
Day 3 Lecture 4

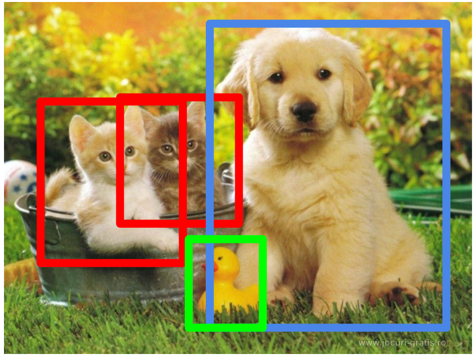
# Object Detection

# Deep ConvNets for Recognition for...

Images (global) Objects (local) Video (2D+T)

Slide Credit: Xavier Giró

# 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

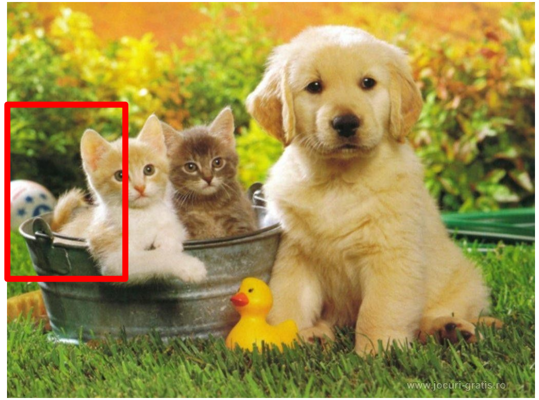

Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

# Object Detection as Classification



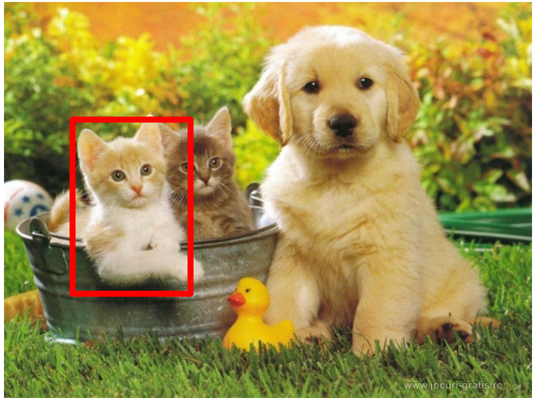

Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

# Object Detection as Classification



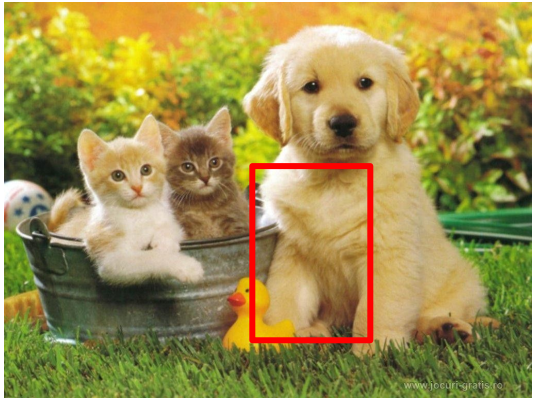

Classes = [cat, dog, duck]

Cat ? YES

Dog ? NO

Duck? NO

# Object Detection as Classification



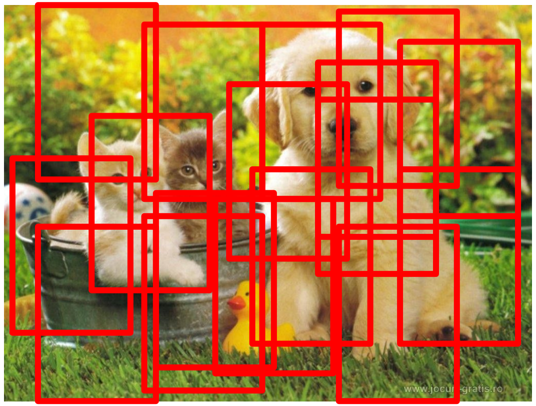

Classes = [cat, dog, duck]

