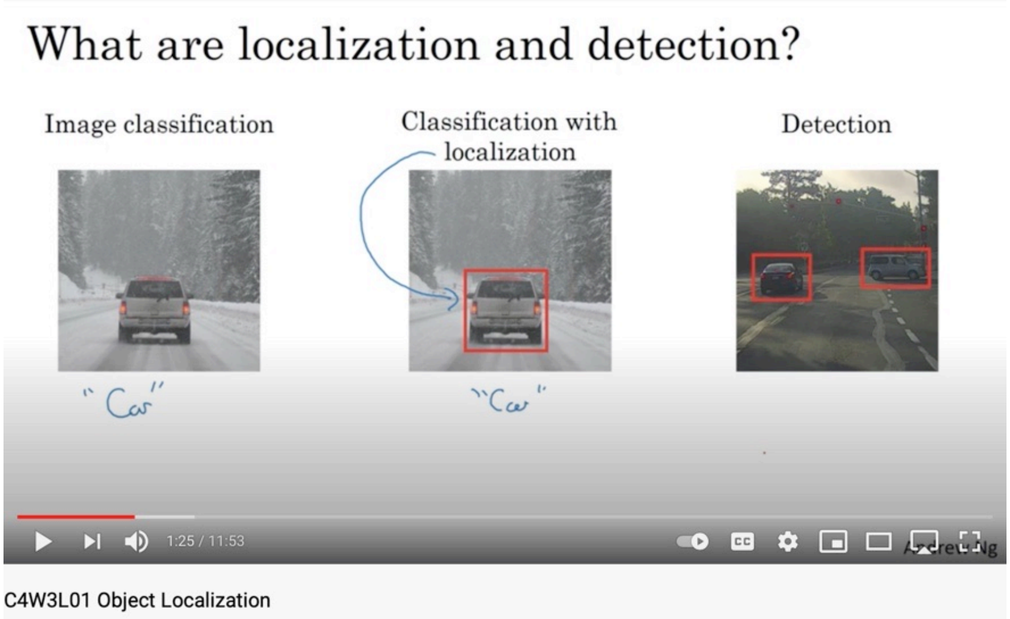


What are localization and detection?

Image classification

"Car"

Classification with localization

"Car"

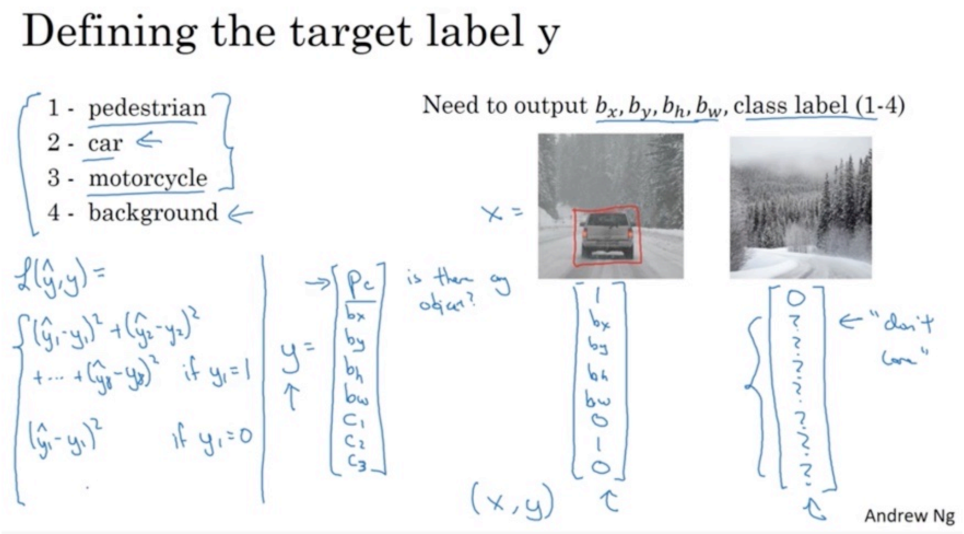
Detection

Andrew Ng

C4W3L01 Object Localization

What is the difference between these problems?

