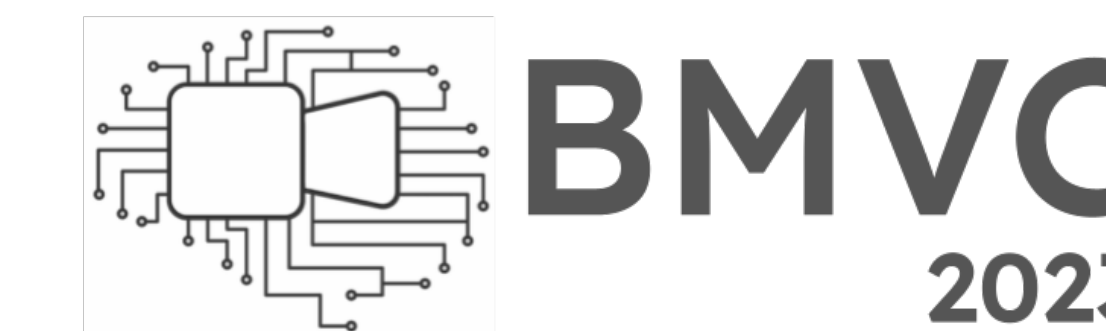


A Comprehensive Crossroad Camera Dataset of Mobility Aid Users

L. Mohr, N. Kirillova, H. Possegger and H. Bischof

{ludwig.mohr,nadezda.kirillova,possegger}@icg.tugraz.at

Graz University of Technology, Institute of Computer Graphics and Vision, Austria



Pedestrian Detection in Traffic Monitoring

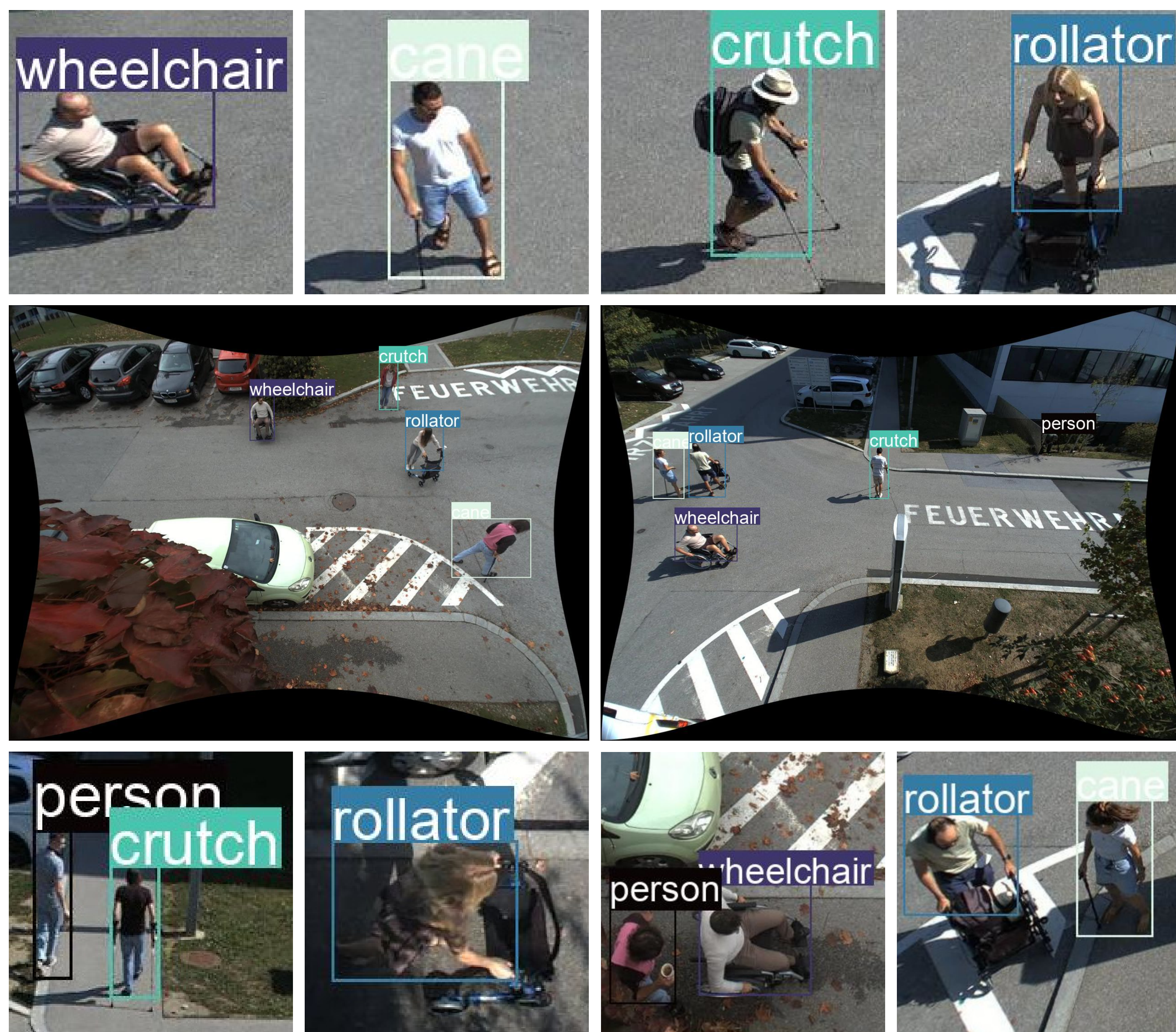
Pedestrians using mobility aids are:

- the most vulnerable group of traffic participants,
- highly underrepresented to not represented at all in commonly available datasets for detection.

We provide a dataset for detection and classification of pedestrians using mobility aids from a traffic monitoring view point, together with baseline evaluations for common classification and detection approaches.

Dataset: Overview

- Captured at a crossroad on the university campus.
- Elevated traffic monitoring view point.
- Captured on three days in summer and autumn.
- Mobility aids present: *wheelchair*, *rollator*, *crutch*, and *walking cane*.
- Captured with the help of 34 able-bodied participants.
- Image resolution of 1280 by 960 pixels.



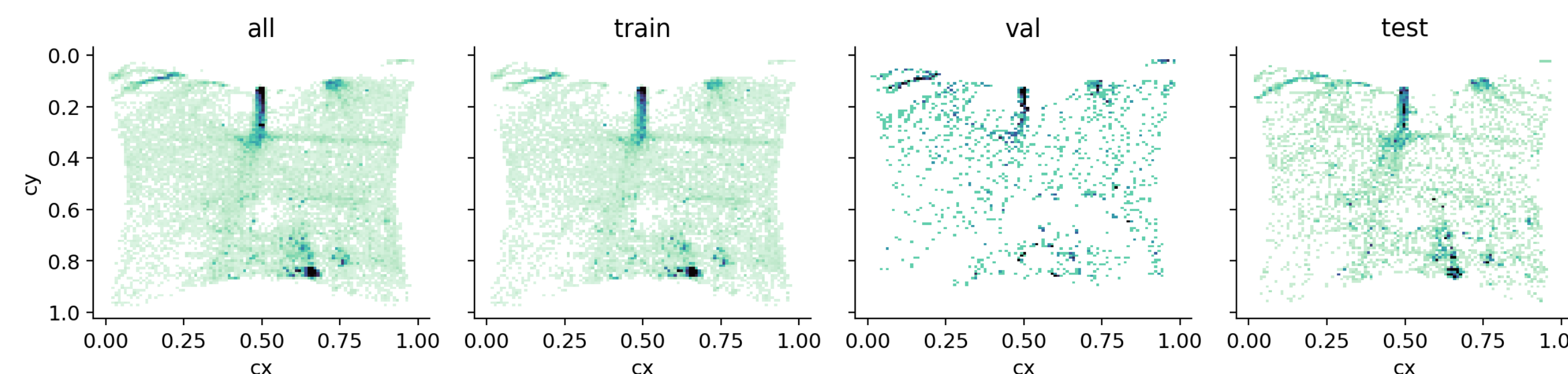
Dataset: Analysis

The dataset consists of 12186 frames in total, with official *training*, *validation*, and *test* splits without temporal overlap. All classes are well represented in each split, visible from the table below listing the number and relative distribution of class annotations per split.

