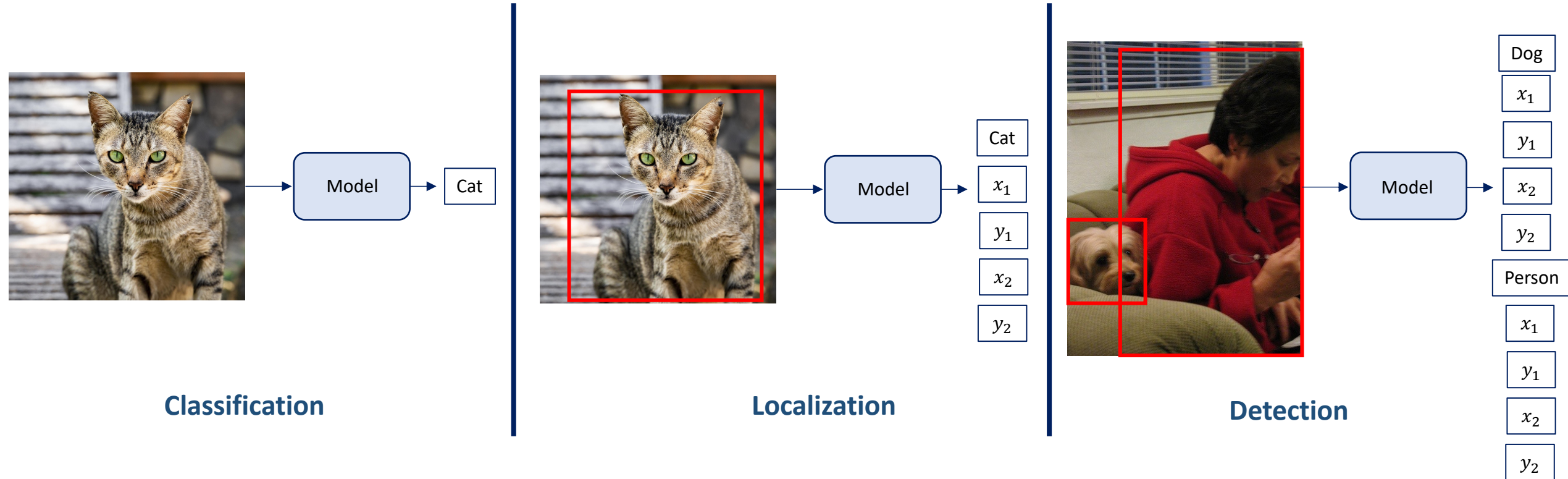# Object detection with YOLO

CMPT 743

Aryan Mikaeili

Object detection

- **Detect** every object in an image
  - **What** the objects are
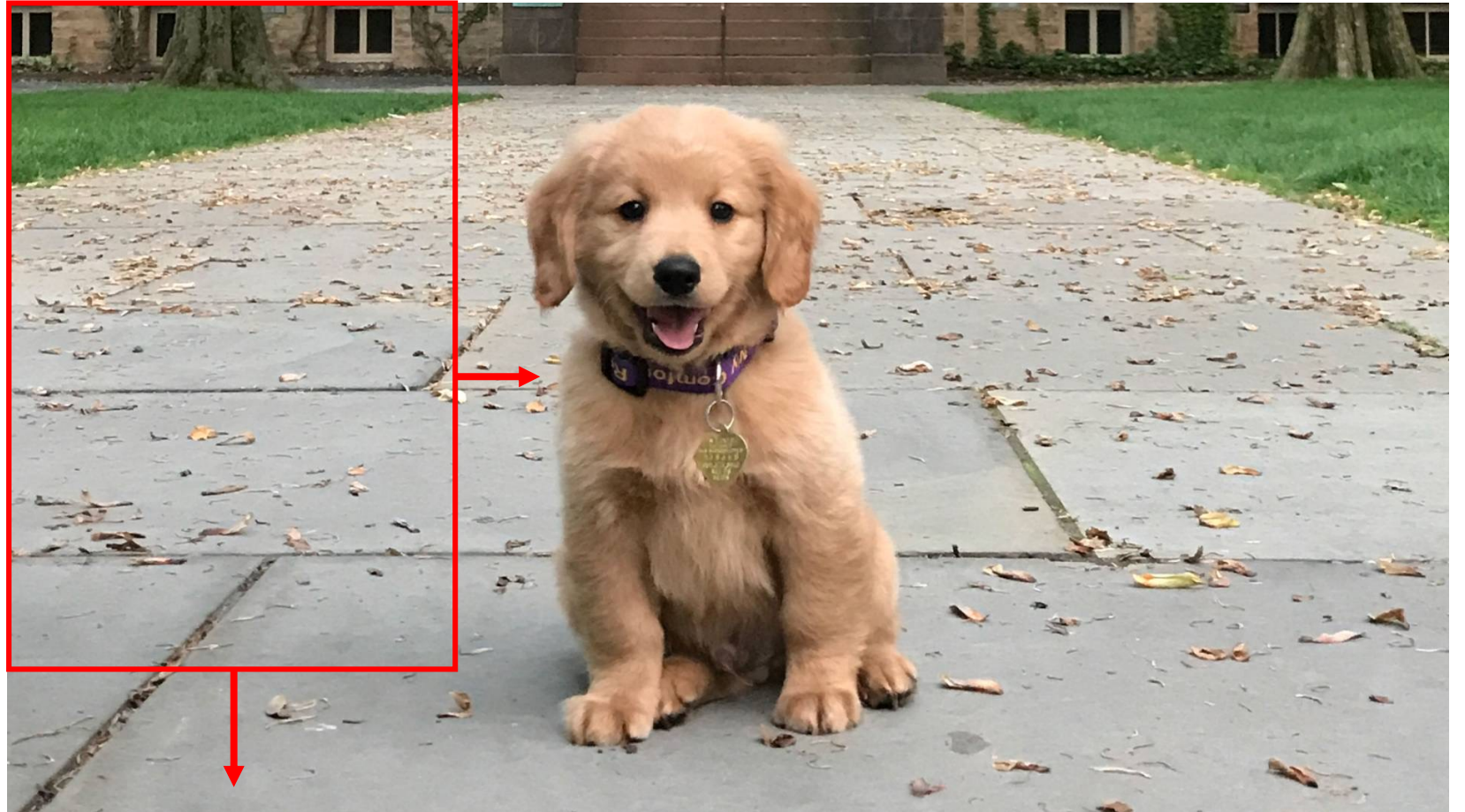 - **Where** the objects are

