May-2-2019 YOLO Formulation (1)

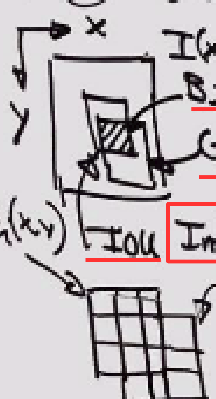
CMPE297 Video Analytics. May 2, 19.

Note: 1° Final Exam: May 21st (Tue) 14:45–17:00
In this Room;

2° Final Project Presentation.
(1) Demo;
(2) P.P.T. Presentation.
(3) Source Code;
(4) Report — IEEE Paper Template.
(See Github / hualili / Cmpe242. "Guidelines for Report ~ V2")

3° Show & Tell

Today's Topics: Yolo Network;
Ref: github / hualili / cmpe297 / 2019S
✓ ① 2019S-39-yolo-reference
✓ ② 2019S-39-lecyolo~ (pdf, ppt)



$I(x,y)$
$B_i(x,y)$ for $i = 1, 2, \ldots, k$
Ground Truth Box
$B_1(x,y)$ — IOU: Intersection Over Union $= \begin{cases} \max IOU = 1 & \text{if Bounding = Ground Truth Box} \\ 0 \leq, < 1 & \cdots (1) \end{cases}$

Ground Truth
Find: $IOU = 4/9 = 4/9$
Example 1:

1. $I(x,y)_{M \times N} \rightarrow G(x,y)$ grid, $\frac{M}{S} \times \frac{N}{S}$, where $S \times S$ is the Total Number of Grids.

2. … $O_i(x,y) \in G(x,y)$

Object O_i(x,y) "falls into" / Belongs to grid G(x,y)

3. 5-parameters for Each Bounding Box
$B_j(x,y)$, $\{x, y, w, h, f(B_j(x,y))\}$

4. $Prob(O_i(x,y))$ Probability of Object $O_i(x,y)$ within the $B_j(x,y)$
?

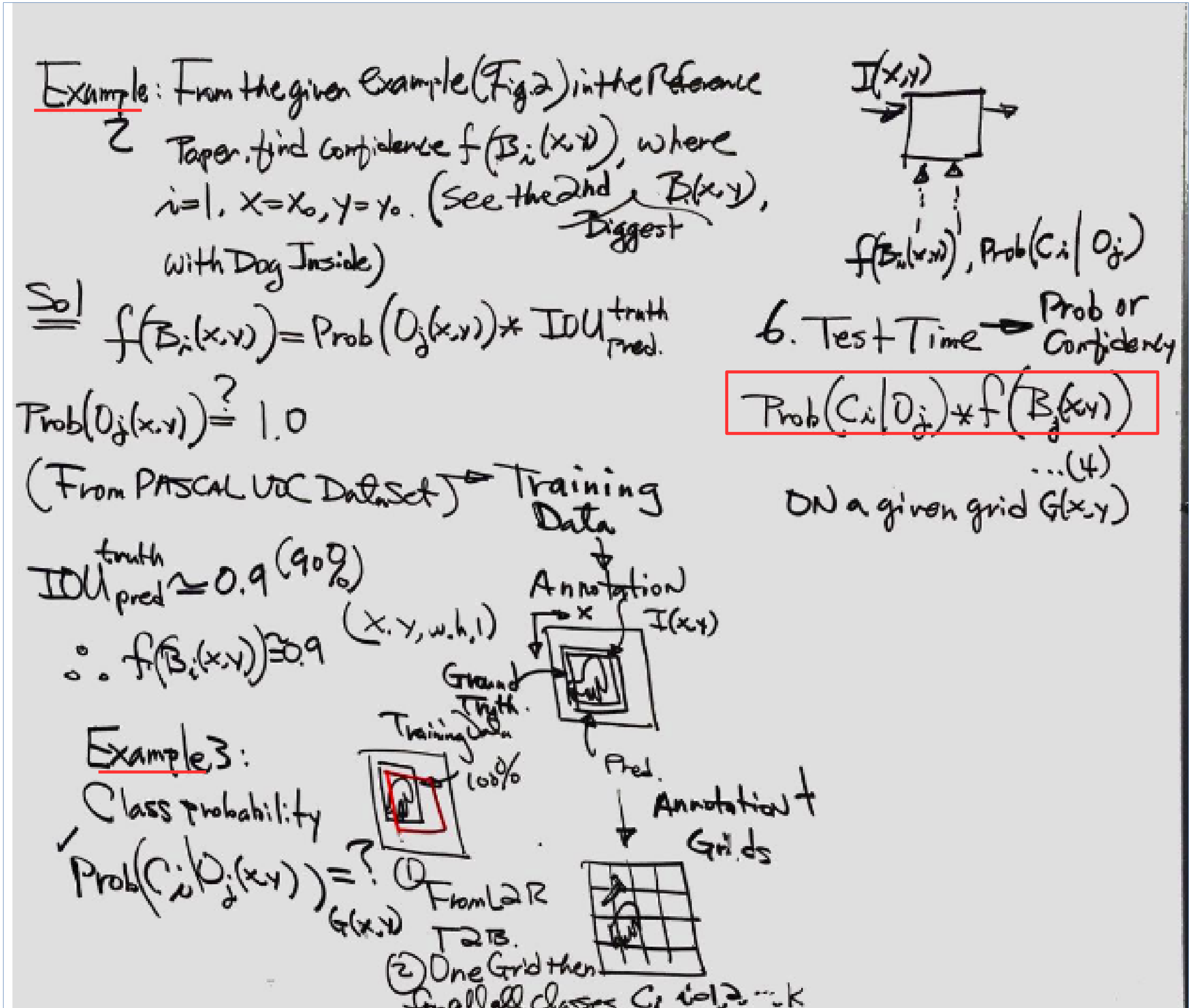
$$IOU = \frac{B_j(x,y) \cap G.Tr.}{\Sigma(\text{Total Area of } B_i(x,y))}$$

$f(B_j(x,y)) = Prob(O_i(x,y) | B_j(x,y)) * IOU \cdots (1)$
?
$f_{max} = 1.0,\ f_{min} = 0;\ \cdots (2)$

5. $Prob(C_i | O_j(x,y))$ ON a Grid $G(x,y)$ $\cdots (3)$
$Prob(C_i | O_j(x,y), G(x,y))$; C_i for Class i, $i = 1, 2, \ldots, M$.
?

Harry Li, Ph.D.

May-2-2019 YOLO Formulation (2)



Example 2: From the given Example (Fig 2) in the Reference Paper, find confidence $f(B_i(x,y))$, where $i=1$, $x=x_0$, $y=y_0$. (See the 2nd Biggest $B(x,y)$, with Dog Inside)

Sol

$$f(B_i(x,y)) = Prob(O_j(x,y)) * IOU^{truth}_{pred.}$$

$$Prob(O_j(x,y)) \overset{?}{=} 1.0$$

(From PASCAL VOC DataSet) → Training Data

$$IOU^{truth}_{pred} \simeq 0.9 \; (90\%)$$

$(x, y, w, h, 1)$

$$\therefore f(B_i(x,y)) \geq 0.9$$

Training Data → Annotation

x

$I(x,y)$

Ground Truth.