Classification with Localization:



Defining the target label y

1 - pedestrian
2 - car
3 - motorcycle
4 - background

Need to output b_x, b_y, b_h, b_w, class label (1-4)

$x =$

$$\mathcal{L}(\hat{y}, y) = \begin{cases} (\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + \dots + (\hat{y}_8 - y_8)^2 & \text{if } y_1 = 1 \\ (\hat{y}_1 - y_1)^2 & \text{if } y_1 = 0 \end{cases}$$

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

is there any object?

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

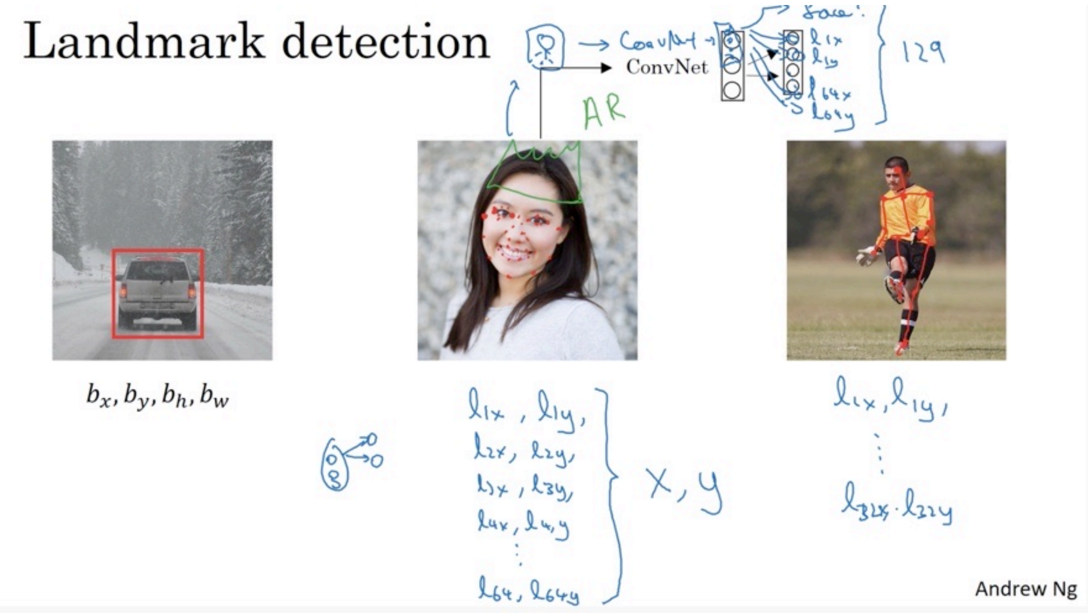
"don't care"

(x, y)

Andrew Ng

What would the loss function be if you used cross-entropy loss for classification outputs and MSE for regression outputs?

What is the theoretical underpinning of this loss?



