

Session-7

OBJECT - DETECTION & LOCALIZATION 2

March 29, 2024



AGENDA

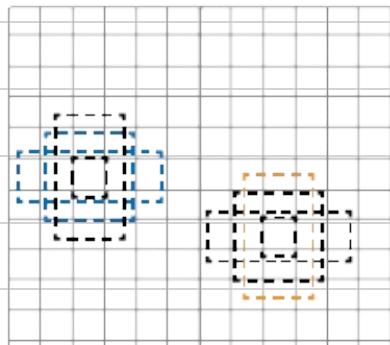
- ① Anchor boxes
- ② Recap R-CNN & Fast R-CNN & Faster R-CNN
- ③ SSD - Yolo V3 → Single object detection

Multiple object detection.

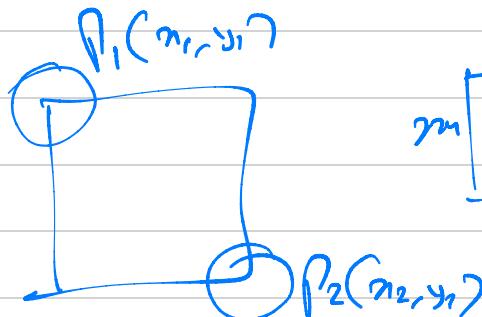
ANCHOR BOXES



Ground truth image and
bounding boxes

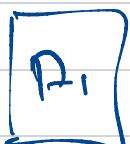
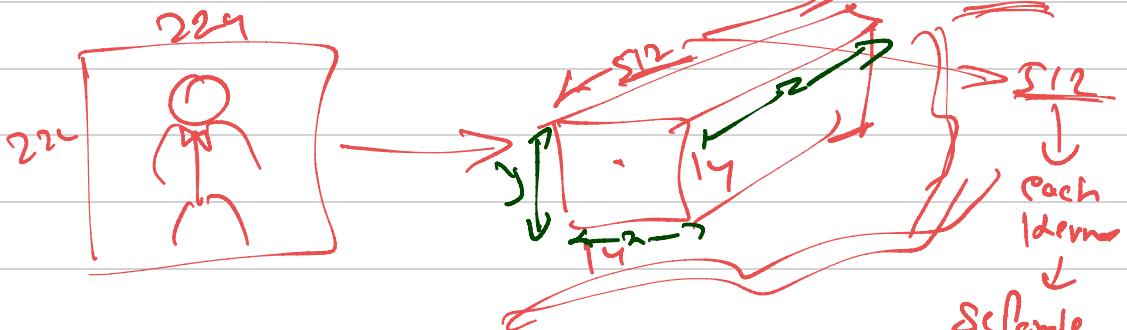


