

# Session-7

## OBJECT

## LOCALIZATION & DETECTION-1

March 27, 2022

Professor: What does the question say?

Me: Derive backpropagation algorithm for a 3-layer neural network.

Professor: then why have you written "Don't worry about it"?



I'm a new employee and I'd like to change my username. How am I supposed to sell our software using this address?

Lorenzo Servantez  
laser@somesoftwarecorp.com



Mr. Servantez,

Unfortunately, all email addresses are automatically generated by the system and cannot be changed. Please, believe me.

Regards,  
Biron Tchaikovsky  
bitch@somesoftwarecorp.com

## AGENDA

- ① Understanding problem of Object localization & detection
- ② IOU - better way to evaluate bounding box prediction
- ③ Non-Maximum Suppression
- ④ Multi-object detection
- ⑤ RCNN }
- ⑥ Fast-RCNN } obsolete
- ⑦ Faster-RCNN } some people use this

$3 \times 3$  — Scissor

Merge / Empend

$1 \times 1$

1

$$\cancel{3 \times 3 \times 128 = 9}$$

128

$128 \rightarrow 32 \geq \underline{\underline{1 \times 1}}$

$224 \times 224 \times 128 \circ (\underline{\underline{1 \times 1 \times 128}}) \times 32$

32

on/off switch



0



0



→

32 chan

:

:

1

1



0

$a_1$	$a_2$	$a_3$
$a_4$	$a_5$	$a_6$
$a_7$	$a_8$	$a_9$

0

$P_1$	$P_2$	$P_3$
$P_4$	$P_5$	$P_6$
$P_7$	$P_8$	$P_9$

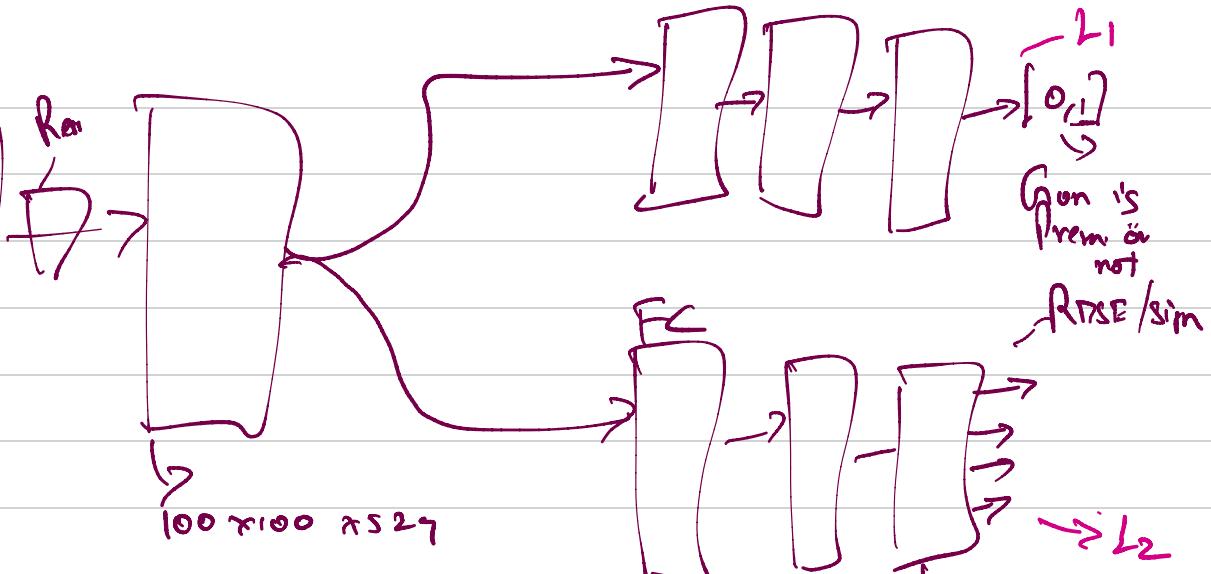
$$56 \times 56 \times 128 \oplus \underbrace{(1 \times 1 \times 128) \times 1}_{\downarrow}$$

