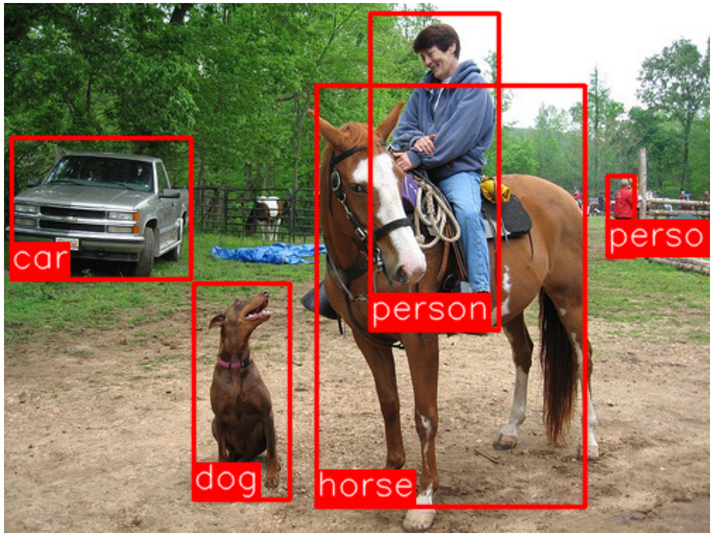
Object detection and segmentation

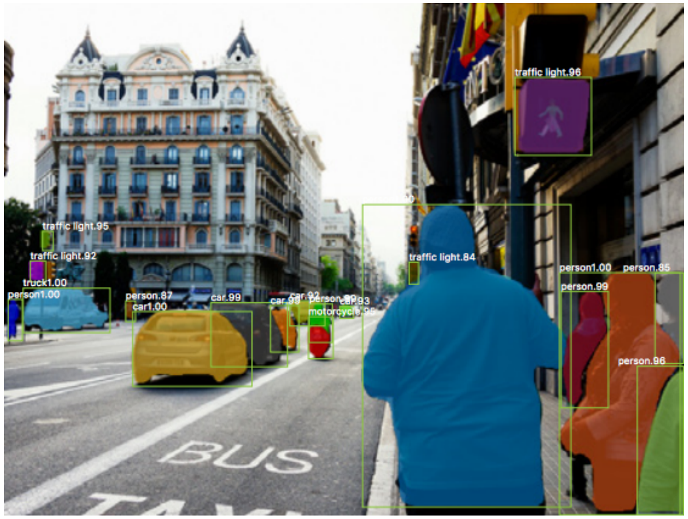
ISTD 50.035

Computer Vision

Object detection / segmentation

- Finding different objects in an image and classify them





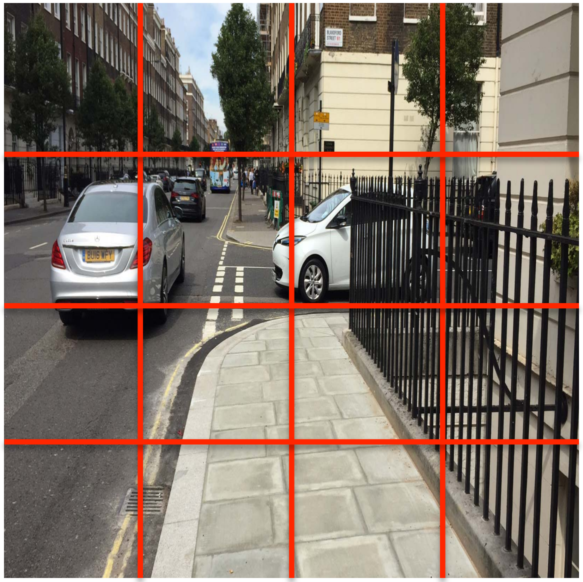
YOLO: You Only Look Once

- Reframe object detection as a single regression problem

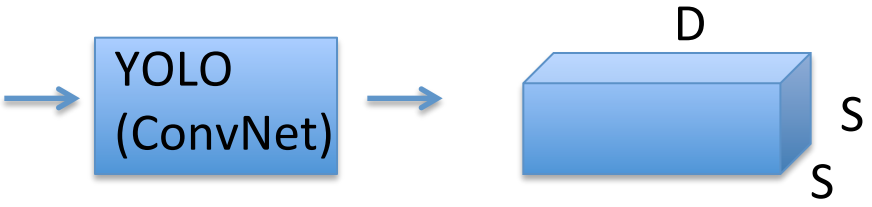
- Very fast: object class probabilities and bounding box coordinates regression in a single forward pass

