# Training Object Detectors if Only Large Objects are Labeled (Supplementary)

Daniel Pototzky[1]
daniel.pototzky@de.bosch.com
Matthias Kirschner[1]
matthias.kirschner@de.bosch.com
Azhar Sultan[1]
azhar.sultan@de.bosch.com
Lars Schmidt-Thieme[2]
schmidt-thieme@ismll.de

[1] Robert Bosch GmbH, Germany
[2] University of Hildesheim, Germany

In this supplementary document, we provide additional ablations on CityPersons and MS COCO (see sections 1 and 2). In section 3 we include a qualitative analysis of pseudo-labels generated by PCIS if only the largest 1% of objects are annotated.

## 1 Ablations on CityPersons

| Probability of downscaling | | | | | |
|---|---|---|---|---|---|
| 100% (=PCIS) | 80% | 60% | 40% | 20% | 0% |
| **13.8** | 15.9 | 19.4 | 21.0 | 30.8 | 45.7 |

Table 1: Effect of downscaling images on the performance of PCIS. Experiments show the miss rate when using CityPersons with 50% of annotations. 'Probability of downscaling' indicates the likelihood that a pseudo-labeled image was scaled in the range 608-1024, the upper limit being the original size. In PCIS, every image is downscaled after pseudo-label generation. If less than 100% of images are downscaled, performance drops substantially.

Table 1 investigates the effect of downscaling training images on the performance of PCIS. In PCIS, every image is downscaled after pseudo-label creation. If this is not done, performance drops substantially. Table 2 analyses the performance of PCIS if the labeled objects are not precisely the largest ones, but only among the largest ones. This can be seen as a test of a real-world scenario, in which an annotator screens a dataset for large objects and only labels a selection of them. Among the largest objects illustrates the case that half of the largest instances are randomly selected and annotated. For example, a random selection of the largest 20% of annotations equals 10% of labels overall. Whether the labeled objects are exactly the largest ones or just among large objects has a small effect on performance.
Figure 1 shows the average number of pseudo-labels per image in the course of training. In general, the number of pseudo-labels per image depends on the dataset, the share of labeled

| Method | 0.5% | 1% | 5% | 10% | 25% |
|---|---|---|---|---|---|
| Exactly largest objects + PCIS | 61.2 | 41.3 | 24.3 | 19.9 | 17.8 |
| Among largest objects + PCIS | 61.4 | 41.7 | 25.0 | 21.9 | 18.6 |

Table 2: MR on CityPersons by PCIS if the labeled objects are not precisely the largest ones, but only among the largest. The percentages in the top row indicate the overall share of labeled objects. Among largest objects illustrates the case that half of the largest instances are randomly selected and annotated. For example, a random selection of the largest 20% of annotations equals 10% of labels overall.

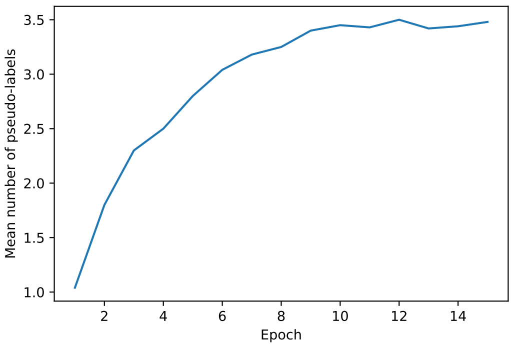


Figure 1: Mean number of pseudo-labels per image in the course of training in case of CityPersons with 50% of annotations. An increasing number of pseudo-labels is generated as the model gradually improves during training.

objects as well as the epoch in training. Figure 2 analyses the number and size of pseudo-

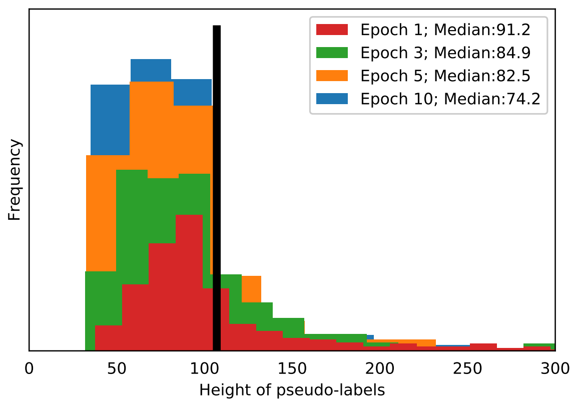


Figure 2: Size and number of pseudo-labels generated during training in case of CityPersons with 50% of annotations. The smallest object annotated in this subset has a height of 105 pixels (see black vertical line). In the course of training an increasing number of pseudo-labels are generated with the median size decreasing from 91.2 to 74.2 pixels.

labels generated in the course of epochs in case of having access to the largest 50% of annotations. The smallest available annotation has a height of 106 pixels, indicated by the black horizontal line. In the early stages of training, few pseudo-labels are created with a median at 91 pixels, only somewhat below the smallest known annotation. As training progresses, more and more pseudo-labels are generated and their median size constantly decreases to 74 pixels. Figure 3 investigates the IOU between pseudo-labels and (unlabeled)

