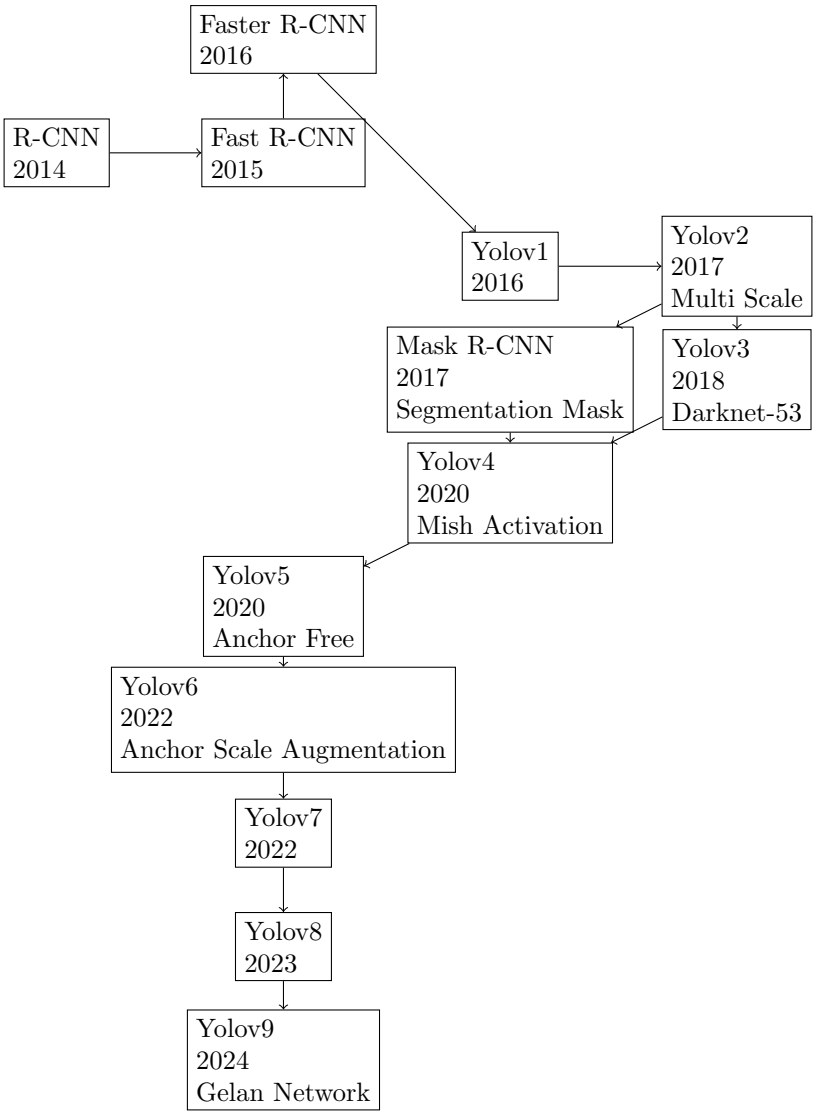
Learning Notes

Felix

June 14, 2024

Contents

1 Timeline

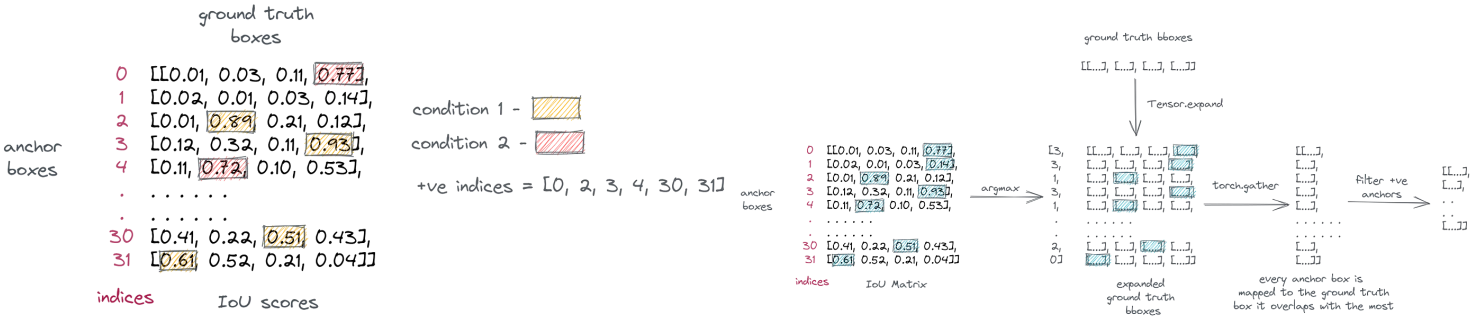


2 Faster Regional-CNN

2.1 Stage 1

CNN Backbone extracts a feature map. Each pixel of the feature map becomes an anchor point. There are 9 different anchor boxes (feature map* 9). Each **anchor box** is compared with the ground truth box. If the score > 0.7, it is a positive box. The highest score wins. To compare IoU is used. For each anchor box, we derive a corresponding ground truth box and filter them based on positive boxes (red). Since each ground truth box has a category, we can apply the same principle to obtain positive categories. Next, we calculate the offsets of the anchor boxes to the ground truth boxes (Fig 1).

Anchor Box + Offset = Region Proposal



(a) IoU

(b) Filter Positive Groundtruth

Figure 1: Stage 1

2.2 Stage 2

The region proposals are divided into a fixed number of regions and then max pooling is applied. This ensures that the regions have the same dimension (ROI). Cross Entropy Loss learns the categories. Regression learns the offsets (Fig 2).

