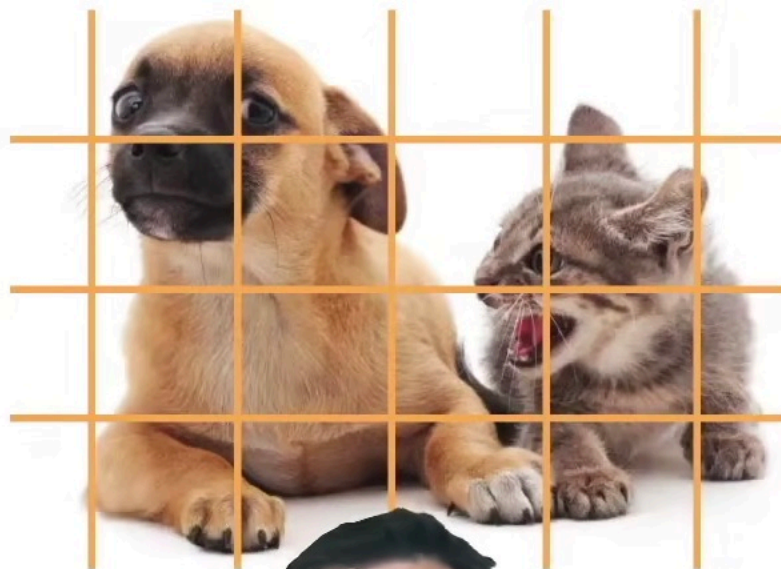


# YOLO (你只看一次)

- SSD中锚框大量重叠，因此浪费了很多计算
- YOLO 将图片均匀分成  $S \times S$  个锚框
- 每个锚框预测  $B$  个边缘框
- 后续版本 (V2,V3,V4...) 有持续改进



