

Multi-sensor rail track detection in automatic train operations

Master's thesis in Data Science

Student: Attila Kovacs

1st Advisor: Lukas Rohatsch (FH Technikum)

2nd Advisor: Daniele Capriotti (M2C Expert Control GmbH)

Interim presentation: 16.10.2023

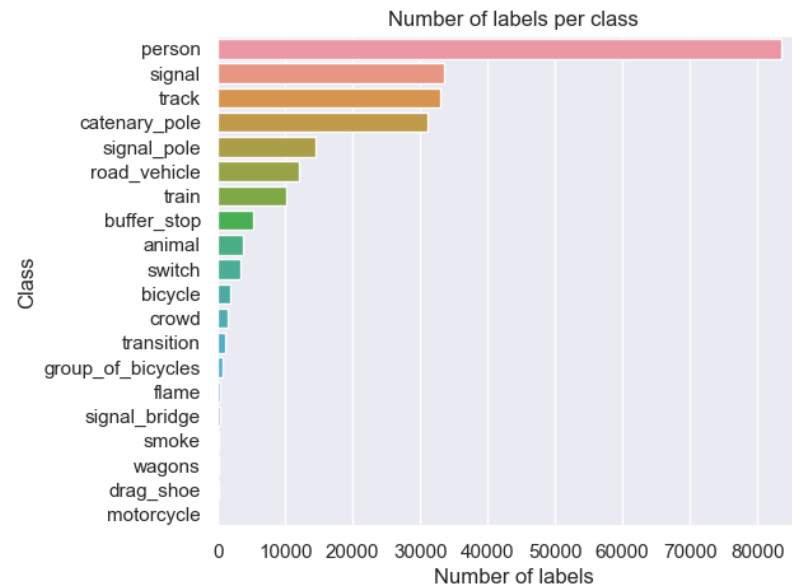


Research questions

- Which modelling technique and which deep learning model can be applied to solve the rail detection problem?
- How can the efficiency of standard, high-resolution, and infrared cameras be compared against each other; does a higher resolution result in a higher accuracy?
- What is the trade-off between model accuracy and speed of providing predictions when applied to a video stream in real time?
- How do deep learning models perform compared to gradient-based thresholding approaches in terms of, e.g., accuracy (share of correctly identified objects) or F1-score (mean of precision and recall)

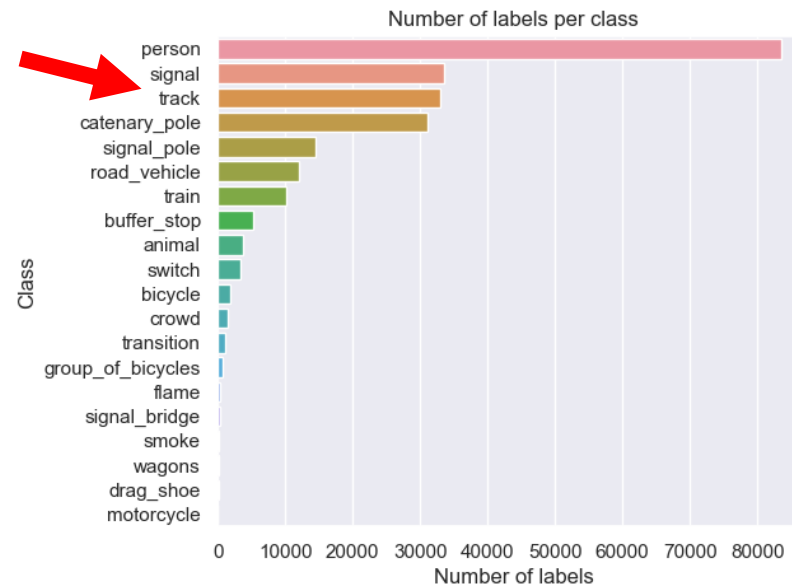
Data source

- Deutsch Bahn / Digitale Schiene Deutschland
(<https://digitale-schiene-deutschland.de/en/news/OSDaR23-multi-sensor-data-set-for-machine-learning>).
- Images were generated between 09.09.2021 and 15.09.2021
- Total number of images: 13.952
- Sensors: 12MP RGB, 5MP RGB, IR cameras, lidar, radar, GPS, inertia sensor



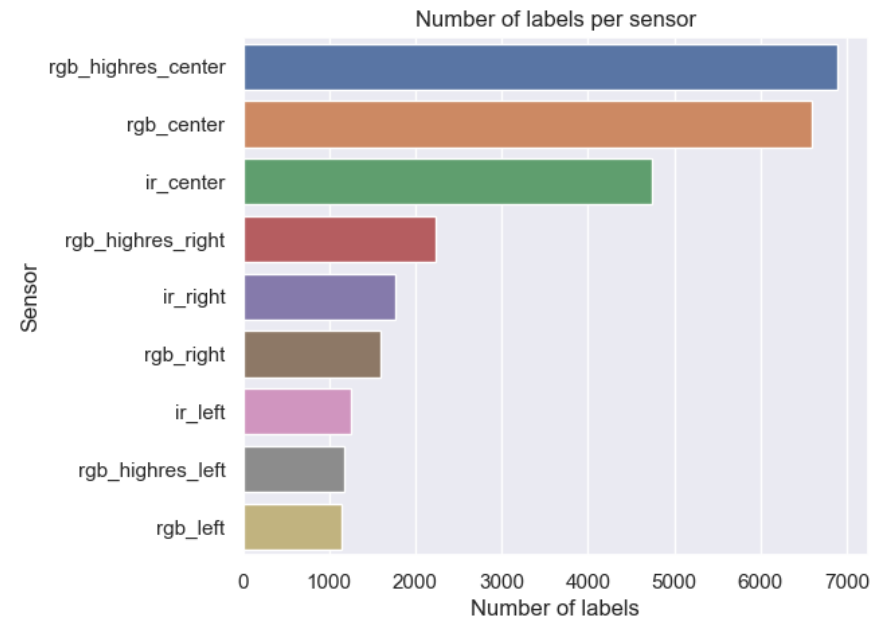
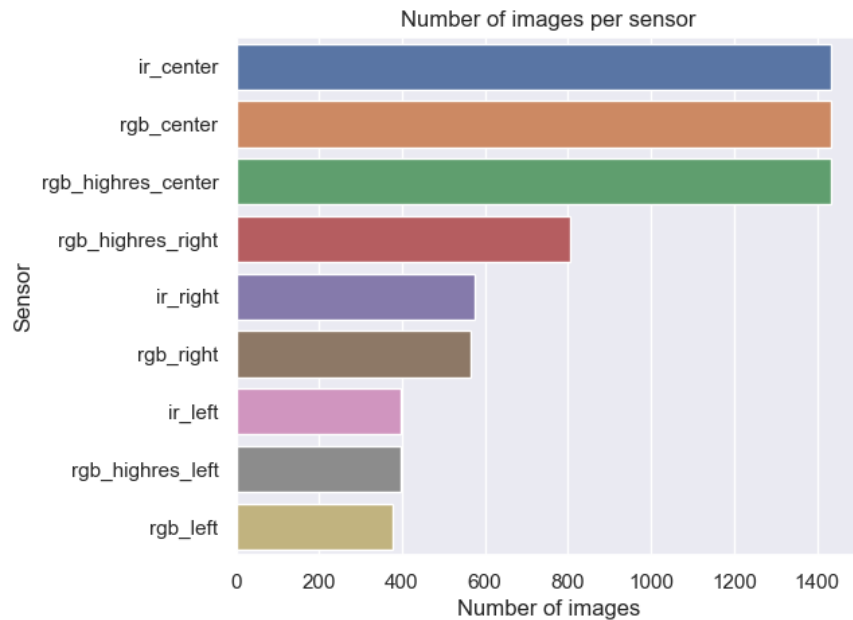
Data source