Classification

Localization

Detection

Previous Methods

- Sliding windows

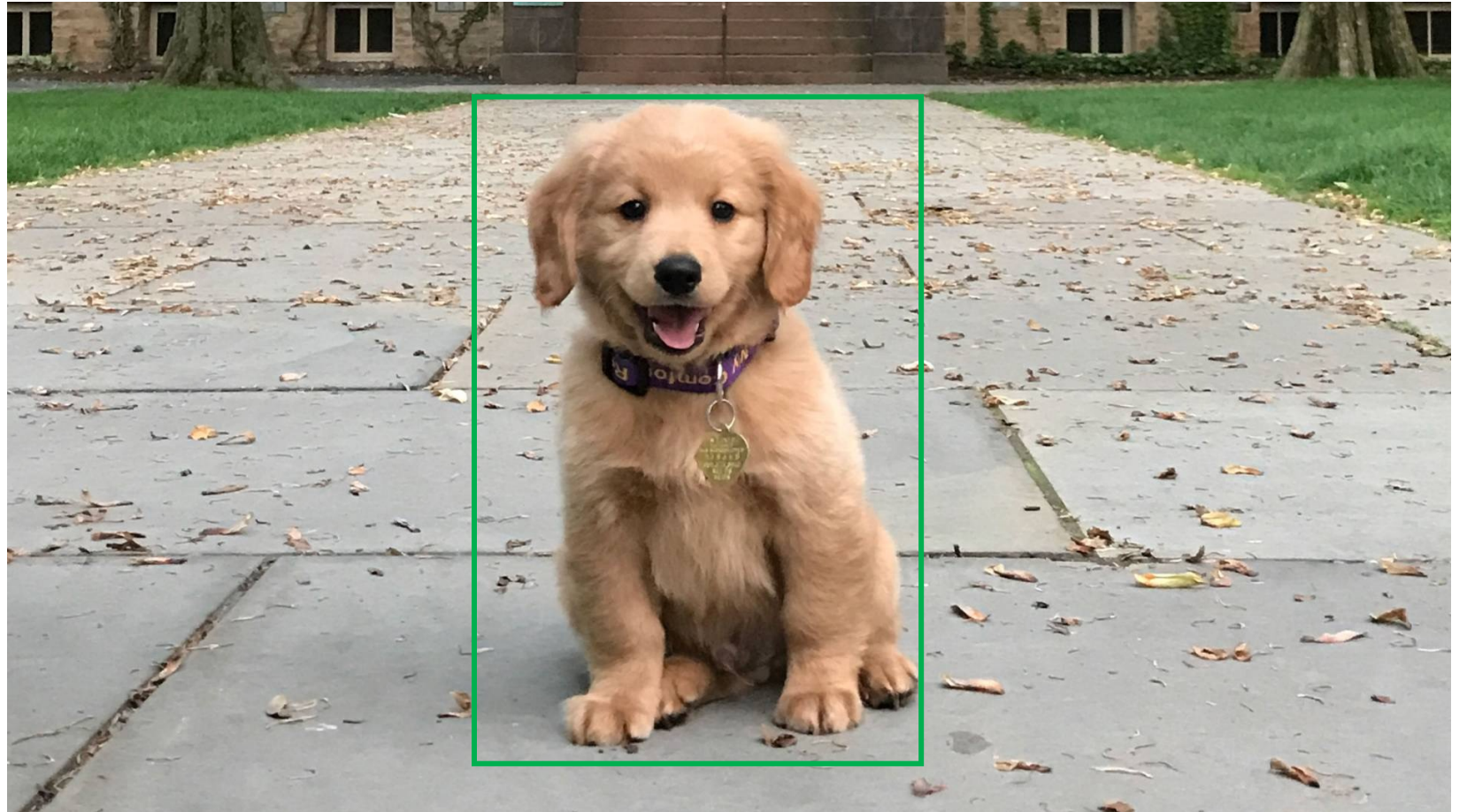# Previous Methods

- Sliding windows

Previous Methods

- Sliding windows

Previous Methods

- RCNN

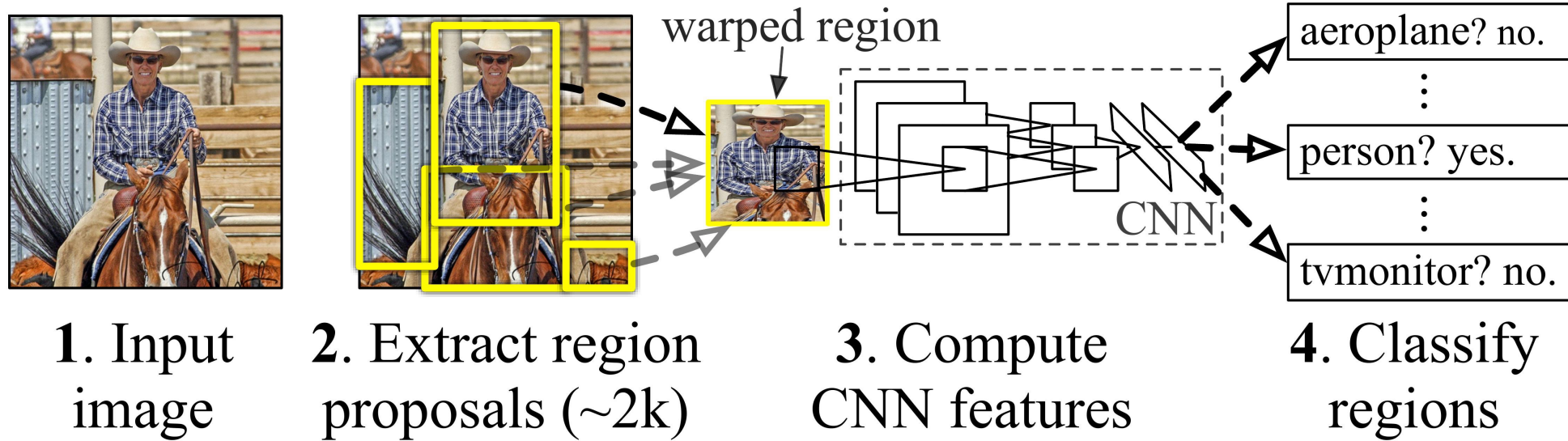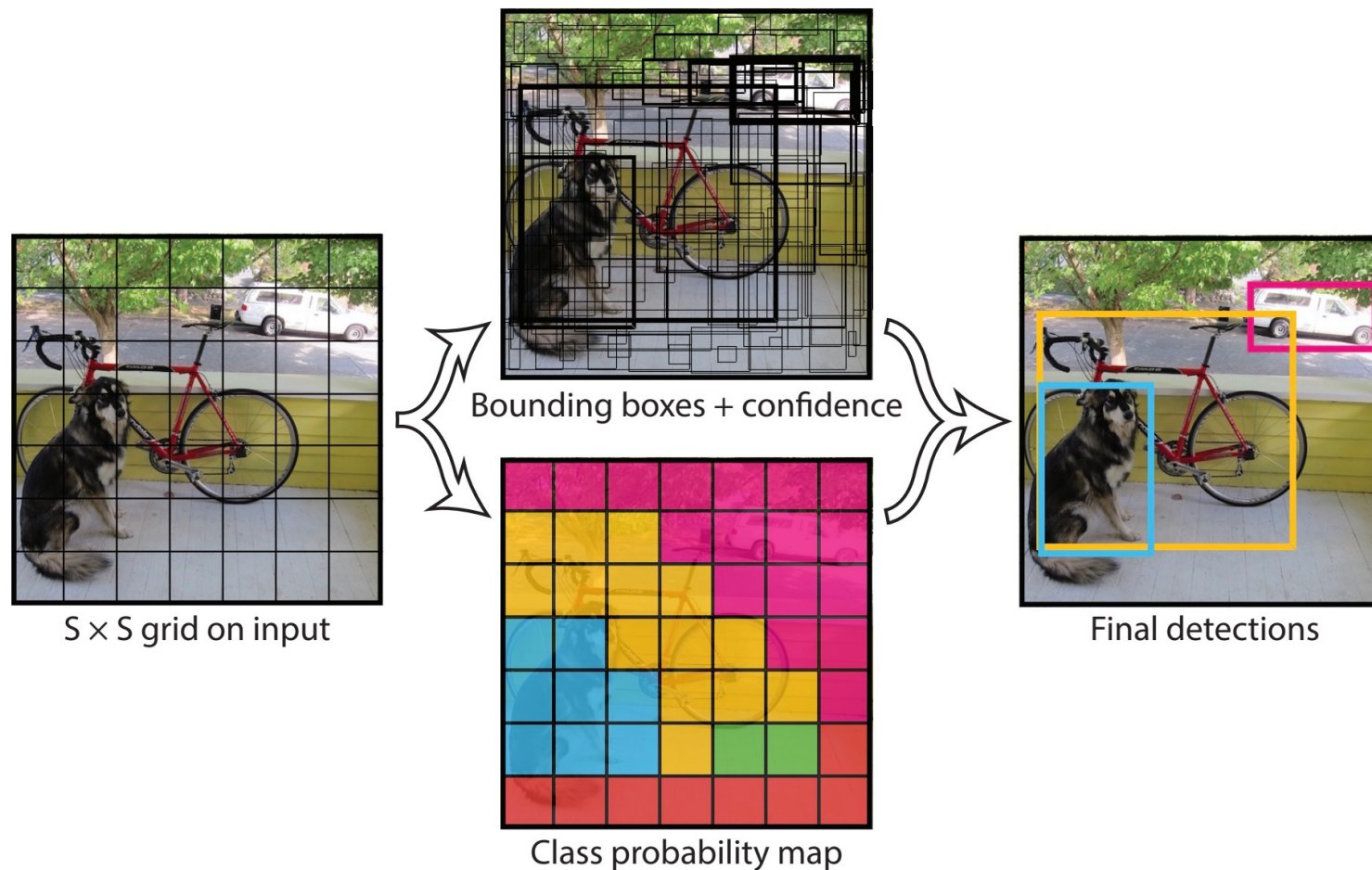**R-CNN:** *Regions with CNN features*

warped region

aeroplane? no.

person? yes.

tvmonitor? no.

CNN

1. Input image   **2**. Extract region proposals (~2k)   **3**. Compute CNN features   **4**. Classify regions

Previous Methods

- RCNN

R-CNN: *Regions with CNN features*

warped region

aeroplane? no.

person? yes.

tvmonitor? no.

CNN

1. Input image

2. Extract region proposals (~2k)

3. Compute CNN features
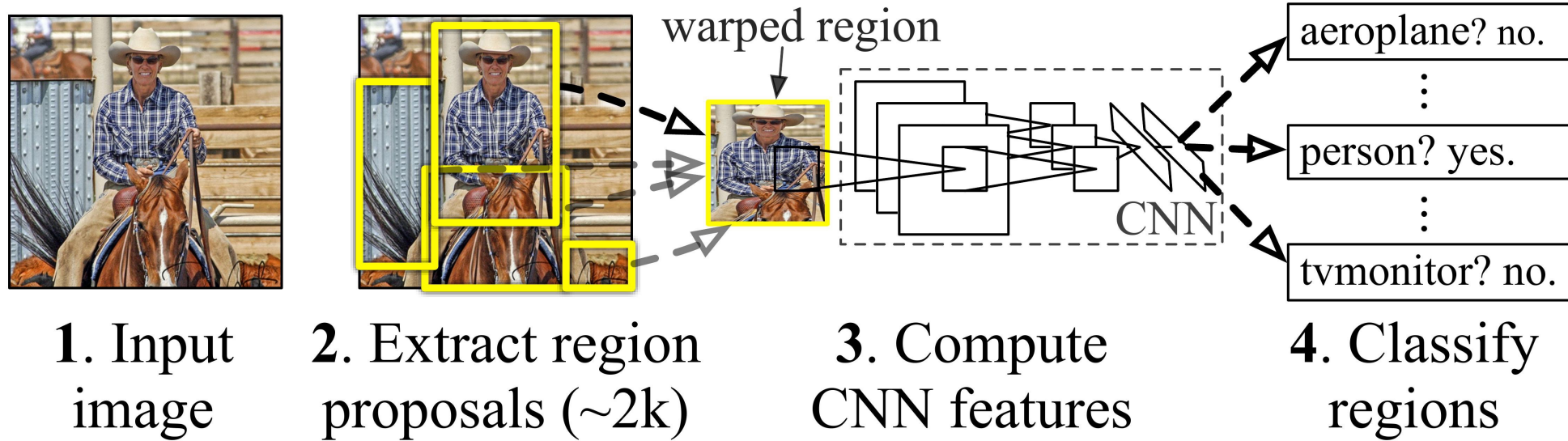
4. Classify regions

- Fast RCNN
- Faster RCNN

YOLO (You only look once)