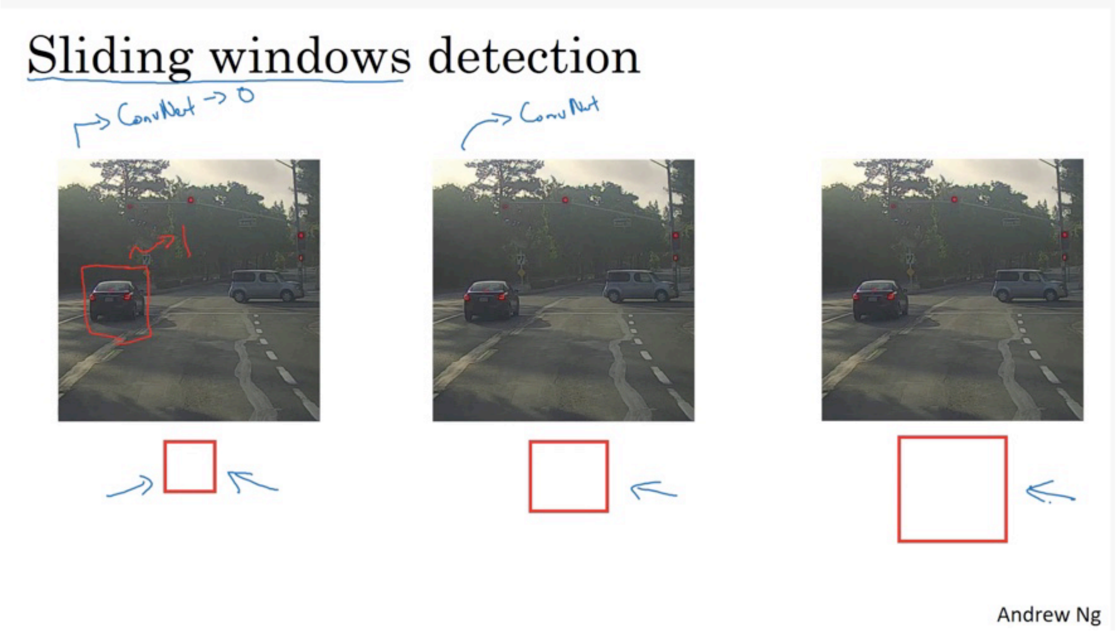
Landmark detection

ConvNet

b_x, b_y, b_h, b_w

Andrew Ng

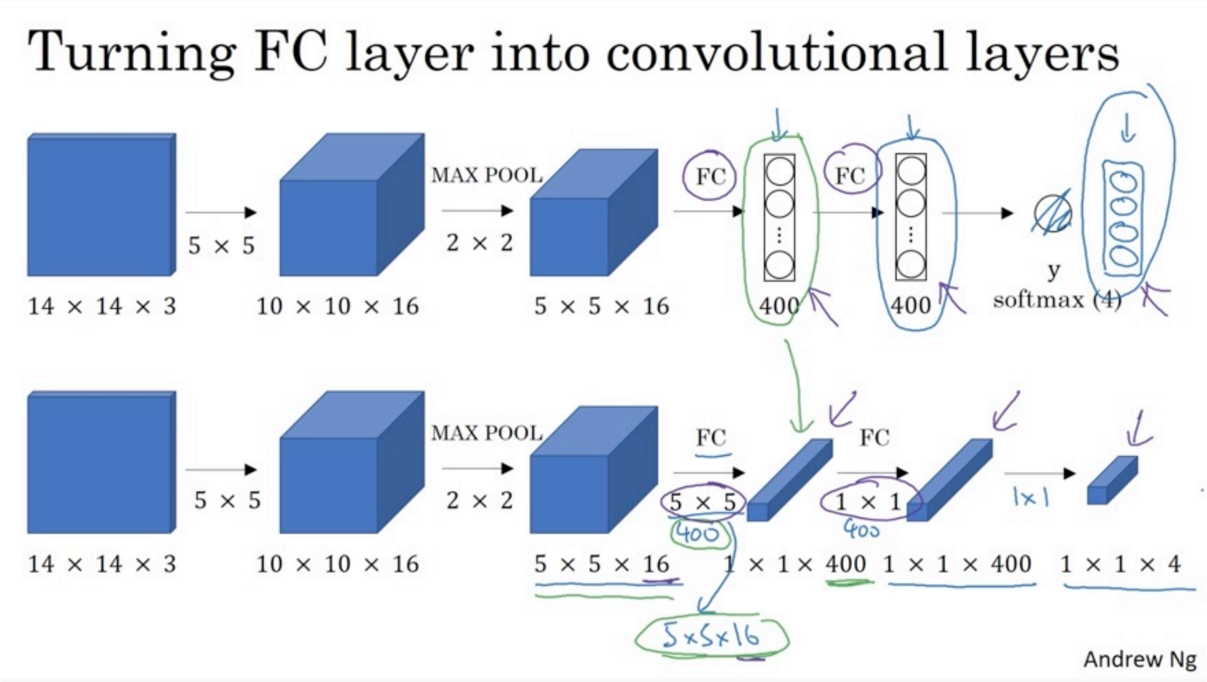
What are some concerns about doing landmark detection as described?



Sliding windows detection

Andrew Ng

How do you pass larger image fragments into the same classifier?

