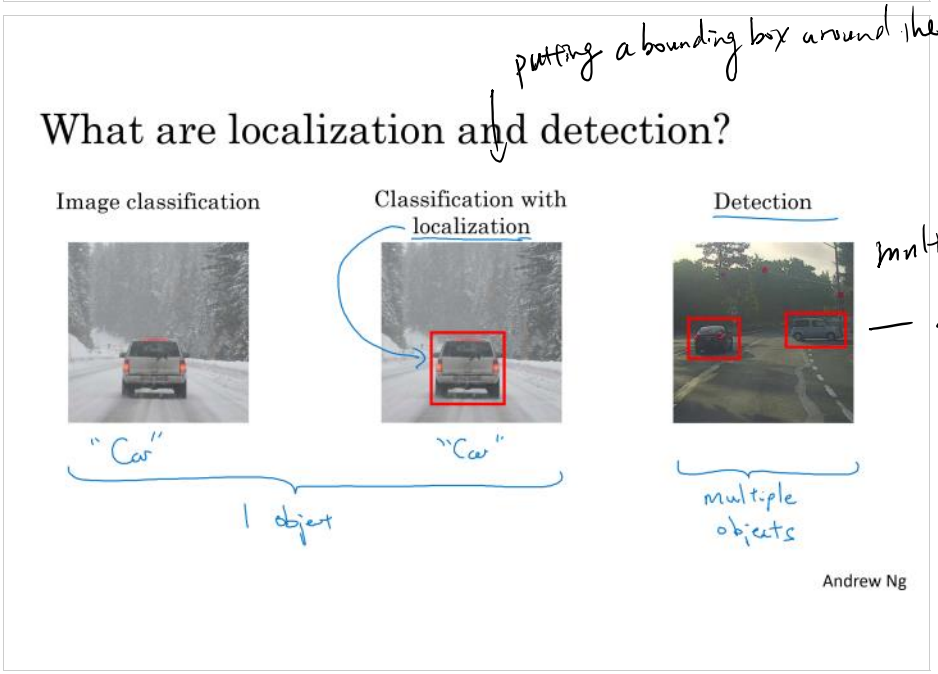
Detection Algorithm

Tuesday, June 12, 2018 1:02 PM

intro

An area that is just exploding



putting a bounding box around the position of the object

multiple objects

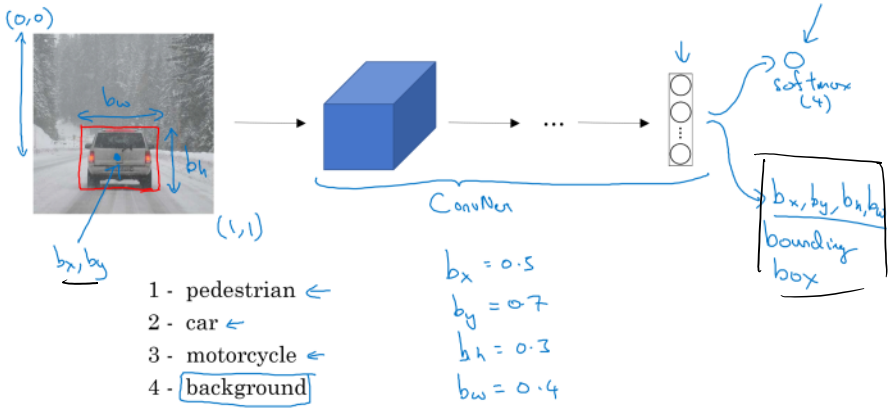
— detect and localize

for auto-driving task

detect different types of object

classification with localization

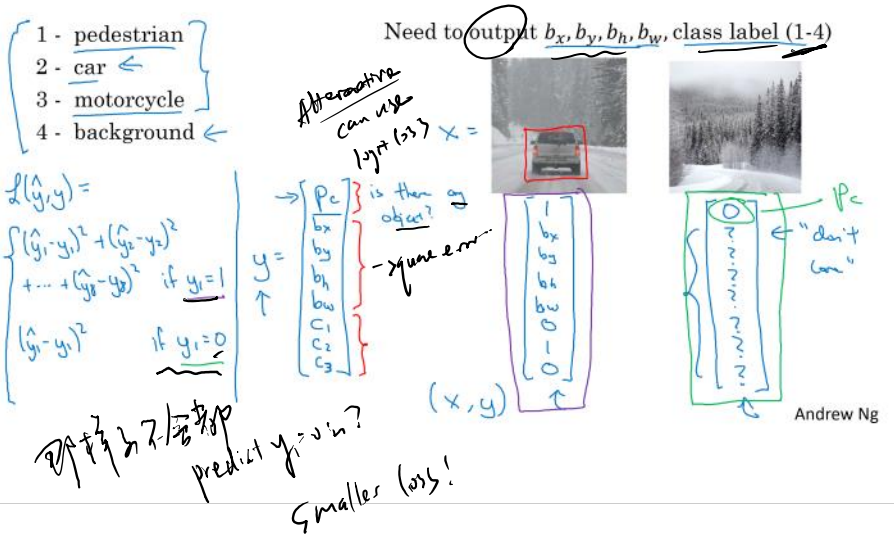
Classification with localization



Andrew Ng

要做的output { training set (class label) include the object type and the 4 variables of location.

Defining the target label y



Andrew Ng

predict $y_1 = ?$

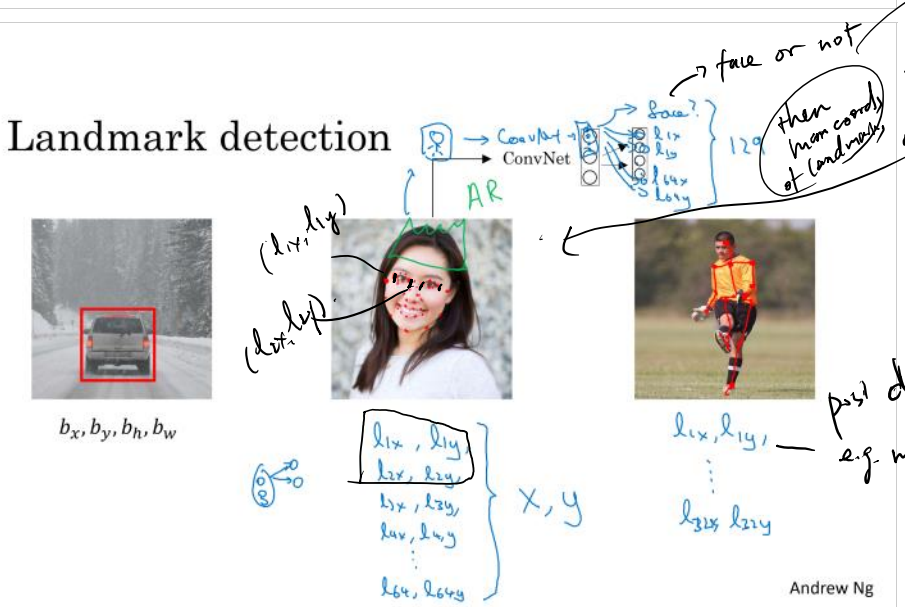
Smaller loss!

landmark

deeplearning.ai

Object Detection

