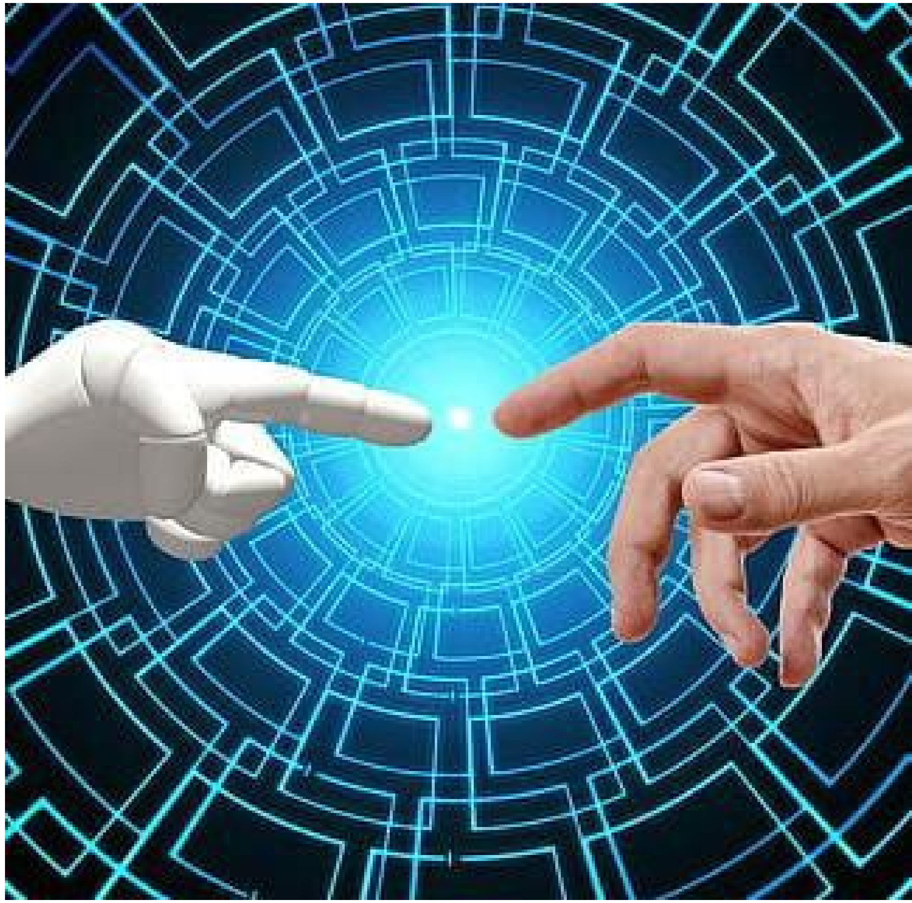
Artificial Intelligence (AI) for Engineering