# 总结

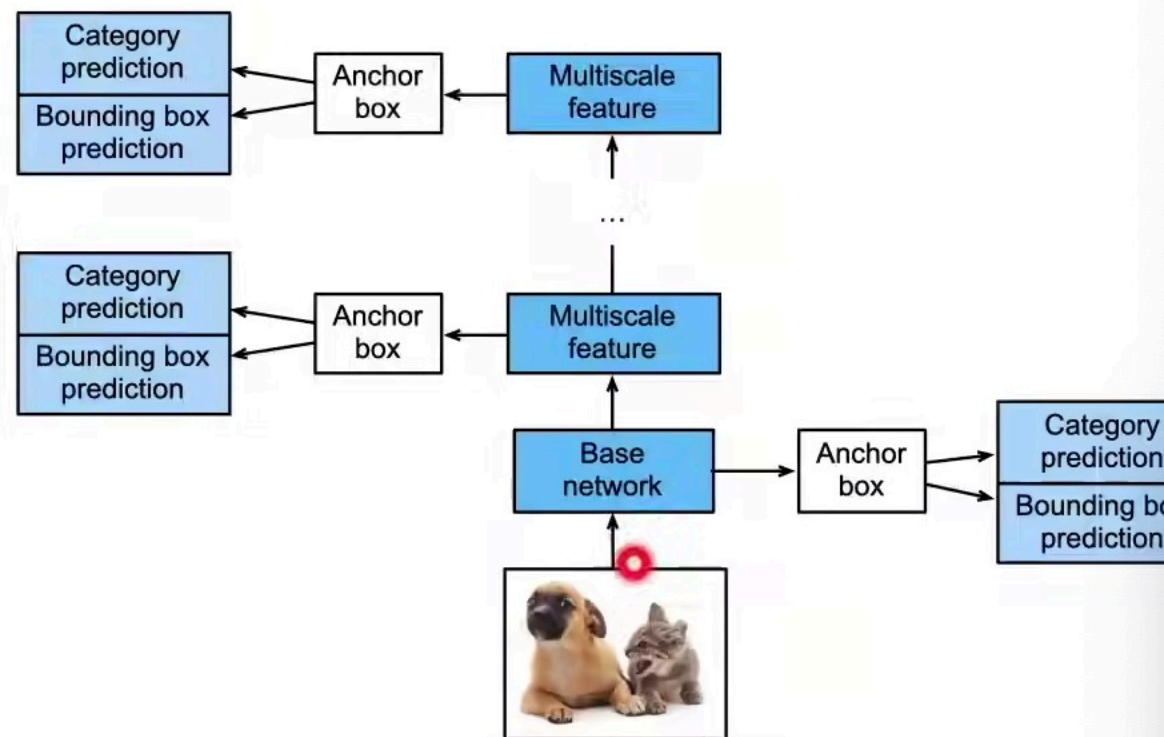


- SSD通过单神经网络来检测模型
- 以每个像素为中心的生成多个锚框
- 在多个段的输出上进行多尺度的检测



# SSD 模型

- 一个基础网络来抽取特征，然后多个卷积层块来减半高宽
- 在每段都生成锚框
  - 底部段来拟合小物体，顶部段来拟合大物体
- 对每个锚框预测类别和边缘框



# 总结

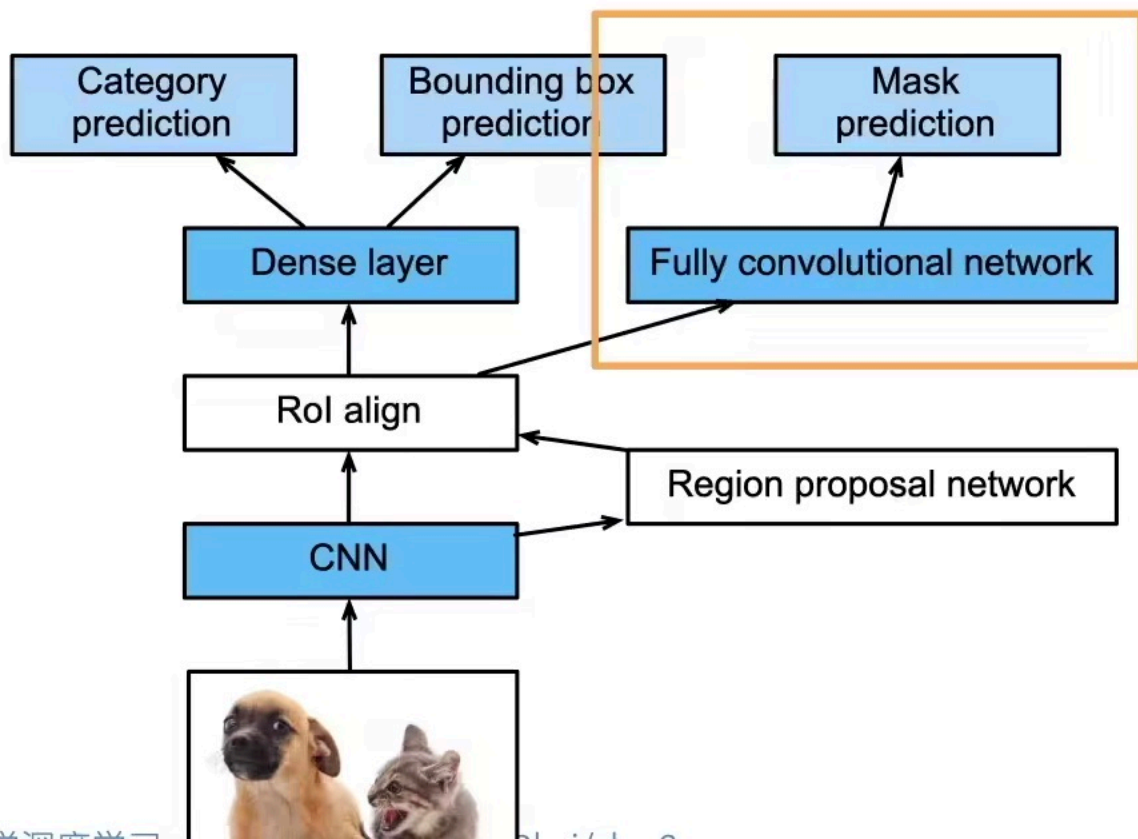
- R-CNN 是最早、也是最有名的一类基于锚框和CNN的目标检测算法
- Fast/Faster R-CNN持续提升性能
- Faster R-CNN 和 Mask R-CNN是在最求高精度场景下的常用算法



