

# The EuroCity Persons Benchmark (ECPB)

Braun, M., Krebs, S., Flohr, F. and Gavrilă, D.

## 1) Contact

Please feel free to contact us with any questions, suggestions or comments:

Markus Braun (Daimler AG, Dep. of Environment Perception)

Email: [contact@eurocity-dataset.tudelft.nl](mailto:contact@eurocity-dataset.tudelft.nl)

## 2) Citation

When using the dataset please add the following citation:

Braun, M., Krebs, S., Flohr, F. and Gavrilă, D.,

EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes,

IEEE transactions on pattern analysis and machine intelligence,

2019

## 3) License agreement

## 4) General - The EuroCity Persons Benchmark

The EuroCity Persons Benchmark provides a benchmark dataset for person detection.

Please see <https://eurocity-dataset.tudelft.nl/eval/overview/overview> for an overview.

We also provide demo videos with annotation examples. Maybe these videos help to decide if this dataset is appropriate for you: <https://eurocity-dataset.tudelft.nl/eval/overview/examples>

All objects were annotated with tight bounding boxes of the complete extent of the entity. If an object is partly occluded, its full extent was estimated (this is useful for later processing steps such as tracking) and the level of occlusion was annotated. We discriminated between no occlusion, low occlusion (10%-40%), moderate occlusion (40%-80%), and strong occlusion (larger than 80%). Similar annotations were performed with respect to the level of object truncation at the image border (here, full object extent was not estimated). For riders, we labeled the riding person and its ride-vehicle with two separate bounding boxes, and annotated the ride-vehicle type. Riderless vehicles of the same type in close proximity were captured by one class-specific group box (e.g. several bicycles on a rack).

We remained with the classical bounding-box convention of labeling the outermost object parts. For every sampled frame, all visible persons were annotated; otherwise, missed annotations could lead to the flawed generation of background samples during training and bootstrapping. Also persons in non-upright poses (e.g. sitting, lying) were annotated or persons behind glass. These cases were tagged separately. A person is annotated with a rectangular (class-specific) ignore region if a person

is smaller than 20px, if there are doubts that an object really belongs to the appropriate class, and if instances of a group can not be discriminated properly. In the latter case, several instances may be grouped inside a single ignore region.

The overall object orientation is an important cue for the prediction of future motion of persons in traffic scenes. We provide this information for all persons larger than 40px (including those riding).

Person depictions (e.g. large poster) and reflections (e.g. in store windows) were annotated as a separate object class. Additional events were tagged at the image level, such a lens flare, motion blur, and rain drops or a wiper in front of the camera.

## Data subsets

We define various data subsets on the overall EuroCity Persons dataset. First, we distinguish a day-time and a night-time data subset, each with its own separate training, validation and test set. Three overlapping data subsets are furthermore defined, considering the ground-truth annotations:

- Reasonable: Persons with a bounding box height greater than 40px which are occluded/truncated less than 40%
- Small: Persons with a height between 30px and 60px which are occluded/truncated less than 40%
- Occluded: Persons with a bounding box height greater than 40px which are occluded between 40% and 80%.

These data subsets can be used to selectively evaluate properties of person detection methods for various sizes or degrees of occlusion.

We split our dataset into training, validation and test by 60%, 10%, and 30% respectively. For the training and validation subsets we provide images as well as groundtruth annotations (see Section 5)).

For the test subset we only provide images. If you want to participate in our benchmark you may upload your detections on the test subset to our evaluation server. For metrics and further details please see our publication (Paper reference in Section 2)).

Feel free to use this dataset for academic research. See the licence for further information. Please add the above mentioned citation to your publication.

If you detect any error in the dataset (missing images, wrong labels etc.) please contact us so that we can improve this dataset. Thank you.

