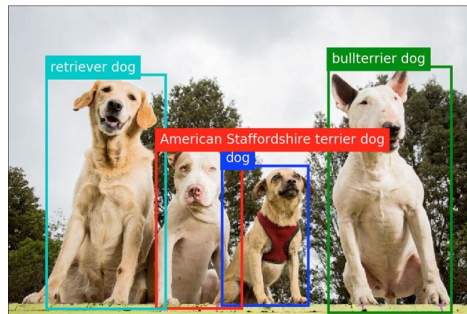
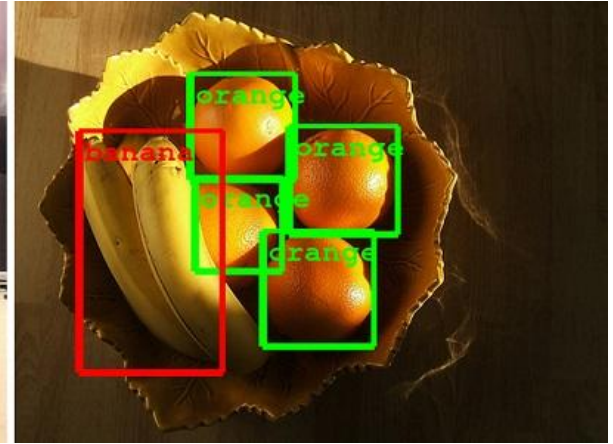
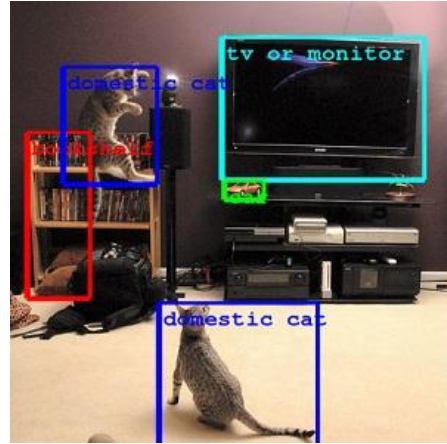
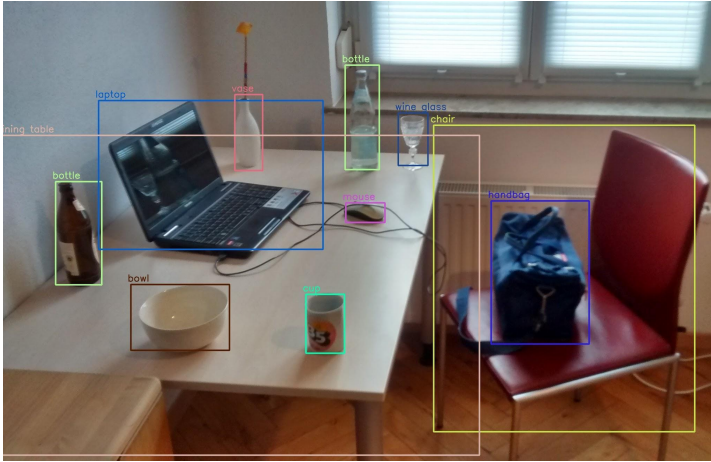


