You Only Live to Look Once
Weapons Object Detection

Camille Porter, Sarah Alkhateeb

Gothenburg University
Computer Science and Engineering Department

CHALMERS UNIVERSITY OF TECHNOLOGY

Introduction

Automating weapon detection using computer vision

Figure 1: Photo of a real life attack.

- Can object detection help in making our lives more secure?
- Can we teach computers to tackle real-life challenges?
  - Sound an alert when a weapon is detected!
  - Help will come faster!

YOLO

YOLO used a single convolutional neural network (CNN) for the whole image. This lets it:

- Extract features
- Predict bounding boxes
- Do non-maximal suppression

All at the same time once for the whole image, hence the look once. It then breaks up the result of the ConvNet with a grid. [0]

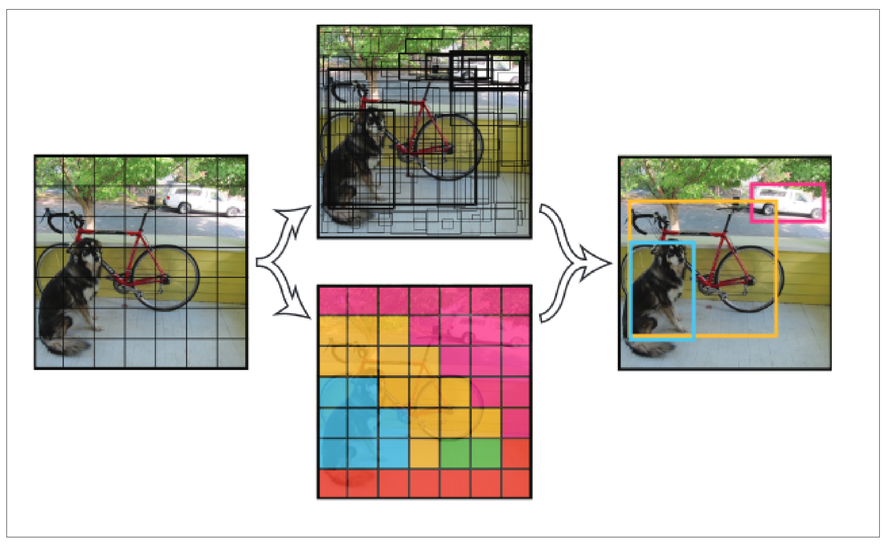

Figure 2: YOLO breaks the image into a grid.[0]

Each grid cell calculates $Y_c = [p_c b_x b_y b_h b_w c_1 c_2]$

- p_c is the probability that an object is in this cell (0 ,1)
- $b_x b_y$ are the coordinates of the center of the box
- $b_h b_w$ are the height and width of the box
- $c_1 c_2$... are the class probabilities

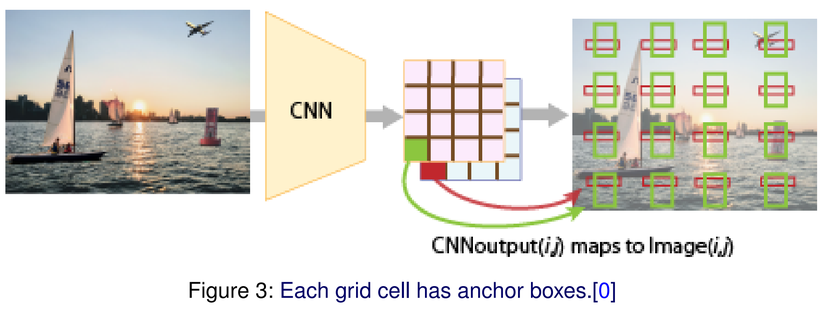

Figure 3: Each grid cell has anchor boxes.[0]

Anchor boxes allow each grid cell to find more than one object. They find objects with different aspect ratios.

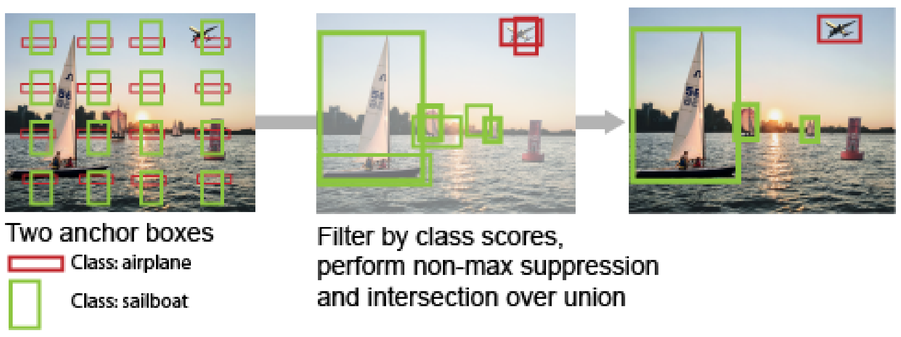

Figure 4: Some of the bounding boxes have non-optimal shape.[0]

It then uses non-maximal suppression (picks the box with the highest p_r).