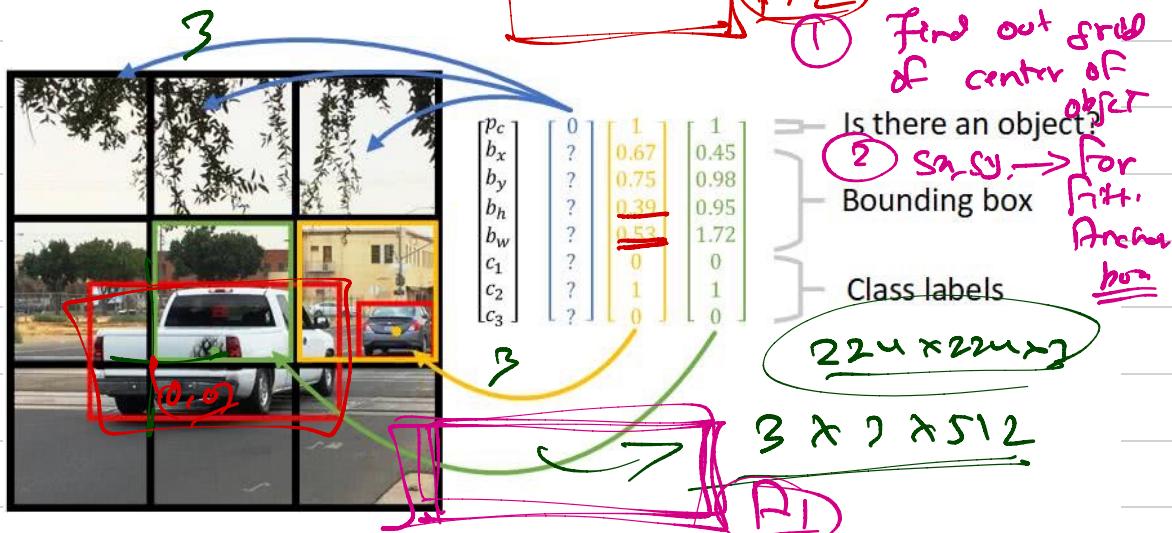
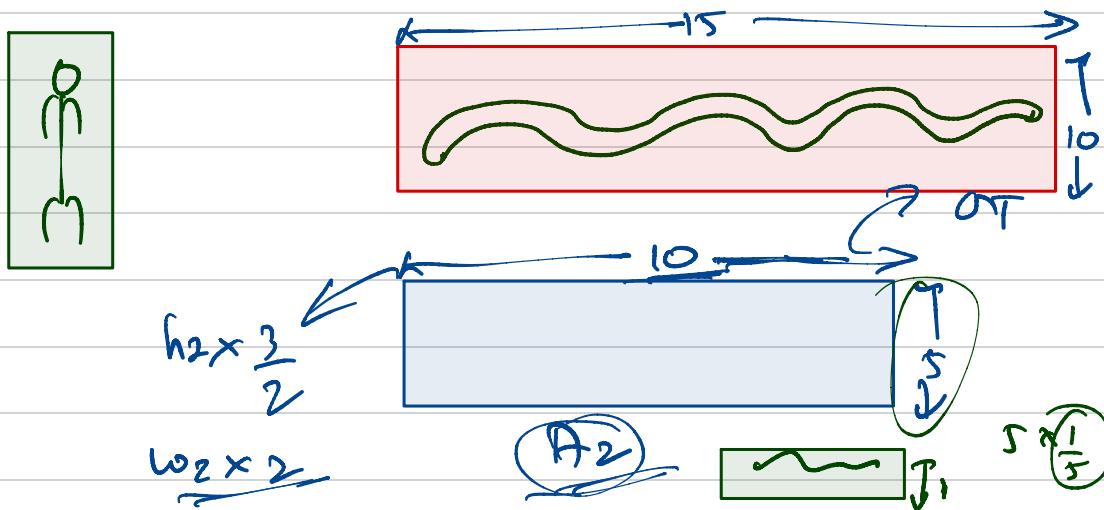
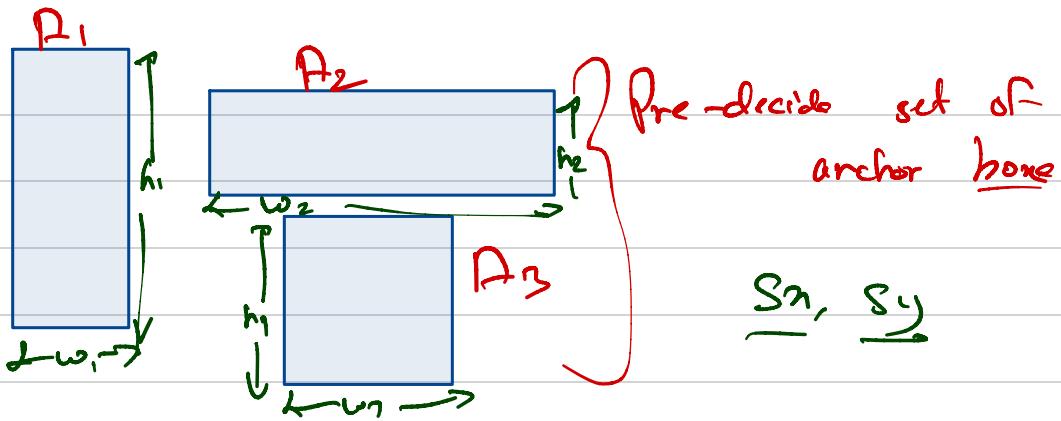
Anchor boxes at each predefined
location in each feature map