Training Data

100%

Pred.

Annotation + Grids

Example 3:

Class probability

✓ $Prob(C_i | O_j(x,y))_{G(x,y)} = ?$ ① From L2R T2B. ② One Grid then for all classes C_i, $i=1,2,\cdots K$

$I(x,y)$

$f(B_i(x,y))$, $Prob(C_i | O_j)$

6. Test Time → Prob or Confidency

$$\boxed{Prob(C_i | O_j) * f(B_j(x,y))} \quad \ldots(4)$$

ON a given grid $G(x,y)$

May-2-2019 YOLO Formulation (3)

Note: Probability Value Defined as Binary Case

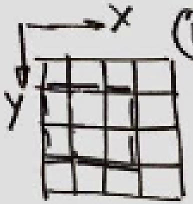
$$Prob(Dog: C_i \mid O_j(x,y), \text{ on } G(x,y)) = 1.0$$

$$\boxed{Tensor = S*S*(B*5+C) \quad \ldots(5)}$$

S×S for Grid G(x,y)

Bounding Box $B_i(x,y)$

Class C_i.



(1) Find Tensor=?

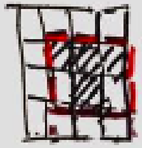
Given 2 Traffic Signs: C_1(Left), C_2(Right).

$B_i(x,y)=2$.

(2) Find Confidency for $B_i(x,y)$.

$f(B_i(x,y)) = Prob(O_j(x,y))$ within the Bounding Box $B_i(x,y)$

Need Annotated Data.