Cat ? NO

Dog ? NO

Duck? NO

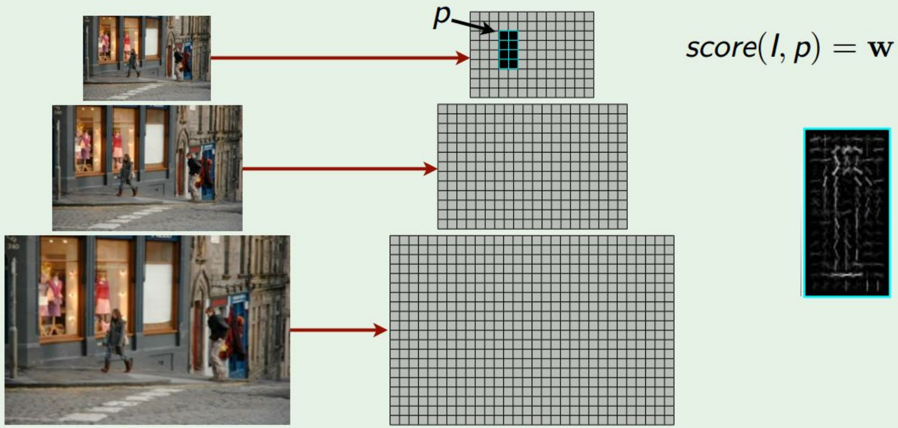
# Object Detection as Classification



Problem:
Too many positions & scales to test

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

# HOG



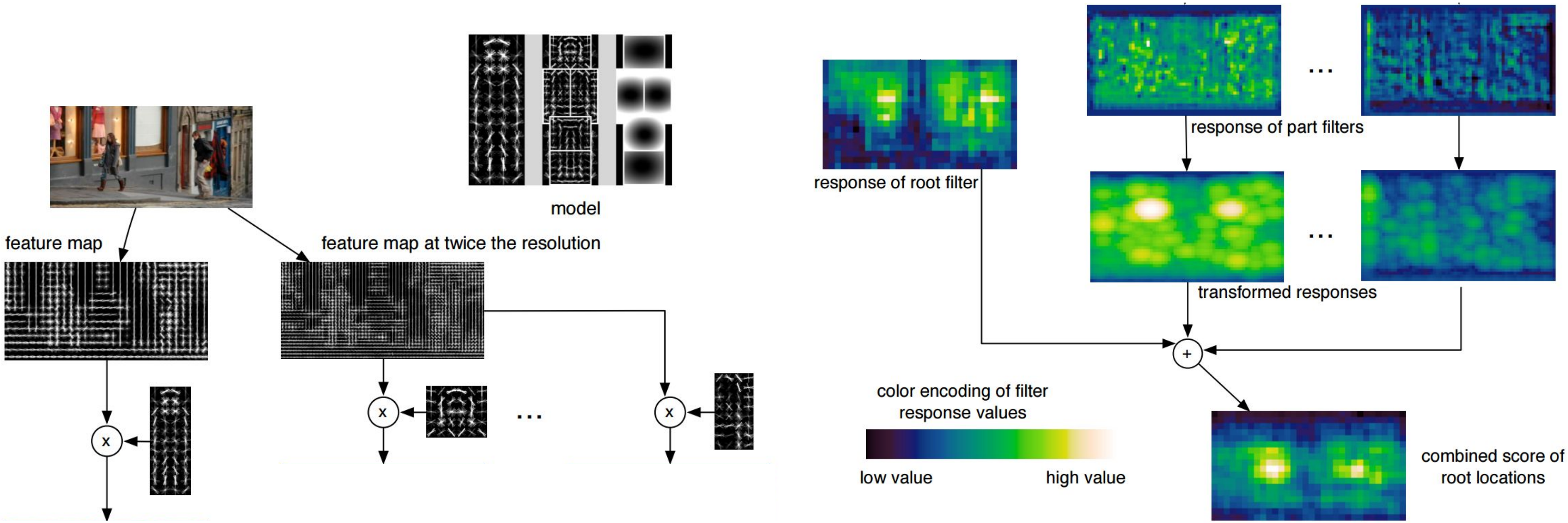

$$score(I, p) = \mathbf{w} \cdot \phi(I, p)$$

Image pyramid HOG feature pyramid

- Compute HOG of the whole image at multiple resolutions
- Score every subwindow of the feature pyramid
- Apply non-maxima suppression

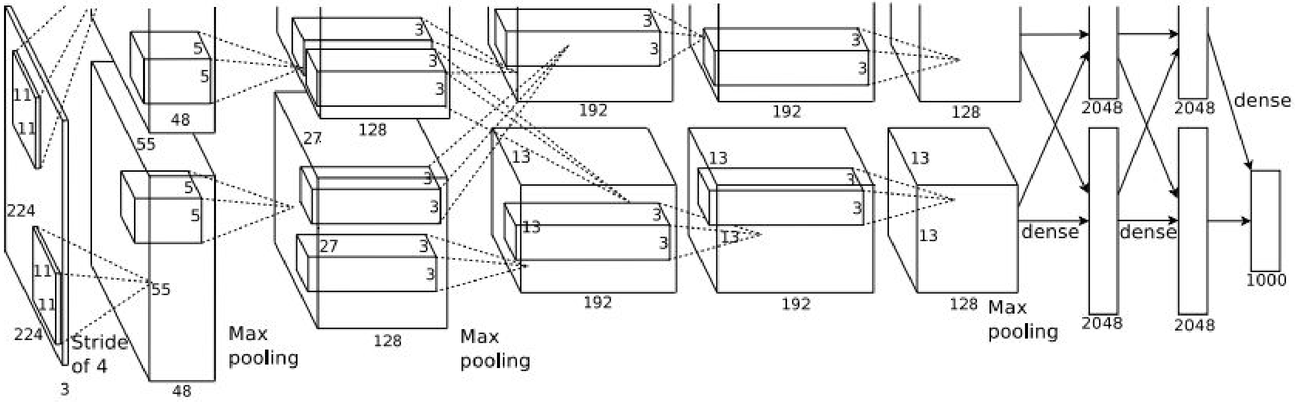
Dalal and Triggs. [Histograms of Oriented Gradients for Human Detection](). CVPR 2005

