问题：回归损失是什么意思？

指的不是交叉熵损失或softmax 损失，指的回归损失一般指L2,L1,有时称为平滑L1损失(smooth L1 loss).一般而言，分类与回归的区别在于你的输出是间断还是连续的，

如果希望输出是一个间断的实例，如在固定的几个类别中分类，这时就要考虑交叉熵或softmax ,SVM loss，

如果输出是连续的，如人体关节图，这时倾向于L2,L1等

## 对象识别：object detection(deep learning core field)

很大的领域，与分类不同，给一张图片，用方框标示对象，并预测该对象从属的类别，与分类不同，一个图片可以有多个不同类别的对象

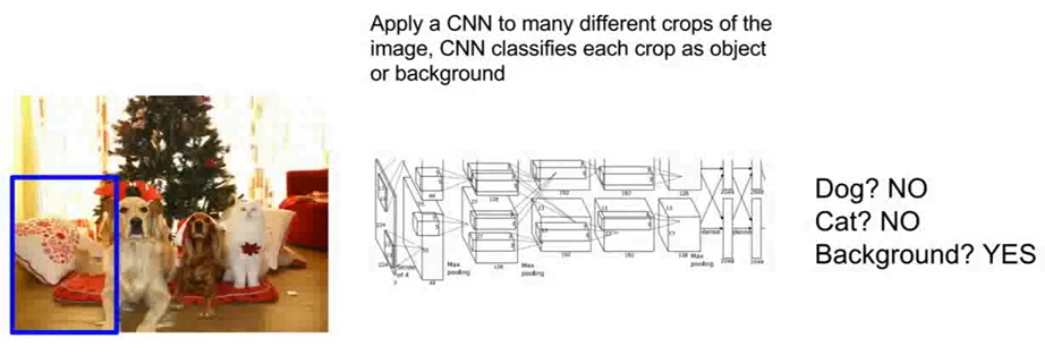
也不同于定位，因为定位只有一个框，而识别有不定数量的框



所以把对象等同于同归问题会比较难，所以用另一种思路来解决这类问题

一种尝试：Sliding Window

并增加一种类型：‘背景’



Region Proposals

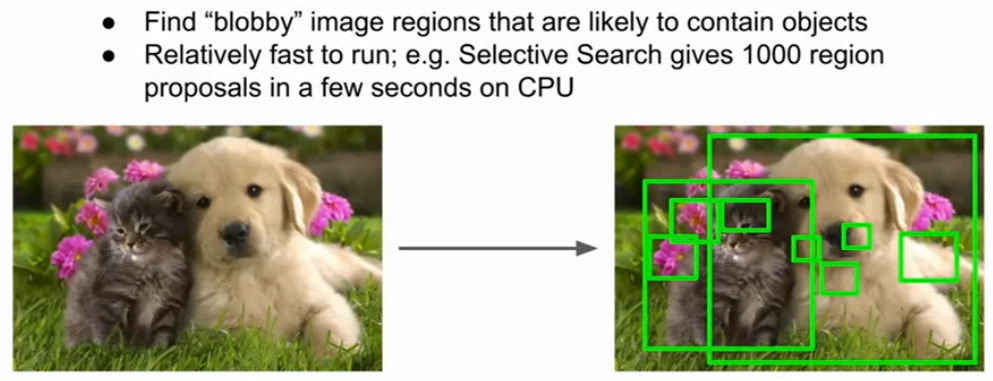
\*\*\* Selective search

Alexe et al, “Measuring the objectness of image windows”, TPAMI 2012

Uijlings et al, “Selective Search for Object Recognition”, IJCV 2013

Cheng et al, “BING: Binarized normed gradients for objectness estimation at 300fps”, CVPR 2014

Zitnick and Dollar, “Edge boxes: Locating object proposals from edges”, ECCV 2014