$Prob(O_j) = 3/6 = 0.50$

(3) Find IoU $= B_i(x,y) \cap$ Ground Truth Box

Ground Truth Box: 3×3 red

$= 4/6 = 2/3 = 0.667$

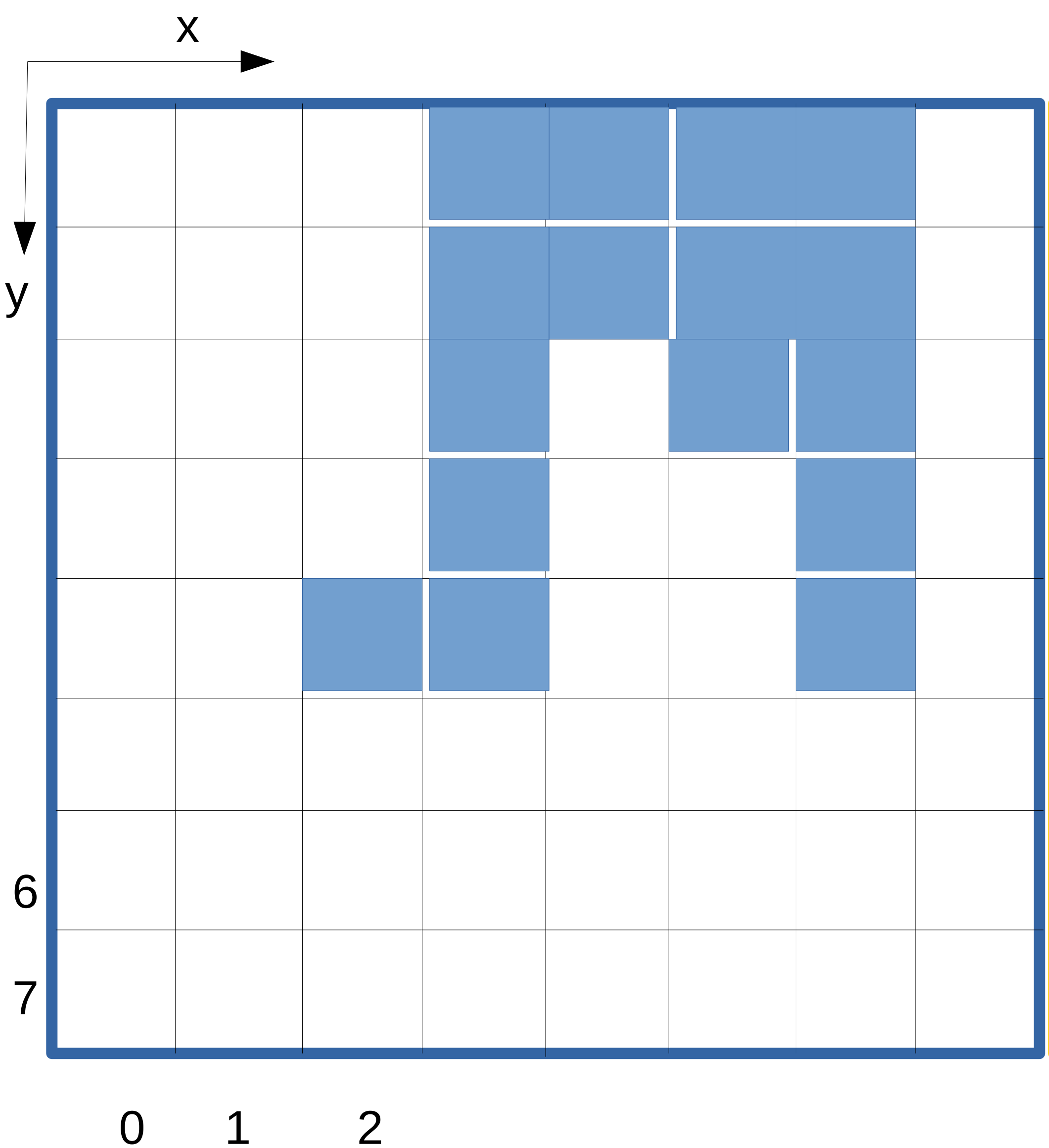
$\therefore f(B_i(x,y)) =$

$$f(B_i(x,y)) = 0.50 * 2/3 = 1/3 = 0.33$$

Example

Page 1

You Only Look Once: Unified, Real-Time Object Detection



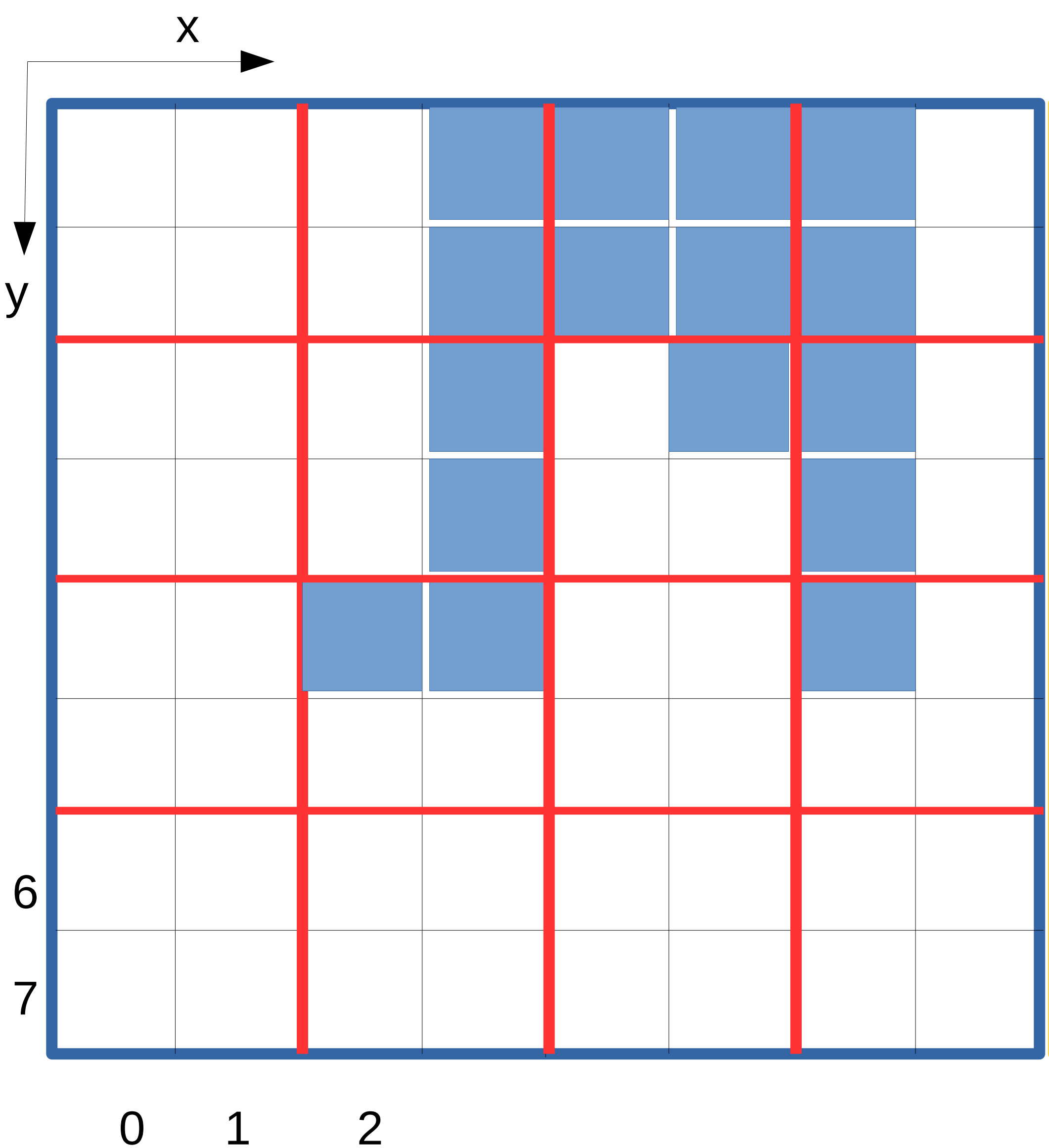
Each pixel is shown here

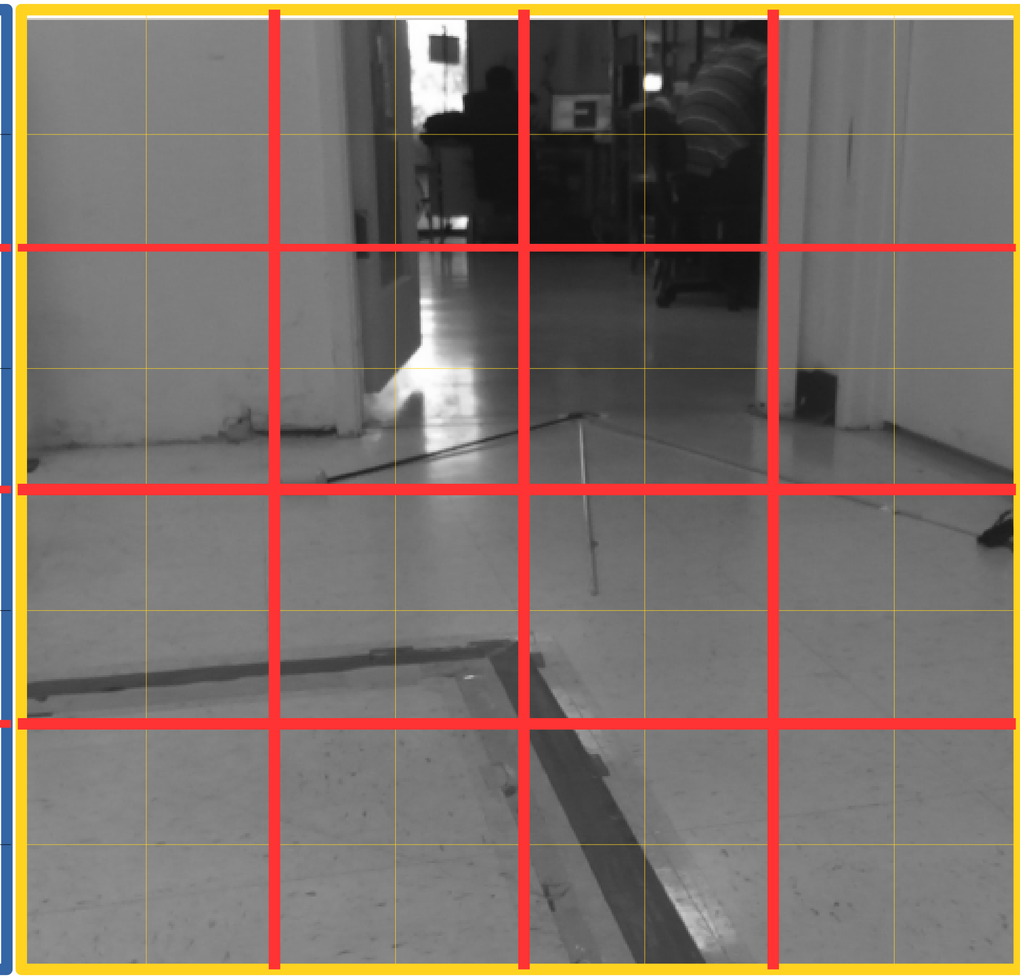
Harry Li, Ph.D.

Example

Page 2

You Only Look Once: Unified, Real-Time Object Detection





(1) Create S by S grid, denoted as G_i(x,y)

Harry Li, Ph.D.

