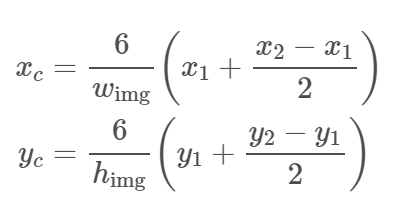
As we know, YOLOv1 takes as input 448×448 images and outputs a S×S×(B⋅5+C)S×S×(B⋅5+C) tensor where S is the grid size, B are the boxes inside each cell of the grid and C is the number of classes.

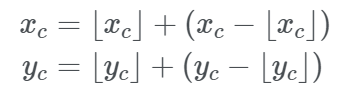
We ad simplified this a bit and keep only the S×S×5 grid. This means our object detector can find one bounding box per grid cell and cannot distinguish between classes.

Last layer of Mobilenetv2 architecture with size 224×224 have the value of S=7

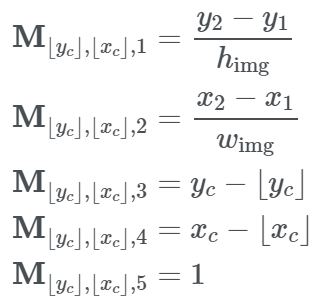
Let xc,yc be the center of the object and wimg,himg size of the unscaled image. Then



Note that we can write xc,yc∈Q as follows:



Let M be our 7×7×5tensor. Then



We put one element at grid position (⌊yc⌋,⌊xc⌋,k) for 1≤k≤5, while all other grid cells remain empty. All entries of MM are between 0 and 1, which makes the use of the sigmoid function possible.

Our model will be much faster than YOLO and only require 500K parameters.

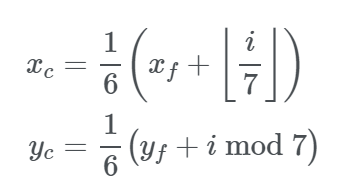
Since block\_16\_project\_BN is a 7×7×112 feature map, the next convolutions should also have 112 filters

After the last convolutional layer, we will add another two convolutional blocks, followed by a convolution with the number of outputs.

The loss function consists of the loss for the coordinates and the loss for the object. Hence, we will use a mix of squared error and binary cross entropy. The coordinate and size loss will only be calculated based on the box with the highest probability.

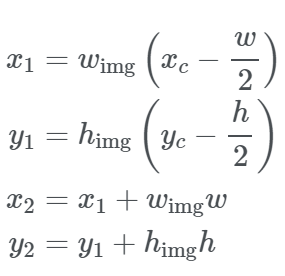
For prediction - first, we flatten the 7×7×5 prediction tensor in order to get five 49×1 vectors xf,yf,w,h,p.

Since xf,yf are offsets, we still have to add the grid position. So we just create a list of numbers from 0 to 48 and calculate for all i∈[0,48]



Next, put all vectors in a 49×5 matrix and select only the boxes with p≥0.5. Furthermore, apply non-maximum suppression.

Finally, the boxes can be displayed on the image with coordinates:



After first training we were left with an accuracy of 85%. Trained model was tuned again to get the accuracy of 87%