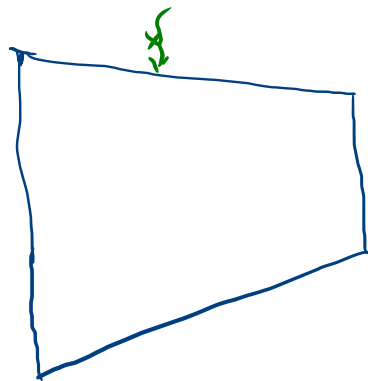
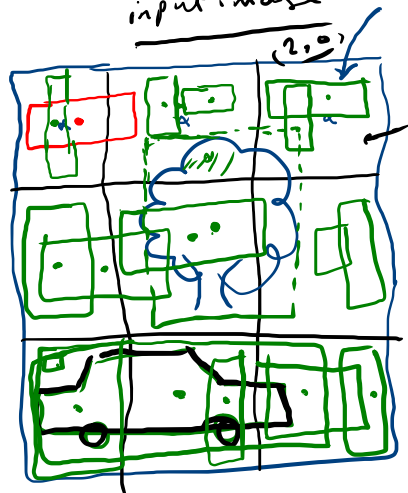


halo v5



2

input image



Num - window : 3x3

Num - Anchor boxes : Num - window x 2

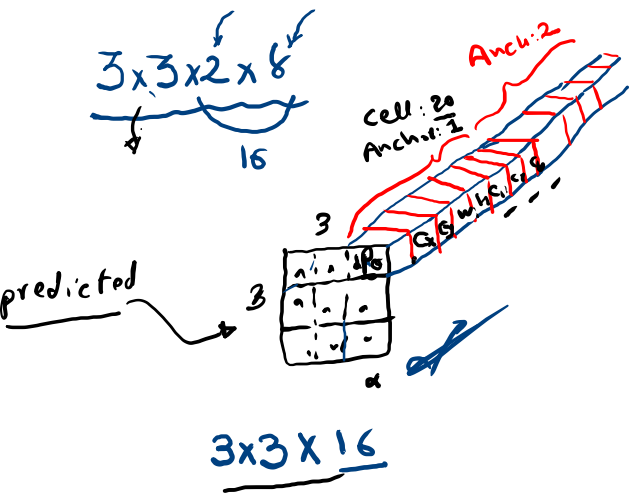
Num - category = 2 : Tree, car

Anchor boxes coordinates :  $\{ \underbrace{p_0}_{x}, \underbrace{c_x, c_y, w, h}_{x}, \underbrace{C_1, C_2, C_b}_{x} \}$

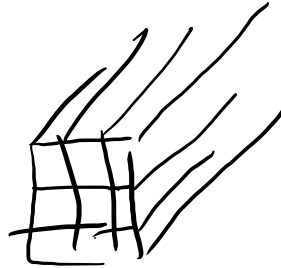
3x3x2x8

! استفاد

5 + Num - cl



GT



loss :

classification loss  $\rightarrow a, c_2$

location loss :  $MSE \rightarrow \begin{matrix} MSE((c_x - c'_x), (c_y, c'_y)) \\ w, h \\ \dots \end{matrix}$

Objective loss  $\rightarrow BCE \rightarrow P_0$

$$loss_t = \lambda_1 Loss_{cls} + \lambda_2 loss_{loc} + \lambda_3 Loss_o$$

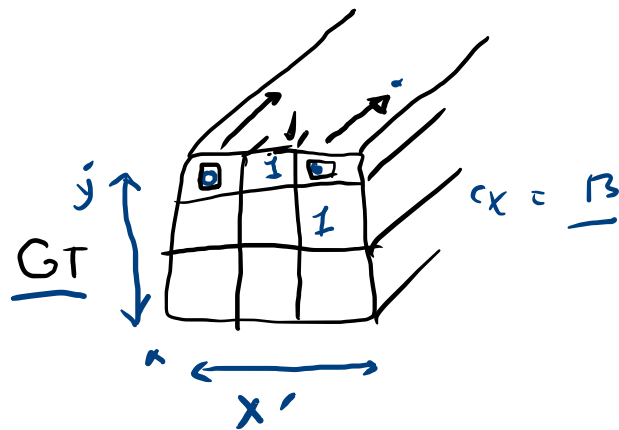
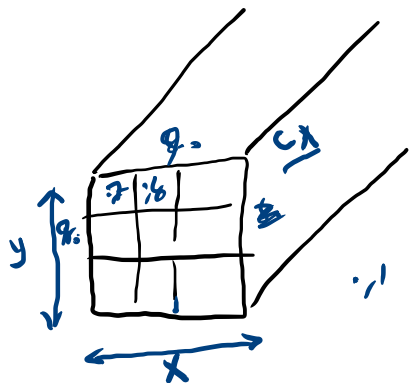
$P_o$  : window , object اضداد

$C_b$  :  $C_1, C_2$

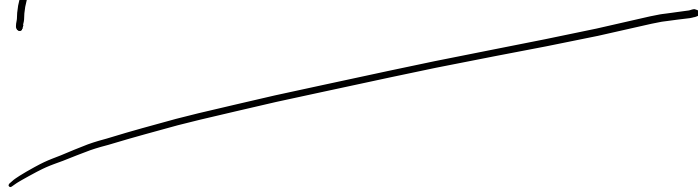
$P_o$  :  $\cdot/k$

$C_{tree} = 1/60$  ,  $C_{ca1} = 9/3$  ,  $C_b = 1/1$

pred



Yolo U5



yolo v5  $\rightarrow$   $\left| \begin{array}{l} \times \underline{\text{Conv}} \quad \alpha \\ \times \underline{\text{BN}} \quad \alpha \\ \times \underline{\text{SiLU}} \quad \alpha \end{array} \right.$

۵۵ و ۵۶

α Backbone  $\rightsquigarrow$  Feature extraction!

$\rightarrow$  استخراج ویژگی ها با اندازه مختلف!

$\rightarrow$  DarkNet 53

α Neck  $\rightsquigarrow$  Feature  $\rightarrow$  Head

در خروجی ها!

α Head  $\rightsquigarrow$  loss

Features  $\rightarrow$  شکل و رنگ را  
بشمارد و میگوید چقدر شبیه است!