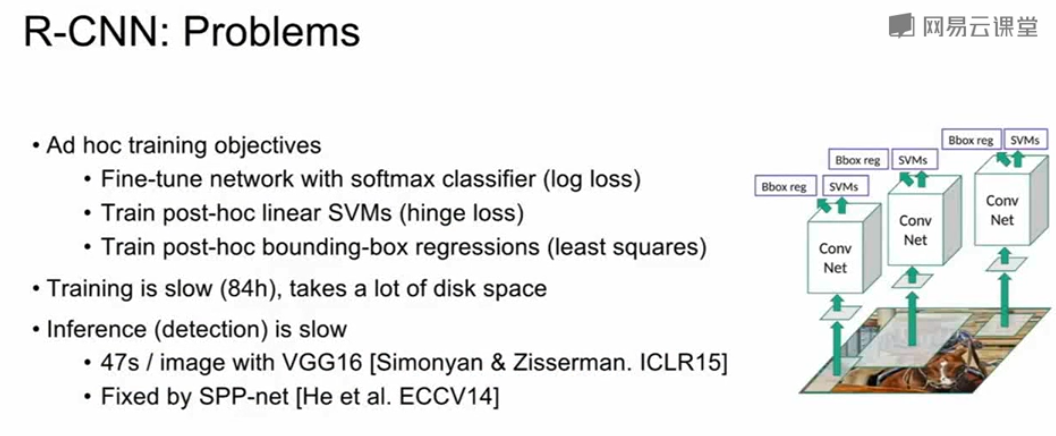
R-CNN

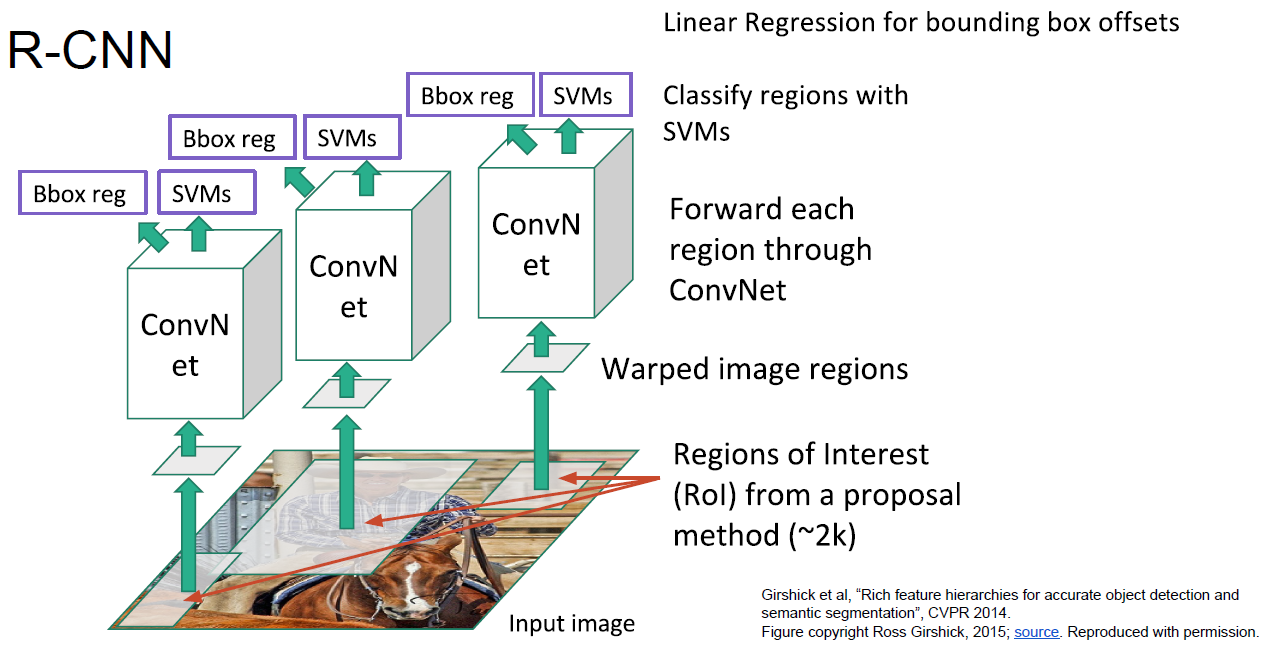
Girshick et al, “Rich feature hierarchies for accurate object detection and

semantic segmentation”, CVPR 2014.