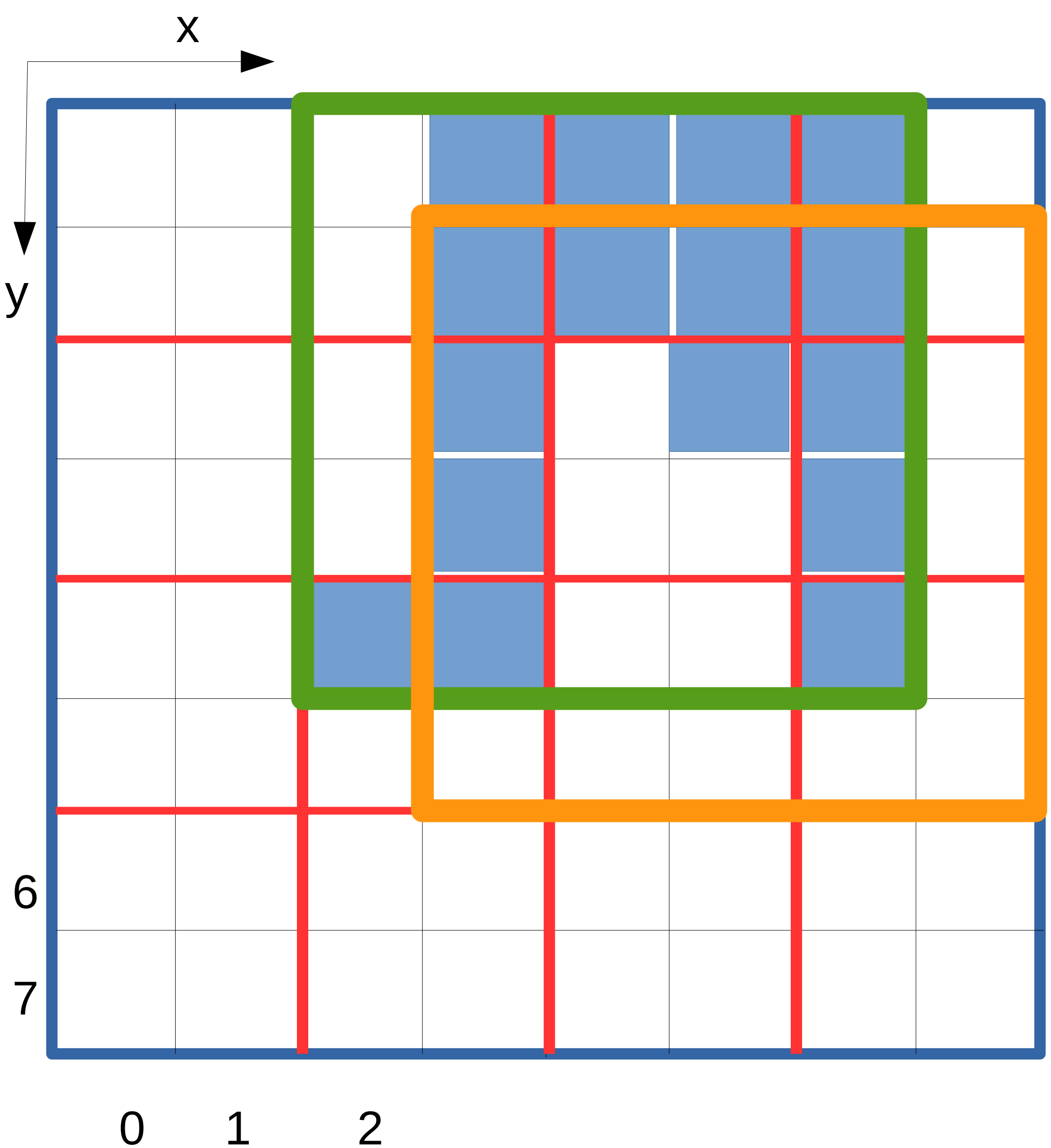
(2) Define bounding box B_i(x,y) = 1, (3) define object (door) as O_i(x,y), (4) define classes C_j = 1.

Example

Calculate Tensor

You Only Look Once:
Unified, Real-Time Object Detection

Page 3



Find: tensor of this example.

Tensor = S * S * (5 * B + C) from equation (5), so

Tensor = 4 * 4 * (5 * 1 + 1) = 96

Find ground truth bounding box:

Based on the input image, and human annotation process, the green box is drawn and shown in this example.

Given one of the bounding boxes shown in this grid, as orange color. Find IOU (intersection over union)

(1) Create S by S grid, denoted as G_i(x,y)

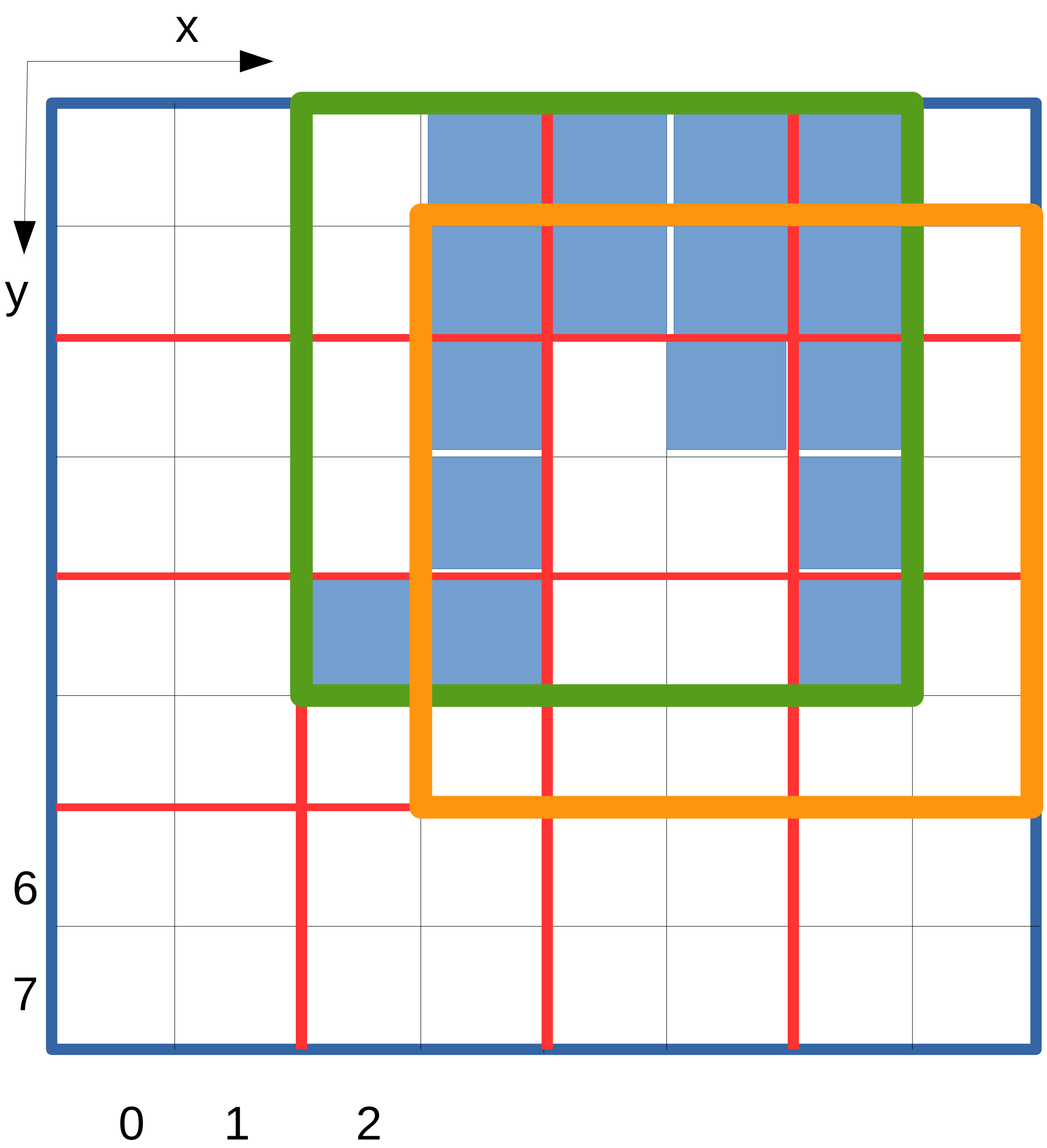
(2) Define bounding box B_i(x,y) = 1, (3) define object (door) as O_i(x,y), (4) define classes C_j = 1.