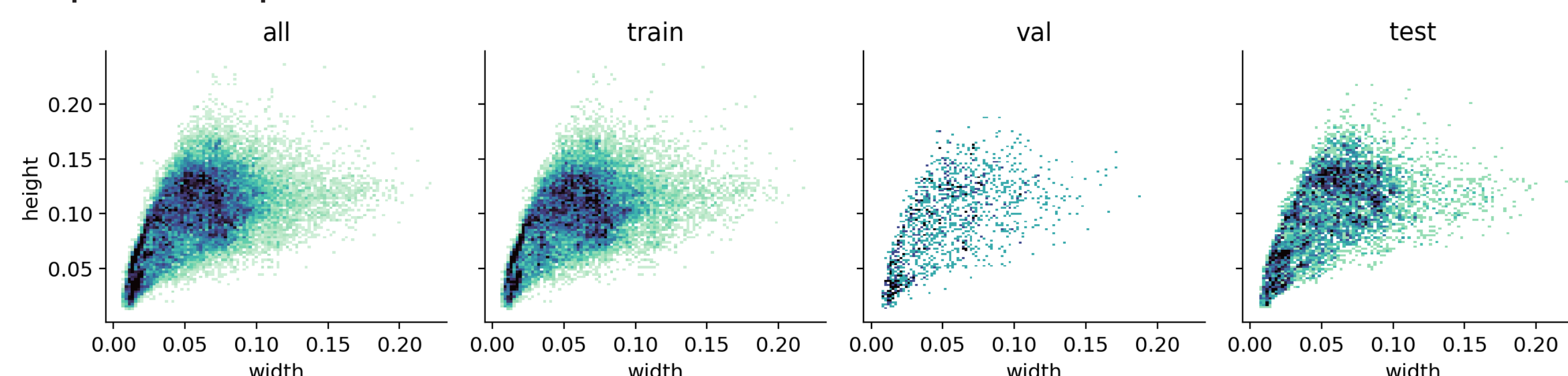
split	frames			annotations									
	total	background	(%)	pedestrian	wheelchair	rollator	crutch	cane	(%)				
train	8456	768	(9.0%)	10518	(48.1%)	3907	(17.9%)	3178	(14.5%)	2466	(11.3%)	1781	(8.2%)
val	784	71	(9.0%)	486	(30.8%)	418	(26.5%)	347	(22.0%)	153	(9.7%)	176	(11.1%)
test	2946	267	(9.0%)	3058	(43.5%)	1660	(23.6%)	787	(11.2%)	979	(13.9%)	540	(7.7%)

The splits represent the dataset well, it does not exhibit bias towards preferred locations for any class relative to the others.

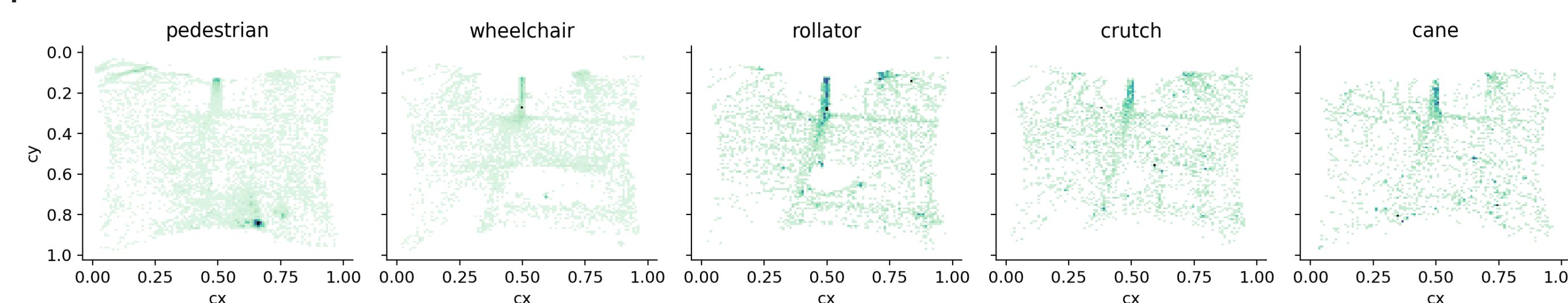
Distribution of bounding box centers for the complete dataset (left) and each split:



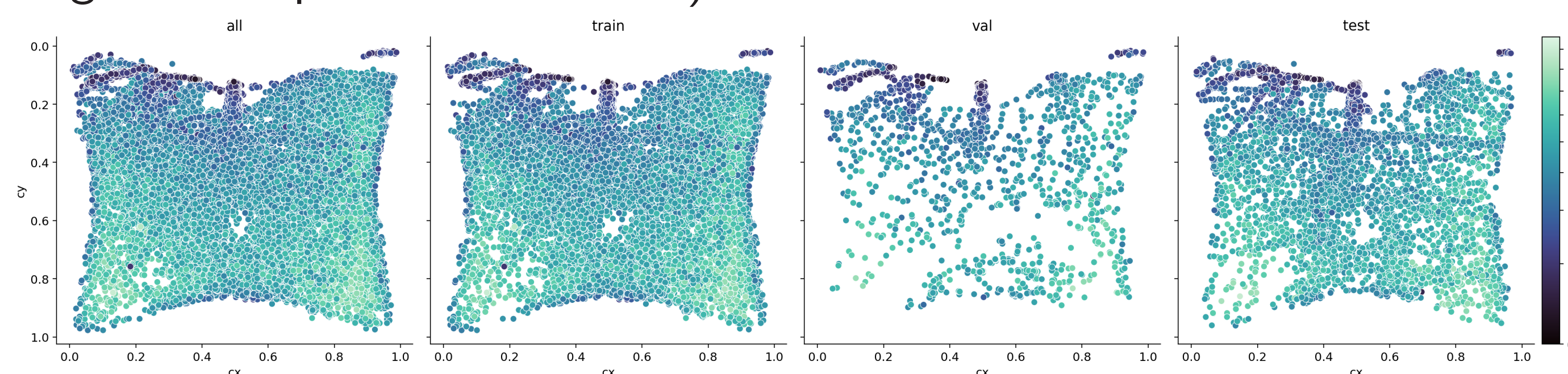
Distribution of bounding box sizes for the complete dataset (left) and the respective splits.



Distribution of labels (centers of respective annotation bounding boxes) per class.



Distribution of annotations, color coded for bounding box size (equivalent length of a square of same area).



Baselines: Classification

Classification performance of several commonly used feature backbones using ground truth boxes as detections. All networks were fine-tuned for 20 epochs using weights pre-trained on ImageNet-1K.

	pedestrian GT 3058		wheelchair GT 1660		rollator GT 787		crutch GT 979		cane GT 540		all GT 7024	
	TP [†]	Acc [†]	TP [†]	Acc [†]	TP [†]	Acc [†]	TP [†]	Acc [†]	TP [†]	Acc [†]	TP [†]	Acc [†]
MobileNetV3 L	3041	0.9944	1642	0.9892	751	0.9543	732	0.7477	424	0.7852	6590	0.9382
MobileNetV3 L (V2)	3044	0.9954	1642	0.9892	744	0.9454	745	0.7610	421	0.7796	6596	0.9391
ResNet18	3053	0.9984	1650	0.9940	764	0.9708	753	0.7992	449	0.8315	6669	0.9495
ResNet34	3056	0.9993	1645	0.9910	775	0.9848	726	0.7416	440	0.8148	6642	0.9456
ResNet50	3057	0.9997	1651	0.9946	763	0.9695	776	0.7926	430	0.7963	6677	0.9506
ResNet50 (V2)	3056	0.9993	1645	0.9910	771	0.9797	764	0.7804	436	0.8074	6672	0.9499
DenseNet201	3053	0.9984	1651	0.9946	764	0.9708	789	0.8059	459	0.8500	6716	0.9562
ResNet152	3054	0.9987	1654	0.9964	764	0.9708	792	0.8090	445	0.8241	6709	0.9552
ResNet152 (V2)	3053	0.9984	1642	0.9892	771	0.9797	788	0.8049	438	0.8111	6692	0.9527
VIT-B/16	3054	0.9987	1648	0.9928	760	0.9657	711	0.7263	397	0.7352	6570	0.9354
VGG16	3054	0.9987	1648	0.9928	758	0.9632	771	0.7875	431	0.7981	6662	0.9485

Baselines: Detection

All models have been fine-tuned on our dataset using the official weights trained on COCO for a maximum of 200 epochs using default hyperparameters of YOLO. Models prefixed with *h* are trained using a hierarchical class structure.