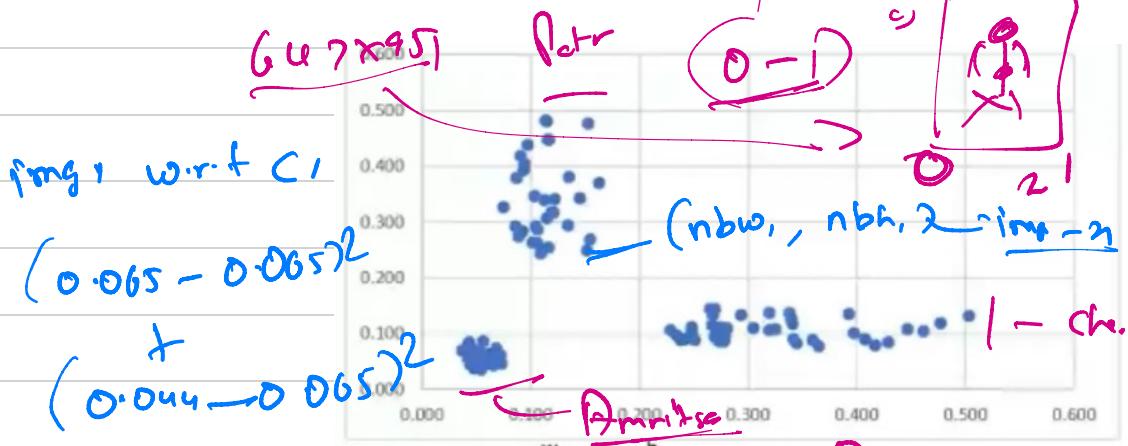
224×3
 m $\rightarrow 14 \times 14 \times 512$

- ① Spatial information
- ② Feature set 1





	W	H	Class	cn	cy	bw	bh	nw	nh	nr	ny	nbw	nbh
img1	647	951	5	337	522	42	41	1	1	0.521	0.549	0.065	0.044
