acm India Council

IIT Gandhinagar
ACM Summer School 2025
AI for Social Good

Tutorial on
Object Detection

Tutor : **Ayush Shrivastava**

Date : 2nd June 2025

SUSTAINABILITY LAB

Image Classification

Image Classification

Image Classification

Cat

Dog

Cat

SUSTAINABILITY LAB

Image Classification

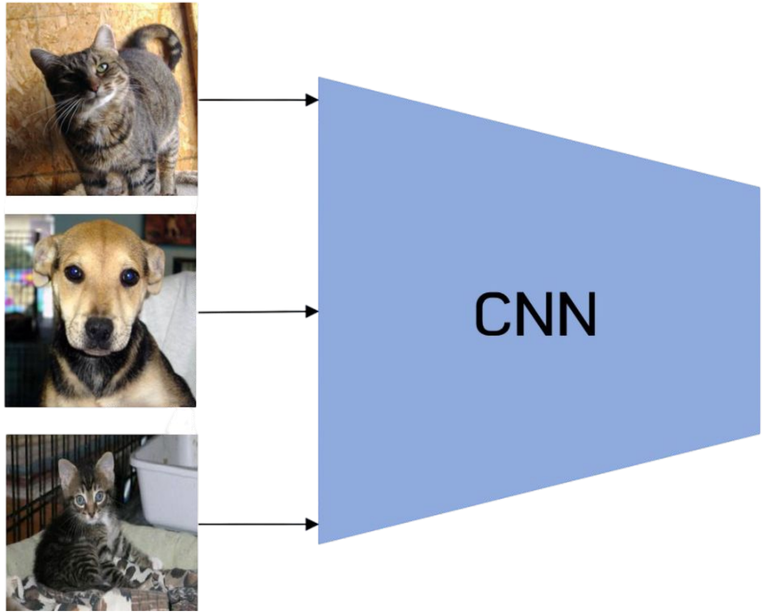
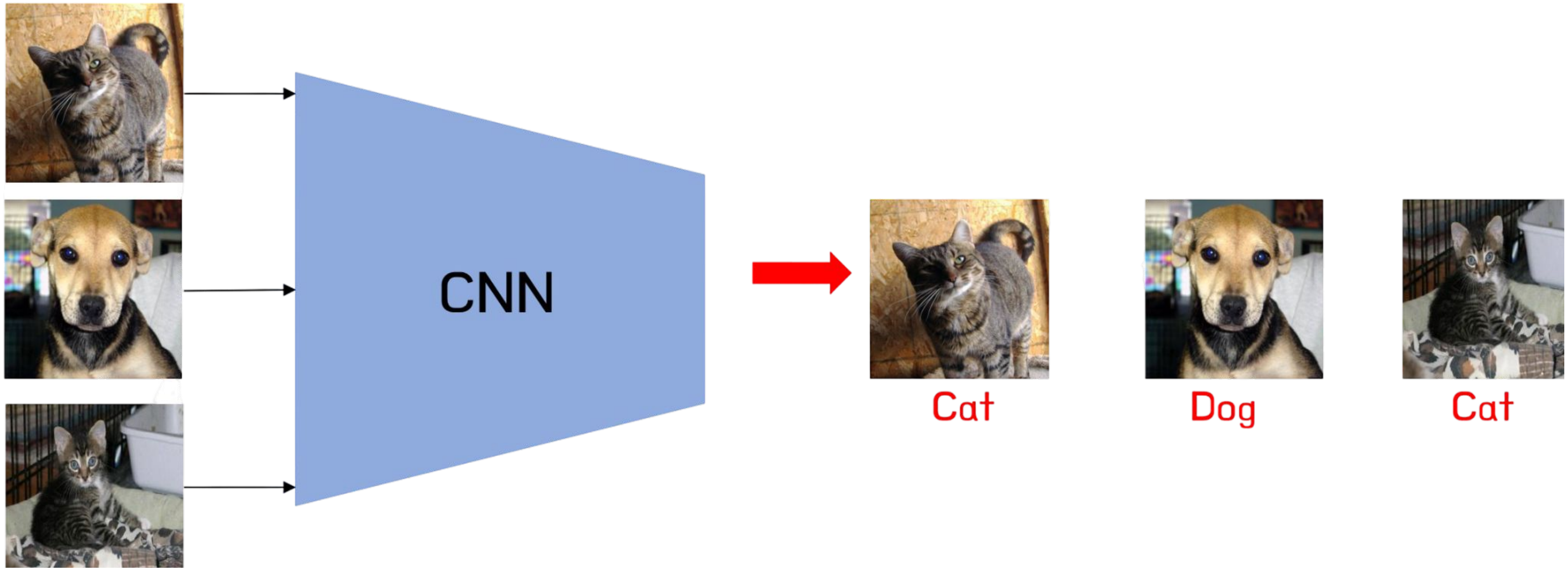