- Deutsch Bahn / Digitale Schiene Deutschland (<https://digitale-schiene-deutschland.de/en/news/OSDaR23-multi-sensor-data-set-for-machine-learning>).
- Images were generated between 09.09.2021 and 15.09.2021
- Total number of images: 13.952
- Sensors: **12MP RGB, 5MP RGB, IR cameras**, lidar, radar, GPS, inertia sensor



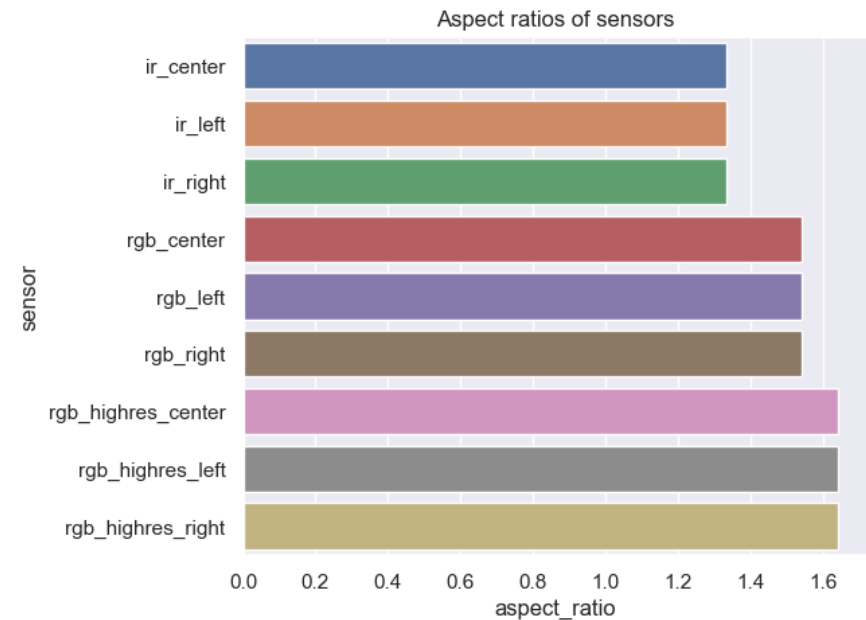
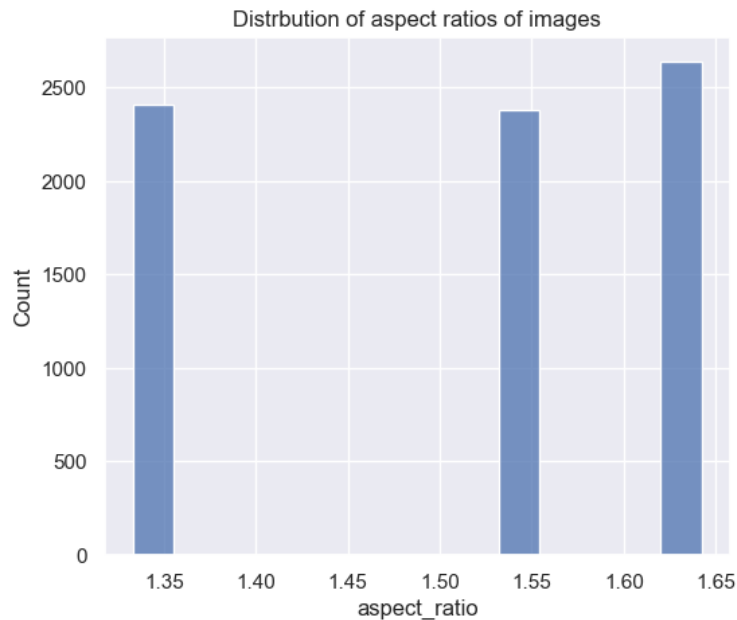
Relevant dataset (7.421 images, 27.386 labels)

- Most images are from forward facing cameras
- The number of labels in the image depends on sensor orientation and type