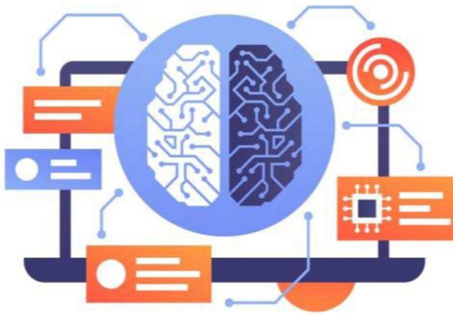


COS40007

Dr. Afzal Azeem Chowdhary

Lecturer, SoCET, Swinburne University of Technology

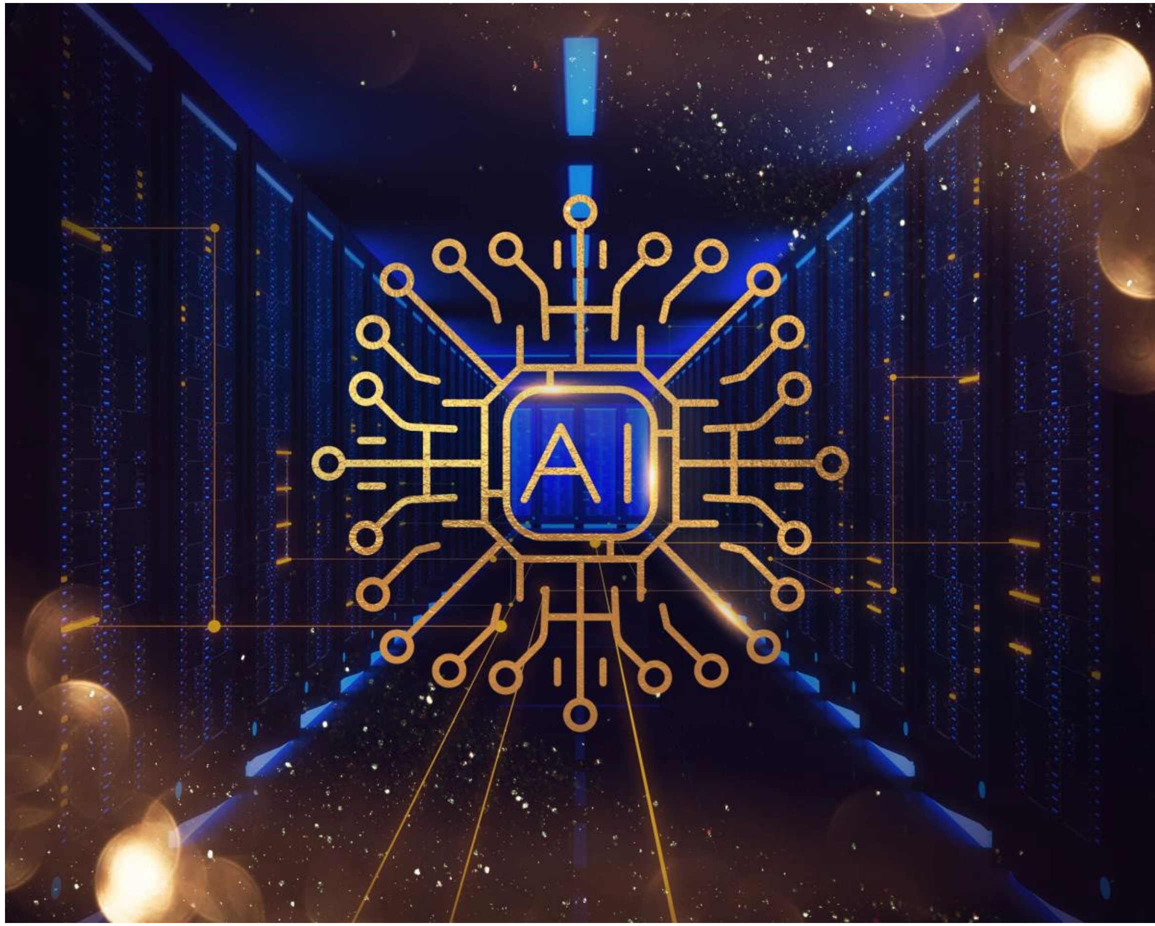
Seminar 6: 8th April 2025



SWINBURNE UNIVERSITY OF TECHNOLOGY

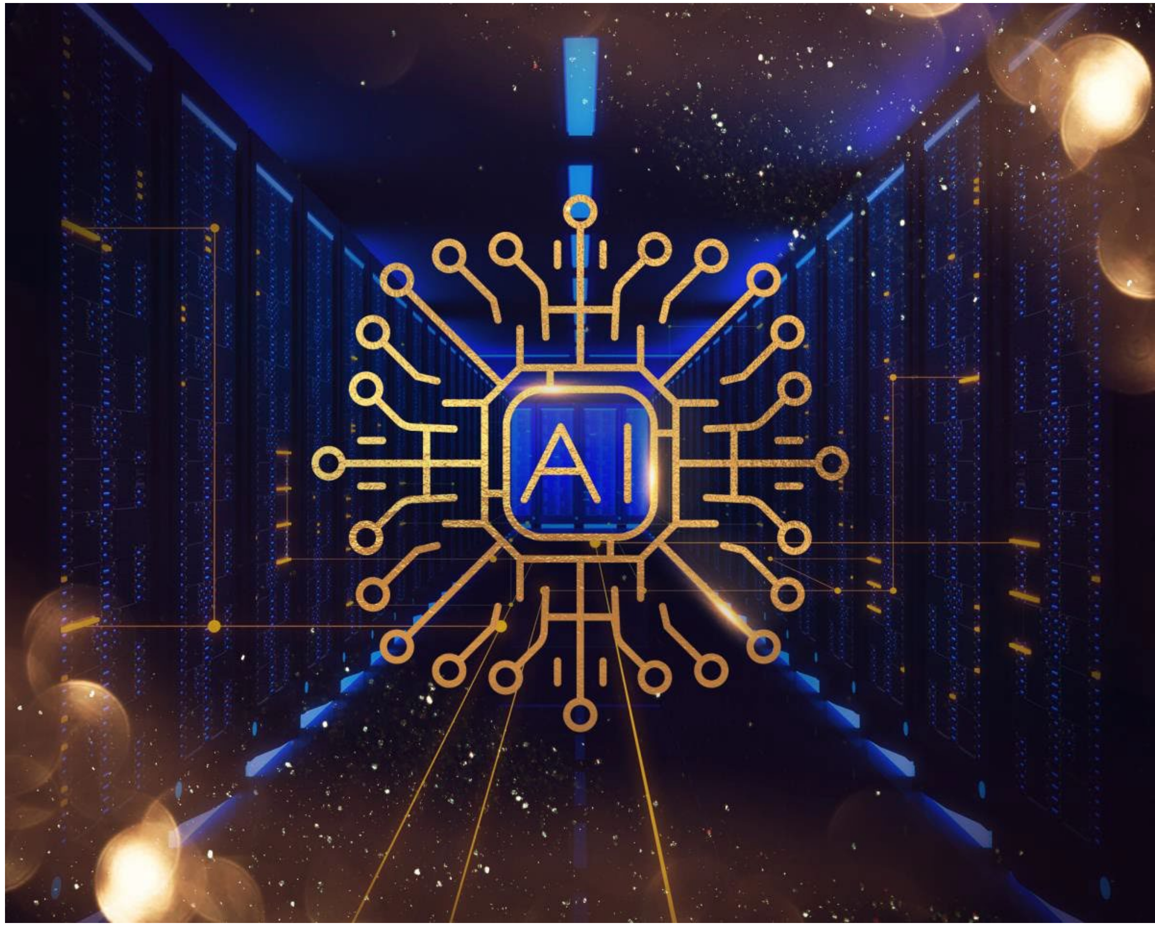
Overview

- ❑ Object Detections
- ❑ YOLO
- ❑ Evaluation of YOLO model
- ❑ Benefit of YOLO
- ❑ Examples of object detections



Required Reading

–Chapter 12 of “Applied Machine Learning and AI for Engineers”



At the end of this you should be able to

- Understand how to perform Object detections using YOLO.
- Understand how to train and evaluate a model using YOLO.
- Understand Evaluation measures for Object Detections.