Harry Li, Ph.D.

Example

Compute IOU

Page 4



Given one of the bounding boxes shown in this grid, as orange color. Find IOU (intersection over union)

Step 1. find the total number of pixels (area) of the ground truth box, A_true = 5 * 5 = 25 (green box), ground truth is given from the annotation process;
Step 2. find the intersection area (pixels) between the bounding box B_i(x,y) and the ground truth box, so A_Intersection = 4 * 4 = 16 from the figure on the left;
Step 3: find IOU (intersection over union) which is the ratio of

$$\boxed{IOU = \frac{A\_Intersection}{A\_true}} \quad \ldots (1)$$

So, the IOU for the left figure:

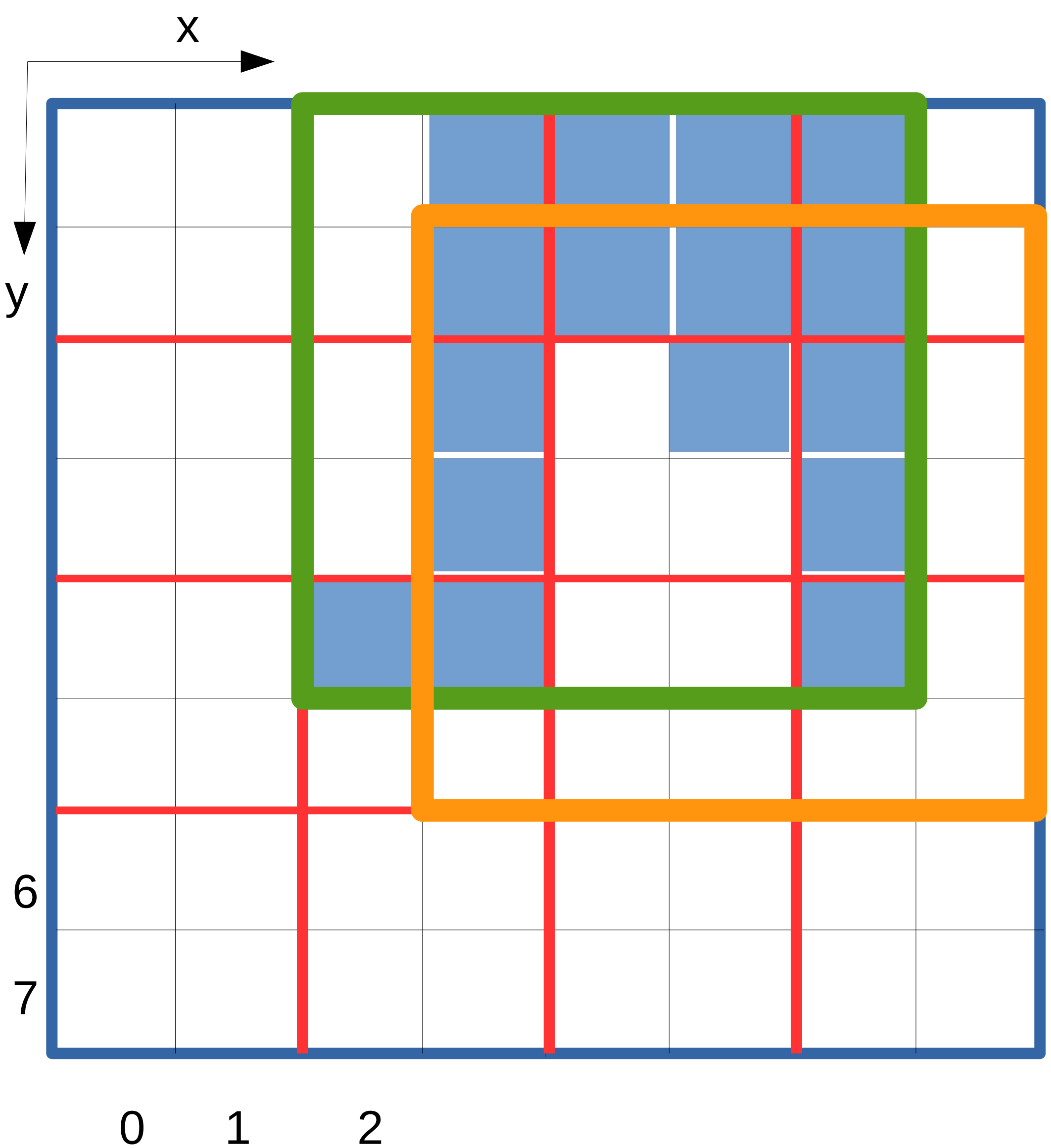
$$IOU = \frac{A\_Intersection}{A\_true} = \frac{16}{25} = 0.64$$

Harry Li, Ph.D.

Example

Page 5

Find Prob(O_i(x,y)|B_j(x,y))



Find the probability of object O_i(x,y), the door, within the bounding box B_j(x,y) as illustrated left in the orange box.