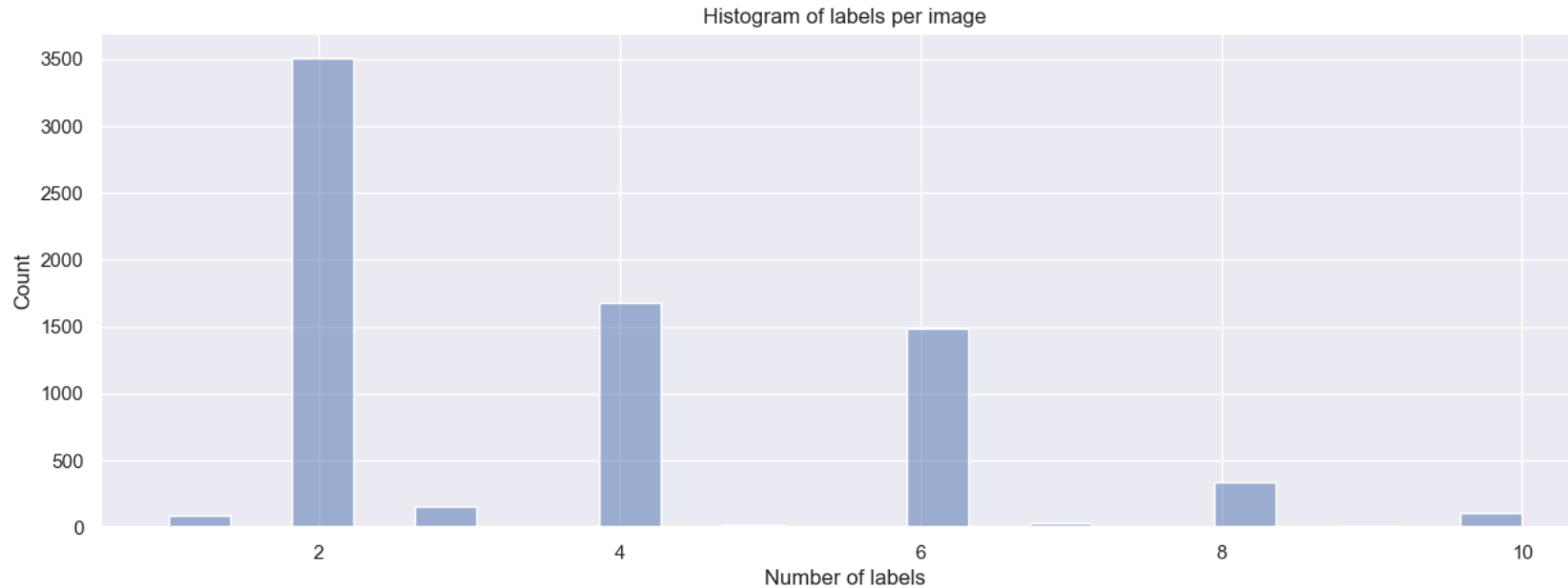
Aspect ratios (= width / height)

- All images are generated with one of three camera types
- Double check whether there are other images in the set



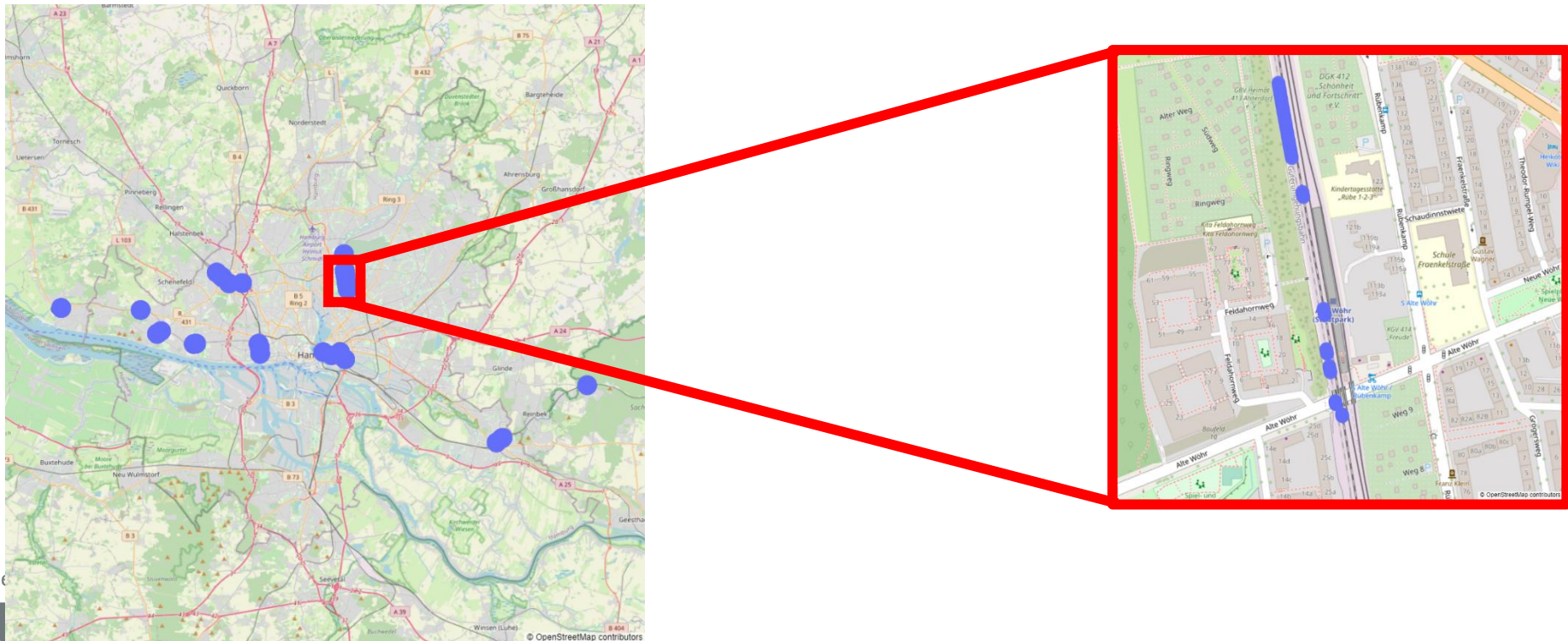
Labels per image

- Most tracks are labelled in pairs
- However, there is also a small number of images with uneven number of tracks



