

# Norm-Aware Embedding for Efficient Person Search

## Supplementary Material

Di Chen, Shanshan Zhang, Jian Yang, Bernt Schiele

### Abstract

The supplementary material complements our work with analysis on detection performance. In addition, we provide more qualitative results on failure cases.

### 1. Detection Performance

Method	Recall	AP
OIM-base w/o re-ID	92.9	86.5
OIM-base	-3.6	-6.8
NAE w/o re-ID	93.3	87.1
NAE	-0.7	-0.3

Table 1. The detection performance of person search model with/without re-ID losses.

To understand how the detection performance is affected by the re-ID task, we conduct ablation study by setting the re-ID loss weight as 0 during training. The result is listed in Table.1. We can see that adding re-ID loss decreases Recall and AP for both OIM-base and our NAE model. This phenomenon indicates that the re-ID task harms the detection performance due to their contradictory objectives, *i.e.* re-ID focuses on the person *uniqueness* while detection focuses on person *commonness*. However, the detection performance of our NAE model is less infected by the re-ID loss, which means disentangling detection and re-ID into embedding norm and angle is an effective strategy to reconcile the two tasks.

### 2. Failure Cases

In Figure 1 we show 5 representative failure cases. For case (a) to (e), the mistakes mainly result from similar appearances shared by the query person and the false positives. This indicates that our embeddings cannot focus on the fine-grained local information, thus not discriminative enough for these hard cases. Another big source of failure cases come from occlusion and crowded scene, as is represented by case (d) and (e). The query features are polluted by occlusions within the region of interest, *e.g.* ac-

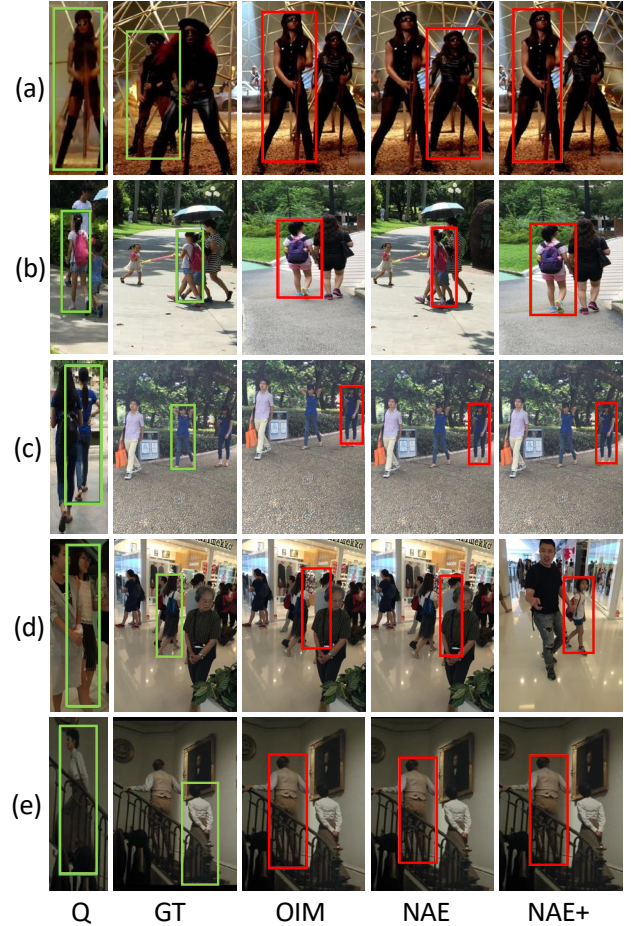


Figure 1. Failure cases for top-1 search results of OIM-base, NAE and NAE+. ‘Q’ stands for the query image and ‘GT’ is short for ground truth.

companying person and handrail. Therefore the search result will shift to the wrong target with the same occlusions. Currently these cases are hard to solve for Faster R-CNN based models. In future works, more sophisticated methods, *i.e.* attention based methods, may solve the problem, which should be an interesting research direction.