Bounding boxes + confidence

S × S grid on input

Class probability map

Final detections

YOLO (You only look once)

- What does the Ground truth data look like?
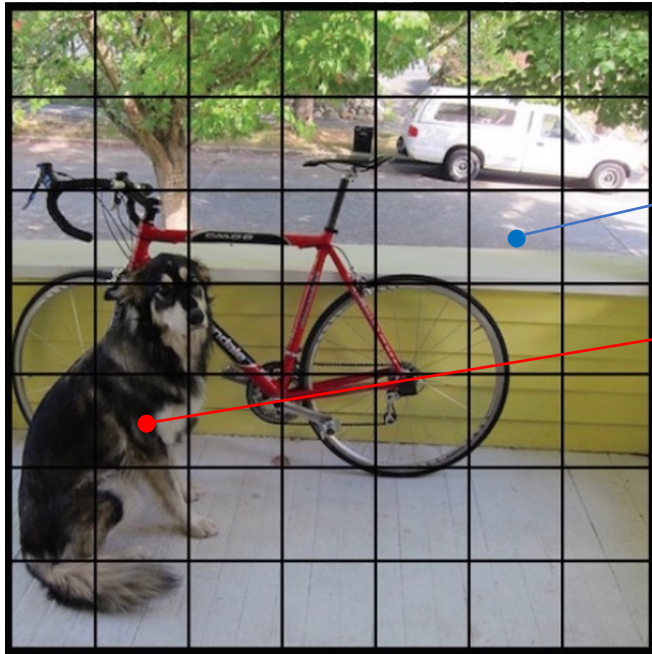
$(\textbf{\textit{Class Probabilities}}, \textbf{\textit{Box Confidence}}, \textbf{\textit{Box Coordinates}})$

20   1   4

$(0, \dots, 0, 0, 0, 0, 0, 0)$

$(0, 1, 0, \dots, 0, 1, 0.2, 0.6, 2.5, 1.8)$

GT Data: $(S \times S \times (CP + 5)) = (7 \times 7 \times 25)$

YOLO (You only look once)

- What does the Prediction look like?

$$(\textit{Class Probabilities}, \textit{Box 1 Confidence}, \textit{Box 1 Coordinates}, \textit{Box 2 Confidence}, \textit{Box 2 Coordinates}))$$
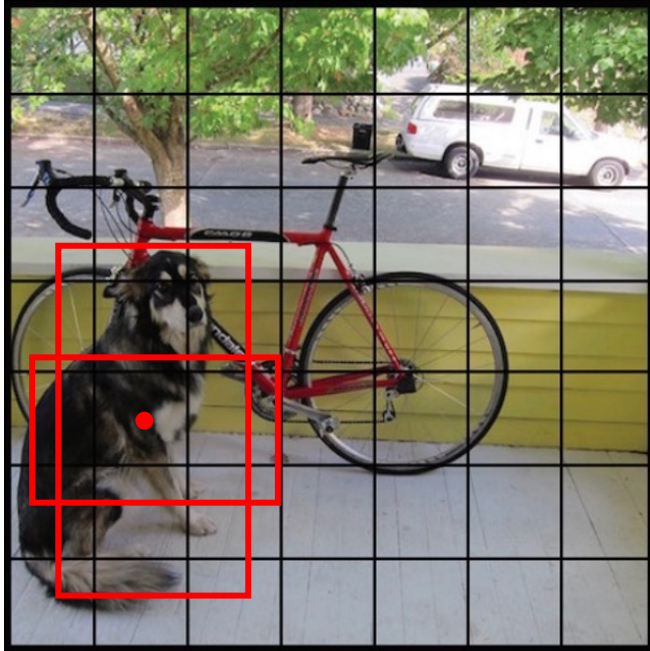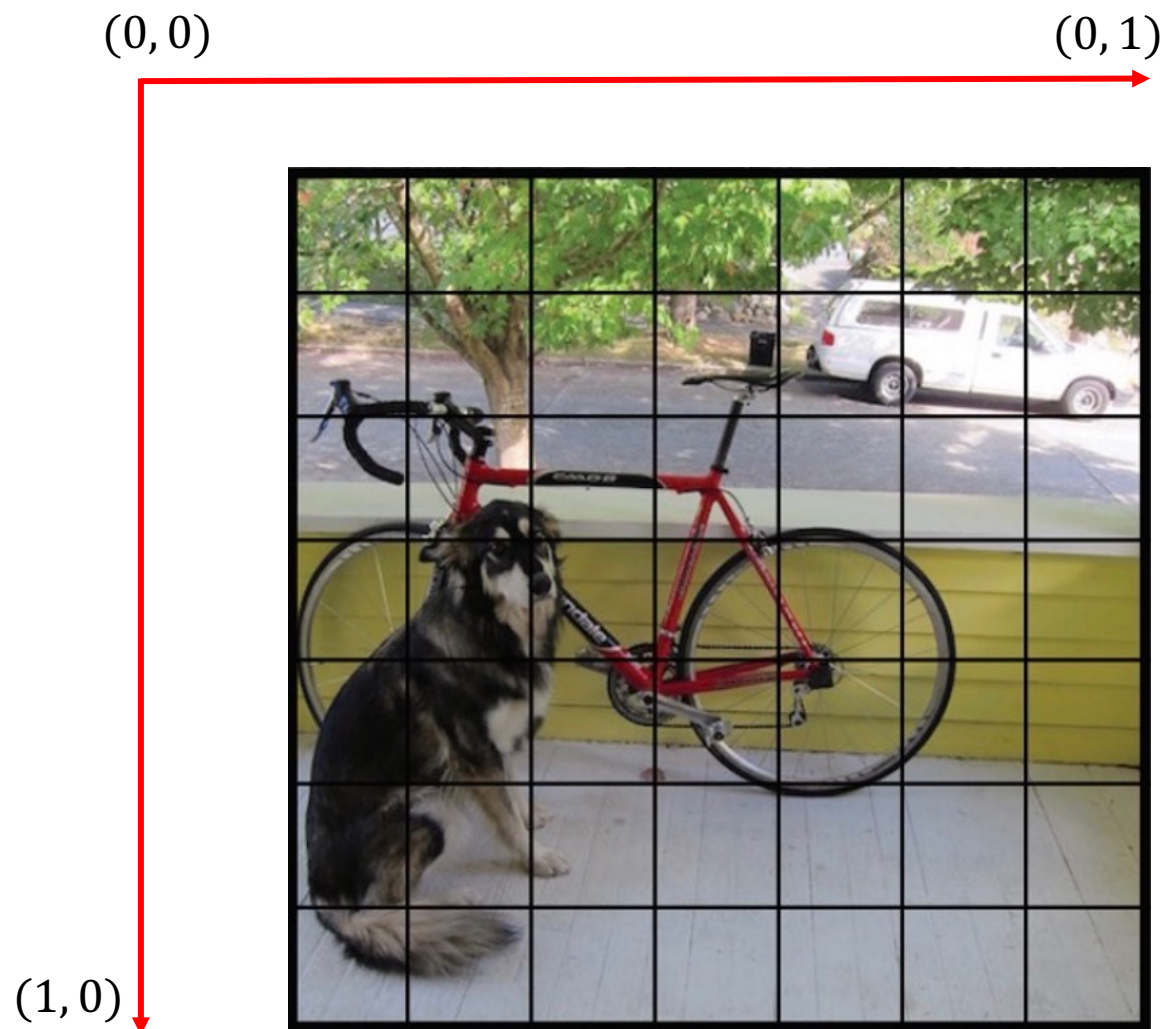
20	1	4	1	4

GT Data: $(S \times S \times B \times (CP + 5)) = (7 \times 7 \times 30)$

YOLO (You only look once)

- Coordinate convention

$(0, 0)$ $(0, 1)$

$(1, 0)$
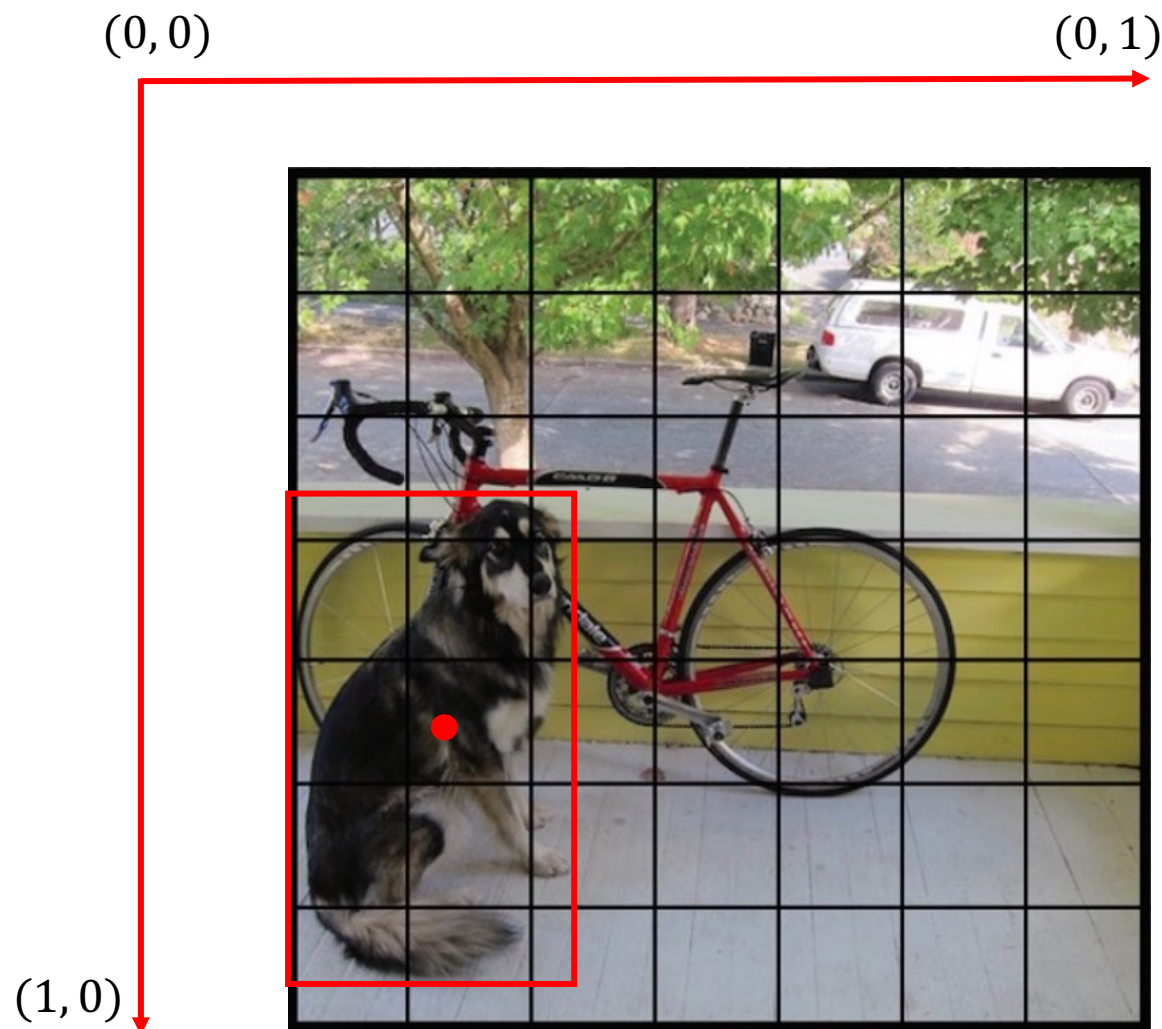
YOLO (You only look once)

- Coordinate convention

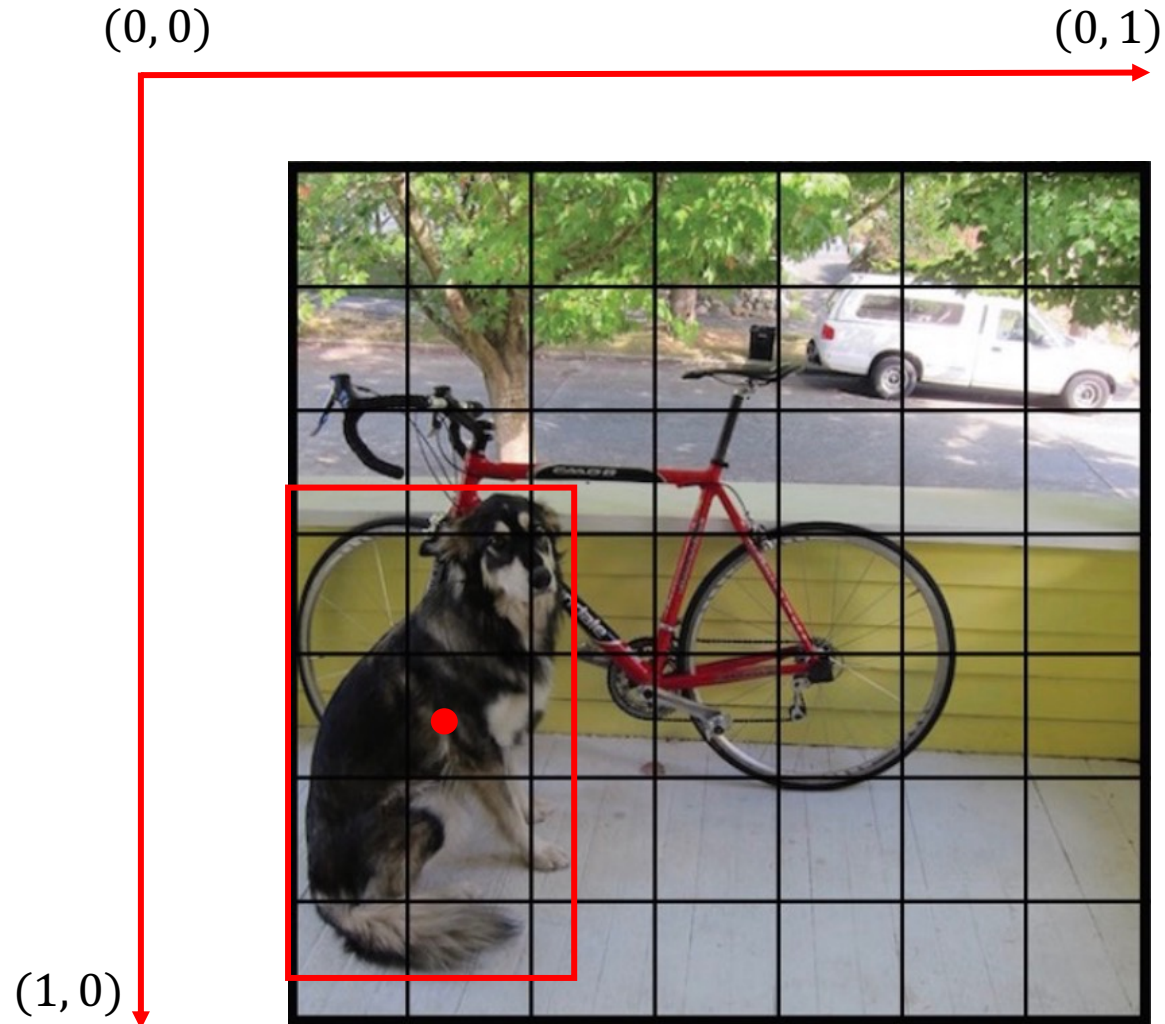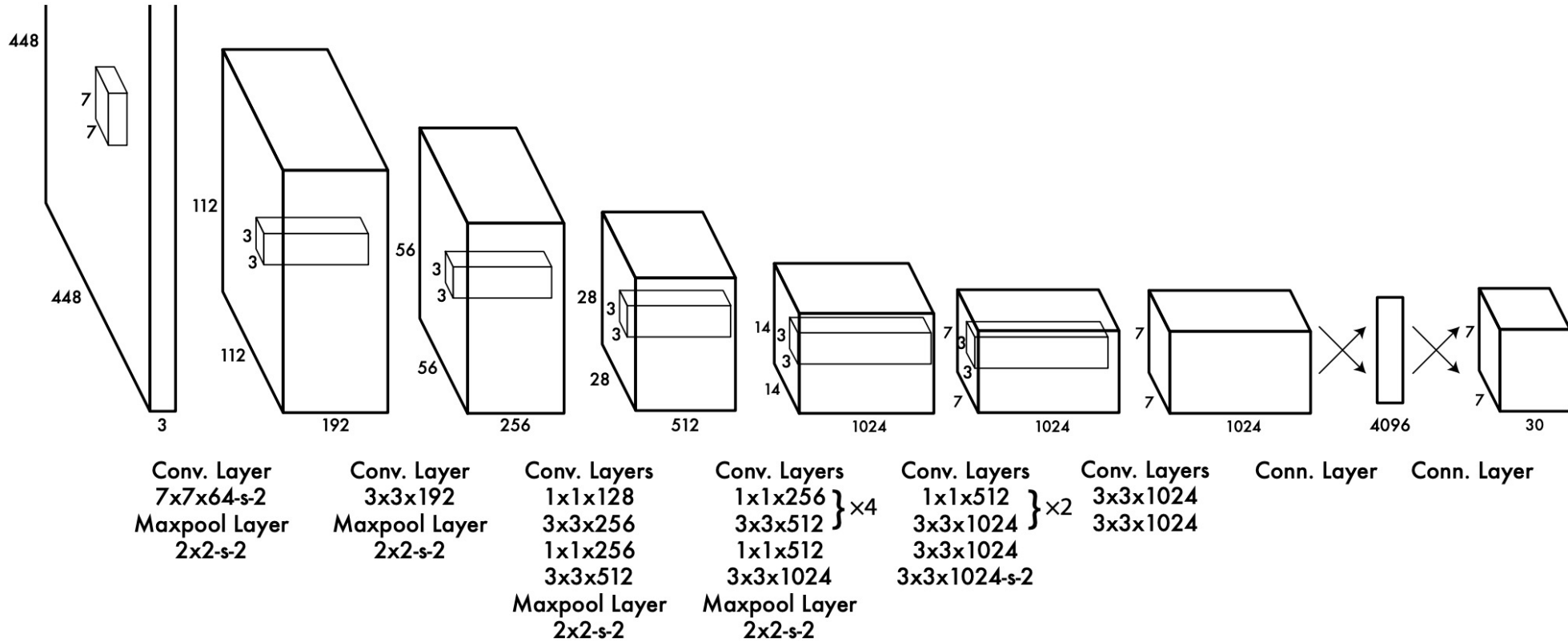(x, y, h, w)