# Deformable Part Model




Felzenszwalb et al, [Object Detection with Discriminatively Trained Part Based Models](), PAMI 2010

# Object Detection with CNNs?

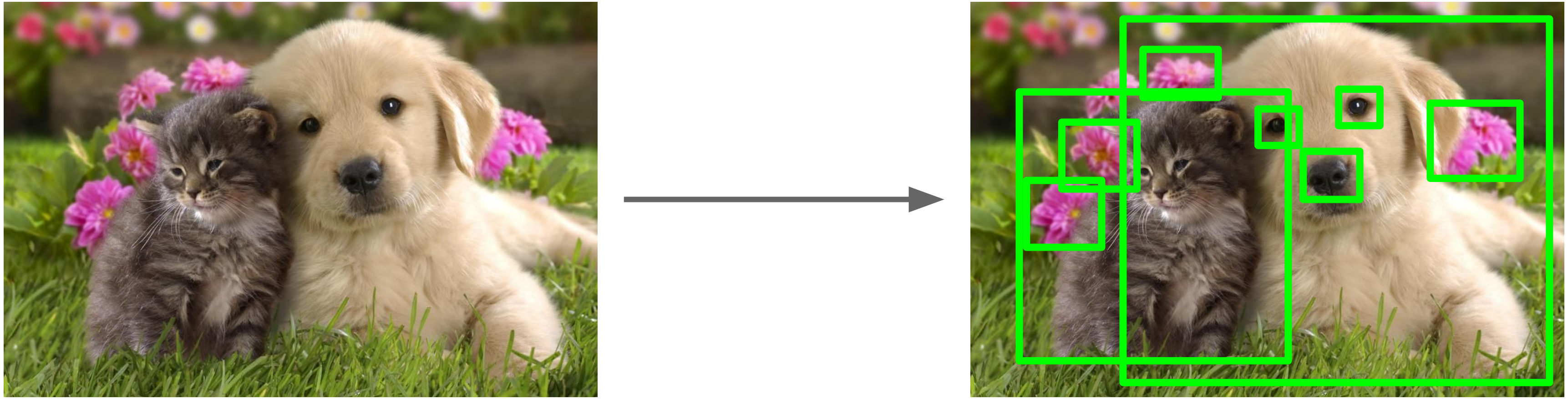



CNN classifiers are computationally demanding. We can't test all positions & scales !

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

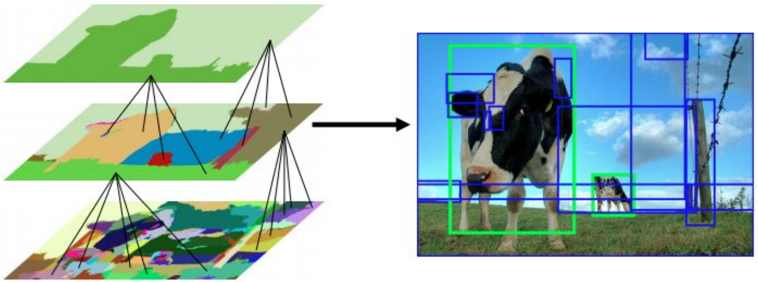
# Region Proposals

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

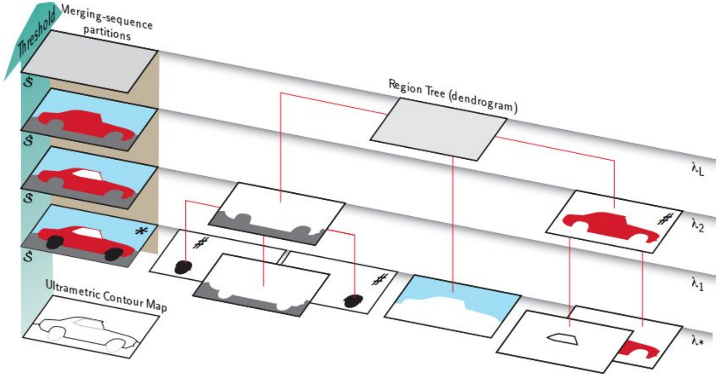


Slide Credit: [CS231n]

# Region Proposals



Selective Search (SS)

