

Computer Vision with CNN

Behnia - Heydari

Amirkabir University of Technology

May 21, 2019

Overview

- 1 Motivation
- 2 ImageNet Challenge
- 3 From Cat's Brain to ResNet (Classification)
- 4 Localization
- 5 Object Detection

Cameras Everywhere

Smartphones

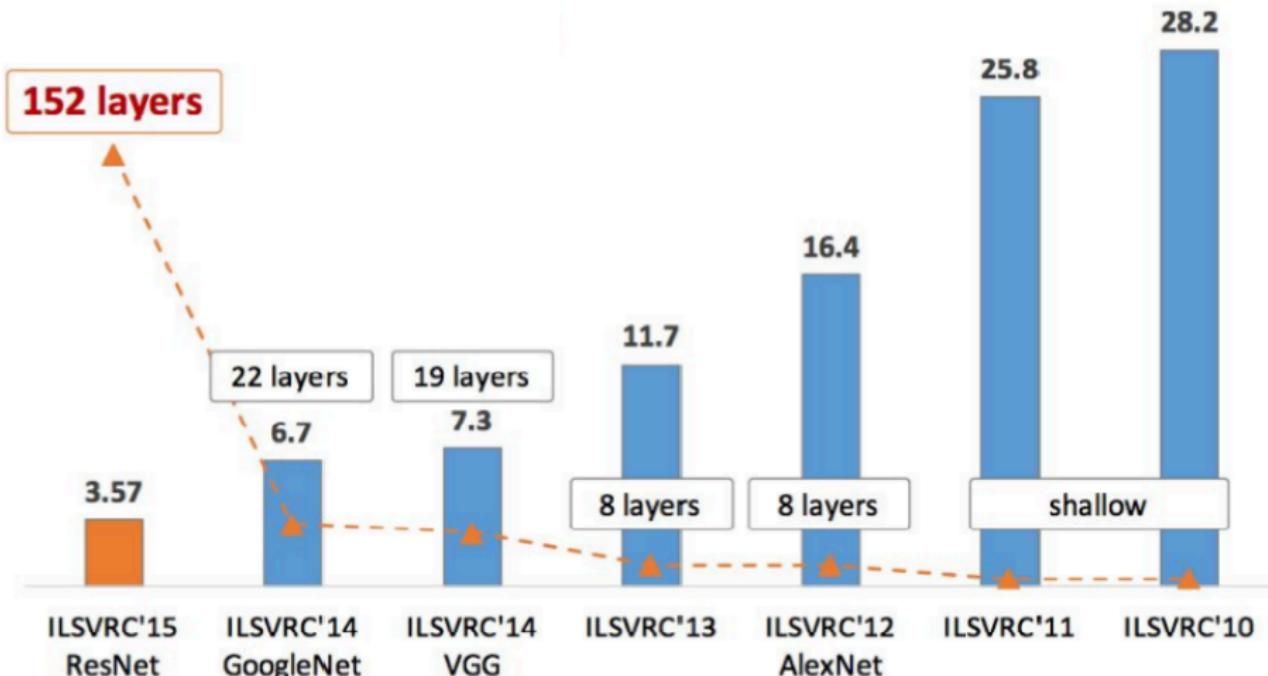
- Exploding number of sensors vs. humans



Neat Dataset



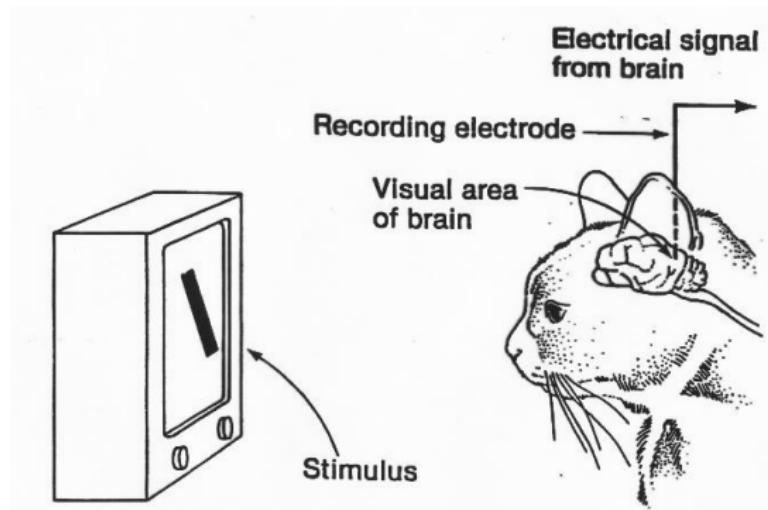
Role of CNN



Cat's Brain

Types of cells:

- simple cells
- complex cells
- hypercomplex cells



AlexNet

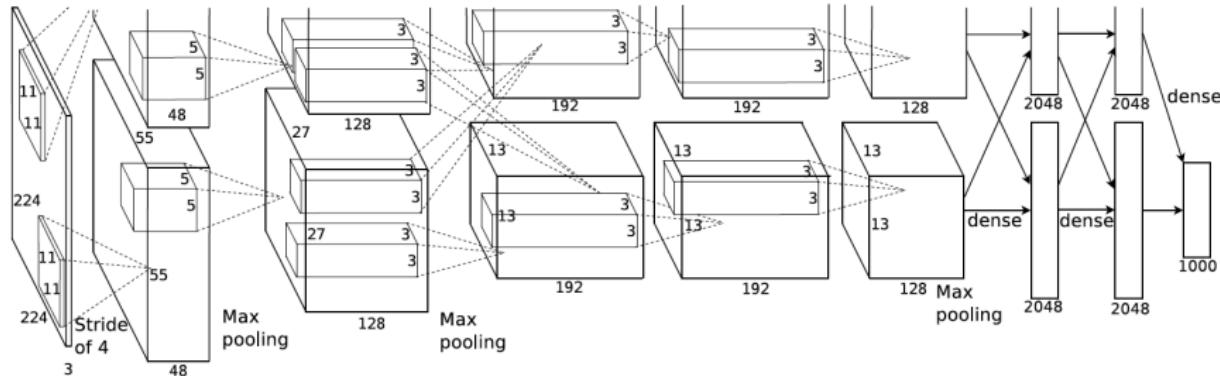


Figure: AlexNet: [Krizhevsky, Sutskever, Hinton] 2012

Convolution

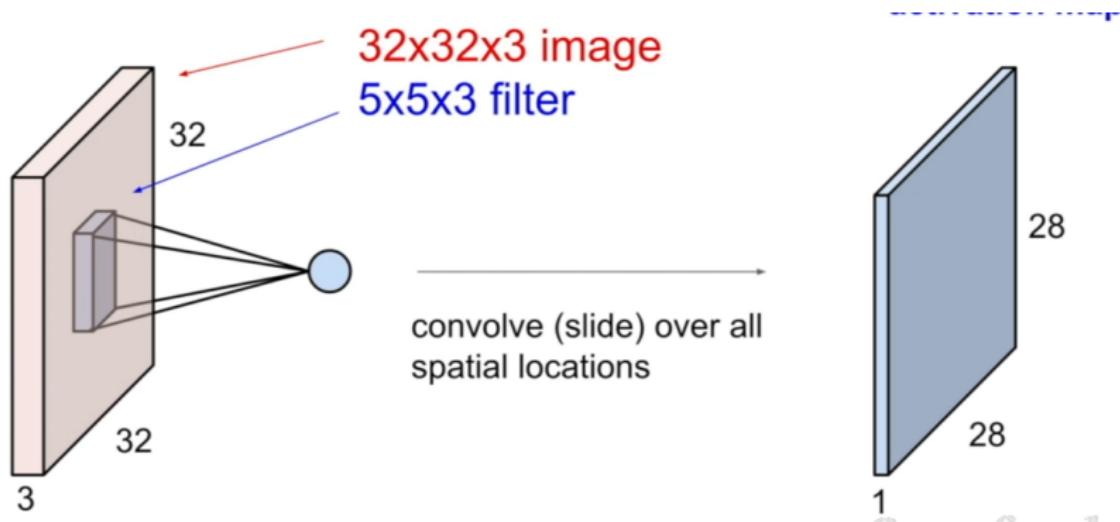


Figure: Convolution Layer

Complexity of Features

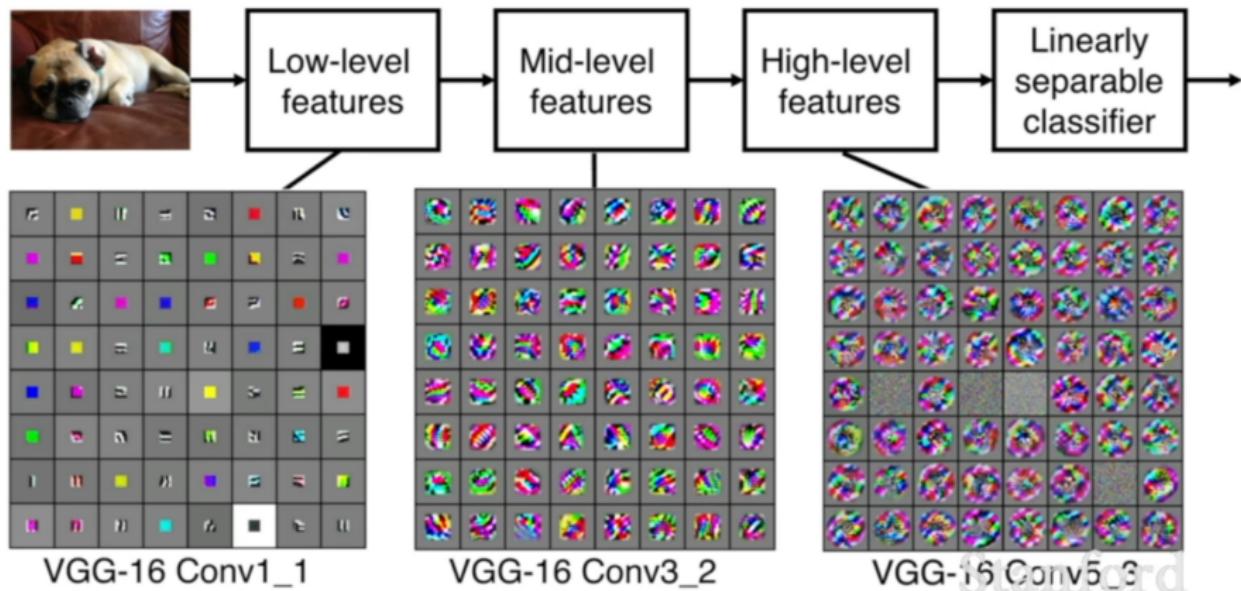


Figure: complexity of features in each layer

Sliding a Filter

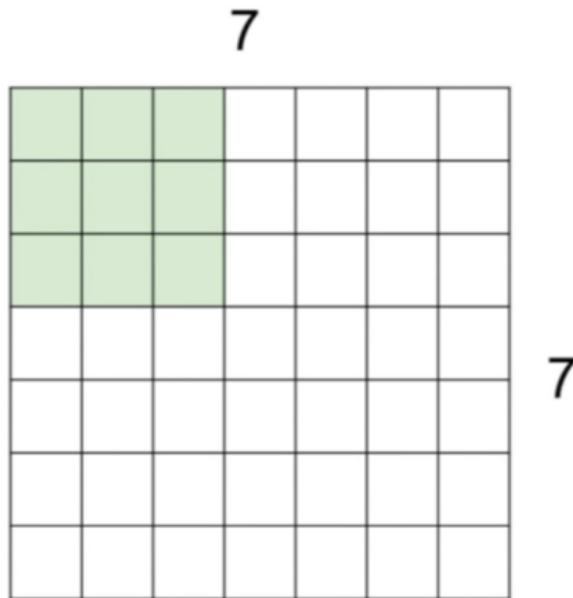


Figure: filtering example

Sliding a Filter (Cont.)

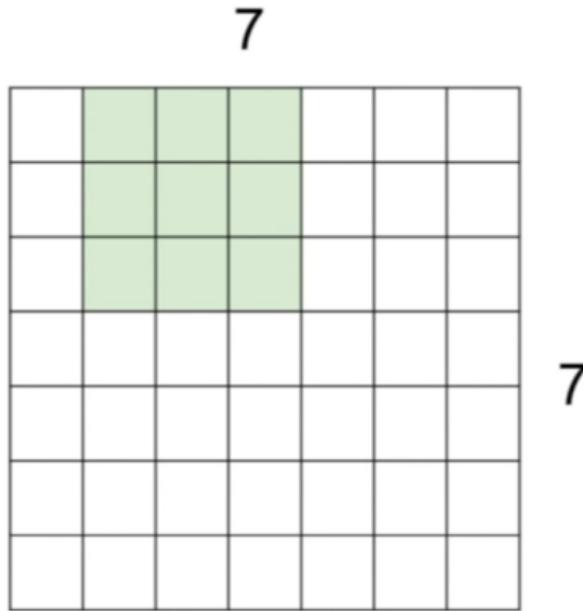


Figure: filtering example

Sliding a Filter (Cont.)