YOLO: You Only Look Once

SxS grid



Each D-dim vector encodes the class probabilities and bounding box coordinates for that cell



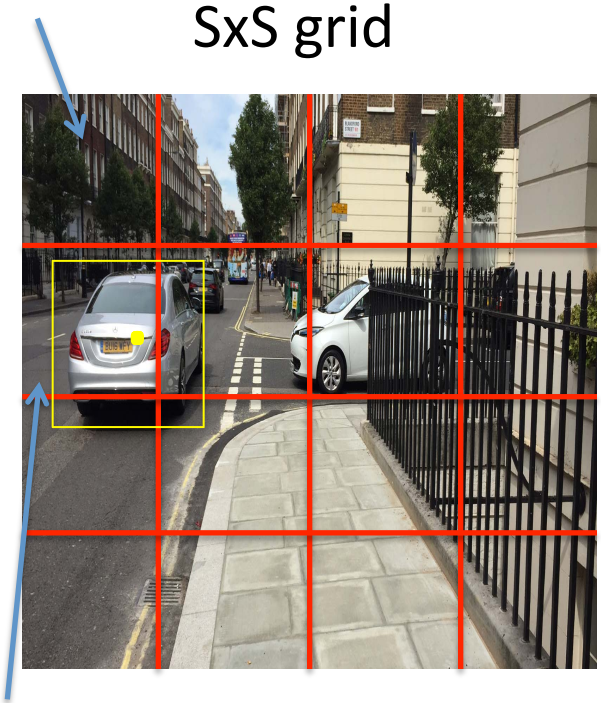
SxSxD regression output

-box confidence (object or not, how accurate is this box)
-x,y,h,w (center coord, width, height)
-class probabilities

YOLO: You Only Look Once

[0,_,_,_,_,_,_]

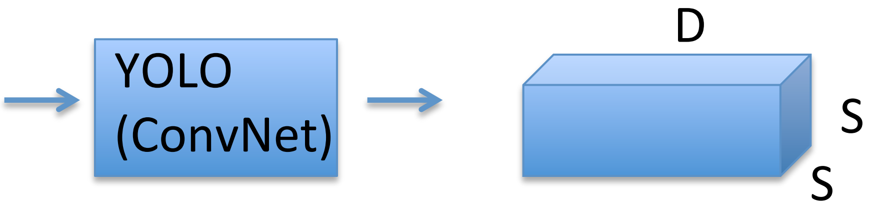
SxS grid



[1,0.9,0.6,0.25,0.25,0,1]

Dog, Car

Each D-dim vector encodes the class probabilities and bounding box coordinates for that cell



SxSxD regression output

-box confidence (object or not, how accurate is this box)

-x,y,h,w (center coord, width, height)

-class probabilities

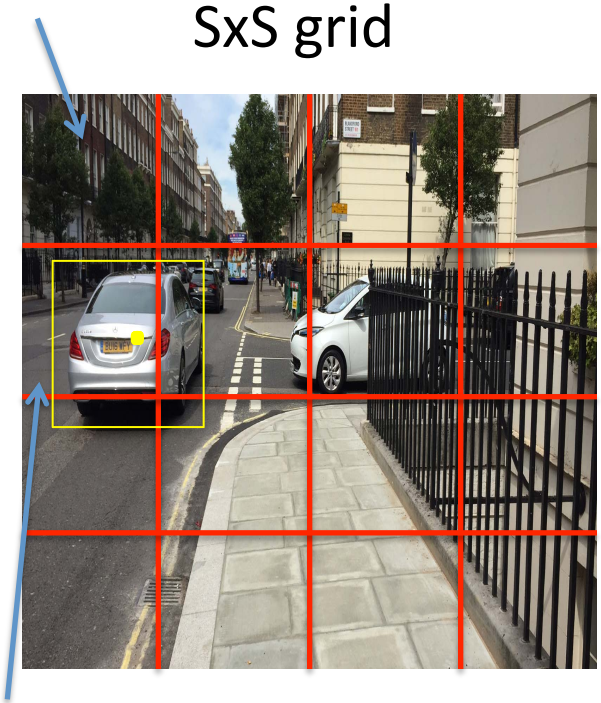
-If the center of an object falls into a grid cell, that gird cell is responsible for detecting that object

