





1.We reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using our system, you only look once (YOLO) at an image to predict what objects are present and where they are.

把目标检测问题，作为一个回归问题，在这个问题里，找到包围盒，同时给包围盒里的物体正确标定类别。

2.YOLO makes less than half the number of background errors compared to Fast R-CNN.

YOLO算法对于背景的误判比Fast R-CNN少。

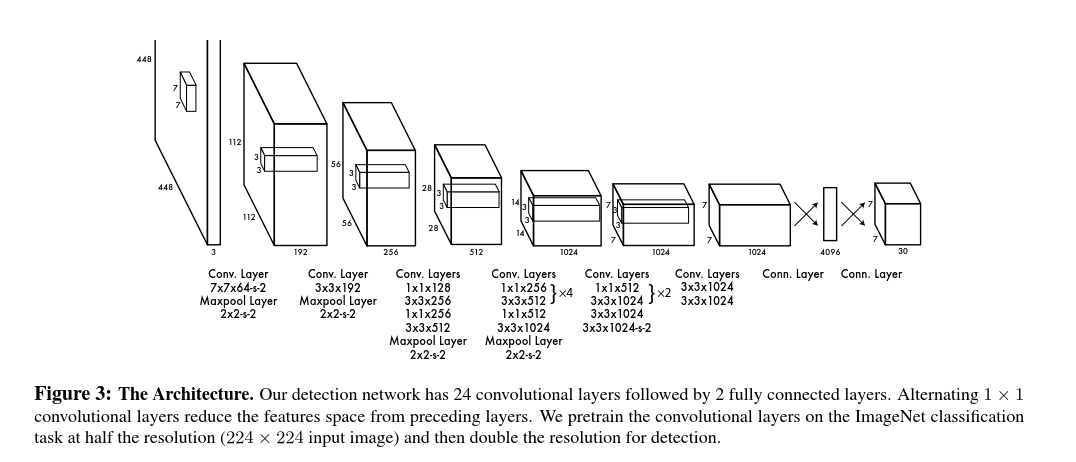
3. YOLO still lags behind state-of-the-art detection systems in accuracy. While it can quickly identify objects in images it struggles to precisely localize some objects, especially small ones. We examine these tradeoffs further in our experiments

YOLO缺陷：相比较于最新的区域提名方法，在检测精度方面仍有欠缺，特别是对一些比较小的目标进检测时。（需要改进）

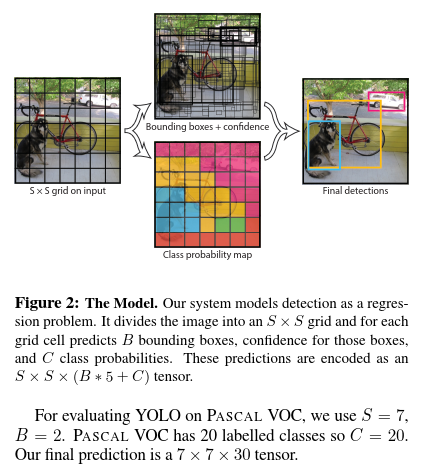
4. Each bounding box consists of 5 predictions: x, y, w, h, and conﬁdence. The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. Finally the conﬁdence prediction represents the IOU between the predicted box and any ground truth box.

**（x,y）**表示包围盒相对于grid的位置。W、H相对于整张图片的宽和高，conﬁdence表示是一个目标的置信度（IOU intersection over union）

5.模型图



6.



7. Our ﬁnal layer predicts both class probabilities and bounding box coordinates. We normalize the bounding box width and height by the image width and height so that they fall between 0 and 1

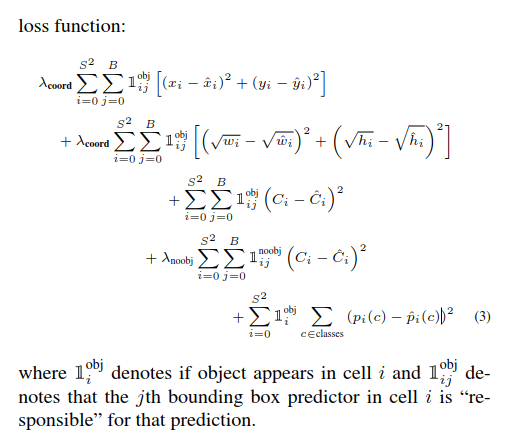
YOYO网络，在预测类别的同时，得出包围盒坐标，w和h被根据图像大小，被归一化到0-1的范围内。

8. We parametrize the bounding box x and y coordinates to be offsets of a particular grid cell location so they are also bounded between 0 and 1

包围盒的坐标x，y是相对于其所在的grid的位置。也是归属于0-1范围内

9.We use sum-squared error because it is easy to optimize, however it does not perfectly align with our goal of maximizing average precision. It weights localization error equally with classiﬁcation error which may not be ideal.Also, in every image many grid cells do not contain any object. This pushes the “conﬁdence” scores of those cells towards zero, often overpowering the gradient from cells that do contain objects. This can lead to model instability, causing training to diverge early on.

解决两个问题：①用平方误差来优化bondingbox②有些grid中可能无检测目标，导致检测算法不稳定的问题。引入两个调节系数λcoord和λnoobj



10.Our learning rate schedule is as follows: For the ﬁrst epochs we slowly raise the learning rate from 10−3 to 10−2. If we start at a high learning rate our model often diverges due to unstable gradients. We continue training with 10−2 for 75 epochs, then 10−3 for 30 epochs, and ﬁnally 10−4 for 30 epochs.

YOLO的学习率

10.Non-maximal suppression can be used to ﬁx these multiple detections

非极大值抑制，来修正多个可能的重复的检测框

11. YOLO imposes strong spatial constraints on bounding box predictions since each grid cell only predicts two boxes and can only have one class. This spatial constraint limits the number of nearby objects that our model can predict. Our model struggles with small objects that appear in groups, such as ﬂocks of birds.

YOLOv1算法缺陷：一个格子只能预测一类，并且只能有两个可能位置。对于成片出现的小目标，难以检测。