Image Classification



SUSTAINABILITY LAB

Exercise : Humans predict the labels

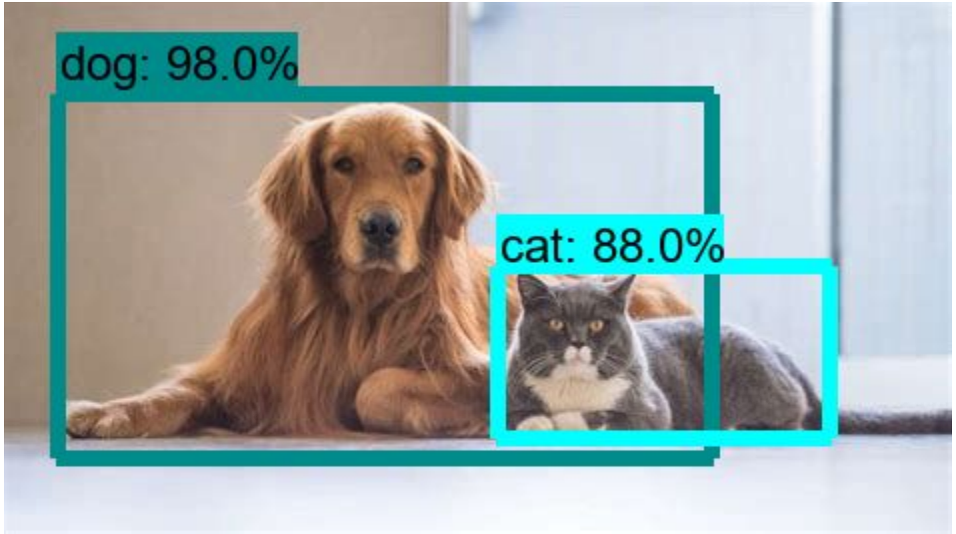
Cat or Dog?

Cat or Dog?

Cat or Dog?

Classification is not enough

While classification tells us what is in an image, it does not tell us where the object is located.



Use Cases of Object Detection

Annotations for Object Detection

For a simple classification task you require two things

1. Image
2. Label

<CAT>

Image

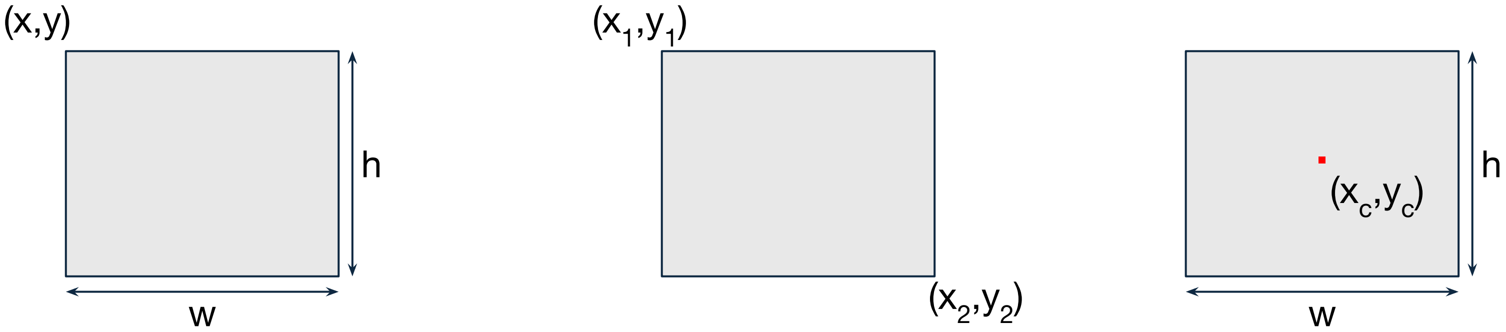
Label

Annotations for Object Detection

For object detection you would need a box and class per object.
What do you need to make a box?

Annotations for Object Detection

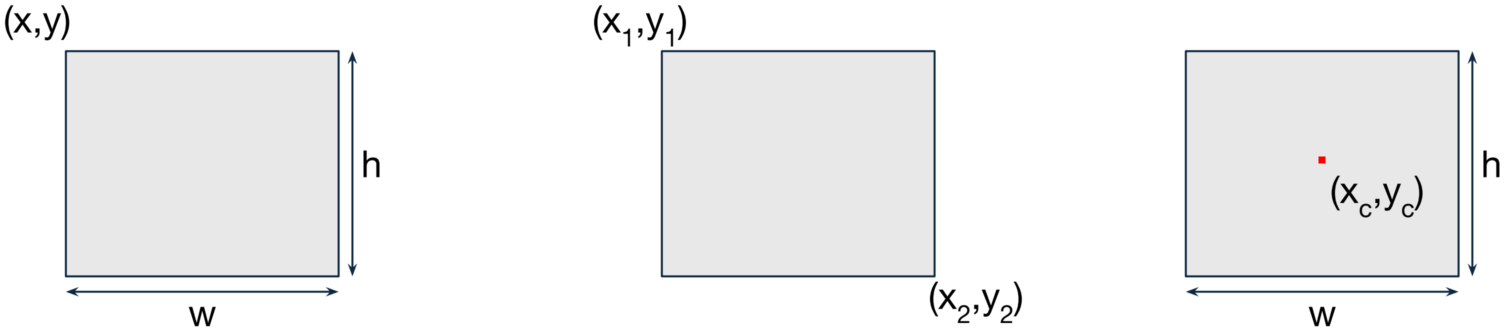
For object detection you would need a box and class per object.
What do you need to make a box?



SUSTAINABILITY LAB

Annotations for Object Detection

For object detection you would need a box and class per object.
What do you need to make a box?