Object Detections

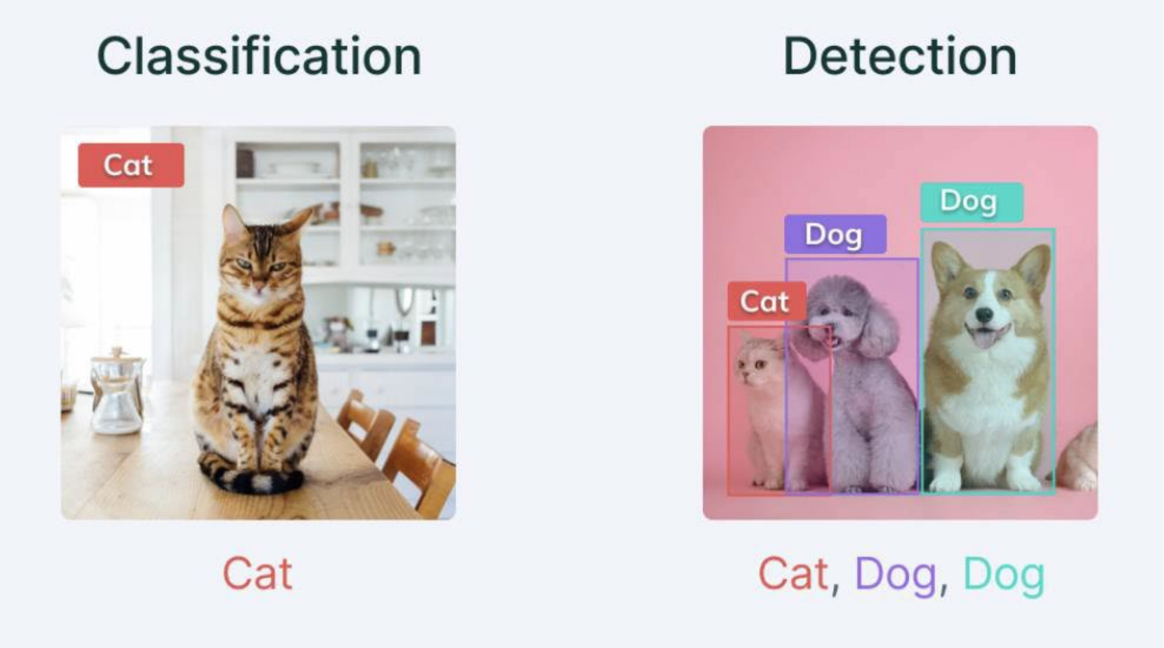
SWINBURNE UNIVERSITY OF TECHNOLOGY

Object Detections

It is a phenomenon in computer vision that involves the detection of various objects (eg. people, cars, chairs, road signs, buildings, and animals) in digital images or videos.
This phenomenon answers two basic questions:

What is the object?
This question seeks to identify the object in a specific image.

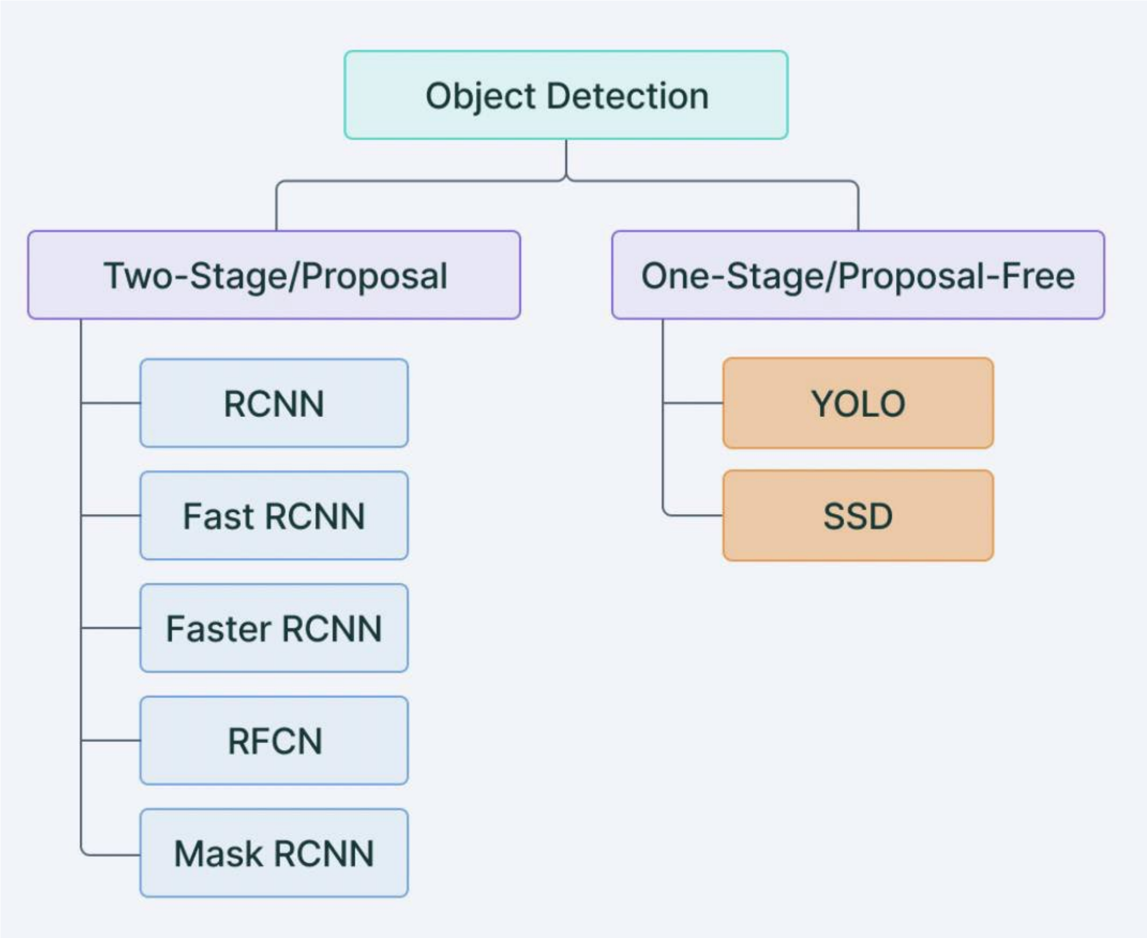
Where is it?
This question seeks to establish the exact location of the object within the image.