Detection performance of selected YOLOv5 models in terms of mean average precision (mAP) metrics mAP@50 and mAP@50-95:

YOLOv5	pedestrian		wheelchair		rollator		crutch		cane		all	
	AP@50	mAP@50-95	AP@50	mAP@50-95	AP@50	mAP@50-95	AP@50	mAP@50-95	AP@50	mAP@50-95	mAP@50	mAP@50-95
small	0.881	0.650	0.983	0.721	0.952	0.693	0.821	0.695	0.674	0.569	0.862	0.666
medium	0.886	0.669	0.986	0.719	0.960	0.706	0.843	0.719	0.639	0.545	0.863	0.672
large	0.891	0.671	0.982	0.727	0.972	0.712	0.840	0.720	0.669	0.568	0.871	0.679
xlarge	0.887	0.682	0.982	0.735	0.954	0.708	0.858	0.732	0.716	0.612	0.880	0.694
h small	0.860	0.637	0.981	0.722	0.941	0.699	0.760	0.666	0.583	0.508	0.825	0.647
h medium	0.853	0.642	0.985	0.734	0.949	0.724	0.796	0.699	0.660	0.582	0.849	0.676
h large	0.859	0.652	0.981	0.736	0.953	0.726	0.788	0.689	0.632	0.548	0.843	0.670
h xlarge	0.852	0.647	0.976	0.747	0.942	0.717	0.820	0.723	0.670	0.597	0.852	0.686

Detection performance in terms of accuracy (ACC), misclassification rate (MCL) and missed (MIS) detections for each class as well as overall false negative (FN) and false positive (FP) detections:

YOLOv5	pedestrian			wheelchair			rollator			crutch			cane			all			
	ACC [†]	MCL [‡]	MIS [‡]	ACC [†]	MCL [‡]	MIS [‡]	ACC [†]	MCL [‡]	MIS [‡]	ACC [†]	MCL [‡]	MIS [‡]	ACC [†]	MCL [‡]	MIS [‡]	ACC [†]	MCL [‡]	FN [‡]	
small	0.874	0.070	0.055	0.967	0.006	0.027	0.940	0.034	0.025	0.653	0.326	0.011	0.693	0.289	0.019	0.859	0.105	430	255
medium	0.880	0.049	0.071	0.965	0.008	0.027	0.924	0.053	0.023	0.664	0.322	0.014	0.598	0.369	0.033	0.863	0.103	440	312
large	0.880	0.049	0.071	0.965	0.008	0.027	0.924	0.053	0.023	0.664	0.322	0.014	0.598	0.369	0.033	0.853	0.103	440	312
xlarge	0.897	0.054	0.048	0.968	0.009	0.024	0.921	0.051	0.028	0.701	0.290	0.009	0.643	0.341	0.017	0.870	0.098	609	227
h small	0.854	0.076	0.070	0.964	0.010	0.027	0.917	0.064	0.019	0.590	0.395	0.014	0.598	0.389	0.013	0.831	0.127	146	295
h medium	0.846	0.105	0.049	0.966	0.015	0.019	0.939	0.047	0.014	0.662	0.333	0.005	0.657	0.330	0.013	0.844	0.126	278	206
h large	0.854	0.094	0.052	0.963	0.016	0.021	0.943	0.041	0.017	0.637	0.349	0.013	0.641	0.337	0.022	0.843	0.124	274	232
h xlarge	0.818	0.123	0.059	0.939	0.033	0.029	0.937	0.039	0.024	0.672	0.322	0.006	0.587	0.394	0.019	0.822	0.141	226	264

Dataset and Code Available Online

Data:



Code:

