

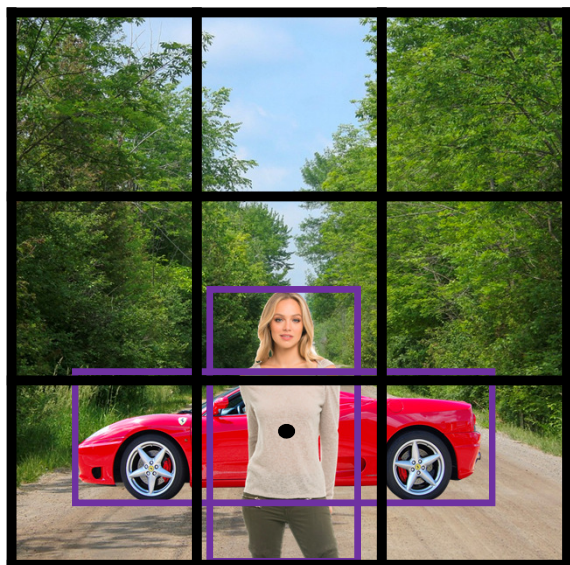


deeplearning.ai

Object Detection

Anchor boxes

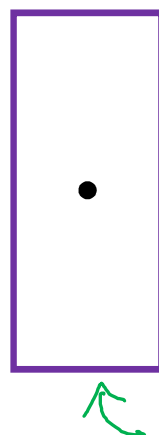
Overlapping objects:



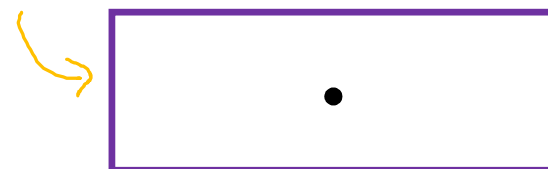
$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Handwritten annotations: A green arrow points from p_c to the center dot in the image grid. A blue arrow points from the b_x, b_y, b_h, b_w group to the purple bounding box around the woman. A blue bracket groups c_1, c_2, c_3 .

Anchor box 1:



Anchor box 2:



$y =$

p_c	Anchor box 1
b_x	
b_y	
b_h	
b_w	
c_1	Anchor box 2
c_2	
\vdots	
c_3	

Handwritten annotations: A green box groups p_c, b_x, b_y, b_h, b_w . An orange box groups c_1, c_2, \dots, c_3 . Blue brackets on the right group the rows for Anchor box 1 and Anchor box 2.

[Redmon et al., 2015, You Only Look Once: Unified real-time object detection]

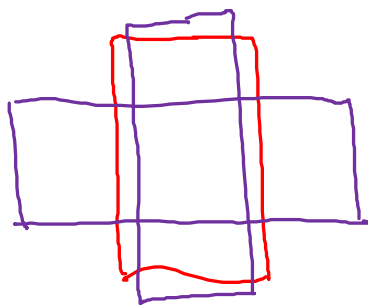
Andrew Ng

Anchor box algorithm

Previously:

Each object in training image is assigned to grid cell that contains that object's midpoint.

Output y :
 $3 \times 3 \times 8$



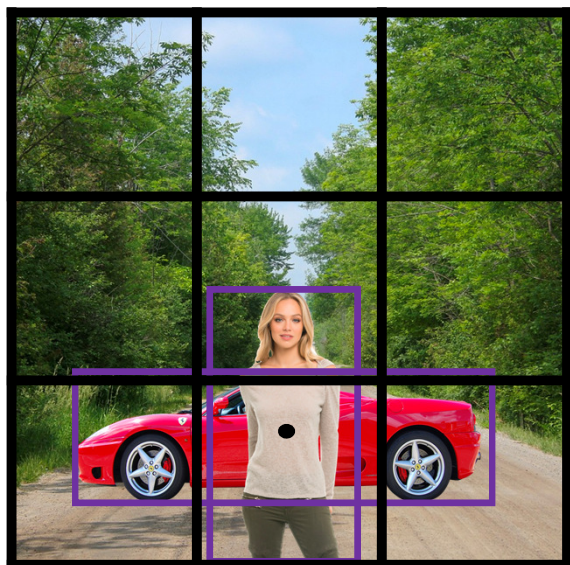
With two anchor boxes:

Each object in training image is assigned to grid cell that contains object's midpoint and anchor box for the grid cell with highest IoU.

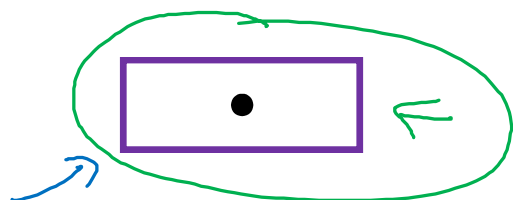
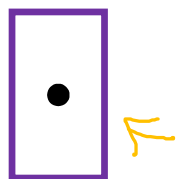
(grid cell, anchor box)

Output y :
 $3 \times 3 \times 16$
 $3 \times 3 \times 2 \times 8$

Anchor box example



Anchor box 1: Anchor box 2:



$y =$

$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Handwritten values for the first vector (orange and green):

$$\begin{bmatrix} 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 1 \\ 0 \\ 0 \\ 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Handwritten values for the second vector (green and blue):

$$\begin{bmatrix} 1 \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Labels for the second vector: "only?" above the first 8 elements, "anchor box 1" for the first 8 elements, and "anchor box 2" for the last 8 elements.