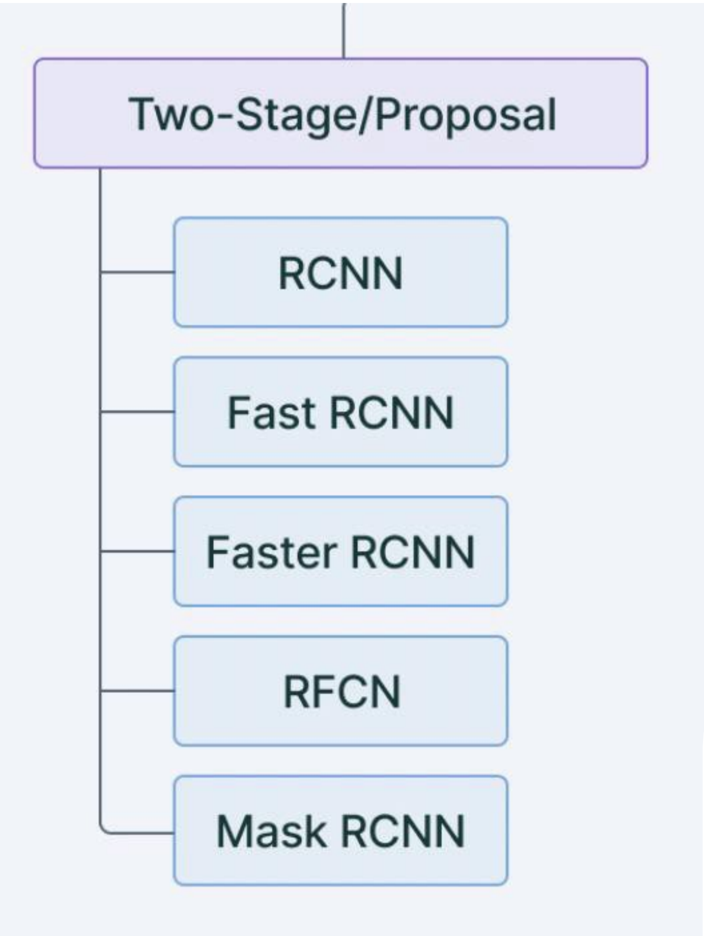
Two-stage Object Detection

Two-stage object detection refers to the use of algorithms that break down the object detection problem statement into the following two-stages:

- Detecting possible object regions.
- Classifying the image in those regions into object classes.
- One way to apply deep learning to the task of object detection is to use region-based CNNs, also known as region CNNs or simply R-CNNs.



Two-stage Object Detection

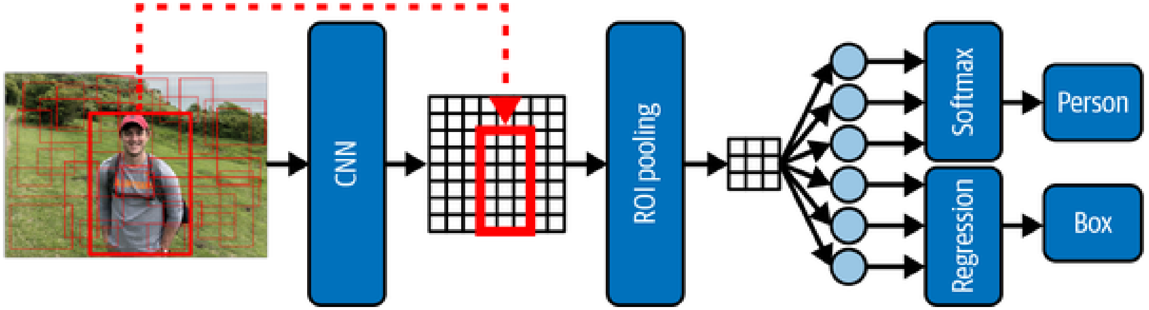


Popular two-step algorithms like Fast-RCNN and Faster-RCNN typically use a Region Proposal Network that proposes regions of interest that might contain objects. The RPN is a shallow CNN that shares layers with the main CNN.

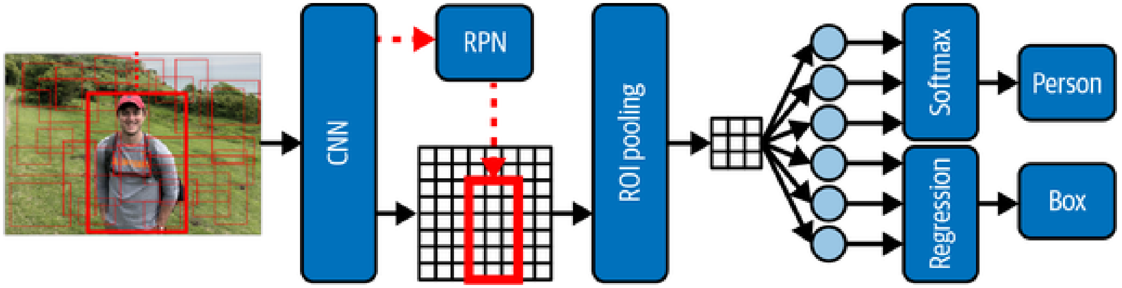
The output from the RPN) is then fed to a classifier that classifies the regions into classes.

While this gives accurate results in object detection with a high mean Average Precision (mAP), it results in multiple iterations taking place in the same image, thus slowing down the detection speed of the algorithm and preventing real-time detection.

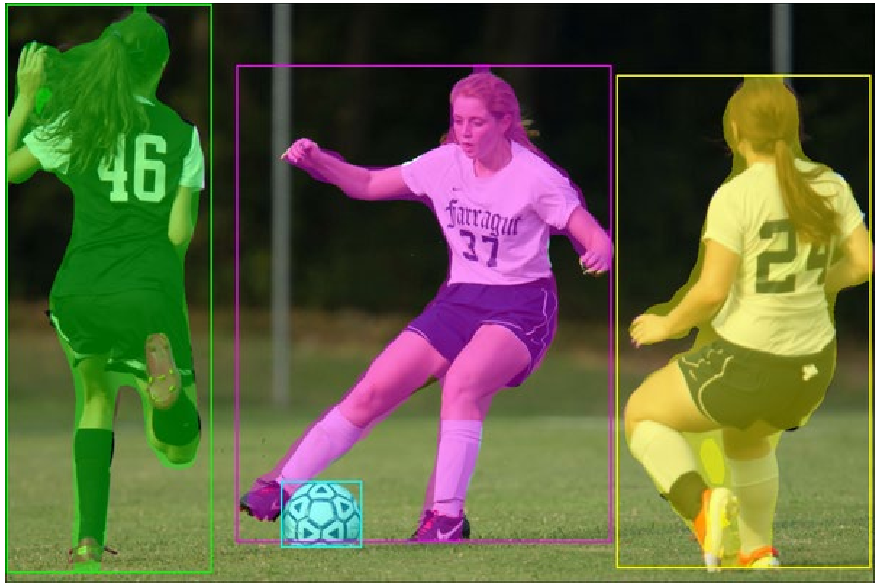
Object Detection



Fast-RCNN



Faster- RCNN



Mask R-CNNs extend Faster R-CNNs by adding *instance segmentation*, which identifies the shapes of objects detected in an image using *segmentation masks*