$$1 \boxed{56} \quad 56 \times B_1$$

$$2 \boxed{56} \quad 56 \times B_2$$

⋮

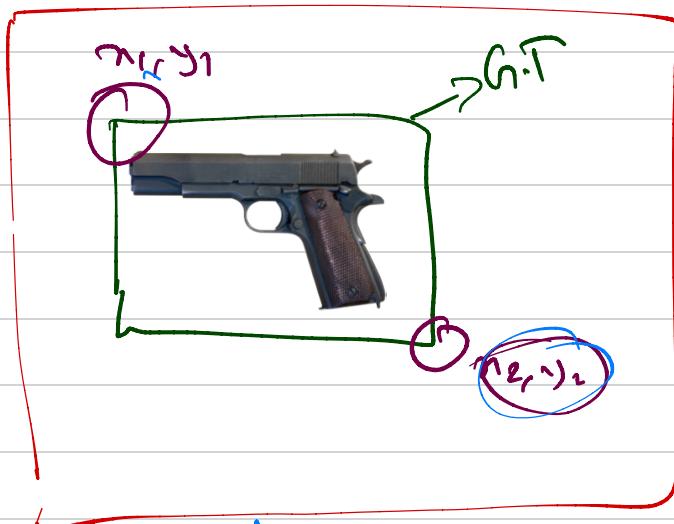
$$128 \boxed{56} \quad 56 \times \boxed{?}_{128}$$



$$\text{Loss} = \alpha \times L_1 + \beta \times L_2 \rightarrow \text{mCE}$$

$\overbrace{\alpha}$  weight  
 $\overbrace{\beta}$  weight  
 $L_1$   
 $L_2$   
 binary Cross entropy

$\overbrace{\alpha}$  weight  
 $\overbrace{\beta}$  weight  
 $u$   
 $u_{\text{out}}$



In Yolo we do not predict co-ords



$[x, y, S, 3]$   
 What we  
 predict  
 in  
 Yolo

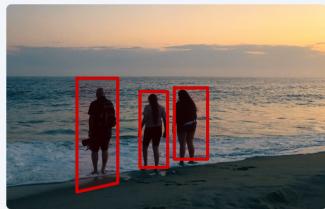
Guarantee → Some will be the  
 ↗ ~~Classifiers~~