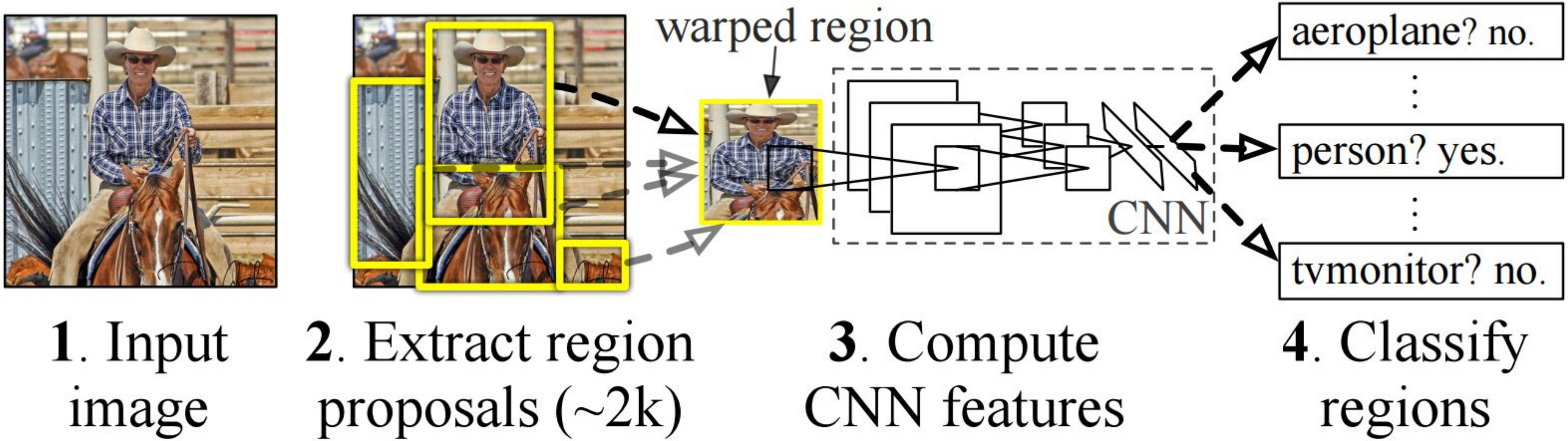



Multiscale Combinatorial Grouping (MCG)