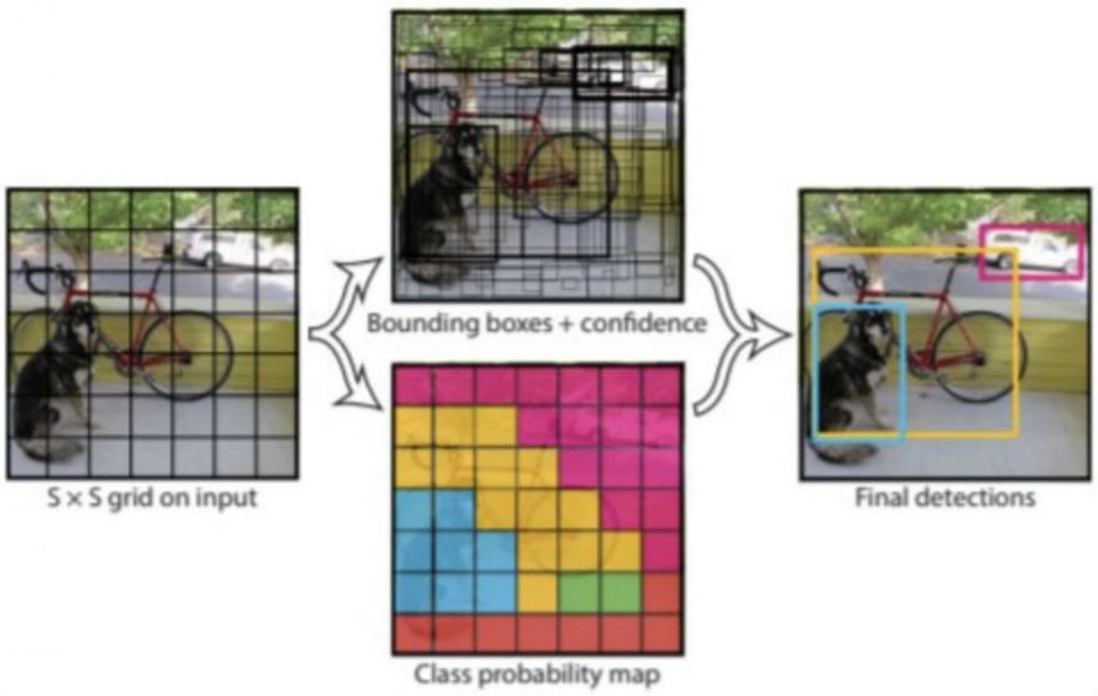
YOLO

YOLO: You Only Look Once

- YOLO is an algorithm proposed by by Redmond et. al in a research article paper titled “You Only Look Once: Unified, Real-Time Object Detection” published in 2015 proposed an alternative to R-CNNs
- In Comparison with other object detection algorithms, YOLO proposes the use of an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once.

How YOLO works

- The YOLO algorithm works by dividing the image into *N* grids, each having an equal dimensional region of *S*x*S*.
- Each of these *N* grids is responsible for the detection and localization of the object it contains.
- Correspondingly, these grids predict *B* bounding box coordinates relative to their cell coordinates, along with the object label and probability of the object being present in the cell.



Bounding Box Regression

A bounding box is an outline that highlights an object in an image.

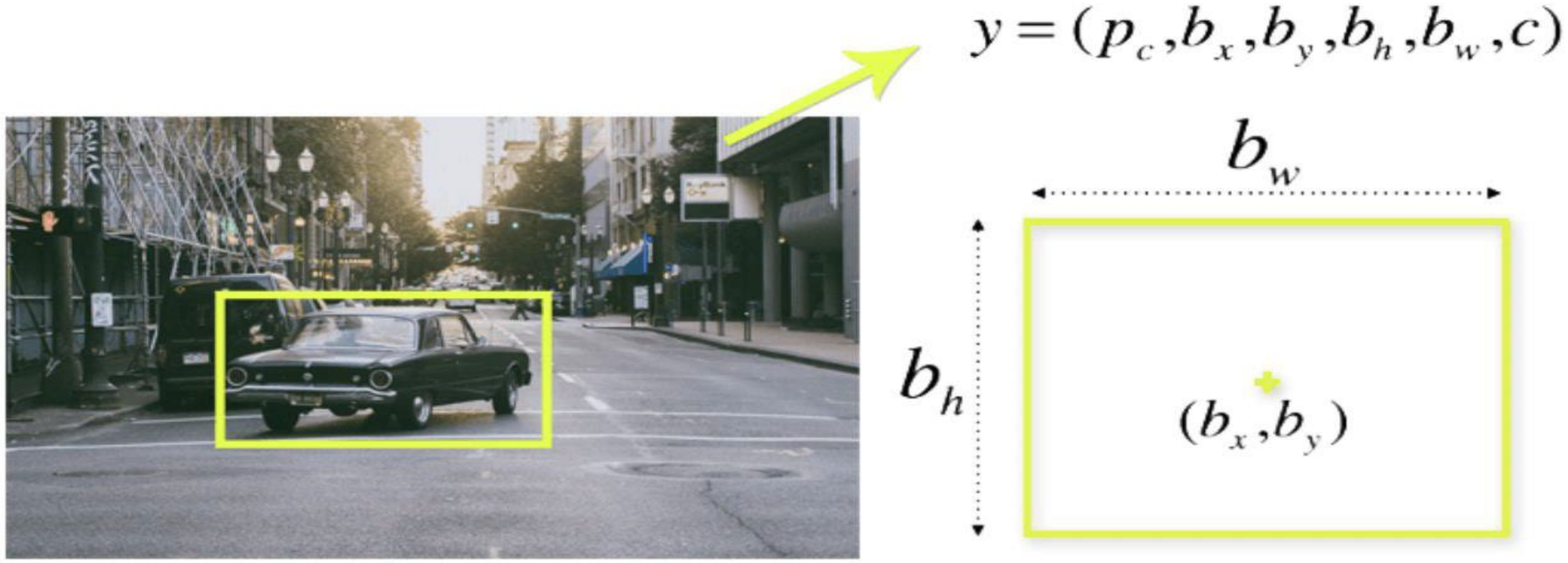
Every bounding box in the image consists of the following attributes:

a. Width (*bw*)

b. Height (*bh*)

c. Class (for example, person, car, traffic light, etc.) - This is represented by the letter *c*.