FIR - Entries  
 ↗ take a pic → Allow/  
 Deny

Production → latency  
 throughput  
 Ram / Core spec

## Semantic Segmentation vs. Instance Segmentation vs. Panoptic Segmentation

Label  
Studio



(a) Image



(b) Semantic Segmentation

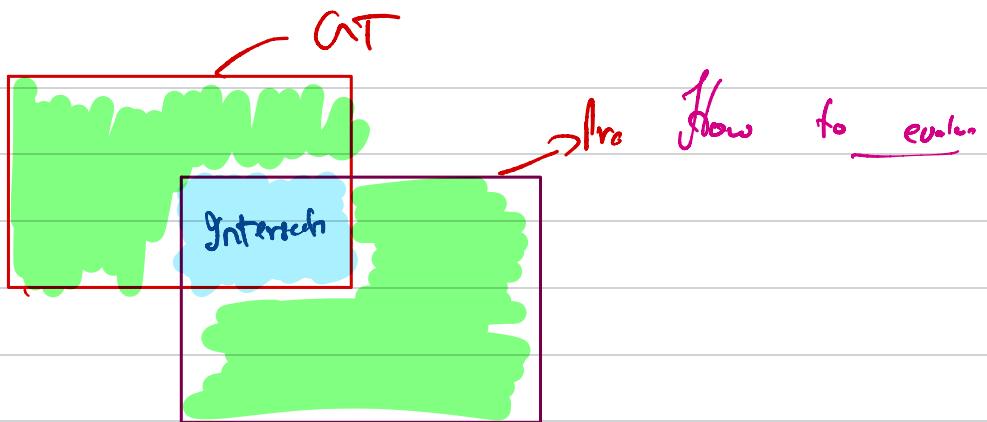


(c) Instance Segmentation



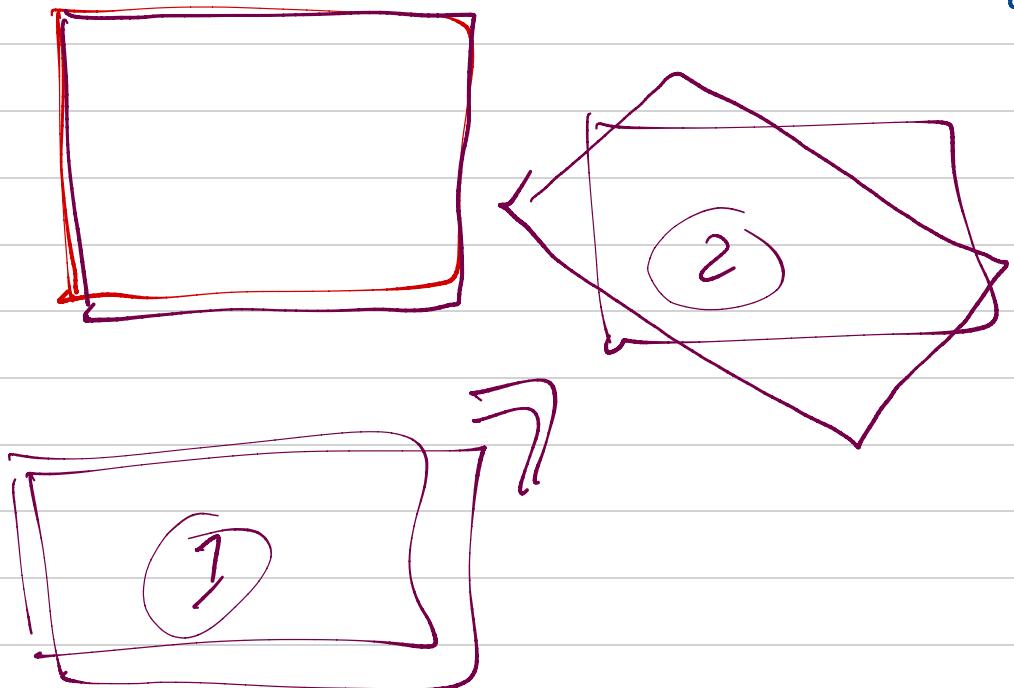
(d) Panoptic Segmentation

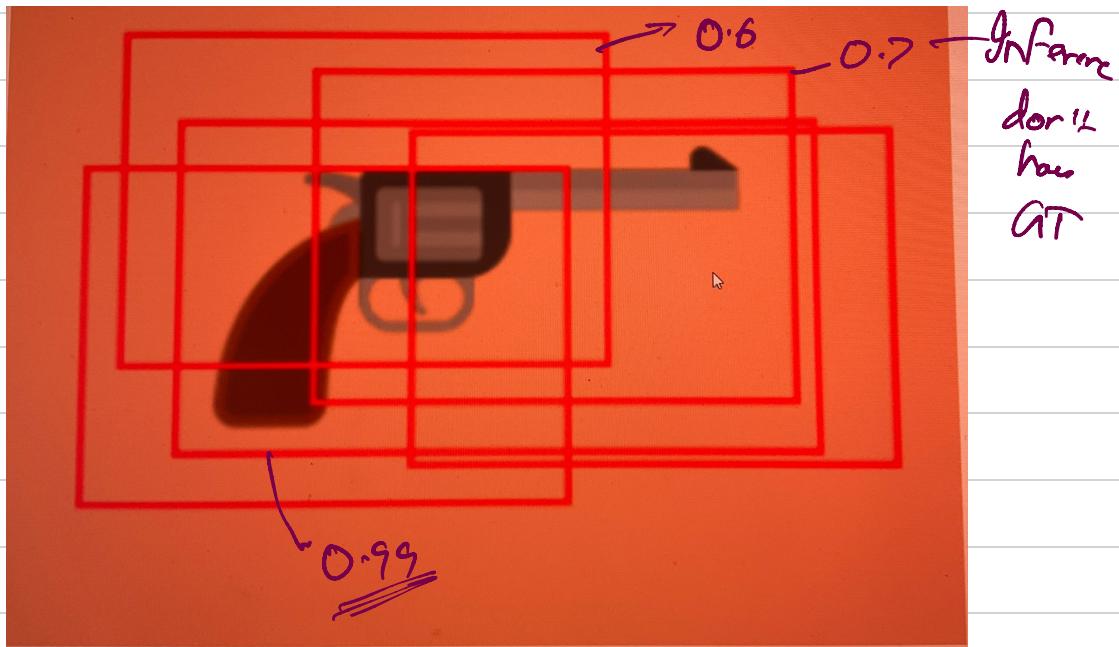
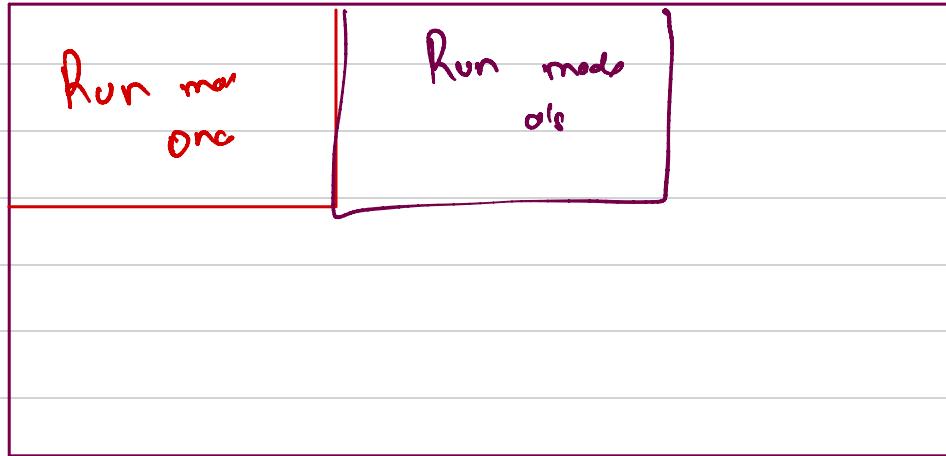
V7 Labs