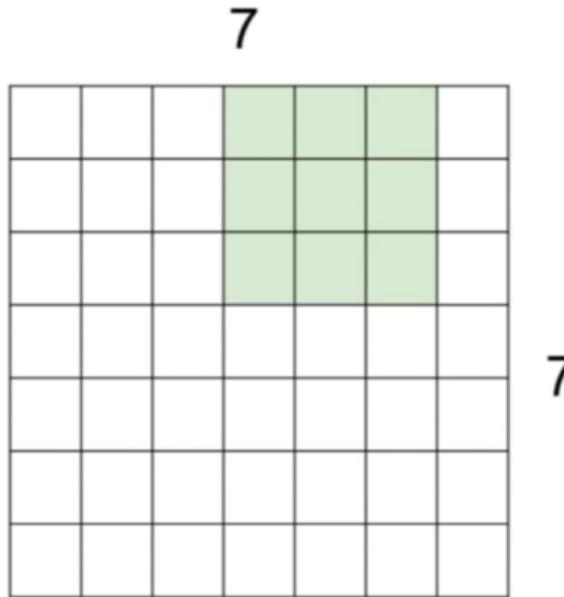


Figure: filtering example

Sliding a Filter (Cont.)

What if using a 3×3 filter with stride 3 ?!

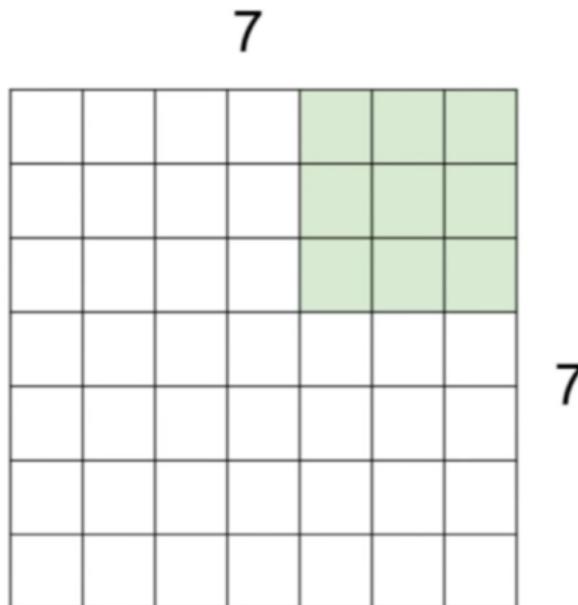


Figure: filtering example

Padding

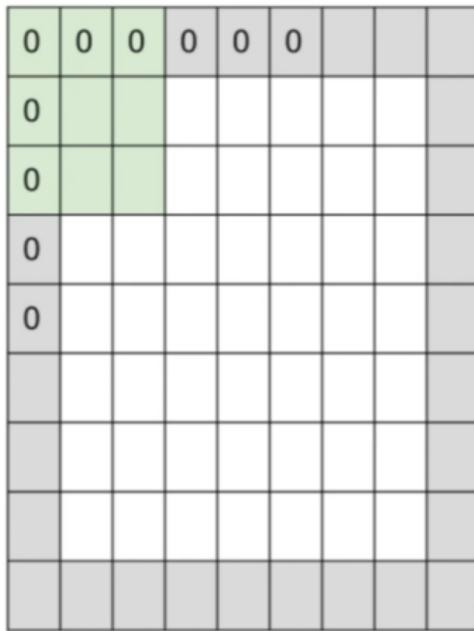


Figure: Padding: control output size

Summary

- Accept a volume of size $N \times N$
- requires 4 hyper parameters:
 - Filter's spatial extent F
 - Filter's Stride S
 - Amount of zero padding P
 - Number of kernels ($\times depth \times$)
- Produces a volume of size $M \times M$
 - $M = \frac{(N-F+2p)}{S} + 1$
 - Number of parameters : $(F \times F \times D + b) \times k$