Step 1. From the ground truth, we know for this object, a particular door, the area (total number of pixels), so we have A_Obj_groundTruth = 16 (blue pixels);

Step 2. Find the total object pixels within the bounding box, A_obj_bounding = 11 (blue pixels within the orange box),

Step 3.

$$\text{Prob(O\_i(x,y) |B\_j(x,y))} = \frac{\text{A\_obj\_bounding}}{\text{A\_Obj\_groundTruth}} \quad \dots (2)$$

So, for

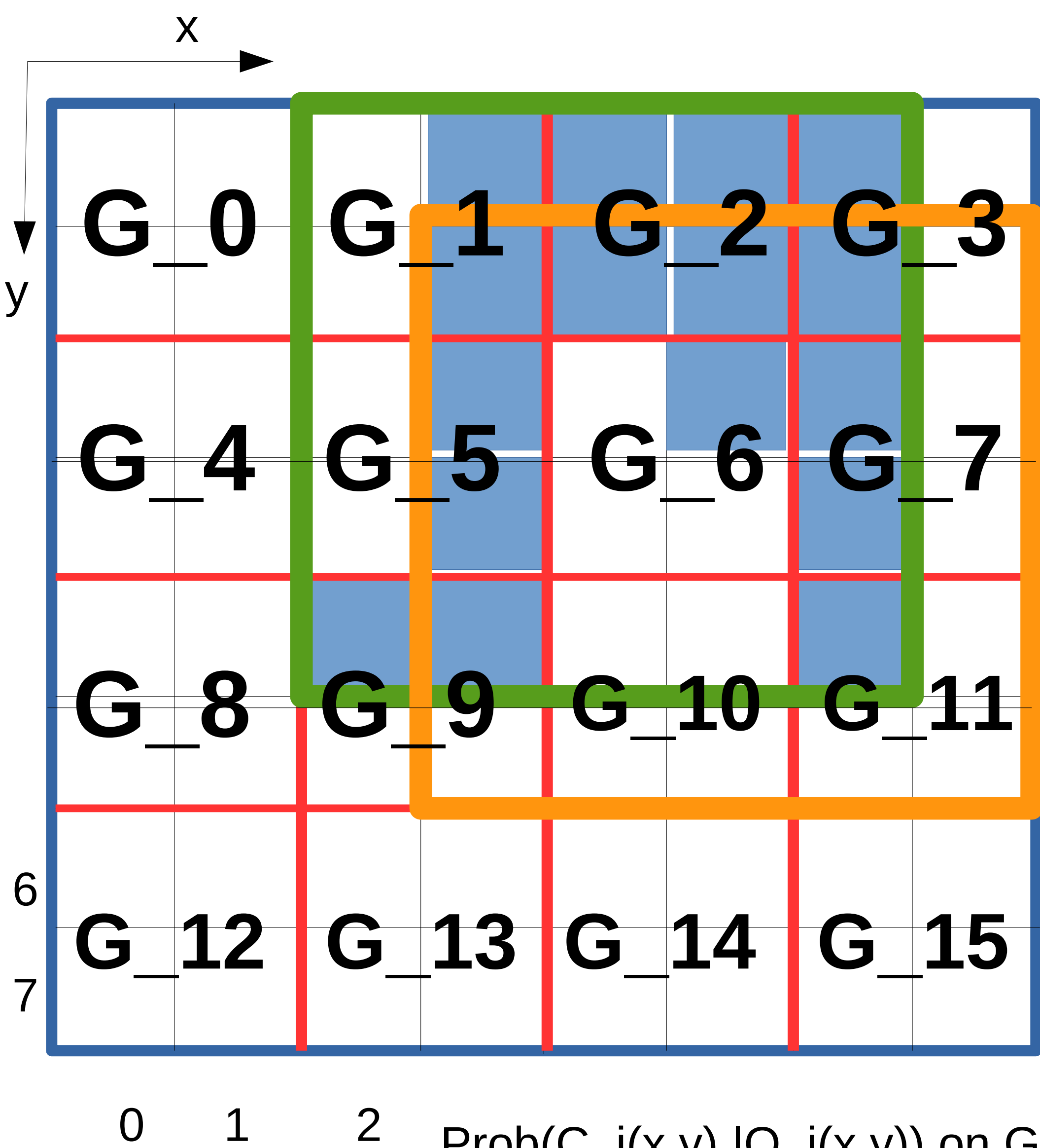
$$\text{Prob(O\_i(x,y) | B\_j(x,y))} = \frac{11}{16} = 0.6875$$

Harry Li, Ph.D.

Example

Page 6

Prob(C_j|O_i(x,y)) on G_k(x,y)



Find the probability of object O_i(x,y), belongs to class C_j , the door, on a given grid G_k(x,y)

Step 1. From the ground truth, we know for this object, a particular door, the area (total number of pixels in blue), so we have A_Obj_groundTruth = 16 (blue pixels), hence at each blue pixel:

$$\text{Prob(O\_i(x,y))} = \frac{1}{\text{A\_Obj\_groundTruth}} \quad .. (2)$$

for (x,y) belongs to the object (blue area). So, Prob(O_i(x,y)) = 1/16;

Step 2.Find the probability of object O_i(x,y), belongs to class C_j, the door, on a given grid G_k(x,y), for k=0, 1, …, 15

Prob(C_j(x,y) |O_i(x,y)) on G0, G1, …, G15 are: 0, 2/16, 4/16, 2/16, 0, 2/16, 1/16, 2/16, 0, 2/16, 0, 1/16, 0, 0, 0, 0

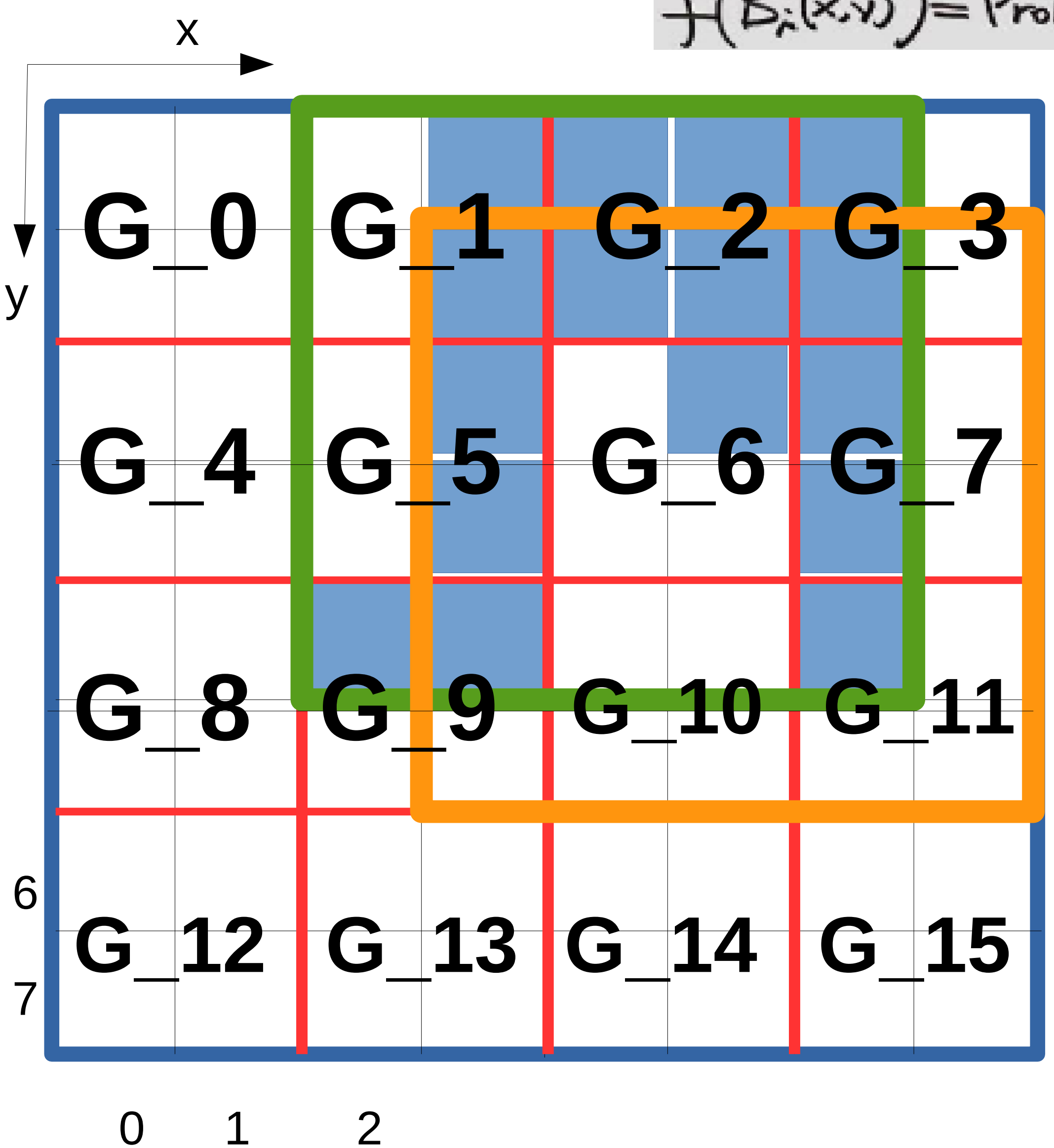
Harry Li, Ph.D.