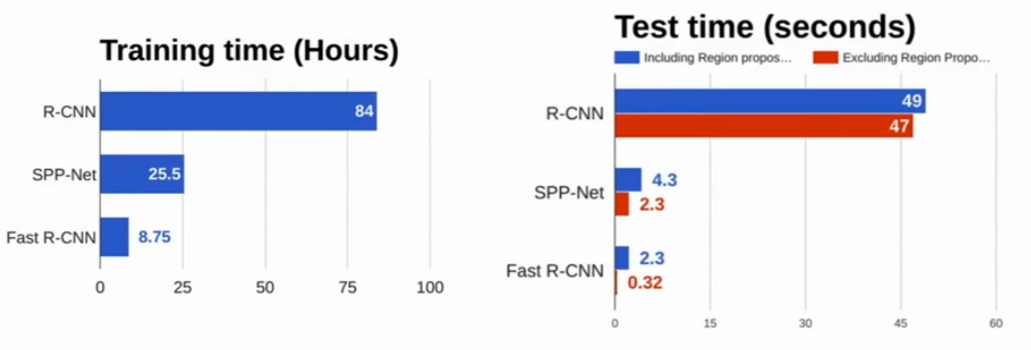
Girshick, “Fast R-CNN”, ICCV 2015.

但有一些问题：

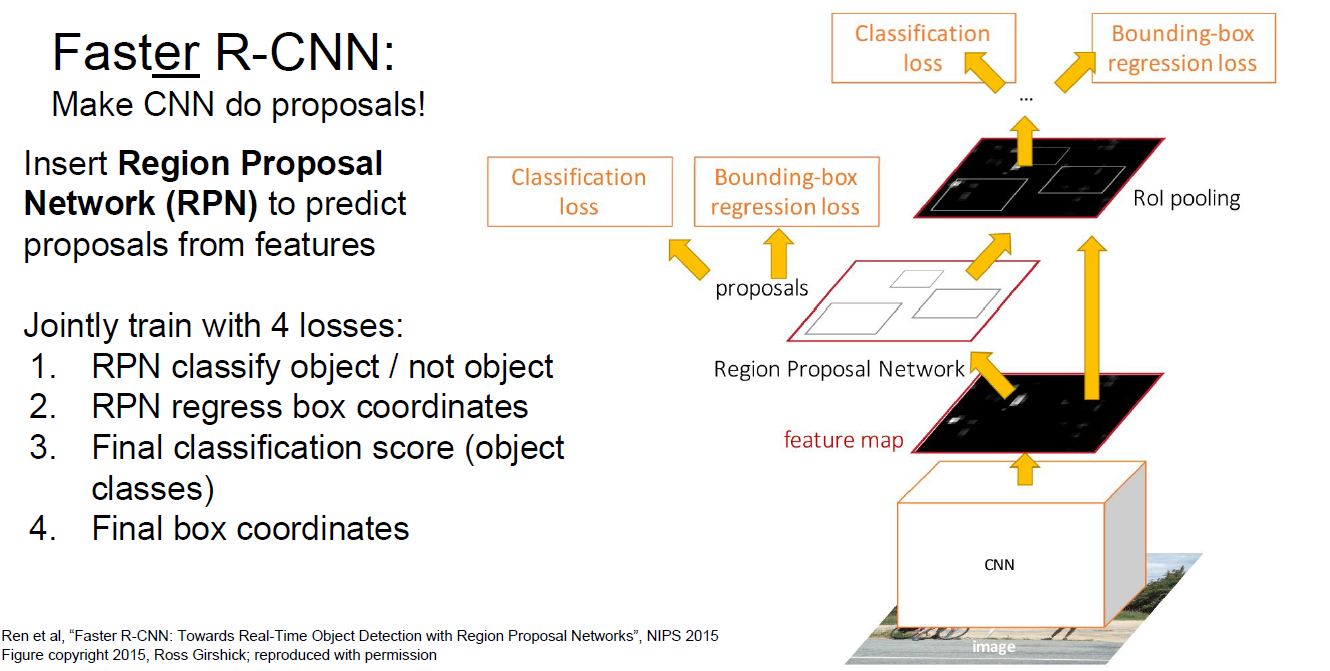




Fast R-CNN

Faster R-CNN

Make CNN do proposals

Ren et al, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, NIPS 2015  
Ren et al, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, NIPS 2015

## 没有proposals的对象识别算法：YOLO/SSD

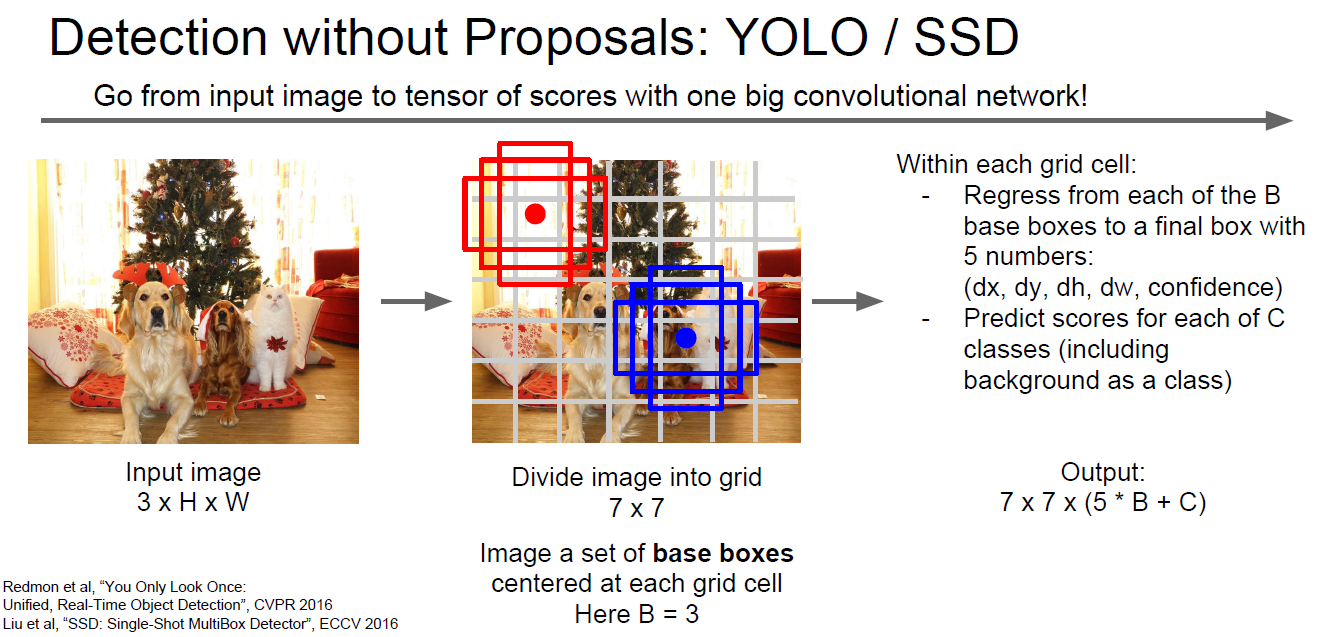
以回归的方式处理

Redmon et al, “You Only Look Once:

Unified, Real-Time Object Detection”, CVPR 2016

Liu et al, “SSD: Single-Shot MultiBox Detector”, ECCV 2016

YOLO(you only look once)