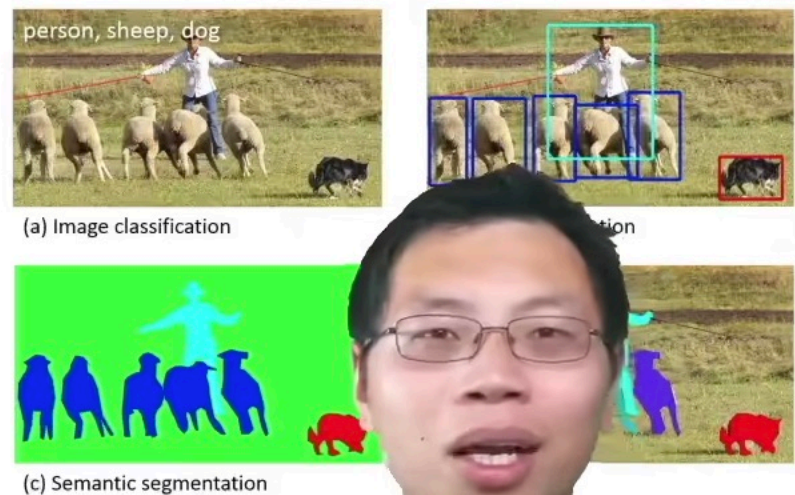


# Mask R-CNN

- 如果有像素级别的标号，使用FCN来利用这些信息

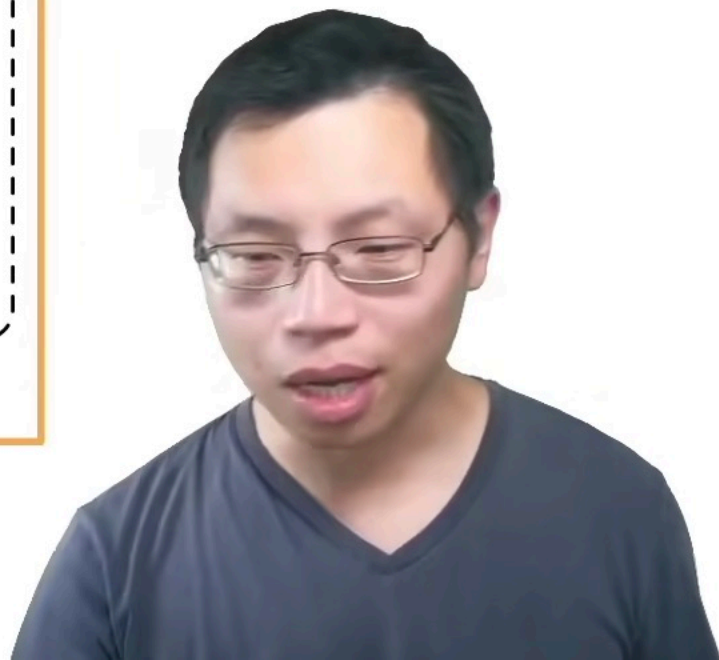
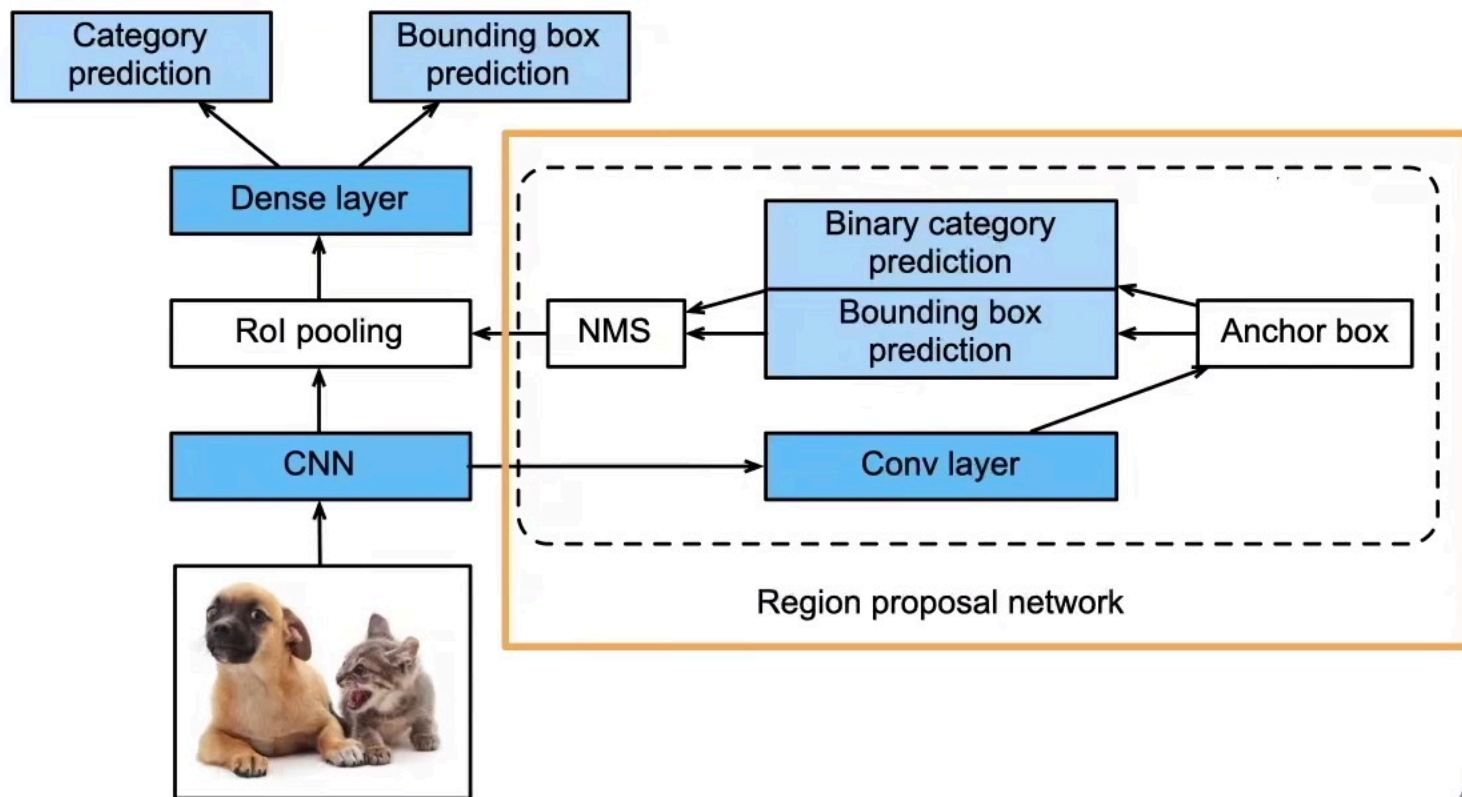