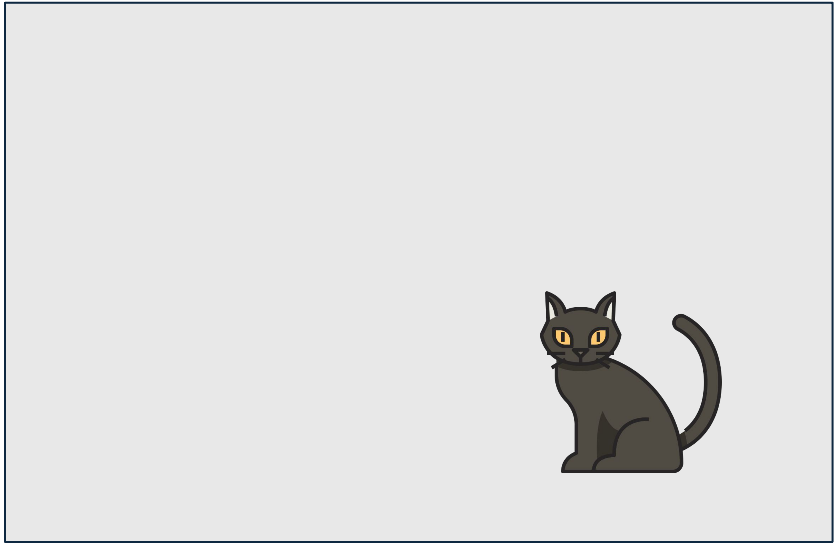
Now an annotation of an object would look like one of these

[<Class id> <x> <y> <w> <h>]

[<Class id> <x_1> <y_1> <x_2> <y_2>] or [<Class id> <x_{min}> <y_{min}> <x_{max}> <y_{max}>]

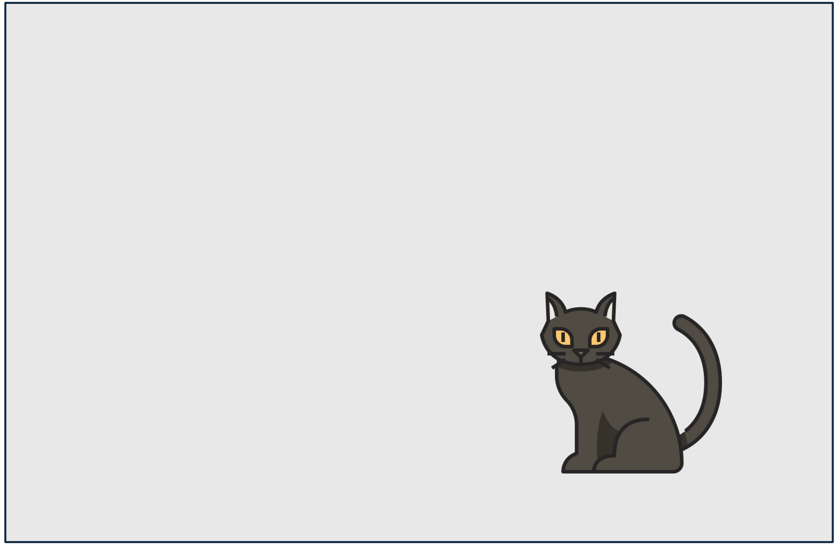
[<Class id> <x_C> <y_C> <w> <h>]

Annotations for Object Detection



We have this image named img.jpg

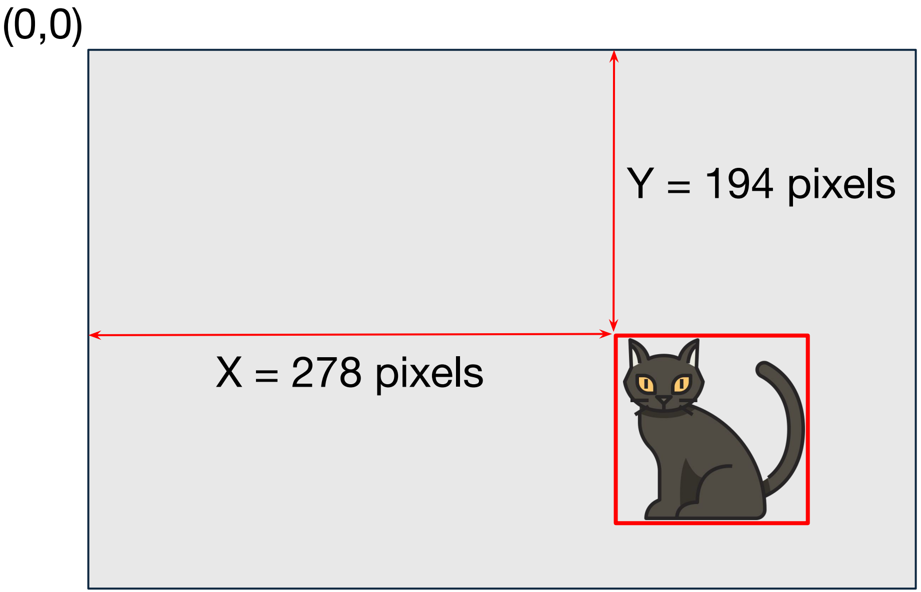
Annotations for Object Detection



Class = CAT (Encoded as 2)

[<2> <?> <?> <?> <?>]