Location of images

- All images/videos were taken around Hamburg

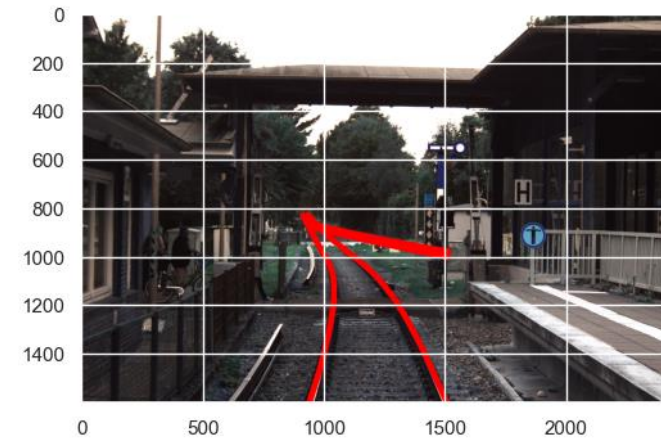
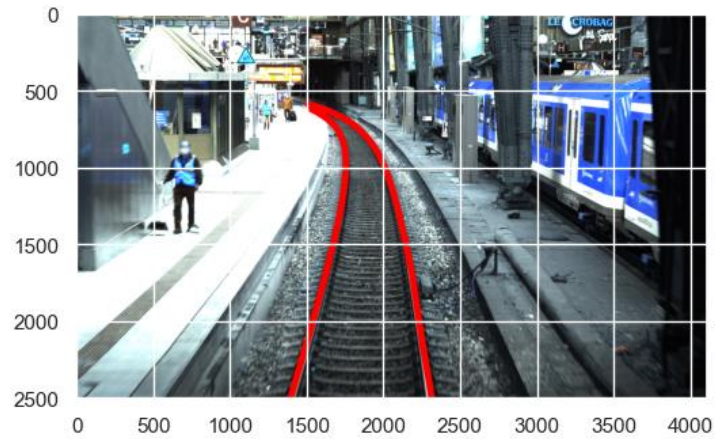
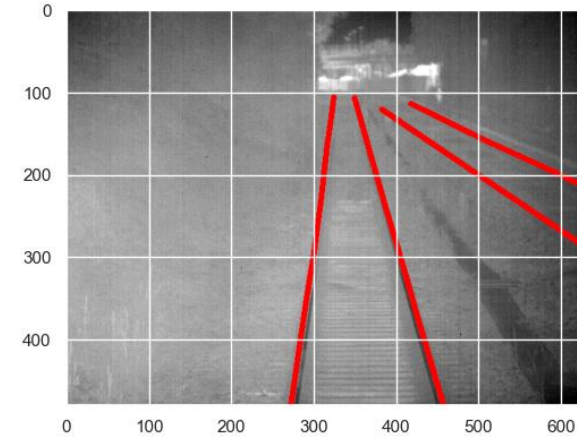
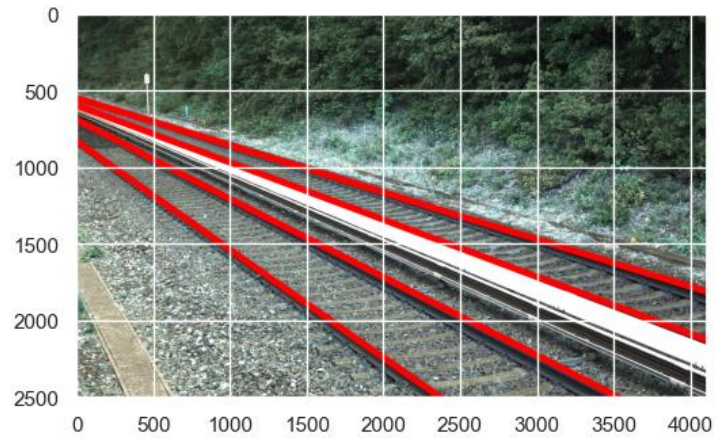