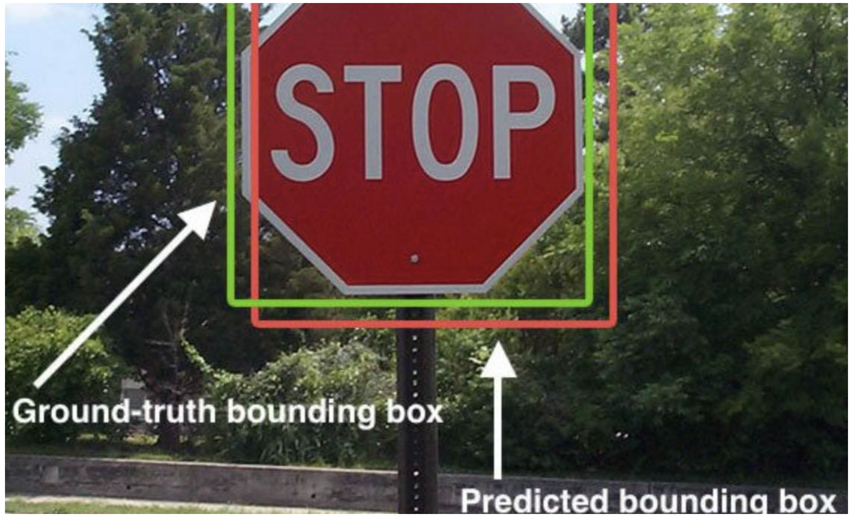
d. Bounding box center (*bx,by*)

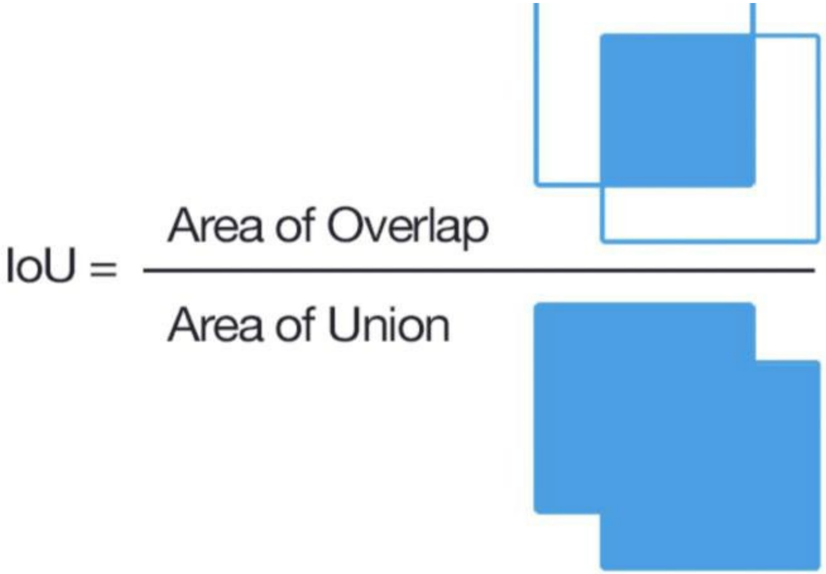


$$y = (p_c, b_x, b_y, b_h, b_w, c)$$

Intersection over Union (IoU)

- Intersection over Union is a popular metric to measure localization accuracy and calculate localization errors in object detection models.
- To calculate the IoU with the predictions and the ground truth, we first take the intersecting area between the bounding boxes for a particular prediction and the ground truth bounding boxes of the same area. Following this, we calculate the total area covered by the two bounding boxes—also known as the Union.

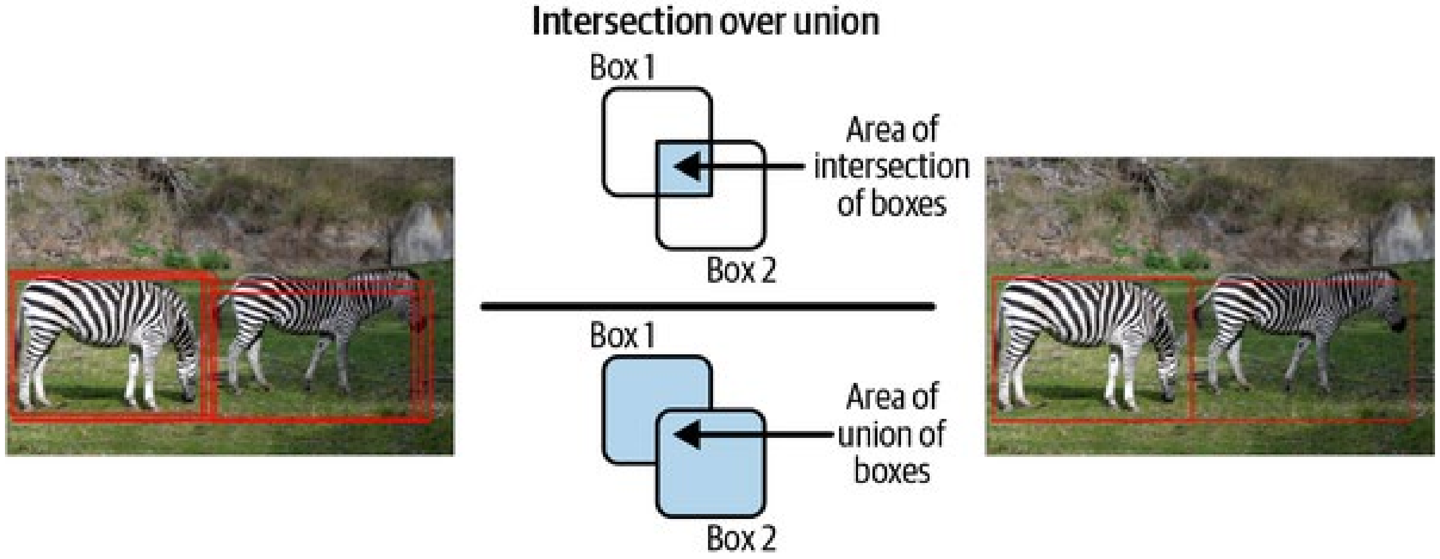




$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Intersection over Union (IoU)

- The intersection divided by the Union, gives us the ratio of the overlap to the total area, providing a good estimate of how close the bounding box is to the original prediction.



