**Public transport access detection**

**1. You Only Look Once: Unified, Real-Time Object Detection**

**-** [**https://arxiv.org/abs/1506.02640**](https://arxiv.org/abs/1506.02640)

**-** This article has a total of 39 references, listed at the end of the article and numerated.

- First the author/authors are mentioned, then the title and optionally the conference or journal and the pages of the reference inside the article or book. At the end the year is mentioned.

- Each article is referenced throughout the article by the number assigned at the end of the article in the following way “[number]”

- This article is relevant to the proposed theme because it presents the YOLO detection system, which can perform object detection (object localization and classification) in real time.

**Abstract**

- short description of the system and how it compares to other methods, together with performance metrics

**1 Introduction**

- illustrate the downsides of other existing detection systems and presents the advantages that make YOLO run faster.

- performance metrics are presented.

- disadvantages of YOLO are also presented.

**2 Unified Detection**

- provides explanations of the output of the network and how the input (image) is processed.

**2.1 Network Design**

**-** the structure and type of the network (inspired by the GoogleLeNet) is presented.

- a fast version is presented

- a figure explaining the architecture is present

**2.2 Training**

- process of training is described, which starts with pretraining

**2.3 Inference**

- describes one pass through the network

**2.4 Limitations of YOLO**

- explains on how small and nearby objects can be missed

- another limitation is that bounding box error is treated the same in small boxes and large boxes

**3 Comparison to Other Detection Systems**

**-** comparison to: Deformable parts models, R-CNN, Fast and Faster R-CNN, Deep MultiBox, OverFeat, MultiGrasp

**4 Experiments**

- different experiments are conducted on Pascal VOC 2007 dataset

**4.1 Comparison to Other Real-Time Systems**

- Deformable parts models, R-CNN, Fast and Faster R-CNN and different versions of YOLO are compared by mean average precision and frames per second

**4.2 VOC 2007 Error Analysis**

**-** deeper analysis of the errors of YOLO and R-CNN

**4.3 Combining Fast R-CNN and YOLO**

- combination of R-CNN and YOLO is analyzed

**4.4 VOC 2012 Results**

- performance on VOC 2012 dataset is presented

**4.5 Generalizability: Person Detection in Artwork**

- YOLO is compared to other systems on Picasso Dataset and the People-Art Dataset (person detection in artwork)

- YOLO has a better performance on these datasets

**5 Real-Time Detection In The Wild**

**-** it is shown that YOLO performs well also on a live camera

**6 Conclusion**

**-** short recap of YOLO advantages

**2. SSD: Single Shot MultiBox Detector**

- <https://arxiv.org/abs/1512.0232>

**-** This article has a total of 26 references, listed at the end of the article and numerated.

- First the author/authors are mentioned, then the title and optionally the conference or journal and the pages of the reference inside the article or book. At the end the year is mentioned and the page in the article where the cited article is referenced.

- Each article is referenced throughout the article by the number assigned at the end of the article in the following way “[number]”

- This article is relevant to the proposed theme because it presents the SSD detection system, which can perform object detection (object localization and classification) in real time. Also, it as an alternative to YOLO.

**Abstract**

**-** description of how SSD eliminates object proposal generation, and all computation is encapsulated in a single pass through the network

- performance metrics are also mentioned

**1 Introduction**

- short description of the state of the art and why it is not suited for real time object detection

- speed/accuracy tradeoff

- description of how the system is better than YOLO

- summary of contributions

**2 The Single Shot Detector (SSD)**

**2.1 Model**

- description of SSD framework for detection

**2.2 Training**

- training process is described: Matching strategy, Training objective, Choosing scales and aspect rations for default boxes, Hard negative mining, Data augmentation

**3 Experimental Results**

- the base network is VGG16 and additional parameter details are provided

**3.1 PASCAL VOC2007**

- on this dataset, SSD is compared to Fast and Faster R-CNN

**3.2 Model analysis**

- experiments to assess the impact on performance

- data augmentation is essential, more default box shapes are better, multiple output layers at different resolutions is better

**3.3 PASCAL VOC2012**

- similar tests are conducted for PASCAL VOC2012

**3.4 COCO**

- similar tests are conducted on COCO

**3.5 Preliminary ILSVRC results**

**-** similar tests are conducted on ILSVRC2014 DET

**3.6 Data Augmentation for small Object Accuracy**

- SSD struggles with small objects, therefore data augmentation improves performance

**3.7 Inference Time**

- time statistics are shown for one pass through the network

**4 Related Work**

- evolution of object detection

**5 Conclusions**

- recap of key features of SSD

**6 Acknowledgment**

- details on how to project started