$(0.6, 0.15, 0.6, 0.3)$

YOLO (You only look once)

- Coordinate convention

$$(x_{cell}, y_{cell}, h_{cell}, w_{cell})$$

$$(0.5, 0.2, 4.2, 2.1)$$

YOLO (You only look once)

| Conv. Layer	Conv. Layer	Conv. Layers	Conv. Layers	Conv. Layers	Conv. Layers	Conn. Layer	Conn. Layer
7x7x64-s-2	3x3x192	1x1x128	1x1x256 }×4	1x1x512 }×2	3x3x1024		
Maxpool Layer	Maxpool Layer	3x3x256	3x3x512	3x3x1024	3x3x1024		
2x2-s-2	2x2-s-2	1x1x256	1x1x512	3x3x1024			
		3x3x512	3x3x1024	3x3x1024-s-2			
		Maxpool Layer	Maxpool Layer				
		2x2-s-2	2x2-s-2				

Detour: Intersection over Union

- How good is a detection?

Detour: Intersection over Union
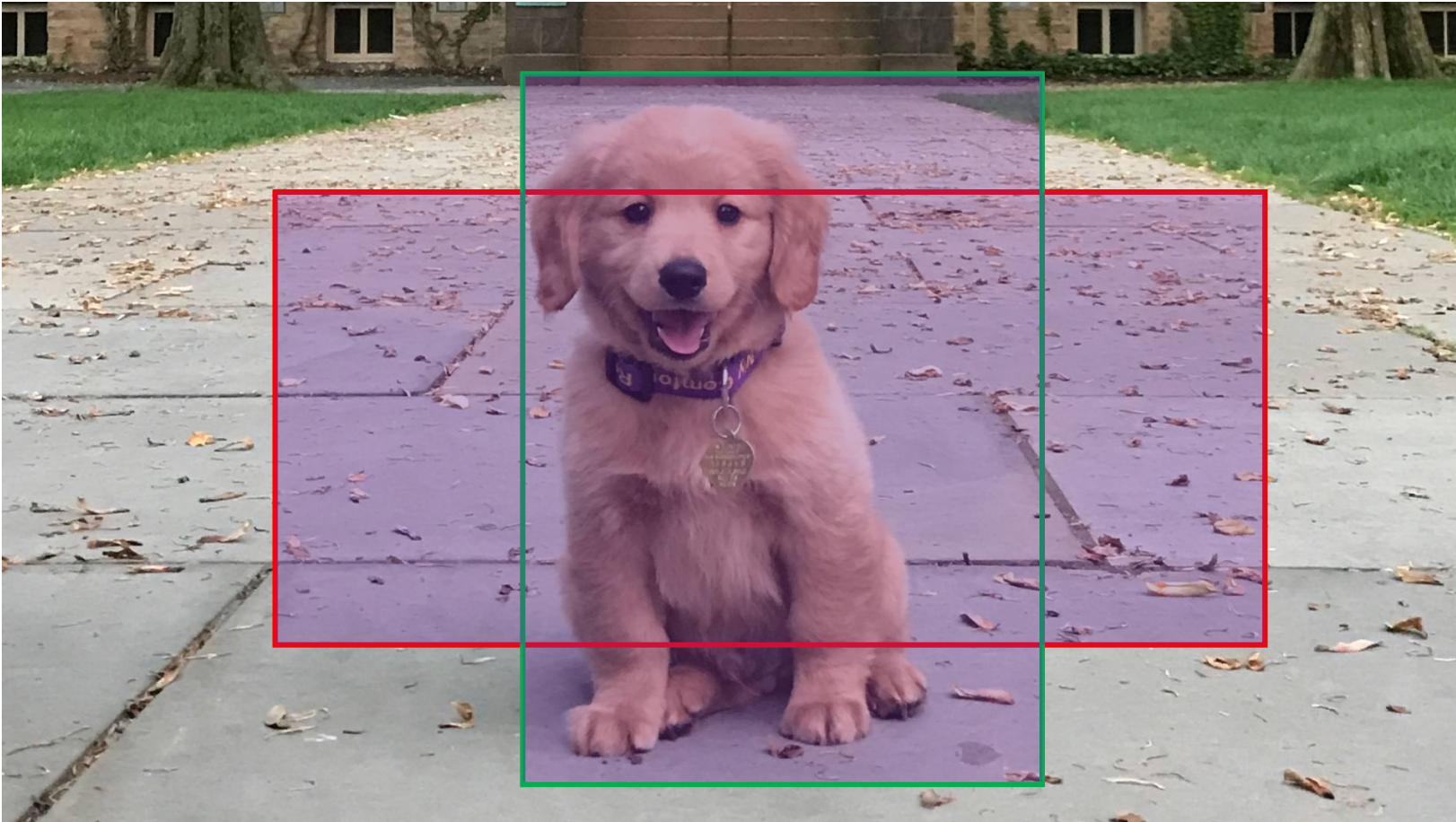
- How good is a detection?
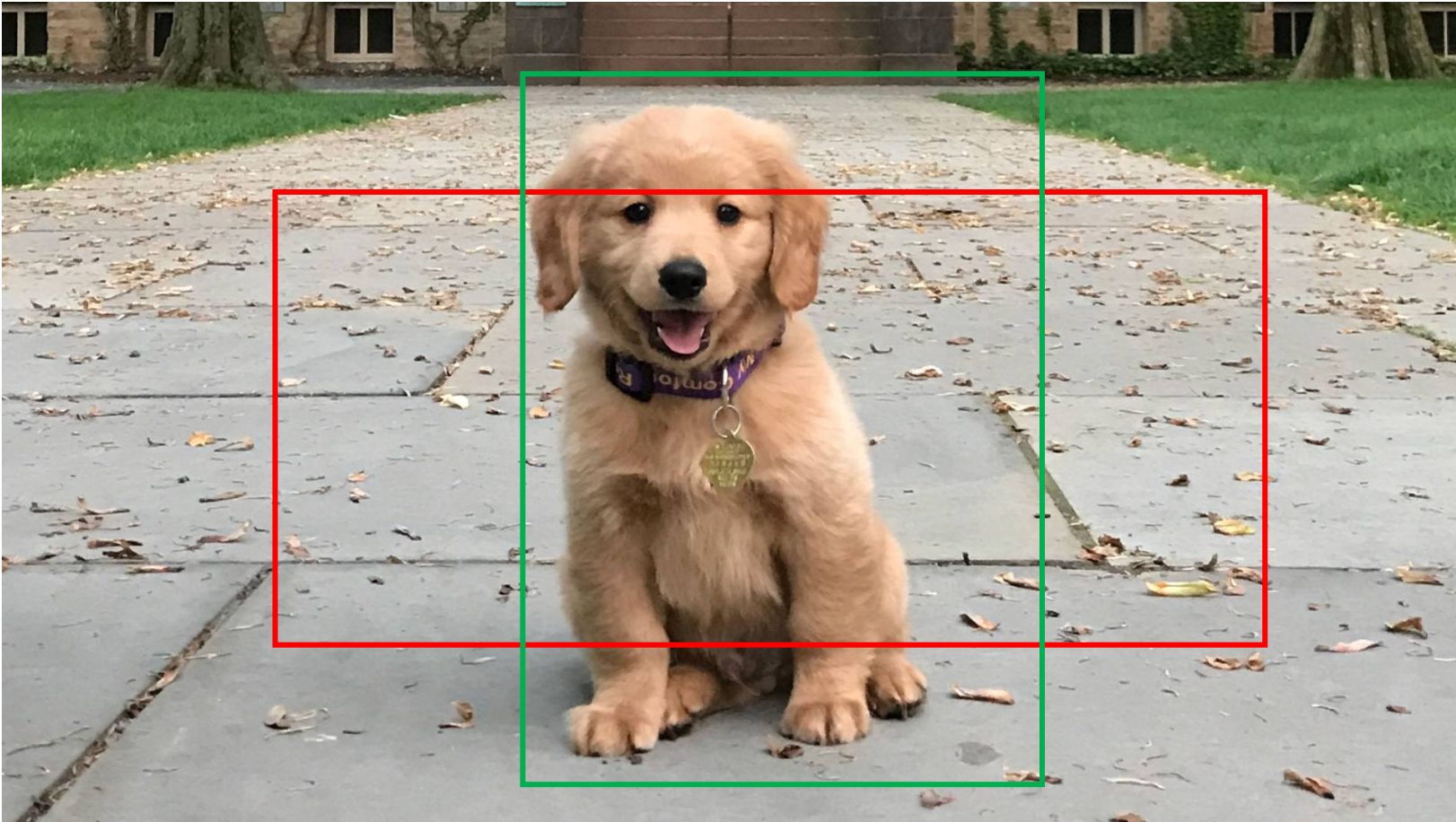
$$IoU = \frac{A_{Intersection}}{}$$