$$= \frac{\text{Intersection}}{\text{Union}}$$

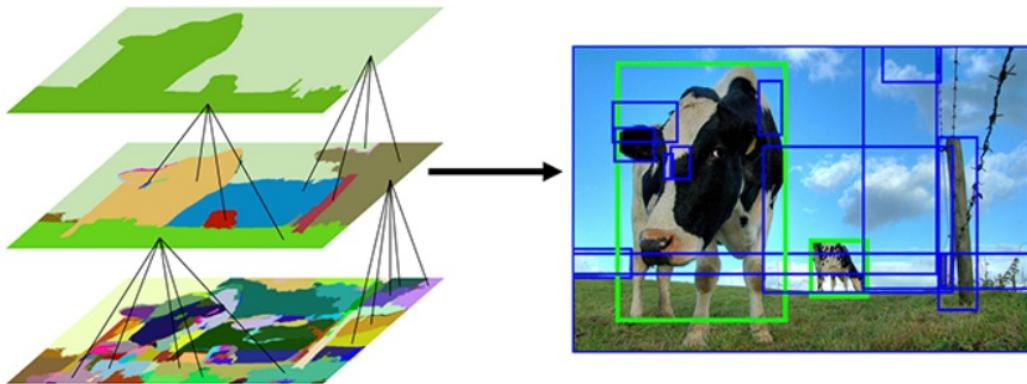
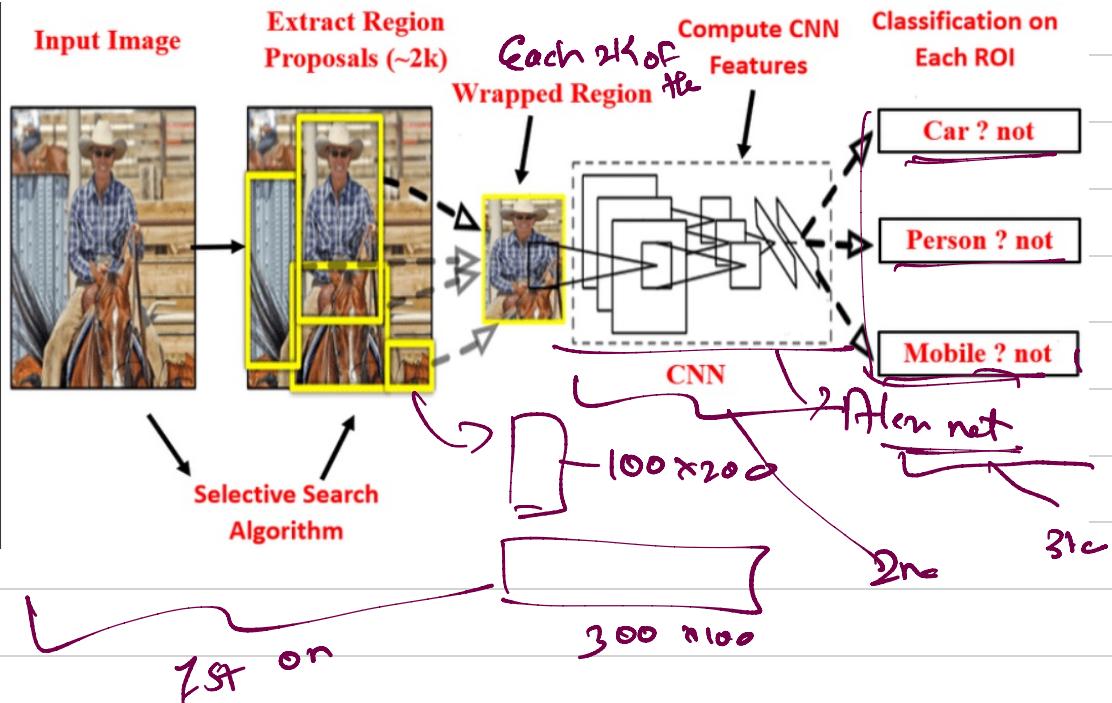
Green + blue color.