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

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

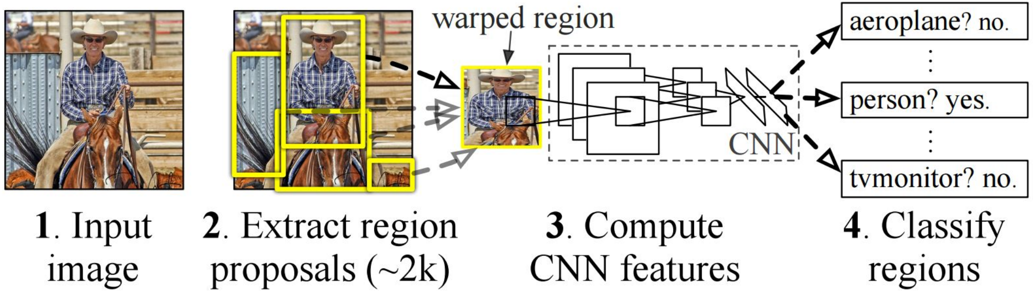
# Object Detection with CNNs: R-CNN




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

# R-CNN

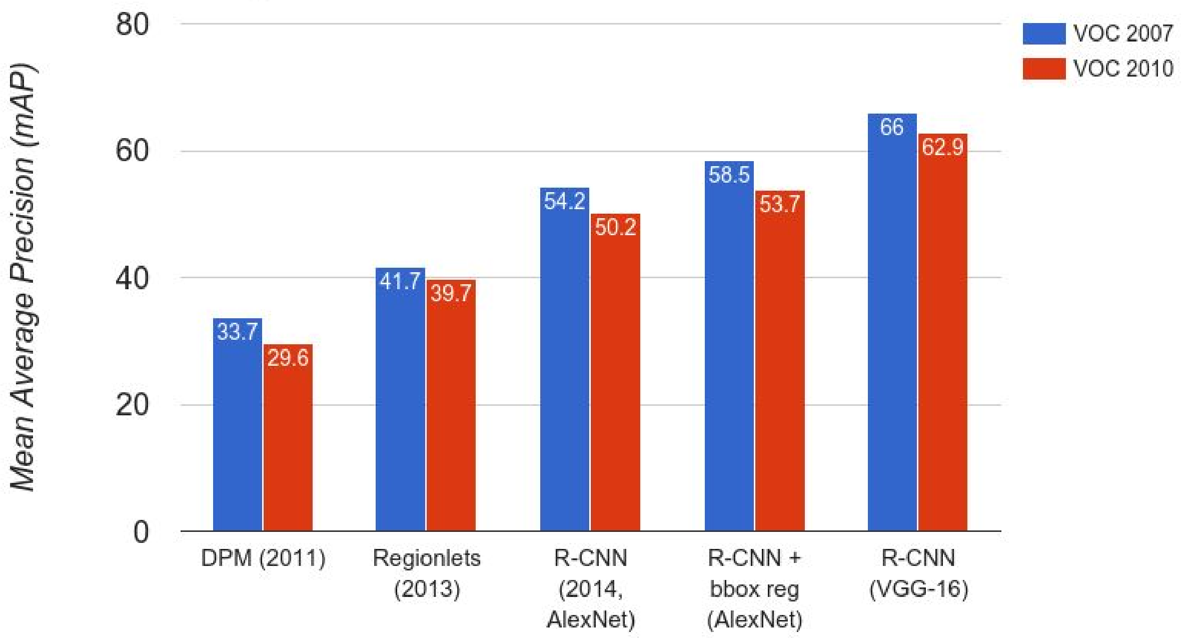
1. Train network on proposals




2. Post-hoc training of SVMs & Box regressors on fc7 features

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

# 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

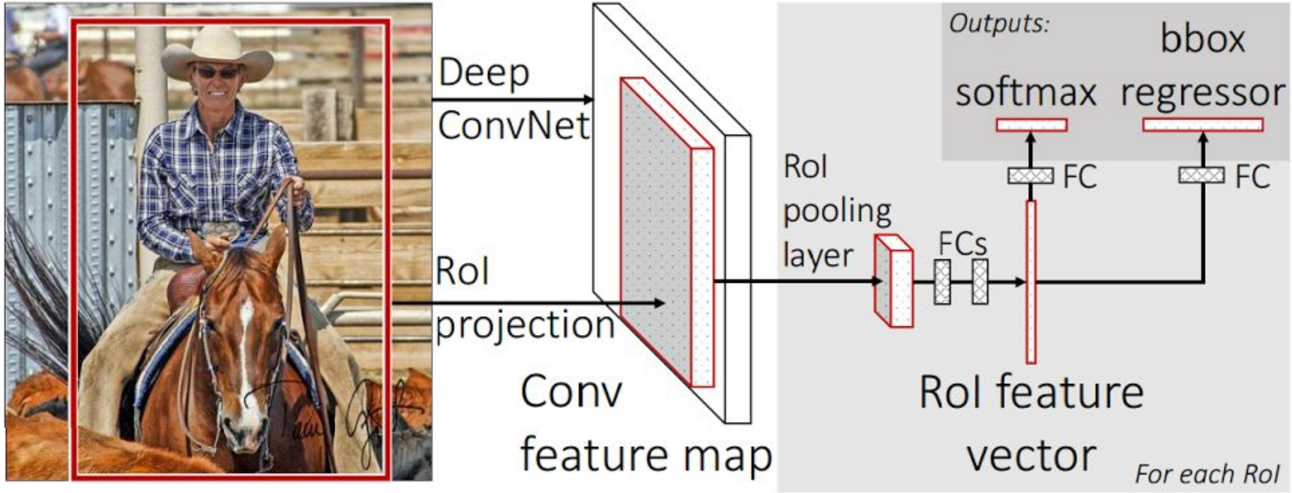
2. SVMs and regressors are post-hoc: CNN features not updated in response to SVMs and regressors

3. Complex multistage training pipeline

Slide Credit: [CS231n]

# Fast R-CNN

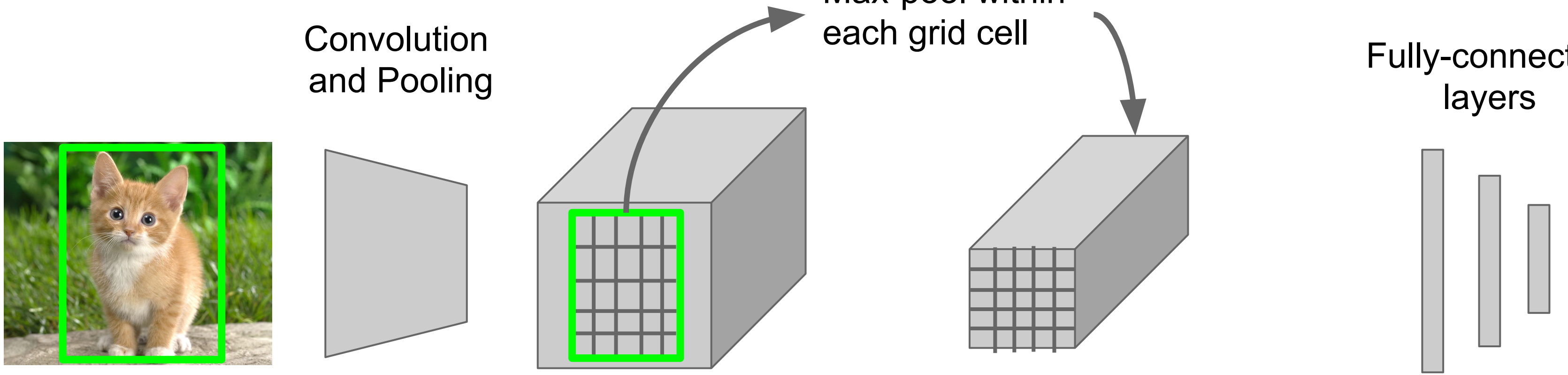
R-CNN Problem #1: Slow at test-time: need to run full forward pass of CNN for each region proposal




