CityPersons: A Diverse Dataset for Pedestrian Detection
Supplementary material

Shanshan Zhang[1,2], Rodrigo Benenson[2], Bernt Schiele[2]
[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[2]Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
shanshan.zhang@njust.edu.cn, firstname.lastname@mpi-inf.mpg.de

1. Content

In this supplementary material, we will show some more illustrations, discussions and experiments for the CityPersons dataset.

- Section 2 shows some examples of our annotations.
- Section 3 provides some analysis regarding the annotations, including height statistics (section 3.1), analysis experiments regarding annotation quality (section 3.2).

2. CityPersons annotation examples

In figure 1, we show some examples of our bounding box annotations and Cityscapes segmentation annotations from different cities. We can see the diversity in terms of people's appearance, clothing, and background objects.

3. Analysis of CityPersons annotations

In this section, we provide some analysis regarding the height statistics and quality of CityPersons annotations.

3.1. Height statistics

In figure 2, we compare the height distribution of CityPersons and Caltech. The CityPersons is more diverse than Caltech in terms of scale:

(1) CityPersons covers a larger range of height, as it consists of larger images.

(2) More than 70% of Caltech pedestrians fall in one single bin [50,100], while CityPersons are more evenly distributed in different scale ranges.

3.2. Quality

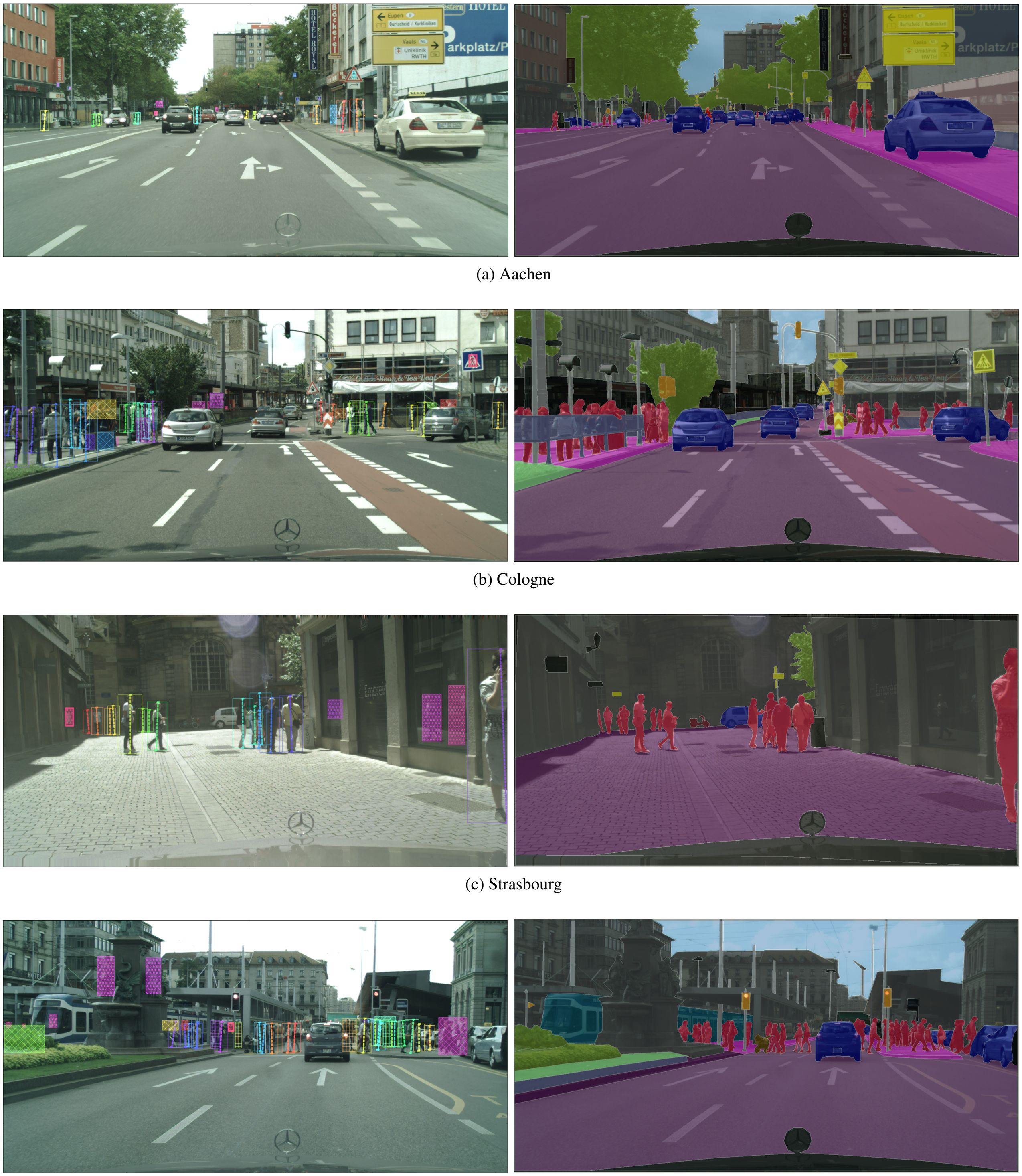
The segment for each person only reflects the visible part, while losing information of the aligned full body. In [1], it is shown that better alignment of training annotations improve the detection quality a lot. Therefore, in this paper we aim to provide high quality well aligned annotations for each pedestrian. On the other hand, as shown in the second section of the main paper, properly handling ignore regions also affects the results, so we also make efforts to label ignore regions over all images.