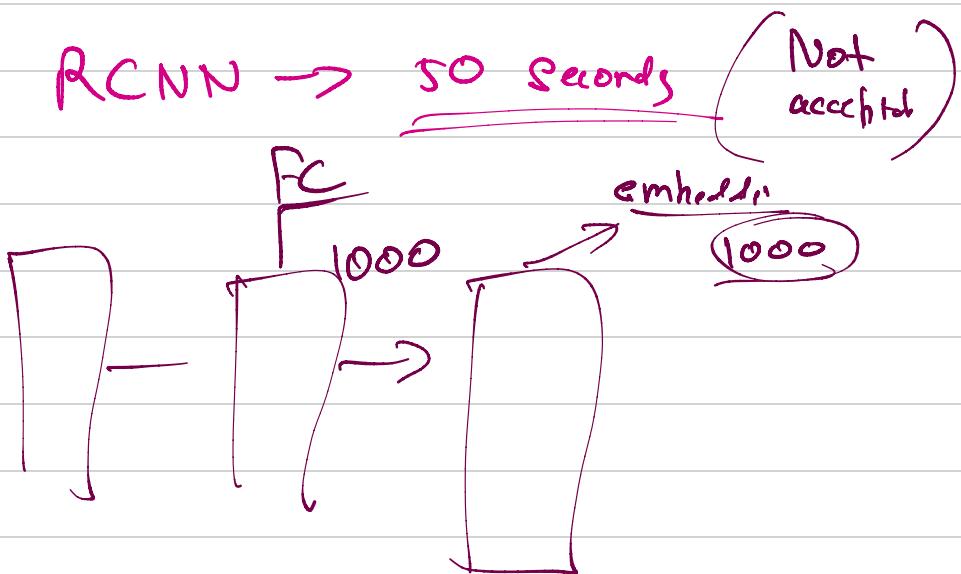
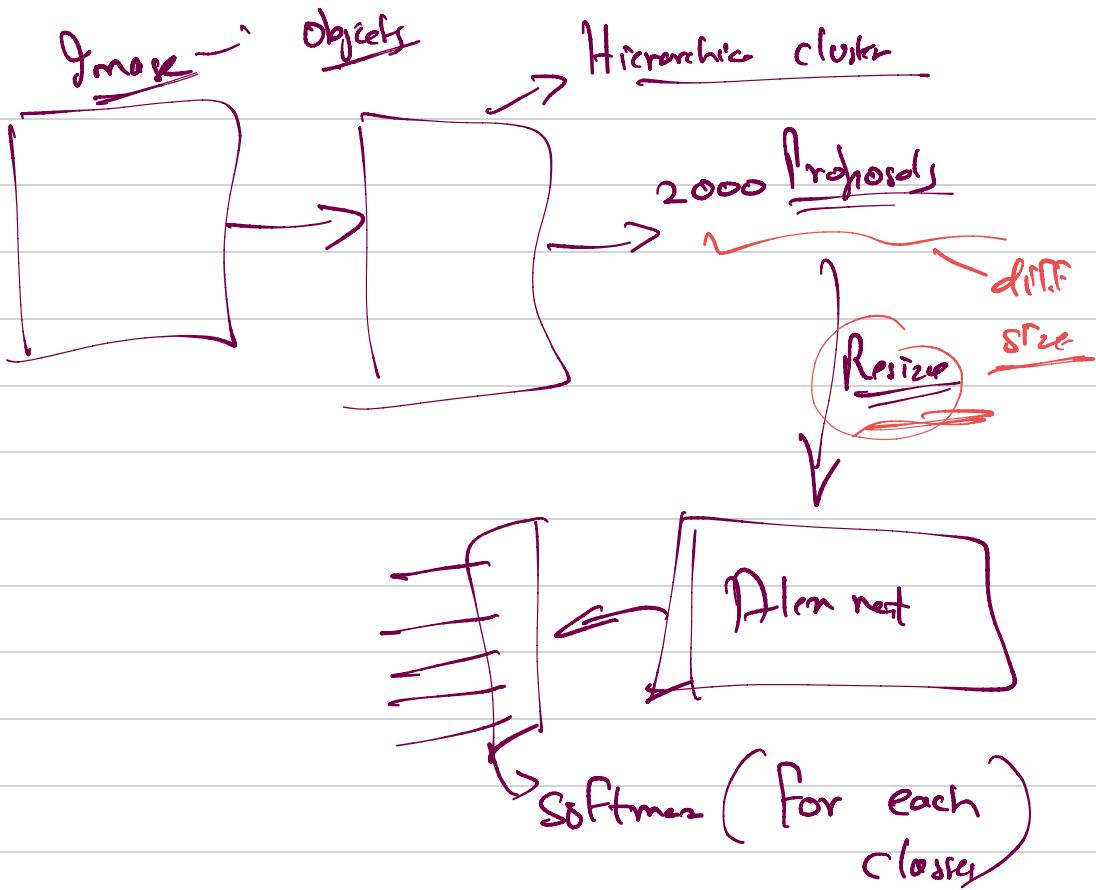


