different extents of noise, *e.g.*, background clutter and person overlaps. We provide in Figs. S2 and S3 the visual examples of positive proposals on the PRW [3] and CUHK-SYSU [2] datasets, respectively. These proposals overlap with the corresponding ground-truth bounding boxes more than 0.5 intersection-of-union (IoU) scores. Nevertheless, they depict detrimental noise that can distract learning discriminative person representations. For both figures, proposals in the left three columns contain background clutter, and the remaining ones depict other pedestrian instances. The LOIM loss assigns adaptive momentum value for each proposal, discouraging severely misaligned proposals to contribute in discriminative feature learning for re-identification.

⋆ Corresponding author.

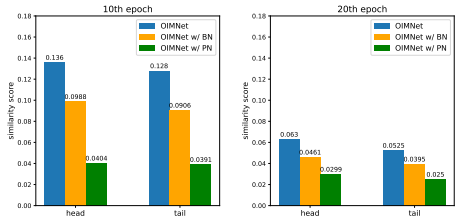

Fig. S1: Average cosine similarity scores on PRW [3] for variants of OIMNet [2]. We divide person IDs into head and tail classes and average the scores among each category over training epochs. Lower similarity score indicates stronger inter-class separateness.

Fig. S2: Positive proposals obtained from PRW [3]. Examples in the top row are sampled at the 10-th epoch, while the bottom ones are sampled at the 20-th epoch.

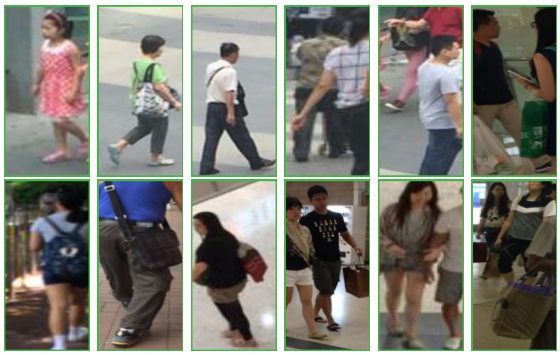

Fig. S3: Positive proposals obtained from CUHK-SYSU [2]. Examples in the top row are sampled at the 10-th epoch, while the bottom ones are sampled at the 20-th epoch.

References

1. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: ICML (2015) 1
2. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. In: CVPR (2017) 1, 2
3. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., Tian, Q.: Person re-identification in the wild. In: CVPR (2017) 1, 2