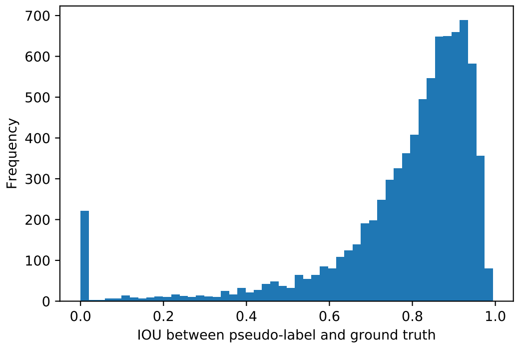


Figure 3: IOU between pseudo-labels and (unlabeled) ground truth objects in case of CityPersons with 50% of annotations. The vast majority of pseudo-labels has a high IOU with a missing annotation. There are also some false positive pseudo-labels, which do not overlap any known annotation. However, it should be noted that some of these detections might actually be correct, and instead, the CityPersons dataset may not be perfectly labeled.

ground truth objects. As can be seen, a large number of pseudo-labels correctly matches an unlabeled object.

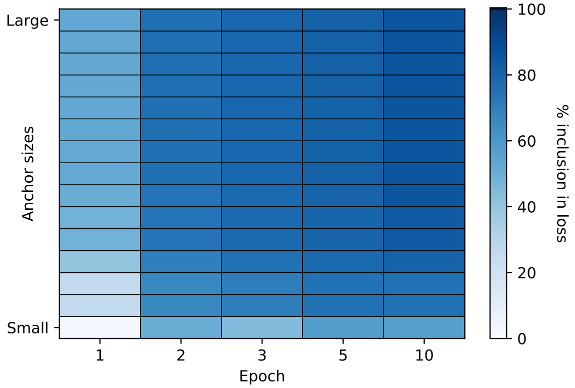


Figure 4: Percentage of anchors included in the loss function depending on anchor size and epoch number when using 50% of CityPersons annotations. The smaller the anchor, the less likely it is to be included. In the course of training, more and more pseudo-labels are created and an increasing share of anchors is included in the loss.

Figure 4 investigates the share of anchors included in the loss function in the course of train-

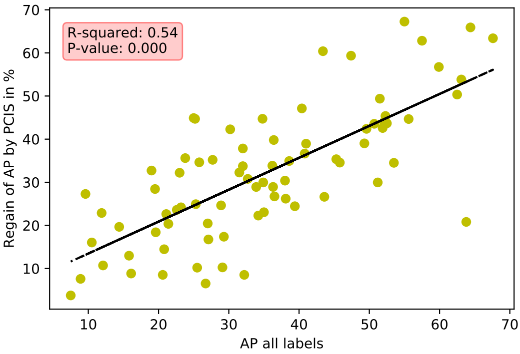


Figure 5: Analysis of factors that affect the performance of PCIS. Our performance metric (y-axis) measures what percentage of the gap in AP between sparse labels and all labels can be regained using PCIS. Each dot corresponds to one class in MS COCO. Applying our method is especially beneficial for classes that achieve a high AP if all labels are available (see x-axis).

ing. In the beginning, only a subset of the anchors is incorporated. As training progresses, more and more pseudo-labels are generated and an increasing amount of anchors is included.

# 2 Ablations on MS COCO

Furthermore, we investigated what properties affect the performance of PCIS by analyzing the results on 80 classes of MS COCO. As a measure for performance, we used what percentage of the gap in AP between sparse labels and all labels can be regained using PCIS. We analyzed the connection of this performance metric with different components using linear regression. Our results indicate that by far the most important factor is the AP per class if all labels are available (see Fig. 5). Similar results as the AP using all labels (although less statistically significant) showed the related factor of AP per class using only sparsely labeled images. Conversely, we found other class statistics such as number of images, number of bounding boxes, median object size, mean object size, distribution of object size to be statistically insignificant, i.e. with a low r-squared and a p-value above 0.05.

# 3 Qualitative Results

Finally, we visualize examples of pseudo-labels (green boxes) generated by PCIS if only the largest 1% of annotations (blue boxes), equivalent to 126 bounding boxes, are available for training on CityPersons (see Fig. 6). In the leftmost image on the top row, example annotations are shown. Their height exceeds 537 pixels, meaning that they are more than ten times larger than the smallest pedestrian considered in the evaluation metric. Although only very few, extremely large pedestrians are labeled, correct pseudo-labels are created for a substantial part of missing annotations.

In addition, we also investigate failure cases (see Fig. 7). These are not specific to our method. False positives and false negatives occur when generating pseudo-labels. For some

Figure 6: Examples for pseudo-labels (green boxes) if only the largest 1% of pedestrians are annotated (blue boxes). Although only very few, extremely large pedestrians are labeled, correct pseudo-labels are created for a substantial part of missing annotations.

strongly overlapping objects, only one pseudo-label is generated (e.g. multiple pedestrians crossing a street). Some occluded objects are not detected at all (e.g. a pedestrian behind a car). In general, failure cases occur more often for objects that are strongly occluded, very small, and which have poor illumination.


Figure 7: Failure cases for pseudo-labels (green boxes) if only the largest 1% of pedestrians are annotated (blue boxes).