Time of images

- All images were taken between 8 AM and 5 PM

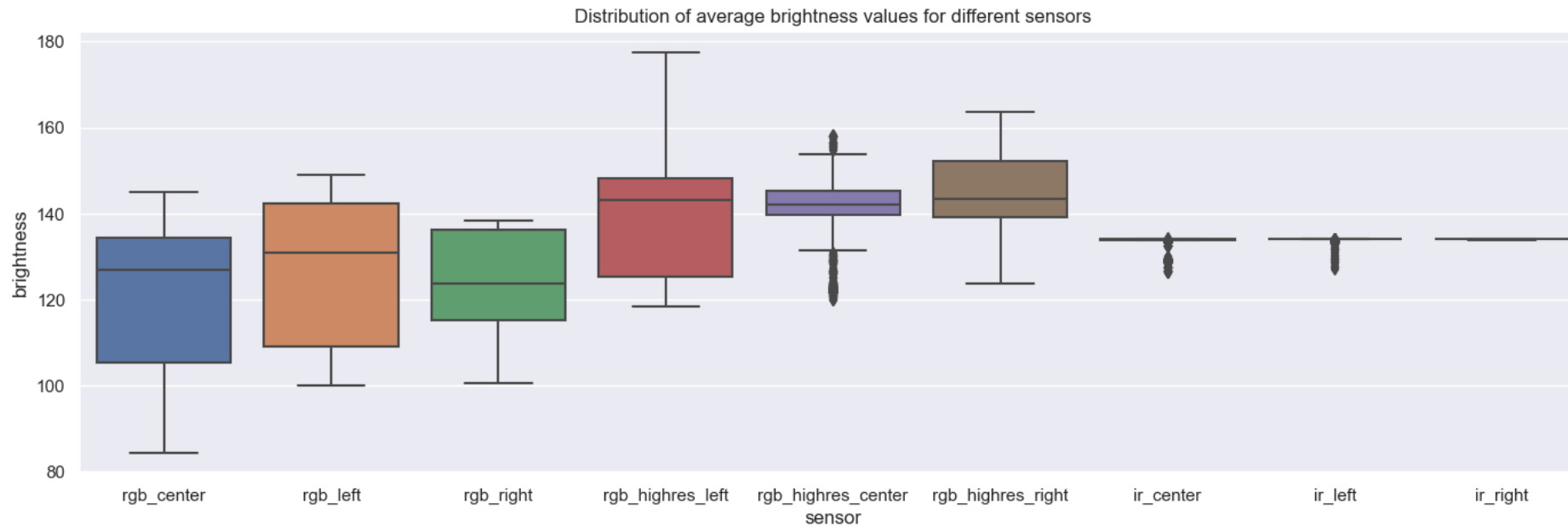


Some examples of labeled images



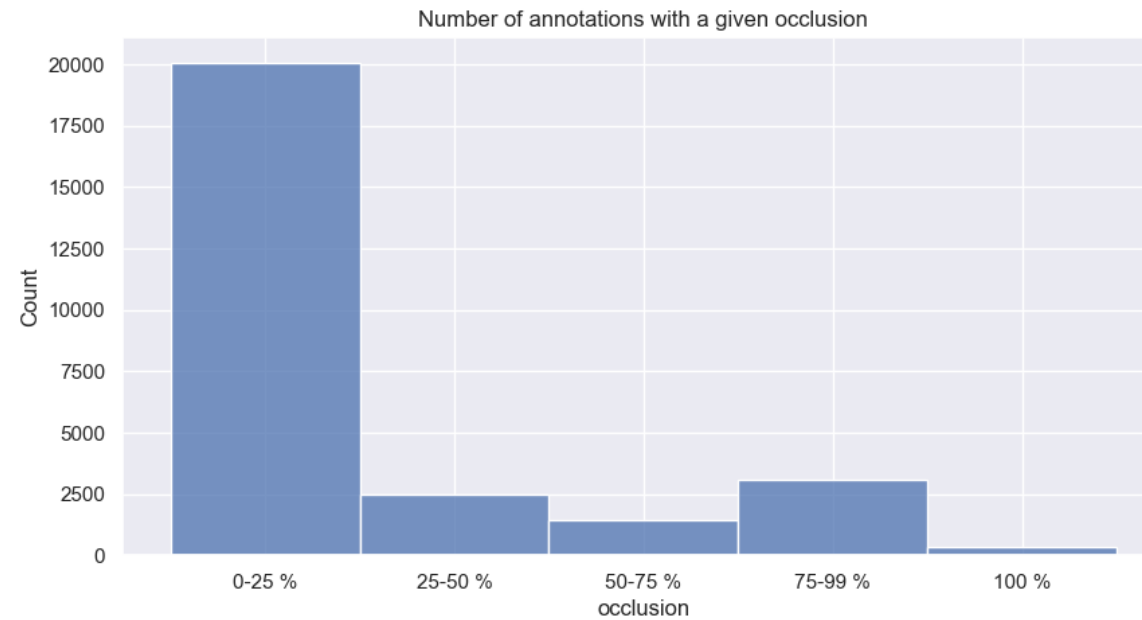
Brightness of images

- RGB images are darker than high resolution images (not pixel related)



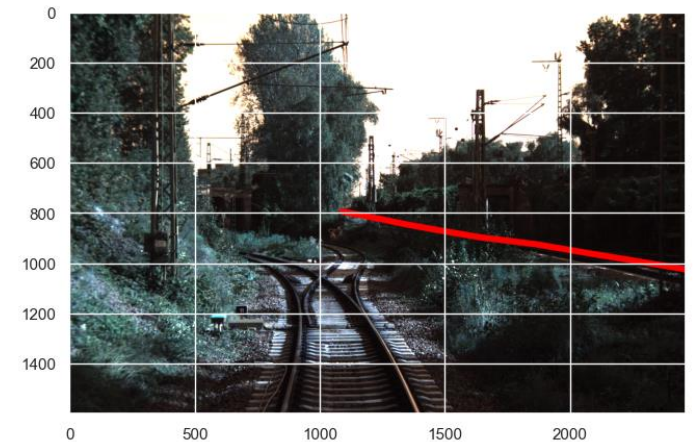
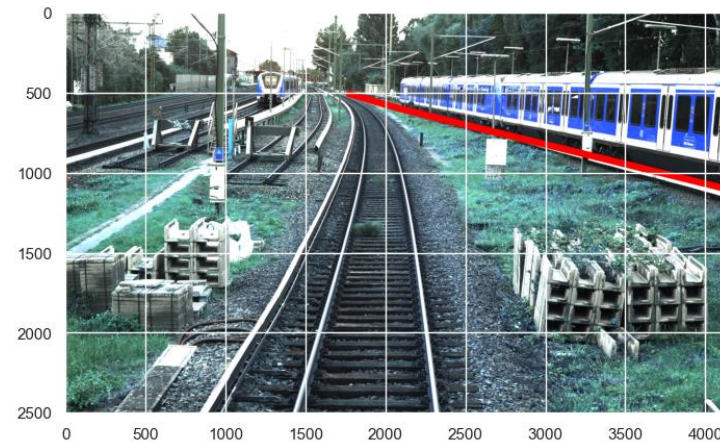
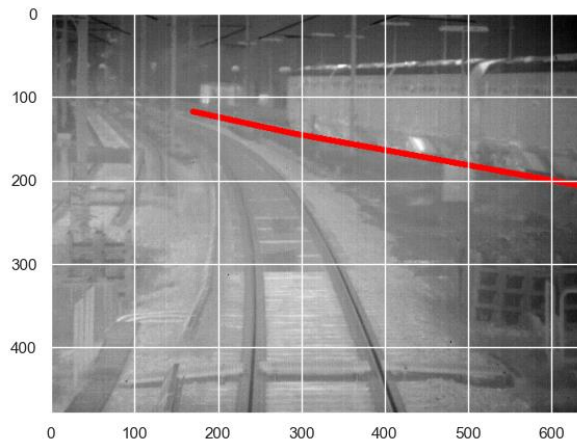
Occlusion

