

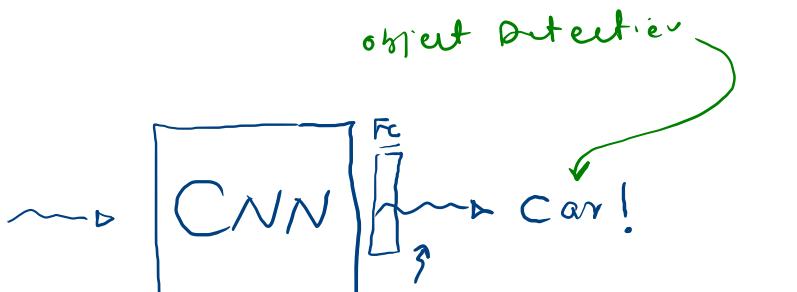
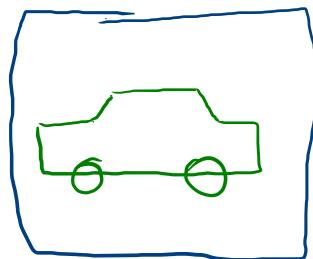
Object



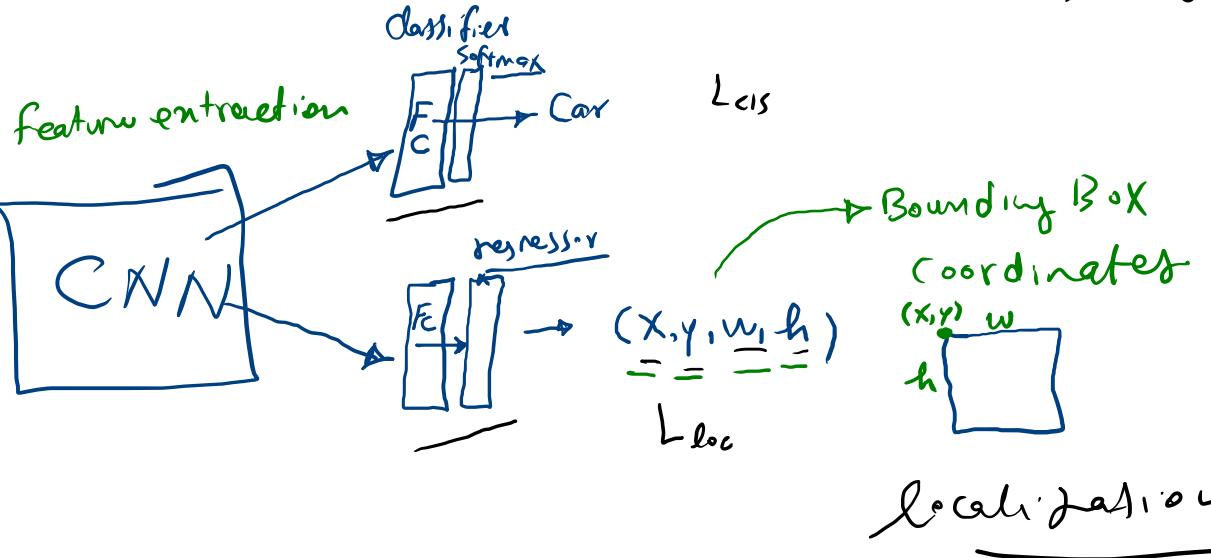
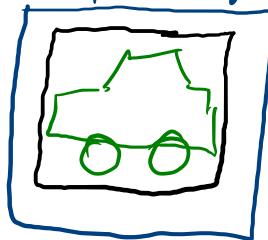
Classification + Localization

Detection.