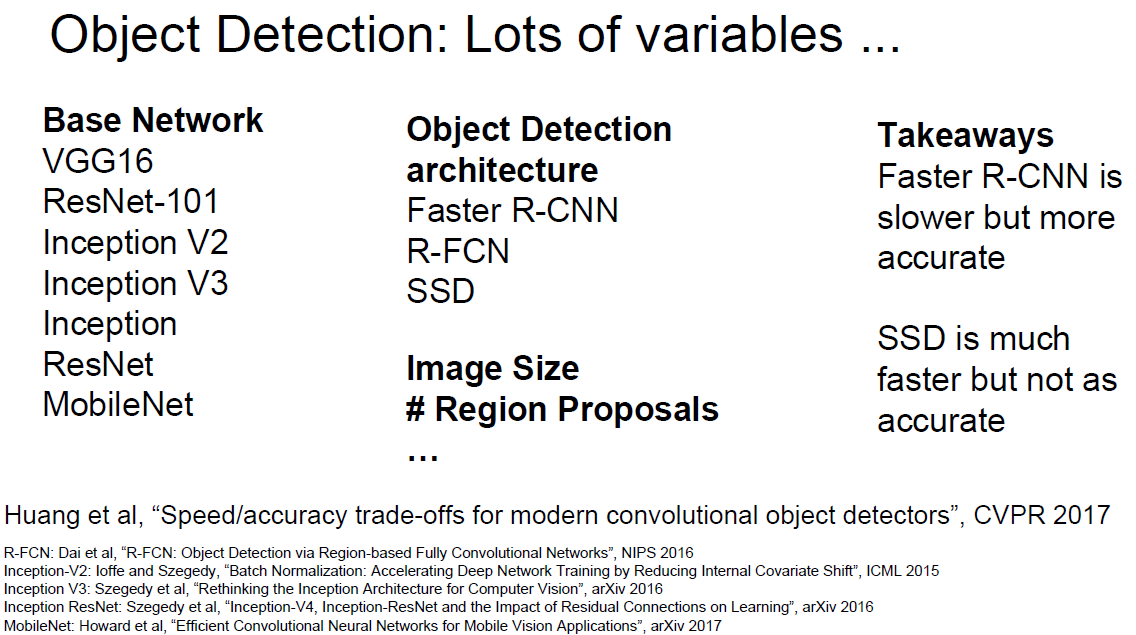
R-FCN: Dai et al, “R-FCN: Object Detection via Region-based Fully Convolutional Networks”, NIPS 2016

Inception-V2: Ioffe and Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, ICML 2015

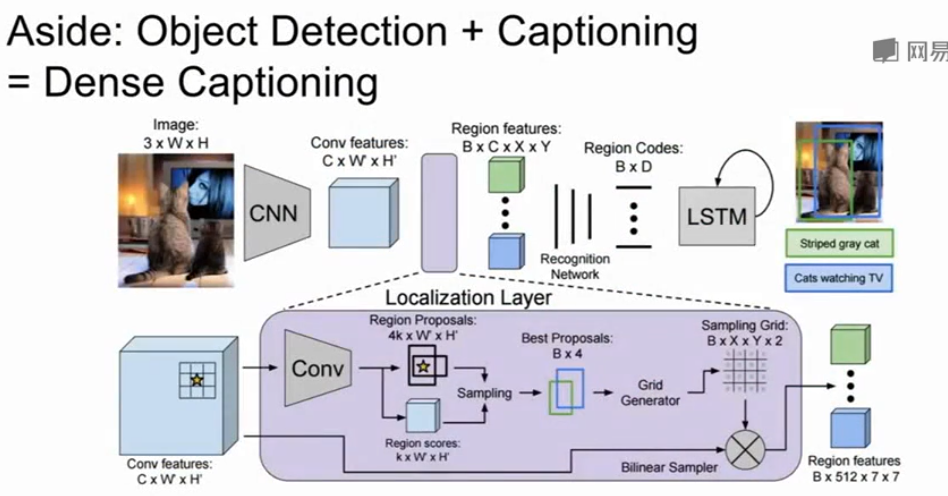
Inception V3: Szegedy et al, “Rethinking the Inception Architecture for Computer Vision”, arXiv 2016

Inception ResNet: Szegedy et al, “Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning”, arXiv 2016

MobileNet: Howard et al, “Efficient Convolutional Neural Networks for Mobile Vision Applications”, arXiv 2017



**Huang et al, “Speed/accuracy trade-offs for modern convolutional object detectors”, CVPR 2017**



Johnson, Karpathy, and Fei-Fei, “DenseCap: Fully Convolutional Localization Networks for Dense Captioning”, CVPR 2016

Figure copyright IEEE, 2016. Reproduced for educational purposes.

Mask R-CNN

He et al, “Mask R-CNN”, arXiv 2017