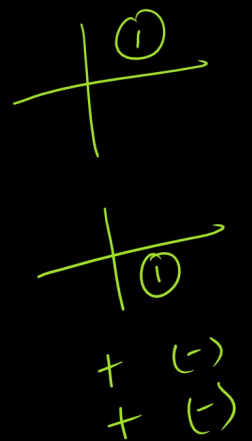
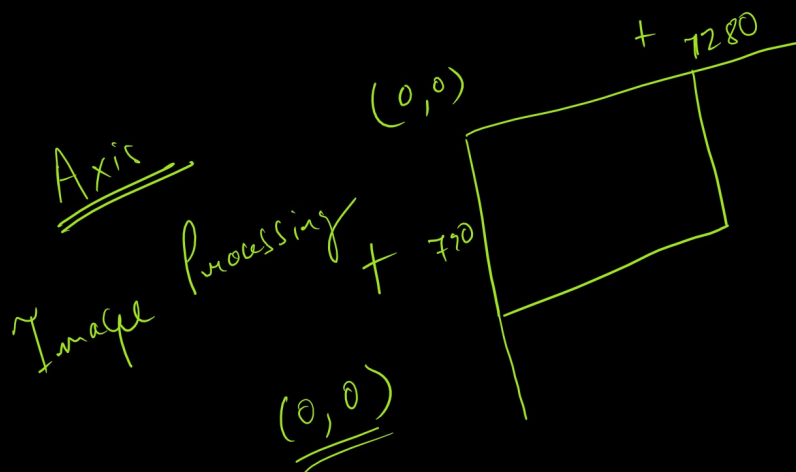
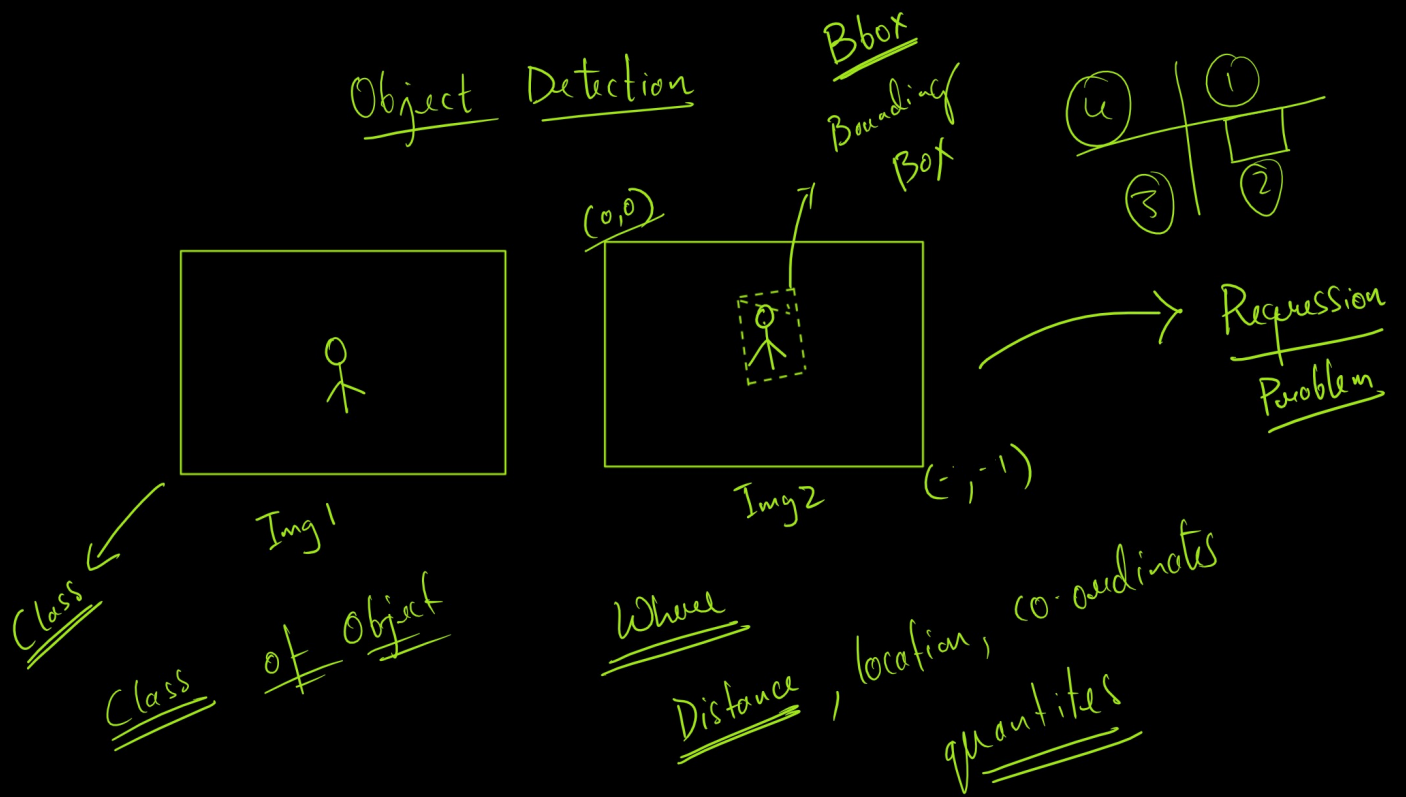


