CityPersons: A Diverse Dataset for Pedestrian Detection

Shanshan Zhang[1,2], Rodrigo Benenson[2], Bernt Schiele[2]
[1]Nanjing University of Science and Technology [2]Max Planck Institute for Informatics
shanshan.zhang@njust.edu.cn, rodrigo.benenson@gmail.com, bernt.schiele@mpi-inf.mpg.de

Data & Benchmark Available Online!
https://bitbucket.org/shanshanzhang/citypersons/

Introduction

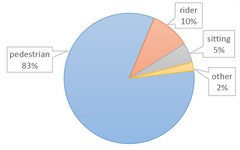
CityPersons

Cityscapes
- Instance masks
- Video frames
- Stereo, GPS, et al.

- Aligned bounding boxes
- Occlusion level estimates
- Fine-grained categories

Statistics

	Train	Val.	Test	Sum
# cities	18	3	6	27
# images	2 975	500	1 575	5 000
# persons	19 654	3 938	11 424	35 016
# ignore regions	6 768	1 631	4 773	13 172



Baseline Detector

Proper adaptations to vanilla FasterRCNN are necessary to obtain competitive results for pedestrian detection.

Detector aspect	MR^O	ΔMR
FasterRCNN-vanilla	20.98	-
+ quantized rpn scales	18.11	+ 2.87
+ input up-scaling	14.37	+ 3.74
+ Adam solver	12.70	+ 1.67
+ ignore region handling	11.37	+ 1.33
+ finer feature stride	10.27	+ 1.10
FasterRCNN-ours	10.27	+ 10.71

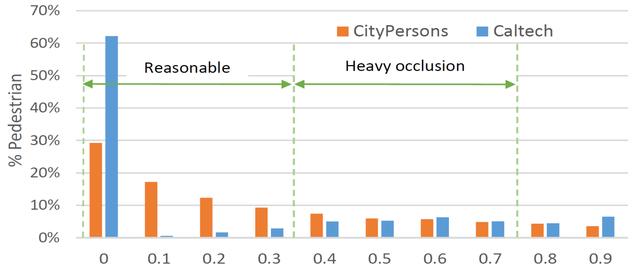
References
[1] Ren et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
[2] Cordts et al. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.
[3] Dollár et al. Pedestrian detection: An evaluation of the state of the art. PAMI, 2012.
[4] Geiger et al. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[5] Zhang et al. How far are we from solving pedestrian detection? In CVPR, 2016.
[6] Benenson et al. Ten years of pedesrian detection, what have we learned? ECCVw, 2014.

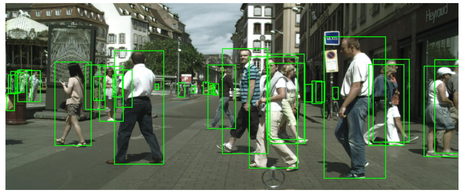
Comparison to Previous Datasets

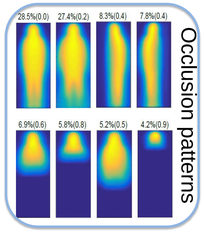
- **More diverse**

	Caltech	KITTI	CityPersons
# country	1	1	3
# city	1	1	18
# season	1	1	3
# person/image	1.4	0.8	7.0
# unique person	1 273	6 336	19 654

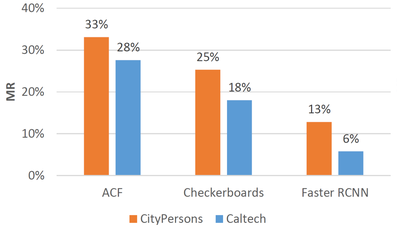
- **More occlusions**

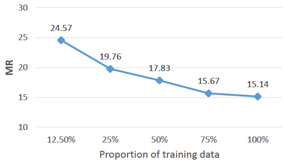






- **More challenging**

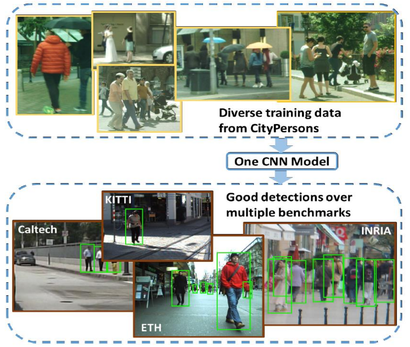




Benefits from CityPersons

- **Diversity enhances generalization ability across datasets**

Train / Test	Caltech	KITTI	CityPersons
Caltech	*10.27*	46.86	**21.18**
KITTI	10.50	*8.37*	**8.67**
CityPersons	**46.91**	51.21	*12.81*
INRIA	11.47	27.53	**10.44**
ETH	57.85	49.00	**35.64**
Tud-Brussels	42.89	45.28	**36.98**
mean MR	29.98	38.04	**20.95**



- **Better pre-training improves**
  - **Detection accuracy (esp. small scale, heavy occlusion)**
  - **Alignment quality**

O/N	Setup	Scale range	IoU	Caltech	CityPersons →Caltech	ΔMR
MR^O	Reasonable	[50, ∞]	0.5	10.3	9.2	+ 1.1
MR^O	Smaller	[30, 80]	0.5	52.0	48.5	**+ 3.5**
MR^O	Heavy occl.	[50, ∞]	0.5	68.3	57.7	**+ 8.6**
MR^N	Reasonable	[50, ∞]	0.5	5.8	5.1	+ 0.7
MR^N	Reasonable	[50, ∞]	0.75	30.6	25.8	**+ 4.8**

Setup	Scale range	IoU	KITTI	CityPersons →KITTI	ΔMR
Reasonable	[50, ∞]	0.5	8.4	5.9	+ 2.5
Reasonable	[50, ∞]	0.75	43.3	39.2	**+ 4.1**
Smaller	[30, 80]	0.5	37.8	27.1	**+ 10.7**

- **Semantic segmentation helps to detect small persons**

Scale range	Baseline	+ Semantic	ΔMR
[50, ∞]	15.4	14.8	+ 0.6
[100, ∞]	7.9	8.0	+ 0.1
[75, 100]	7.2	6.7	+ 0.5
[50, 75]	25.6	22.6	**+ 3.0**

(a) Original image (b) Semantic map

CityPersons: a new dataset for pedestrian detection
- Rich & diverse
- Improves quality of existing models
- New challenges motivate future research

More data to explore!