img2	783	755	0	614	433	44	40	1	1	0.784	0.574	0.056	0.054



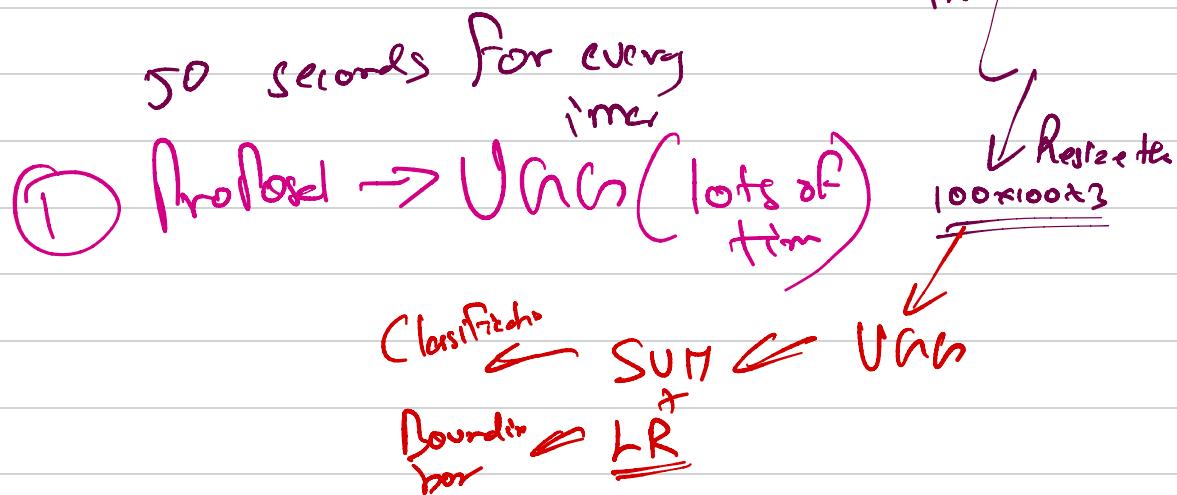
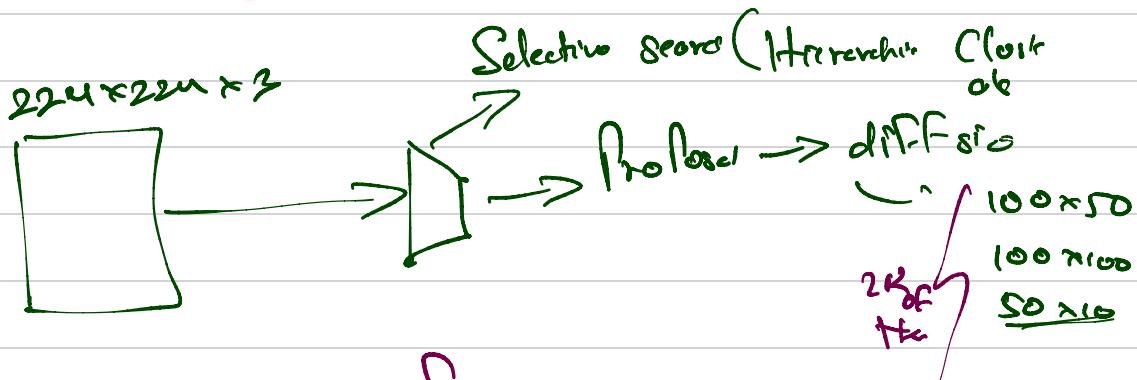
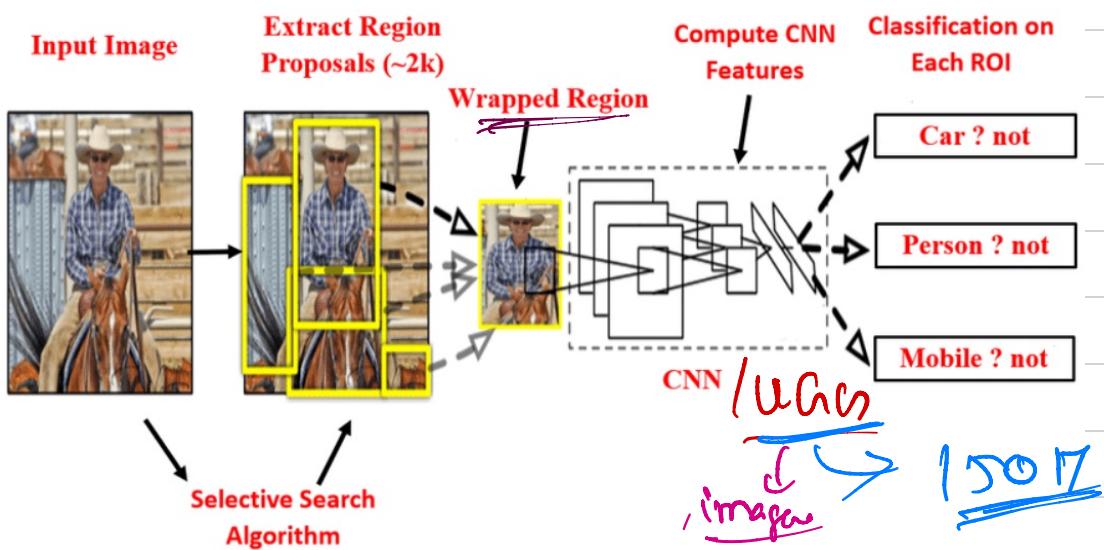
most strength
ht / ut of & class