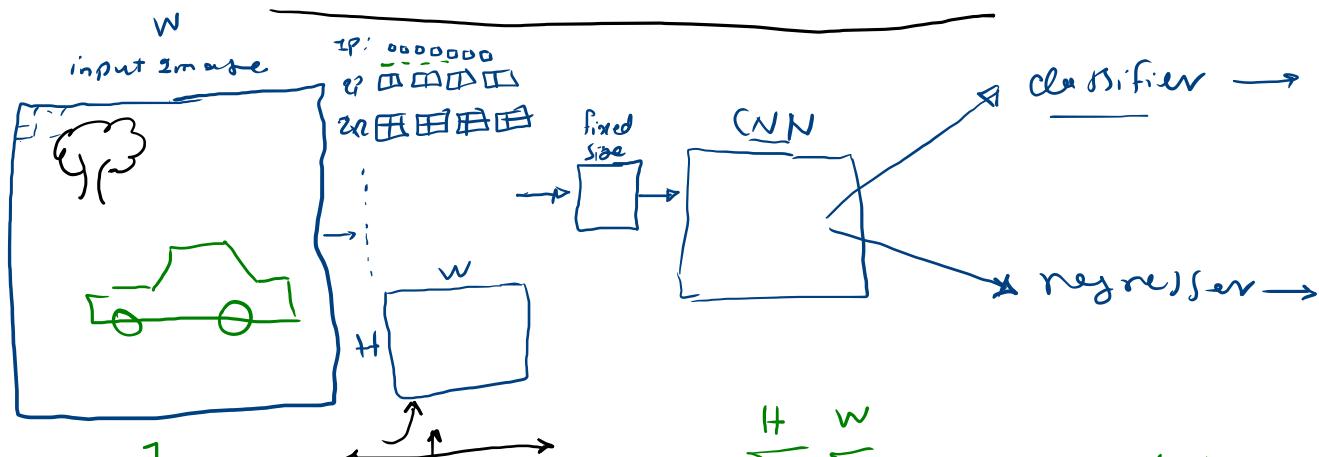
what?  
input Img



where?  
input Img



جِهَادُ الْجَمَعَةِ، object  
عِنْدَكُمْ



$$\text{number of F images} = \sum_{h=1}^H \sum_{w=1}^W (w-w+1)(H-h+1) = \frac{H(H+1)W(W+1)}{2}$$

$$\left\{ \begin{array}{l} W=224 \\ H=224 \end{array} \right. \rightarrow \frac{224(225)224(225)}{2} = 635 \text{ million.}$$