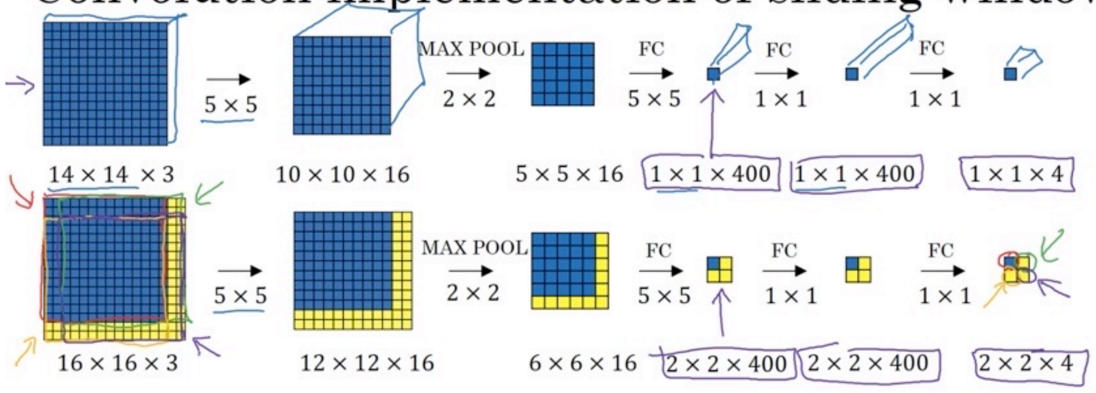


Turning FC layer into convolutional layers

Andrew Ng

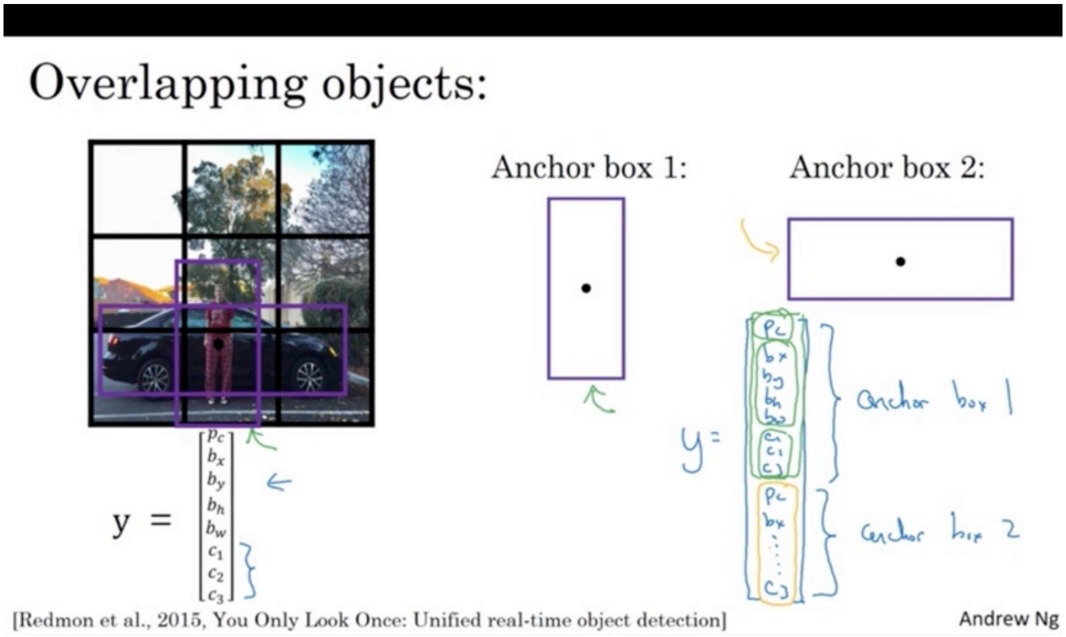
By changing these fully connected layers into convolutional layers, what does that now allow us to do?