Annotation aspect	MR	ΔMR
segment bounding boxes	22.54	-
+ ignore regions	21.31	+ 1.23
+ better boxes	15.14	+ 6.17
our annotations	15.14	+ 7.40

Table 1: The effects on performance of using high quality training annotations. CityPersons validation set evaluation. Training with our aligned bounding box annotations and ignore region annotations gives better performance than training with segment bounding box annotations.

In table 1, we show that our high quality annotations improve the performance by ~7 pp, among which ~6 pp is gained from better alignment, and another ~ 1pp from ignore regions handling.

Another argument for our aligned bounding box annotations is the comparison of performance on an external benchmark (Caltech) using two types of training annotations. From table 2, we can see the model trained with segment bounding boxes fails not only on CityPersons, but also on Caltech. The reason is other benchmarks, e.g. Caltech, also provide aligned bounding box annotations. Therefore, using our annotations helps to train a better generalizable model over multiple benchmarks.



(a) Aachen

(b) Cologne

(c) Strasbourg

(d) Zurich

Figure 1: Examples of annotations from different cities. Left: our bounding box annotations; right: Cityscapes segmentation annotations. For visualization, we use different masks for pedestrians/riders, sitting persons, other persons, group of people, and ignore regions.

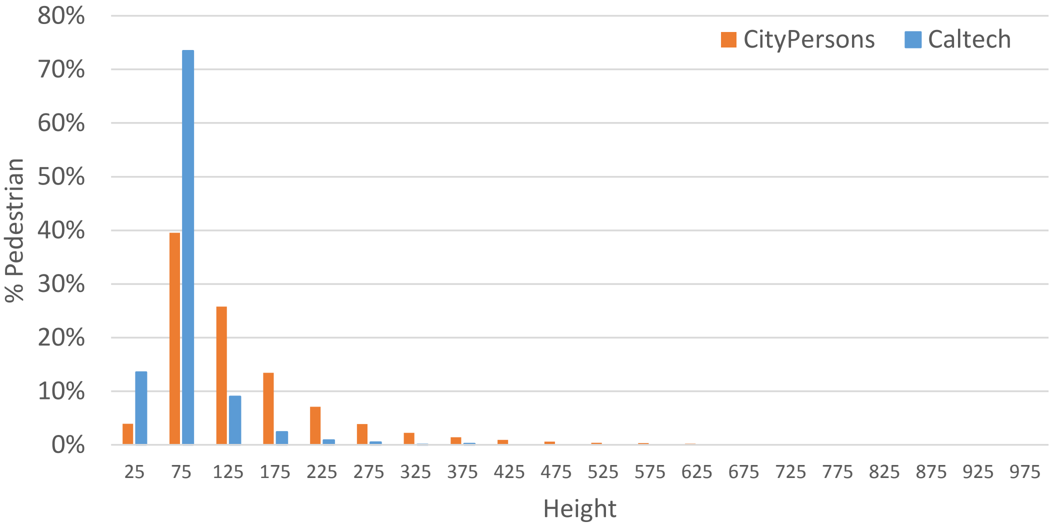


Figure 2: Height distributions of CityPersons and Caltech.

Train anno. / Test set	Seg. bounding box	Aligned bounding box
Caltech	37.5	26.9
CityPersons	22.5	15.1

Table 2: Comparison of performance using two types of training annotations. Numbers are MR on CityPersons validation set; and MR^O on Caltech test set. Using our aligned bounding box for training obtains better quality on both Caltech and CityPersons.

References

[1] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele. How far are we from solving pedestrian detection? In *CVPR*, 2016. 1