



deeplearning.ai

## Object Detection

## Object localization

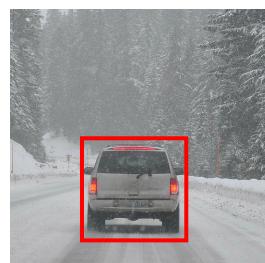
1

## What are localization and detection?

Image classification



Classification with  
localization



Detection

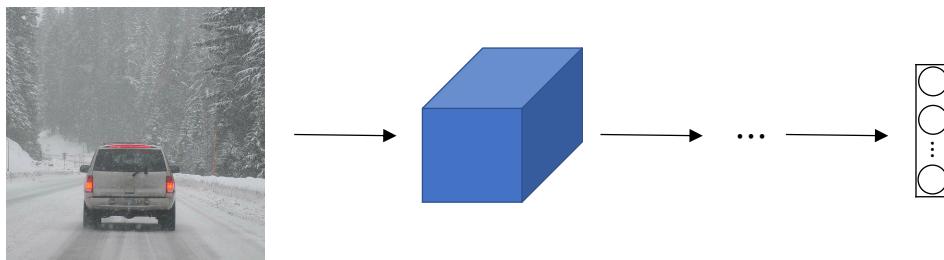


Andrew Ng

2

1

## Classification with localization



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

Andrew Ng

3

## 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)



Andrew Ng

4



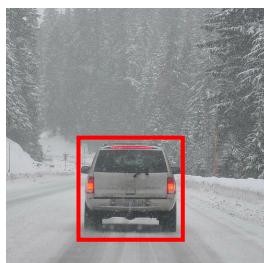
deeplearning.ai

# Object Detection

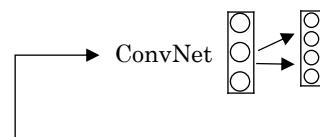
## Landmark detection

5

### Landmark detection



$b_x, b_y, b_h, b_w$



Andrew Ng

6

3



deeplearning.ai

## Object Detection

### Object detection

7

## Car detection example

Training set:



x

y

1

1

1

0

0

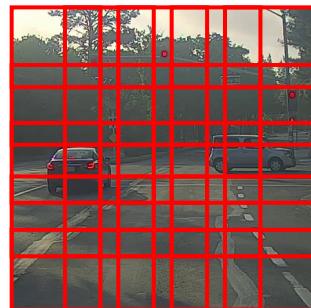


Andrew Ng

8

4

## Sliding windows detection



Andrew Ng

9



deeplearning.ai

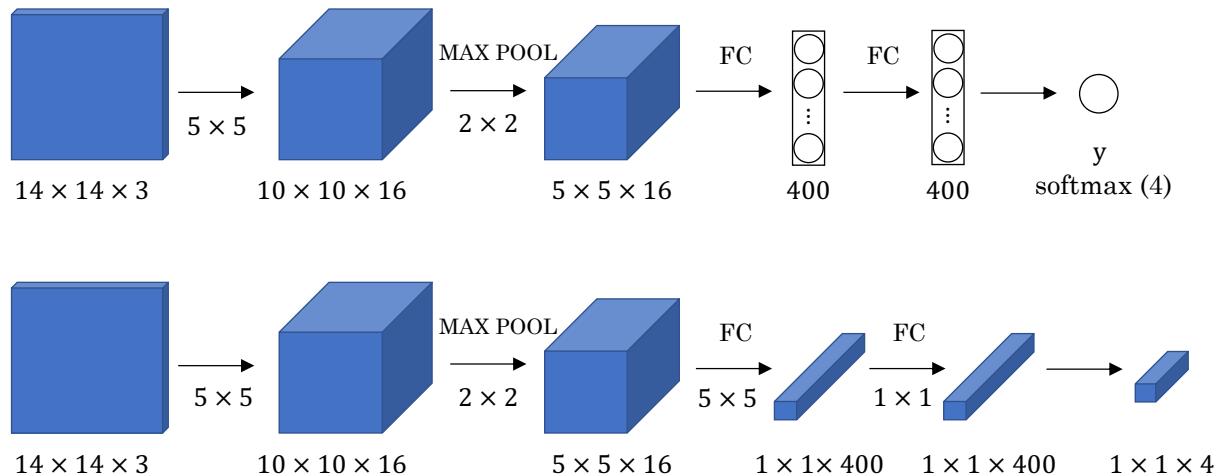
## Object Detection

---

### Convolutional implementation of sliding windows

10

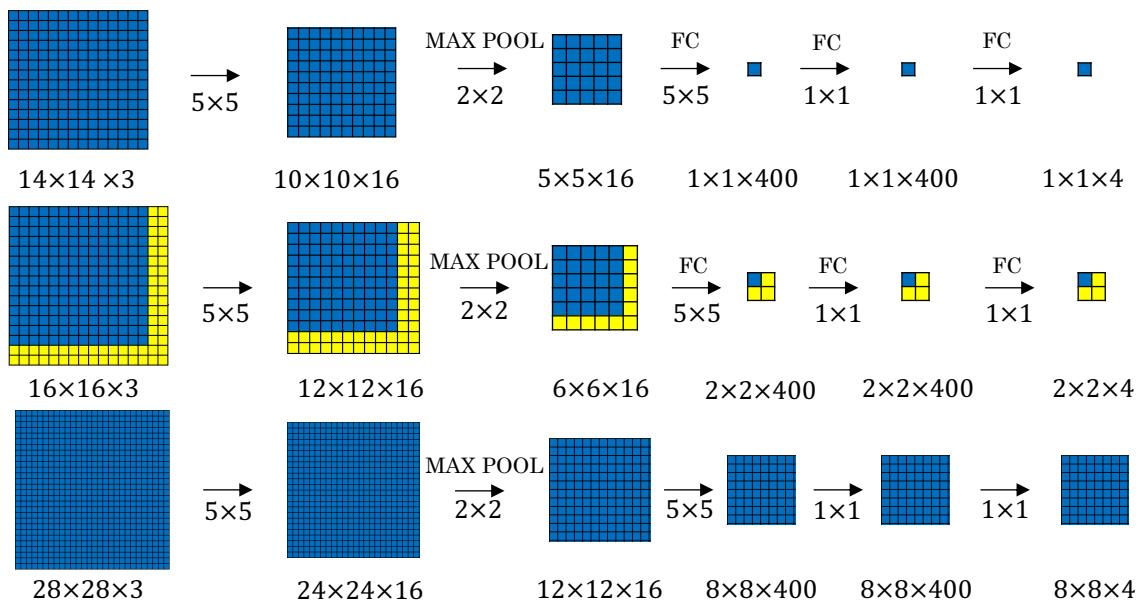
## Turning FC layer into convolutional layers