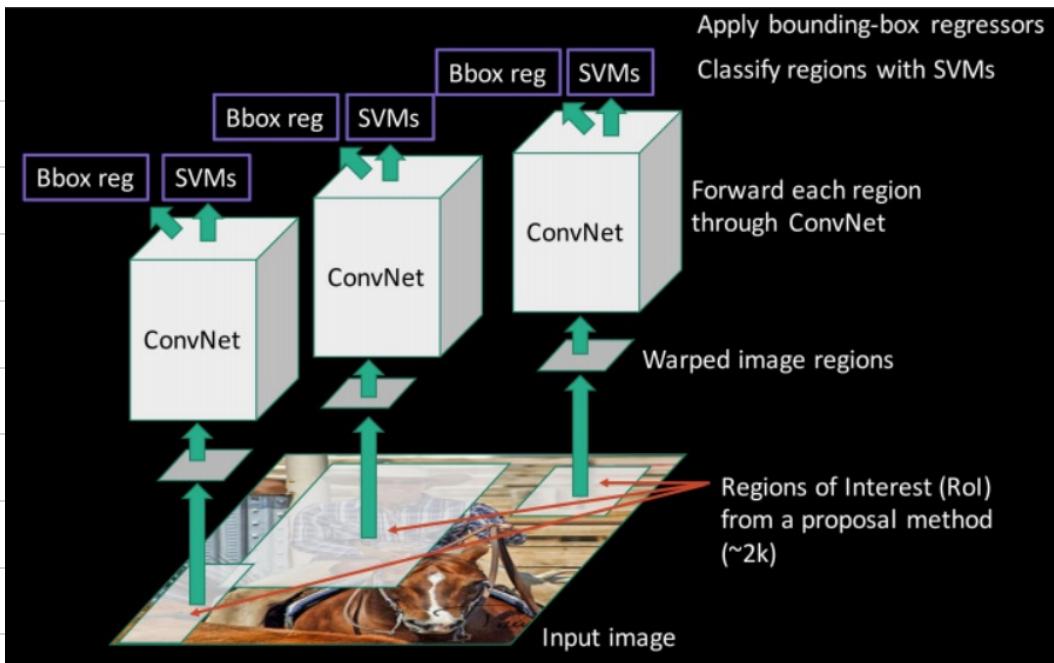


Classifie → highest  
R<sub>45</sub>

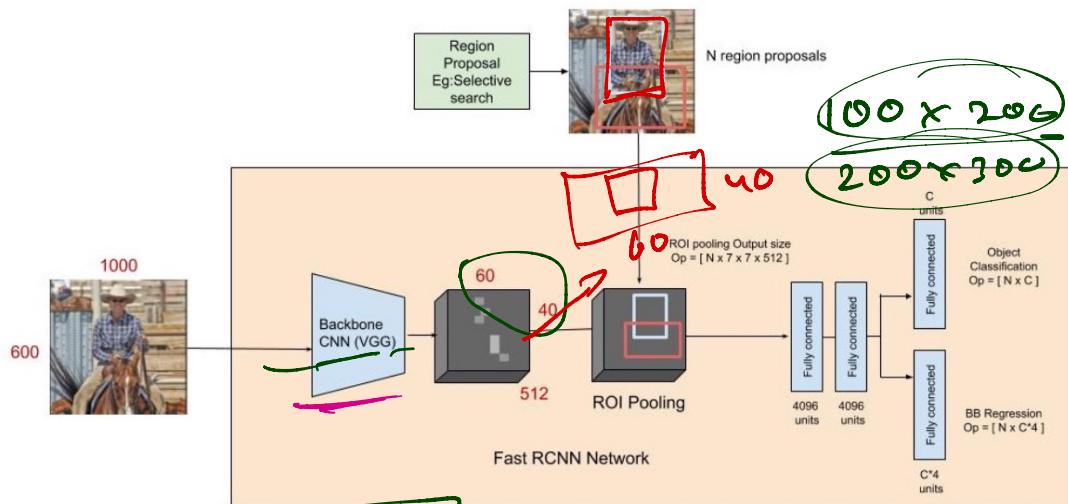
RCNN → Dec 2013







# Fast-RCNN

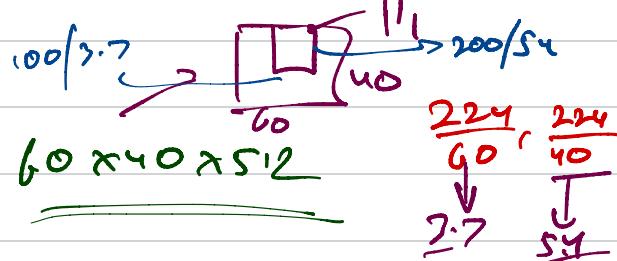


$60 \times 40 \times 512$

Attnv

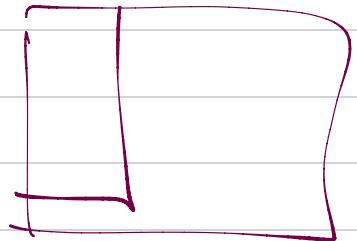
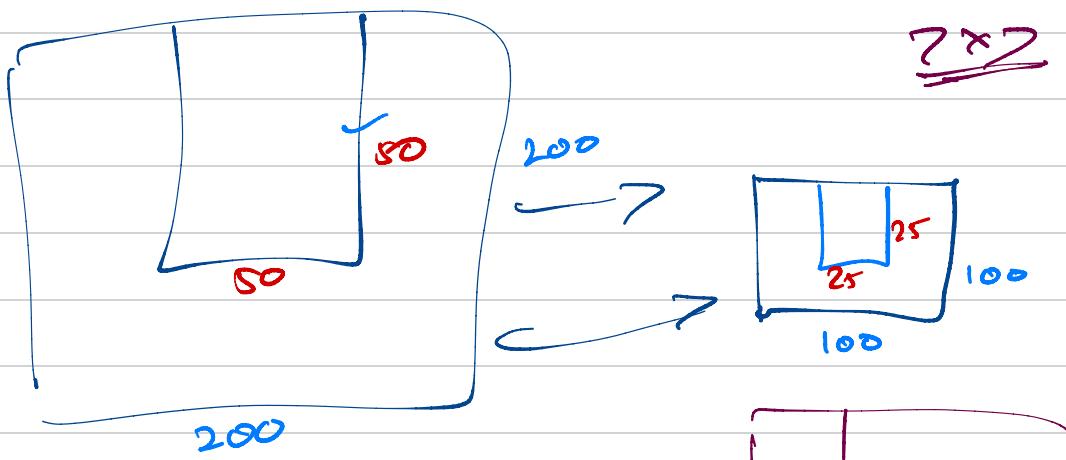
Classified

Fast RCNN  $\rightarrow$  Combin



$\rightarrow$  R-Proposed  $\rightarrow$  2K Proposals

$$\frac{100/3.7 \times 200/5.4}{2} > \frac{100 \times 200}{300 \times 400}$$



Region Pooling  $\rightarrow$  Output size  $2 \times 2$

Region proposal



0.88	0.44	0.14	0.16	0.37	0.77	0.96	0.27
0.19	0.45	0.57	0.16	0.63	0.29	0.71	0.70
0.66	0.26	0.82	0.64	0.54	0.73	0.59	0.26
0.85	0.34	0.76	0.84	0.29	0.75	0.62	0.25
0.32	0.74	0.21	0.39	0.34	0.03	0.33	0.48
0.20	0.14	0.16	0.13	0.73	0.65	0.96	0.32
0.19	0.69	0.09	0.86	0.88	0.07	0.01	0.48
0.83	0.24	0.97	0.04	0.24	0.35	0.50	0.91