All About Detectors

Muhammad Kamran Janjua

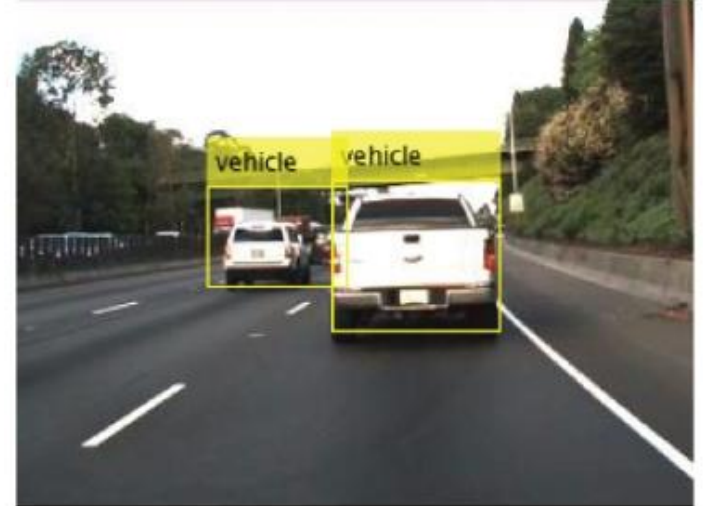


Object Detection



Visual Recognition

Object Detection - What and Why?



1. Recognize/Identify Objects in Scenes - Images/Videos
2. Reason about the information in a quantified manner

The Detection Algorithm

1. Classical Techniques

- a. Feature Detectors + Component Labelling
- b. Binarization + Boundary Detection
- c. Skeleton Detection + Corner Detection

2. ML Techniques

- a. Feature Detectors + Regression
- b. Feature Extraction + Pixel-wise Classification
- c. Boosting (Ensemble Approach)

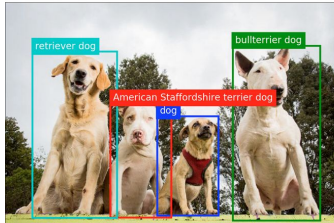
3. DL Techniques

- a. Blackbox - works best

Problem Formulation - 1

Detection as a Regression Problem

Too Many Outputs to Handle [Difficult - too computationally expensive]. The outputs vary for each image.



Object Detection



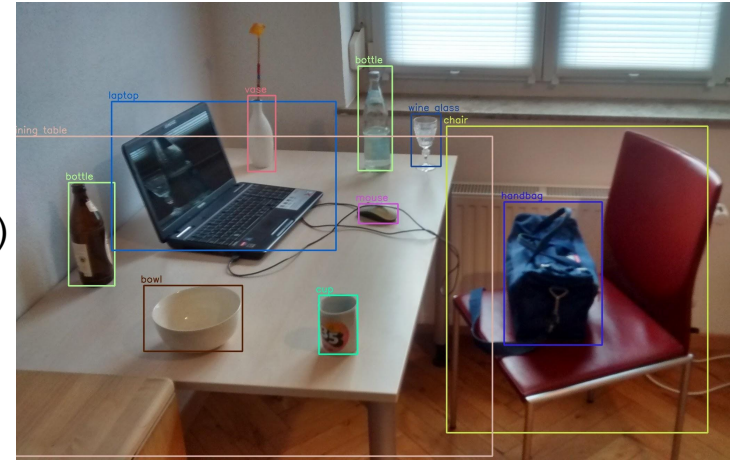
Visual Recognition

(4 DOGs) $4 \times (x,y,w,h)$ - 16 numbers

CHAIR (x, y, w, h)
BOWL (x, y, w, h)
CUP (x, y, w, h)
BAG (x, y, w, h)
MOUSE (x, y, w, h)

.
. .
.