Convolution implementation of sliding windows



[Sermanet et al., 2014, OverFeat: Integrated recognition, localization and detection using convolutional networks]

Andrew Ng



Overlapping objects:

Anchor box 1:

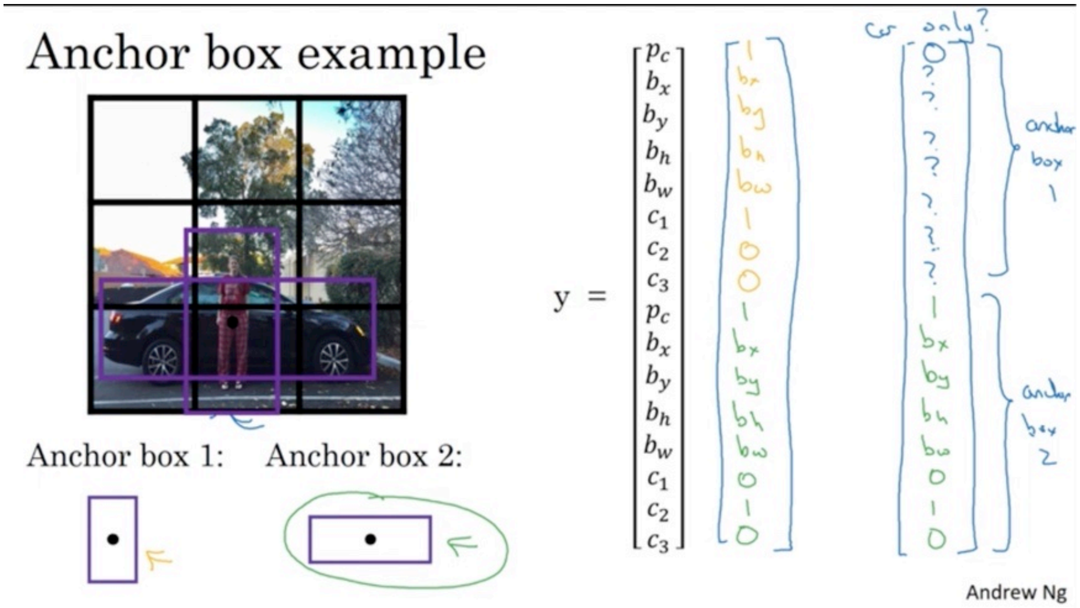
Anchor box 2:

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

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ \vdots \\ c_3 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{anchor box 1} \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \text{anchor box 2} \end{matrix}$$

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

Andrew Ng



Anchor box example

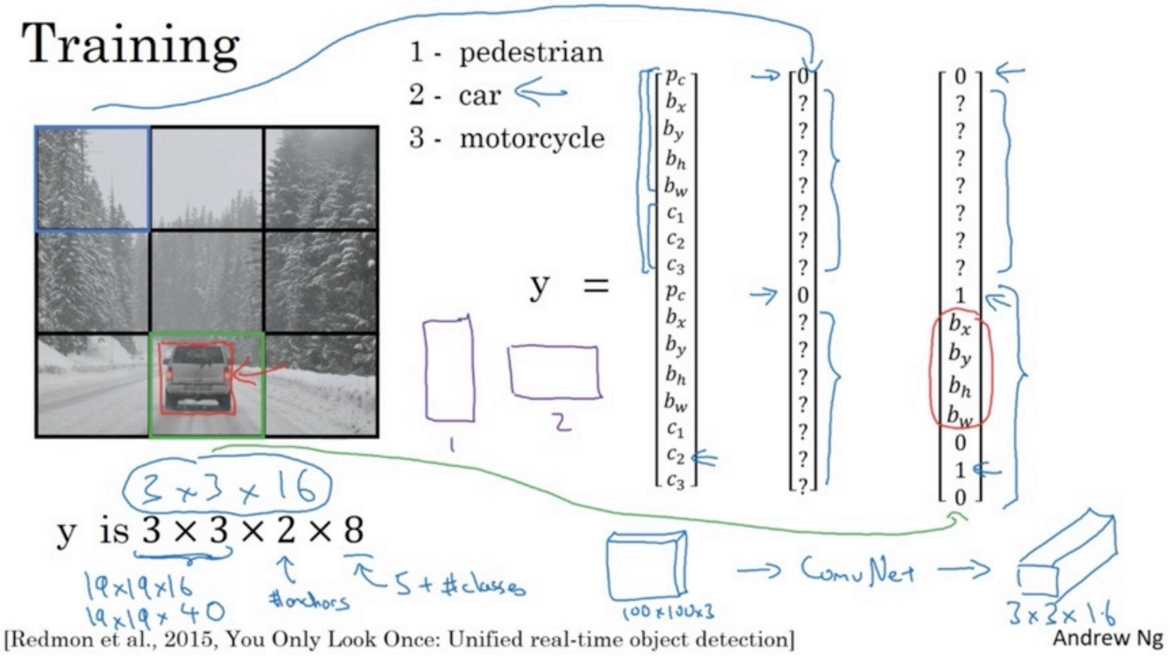
Anchor box 1: Anchor box 2:

car only?