Landmark detection



use: recognize facial expression etc

- face or not
- then mark coords of landmarks
- export important points of the image
- want: where's the corner in someone's eyes.

can extract many points

pose detection

e.g. midpoint of chest ...

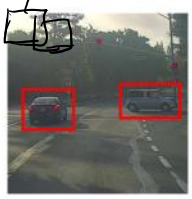
object detection

deeplearning.ai

Object Detection

Object detection

Car detection example



Training set:

x	y
	1
	1
	1
	0
	0

Andrew Ng

Sliding windows detection algorithm.

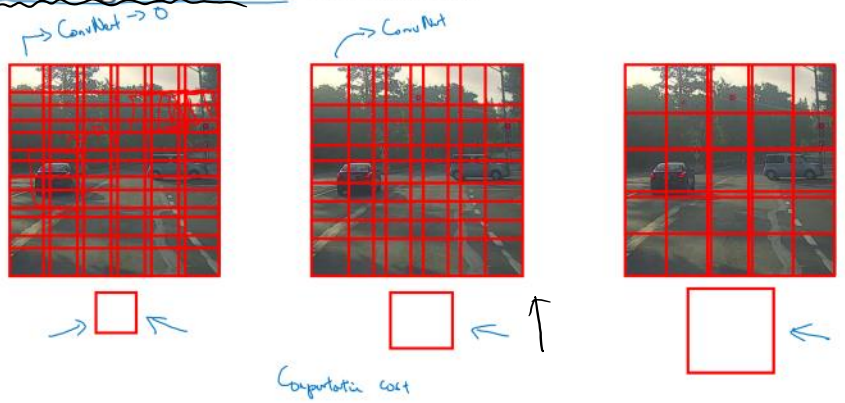
small region
— pass many crop of the image.

- - >

large region (choose a larger window)

sliding window detection

Sliding windows detection



Andrew Ng

large computational cost!