Detour: Intersection over Union

- How good is a detection?
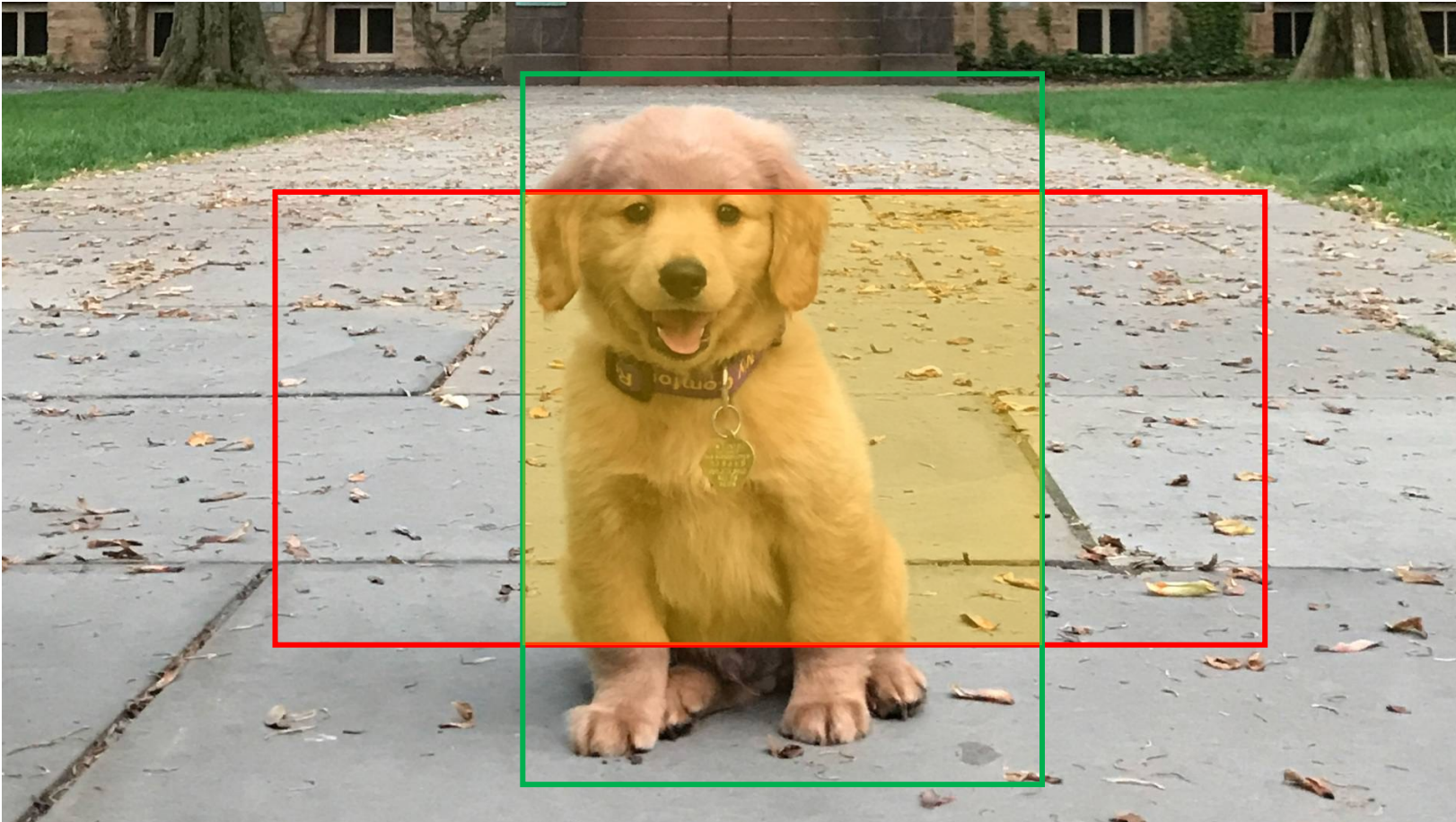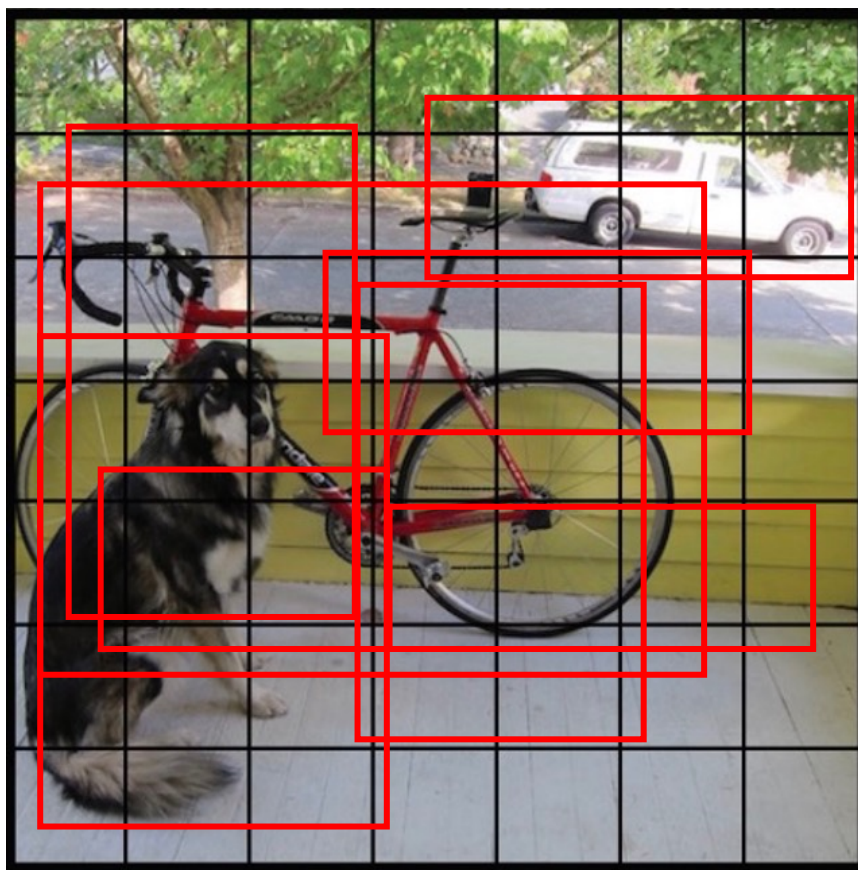

