Bounding Box Dataset Augmentation for Long-range Object Distance Estimation

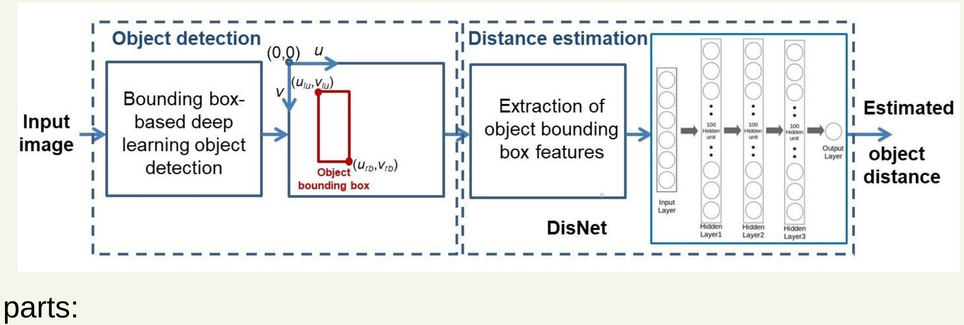
Marten Franke, Vaishnavi Gopinath, Chaitra Reddy, Danijela Ristić-Durrant, Kai Michels

Institute of Automation

Introduction

Autonomous long-range obstacle detection and distance estimation plays an important role in numerous applications such as railway applications when it comes to locomotive drivers support or developments towards driverless trains. To overcome the problem of small training datasets, this paper presents two data augmentation methods for training the Artificial Neural Network (ANN) named DisNet to perform reliable long-range distance estimation.

DisNet: object bounding box-based distance estimation

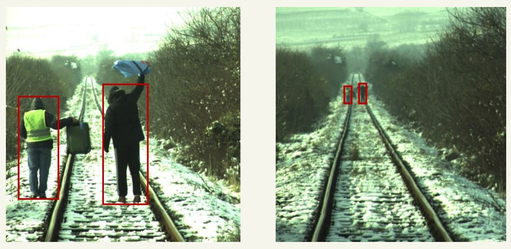


- Two parts:
  - Deep learning-based Object Detection (OD), and
  - ANN-based distance estimation named DisNet
- Different object detectors can be integrated into the system such as YOLOv3.
- Object Detection outputs are bounding boxes (BBs) of detected objects in the image.
- Based on the resulting object BBs size features, DisNet gives the estimated distance of the object to the camera as output.
- ***Initial DisNet*** – The initial training of DisNet was done by using the parameters of manually extracted BBs of 2000 objects (of the classes person and car), which were in the distance range 0 m - 60 m from the static camera (camera mounted on a test stand).

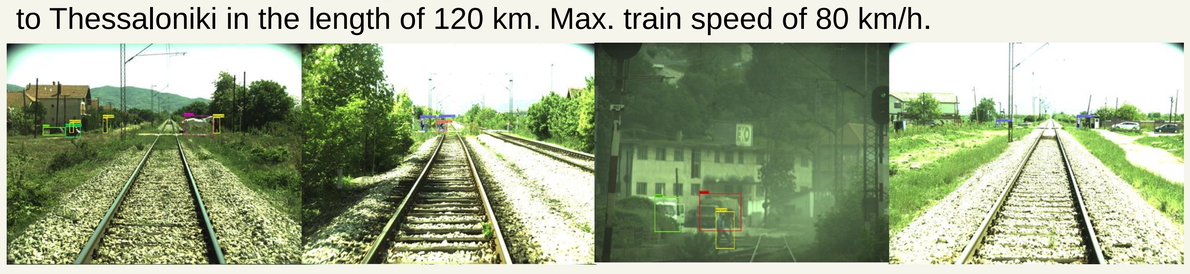
Custom long-range dataset generation

In order to re-train YOLO for the purpose of long-range OD, custom long-range railway dataset was generated.

- ***Static field tests*** – Image recording in the field tests on the Serbian railway location of the straight rail tracks in length of about 1100 m in different times of the day and night and in different weather conditions.
- Two persons imitated potential obstacles on the rail tracks while walking along the rail tracks 1000 m away from the cameras and 1000 m back, towards the cameras. At every 5 m they signalized in a particular way, so that frames recorded at moments of the signalization could be used for the dataset generation.

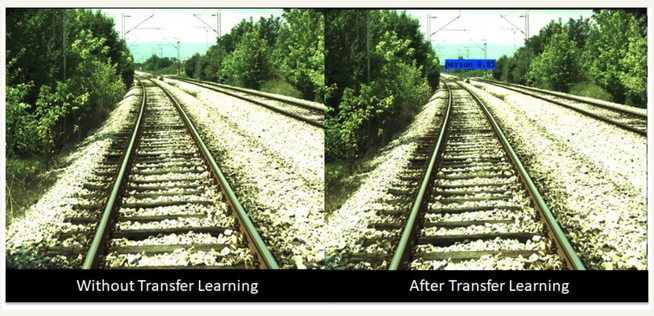


- ***Dynamic field tests*** – Cameras integrated into sensors' housing mounted on the frontal profile below the headlights of the moving locomotive Serbia Cargo type 444. The freight train with 21 wagons on the Serbian part of the Pan European corridor X to Thessaloniki in the length of 120 km. Max. train speed of 80 km/h.