Using sliding window detection.

(before: simple classifier + sliding window)

Now: NN + sliding → unbearable computation

Next: convolutional sliding window!

Solve the problem.

window size ↑

conv sliding windows

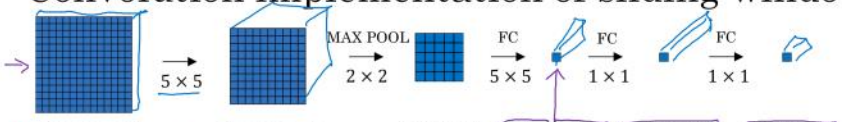
Turning FC layer into convolutional layers

Mathematically, the same as fully connected softmax activation.

4 output classes

filter size $1 \times 1 \times n_c$

Convolution implementation of sliding windows



running on the right as ✓ top left (14×14×3)

running on
→ same result as ✓ top left 14×14×3

share parameters
— share computation

→ running sliding window on top-left 14×14×3

用 conv net. 后不会再放比高地把图裁开了一次次 run:

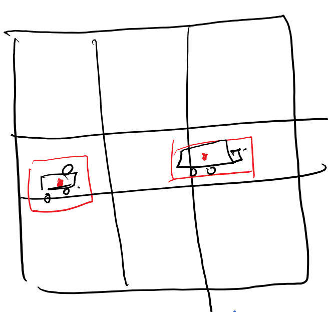
previous Sequentially.   make prediction for all regions simultaneously

个结 y 是 $\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ 之类的 吧

problem: not output the accurate bounding box.

Bounding box precision

YOLO Algorithm   You only looks once.



- Divide image into 9 grid cells
- Labels for training for each grid cell

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

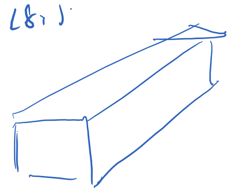
Assign the car to the grid cell containing the ...

meaning of these dimensions See the first few slides. they label the location, type of object

Assign the car to the grid cell containing the midpoint.

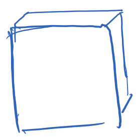
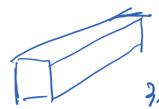
$\begin{bmatrix} b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$ $\begin{bmatrix} \vdots \\ ? \end{bmatrix}$ $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ / object

(8, 1)

output 

3×3×8 ↳ length of y

3×3 grid  → is there an object? where it is

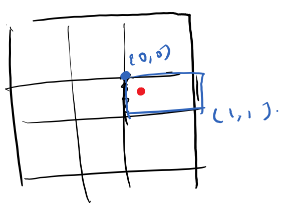
 → conv → maxpool ... →  3×3×8

As long as ≤ 1 object each grid cell, this algorithm should work fine

What's special You are NOT implementing this algorithm 9 times for each grid.

You share computation among grids — efficient

One more detail: How to code the bounding box?



$\begin{bmatrix} 1 \\ 0.4 \\ 0.3 \\ 0.9 \\ 0.5 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ b_x: the location is specified relative to the grid cell it's in.

(HC: 所以应该是0, 1之间).

there are some other parameterization

The YOLO paper is hard to read ... [warning]

intersection union

(Evaluate) Object detection model.

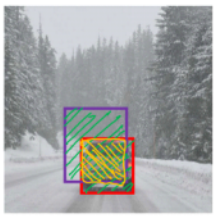
(Evaluate) Object detection model.

deeplearning.ai

Object Detection

Intersection over union

Evaluating object localization



Intersection over Union (IoU)

= Size of ▨ / Size of ▨

"Correct" if IoU ≥ 0.5 ← the rule of thumb.

0.6 ←

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

Andrew Ng

non max suppression

deeplearning.ai

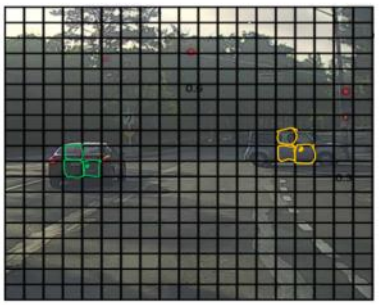
Object Detection

Non-max suppression

Non-max suppression example

Andrew Ng

Non-max suppression example

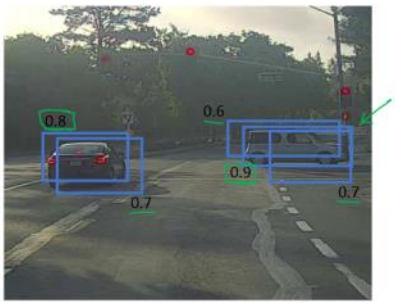


19x19

Andrew Ng

The car's size span through grids

Non-max suppression example



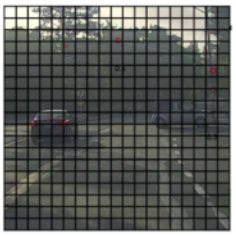
P_c

multiple detection of one car (object)

takes

Andrew Ng

Non-max suppression algorithm



19× 19

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 IoU ≥ 0.5 with the box output in the previous step