- Intersection over union ensures that the predicted bounding boxes are equal to the real boxes of the objects.
- This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of the objects (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly.

Object Detections Evaluation

- True positive: correct class prediction AND IoU > 50%.
- False positive: wrong class OR IoU < 50%.
- False negative: missed (not detected) object
- Only one detection can be matched to an object.

$$z_{ij} = \begin{cases} 1 & b_{ij} \text{ is a True Positive} \\ 0 & b_{ij} \text{ is a False Positive} \end{cases}$$

Average Precision

- **Average Precision** is calculated as the area under a precision vs recall curve for a set of predictions.
- **Recall** is calculated as the ratio of the total predictions made by the model under a class with a total of existing labels for the class.
- **Precision** refers to the ratio of true positives with respect to the total predictions made by the model.
- The *area under the precision vs recall curve* gives us the Average Precision per class for the model. The average of this value, taken over all classes, is termed as mean Average Precision (mAP).

Note: In object detection, precision and recall are not for class predictions, but for predictions of boundary boxes for measuring the decision performance.

- An IoU value > 0.5. is taken as a positive prediction, while an IoU value < 0.5 is a negative prediction.

Architecture of YOLO

Inspired by the GoogleNet architecture, The first YOLO architecture has a total of 24 convolutional layers with 2 fully connected layers at the end.