Pooling

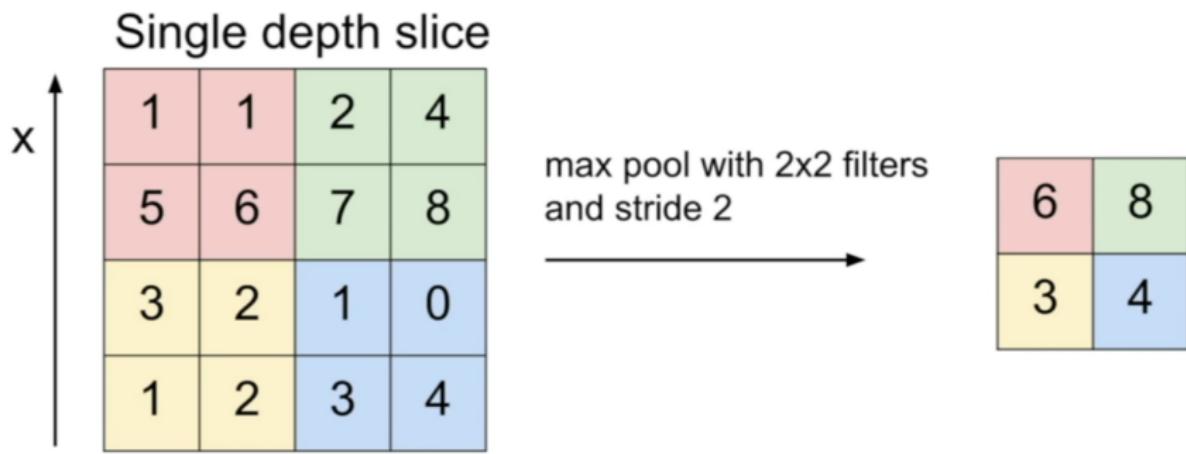


Figure: Max-pooling

Fully Connected Layer

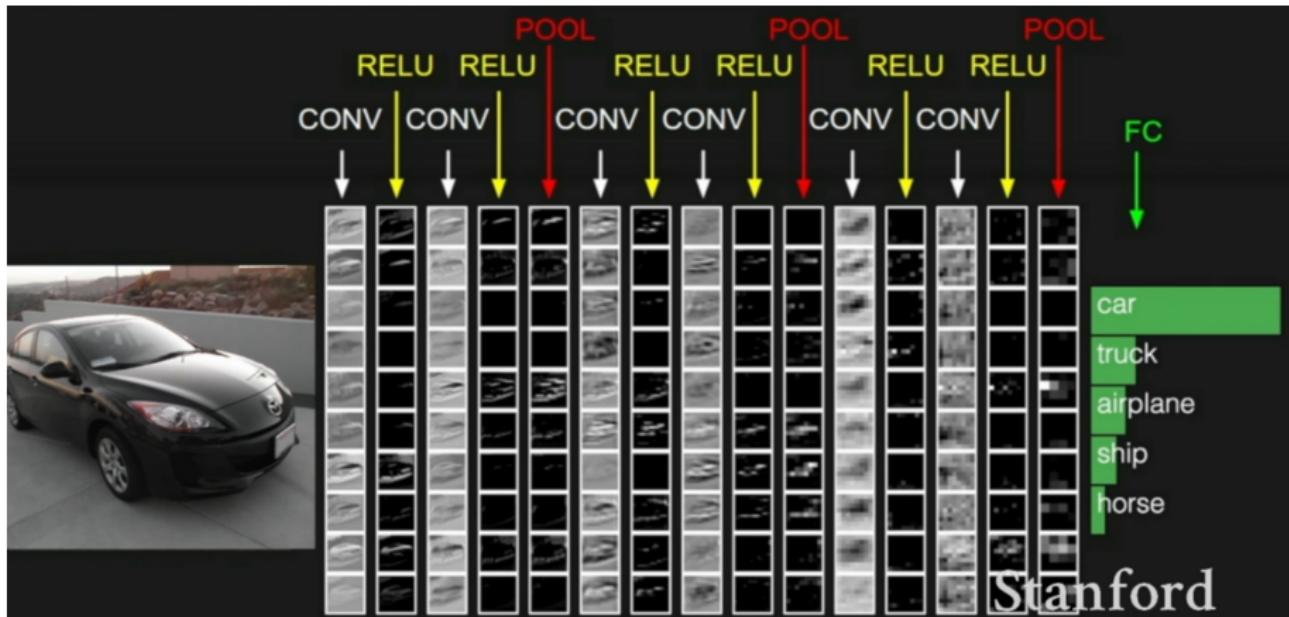
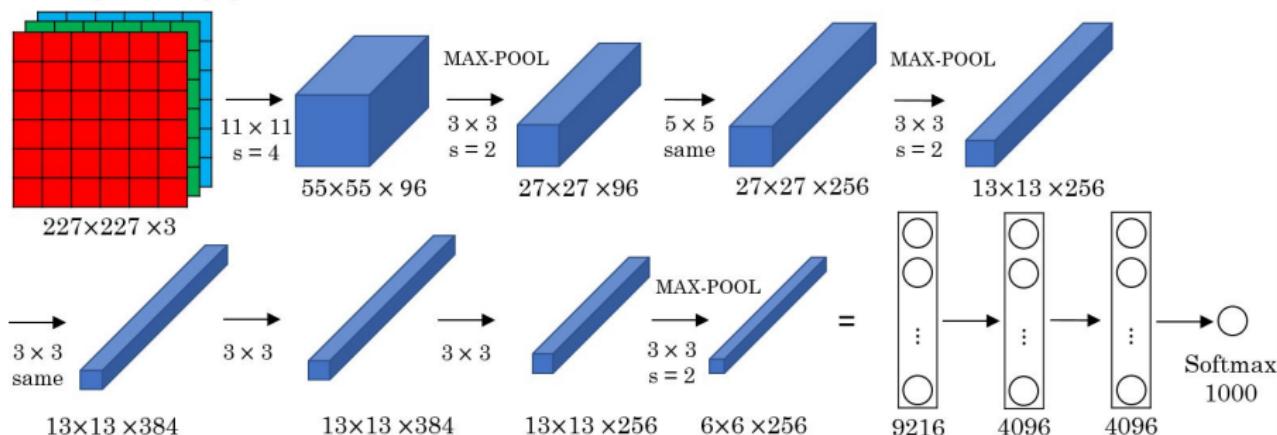


Figure: Fully Connected Layer

AlexNet

AlexNet



Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks

Andrew Ng

Figure: AlexNet architecture

Comparing different Structures

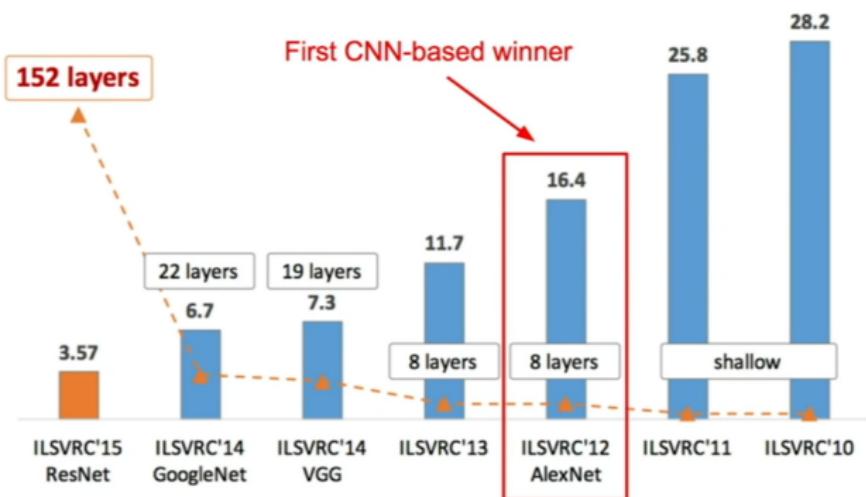
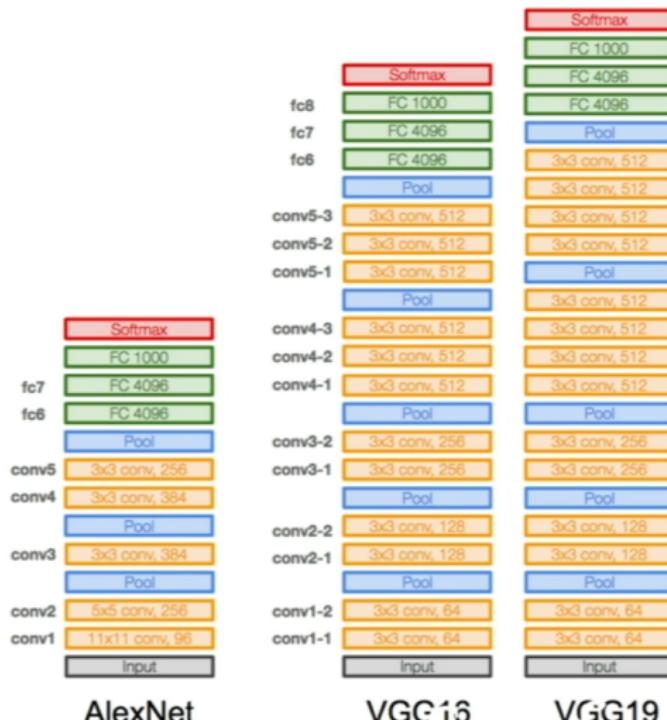