## 5) Setup

Clone this repository to any {root} directory.

```
git clone https://github.com/SomeNeurons/ECPB.git {root}
```

Download all of the following files and extract them to the given destination directory

Data destination path	Link	MD5 hash sum
{root}/data/day/img/test	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpda%20day_img_test.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpda%20day_img_test.zip</a>	bfe692154144e8fdd0913061b63bc94d

{root}/data/day/img/train	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_img_train.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_img_train.zip</a>	bdaca684ffae5bf9bba95b6a9934ade2
{root}/data/day/img/val	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_img_val.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_img_val.zip</a>	5598524609a00ad9808624d7e7b4b6a1
{root}/data/day/labels/train	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_labels_train.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_labels_train.zip</a>	73baae652564cbadb14a2fd8e3f61c4d
{root}/data/day/labels/val	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_labels_val.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_day_labels_val.zip</a>	4b7cd3953c002c7a5235db807e6a326d
{root}/data/night/img/test	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_test.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_test.zip</a>	bfe692154144e8fdd0913061b63bc94d
{root}/data/night/img/train	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_train.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_train.zip</a>	c5cee39a4aa9c3488e429801531d1f34
{root}/data/night/img/val	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_val.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_img_val.zip</a>	f151b4d290a6fe152f03bd587b7575d6
{root}/data/night/labels/train	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_labels_train.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_labels_train.zip</a>	76bcfca00f61317b86fea9284f07093e
{root}/data/night/labels/val	<a href="http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_labels_val.zip">http://eurocity-dataset.tudelft.nl/eval/downloadFiles/downloadFile/index?file=ecpdata%2FECP_night_labels_val.zip</a>	fc708c6481f5cc0d85c359970db178a6

Note that we only provide images and no labels for the test datasets.

## 6) Dataset Structure

The folder structure of the EuroCity Persons Benchmark is organized as

***{root}/data/{time}/{type}/{split}/{city}/{city}\_{frame}.{ext}***

The meaning of the individual elements is:

- **root**: the root folder of the EuroCity Persons Benchmark. Our scripts are using paths relatively to this folder.
- **time**: the time of day, i.e. 'day' or 'night'
- **type**: the type/modality of data, e.g. 'labels' for ground truth, or 'img' for 8 bit png images.

- **split**: the split, i.e. 'test', 'train' or 'val'
- **city**: the city the recording took place in
- **frame**: the frame number using 5 digits.
- **ext**: the fileextension, 'png' for images, 'json' for groundtruth labels

## 7) JSON-Layout

### General

Within this dataset each frame is represented by one json-file with a predefined format. The name of the json-file follows the scheme: **{city}\_{frame}.json**

We use the same format for groundtruth and detection files.

Attention: During the evaluation process the correspondence is performed based on the filename of the detection and ground truth files.

Ground truth annotation and detection fields	Variable type	Meaning
x0, x1, y0, y1	[unsigned integer]	Defining the position and size of the bounding box by its upper left (x0, y0) and bottom right point (x1, y1) x0, x1 in [0, 1920] y0, y1 in [0, 1024] object width w = x1 – x0 object height h = y1 – y0
identity	[string]	Specifies the object class. Possible values are ["pedestrian", "rider", "person-group-far-away", "rider+vehicle-group-far-away", "bicycle-group", "buggy-group", "motorbike-group", "scooter-group", "tricycle-group", "wheelchair-group"] for the top-level entities and ["bicycle ", "buggy ", "motorbike ", "scooter ", "tricycle ", "wheelchair "] for the ride-vehicles (child-entities of riders).  "pedestrian" and "rider" are used for single instances of the corresponding class. Ignore regions are labeled with "person-group-far-away" for pedestrians and "rider+vehicle-group-far-away" for riders (see Section 4) and 9) for the meaning of ignore regions). The vehicle-groups are used for vehicles without corresponding rider. Several vehicles of the same type may be grouped together in a single box.
Orient	[float]	The orientation of a person or the ride-vehicle.
<b>Groundtruth only fields</b>		
tags	[list of strings]	List (possible empty) of specified tags. Possible tags are ["occluded>10", "occluded>40", "occluded>80",

		"truncated>10", "truncated >40", "truncated >80", "sitting-lying", "skating", "behind-glass", "depiction", "unsure-orientation"]
children	[list]	For riders we provide the bounding box for the ride-vehicle as a subentity in this children list. This subentity uses the same fields for the bounding box coordinates, the identity and the orientation value.
<b>Detection only fields</b>		
score	[float]	Detection score (confidence) of the given detection.