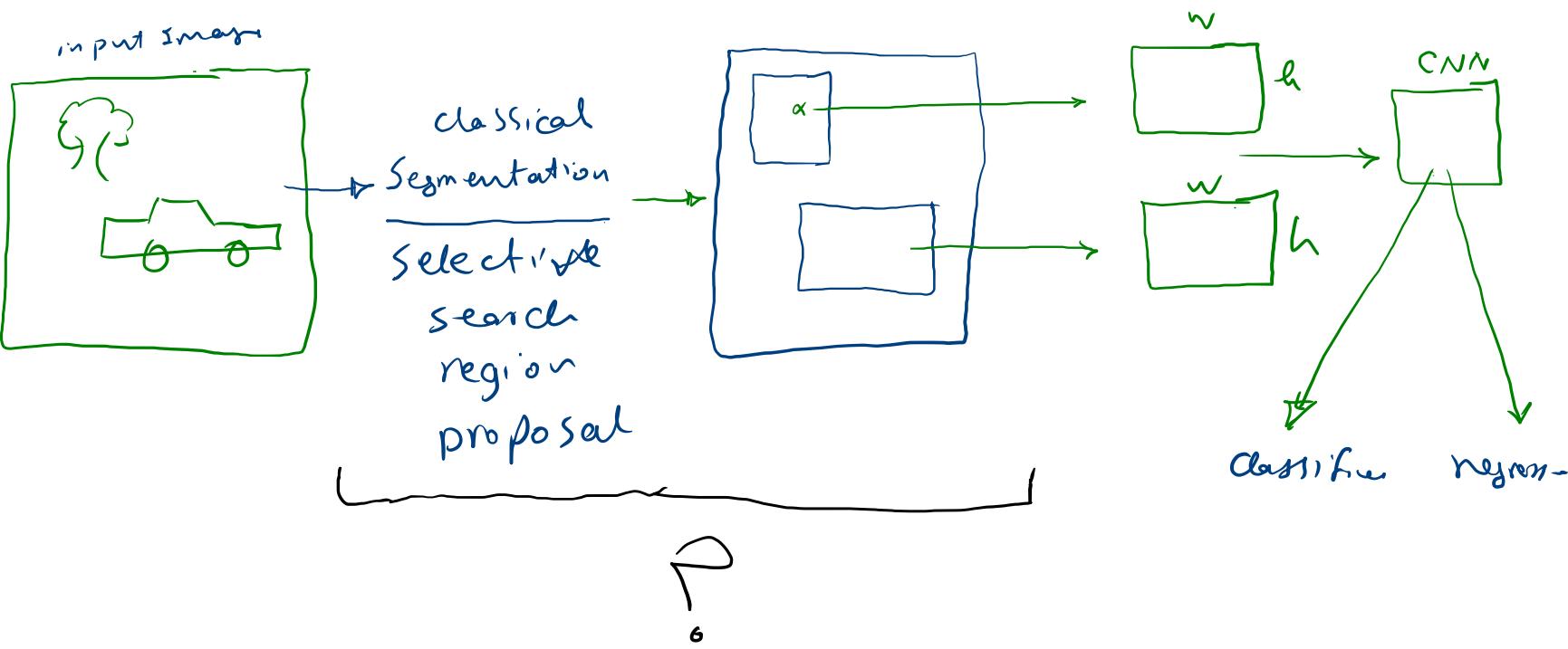
$\text{! } \text{cur} \leftarrow \text{computationnel } \text{bus}$   
 $\text{! } \text{cur} \leftarrow \text{cost}$

real time  $\rightarrow$  inference  $\leftarrow$  object detection; inference  $\leftarrow$   
 $\text{! } \text{cur}$   $\downarrow$   
inference time

! process 1 frame  $\leftarrow$  object detector

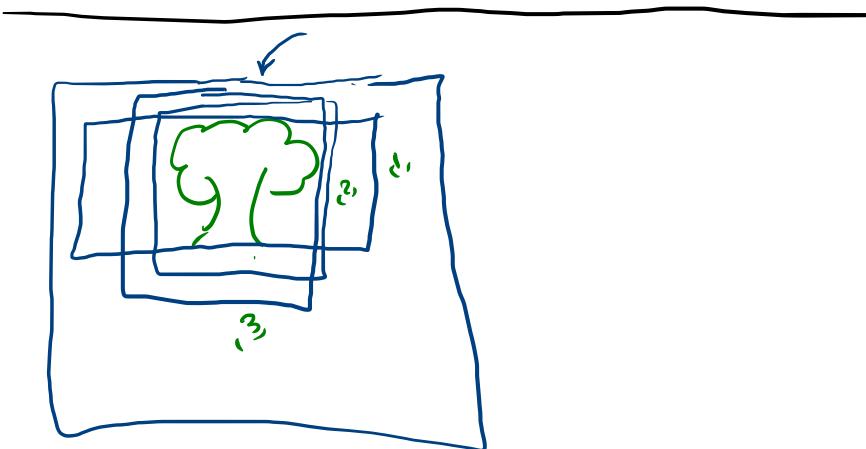
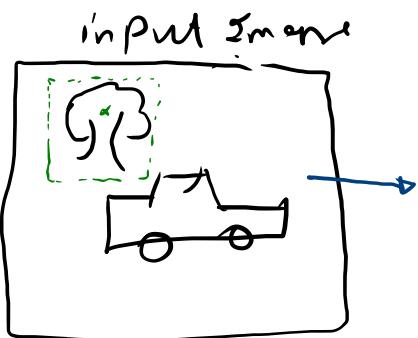
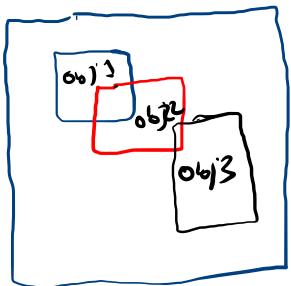
? Rel