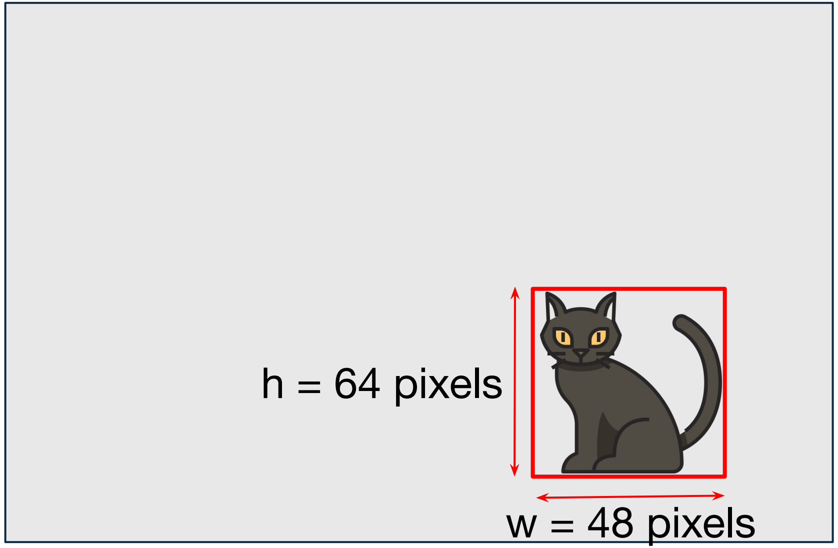
Annotations for Object Detection



Top corner of bounding box lies at (278,194)

[<2> <278> <194> <?> <?>]

Annotations for Object Detection



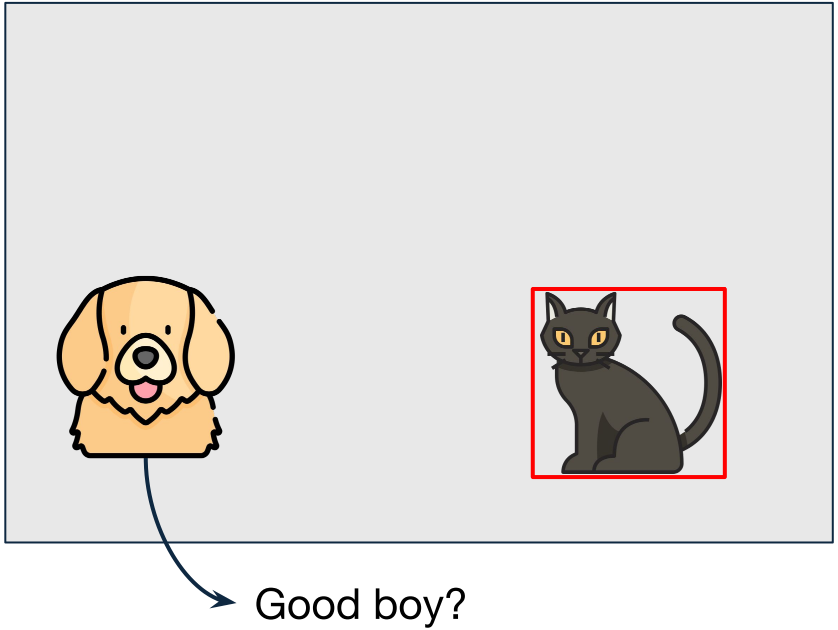
Top corner of bounding box lies at (278,194)

[<2> <278> <194> <48> <64>]

Annotations for Object Detection

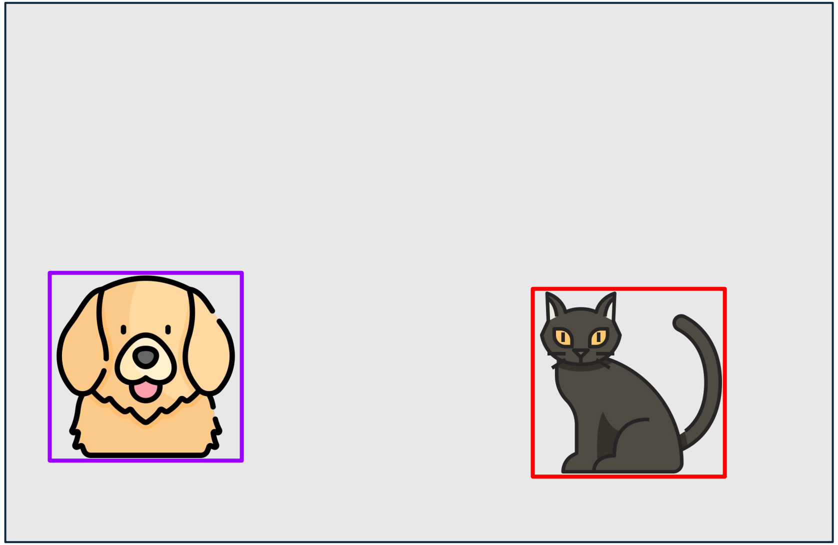
[<2> <278> <194> <48> <64>]

Annotations for Object Detection



[<2> <278> <194> <48> <64>]