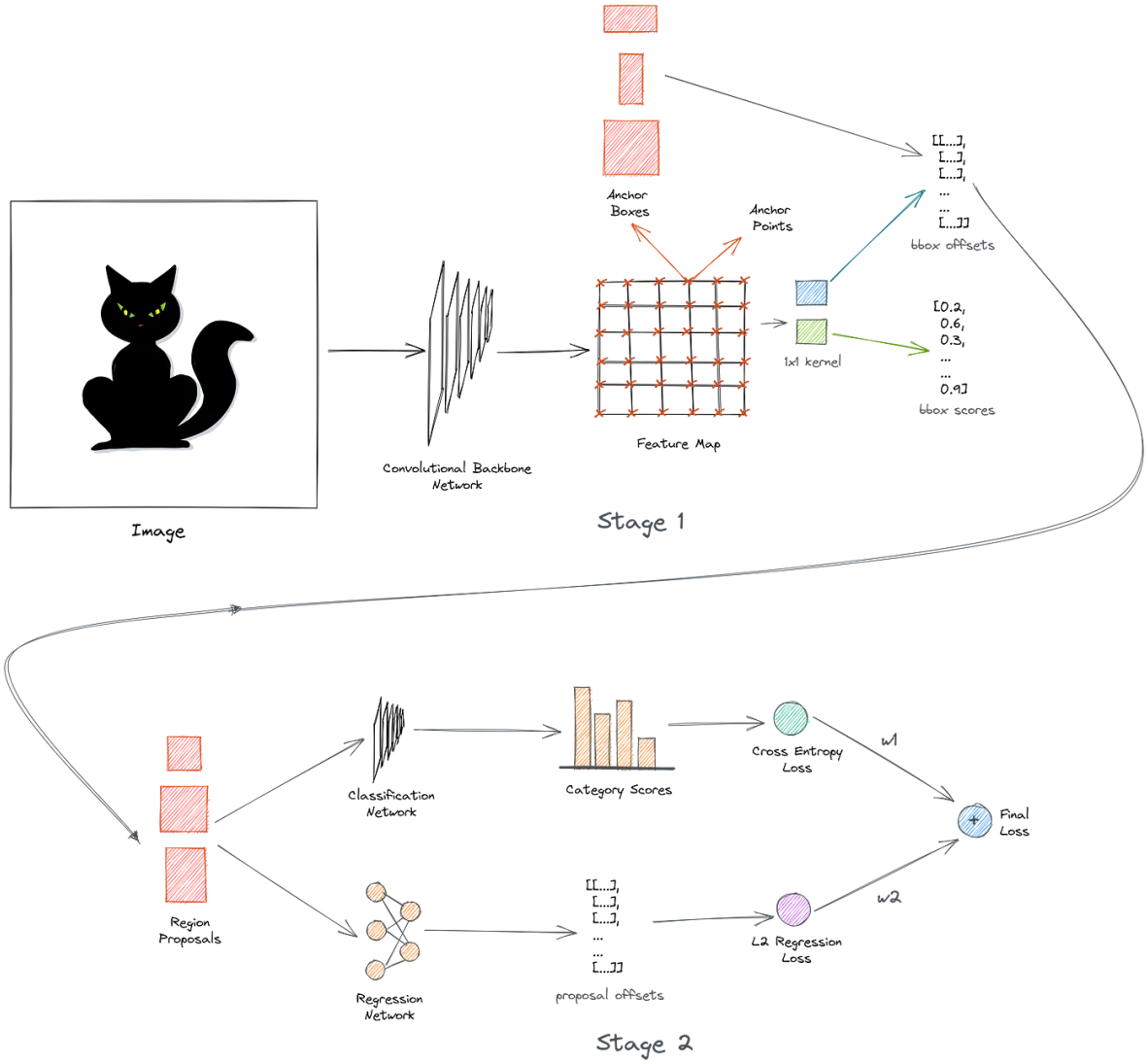


Figure 2: Faster RCNN

2.3 Sources

1. TowardsDataScience

3 Yolo You Only Look Once

Yolo1 splits the image in a 7x7 grid. Each cell predicts a bounding box.

In Yolo2 different scales to recognize smaller objects are used. Predicts anchor boxes instead of bounding boxes. The idea is, it is easier to predict the offset of predefined bounding boxes than to classify the bounding box itself. Uses Darknet-19 (Fig 4) and BatchNormalization.