- Most of the labels have a good visibility



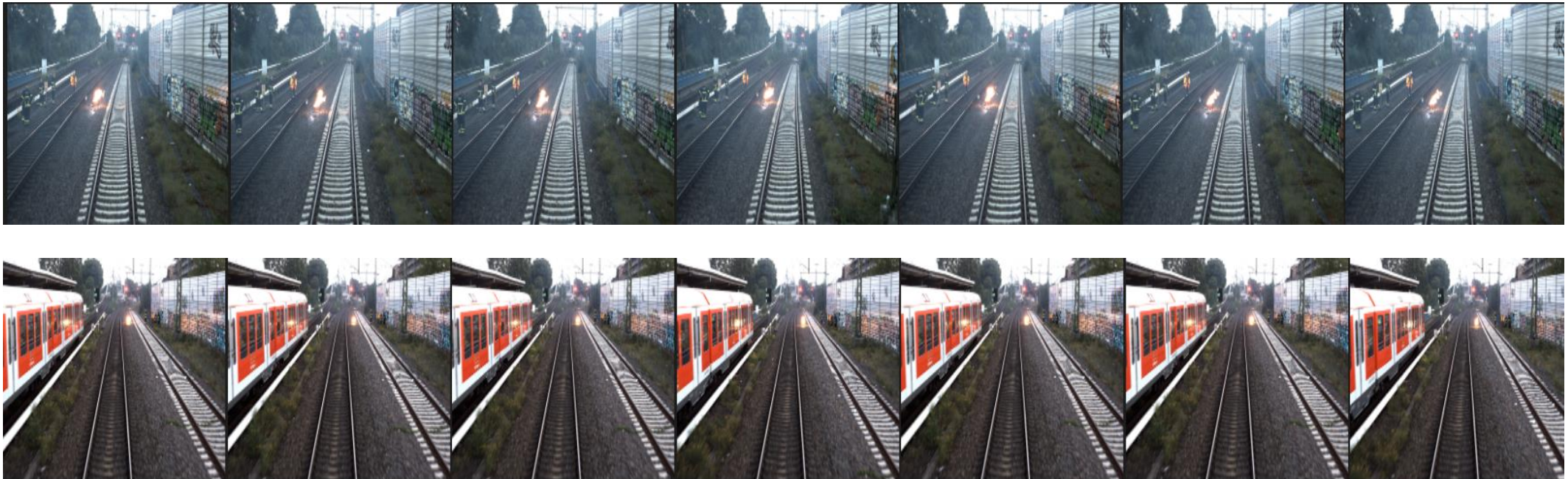
Examples of occlusion = 100%

- Most of the labels have a good visibility



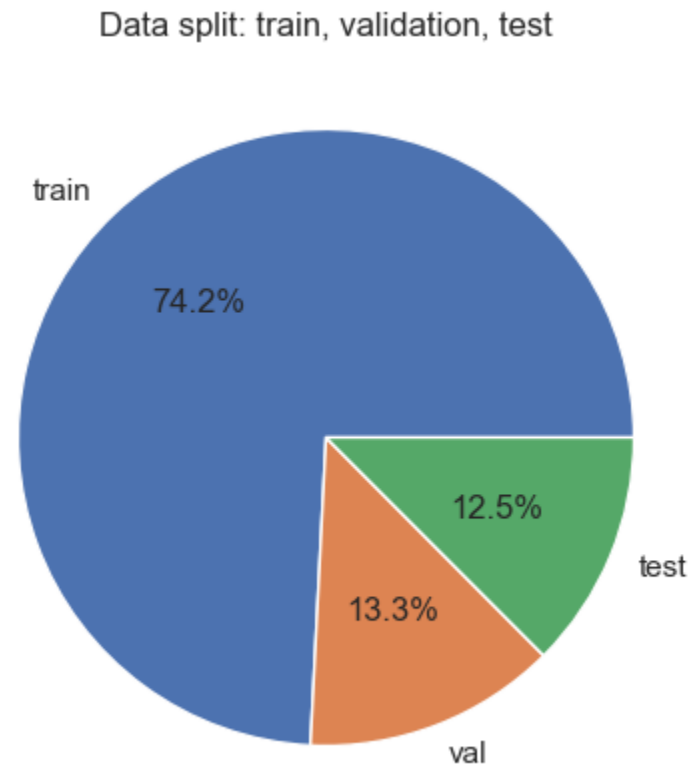
Splitting into train, validation and test set

- Many images are very similar which needs to be taken into account

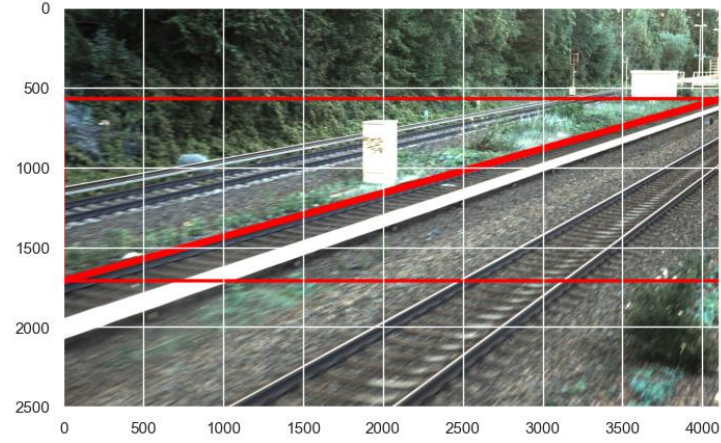
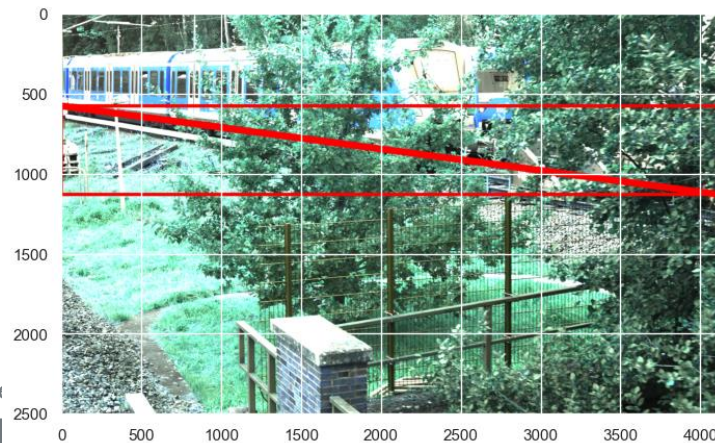
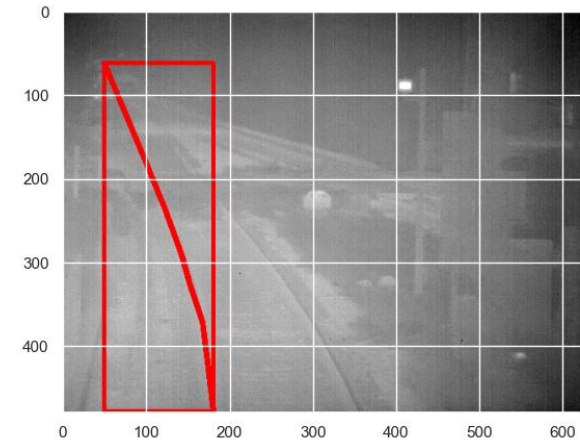
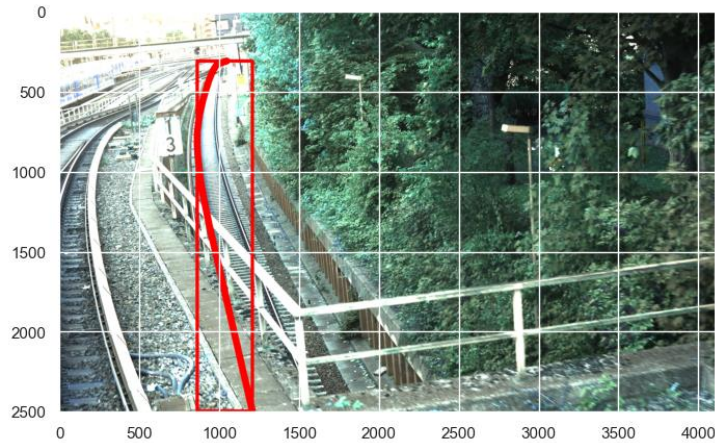