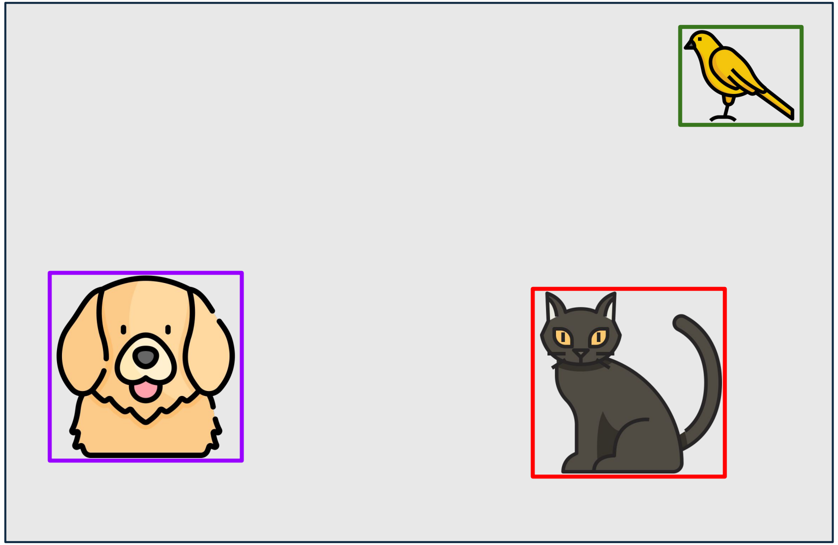
Annotations for Object Detection



[<2> <278> <194> <48> <64>]

[<1> <18> <190> <45> <69>]

Annotations for Object Detection

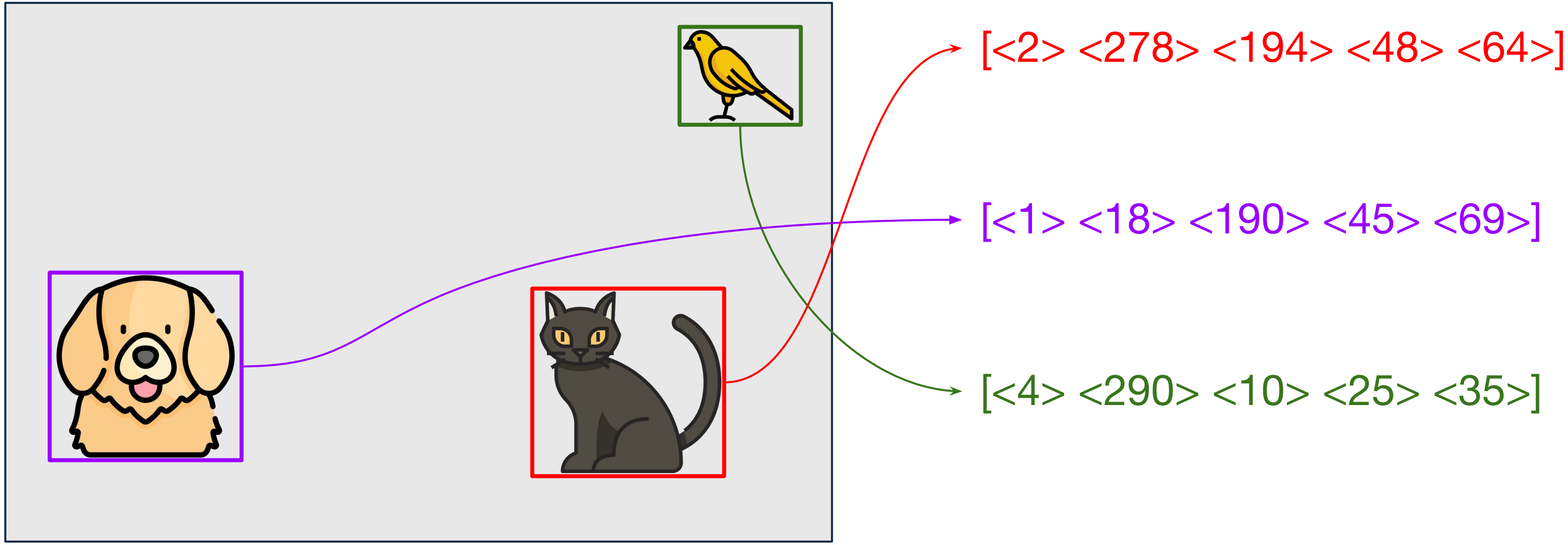


[<2> <278> <194> <48> <64>]

[<1> <18> <190> <45> <69>]

[<4> <290> <10> <25> <35>]

Annotations for Object Detection



Each object of interest gets a bounding box annotation.

SUSTAINABILITY LAB

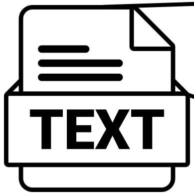
Annotations for Object Detection

img.jpg

img.txt

Annotations for Object Detection

img.jpg



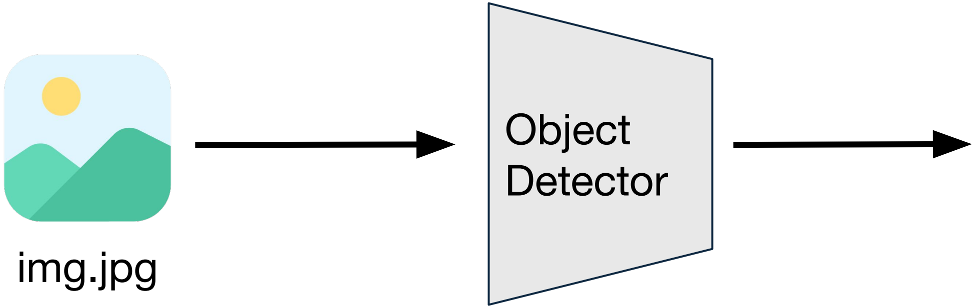
img.txt

2, 278, 194, 48, 64
1, 18, 190, 45, 69
4, 290, 10, 25, 35

Image

Annotation

Object Detector Output



x	y	w	h	conf	class

Classification Metrics

Let's say you train a classifier.