---



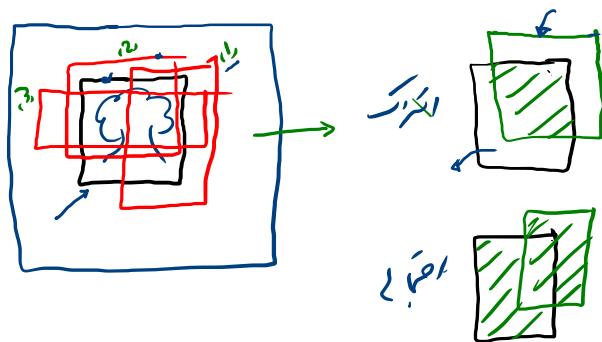
# Selective Search region proposal algorithm

greedy



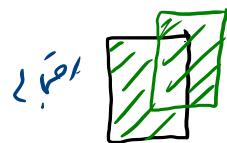
IoU = Intersection of Unions

$\frac{\text{NMS}}{\text{Non maximal suppression!}}$

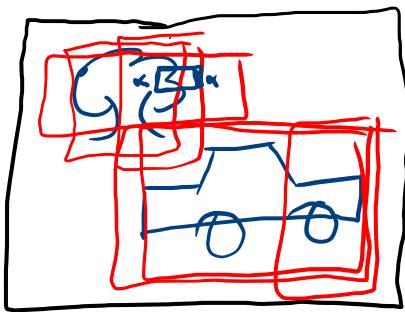
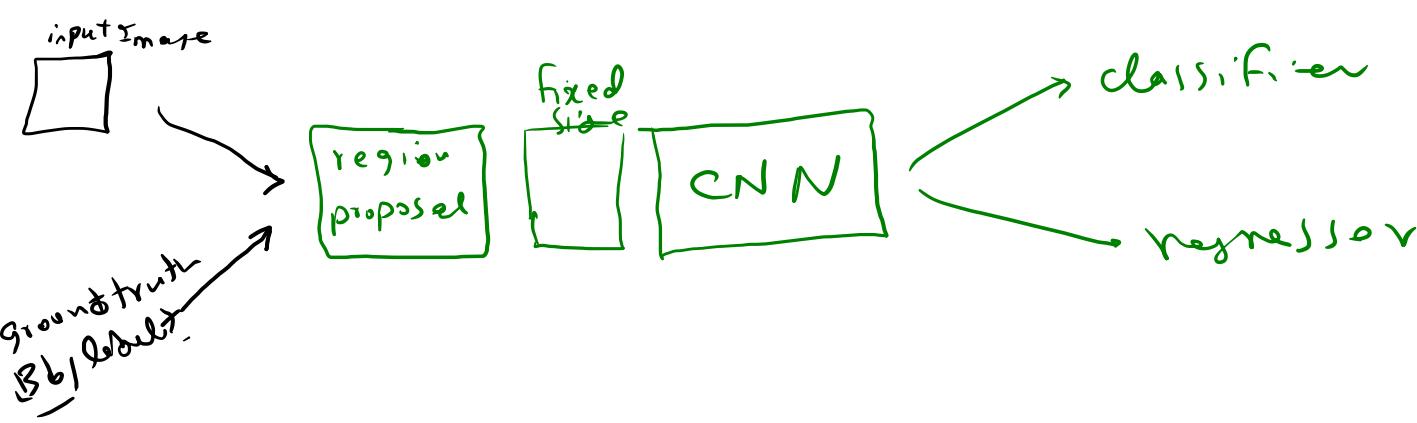