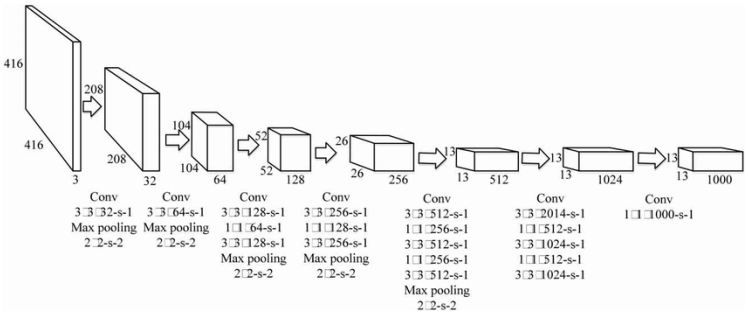


Figure 4: DarkNet

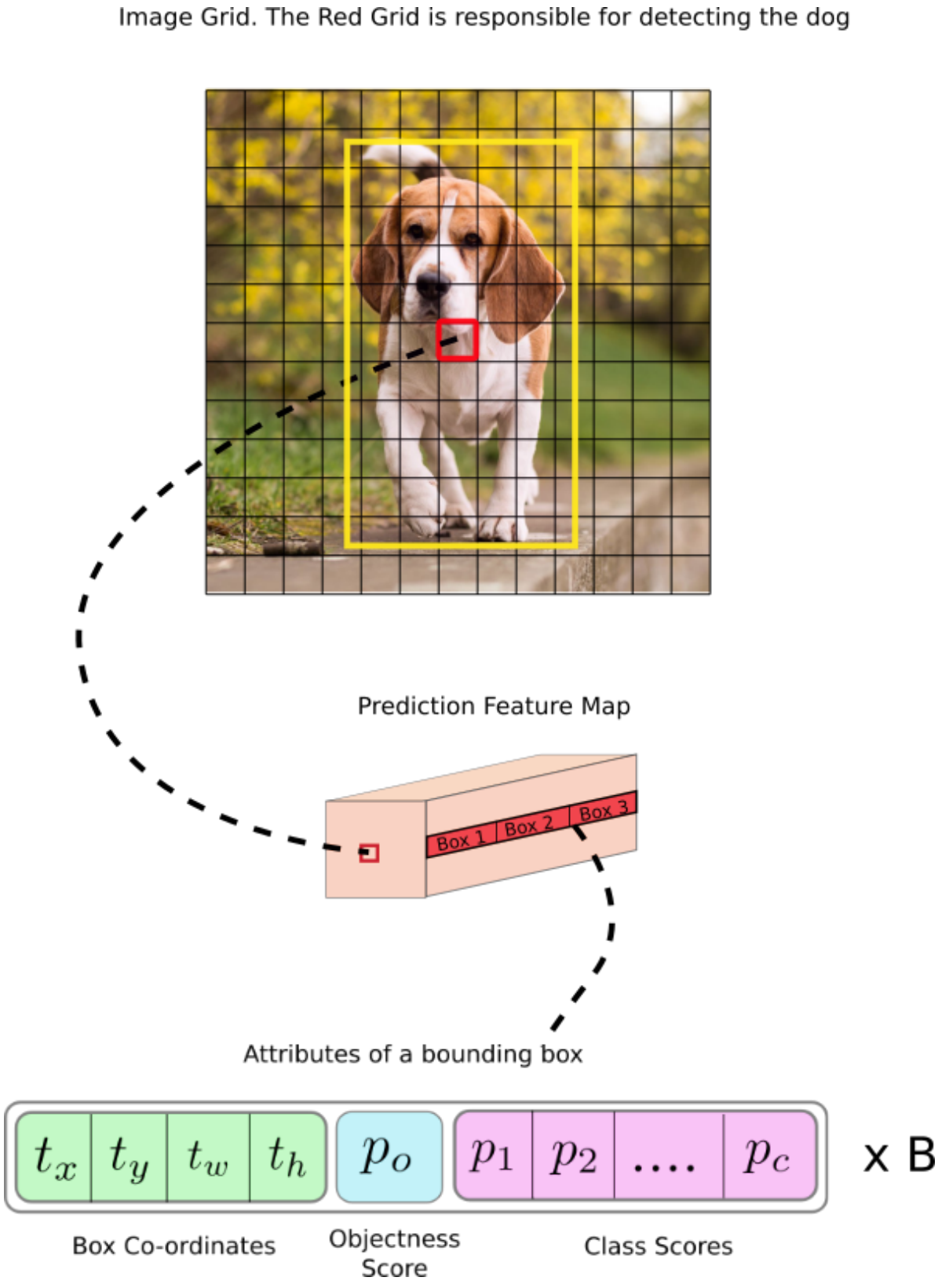


Figure 3: Image Grid

Yolo3 changes to Darknet-53 and uses three scales by using three different sizes of anchor boxes.

Yolo4 has been developed by Chien-Yao Wang, developer of CSPBlock. It uses the mish action function and calculates the CIOU, complete intersection over union. It uses CSPDarknet53 (Head) + Spatial Pyramid Pooling + PANet(Neck)

Yolo5 is anchorfree.

Yolo7 has been developed again by Chien-Yao Wang.

Yolo8 uses self attention.

Yolo9 is based on Yolo7. It introduces Programmable Gradient Information and improves the backbone using Generalized Efficient Layer Aggregation Network (GELAN). YoloE > YoloC > (YoloM > YoloS It can perform classification, instance segmentation and panoptic segmentation. It has three branches: main Branch, auxiliary branch and multi-evel auxiliary branch which combines gradients from different prediction heads to consider all object sizes Non-maximum-suppression is used to remove duplicate predictions.

3.1 Sources

1. How to implemenet Yolo
2. Yolo Object Detection
3. CASPNet-Backbone