Example

Page 7

Find Confidence

$$f(B_i(x,y)) = Prob(O_j(x,y)) * IOU_{pred}^{truth}$$



Note we write Prob(O_i(x,y)) as simplified version of Prob(O_i(x,y) |B_j(x,y)) on example page 5, so we have

$$\boxed{f(B\_i(x,y) = Prob(O\_i(x,y)) * IOU} \quad \ldots (3)$$

Note IOU $_{pred}^{truth}$ is written in a simplified notation as IOU

From example pp. 5, we have

$$Prob(O\_i(x,y) \mid B\_j(x,y)) = \frac{11}{16} = 0.6875$$

And from example pp. 4, we have

$$IOU = \frac{A\_Intersection}{A\_true} = \frac{16}{25} = 0.64$$

So,

f(B_i(x,y) = Prob(O_i(x,y)) * IOU = 0.6875 * 0.64 = 0.44

Example

Find $\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}}$

Page 8

Based on the example discussion, so far, we have

$\boxed{\text{Prob(C\_j | O\_i(x,y)) * f(B\_i(x,y)} \quad \ldots (8)}$

Note Prob(C_j | O_i(x,y)) is a probability on a given grid G_i(x,y), or can be written as Prob(C_j | O_i(x,y) and G_k(x,y), where we use simplified notation without G_k(x,y)

And again use simplified notation IOU

From example pp. 6, we have

$$\text{Prob(O\_i(x,y))} = \frac{1}{\text{A\_Obj\_groundTruth}}$$

Which is on each individual pixel, so for the bounding box B_i(x,y), we can evaluate

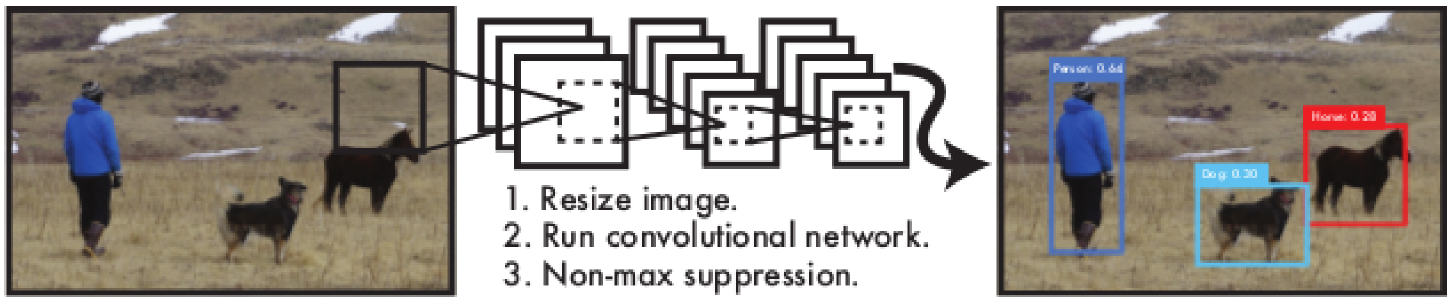
Prob(O_i(3,1) | (3,1)) + Prob(O_i(4,1) | (4,1)) + Prob(O_i(5,1) | (5,1)) + Prob(O_i(6,1) | (6,1)) + Prob(O_i(3,2) | (3,2)) + Prob(O_i(5,2) | (5,2)) + Prob(O_i(6,2) | (6,2)) + Prob(O_i(3,3) | (3,3)) + Prob(O_i(6,3) | (6,3)) + Prob(O_i(3,4) | (3,4)) + Prob(O_i(6,4) | (6,4)) = 1/16 + … + 1/16 = 11 * 1/16 = 11/16

So,
Prob(C_j | O_i(x,y)) * f(B_i(x,y) = 11/16 * 0.44 = 0.3025

You Only Look Once:
Unified, Real-Time Object Detection

https://arxiv.org/pdf/1506.02640v5.pdf

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi
University of Washington, Allen Institute for AI,Facebook AI Research http://pjreddie.com/yolo/



1. A single neural network predicts bounding boxes and class probabilities. 2. Base YOLO model runs at 45 FPS. A smaller version of the network, Fast YOLO, runs astounding 155 FPS second, outperforms DPM (deformable parts models) and R-CNN.

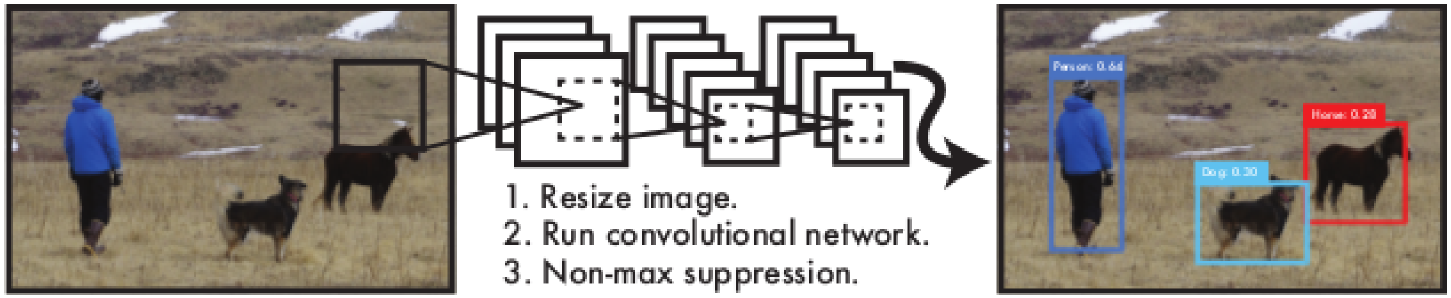
Figure 1: The YOLO Detection System. (1) resizes image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the result by the model's confidence.