How do you estimate as classifier performs well?
What do you look at?

Classification Metrics

Let's say you train a classifier.

How do you estimate as classifier performs well?
What do you look at?

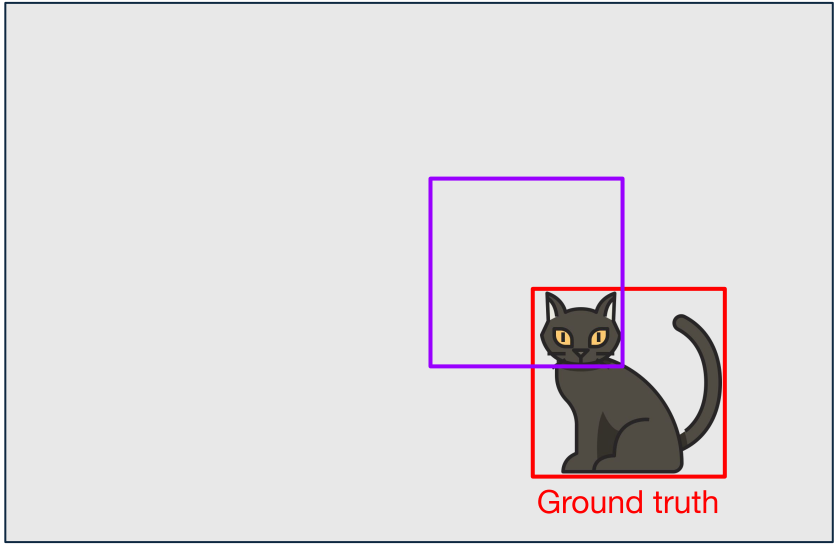
- Accuracy
- Precision
- Recall
- F1 Score

Metrics for Bounding Box

I train two Object Detectors

Metrics for Bounding Box

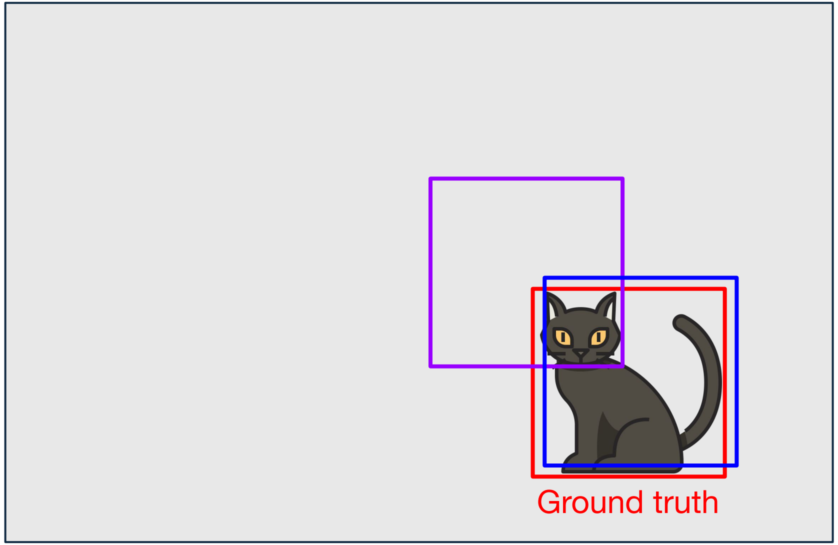
I train two Object Detectors.
My first object detection algorithm predicts the **purple** bounding box



Is **purple** box a good prediction?

Metrics for Bounding Box

I train two Object Detectors.
My first object detection algorithm predicts the **purple** bounding box
My second object detection algorithm predicts the **blue** bounding box

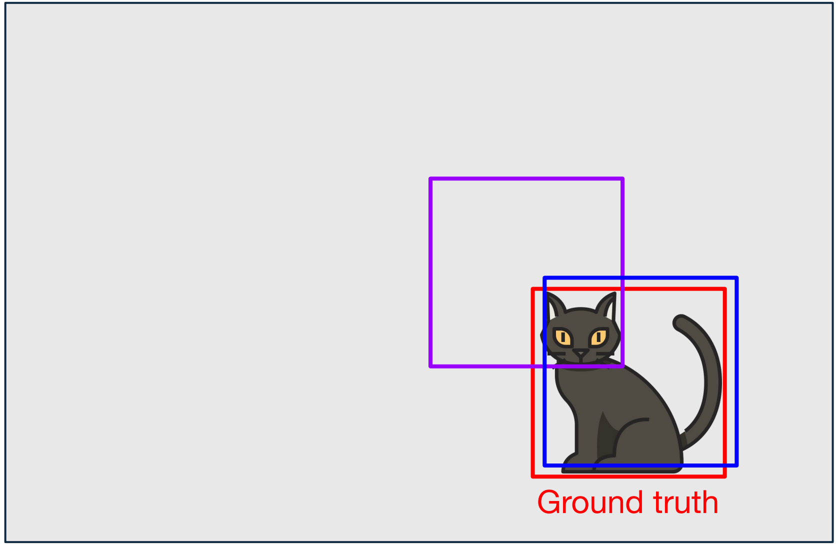


Is **purple** box a good prediction?