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

Girshick [Fast R-CNN](). ICCV 2015

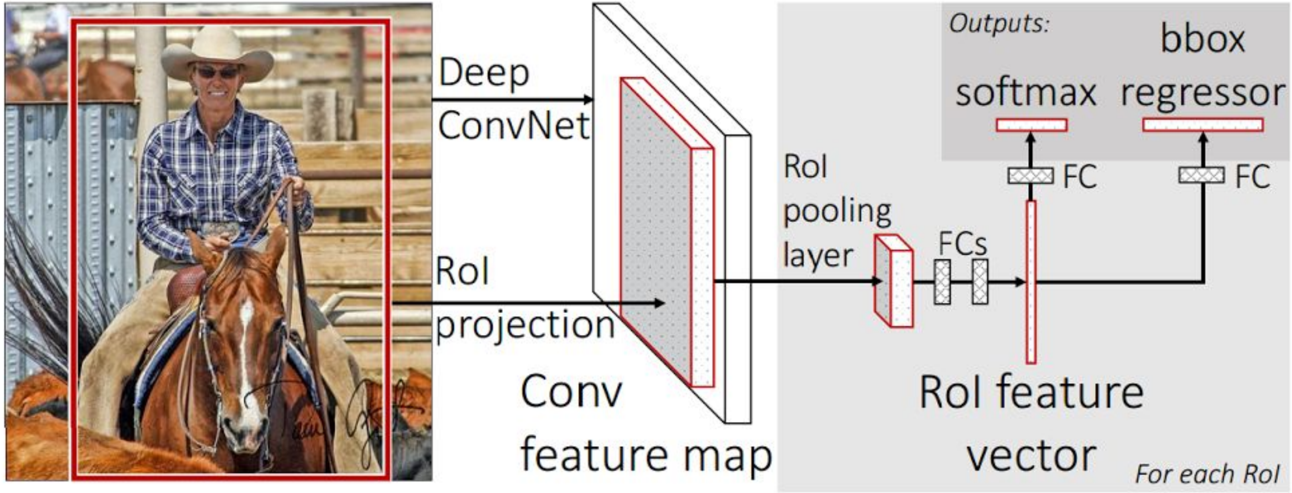
# Fast R-CNN




Girshick [Fast R-CNN](). ICCV 2015

# Fast R-CNN

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




Solution: Train it all at together E2E

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]