$\ast (x,y,w,h)$

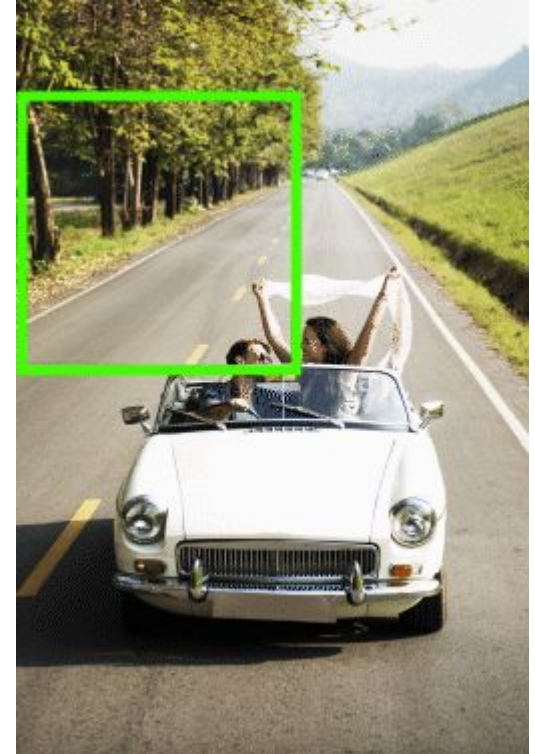


Problem Formulation - 2

Sliding Window - Classification

Apply DNN to a variety of locations, classify each location as object or background. Too computationally expensive.

If sliding window of size $[w', h']$, the image of size $[w, h]$ then locations to apply $[BATCH_SIZE * w.h/w'.h'] +$ classification cost.



Problem Formulation - 3

Region Proposals

Better to find regions most likely to contain the object - ROI and then check if background or object.

Fast to run, reduces the image to mere n blobs.

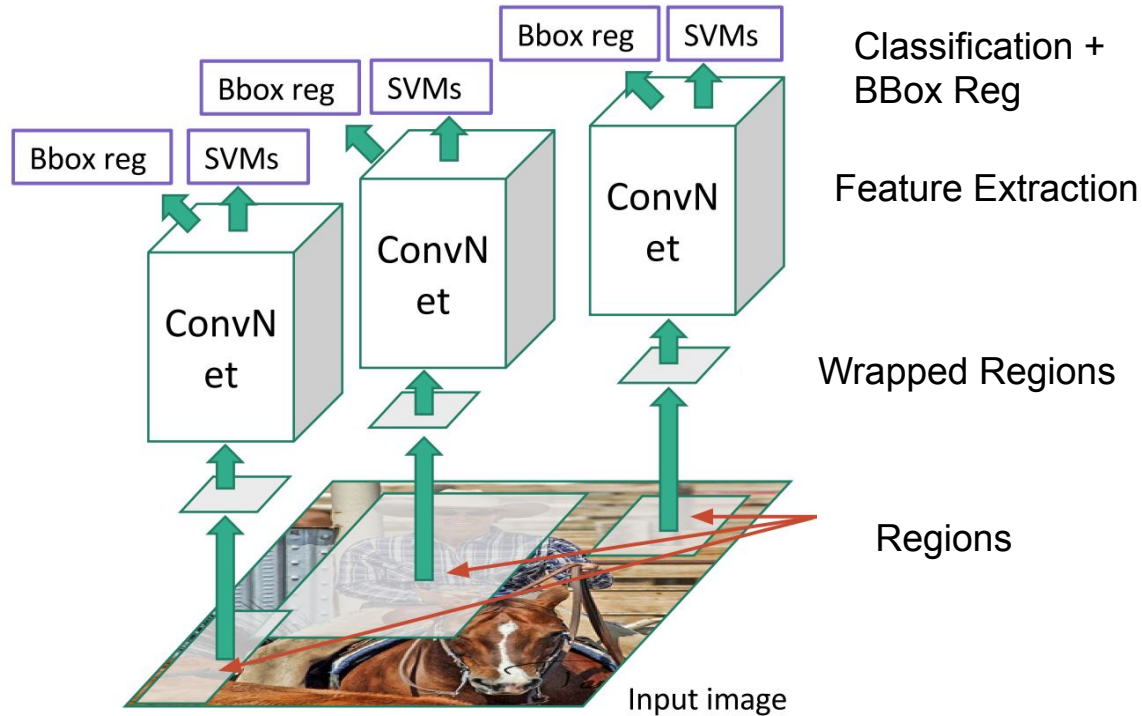
Can apply selective search: a technique for region proposal.

This technique is backbone of many standard detection architectures.