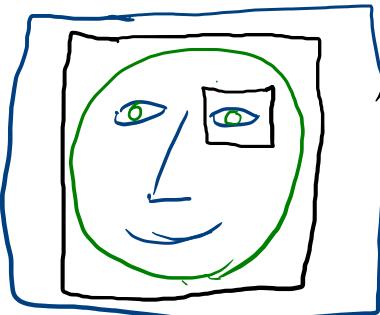
IoU =

$$\frac{\text{أثر المربع المختار}}{\text{مربع المختار} + \text{المربع المطرد}} \quad \begin{cases} \text{IoU}(1) : 0.7 \\ \text{IoU}(2) : 0.8 \\ \text{IoU}(3) : 0.3 \end{cases}$$



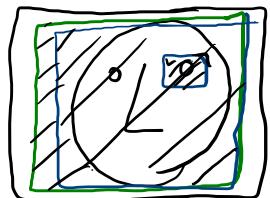
انتخاب المربع المختار  
 مراد بـ Bb مارسی بـ BS مارسی بـ BS  
 Selective search region algorithm





eye\*

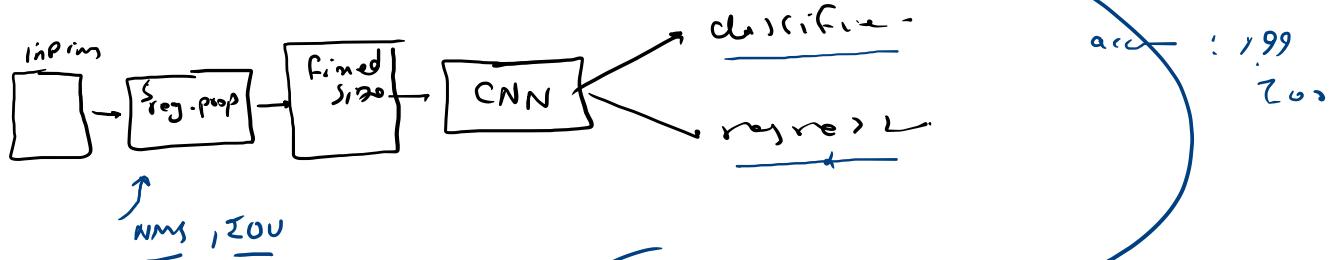
Face



Face

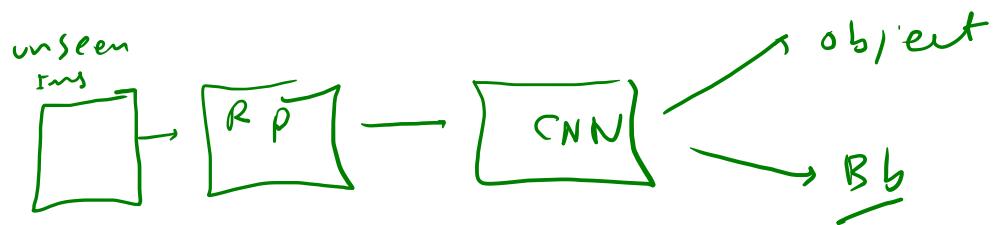
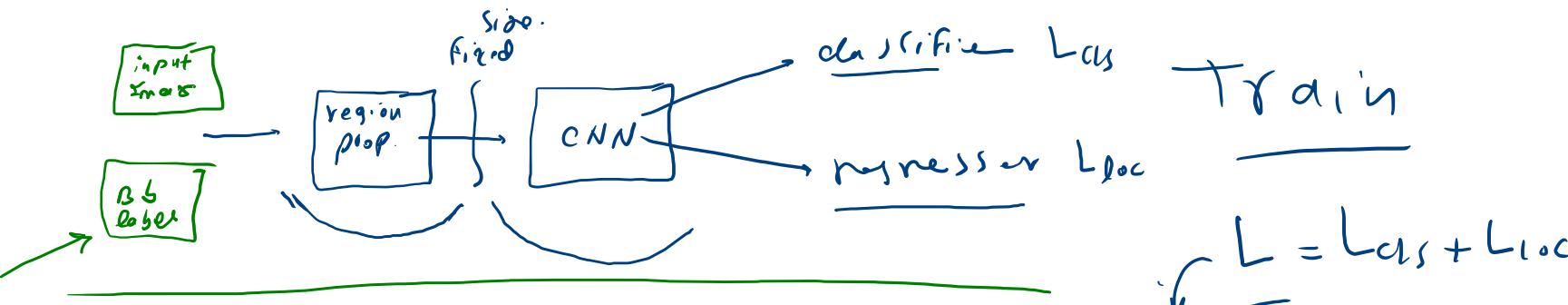
$$IoU_{\text{eye}} = \underline{\hspace{2cm}} = .1$$