Transfer Learning of object detector

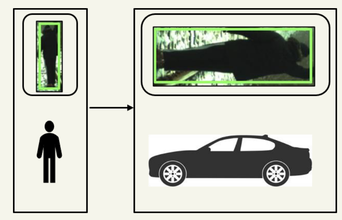
The initial YOLO model trained with COCO dataset for the purpose of general OD was re-trained with custom long-range dataset to detect objects in railway environments with long distance range of up to 1000 m.



Data augmentation for re-training of DisNet

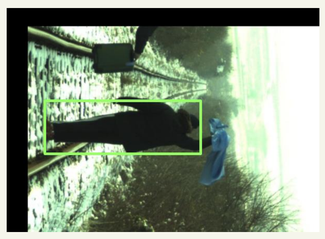
Re-training of *Initial DisNet* using an augmented BB long-range dataset to improve the reliability of BB-based object distance estimation.

Image transformation-based BB data augmentation – to obtain the BBs of objects of one class (so-called transformed object class) by transforming the BBs of objects of another class for which the BBs sizes and corresponding ground truth distance were a priory known (so-called reference object class).

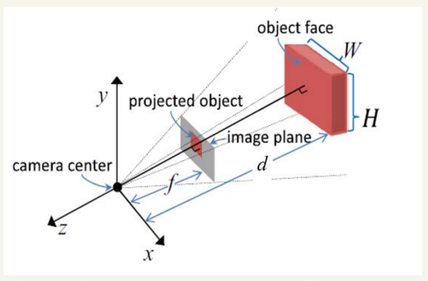


Reference object class person

Rotation
Scaling
Translation



Transformed object class car



d - object distance, *f* - camera focal length, *W&H* - real-world dimensions (width&height) of the object

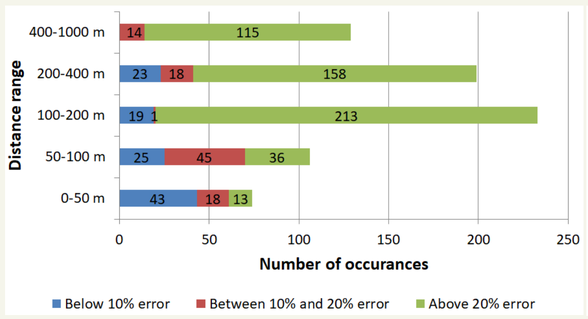
Projective transformation-based BB data augmentation – use of projective geometry to augment BB dataset so to generate synthetic object BBs corresponding to different distances

$$BB\ height: w = f\left(\frac{H}{d}\right), \quad BB\ width: w = f\left(\frac{W}{d}\right)$$

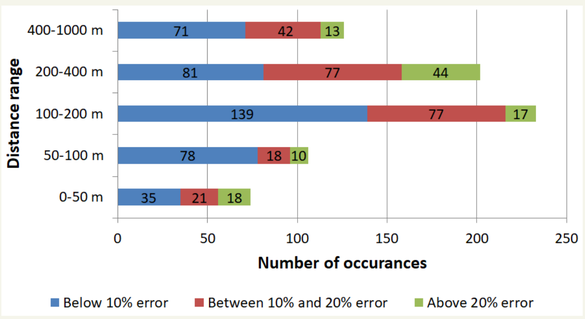
Finally, an augmented BB dataset of about 10000 BBs with corresponding distances from the range 0 m - 1000 m was obtained. This augmented dataset was used for re-training of Initial DisNet. Resulting retrained DisNet is referred to as *DisNet* in the following evaluation section.

Evaluation

The evaluation of DisNet re-trained with the augmented BB dataset was done on testing images recorded in SMART dynamic field tests. In total, 741 BBs extracted by the re-trained YOLO object detector were used for the evaluation of DisNet. Out of these 741 BBs, 654 were BBs of objects from class *person* and 87 were BBs of objects from the class *car*. For all 741 BBs the corresponding ground truth distances D_{GT} were known as calculated in dynamic field tests using the GPS coordinates of the moving train and Google maps GPS coordinates of the objects (obstacles) locations.



Initial DisNet distance estimations



DisNet distance estimations

$$RMSE = \sqrt{\frac{\sum(D_{est} - D_{GT})^2}{N}}$$

RMSE	Car	Person	Total
Initial DisNet	38.20%	38.21%	38.20%
DisNet	13.72%	10.52%	10.90%

$D_{GT=524\ m}$

This research received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation program under Grant No. 881784

Universität Bremen