Splitting data set

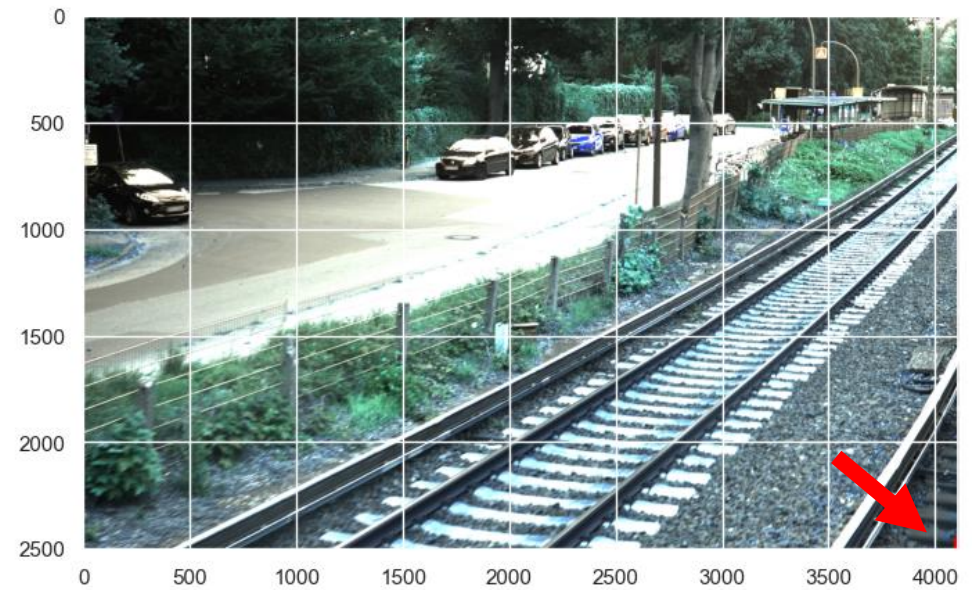
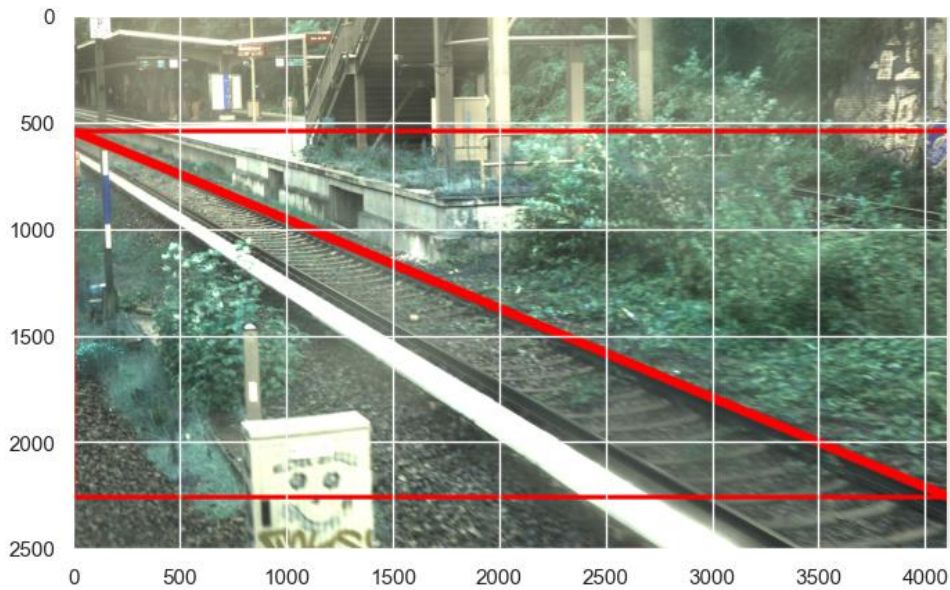
- Given the large similarity between images, we split data by randomly assigning videos to either train, validation or test set
- Advantage: fair testing as we have a „data leakage“ between the sets
- Disadvantage: validation batches might have low variety