COCO



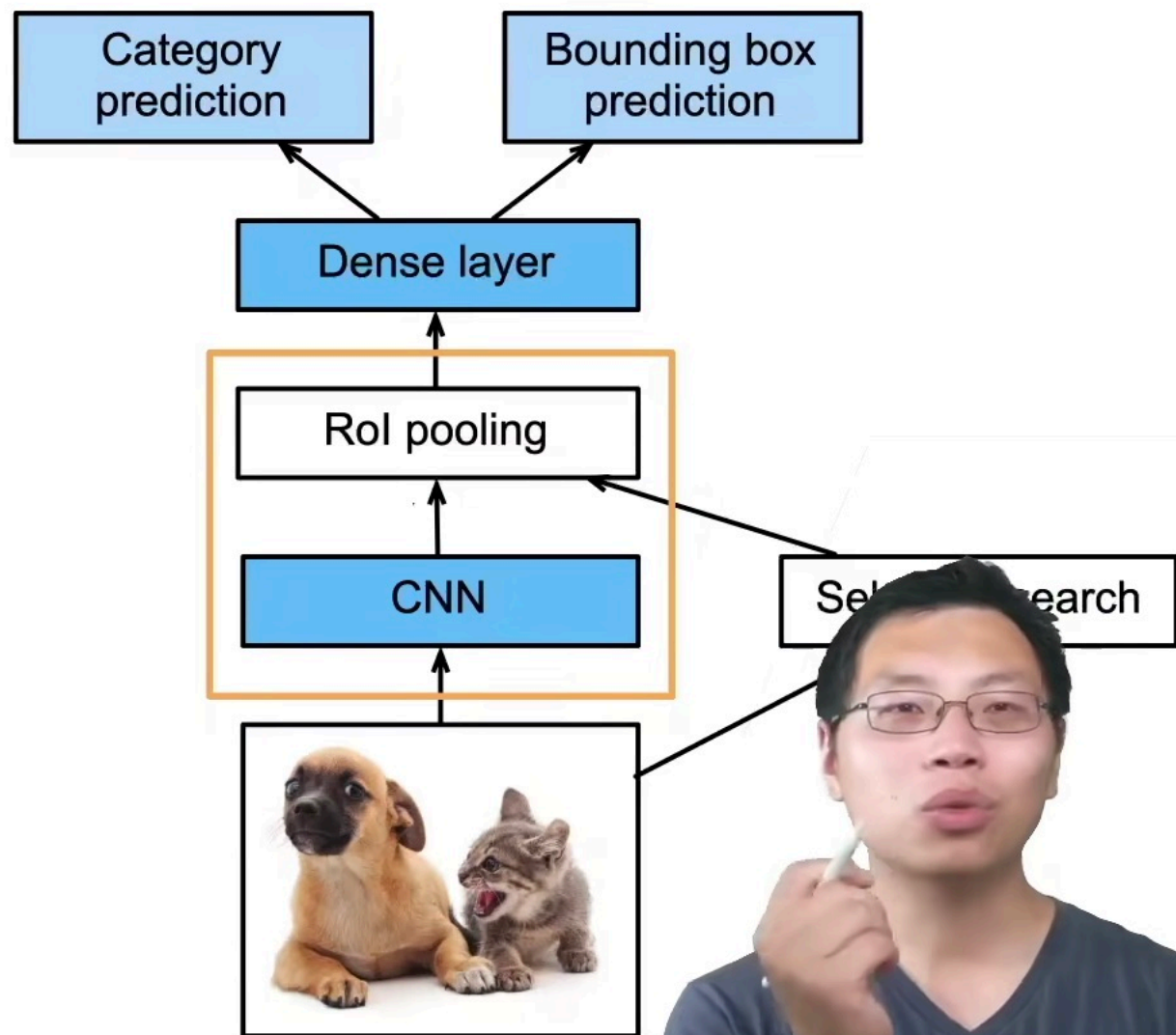
# Faster R-CNN

- 使用一个区域提议网络来替代启发式搜索来获得更好的锚框



# Fast RCNN

- 使用CNN对图片抽取特征
- 使用RoI池化层对每个锚框生成固定长度特征



# 兴趣区域 (RoI) 池化层

- 给定一个锚框，均匀分割成  $n \times m$  块，输出每块里的最大值
- 不管锚框多大，总是输出  $nm$  个值