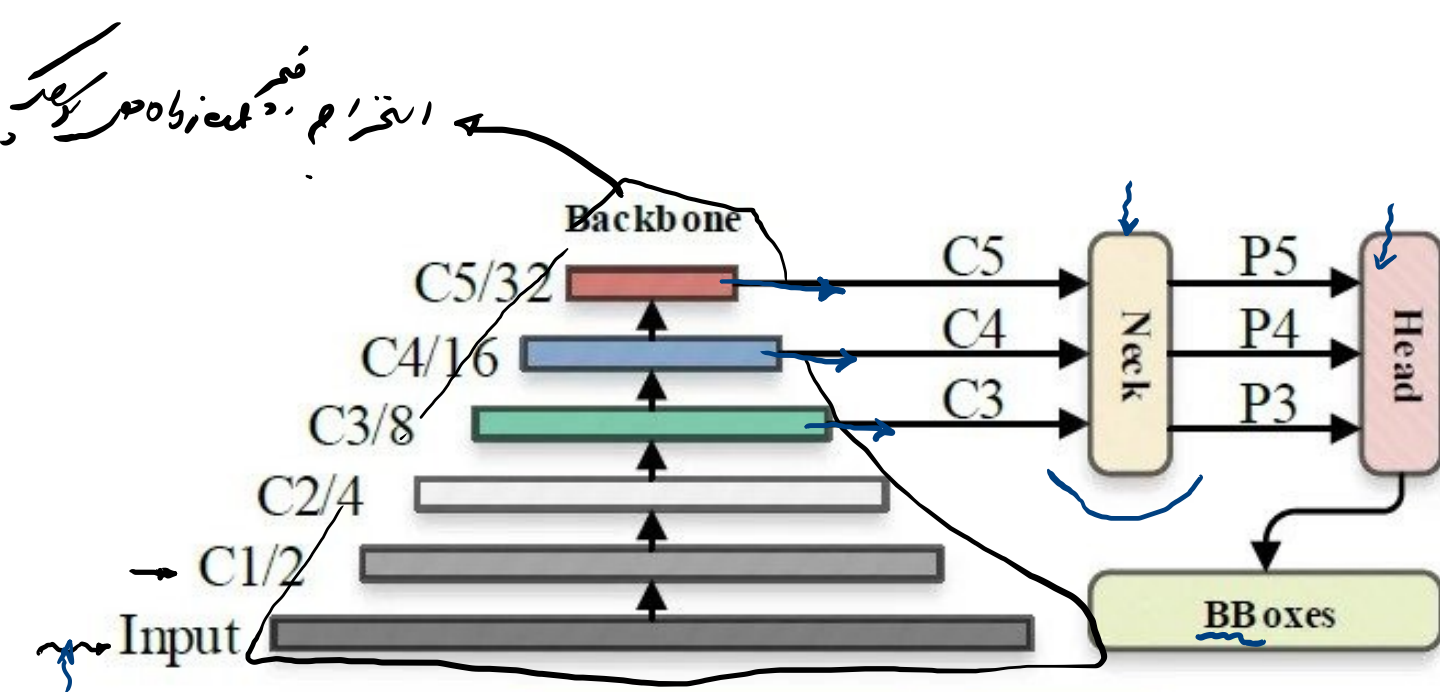
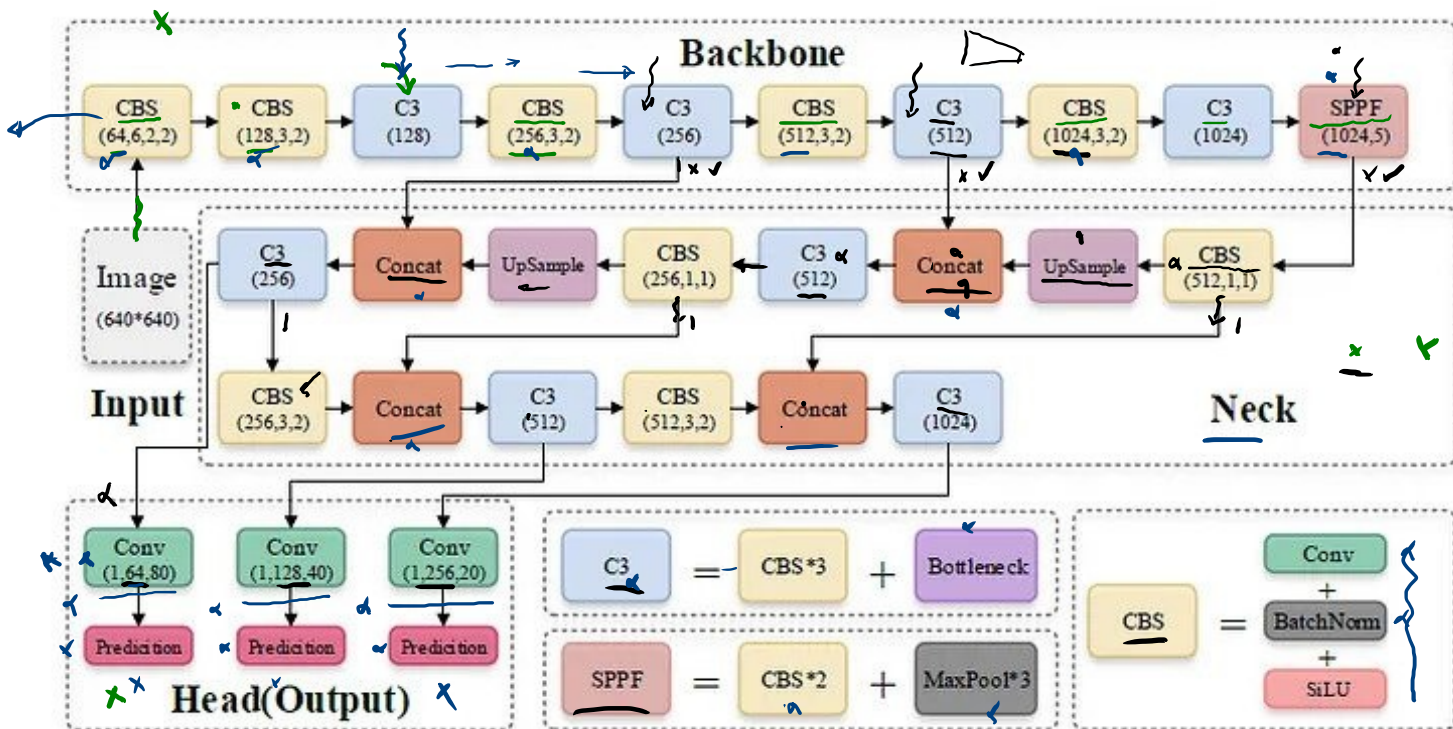
