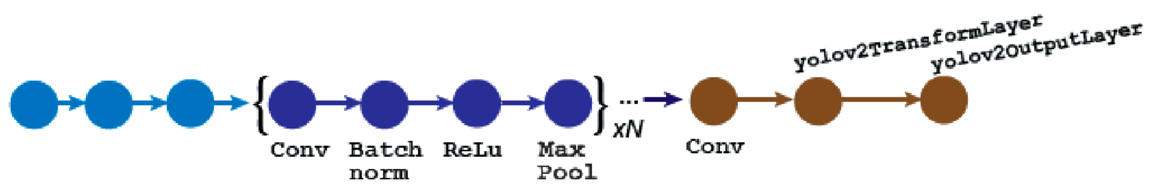
DTU Compute
Department of Applied Mathematics and Computer Science

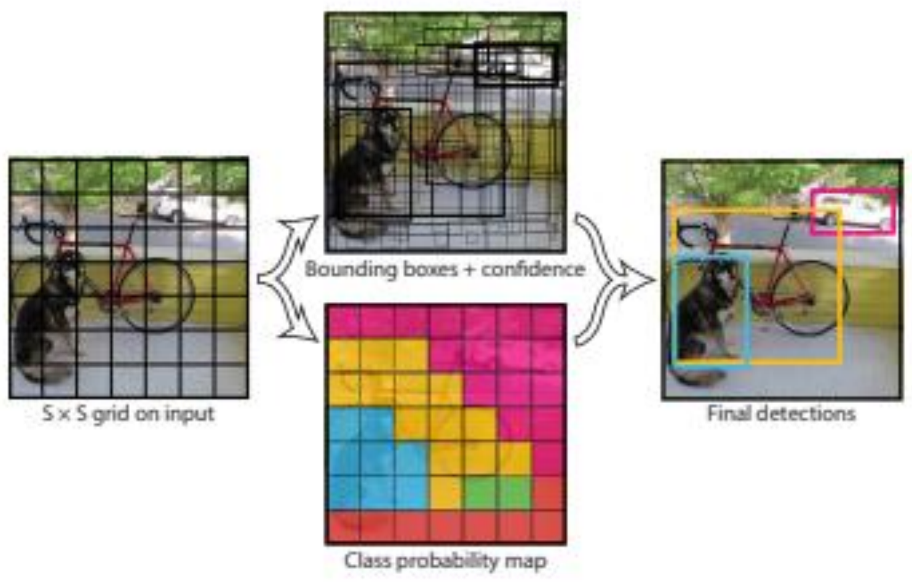
Real Time Object Detection Using YOLO

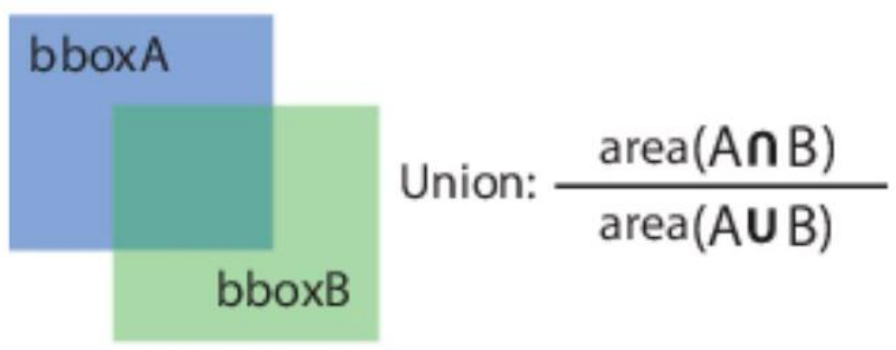
Andrew Blumensen s194139, Marcus Jepsen s195780 and Daniel Jalel s194291

1 Introduction

This project aims towards achieving real time CNN segmentation to identify tanks in drone footage. This was attempted with the use of the YOLO Object detection algorithm, which uses an individual neural network to identify and classify the content for every frame of the video. The network is trained on a collection of drone images of tanks and is later applied to real 30-fps recording.