## 8) Scripts

### **{root}/detect.py**

Run this script to generate detection files in a format readable by the eval.py script and our evaluation server. Select the appropriate subset by setting time (day or night) and mode (val or test) in line 44

```
run_detector_on_dataset(time='day', mode='val')
```

It automatically calls the mock\_detector method in line 9 and creates a single detection file for every image of the given subset.

You may return the results of your detector instead of the mock\_detections while keeping the given format for the detections. You can use the eval.py script for evaluation on the validation dataset or upload the zipped results on the test dataset to the benchmark server for online comparison.

### **{root}/eval.py**

Use this script to evaluate your results on the validation dataset.

In line 144 you may select the appropriate data subset and the class to be evaluated (pedestrian or rider).

```
eval(time='day', mode='val', eval_type='pedestrian')
```

By default predictions are loaded from

```
{root}/data/mock_detections/{time}/{split}
```

which is also the default output path of the detect.py script.

After the evaluation you may find Miss-rate/false-positives-per-image curves for every data subset (reasonable, small, occluded) in

```
{root}/results.
```

The evaluation script on the evaluation server is identical with the eval.py script, to sustain a transparent benchmark proceeding.

### **{root}/dataconverter.py**

Provides a simple example to convert our ECP format to KITTI and back again.

### **{root}/create\_tfrecords.py**

This script may be used to create tfrecords from our ECP data usually applied in tensorflow trainings.

Select the time (day or night) in lines 19, 21 and 23.

If you want to use ignore regions (person-groups, rider-groups), you have to adapt line 122 to additionally load these classes. By default only pedestrian- and rider-boxes are loaded.

By default rider boxes are loaded without including the ridden vehicle. This may be changed in line 129.

## 9) Evaluation

### Offline

For offline evaluation on the validation subset, run the eval.py script as described in Section 8). Miss-rate/false-positives-per-image curves are plotted.

### Online benchmark

Once you want to participate in the official online benchmark, please run your approach on the provided test images (e.g. using the detect.py script) and upload your zipped results to our evaluation server: <https://eurocity-dataset.tudelft.nl/eval/submissions/submit>.

The zip has to contain a single detections file per test image. The correspondence is done based on the filename. Please see

[https://github.com/SomeNeurons/ECPB/blob/master/data/examples/detections/roma\\_01393.json](https://github.com/SomeNeurons/ECPB/blob/master/data/examples/detections/roma_01393.json) for an example of a complete result file containing detections for pedestrians and riders.

The evaluation settings of the online benchmark and the eval.py script are identical.

### Evaluation settings

For each image the set of all detections is compared to the groundtruth annotations by utilizing a greedy matching algorithm. An object is considered as detected (true positive) if the Intersection over Union (IoU) of the detection and groundtruth bounding box exceeds a pre-defined threshold. Due to the high non-rigidness of pedestrians we follow the common choice of an IoU threshold of 0.5. Since no multiple matches are allowed for one ground-truth annotation, in the case of multiple matches the detection with the largest score is selected, whereas all other matching detections are considered false positives. After the matching is performed, all non matched ground-truth annotations and detections, count as false negatives and false positives, respectively.

Neighboring classes and ignore regions are used during evaluation. Neighboring classes involve entities that are semantically similar, for example bicycle and moped riders. Some applications might require their precise distinction (enforce) whereas others might not (ignore). In the latter case, during matching correct/false detections are not credited/penalized. If not stated otherwise, neighboring classes are ignored in the evaluation. In addition to ignored neighboring classes all persons annotations with the tags behind glass or sitting-lying are treated as ignore regions. Further, ignore regions are used for cases where no precise bounding box annotation is possible (either

because the objects are too small or because there are too many objects in close proximity which renders the instance based labeling infeasible). Since there is no precise information about the number or the location of objects in the ignore region, all unmatched detections which share an intersection of more than 0.5 with these regions are not considered as false positives.