$$IoU = \frac{A_{Intersection}}{A_{Union}}$$
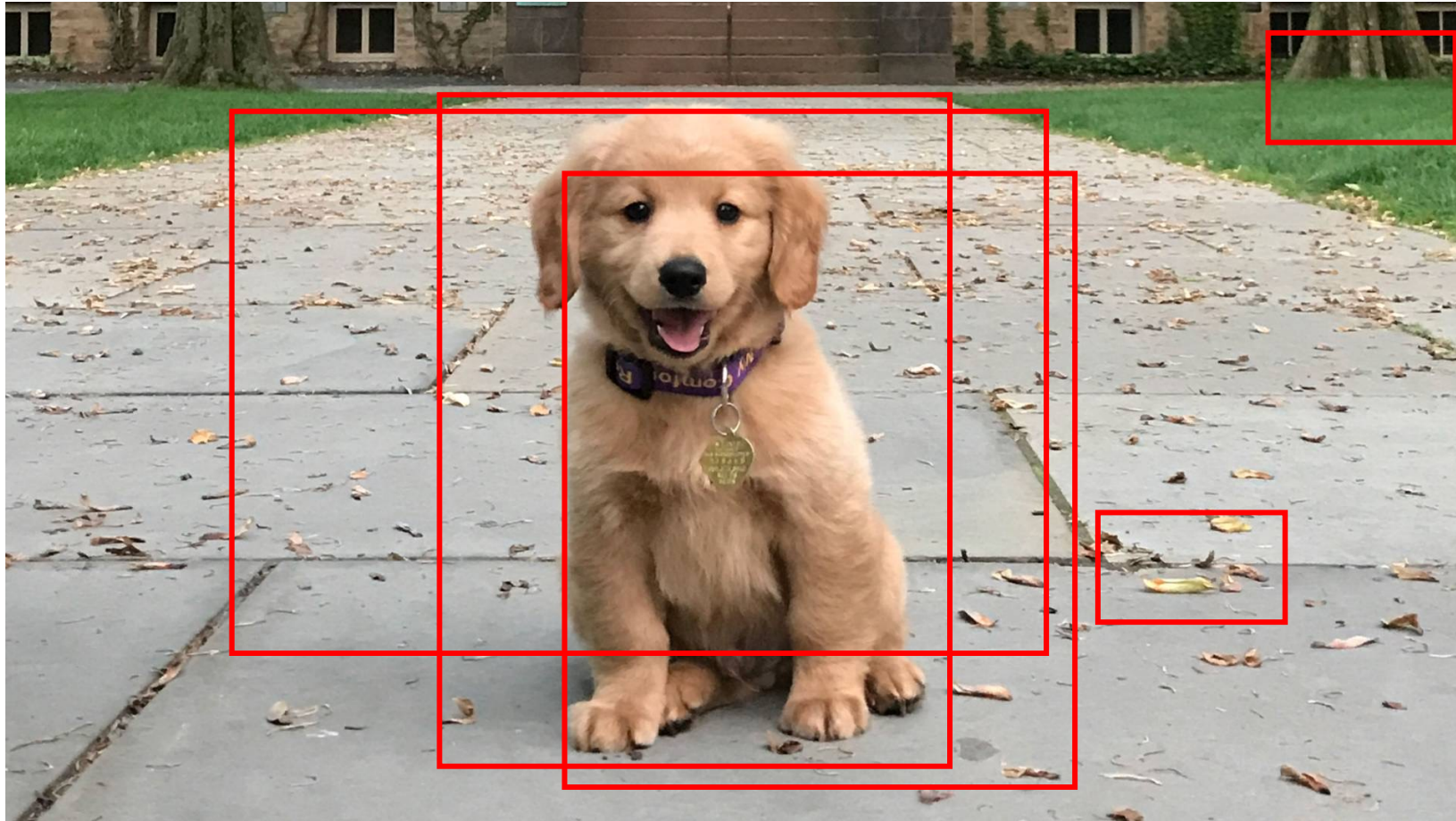
YOLO (You only look once)

- Loss function

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

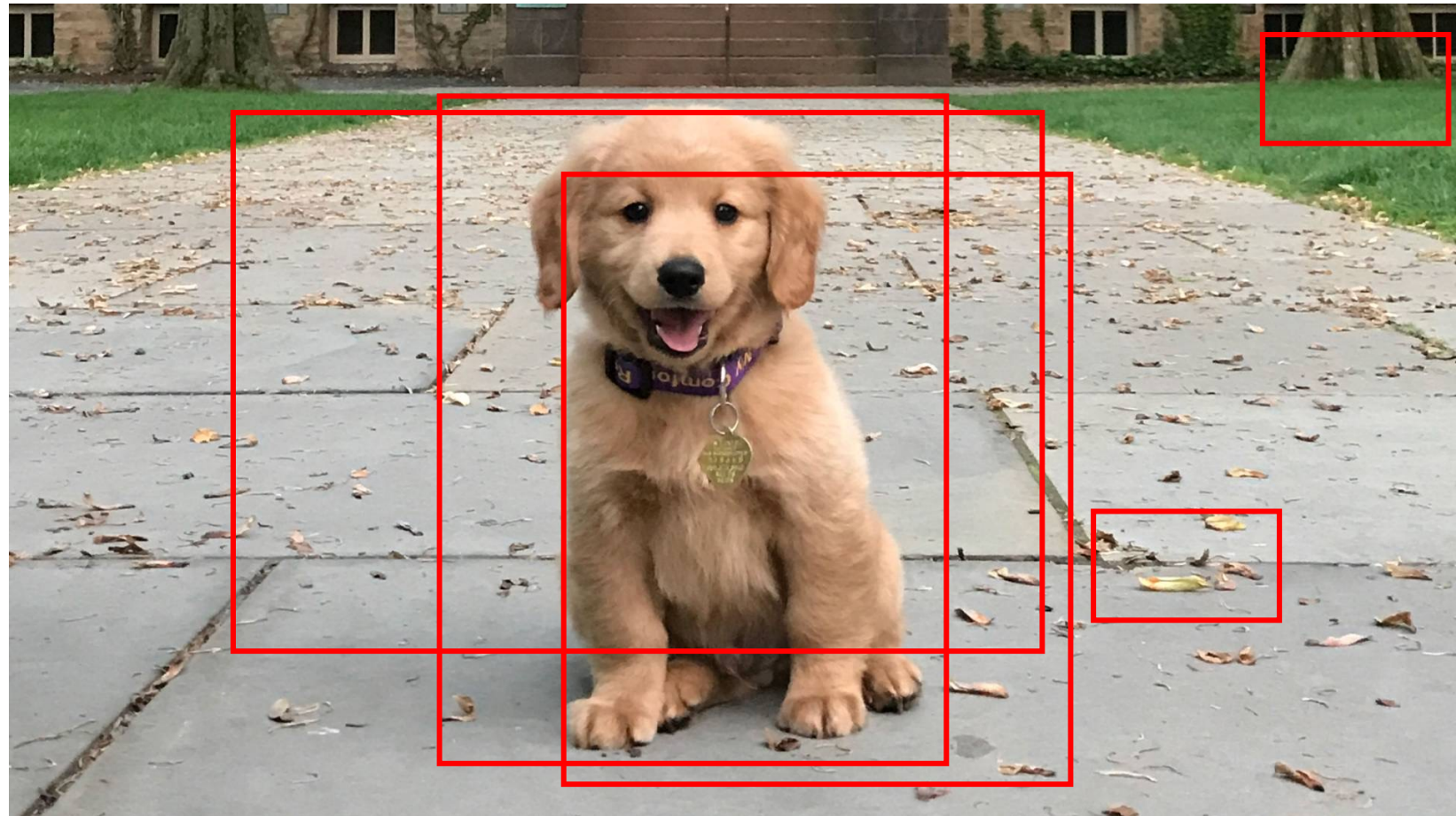# YOLO (You only look once)

- Inference

Detour: Non maximum suppression

Detour: Non maximum suppression
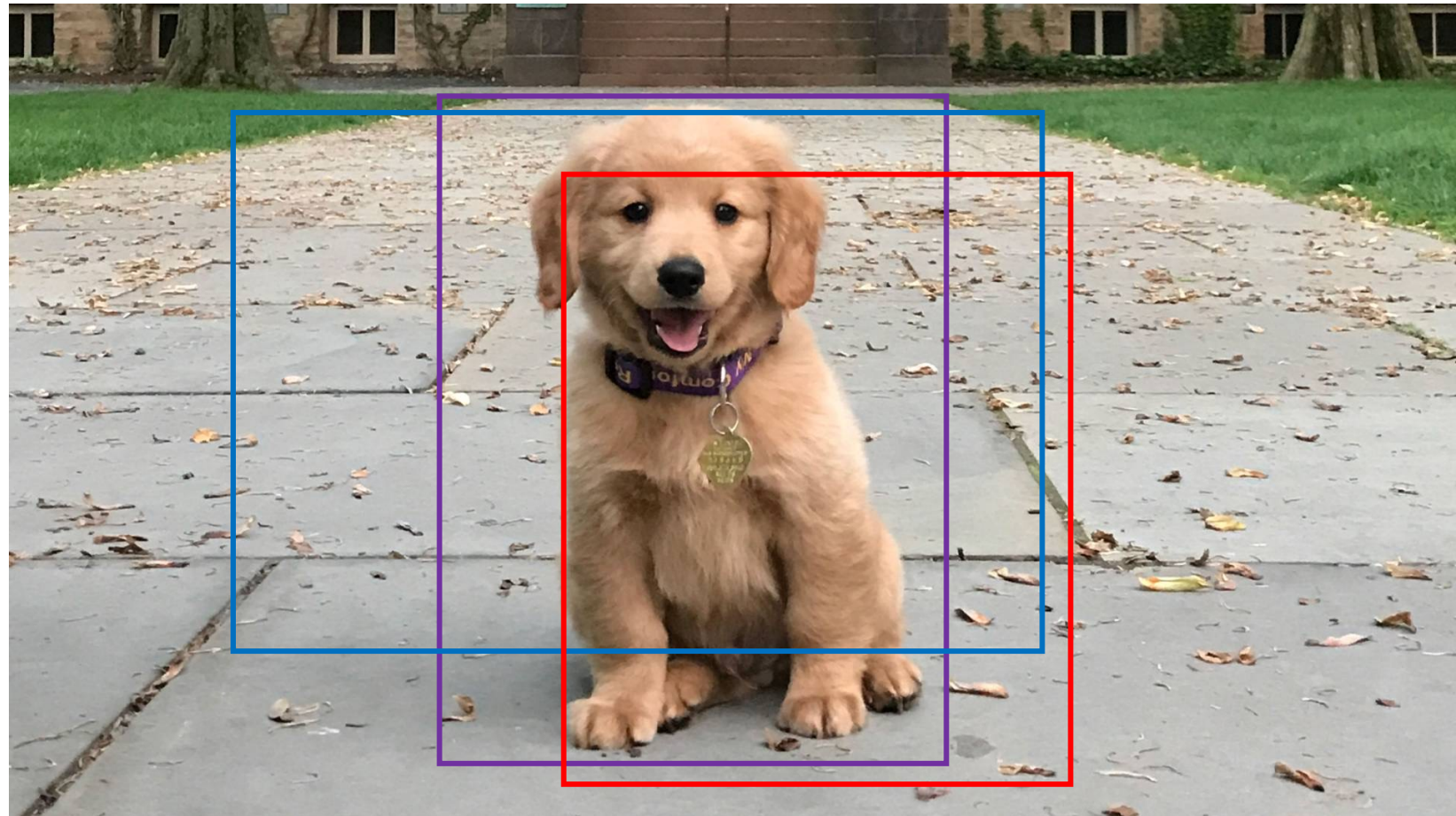
- Discard low confidence predictions