Agenda

- 1) Object Detection Basics
- 2) Metrics
- 3) Regression
- 4) Sliding Window
- 5) Image Pyramids

Object Detection

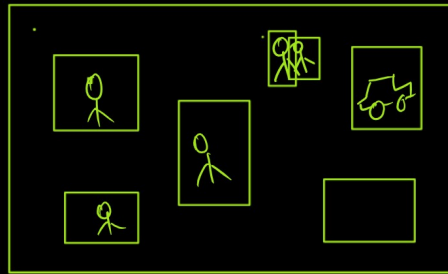


Bounding



1 single object

(0,0)



(x₁, y₁)



(x₂, y₂)

1

Crowd Counting

↳ no of Persons



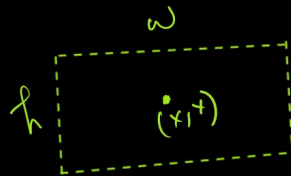
→ Create a Rectangle

Prob Stat

1) Class

2) Reg

CNN → Flatten → FC → FC → Softmax



Variables

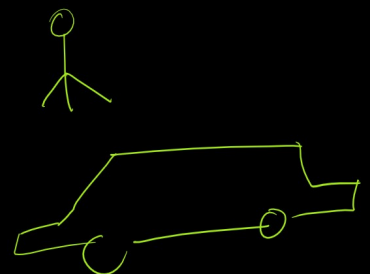
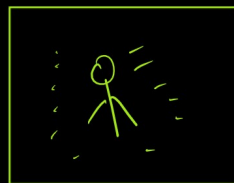


Mid Point
h, w

x, y, h, w
= h → Points

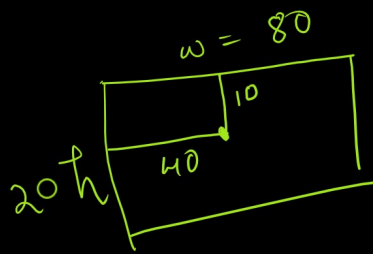
Object Detection = Classification + Regression
(Class) (BBBox coordinates)
(4 values)

{ mid, x }



Bbox = Object + Background

(h, w)



Coin

Case 1

w, h Case 2

mid point (x, y)

(h, w)

4 values

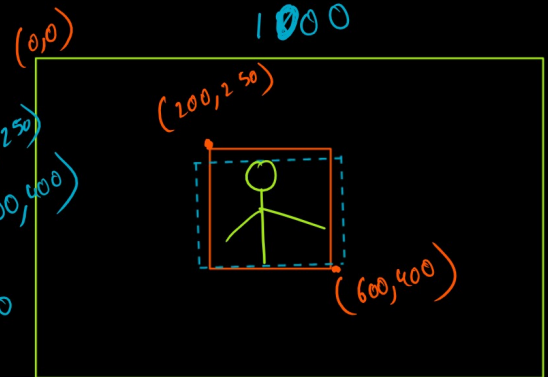
(x_1, y_1)
 (x_2, y_2)



