#DLUPC

Day 2 Lecture 1

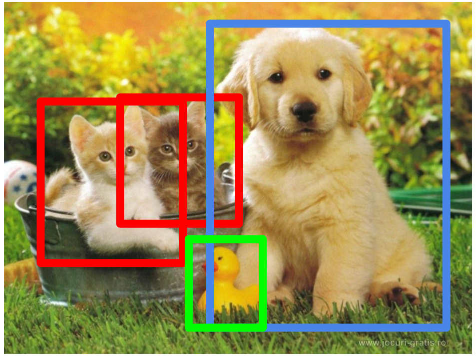
Object Detection

http://bit.ly/dlcv2018

Míriam Bellver
miriam.bellver@bsc.edu

PhD Candidate
Barcelona Supercomputing Center

Object Detection



CAT, DOG, DUCK

The task of assigning a **label** and a **bounding box** to all objects in the image

Object Detection as Classification

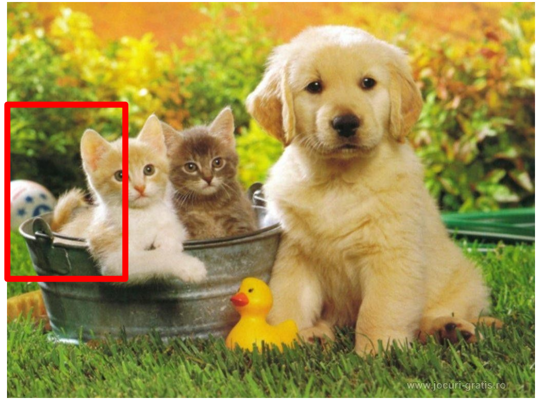
Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

Object Detection as Classification



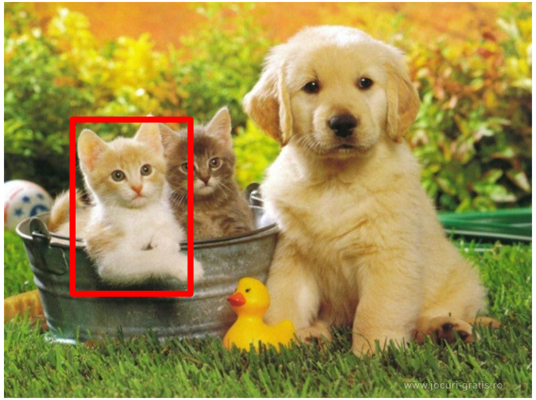
Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

Object Detection as Classification



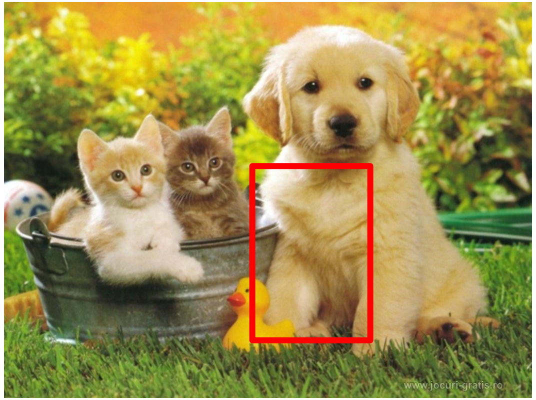
Classes = [cat, dog, duck]

Cat ? YES

Dog ? NO

Duck? NO

Object Detection as Classification



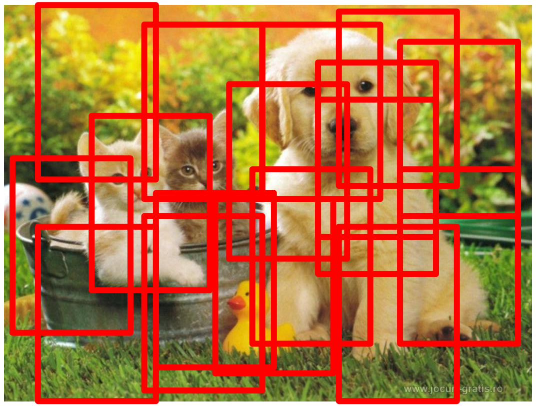
Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

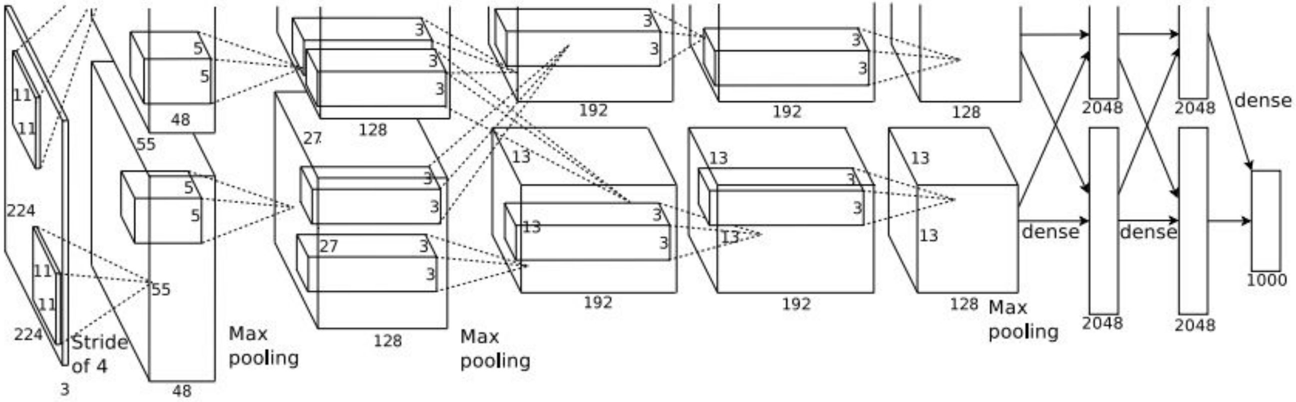
Object Detection as Classification



Problem:
Too many positions & scales to test

Solution: If your classifier is fast enough, go for it

Object Detection with ConvNets?



Convnets are computationally demanding. We can’t test all positions & scales !

Solution: Look at a tiny subset of positions. Choose them wisely :)

Object Detection: Datasets

20 categories
6k training images
6k validation images
10k test images

80 categories
200k training images
60k val + test images

200 categories
456k training images
60k validation + test images

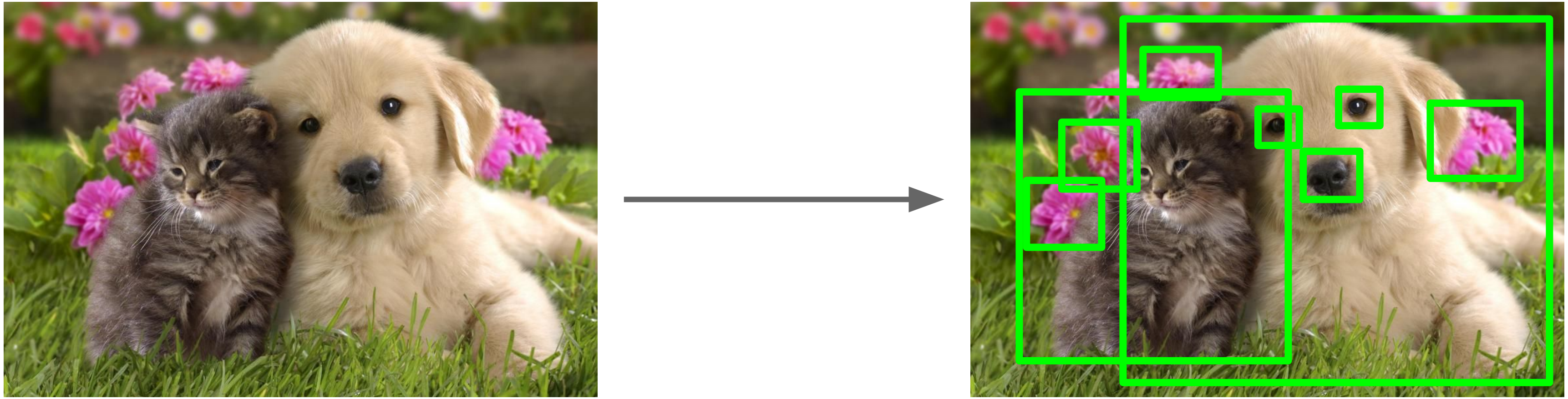
Outline

Proposal-based methods

Proposal-free methods

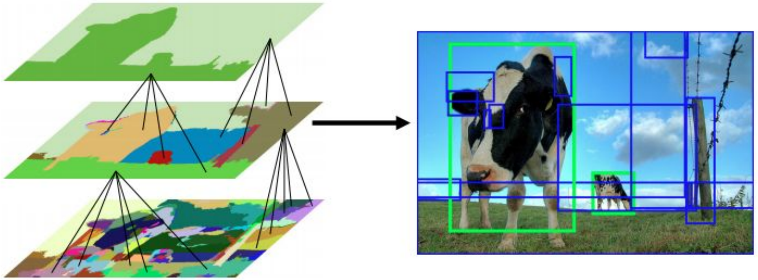
Region Proposals

- Find “blobby” image regions that are likely to contain objects
- “Class-agnostic” object detector

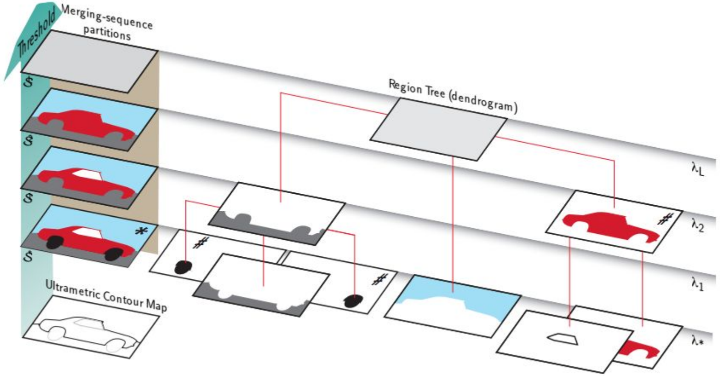


Slide Credit: [CS231n]

Region Proposals



Selective Search (SS)

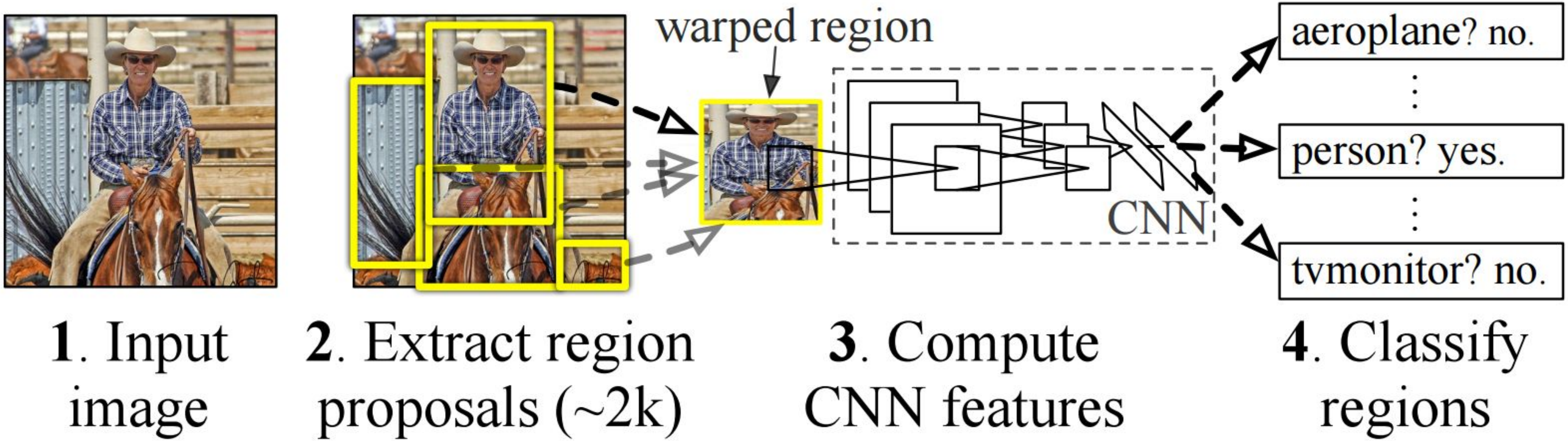


Multiscale Combinatorial Grouping (MCG)

[SS] Uijlings et al. Selective search for object recognition. IJCV 2013

[MCG] Arbeláez, Pont-Tuset et al. Multiscale combinatorial grouping. CVPR 2014

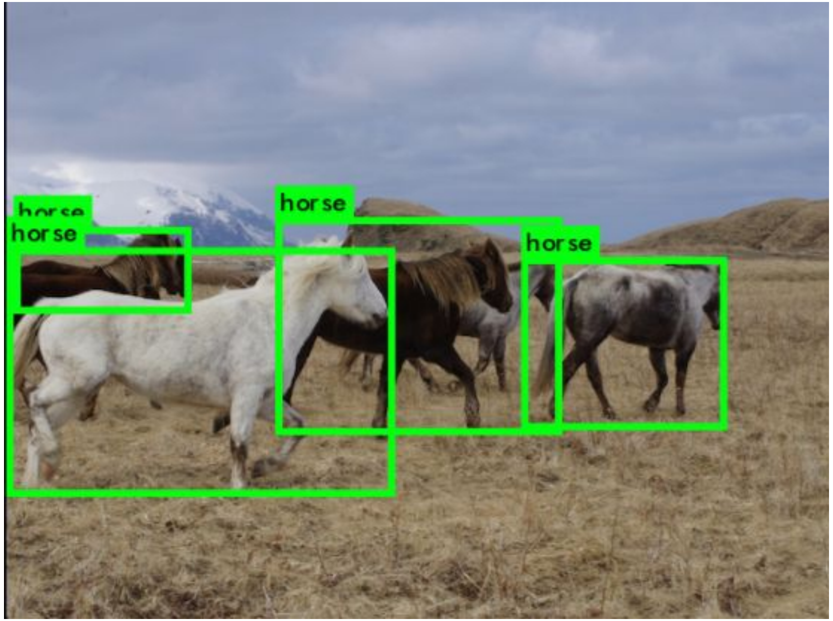
Object Detection with Convnets: R-CNN



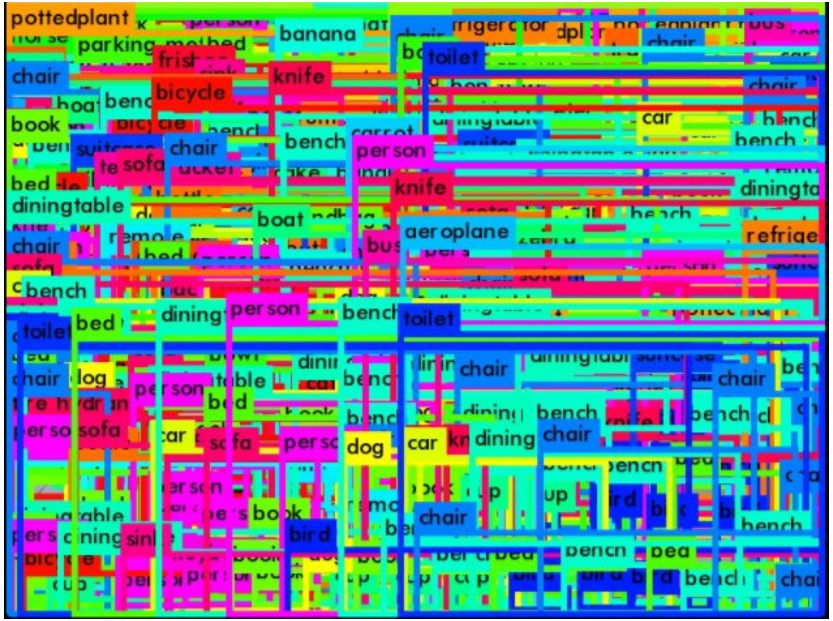
Girshick et al. Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR 2014

R-CNN

We expect:

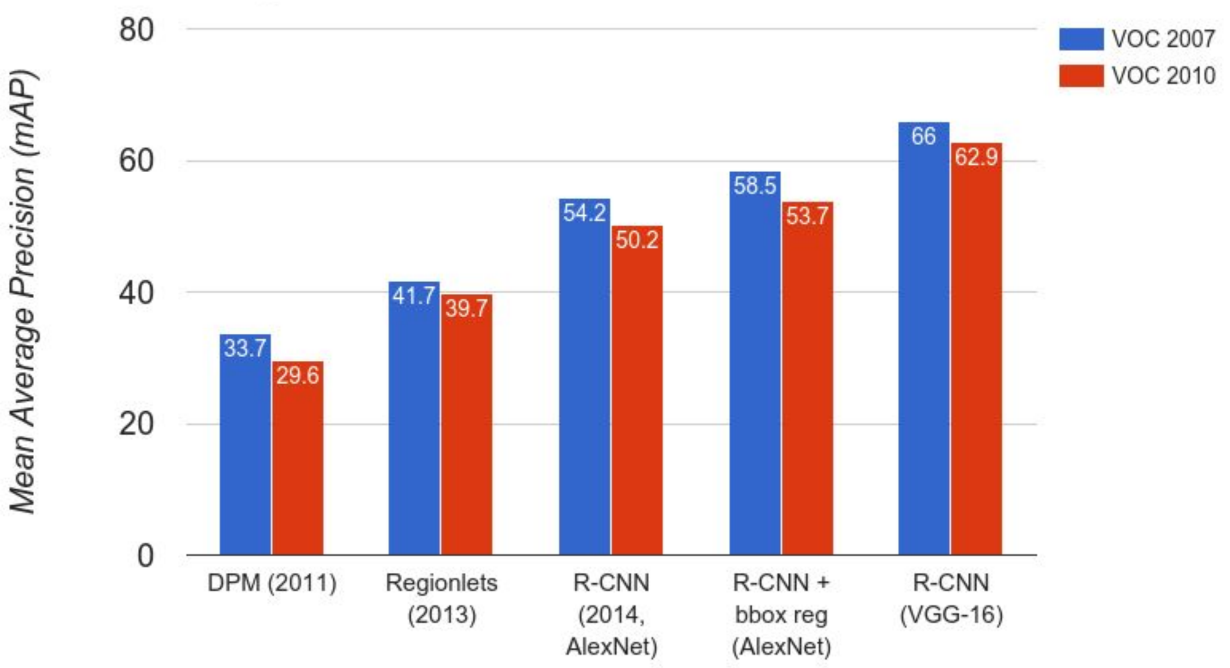


We get:



Non Maximum Suppression + score threshold

R-CNN



Girshick et al. Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR 2014

R-CNN: Problems

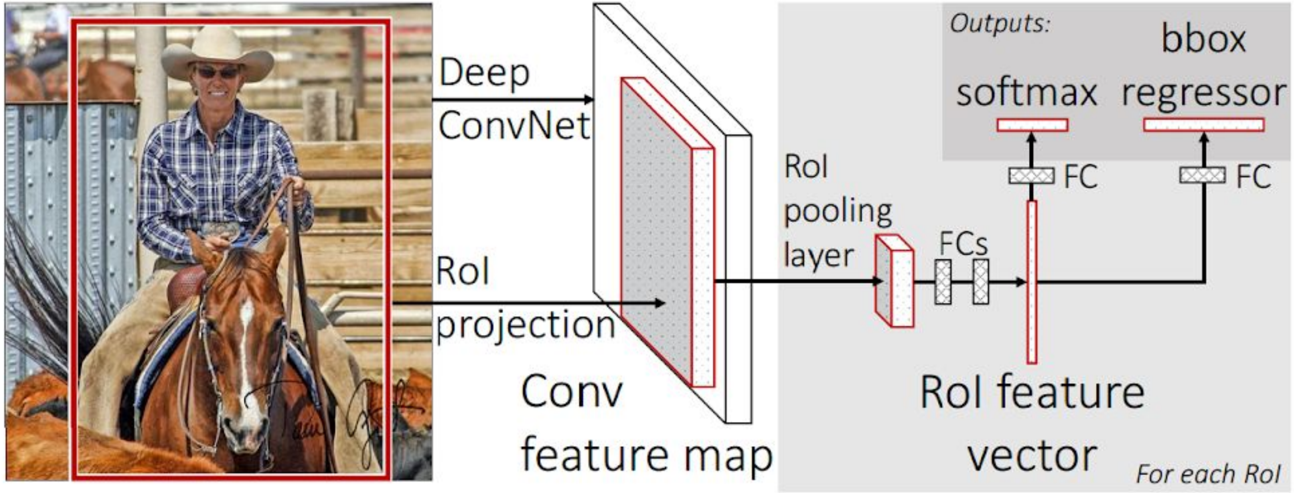
1. Slow at test-time: need to run full forward pass of CNN for each region proposal

2. SVMs and regressors are post-hoc: CNN features not updated in response to SVMs and regressors

Slide Credit: [CS231n]

Fast R-CNN

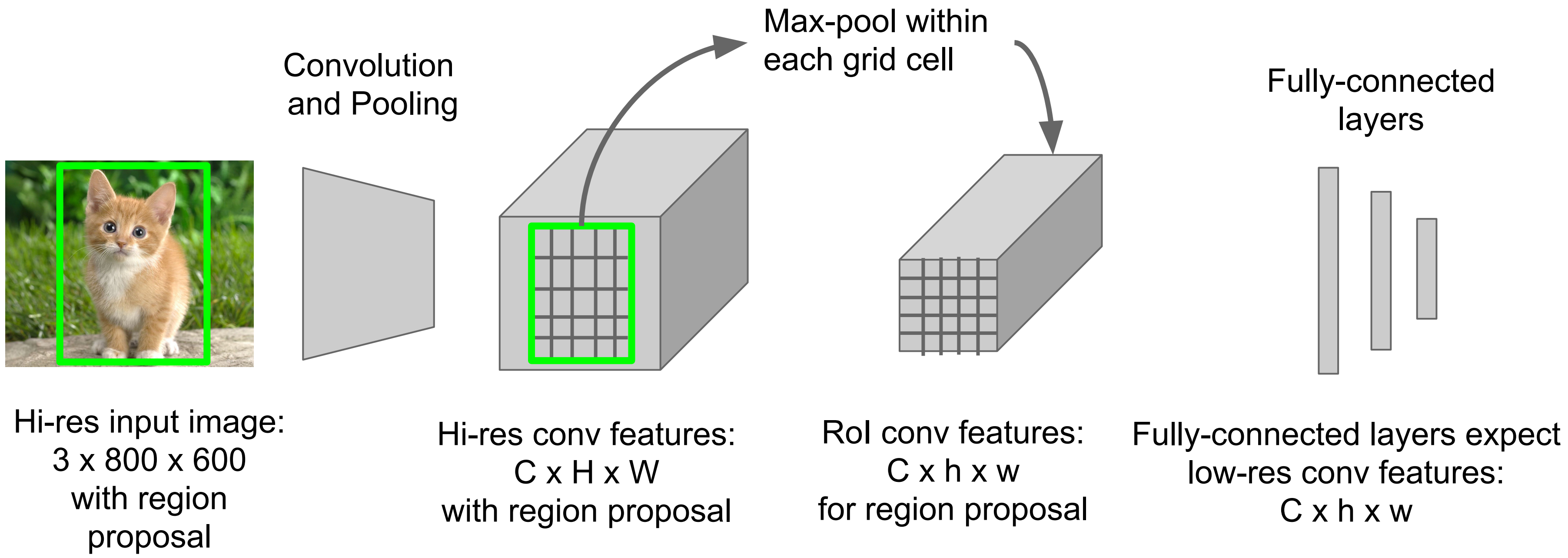
R-CNN Problem #1: Slow at test-time: need to run full forward pass of CNN for each region proposal



Solution: Share computation of convolutional layers between region proposals for an image

Girshick Fast R-CNN. ICCV 2015

Fast R-CNN: Sharing features

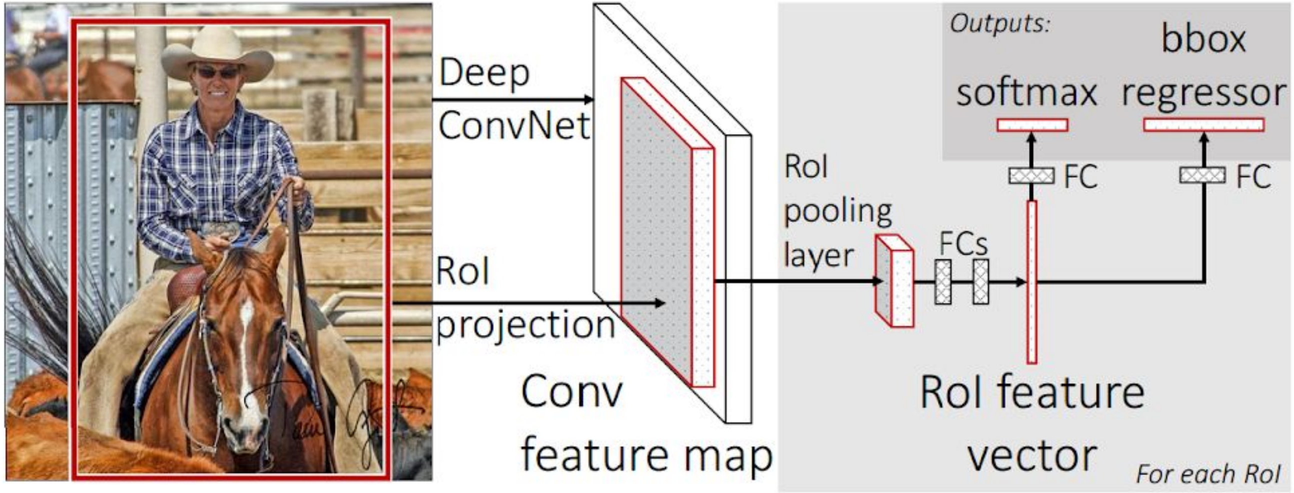


Girshick Fast R-CNN. ICCV 2015

Slide Credit: CS231n

Fast R-CNN

R-CNN Problem #2&3: SVMs and regressors are post-hoc. Complex training.



Solution: Train it all together end to end

Girshick Fast R-CNN. ICCV 2015

Fast R-CNN

		R-CNN	Fast R-CNN
Faster!	Training Time:	84 hours	**9.5 hours**
	(Speedup)	1x	**8.8x**
FASTER!	Test time per image	47 seconds	**0.32 seconds**
	(Speedup)	1x	**146x**
Better!	mAP (VOC 2007)	66.0	**66.9**

Using VGG-16 CNN on Pascal VOC 2007 dataset

Slide Credit: [CS231n]

Fast R-CNN: Problem

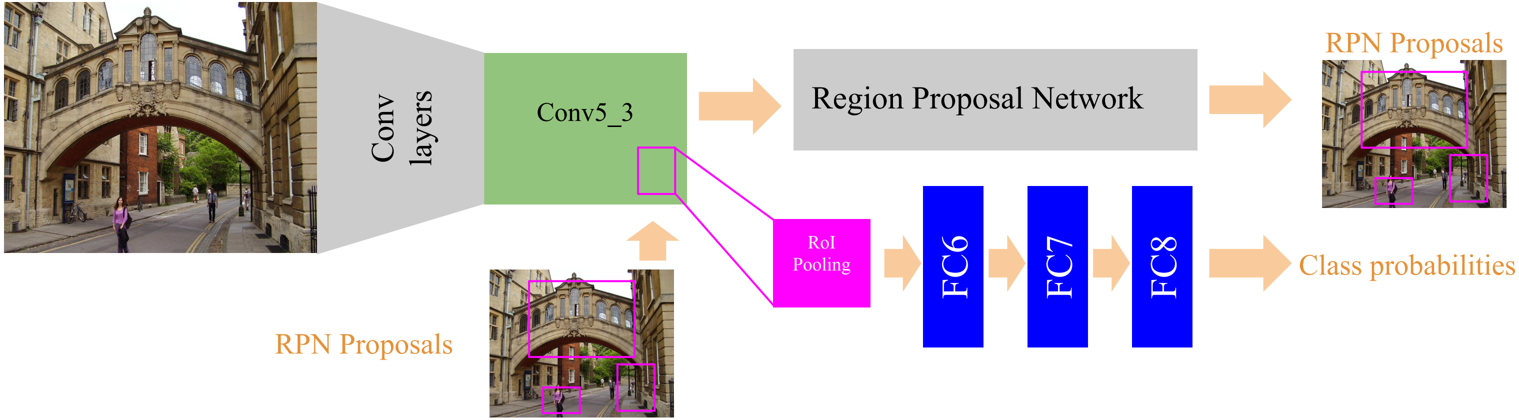
Test-time speeds don't include region proposals

	R-CNN	**Fast R-CNN**
Test time per image	47 seconds	**0.32 seconds**
(Speedup)	1x	**146x**
Test time per image with Selective Search	50 seconds	**2 seconds**
(Speedup)	1x	**25x**

Slide Credit: CS231n

Faster R-CNN

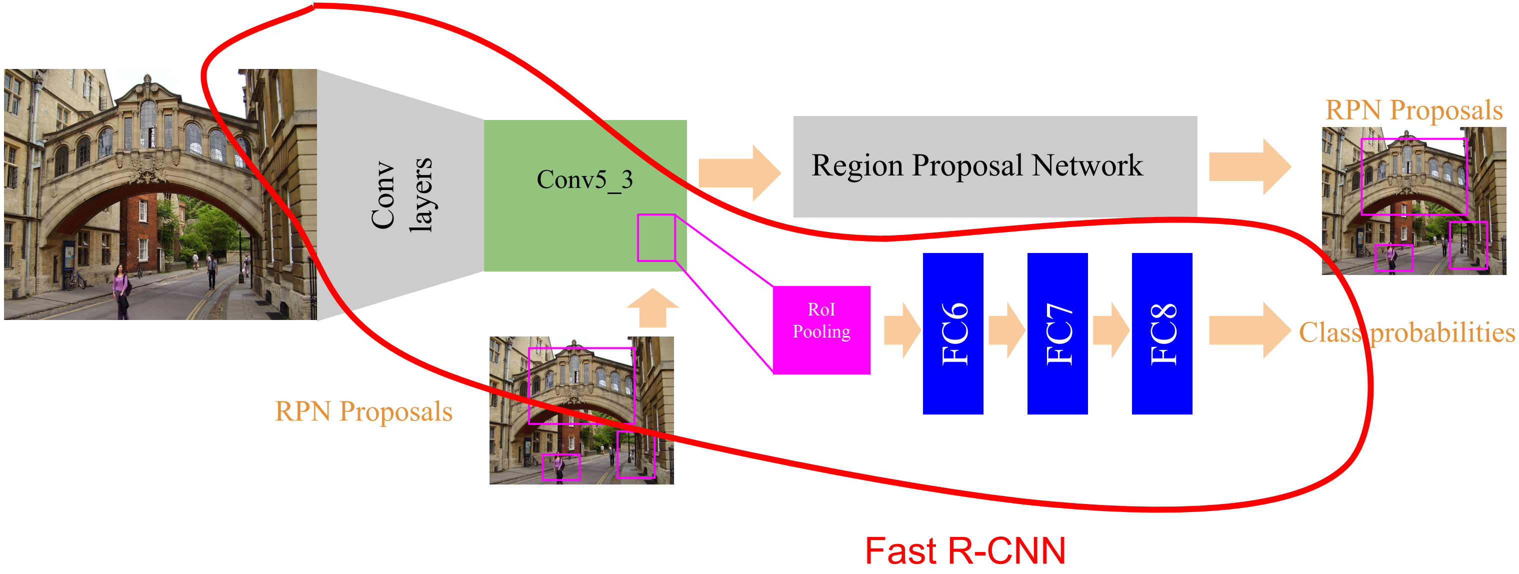
Learn proposals end-to-end sharing parameters with the classification network



Ren et al. [Faster R-CNN: Towards real-time object detection with region proposal networks.]() NIPS 2015

Faster R-CNN

Learn proposals end-to-end sharing parameters with the classification network

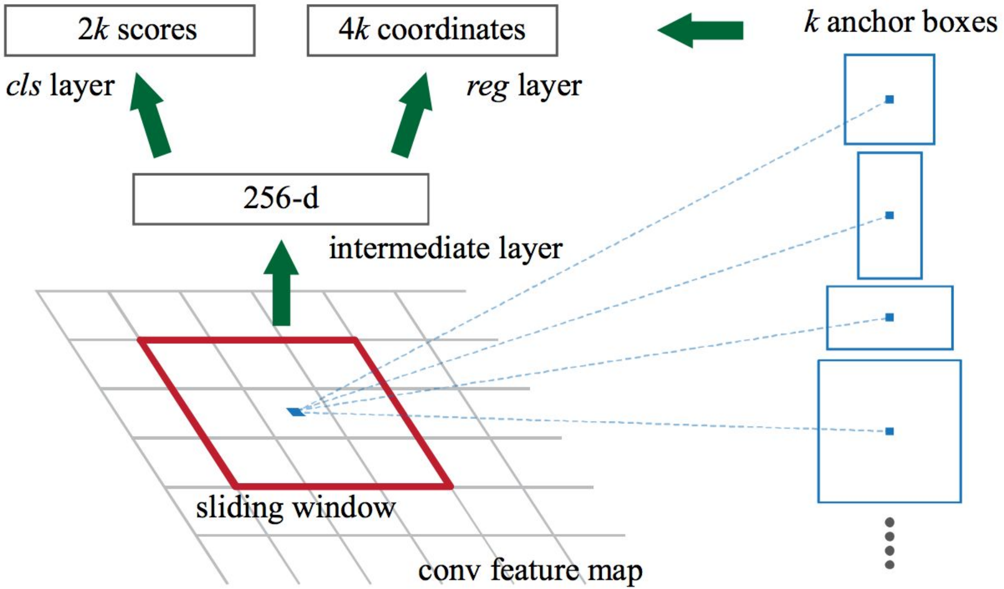


Ren et al. [Faster R-CNN: Towards real-time object detection with region proposal networks.]() NIPS 2015

Region Proposal Network

Bounding Box Regression

Objectness scores
(object/no object)



In practice, k = 9 (3 different scales and 3 aspect ratios)

Ren et al. [Faster R-CNN: Towards real-time object detection with region proposal networks.](#) NIPS 2015

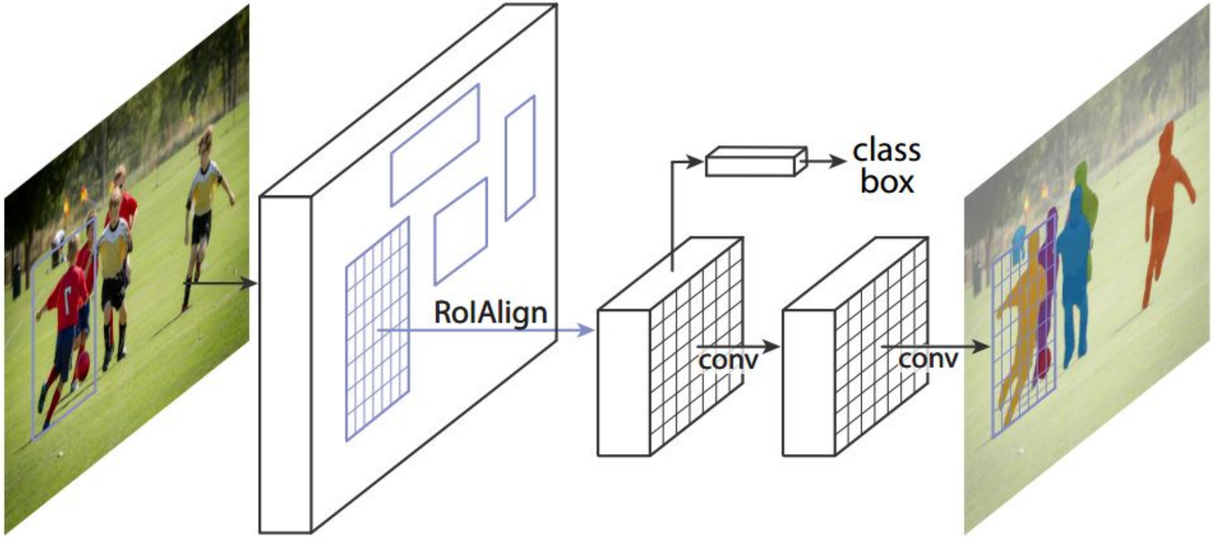
Faster R-CNN

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	**0.2 seconds**
(Speedup)	1x	25x	**250x**
mAP (VOC 2007)	66.0	**66.9**	**66.9**

Ren et al. Faster R-CNN: Towards real-time object detection with region proposal networks. NIPS 2015

Slide Credit: CS231n

Mask R-CNN



He et al. [Mask R-CNN](). ICCV 2017

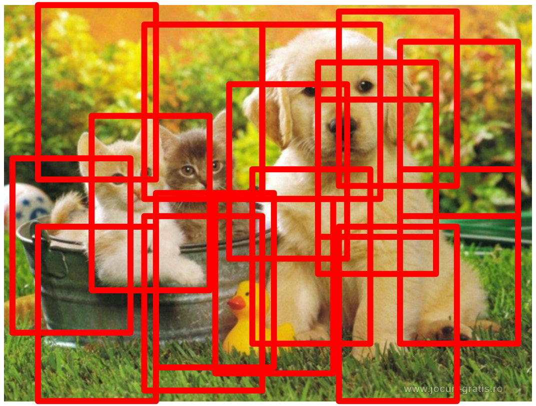
Outline

Proposal-based methods

Proposal-free methods

One-stage methods

Previously… :

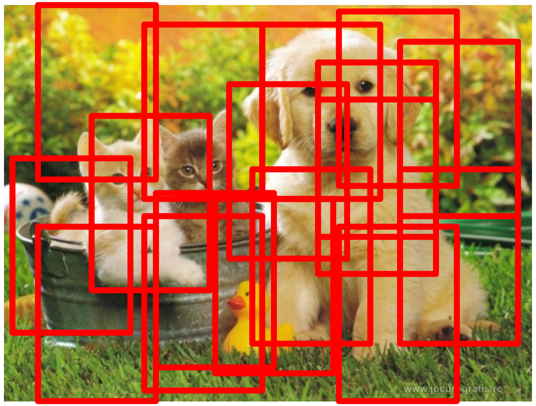


Problem:
Too many positions & scales to test

Solution: If your classifier is fast enough, go for it

One-stage methods

Previously… :

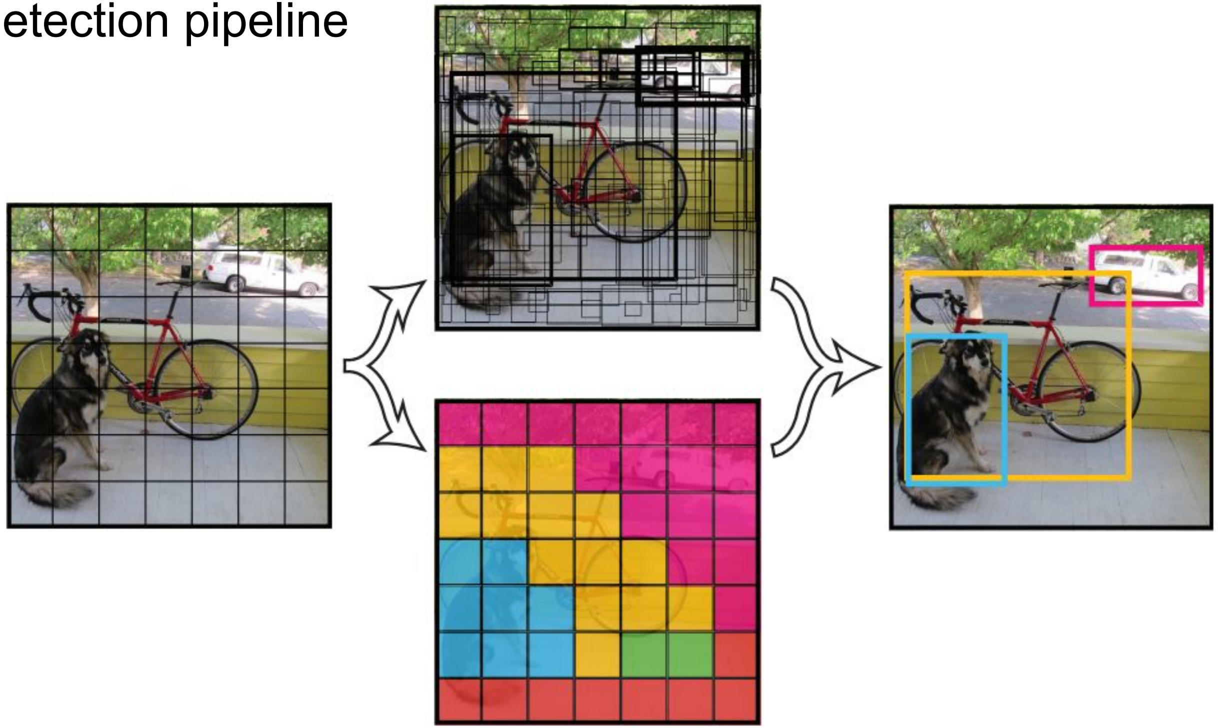


Problem:
Too many positions & scales to test

Modern detectors parallelize feature extraction across all locations.
Region classification is not slow anymore!

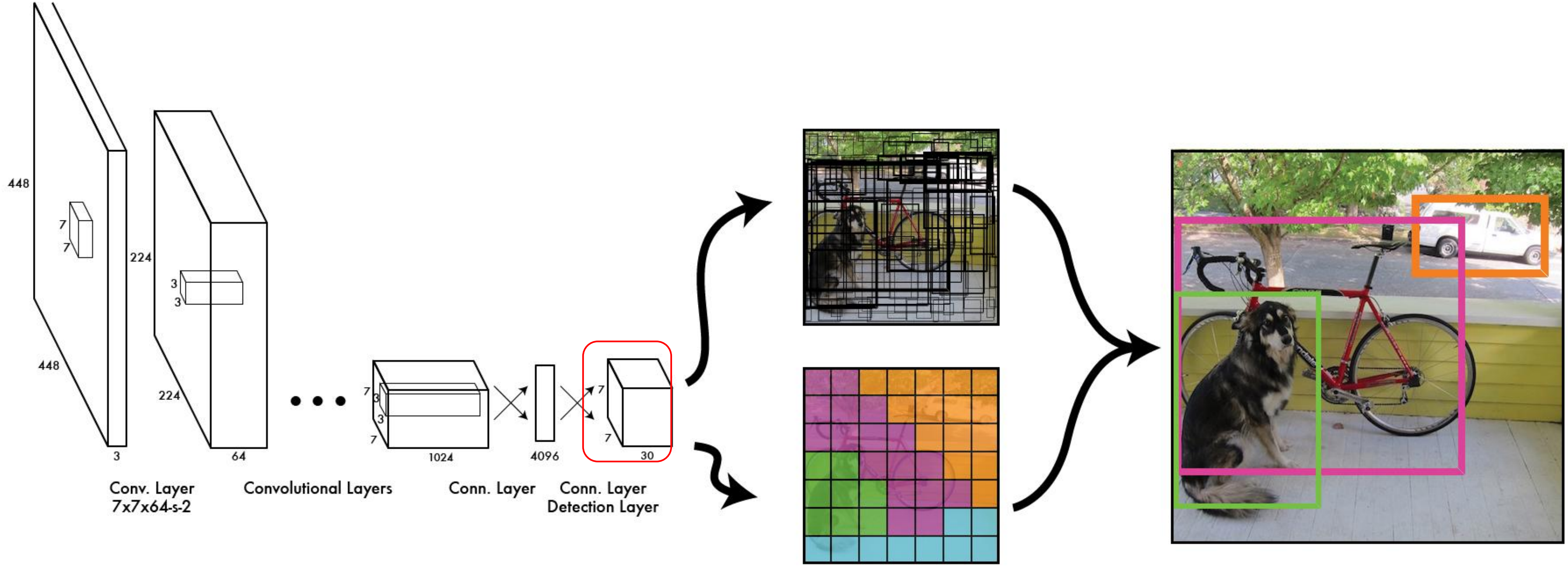
YOLO: You Only Look Once

Proposal-free object detection pipeline



Redmon et al. [You Only Look Once: Unified, Real-Time Object Detection,](#) CVPR 2016

YOLO: You Only Look Once

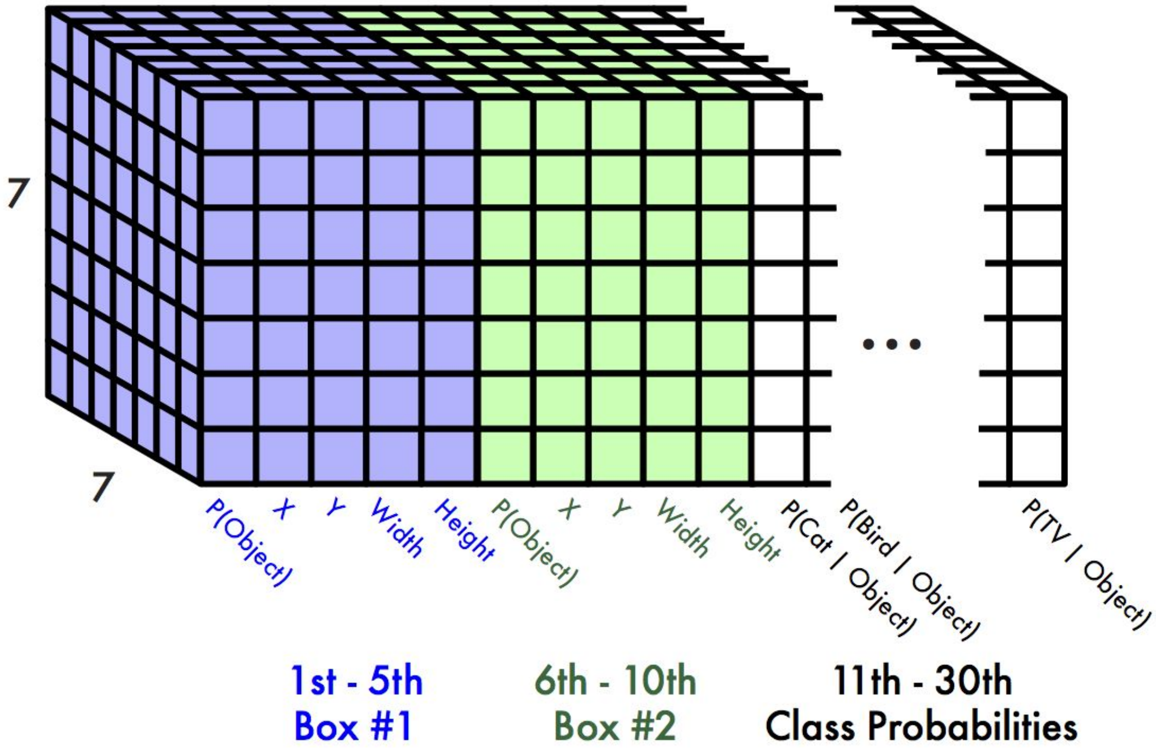


Redmon et al. [You Only Look Once: Unified, Real-Time Object Detection,](#) CVPR 2016

YOLO: You Only Look Once

Each cell predicts:

- For each bounding box:
  - 4 coordinates (x, y, w, h)
  - 1 confidence value
- Some number of class probabilities



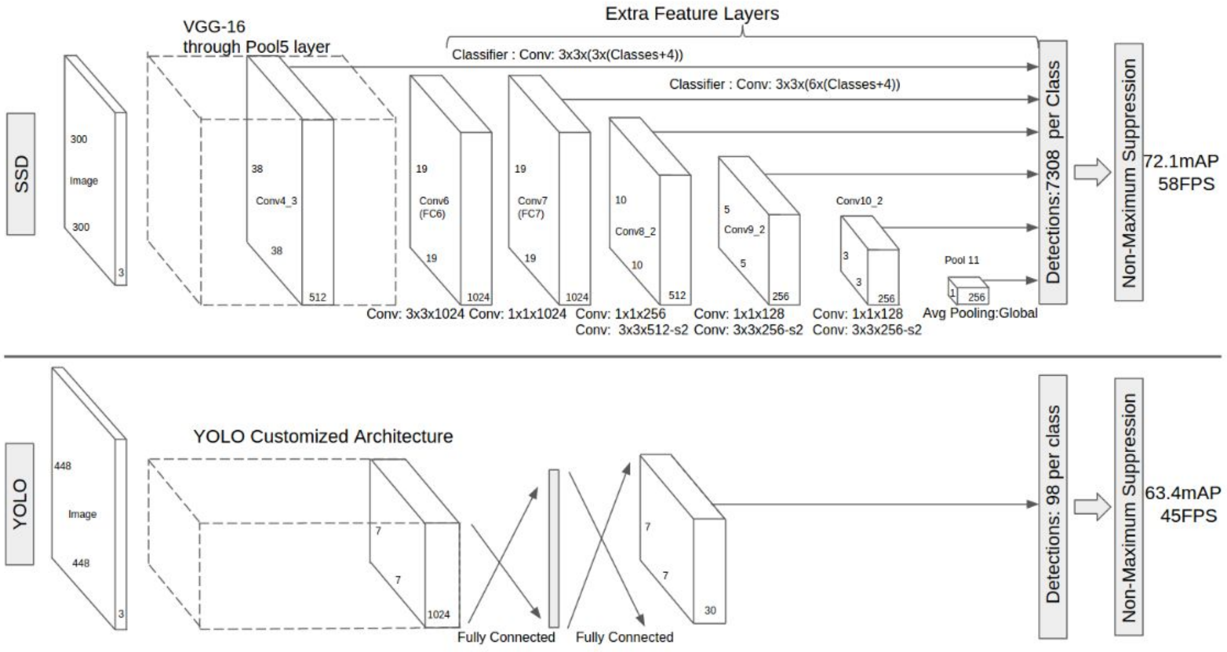
For Pascal VOC:

- 7x7 grid
- 2 bounding boxes / cell
- 20 classes

7 x 7 x (2 x 5 + 20) = 7 x 7 x 30 tensor = **1470 outputs**

SSD: Single Shot MultiBox Detector

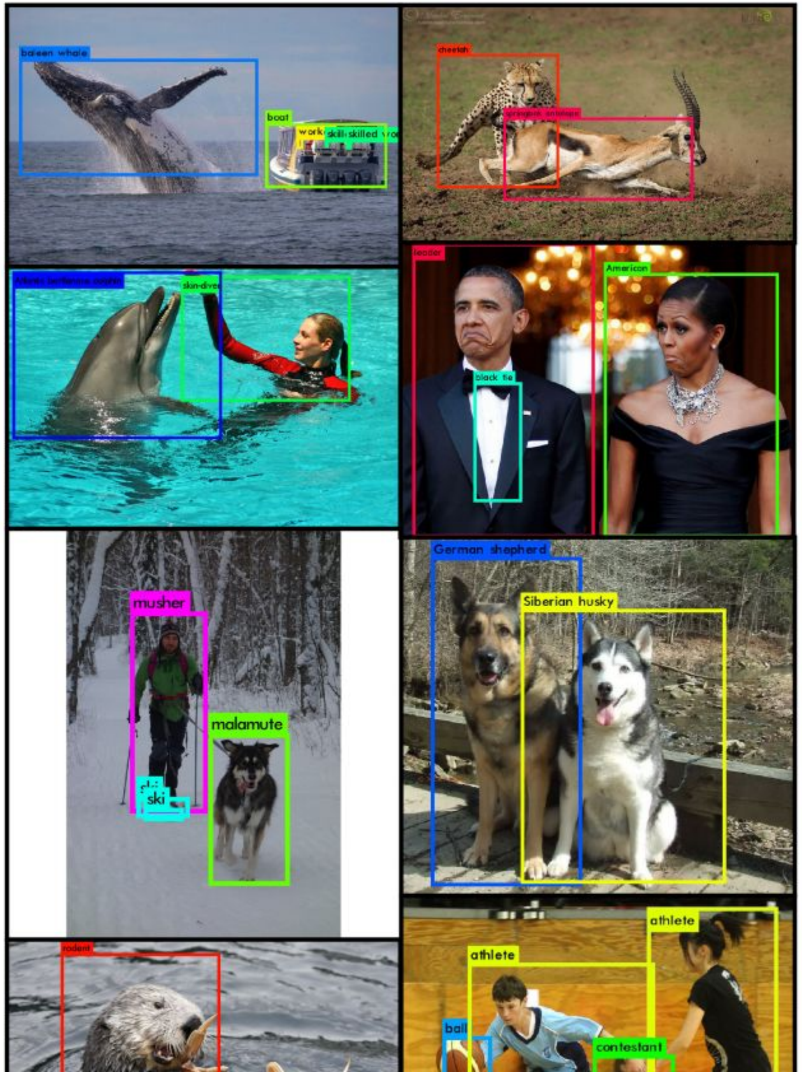
Same idea as YOLO, + several predictors at different stages in the network