What about this **blue** one?

Metrics for Bounding Box

I train two Object Detectors.
My first object detection algorithm predicts the **purple** bounding box
My second object detection algorithm predicts the **blue** bounding box

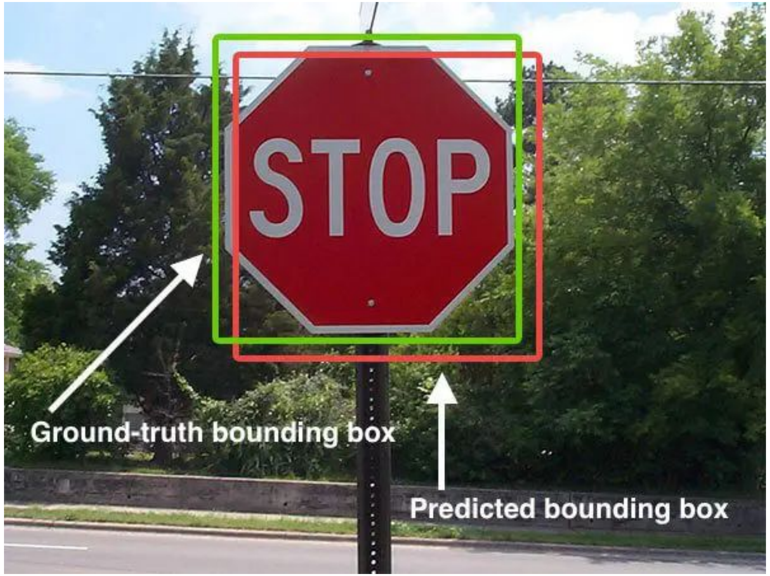


Is **purple** box a good prediction?

What about this **blue** one?

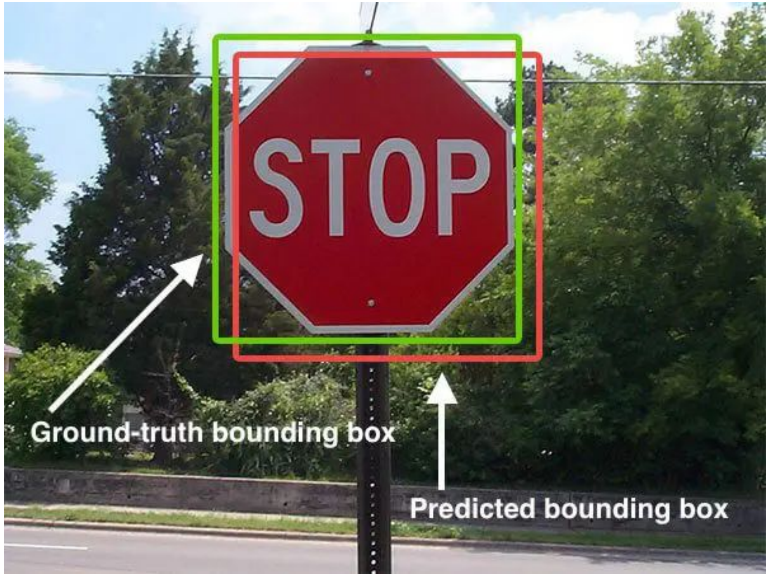
Why is blue better? How do you measure the fitness or accuracy of a bounding box?

IoU : Intersection over Union

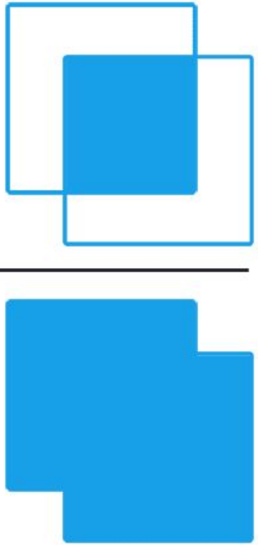


Source : https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

IoU : Intersection over Union



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



Source : https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

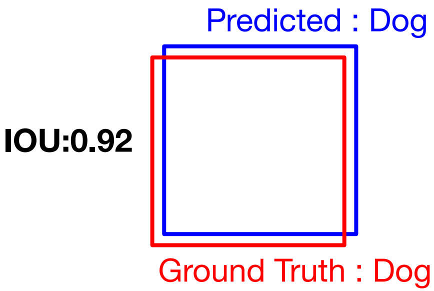
IoU : Intersection over Union

Generally an IOU of 0.5 is considered good enough but it could vary based on the use-case

Source : https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

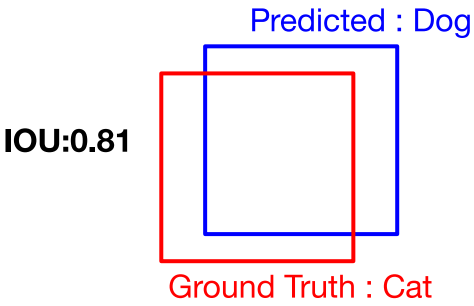
SUSTAINABILITY LAB

True Positives



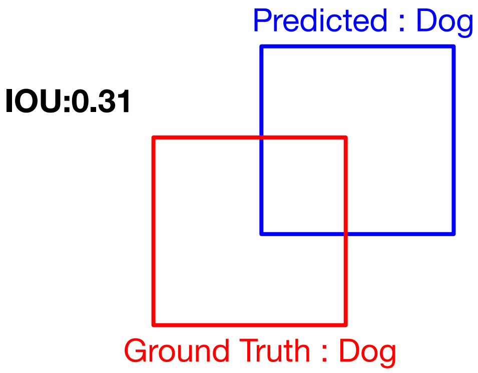
- IOU > Threshold
- Class label of ground truth matches with predicted label
- Model predicted the right object and also predicted it at the correct location