Andrew Ng

anchor boxes

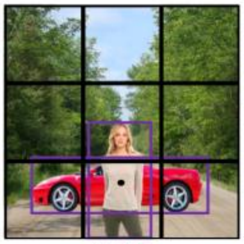
deeplearning.ai

Object Detection

Anchor boxes

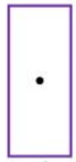
Overlapping objects:

pre-defined two boxes
=> associate the prediction with two boxes

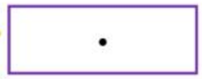


$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Anchor box 1:



Anchor box 2:



$y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3, p_c, b_x, \ldots, c_3]$ — anchor box 1 / anchor box 2

pre-defined shape! How?

after getting bx, by, bh, bw
后，以它为参数做一个 anchor box
=> 然后计算 anchor box
和 label 之间的 IoU ..
吧 (?)

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

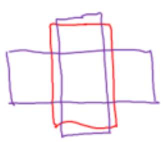
Andrew Ng

Anchor box algorithm

Previously:

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

Output y: 3×3×8



With two anchor boxes:

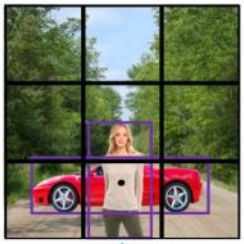
Each object in training image is assigned to grid cell that contains object's midpoint and anchor box for the grid cell with highest IoU.

(grid cell, anchor box)

Output y:
3×3×16
3×3×2×8

Andrew Ng

Anchor box example



Anchor box 1: Anchor box 2:

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

Car only?

anchor box 1

anchor box 2

Andrew Ng

每种东西一个类型的 anchor box

other problem

two objects in a grid that same type have the same anchor box

→ 这个 algorithm 暂未解决

Anchor box: Allow algorithm to specialize better.

? How to choose the anchor boxes? By hand

Advanced: use the k-means algorithm to select Anchor box (?)

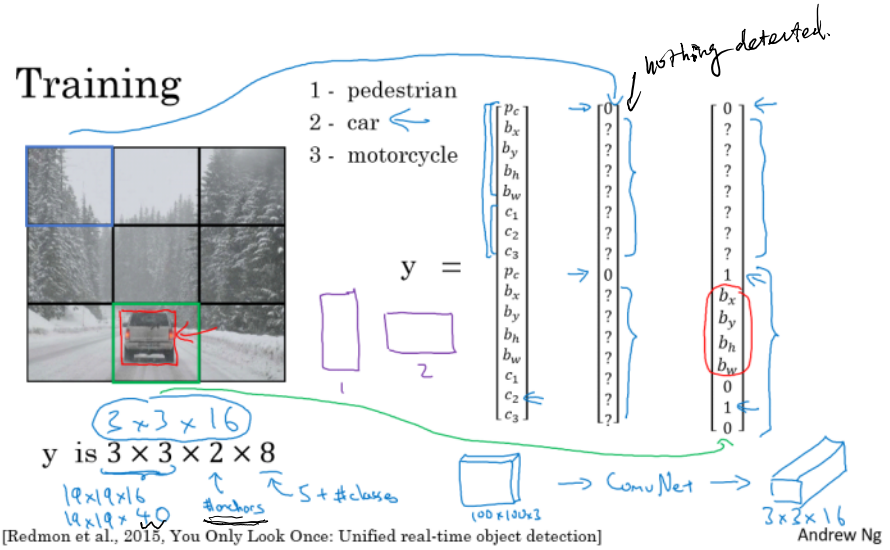
later YOLO papers ...