-Each grid cell detects only one object (small cell is used)

YOLO: You Only Look Once

[0,_,_,_,_,_,_]

SxS grid



[1,0.9,0.6,0.25,0.25,0,1]

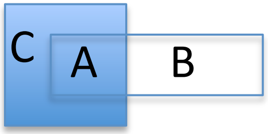
Dog, Car

Center (x,y): relative to the grid, normalized to [0,1]
(w,h): relative to image size, [0,1]

Box confidence: Pr(object) * IOU (pred, truth)

- No object -> box confidence = 0
- Higher IOU -> potentially more accurate bbox

-If the center of an object falls into a grid cell, that gird cell is responsible for detecting that object
-Each grid cell detects only one object (small cell is used)

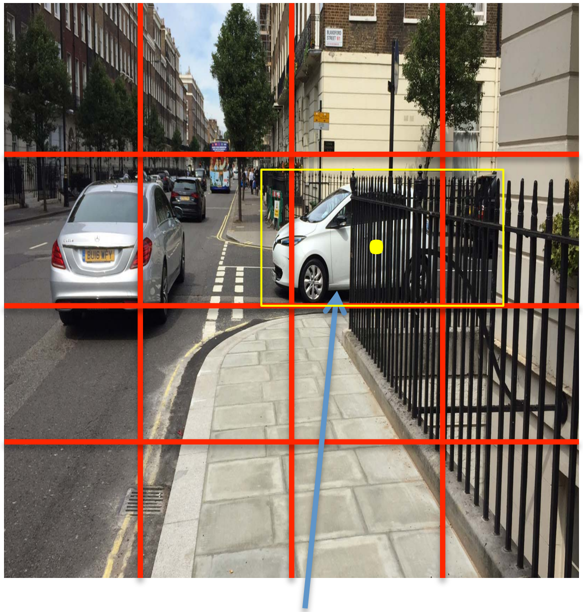


Intersection over Union:

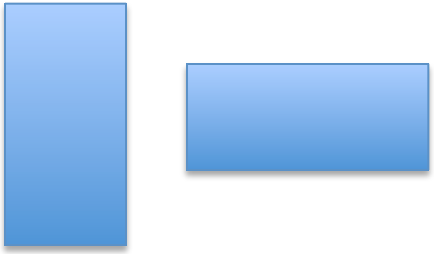
IOU= A / (A+B+C)
Usually requires > 0.5 for overlapping

YOLO: You Only Look Once

SxS grid

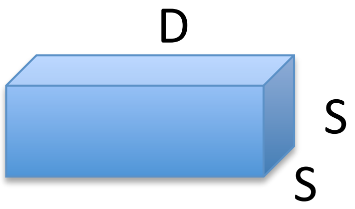


Predict two bounding boxes (anchors) per grid cell



Training: only want one bbox predictor to be responsible for each object: one with highest IOU

$$[\underbrace{0,\_,\_,\_,\_}_{\text{bbox 0}},\underbrace{1,0.6,0.6,0.24,0.3}_{\text{bbox 1}},0,1]$$