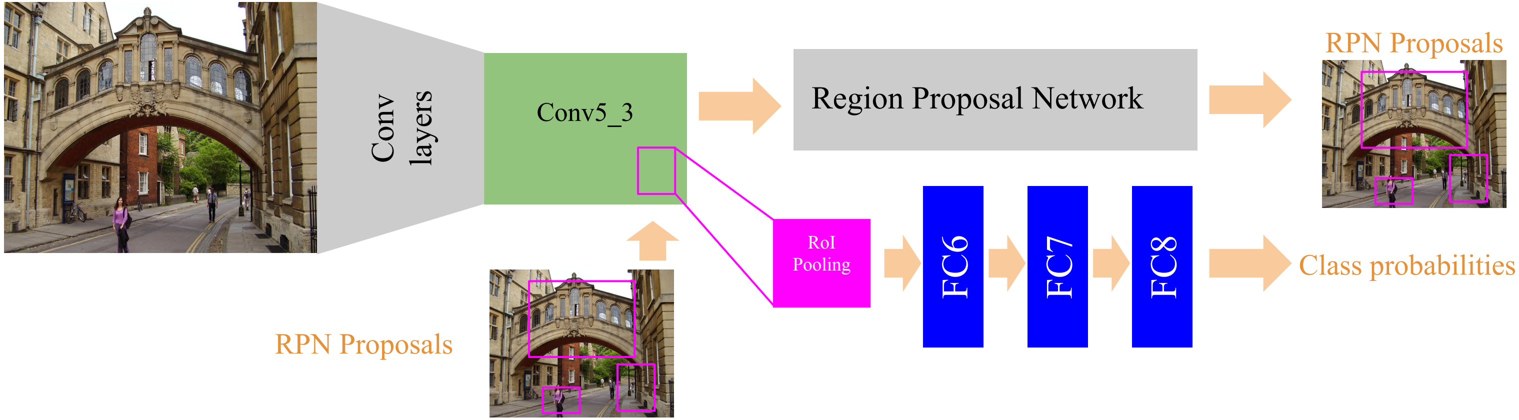
# Fast R-CNN: Problem

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** |

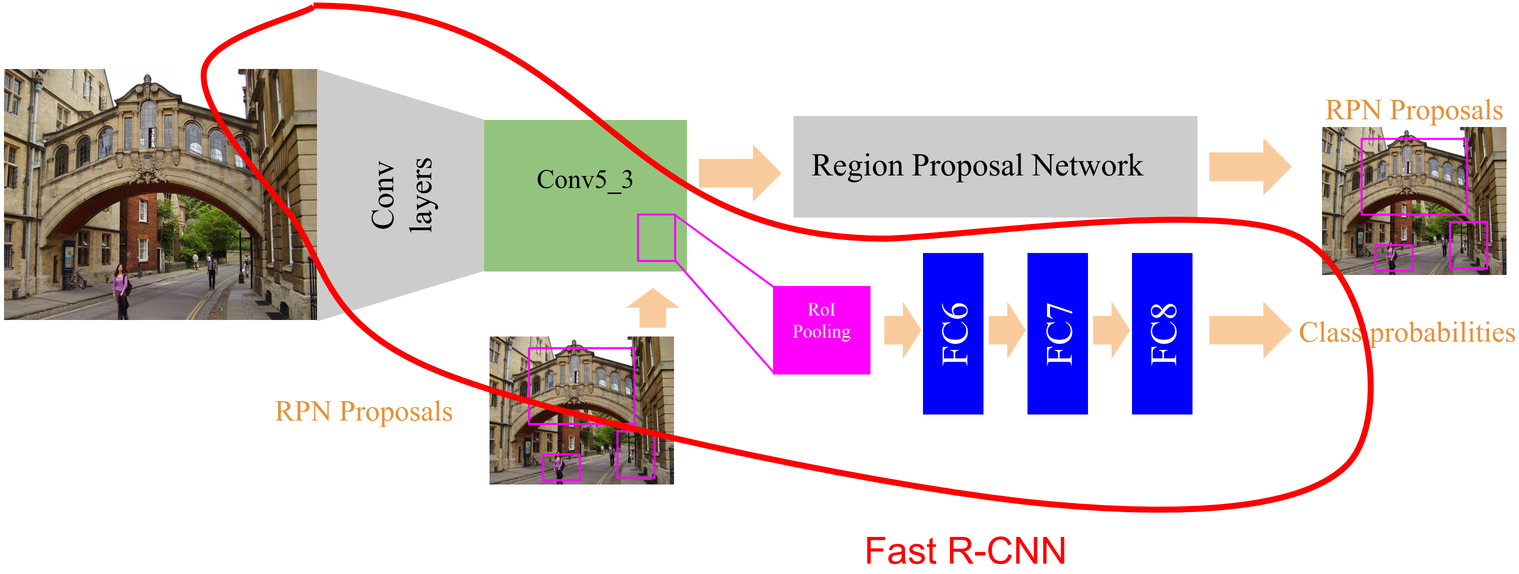
Slide Credit: CS231n

# Faster R-CNN




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

# Faster R-CNN

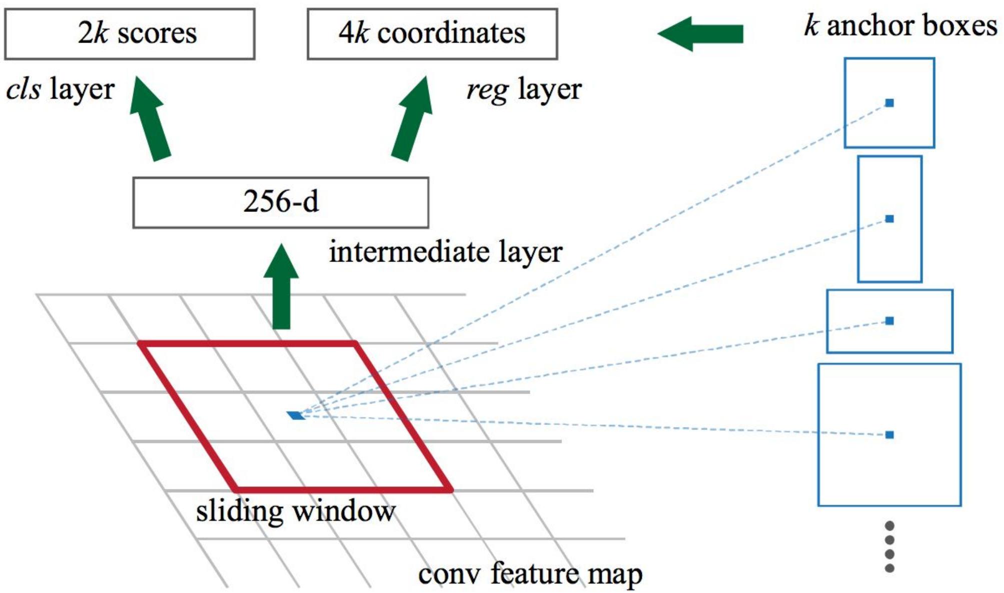



Ren et al. [Faster R-CNN: Towards real-time object detection with region proposal networks.]() NIPS 2015

# 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

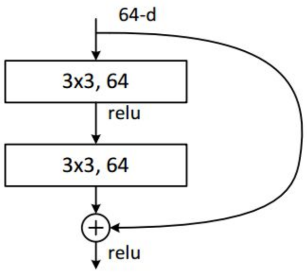
| | **R-CNN** | **Fast R-CNN** | **Faster R-CNN** |
|---|---|---|---|
| Test time per image (with proposals) | 50 seconds | 2 seconds | **0.2 seconds** |
| (Speedup) | 1x | 25x | **250x** |
| mAP (VOC 2007) | 66.0 | **66.9** | **66.9** |

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

Slide Credit: CS231n

# Faster R-CNN

- Faster R-CNN is the basis of the winners of COCO and ILSVRC 2015 object detection competitions.