$$IoU_{\text{Face}} = \underline{\hspace{2cm}} = .7$$



R-CNN

Region based Convolutional neural network



Pix2pixMLL object  
detection は どう

OD performance met "nich-"  
?

  $\rightarrow$  FPS : Frame Per Second



$$\left\{ \begin{array}{l} \text{inference time} = \frac{1s}{7} = \frac{1000ms}{7} \\ \text{inference time} \approx \underline{150ms} \end{array} \right.$$

classification +  
overlapping (regression)

② mAP : mean Average precision:

precision

PR curve

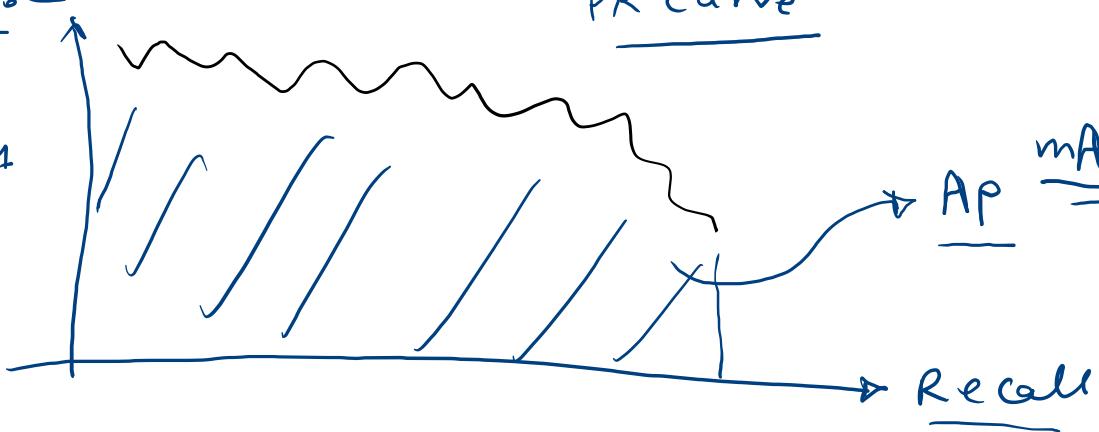
[0, 100]