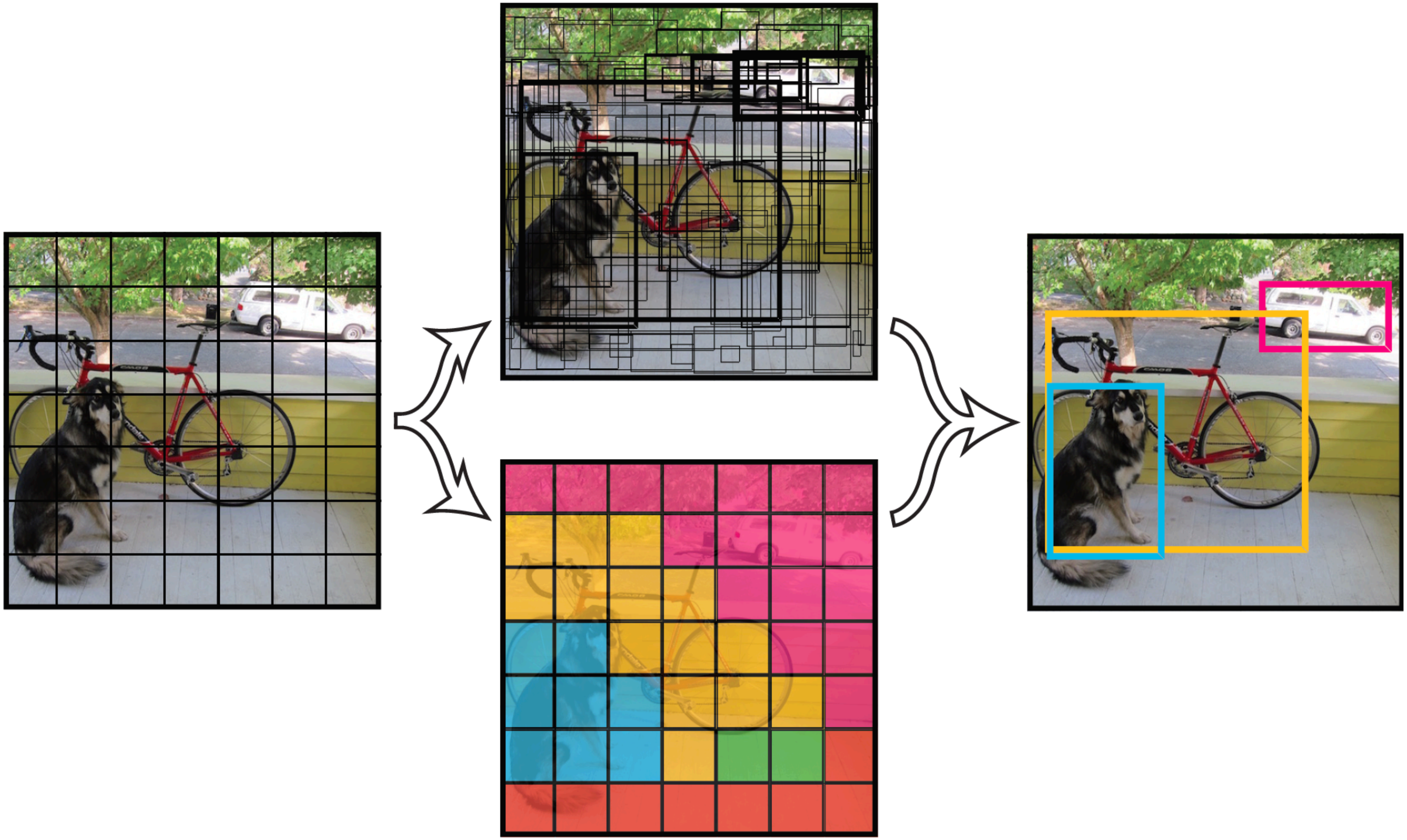


YOLO output: SxSxD tensor

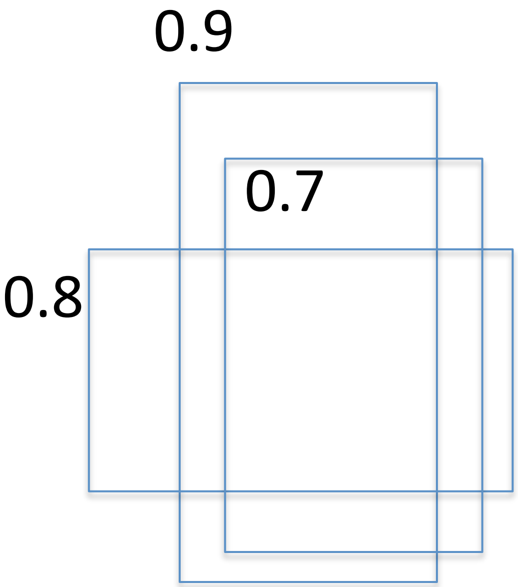
D=Bx5 + C

B = 2 (two bbox)