→ Find out Euclidean distance for a given training data → wrt A₁, A₂, D₃

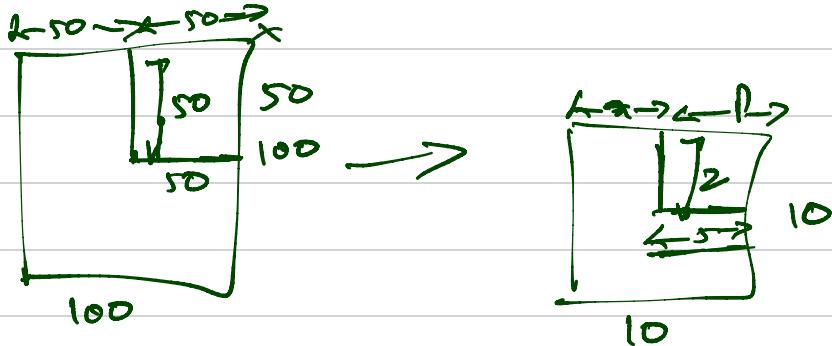
level - 1 Data onset

RCNN

2014 (late 2017)



FAST-RCNN

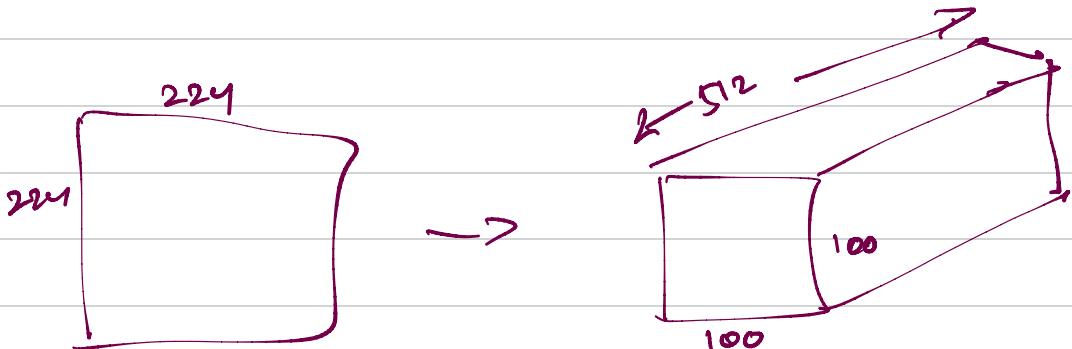


$$\begin{aligned} n &= 5 \\ P &= 5 \\ Z &= 5 \end{aligned}$$

① - Port 1



VGG $\rightarrow 100 \times 100 \times 512$



$P_1 \rightarrow 5 \times 6 \times 512$

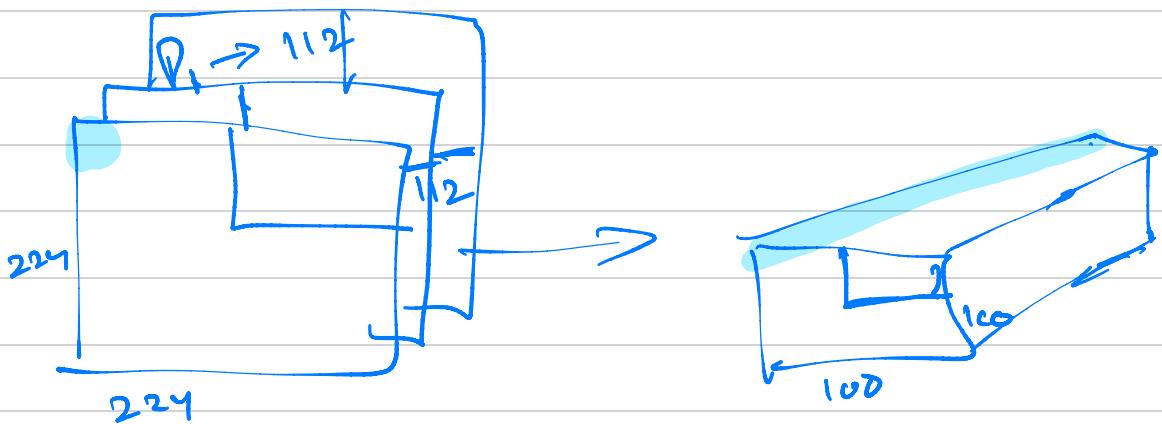
$P_2 \rightarrow 8 \times 7 \times 512$

Common Size

Port 2

① → 2K Proposals → Varred sizes

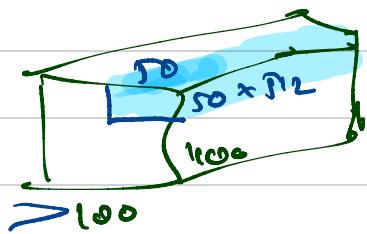
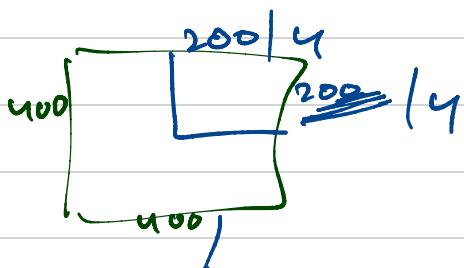
② → $100 \times 100 \times 512$ ↗ Project