Figure: Comparing Structures

Deeper Networks, Smaller Filters



Comparing

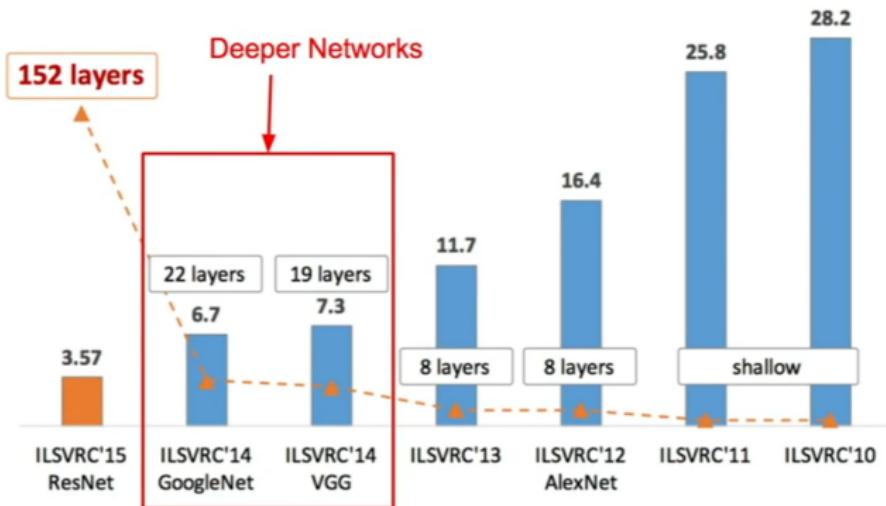
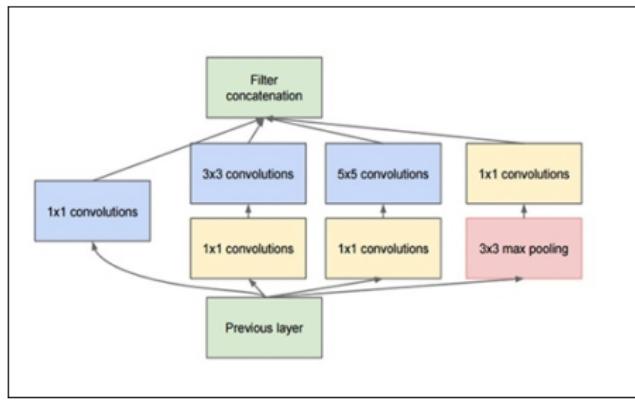
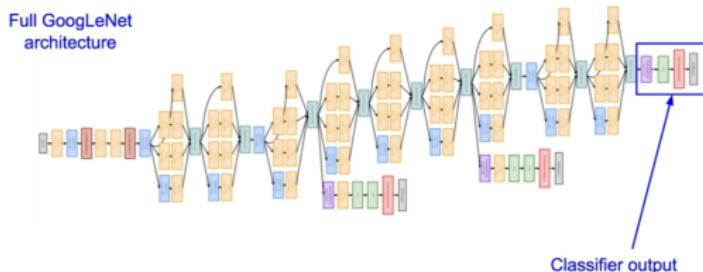