Data Preprocessing

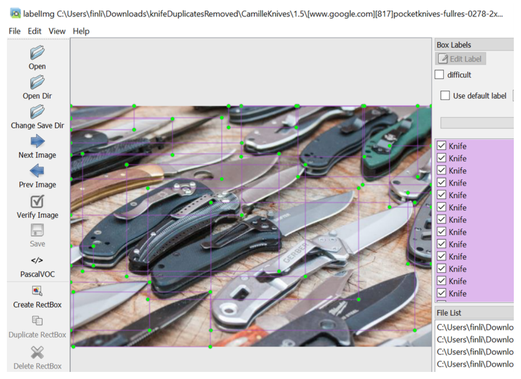

Figure 5: labelImg tool to label an image with lots of knives.

- Images of guns and knives downloaded from google
- Labeled using labelImg tool
- After labeling in YOLO format, we had to change the names and clean the data before YOLO would accept it

Results

First attempt to train YOLO didn't work well.

- 2000 images per category (guns and knives)
- Many of the images were too easy to categorize. They were centered with a white background.
- All objects were categorized as guns.

Figure 6: Knife mislabeled as gun.

Figure 7: Correctly labeled knife.

Second attempt was better.

- More images that were more similar to security videos (grainy, not white background)
- 3,500 images per category
- More objects were labeled correctly.

"YOLOv4 trains with a two-stage approach. First, it modifies the original image to create the deception that there is no object on the image. Second, the network is trained to detect this adversarial modified image in the standard way". [0]

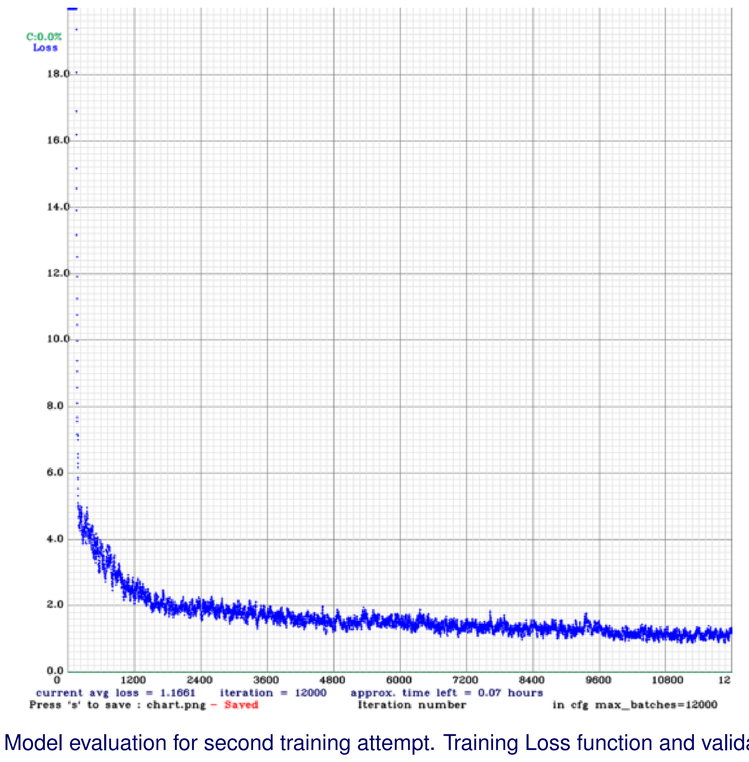

Figure 8: Model evaluation for second training attempt. Training Loss function and validation mAP

- Ran 12000 iterations but had overfitting after 4000 iterations.
- Achieved validation mAP of $\approx$ 83% and a training loss $\approx$ 1.1 at 4000th iteration.

References

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once:Unified, Real-Time Object Detection. https://pjreddie.com/media/files/papers/yolo.pdf

Anchor boxes for object detection. MathWorks. https://www.mathworks.com/help/vision/ug/anchor-boxes-for-object-detection.html

YOLOv4: https://mc.ai/reviewing-yolov4/: :text=YOLOv4

youremailhere@student.chalmers.se