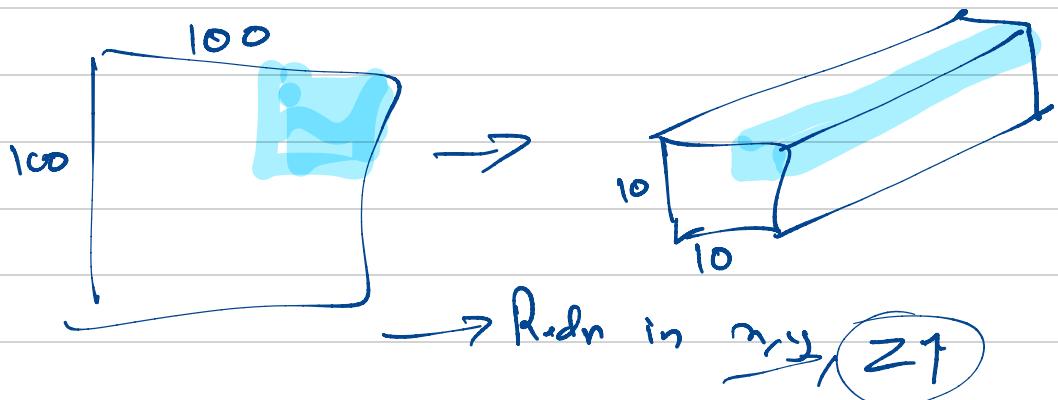
$P_1 \rightarrow 112 \times 112 (R, G, B) \rightarrow 512 \text{ chan}$

$400 \times 400 \times 3 \rightarrow$

$100 \times 100 \times 512$



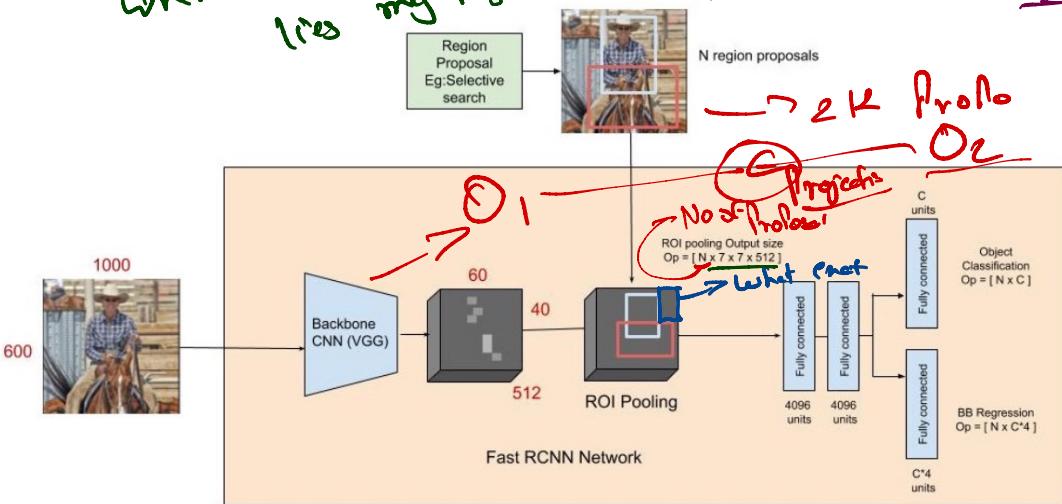
$$1, \underline{200 \times 200 \approx 3} \rightarrow \underline{50 \times 50 \approx 512}$$



What region of my Feature map
corresponds to Proposals

where exactly on feature map
lies my Regions

→ Scaling - $O(1)$



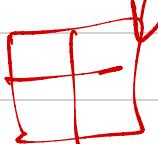
2x2

REGION POOLING

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	0.61	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



$2 \times 2 \times 52$

Region proposal

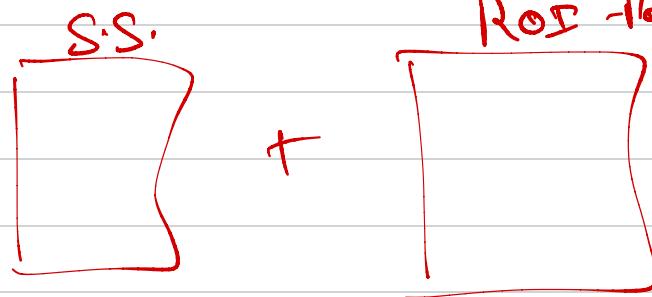
0.88	0.44	0.14	0.16	0.37	0.77	0.96	0.27
0.19	0.45	0.57	0.16	0.63	0.29	0.71	0.70
0.66	0.26	0.82	0.64	0.54	0.73	0.59	0.26
0.85	0.34	0.76	0.84	0.29	0.75	0.62	0.25
0.32	0.74	0.21	0.39	0.34	0.03	0.33	0.48
0.20	0.14	0.16	0.13	0.73	0.65	0.96	0.32
0.19	0.69	0.09	0.86	0.88	0.07	0.01	0.48
0.83	0.24	0.97	0.04	0.24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