Andrew Ng

11

## Convolution implementation of sliding windows

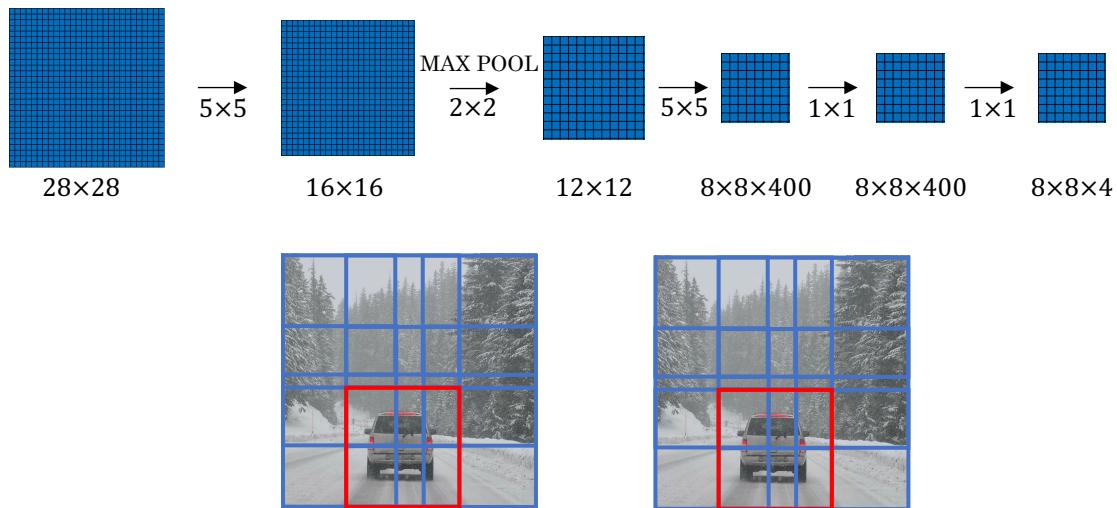


[Sermanet et al., 2014, OverFeat: Integrated recognition, localization and detection using convolutional networks]

Andrew Ng

12

## Convolution implementation of sliding windows



Andrew Ng

13



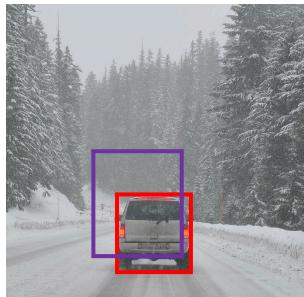
deeplearning.ai

## Object Detection

Intersection  
over union

14

## Evaluating object localization



“Correct” if  $\text{IoU} \geq 0.5$

More generally, IoU is a measure of the overlap between two bounding boxes.

Andrew Ng

15



deeplearning.ai

## Object Detection

## Non-max suppression

16

## Non-max suppression example



Andrew Ng

17

## Non-max suppression example

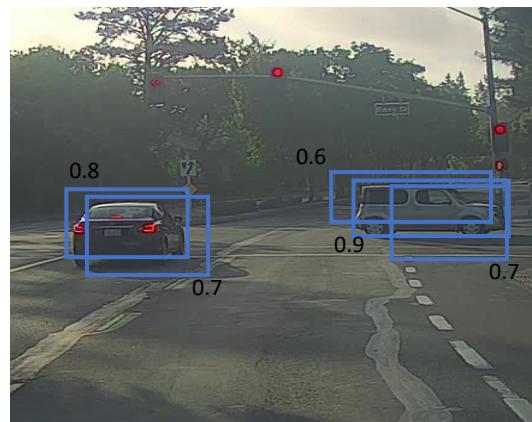


19x19

Andrew Ng

18

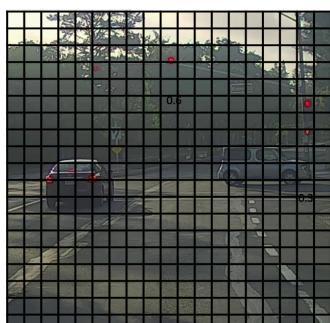
## Non-max suppression example



Andrew Ng

19

## Non-max suppression algorithm



Each output prediction is:

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

Discard all boxes with  $p_c \leq 0.6$

While there are any remaining boxes:

- Pick the box with the largest  $p_c$   
Output that as a prediction.
- Discard any remaining box with  
 $\text{IoU} \geq 0.5$  with the box output  
in the previous step

Andrew Ng

20



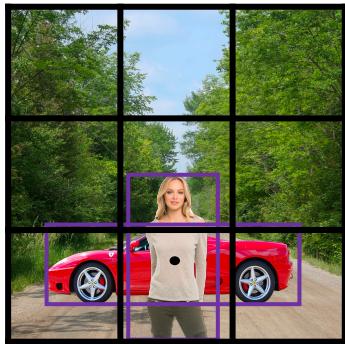
deeplearning.ai

# Object Detection

## Anchor boxes

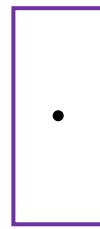
21

### Overlapping objects:

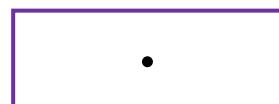


$$\mathbf{y} = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Anchor box 1:



Anchor box 2:



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

Andrew Ng

22

## Anchor box algorithm

Previously:

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

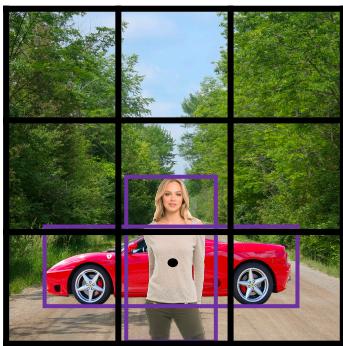
With two anchor boxes:

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.