Various Types of Detectors

1. RCNN
2. SPP-Net
3. Fast-RCNN
4. Faster-RCNN
5. YOLO/SSD

RCNN



Problems with Slow-RCNN

Inefficient

Slow Training

Slow Testing

Takes up lot of disk space (ROIs)

SPP-Net

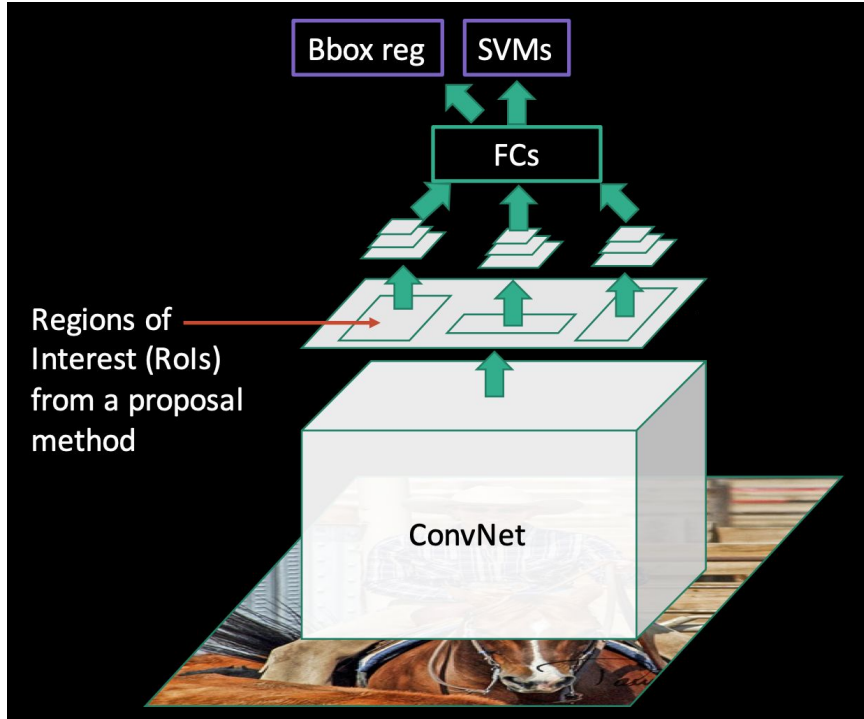


Image taken from: Ross Girshick, 2015;

3 Trainable Layers

Let conv-net features be used to extract ROIs.

Inference time is faster.

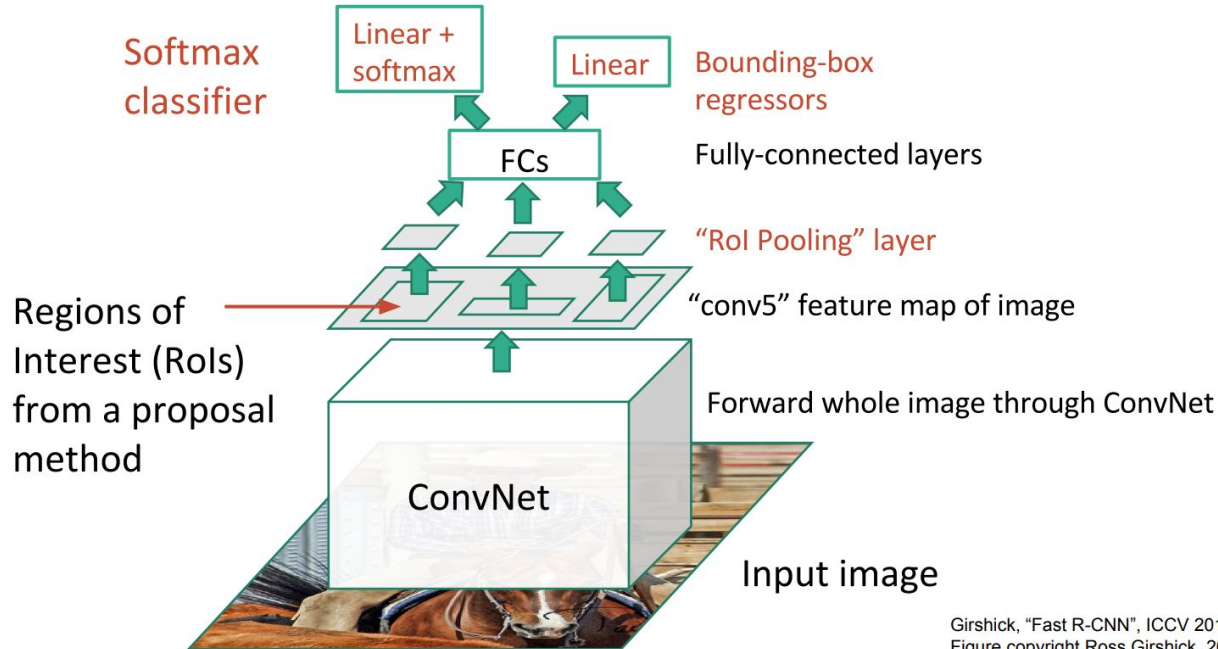
New Module introduced: SPP
Spatial Pyramid Pooling - a pyramid of pooling layers render a fixed length feature maps.