Liu et al. [SSD: Single Shot MultiBox Detector](), ECCV 2016

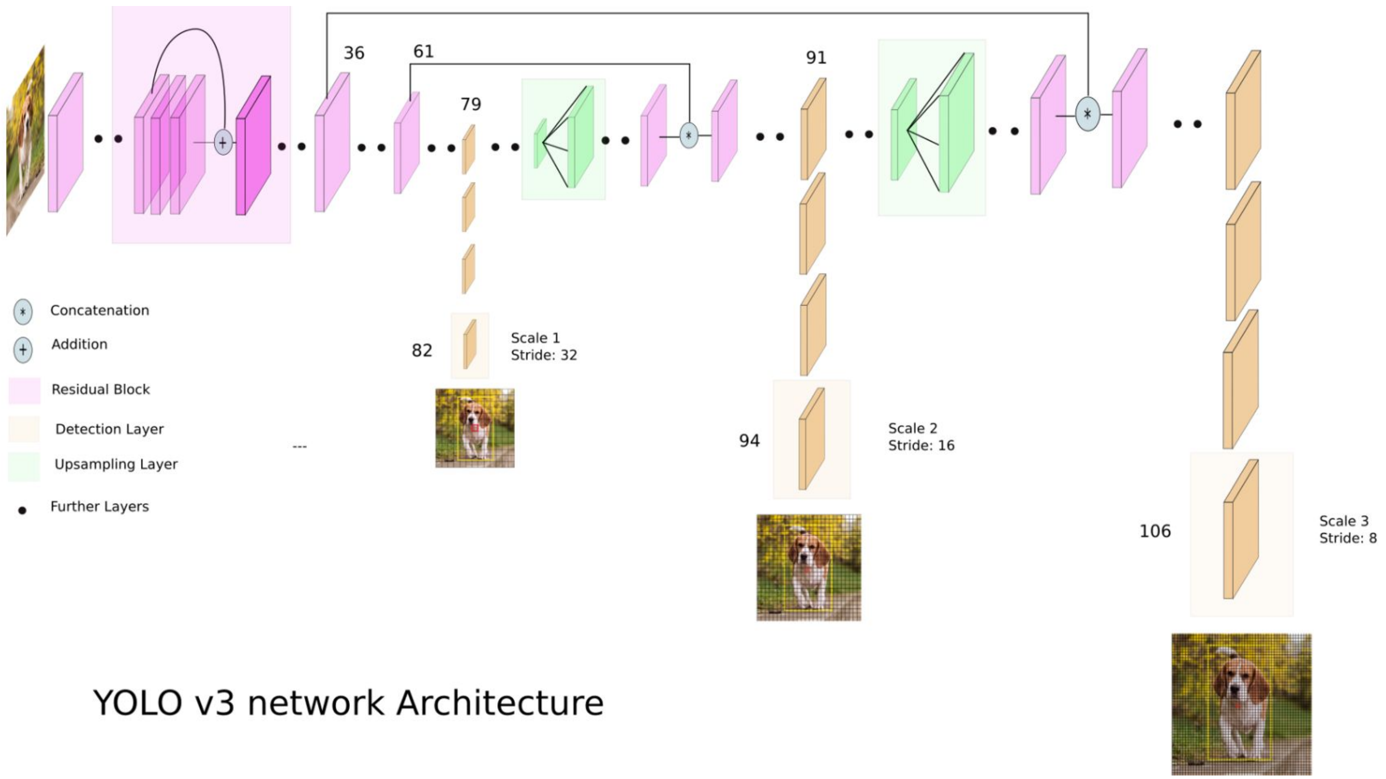
YOLOv2

	YOLO								YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓	✓
anchor boxes?				✓	✓				
new network?					✓	✓	✓	✓	✓
dimension priors?						✓	✓	✓	✓
location prediction?						✓	✓	✓	✓
passthrough?							✓	✓	✓
multi-scale?								✓	✓
hi-res detector?									✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	**78.6**



Redmon & Farhadi. [YOLO900: Better, Faster, Stronger](). CVPR 2017

YOLOv3

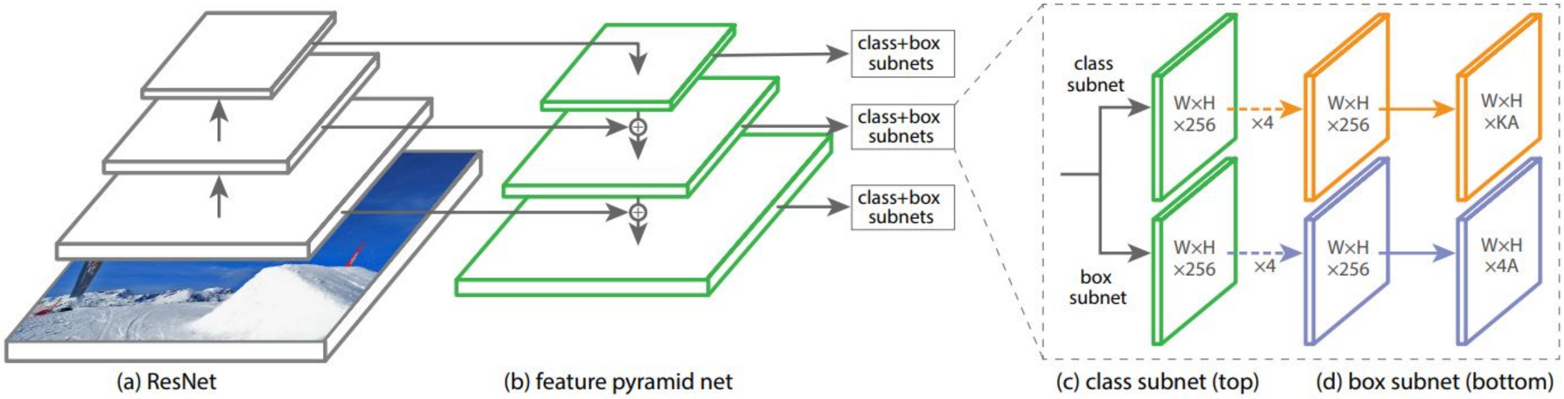


YOLO v3 network Architecture

YOLO v2
- \+ residual blocks
- \+ skip connections
- \+ upsampling
- \+ detection at multiple scales

RetinaNet

Matching proposal-based performance with a one-stage approach

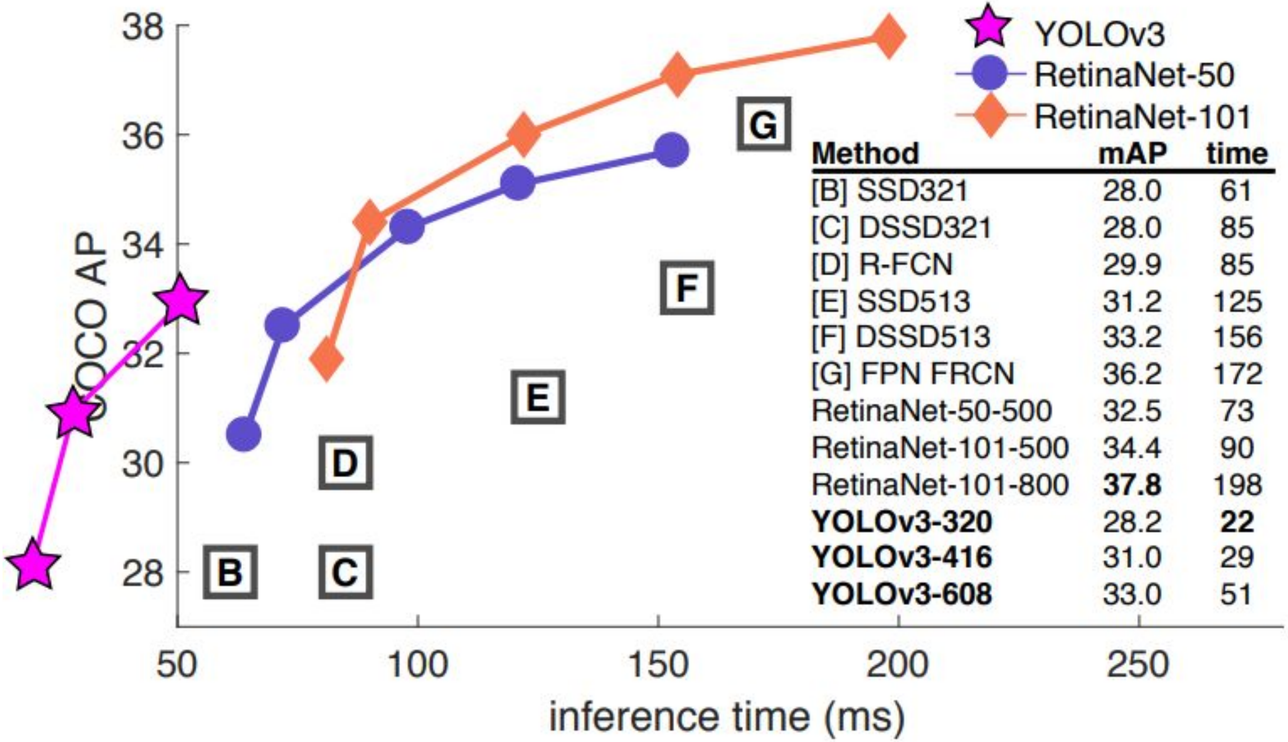


Problem of one-stage detectors? They evaluate many candidate locations but only a few have objects ---> **IMBALANCE**, making learning inefficient

Key idea is to lower loss weight for well classified samples, increase it for difficult ones.

Lin et al. [Focal Loss for Dense Object Detection.](#) ICCV 2017

Overview



Method	mAP	time
[B] SSD321	28.0	61
[C] DSSD321	28.0	85
[D] R-FCN	29.9	85
[E] SSD513	31.2	125
[F] DSSD513	33.2	156
[G] FPN FRCN	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	**37.8**	198
YOLOv3-320	28.2	**22**
YOLOv3-416	31.0	29
YOLOv3-608	33.0	51

Summary

Proposal-based methods

- R-CNN
- Fast R-CNN
- Faster R-CNN
- Mask R-CNN

Proposal-free methods

- YOLO
- SSD
- RetinaNet

Questions?