Detour: Non maximum suppression

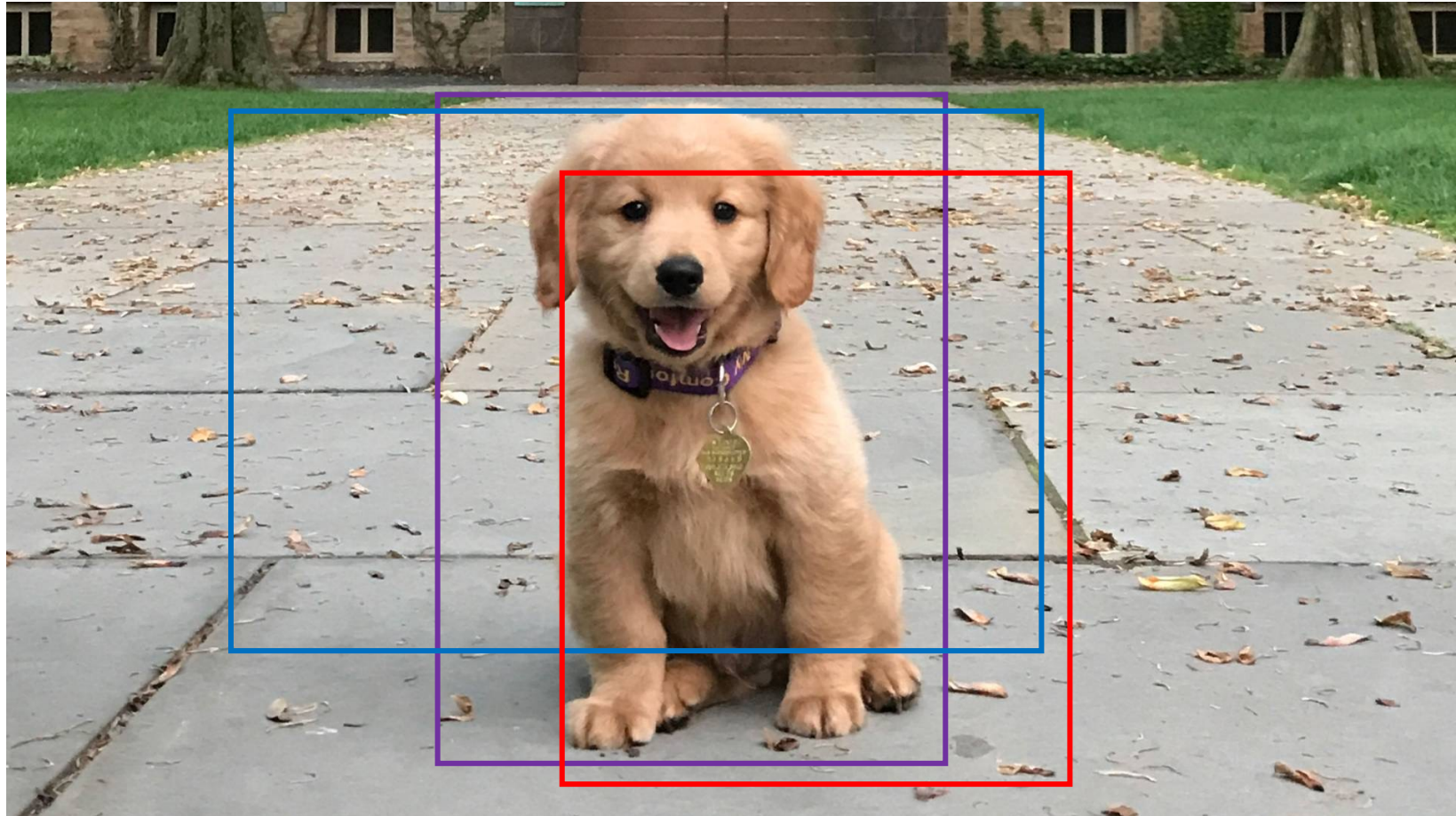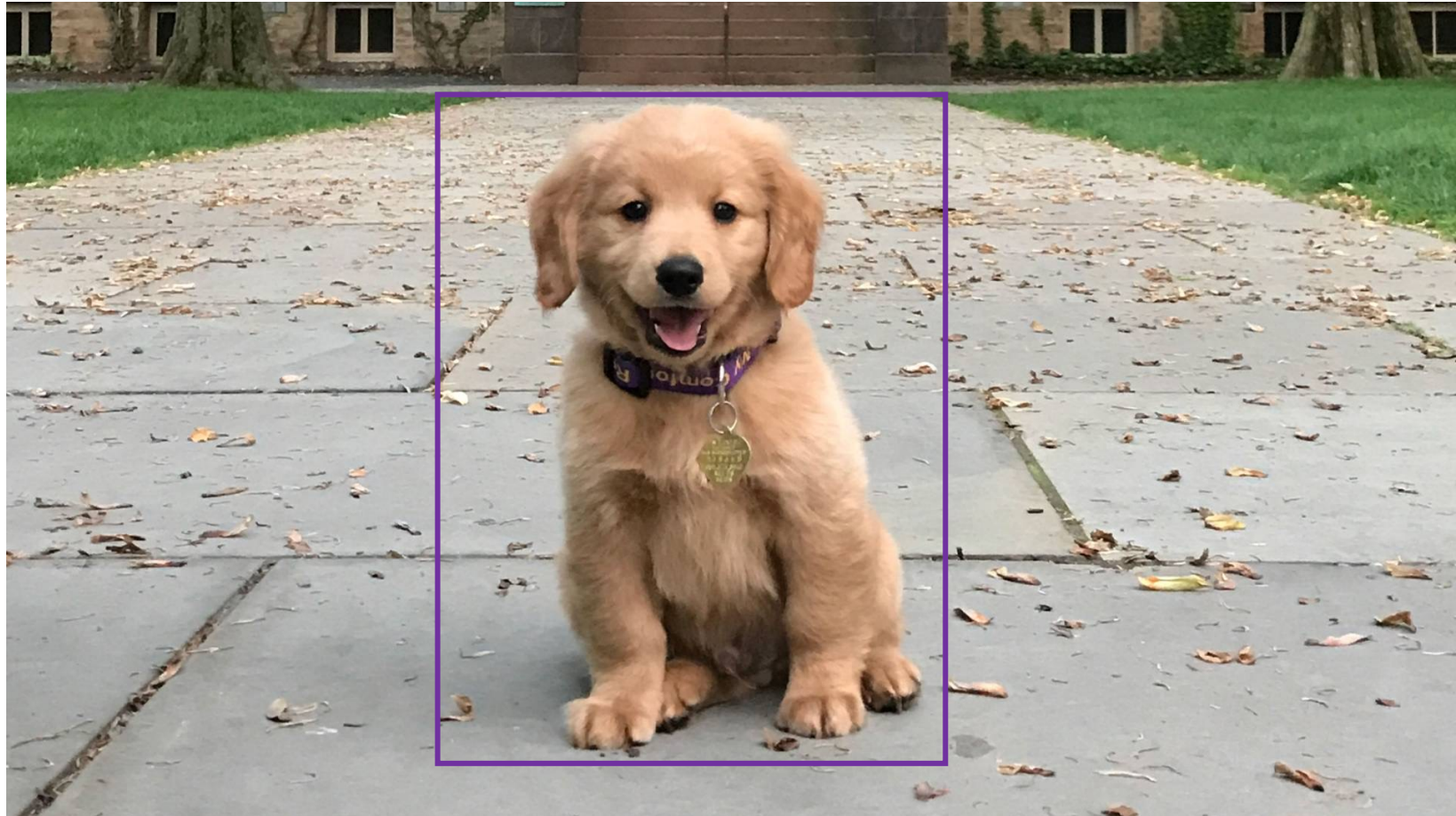- Sort remaining boxes by Confidence score

0.9
0.7
0.4

Detour: Non maximum suppression

- Compute IoUs of all boxes with the highest confidence box

- If IoU > threshold discard box

Detour: Non maximum suppression

- Final result
- Repeat for every class

References

- https://www.youtube.com/watch?v=6ykvU9WuIws
- https://www.youtube.com/watch?v=t-phGBfPEZ4&list=PLhhyoLH6Ijfw0TpCTVTNk42NN08H6UvNq&index=1
- https://arxiv.org/pdf/1506.02640.pdf