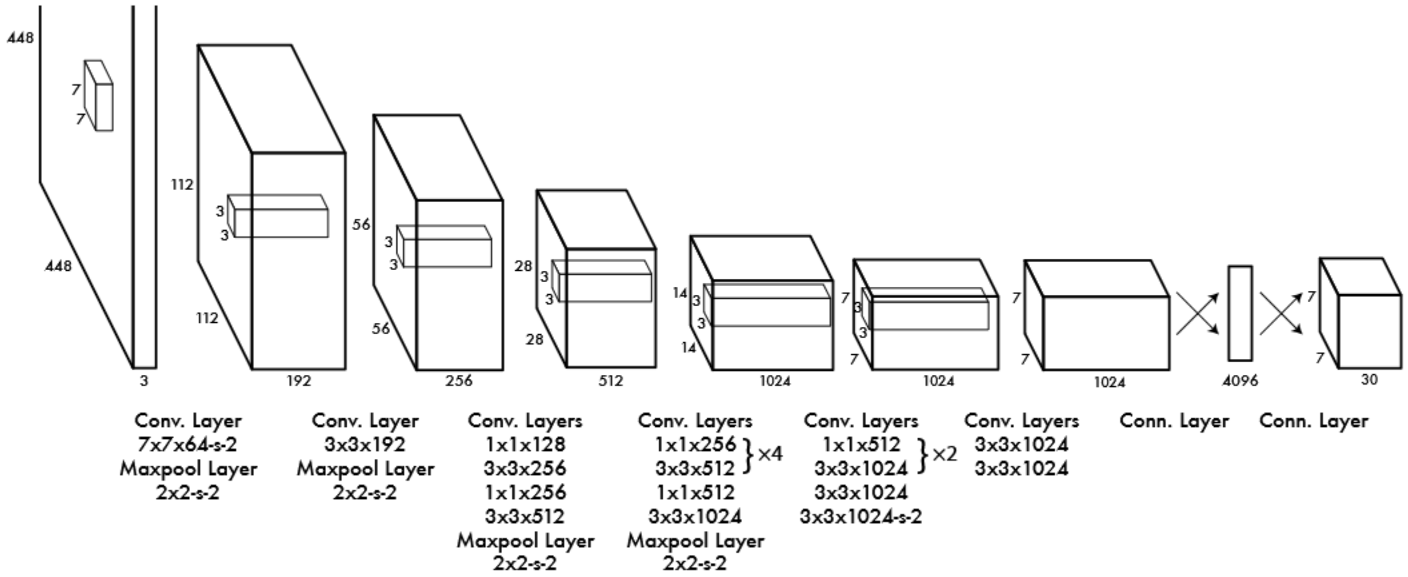


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

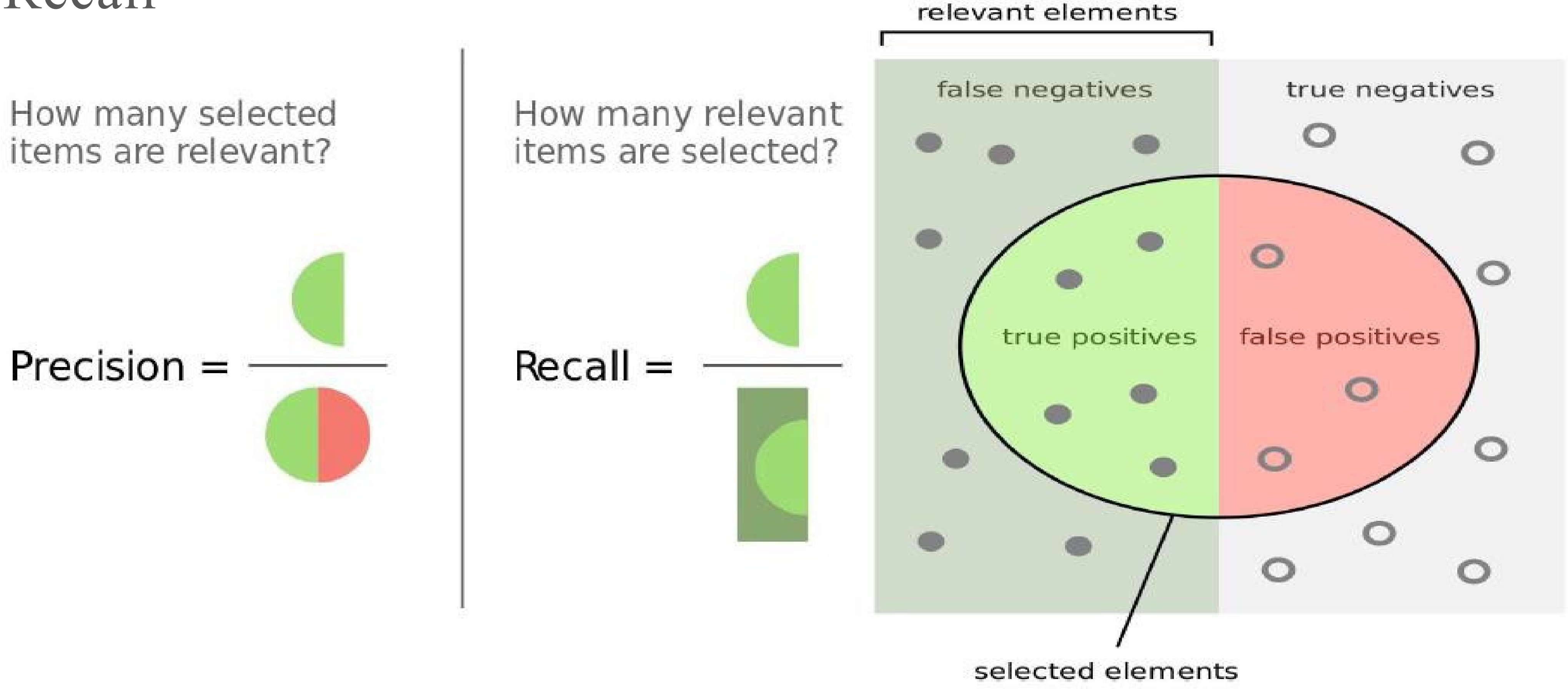
Benefit of YOLO

YOLO algorithm is important because of the following reasons:

- **Speed:** This algorithm improves the speed of detection because it can predict objects in real-time.
- **High accuracy:** YOLO is a predictive technique that provides accurate results with minimal background errors.
- **Learning capabilities:** The algorithm has excellent learning capabilities that enable it to learn the representations of objects and apply them in object detection.

mAP (mean Average Precision)

Mean Average Precision (mAP) across all classes, based on Average Precision (AP) per class, based on Precision and Recall

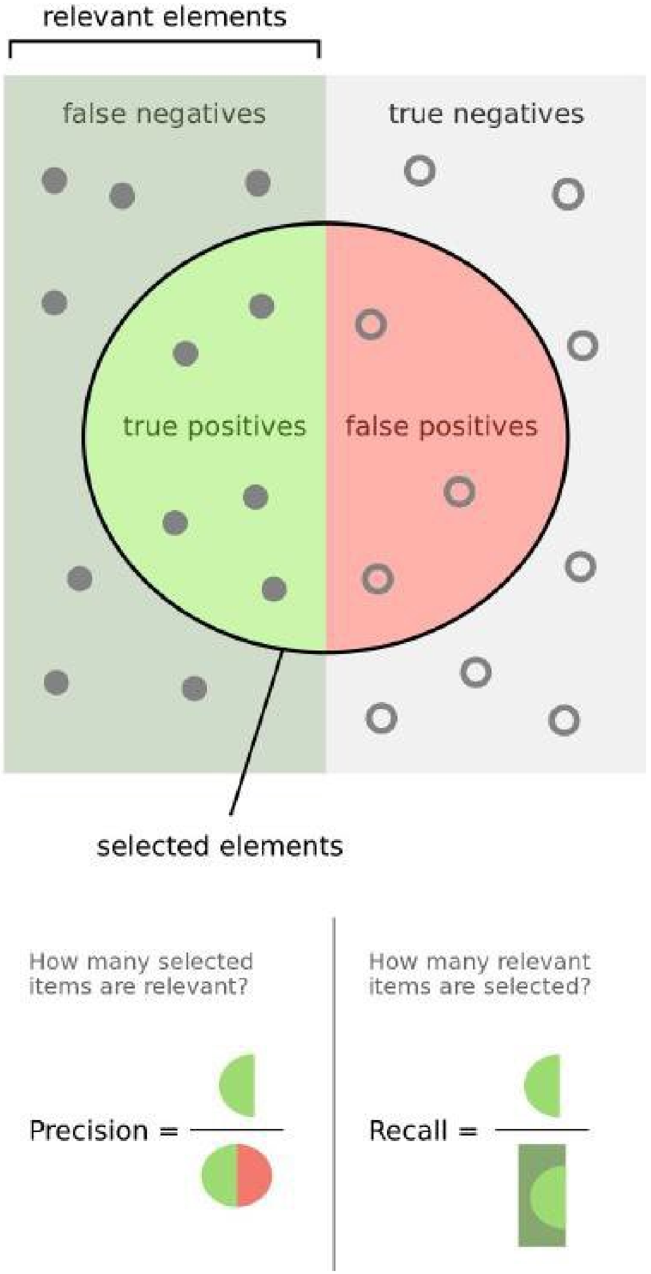


Precision and Recall

$$Recall(t) = \frac{\sum_{ij} 1[s_{ij} \geq t] z_{ij}}{N}$$

$$Precision(t) = \frac{\sum_{ij} 1[s_{ij} \geq t] z_{ij}}{\sum_{ij} 1[s_{ij} \geq t]}$$

$$z_{ij} = \begin{cases} 1 & b_{ij} \text{ is a True Positive} \\ 0 & b_{ij} \text{ is a False Positive} \end{cases}$$



Learn, Practice and Enjoy the AI journey