class 1

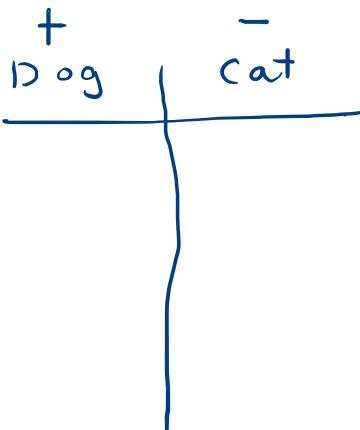
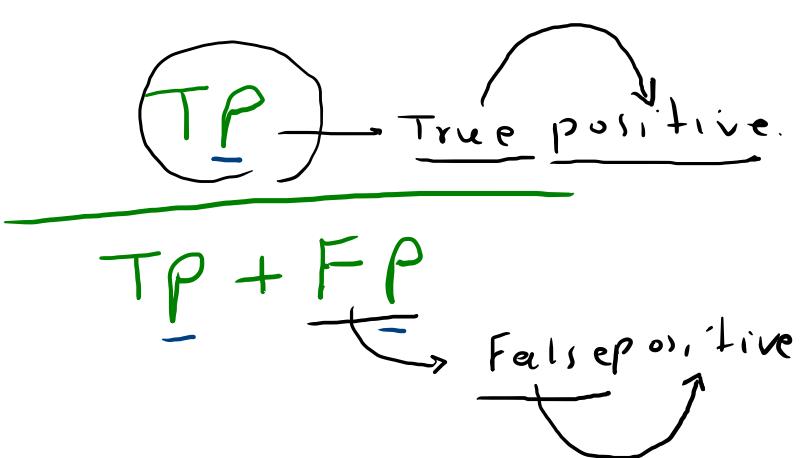
AP