Tutorial: https://pjreddie.com/darknet/yolo/

GitHub, Inc. (US)

```
git clone https://github.com/pjreddie/darknet
```

YOLO Model



1. Object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities.

A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance.

YOLO Formulation (1)

1. Yolo divides I(x,y) into an S × S grid G(x,y).

2. If the center of an object O_i(x,y), for i = 1, 2, ... , K, falls into a grid G(x,y), then that G(x,y) is responsible for detecting objects O_i(x,y).

3. Each grid G(x,y) predicts bounding boxes B_j(x,y). To detect objects O_i(x,y), G(x,y) places bounding box B_j(x,y), for j = 1, 2, ... , M. Each B_j(x,y) consists of 5 predictions: {x, y, w, h, f(B_j(x,y)}, where f(B_j(x,y)) is defined as confidence.

4. Define confidence

$$\boxed{f(B\_j(x,y)) = Prob(O\_i(x,y)) * (IOU\_truth)^{pred} \quad \ldots (1)}$$

where IOU is Intersection Over Union. Calculate the confidence for B_j(x,y):

$$\boxed{f(B\_j(x,y)) = \begin{array}{l} 0 \text{ if } O\_i(x,y) \text{ is null (no object).} \\ \text{equal to IOU between the predicted box and the ground truth.} \end{array}} \quad \ldots (2)$$

5. Each grid G(x,y) also predicts C_i, i = 1, 2, ..., N, define conditional class probabilities,

$$\boxed{Prob(C\_i \,|O\_j(x,y)) \quad \ldots (3)}$$

These probabilities are conditioned on the grid cell G(x,y) containing an object O_j(x,y).

YOLO Formulation (2)

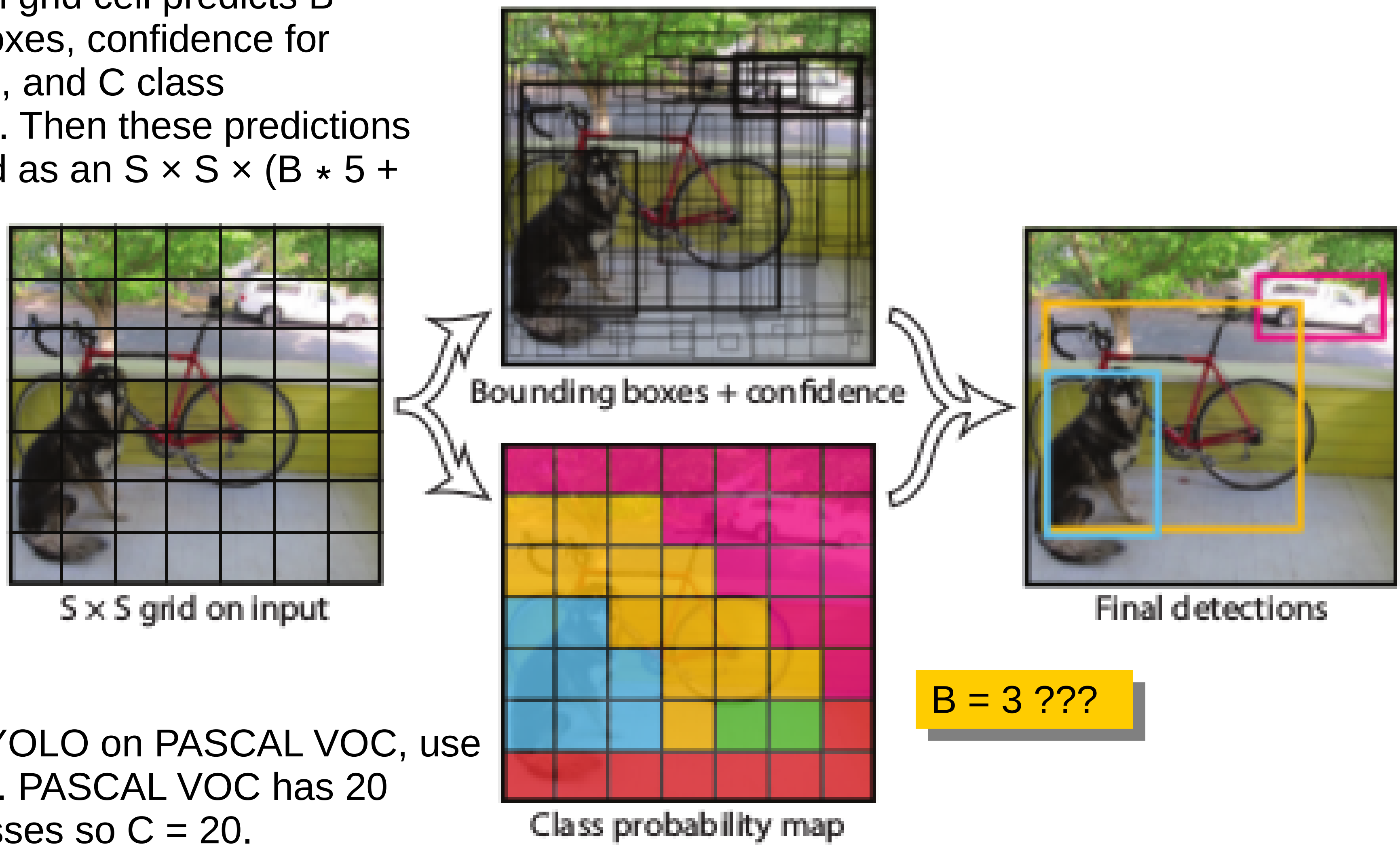
We only predict one set of class probabilities per grid cell G(x,y), regardless of the number of boxes B.
At test time we multiply the conditional class probabilities and the individual box confidence predictions,

$$\boxed{\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}}} \quad \ldots (4)$$

class-specific confidence scores for each box. These scores encode both the probability of that class appearing in the box and how well the predicted box fits the object.

Example YOLO Formulation (3)

Divides the image into an S × S grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. Then these predictions are encoded as an S × S × (B ⁎ 5 + C) tensor.



B = 3 ???

Evaluating YOLO on PASCAL VOC, use S = 7, B = 2. PASCAL VOC has 20 labelled classes so C = 20.
Our final prediction is a 7 × 7 × 30 tensor.