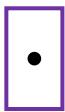
Andrew Ng

23

## Anchor box example



Anchor box 1:    Anchor box 2:



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

Andrew Ng

24



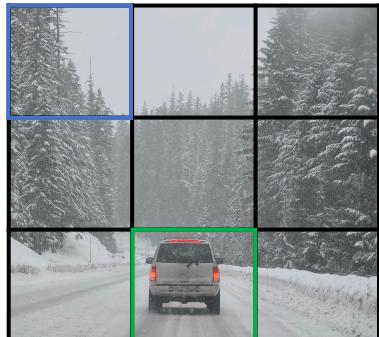
deeplearning.ai

## Object Detection

### Putting it together: YOLO algorithm

25

## Training



1 - pedestrian

2 - car

3 - motorcycle

$y =$

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

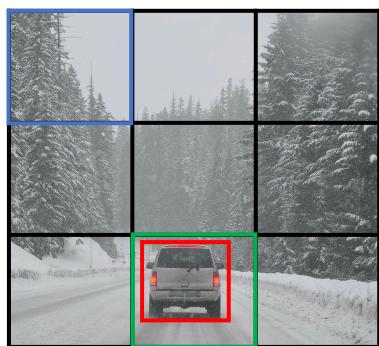
$y$  is  $3 \times 3 \times 2 \times 8$

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

Andrew Ng

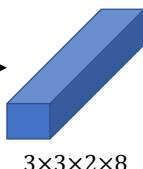
26

## Making predictions



→

...

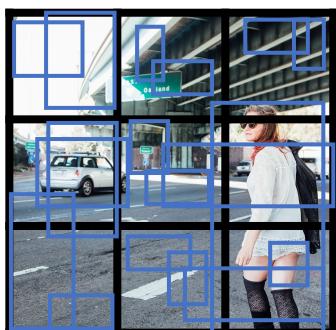
 $y =$ 

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

Andrew Ng

27

## Outputting the non-max suppressed outputs



- For each grid call, get 2 predicted bounding boxes.
- Get rid of low probability predictions.
- For each class (pedestrian, car, motorcycle) use non-max suppression to generate final predictions.

Andrew Ng

28



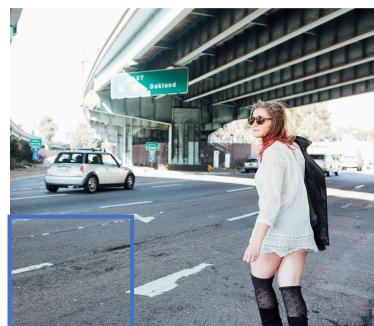
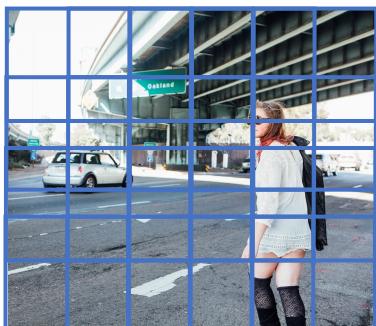
deeplearning.ai

## Object Detection

### Region proposals (Optional)

29

### Region proposal: R-CNN



[Girshik et. al, 2013, Rich feature hierarchies for accurate object detection and semantic segmentation] Andrew Ng

30

## Faster algorithms

R-CNN: Propose regions. Classify proposed regions one at a time. Output label + bounding box.

Fast R-CNN: Propose regions. Use convolution implementation of sliding windows to classify all the proposed regions.

Faster R-CNN: Use convolutional network to propose regions.

[Girshik et. al, 2013. Rich feature hierarchies for accurate object detection and semantic segmentation]

[Girshik, 2015. Fast R-CNN]

[Ren et. al, 2016. Faster R-CNN: Towards real-time object detection with region proposal networks]

Andrew Ng