Figure: Comparing Structures

GoogLeNet

Make deeper networks computationally inexpensive



Comparing

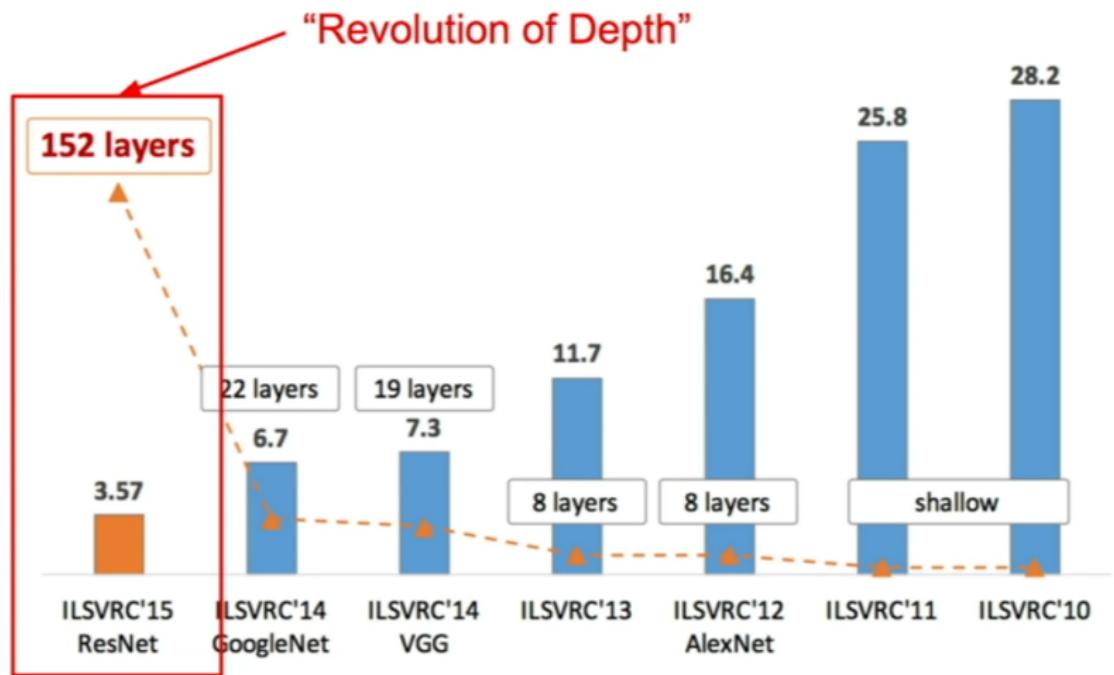


Figure: Comparing Structures

ResNet

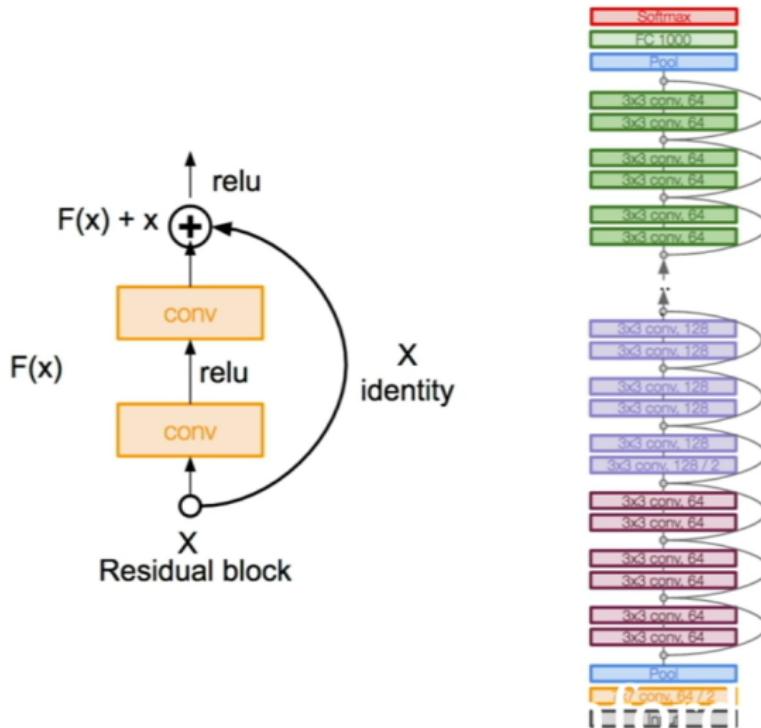


Figure: ResNet Architecture

Comparing Structures

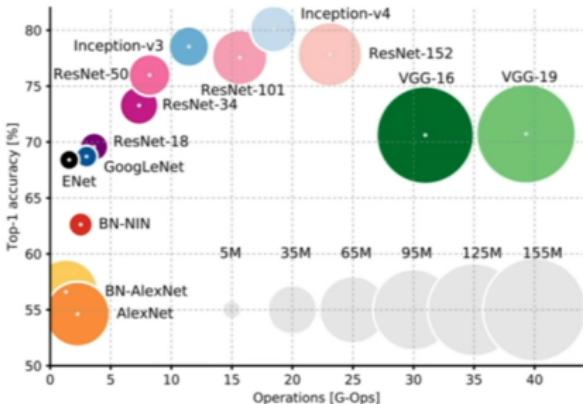
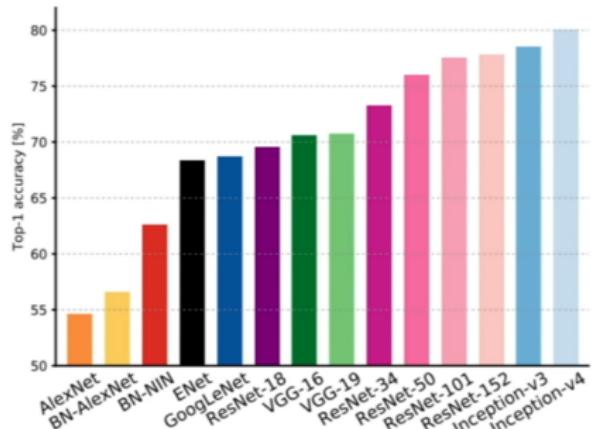


Figure: Comparison

Classification + Localization



GRASS, CAT,
TREE, SKY

No objects, just pixels



CAT

Single Object



DOG, DOG, CAT

Multiple Object

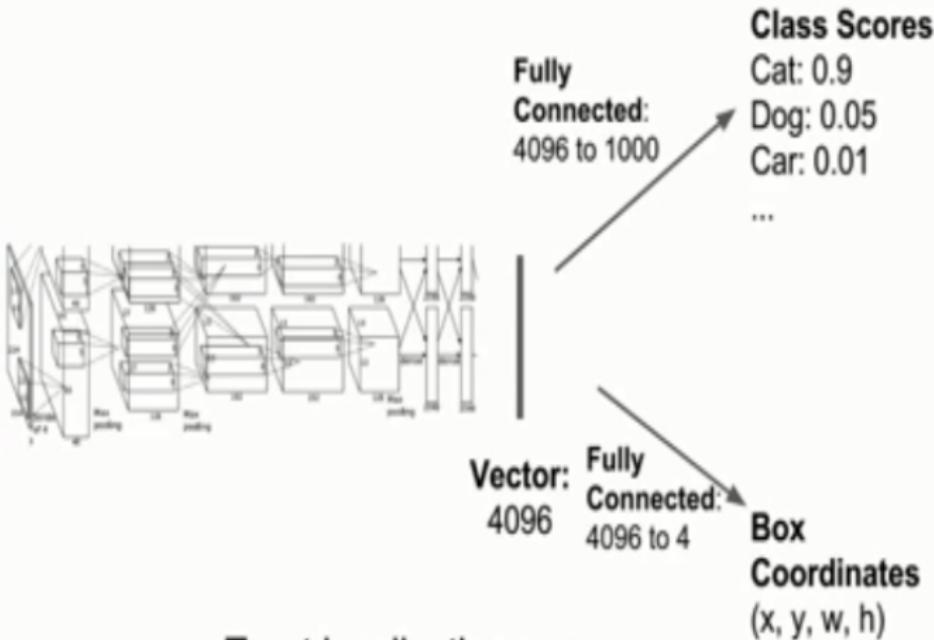


DOG, DOG, CAT

Classification + Localization

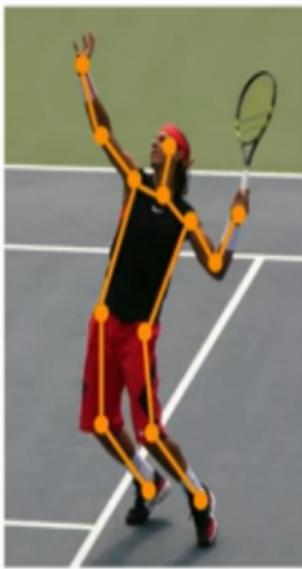


This image is CC0 public domain



Treat localization as a regression problem!

Example



This image is licensed under CC-BY 2.0.