He et al. [Deep residual learning for image recognition](). arXiv 2015

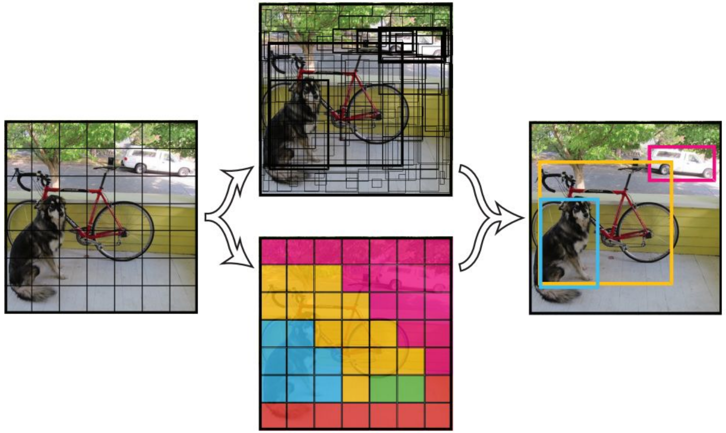
# YOLO: You Only Look Once

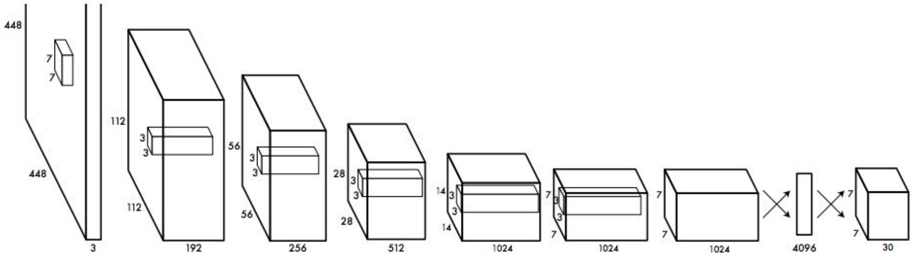
Divide image into S x S grid

Within each grid cell predict:
B Boxes: 4 coordinates + confidence
Class scores: C numbers

Regression from image to
7 x 7 x (5 * B + C) tensor

Direct prediction using a CNN

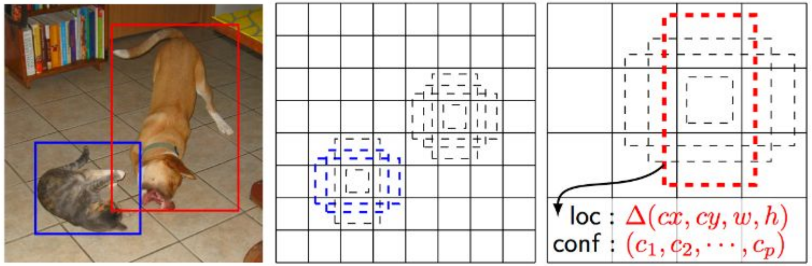




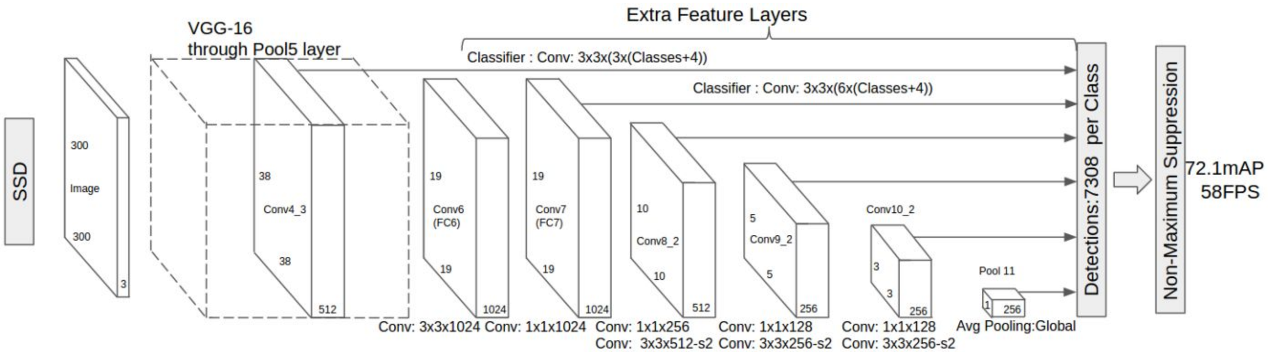

Redmon et al. You Only Look Once: Unified, Real-Time Object Detection, CVPR 2016

Slide Credit: CS231n

# SSD: Single Shot MultiBox Detector




(a) Image with GT boxes (b) $8 \times 8$ feature map (c) $4 \times 4$ feature map




Liu et al. [SSD: Single Shot MultiBox Detector](), arXiv 2015

# SSD: Single Shot MultiBox Detector

| System | VOC2007 test *mAP* | FPS (Titan X) | Number of Boxes |
|---|---|---|---|
| Faster R-CNN (VGG16) | 73.2 | 7 | 300 |
| Faster R-CNN (ZF) | 62.1 | 17 | 300 |
| YOLO | 63.4 | 45 | 98 |
| Fast YOLO | 52.7 | 155 | 98 |
| SSD300 (VGG) | 72.1 | 58 | 7308 |
| SSD300 (VGG, cuDNN v5) | 72.1 | 72 | 7308 |
| SSD500 (VGG16) | 75.1 | 23 | 20097 |

Training with Pascal VOC 07+12

Liu et al. SSD: Single Shot MultiBox Detector, arXiv 2015

# Resources

- Related Lecture from CS231n @ Stanford [slides][video]
- Caffe Code for:
  - R-CNN
  - Fast R-CNN
  - Faster R-CNN [matlab][python]
- YOLO
  - Original (Darknet)
  - Tensorflow
  - Keras
- SSD (Caffe)