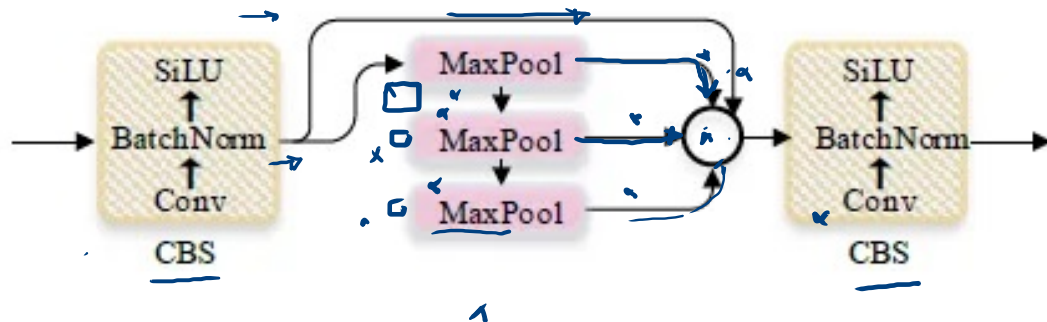
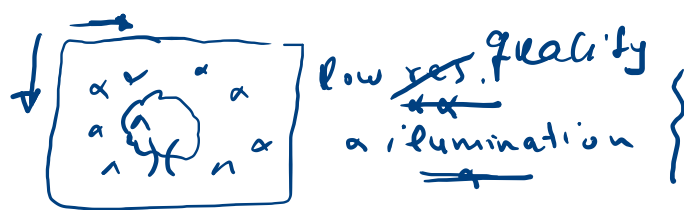