Represent pose as a set of 14 joint positions:

- Left / right foot
- Left / right knee
- Left / right hip
- Left / right shoulder
- Left / right elbow
- Left / right hand
- Neck
- Head top

Figure: Body pose estimation

Fancier Task

Object Detection



GRASS, CAT,
TREE, SKY



CAT



DOG, DOG, CAT

No objects, just pixels

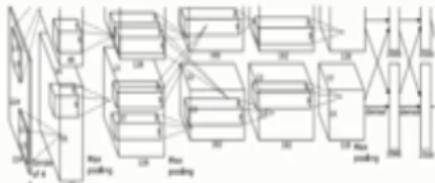
Single Object



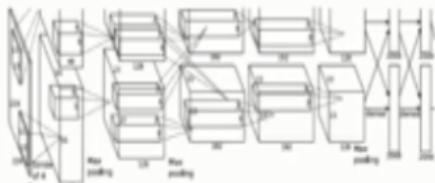
DOG, DOG, CAT

Multiple Object

Object detection



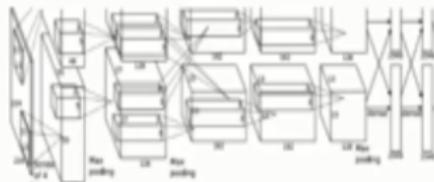
CAT: (x, y, w, h)



DOG: (x, y, w, h)

DOG: (x, y, w, h)

CAT: (x, y, w, h)



DUCK: (x, y, w, h)

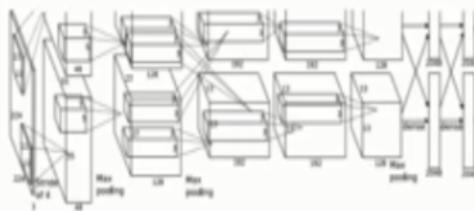
DUCK: (x, y, w, h)

...

Figure: New approach is needed

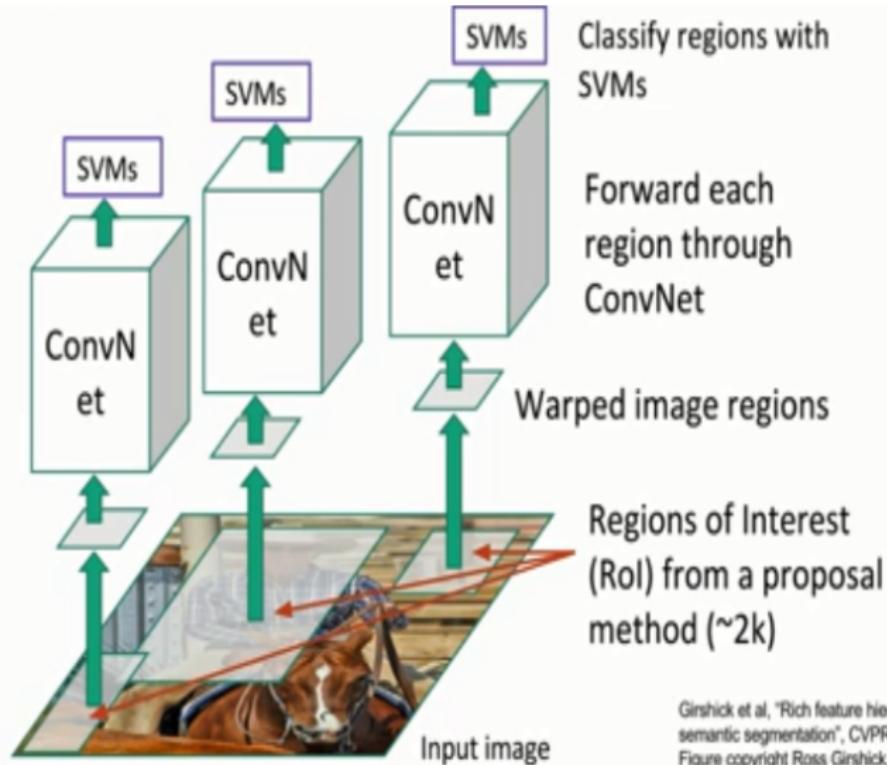
Object detection

Apply a CNN to many different crops of the image, CNN classifies each crop as object or background



Dog? NO
Cat? NO
Background? YES

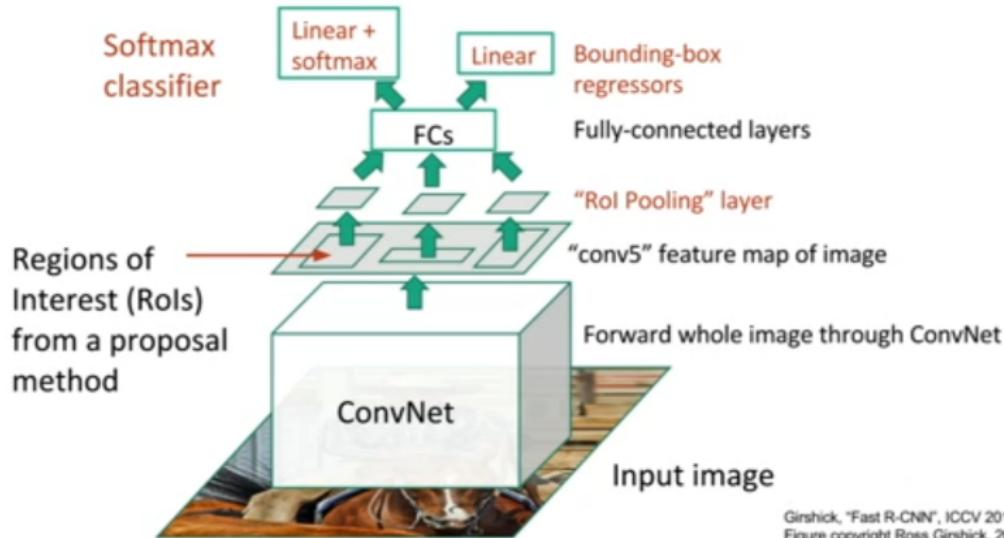
RCNN



Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [http://csail.mit.edu/people/girshick/pubs/cvpr15_rcnn.pdf](#). Reproduced with permission.

Fast RCNN

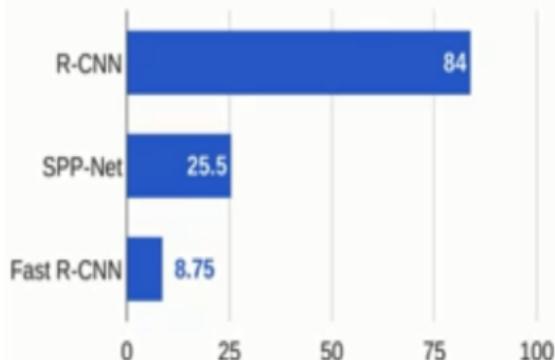
Fast R-CNN



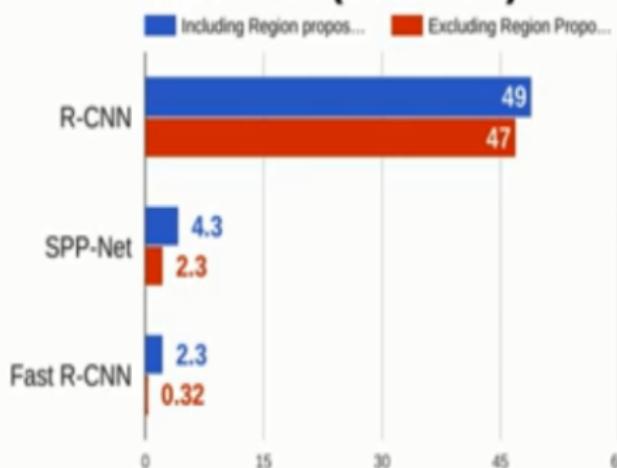
Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015. <http://csail.mit.edu/people/girshick/pubs.html>