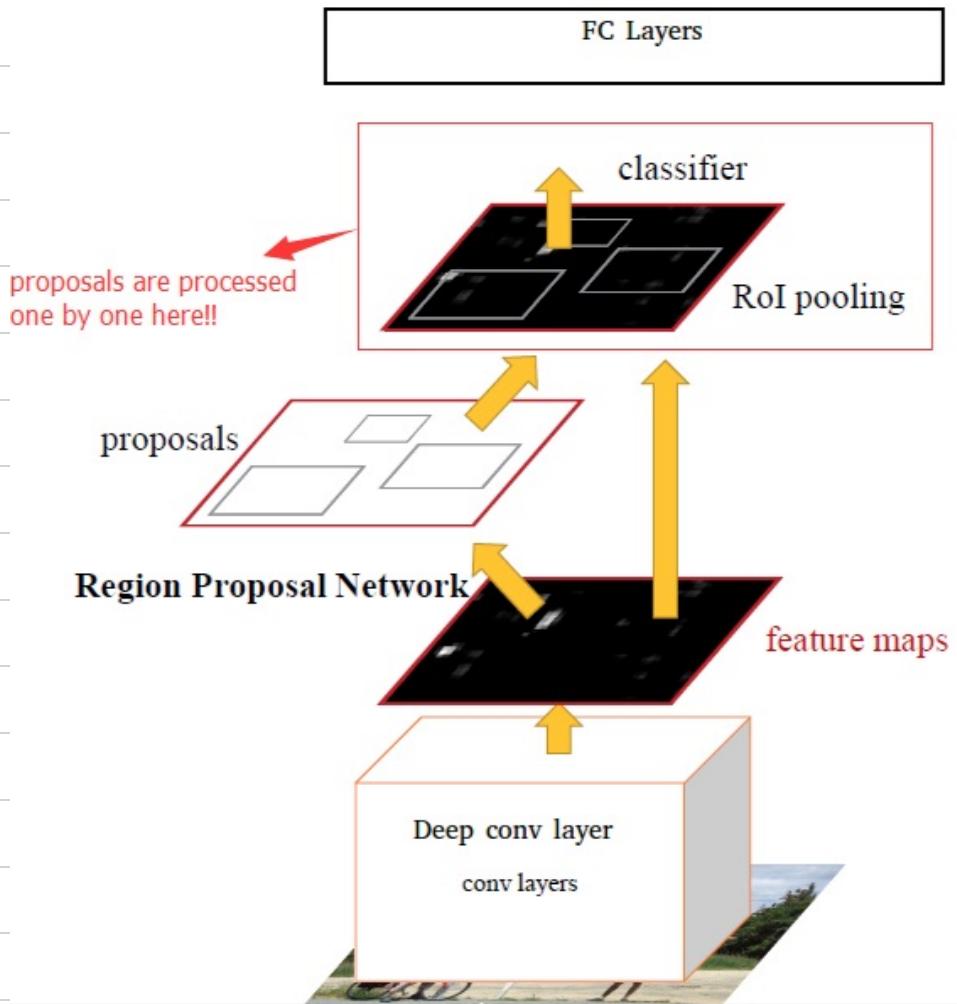
Input Image

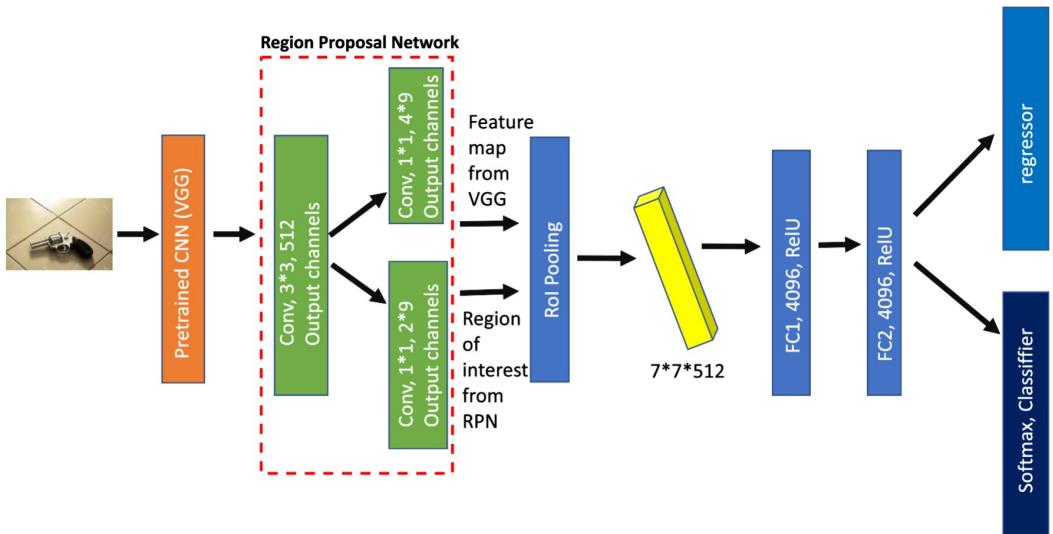


Region proposal

0.88	0.44	0.14	0.16	0.37	0.77	0.96	0.27
0.19	0.45	0.57	0.16	0.63	0.29	0.71	0.70
0.66	0.26	0.82	0.64	0.54	0.73	0.59	0.26
0.85	0.34	0.76	0.84	0.29	0.75	0.62	0.25
0.32	0.74	0.21	0.39	0.34	0.03	0.33	0.48
0.20	0.14	0.16	0.13	0.73	0.65	0.96	0.32
0.19	0.69	0.09	0.86	0.88	0.07	0.01	0.48
0.83	0.24	0.97	0.04	0.24	0.35	0.50	0.91

# FASTER - Rcnn





External region proposals method (Selective Search Algorithm)	✓	✓	✗
	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image	50 seconds	2 seconds	0.2 seconds
Speed-up	1x	25x	250x
mAP (VOC 2007)	66.0%	66.9%	66.9%