Bounding Box Regression

L1 & L2

Ground Truth = (x_1, y_1) (x_2, y_2)
Predicted = (x_p, y_p) (x_p, y_p)
Distance = 0
Assume $(x_1, y_1) = (200, 250)$
 $(x_2, y_2) = (600, 400)$
1000



	x_1	y_1	x_2	y_2	L2 LOSS			
GT	200	250	600	400	$(200-0)^2$	$(250-0)^2$	$(600-800)^2$	$(400-600)^2$
Prediction	0	0	800	600	$(200-100)^2$	$(250-150)^2$	$(600-700)^2$	$(400-450)^2$
	210	240	590	405				

200 250 600 400

$= 182500$
 \downarrow
 $= 32500$
 \downarrow
 $= 5000$
 \downarrow
 $= 0$
close

Ground Truth
Truth

loss function

5 classes

dog
cat

less
Hold

25 40
 \checkmark Dog AP = 53
 \checkmark Cat AP = 60

60

80

90

\checkmark House = 67
 \checkmark person = 80
 \checkmark wheel = 63

Sliding Window

Window Size (3,4)
 = (x,y)
 2,2
 Scanning of the
 entire image.



Image Pyramid

Object Size

Window Size is always
 less than the
 size of image.

More Edges
 More Bottom
 Size OR Scale
 Find → Then Localize



Fixed Window Size
Diffrent Image Resolution

Sliding Window w/ Image Pyramids

New Concepts

1) Edge Detection \rightarrow (B/W)

2) Selective Search \rightarrow Colours

Region Proposals

Subpixels Strandling

RGB

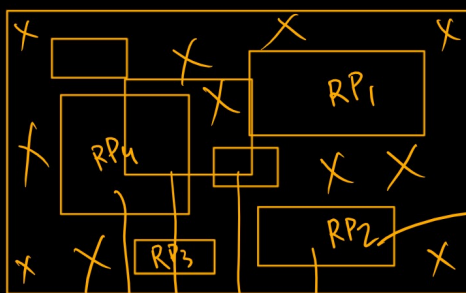
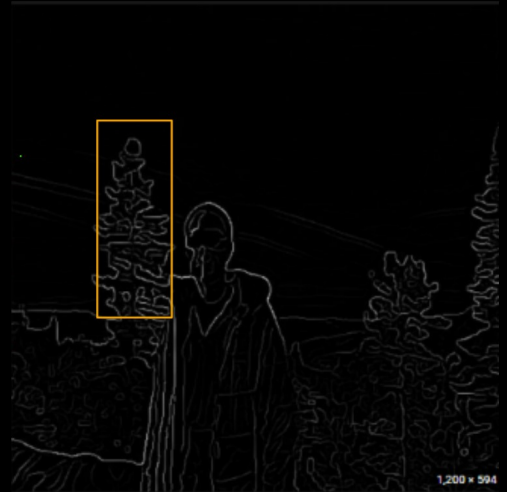
B/W

RGB

VGG

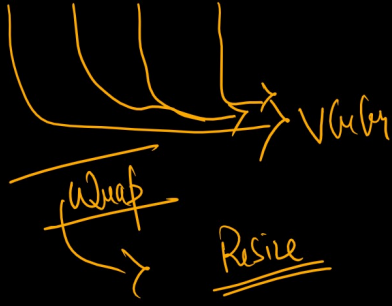
Resnet

\rightarrow RGB \rightarrow 1st layer



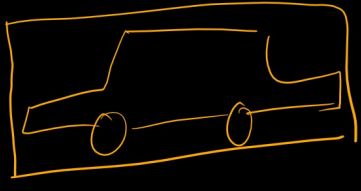
- 1) Edge Boxes
- 2) Selective Search

Region Proposal \rightarrow (Object)



Input Size Constraint

RP \rightarrow WRAP \rightarrow Value

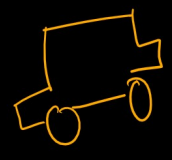


Wrap

RCNN, FAST RCNN
*

Faster RCNN
✓

Resize



(224, 224)

Blurp

distortion
Aspect Ratio

List (x_1, y_1)