A visual illustration representing the different steps the yolo algorithm uses to accomplish its conclusion. The algorithms architecture for all epochs. The segmentation and categorization of all segments in the frame. Ending with evaluation of intersections between prediction and truth

The test data is a 6-second video (180 images) that shows the movement of tanks in Ukraine. The network was initially trained with 25 epochs, however, the results got quite poor. An example of these results can be seen on the figure to the right, where only two out of four tanks were detected. Furthermore, the confidence of the model was quite low in general and in some frames no tanks could be detected at all. On the other hand. a trial model was made with 180 epochs which resulted in overfitting and only few tanks detected. For the final model, the number of epochs chosen was 80 and this lead to the best fit.

2 Yolo Algorithm

Older detection systems, like R-CNN, apply models to multiple locations and scales in an image which takes computation time and is ineffective in real time object detection. The YOLO algorithm is a newer approach to object detection which only uses a single neural network for the whole image. It is built to solve object detection as a regression problem by spatially separating bounding boxes and class probabilities. This allows the Yolo algorithm to process between 45 to 150 frames per second which is 100 times faster than Fast R-CNN.

3 Training and testing

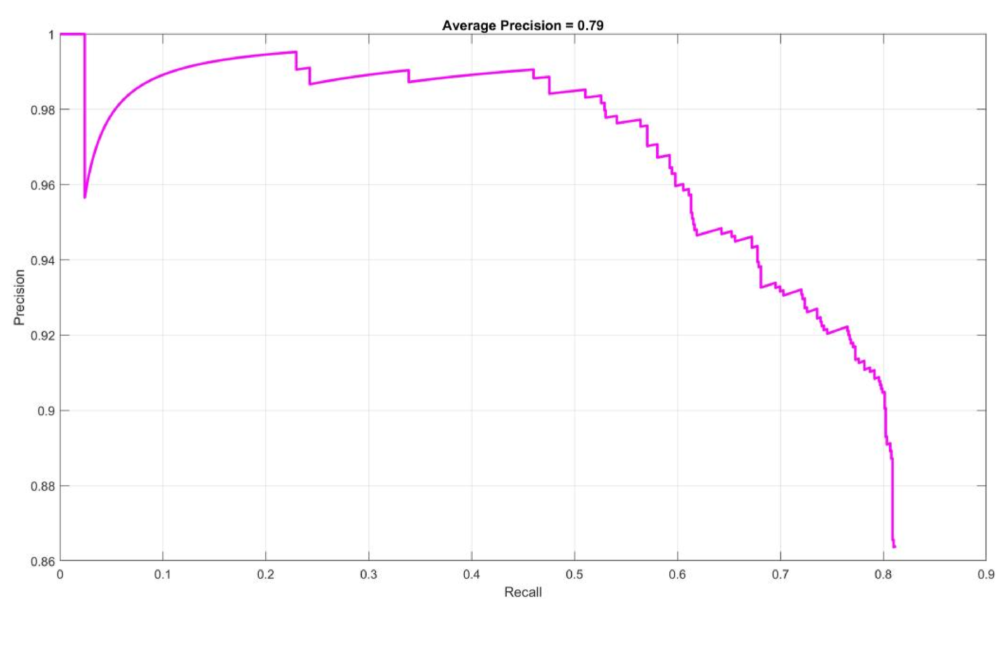
The training data consists of 81 aerial images of tanks which was given ground truth bounding boxes by hand. The mean average precision of the model was calculated to measure the performance of the network. For this project, a predicted box was a true positive if the intersection of the predicted box and the ground truth box was at least 0.4.

This image shows the results of the YOLO algorithm with 25 epochs used on real drone footage from Ukraine. The figure shows false negatives from the model, as two tanks are not labeled.

This image shows the final results of the YOLO algorithm with 80 epochs on the same drone footage. The false negatives from the model has now been corrected, identified and labeled.

4 Performance and conclusion

The (recall, precision) graph gives an mAP of 0.79. The mAP is defined as the area under the curve, and the area is quite large in our case. This result is based on setting an IoU to 0.4. For this project, an IoU of 0.4, as well as an mAP of 0.79, is deemed successful since the main objective is to just identify the tanks within the frame. If one ought to make an object detection algorithm with the purpose of firing missiles accurately toward the tanks, then one would probably want a higher IoU and a much higher resulting mAP in order to accurately be able to hit the tanks. .



All in all, the YOLO algorithm is a fast alternative to R-CNN with a bit lower local accuracy but much less computation time which makes it optimal for use object detection in real time. It is, as many other CNN´s, depended on a lot of data quality which inherently is not easy to acquire and takes a great deal of time to label.