$2 \times 2 \times 511$

Input Image

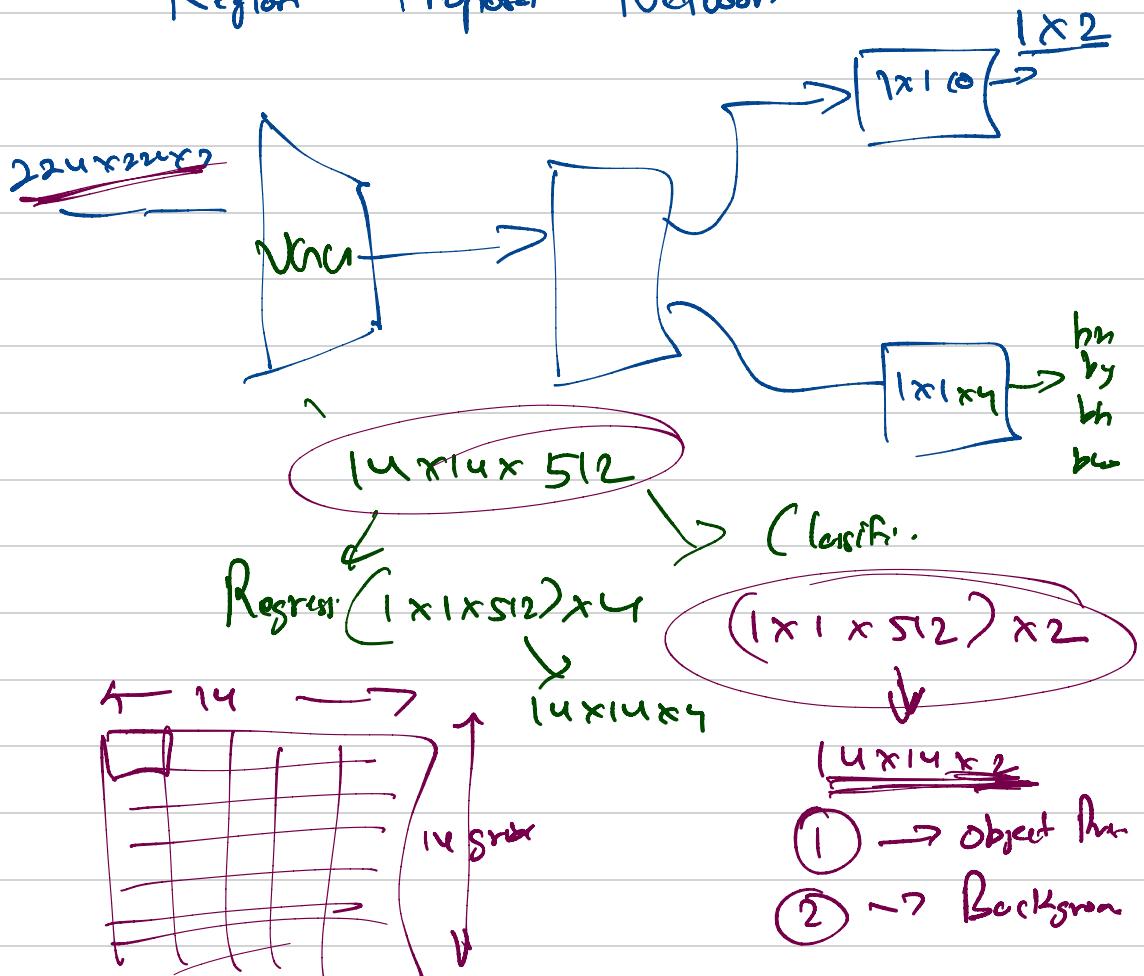


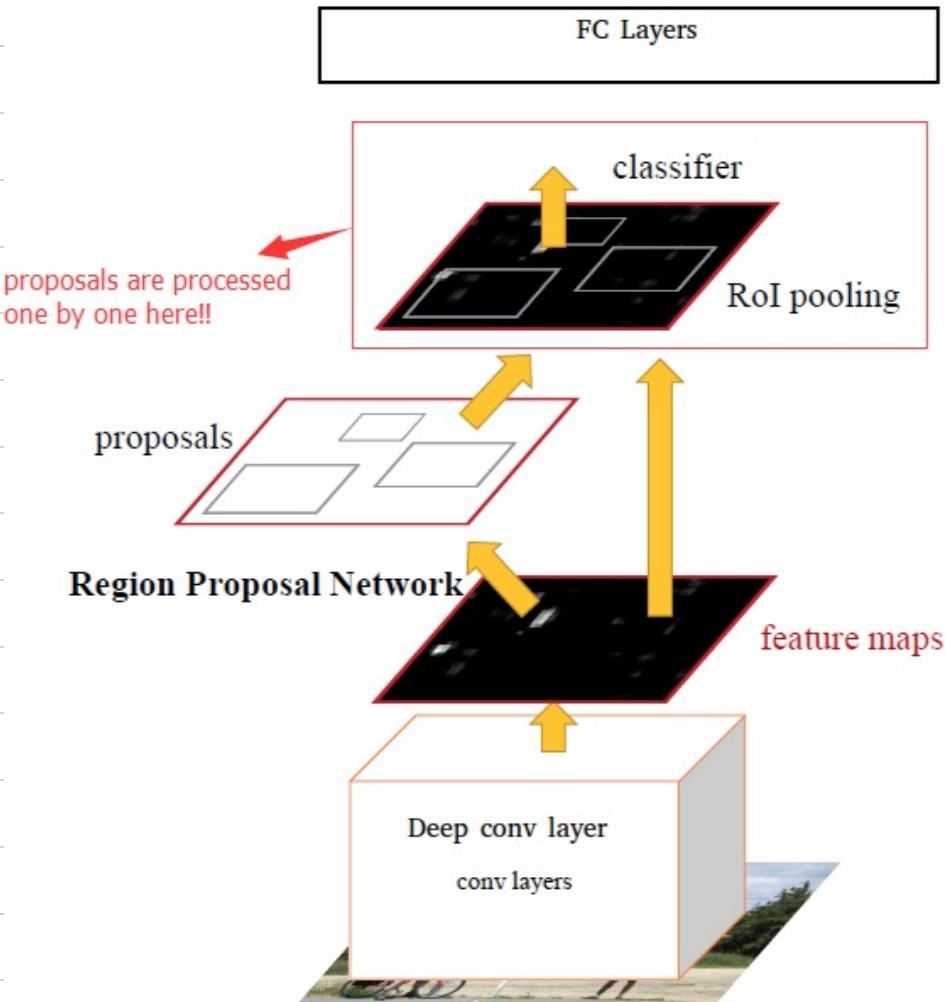
Fast - RCNN FASTER - RCNN

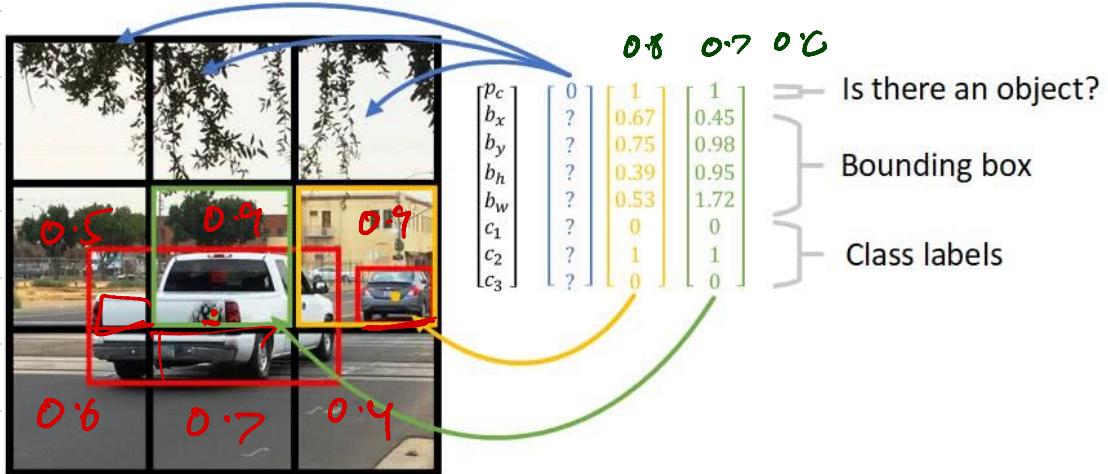


Region Proposal Network

Objectness score
background score







luxury

$$\Pr \left[\text{L} \right] = ?$$

$$400 \times 400 \times 3 \rightarrow \boxed{10 \times 10 \times 512}$$

