

Object Localization

$$\text{Output} = [P_c \ b_x \ b_y \ b_h \ b_w \ c_1 \ c_2 \ c_3]^T$$

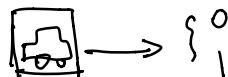
if $P_c = 0$, dont care rest output

whether object exists coordinate of bounding box + squared error class
↓
logit regression err ↓
log likelihood loss

Landmark Detection

- similar to object localization, instead of b_x b_y b_h b_w .
we just need multiple point coordinate $[P_c, l_1x \ l_1y \ l_2x \ l_2y \ \dots]^T$

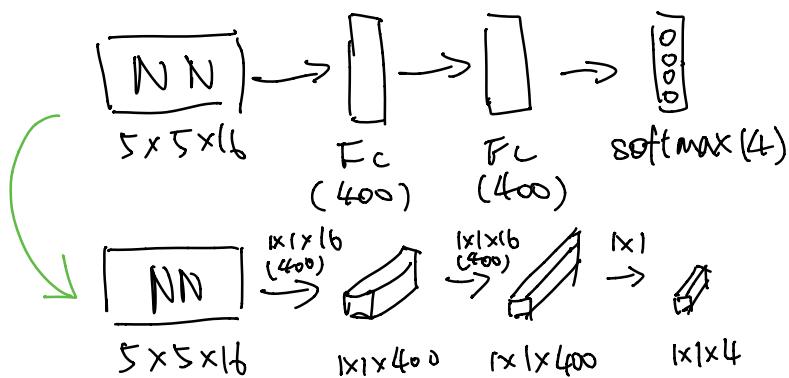
Sliding Window Algorithm



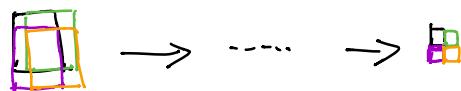
1. Train a convnet that output 0 / 1 for a car
2. feed cropped image into convnet

Sliding Window can be implemented using convolution

use convolution to replace FC layers

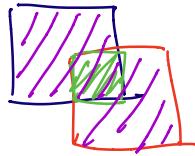


sliding window convolution:



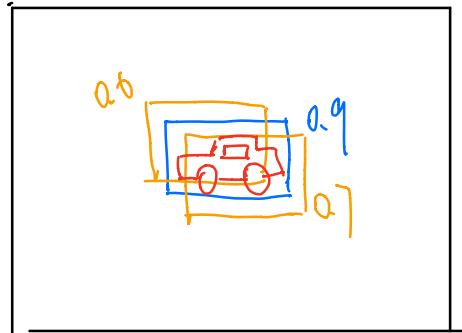
Intersection over union (IOU)

$$= \frac{\text{size of } \textcolor{green}{\square}}{\text{size of } \textcolor{purple}{\square}}$$



Convention $\begin{cases} \text{iou} \geq 0.5 \rightarrow 1 \\ \text{else} \quad \quad \quad \rightarrow 0 \end{cases}$

Non Max Suppression :



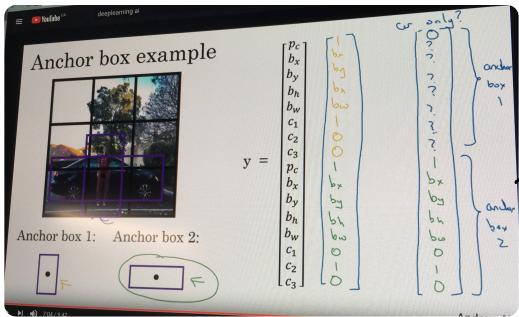
only keep the box with the highest probability (0.9 in this case)
 \Rightarrow Discard non-max boxes

Non-Max algo :

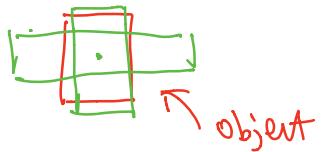
Discard $P_c \leq 0.6$

while there is remaining box:

- pick box B with largest P_c .
- Discard rest with $\text{iou} \geq 0.5$ with B .
- repeat



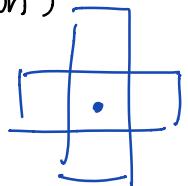
eg. 2 anchor box



Anchor Box [Redmon et al., 2015. Unified real-time object de...]

Problem Anchor Box solve: 2 objects in one grid cell
(detection specialization)

$$y = \underbrace{[p_c \ b_x \ b_y \ b_h \ b_w \ c_1 \ c_2 \ c_3]}_{\text{Box 1}} \underbrace{[p_c \ b_x \dots \ c_3]}_{\text{Box 2}}^T$$