YOLO: You Only Look Once



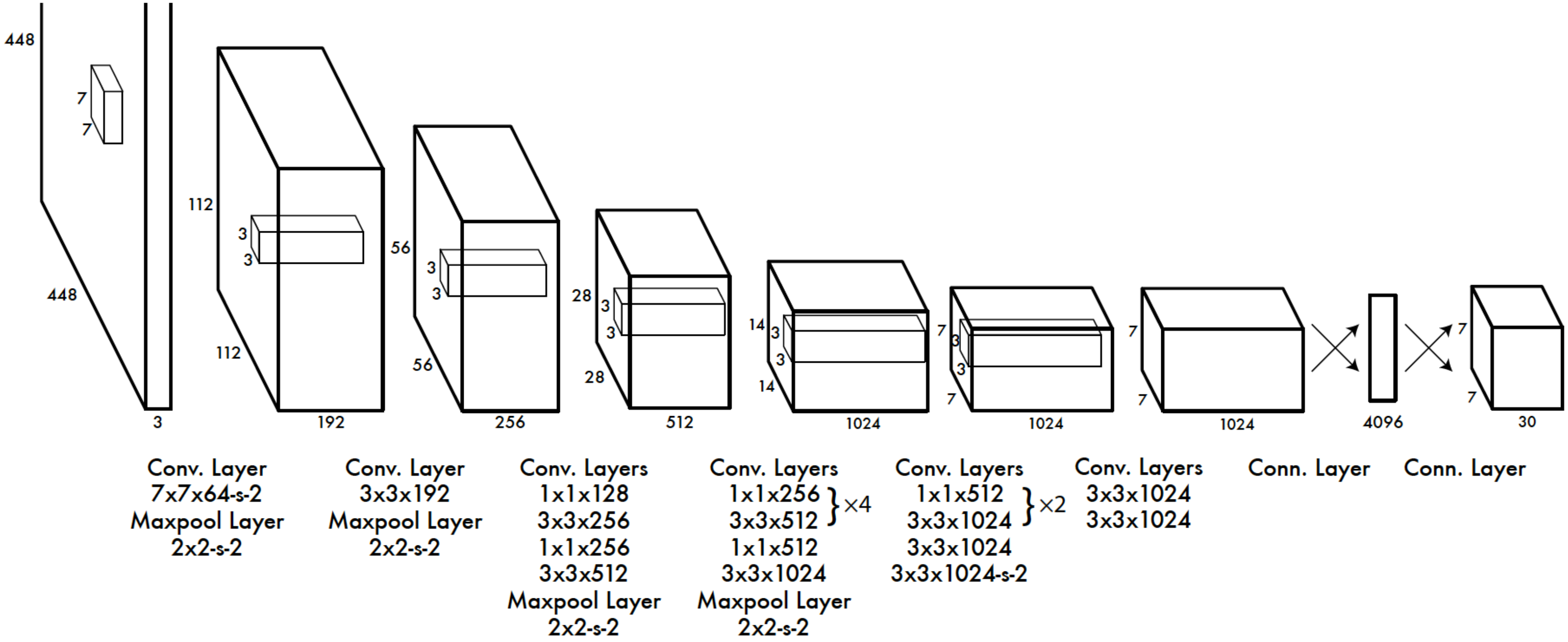
Non maximum suppression



Many detection per object -> choose one

1) Discard bbox_confidence <= 0.6
2) Select one with highest bbox_confidence
3) Discard bbox with IOU >= 0.5
4) Goto (2)

YOLO: You Only Look Once



20 labeled classes

YOLO: You Only Look Once

Training: multi part loss

$$
\begin{aligned}
&\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \\
&+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2
\end{aligned}
$$

where $\mathbb{1}_i^{\text{obj}}$ denotes if object appears in cell i and $\mathbb{1}_{ij}^{\text{obj}}$ denotes that the jth bounding box predictor in cell i is “responsible” for that prediction.