$$y = \begin{array}{c|c|c} p_c & 1 & 0 \\ b_x & b_x & ? \\ b_y & b_y & ? \\ b_h & b_h & ? \\ b_w & b_w & ? \\ c_1 & 1 & ? \\ c_2 & 0 & ? \\ c_3 & 0 & ? \\ p_c & 1 & 1 \\ b_x & b_x & b_x \\ b_y & b_y & b_y \\ b_h & b_h & b_h \\ b_w & b_w & b_w \\ c_1 & 0 & 0 \\ c_2 & 1 & 1 \\ c_3 & 0 & 0 \end{array} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{anchor box 1} \\ \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{anchor box 2} \end{matrix}$$

Andrew Ng

YOLO Training



Training

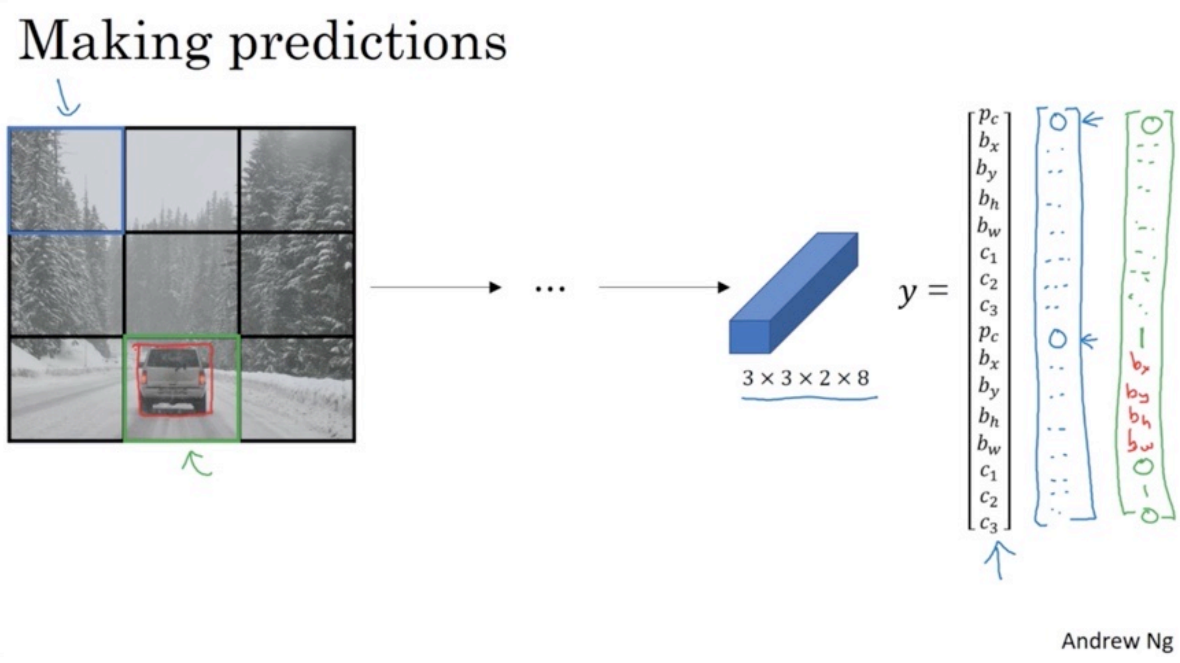
1 - pedestrian
2 - car
3 - motorcycle

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} \quad \begin{bmatrix} 0 \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ 0 \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \end{bmatrix} \quad \begin{bmatrix} 0 \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

y is $3 \times 3 \times 2 \times 8$

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

Andrew Ng



Making predictions

$3 \times 3 \times 2 \times 8$

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

Andrew Ng