Problems with SPP-Net

Takes a lot of training time

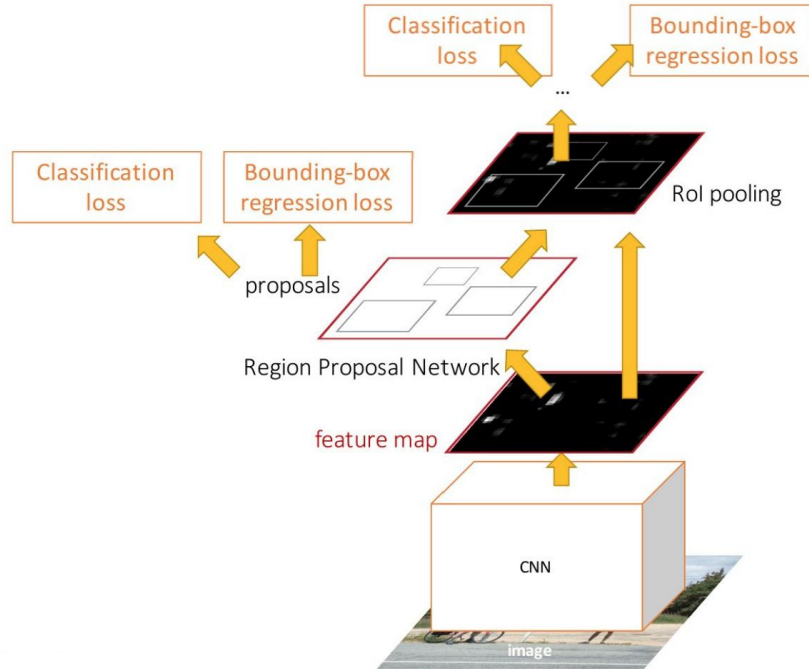
Freezes everything below ROIs so parameter updation is not possible - problem

Fast-RCNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015

Faster-CNN



Add a RPN.

Let CNN propose regions.

4 losses - joint training.

Network REDUCED.

Fast, Efficient.

YOLO/SSD - A Tiny Overview

No RPNs, no sliding windows.

Image, divide into a set of grid, assume a set of base boxes centered at each grid cell, do regression to a final box from each of the base box.

Gets you 5 outputs = (x, y, w, h, c) where c = confidence

Predict scores for each of class including background class.

Faster than Faster-RCNN, but not much accurate.

Redmon et al, "You Only Look Once: Unified, Real-Time Object Detection"

CVPR 2016 Liu et al, "SSD: Single-Shot MultiBox Detector", ECCV 2016

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

Backbones - Network Architectures

1. AlexNet
2. VGGNet
3. ResNet
4. Inception
5. DenseNet

THANK YOU!