Comparing

Training time (Hours)



Test time (seconds)



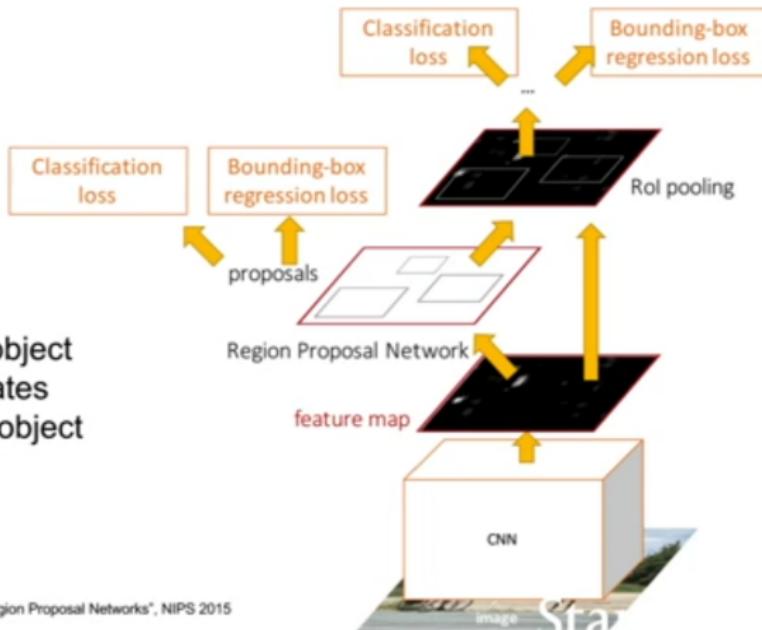
Even Faster

Faster R-CNN: Make CNN do proposals!

Insert **Region Proposal Network (RPN)** to predict proposals from features

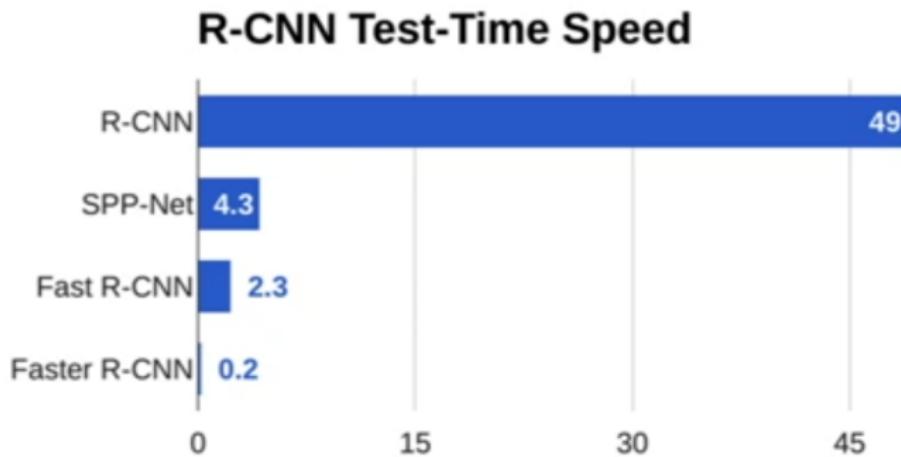
Jointly train with 4 losses:

1. RPN classify object / not object
2. RPN regress box coordinates
3. Final classification score (object classes)
4. Final box coordinates

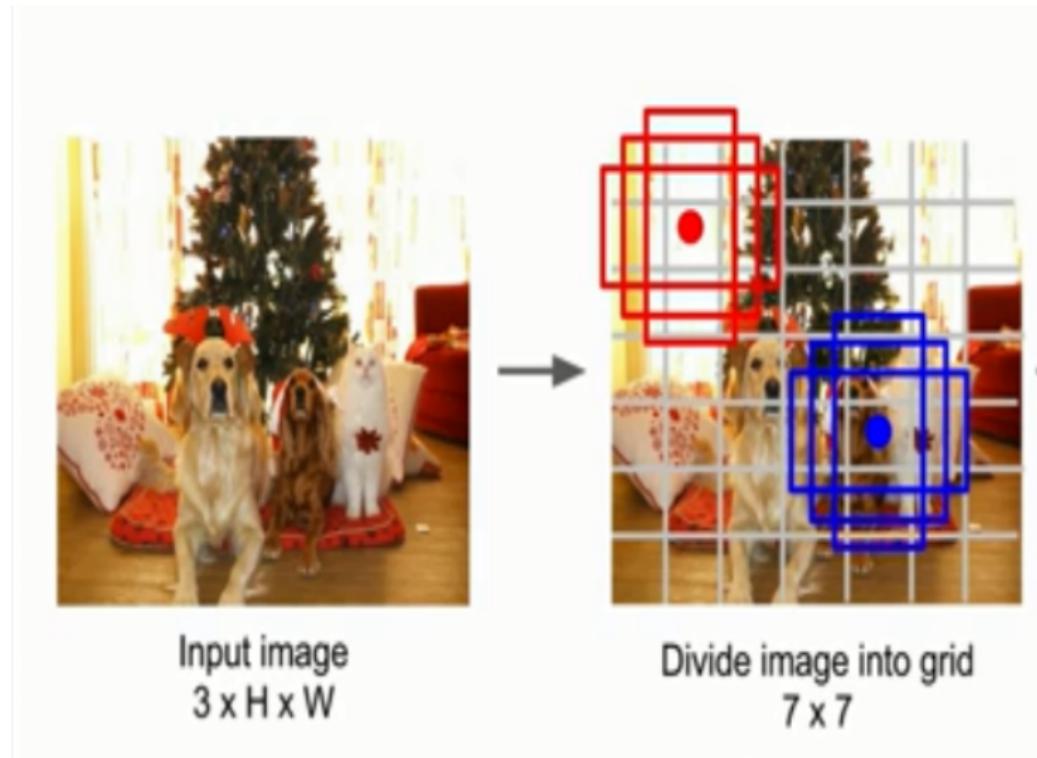


Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NIPS 2015
Figure copyright 2015, Ross Girshick; reproduced with permission

Faster R-CNN: Make CNN do proposals!



Yolo / SSD



References:

- Stanford CS231n

Any Question?