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

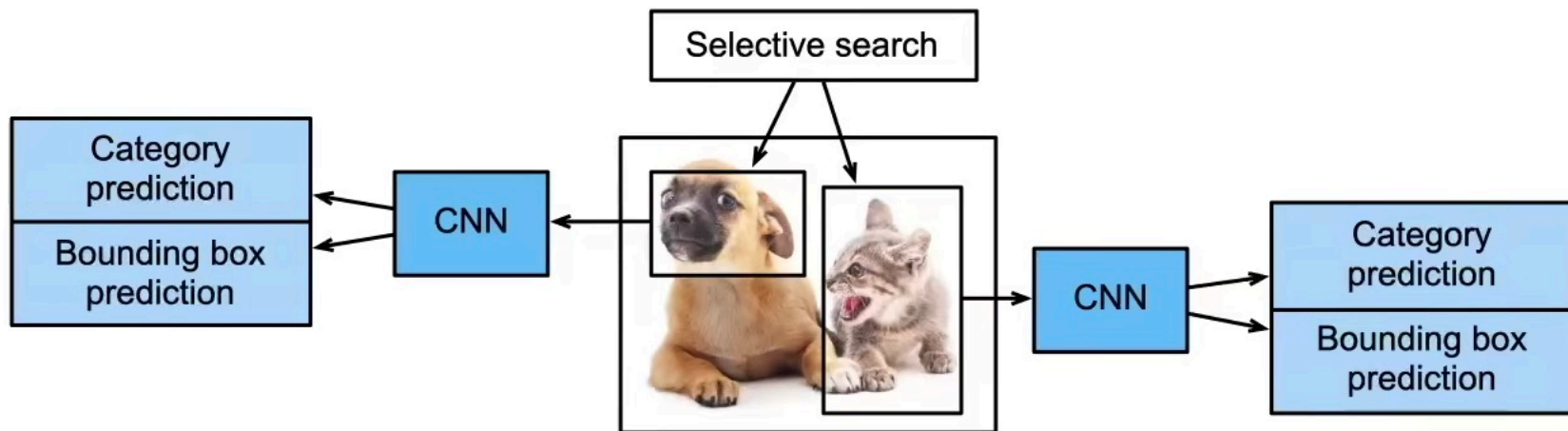
2 x 2 RoI  
Pooling

5	6
9	10





# R-CNN



- 使用启发式搜索算法来选择锚框
- 使用预训练模型来对每个锚框抽取特征
- 训练一个SVM来对类别分类
- 训练一个线性回归模型来预测边缘框偏移