Orientation of labels

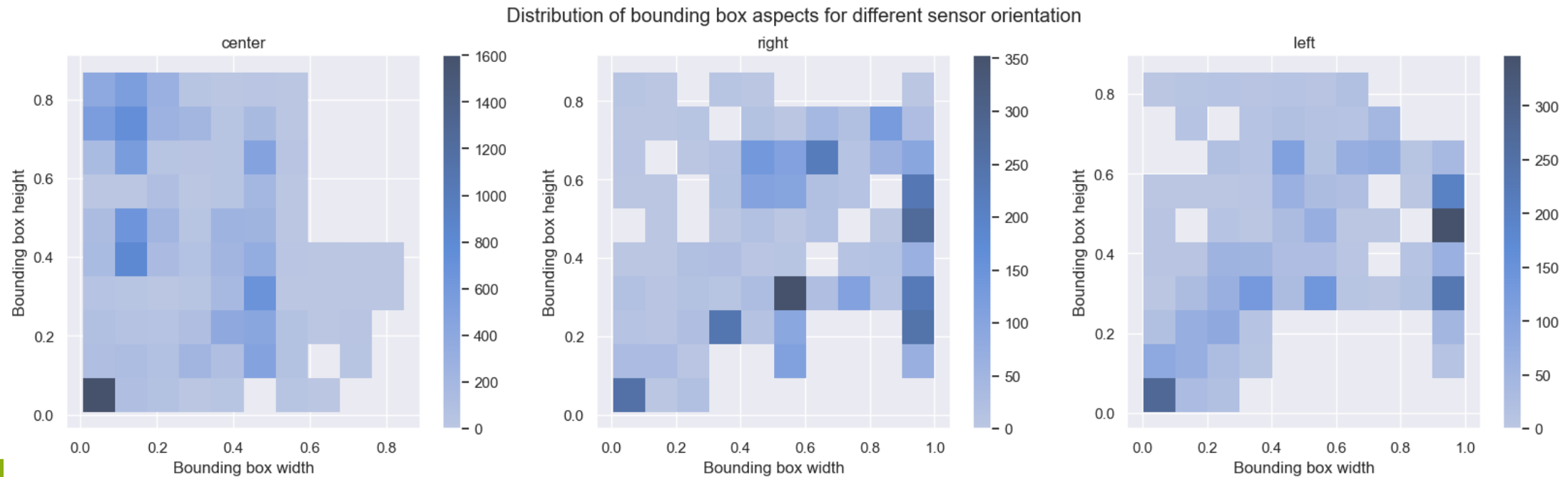


Size of labels

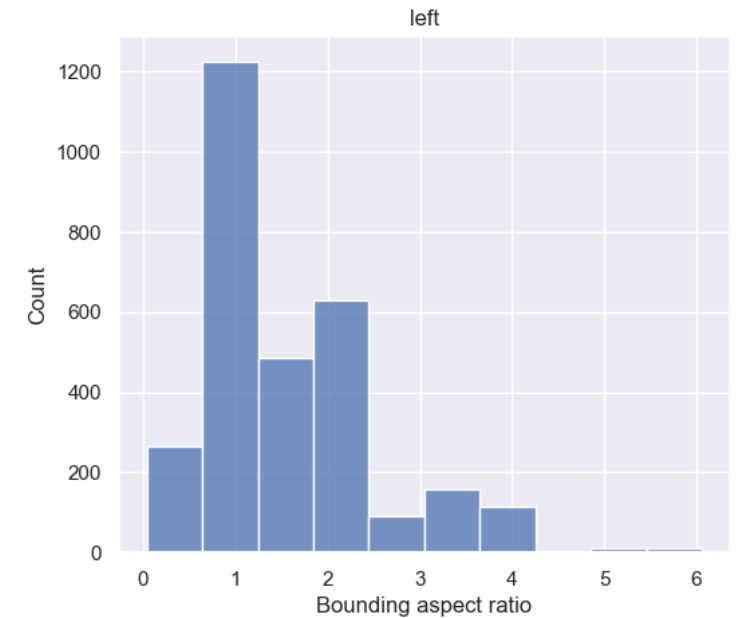
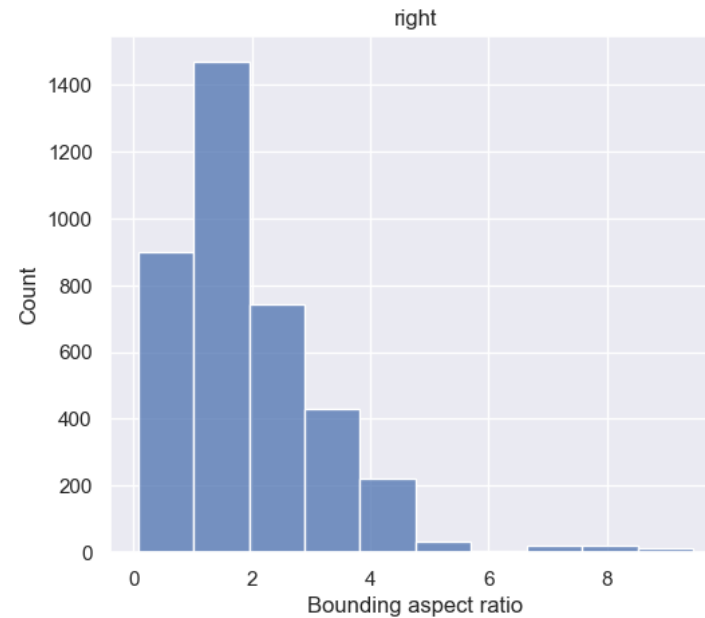
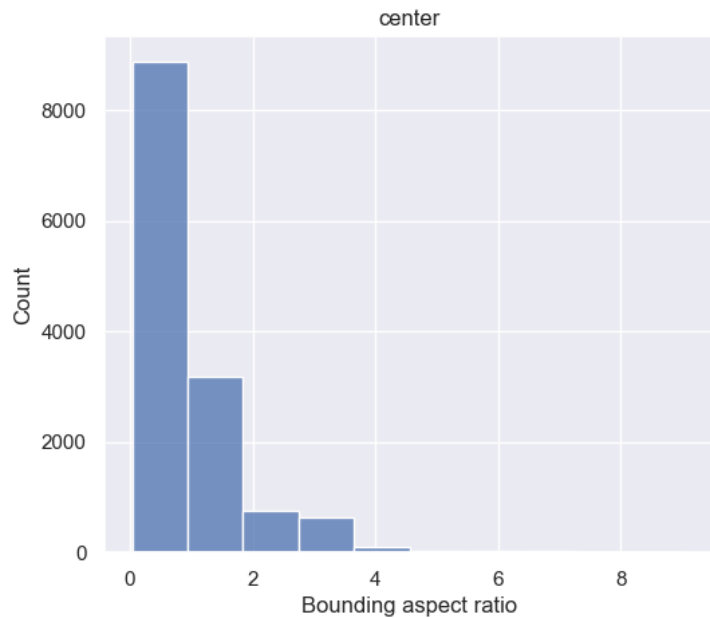


Histogram of bounding box shapes and sizes

- Bounding boxes format is related to the orientation of sensors



Aspect ratio of bounding boxes



Next steps

- Generate label masks for images
- Establish baseline
 - Research gradient thresholding approaches
 - Select and implement approach
 - Analyze results
- Select image segmentation model

Q&A