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

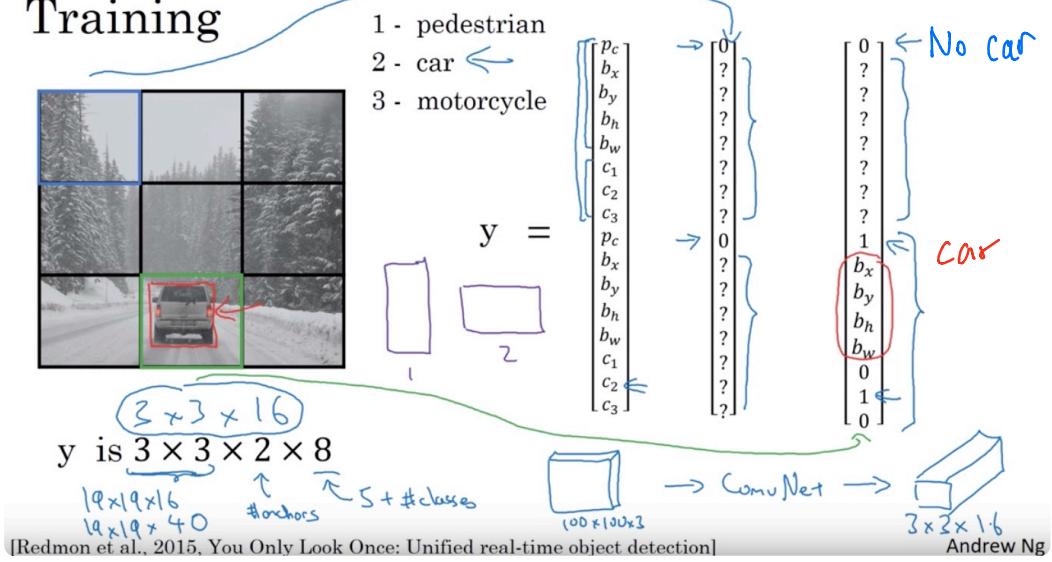
Output: $3 \times 3 \times 8$

With 2 anchor box: Each object in training image is assigned to a grid cell that contains object's midpoint and anchor box for the grid cell with highest IOU.

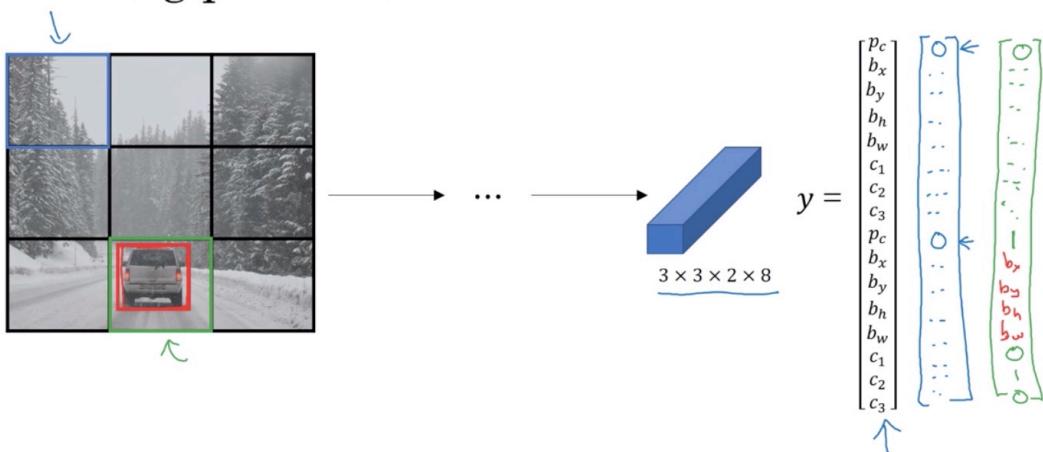
Output: $3 \times 3 \times (8 \times 2)$ (grid cell, anchor box)

YOLO Algorithm \rightarrow You Only Look Once

Training



Making predictions



Outputting the non-max suppressed outputs

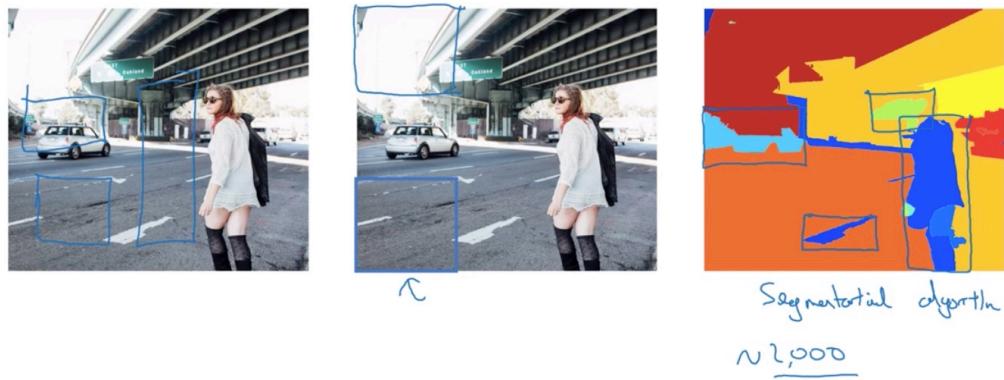


- For each grid cell, get 2 predicted bounding boxes.
- Get rid of low probability predictions.
- For each class (pedestrian, car, motorcycle) use non-max suppression to generate final predictions.

Region Proposal (R-CNN)

- only run CNN on region that likely to be object . \rightarrow still much slower YOLO

Region proposal: R-CNN



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

R-CNN: Propose Regions - classify proposed regions one at a time .
output label + bounding Box

Fast R-CNN: Propose Regions : then use convolution implementation of sliding window to classify all proposed regions

Faster R-CNN : use Convnet to propose Region