$$\underline{\underline{mAP}} = \frac{AP_1 + AP_2 + \dots + AP_N}{N}$$

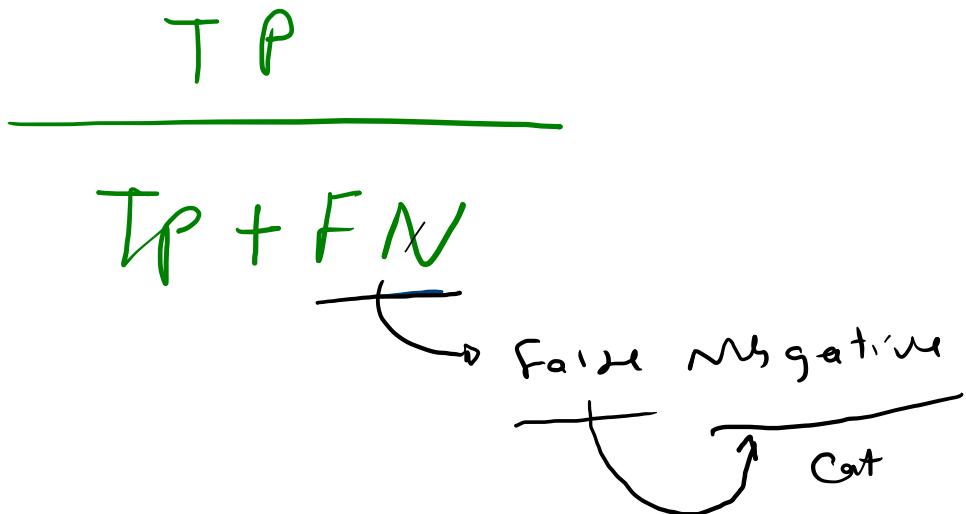
Recall



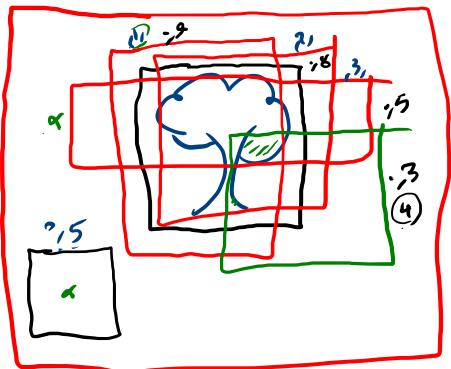
precision =



recall =



+ : Tree  
- : background



Tree  $\rightarrow$  1

? Car, 1, 5  $\rightarrow$  1

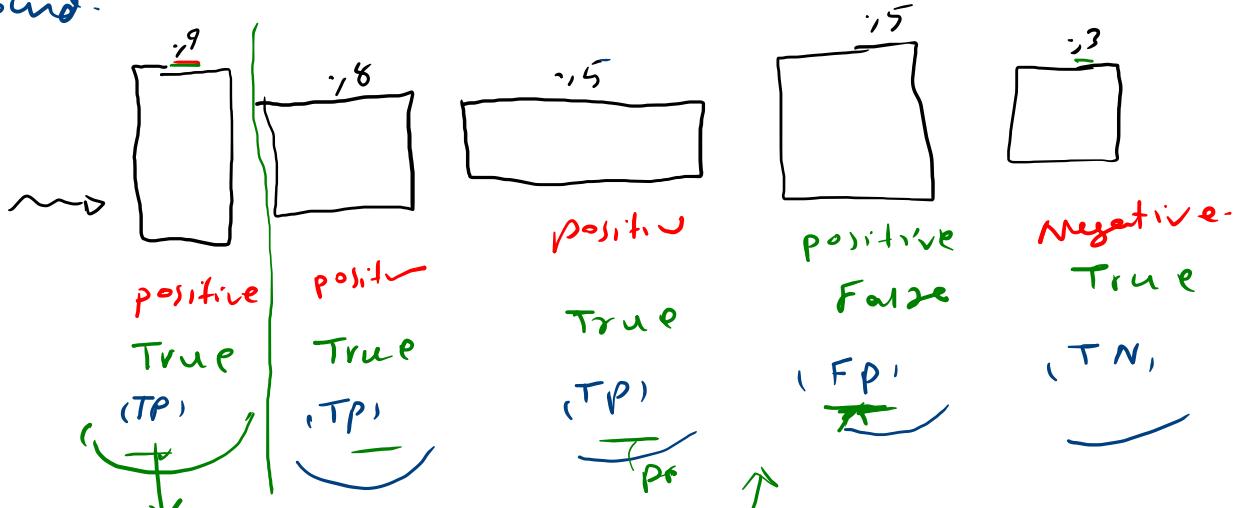
positive  $\rightarrow$

Negative  $\rightarrow$

? Car, 1, 5  $\rightarrow$  1

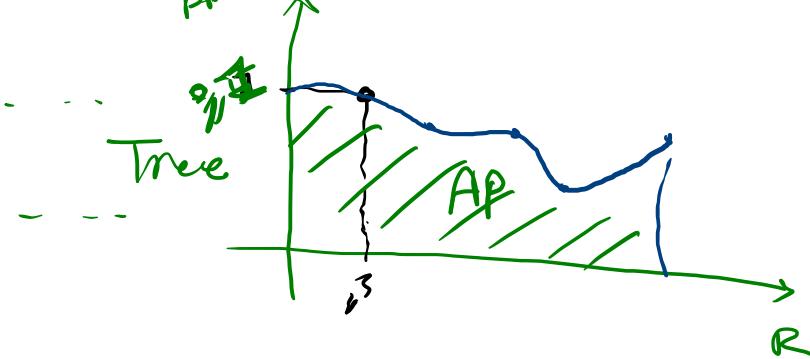
True  $\rightarrow$

False  $\rightarrow$



$$P = \frac{1}{1+0} = 1$$

$$R = \frac{1}{3} = 0,3$$



$$\sqrt{L} \Rightarrow L_{cls} = CE$$

$$\therefore L_{reg: loc}$$

predicted  $B_b$

ground truth  $B_b$

$$L_{loc} = \sum L_{tf \text{ smooth}}(t_i - v_i)$$

$$L_{tf \text{ smooth}}(t) \left\{ \begin{array}{ll} 0,5x^2 & |x| < 1 \\ |x| - 0,5 & \text{otherwise} \end{array} \right.$$

RCNN → Fast RCNN → Faster RCNN

25X

250X

Test time → 50s , 2s , 2s  
per frame

The End