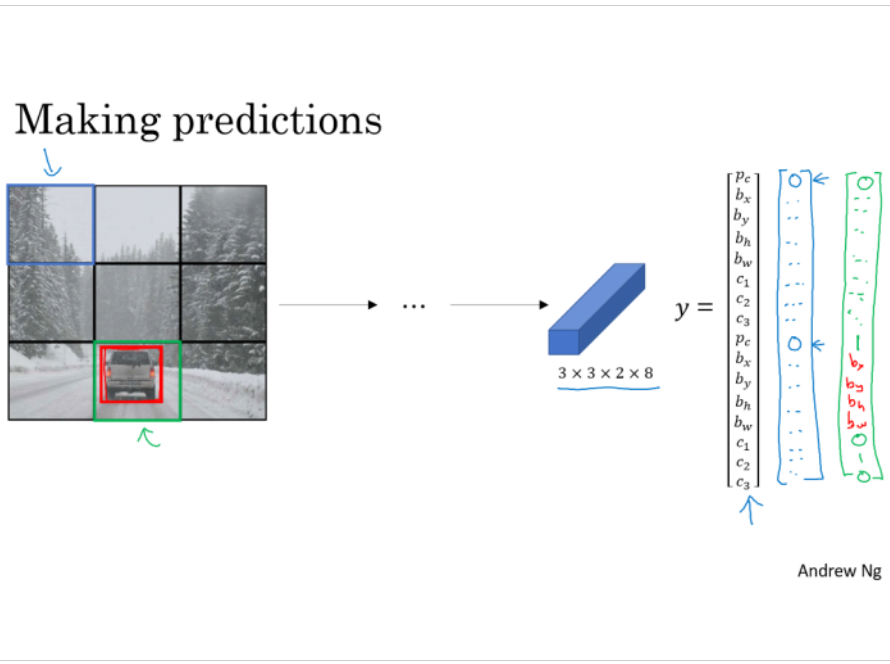
yolo

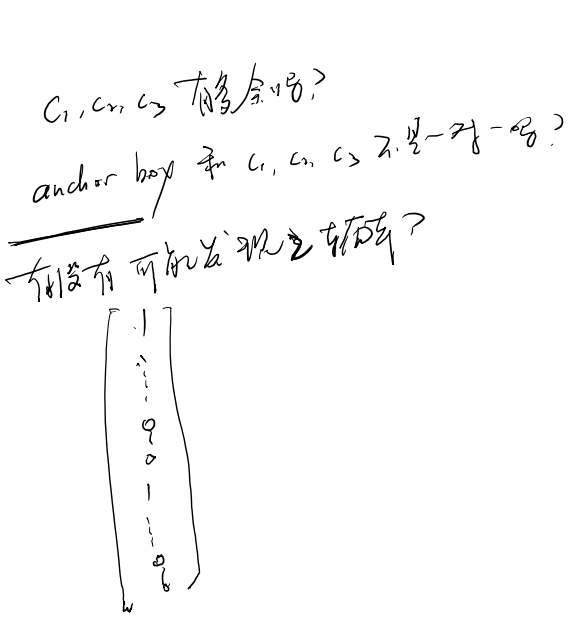
deeplearning.ai

Object Detection

Putting it together: YOLO algorithm


Training
1 - pedestrian
2 - car
3 - motorcycle
y =
nothing detected.
y is 3 × 3 × 2 × 8
3×3×16
19×19×16
19×19×40
#anchors
5 + #classes
100×100×3
ConvNet
3×3×16
[Redmon et al., 2015, You Only Look Once: Unified real-time object detection]
Andrew Ng


Making predictions
3 × 3 × 2 × 8
y =
Andrew Ng



Outputting the non-max supressed 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.

Some bounding boxes can go outside the grid — ? how

Andrew Ng

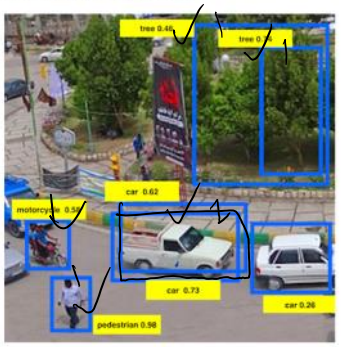
✖ 7. In the YOLO algorithm, at training time, only one cell ---the one containing the center/midpoint of an object--- is responsible for detecting this object.

0 / 1 points

Yes (I don't understand what this means).

✖ 9. Suppose you run non-max suppression on the predicted boxes above. The parameters you use for non-max suppression are that boxes with probability ≤ 0.4 are discarded, and the IoU threshold for deciding if two boxes overlap is 0.5. How many boxes will remain after non-max suppression?

0 / 1 points



✖ 10. Suppose you are using YOLO on a 19x19 grid, on a detection problem with 20 classes, and with 5 anchor boxes. During training, for each image you will need to construct an output volume y as the target value for the neural network; this corresponds to the last layer of the neural network. (y may include some "?", or "don't cares"). What is the dimension of this output volume?

0 / 1 points

$19 \times 19 \times 5 \times 25$

5 : num of classes.