Figure 1. The default inference flowchart of YOLOv5.

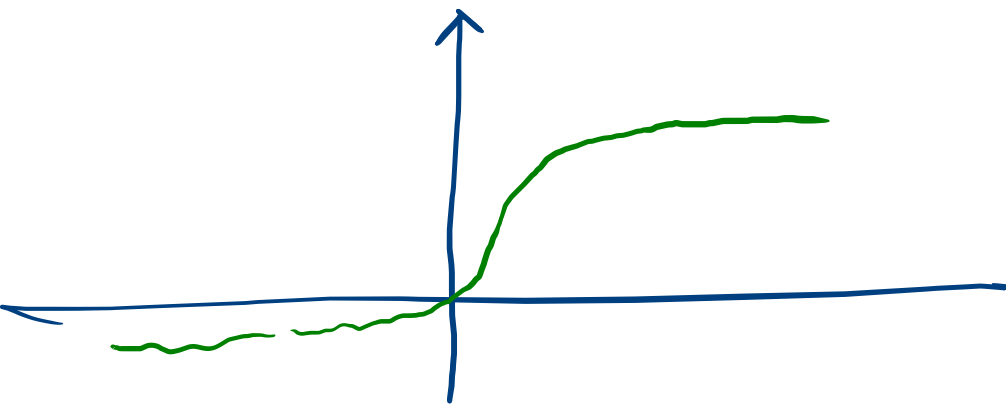


Sppf

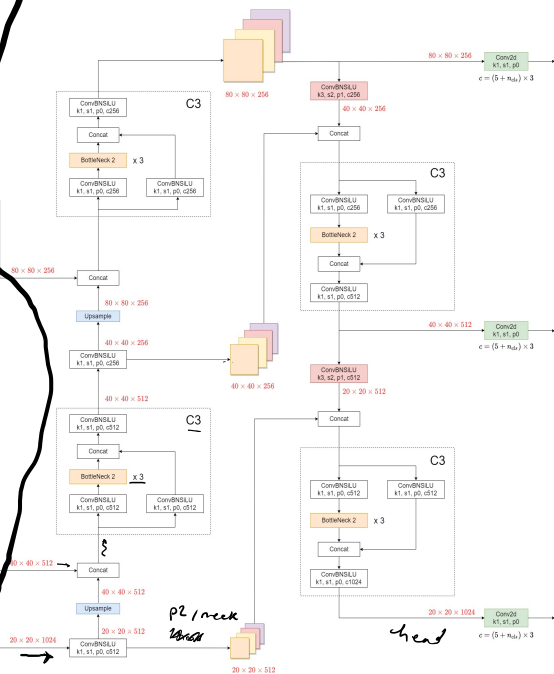
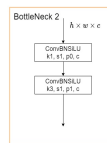
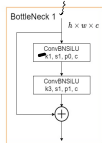


نقشه‌ی شبکه‌ی استخراج تصویر از داده‌های رادار

SILU  $\rightarrow$  SILU (X) = X 6 (X)



The diagram illustrates the YOLOv3 architecture, showing the flow from input image to final detection output. The input image is a 640 x 640 x 3 RGB image. It is processed by three parallel paths (P1, P2, P3) to generate feature maps of different resolutions. The main path (P1) consists of several stages, each containing a ConvBNReLU layer, a Bottleneck layer, and a Concat layer. The output of the main path is a 30 x 30 x 1024 feature map. The other two paths (P2 and P3) are similar but with different resolutions. The output of the main path is then processed by a SPPF layer to produce the final 30 x 30 x 1024 feature map. The diagram also shows the flow of feature maps between the different paths and the final output.



head

Conv1

$P_0$

$\rightarrow$  B BCE

Conv2

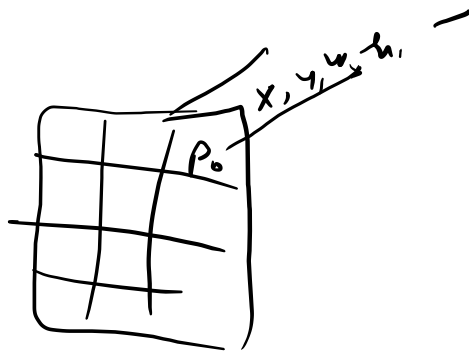
$\rightarrow$

$w, h, x, y$

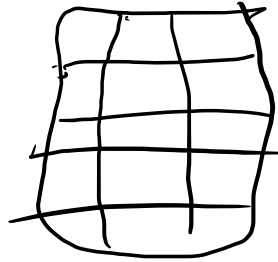
$\alpha \text{ ms } \in$

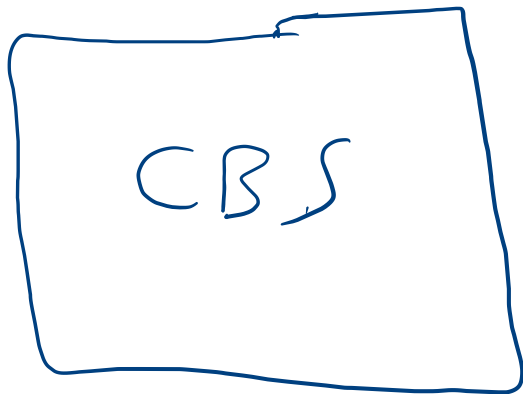
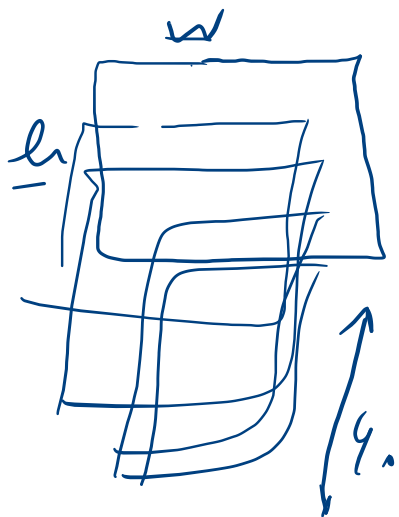
Conv3

$\rightarrow c_1, c_2, \dots, c_n \rightarrow c \in$



$sw = 20 \times 20$   
 $sw = 10, 10$   
 $sw = 8, 8$





$C_- = \frac{WX \text{ input chan}}{120}$