False Positives



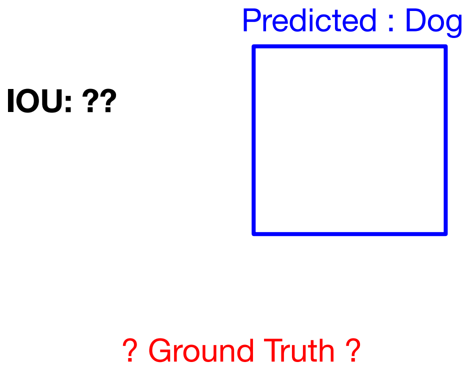
- IOU > Threshold
- Class label of ground truth do not match with predicted label
- Model predicts the wrong object even though the bounding box location is correct

False Positives



- IOU < Threshold
- Class label of ground truth matches with predicted label
- Model predicts the object correctly but the bounding box location is not good enough

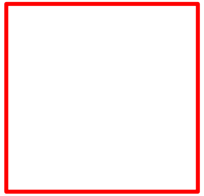
False Positives



- No matching ground truth exists for the predicted bounding box
- Model predicts object when none exists

False Negatives

? Predicted ?

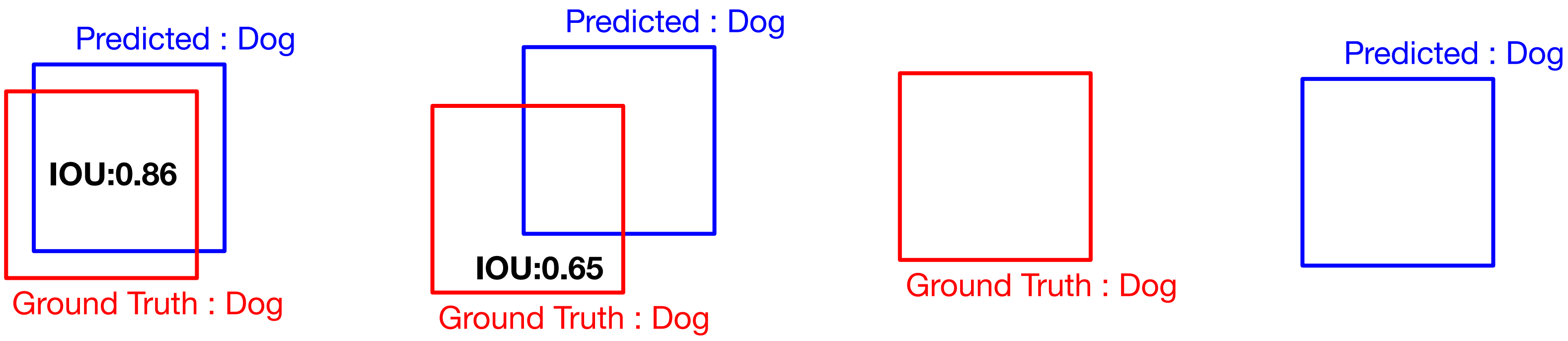


Ground truth : Human

- Ground truth bounding boxes without any predicted box
- Model unable to detect the existence of object

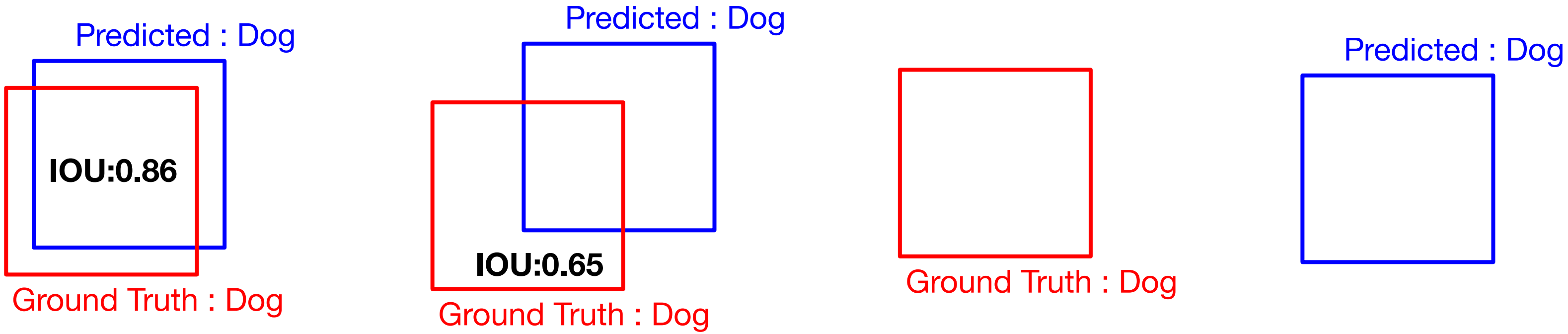
Object Detection Metrics

Precision and Recall are predicted per image given an **IOU threshold = 0.5**.



Object Detection Metrics

Precision and Recall are predicted per image given an **IOU threshold = 0.5**.



True Positive (TP)

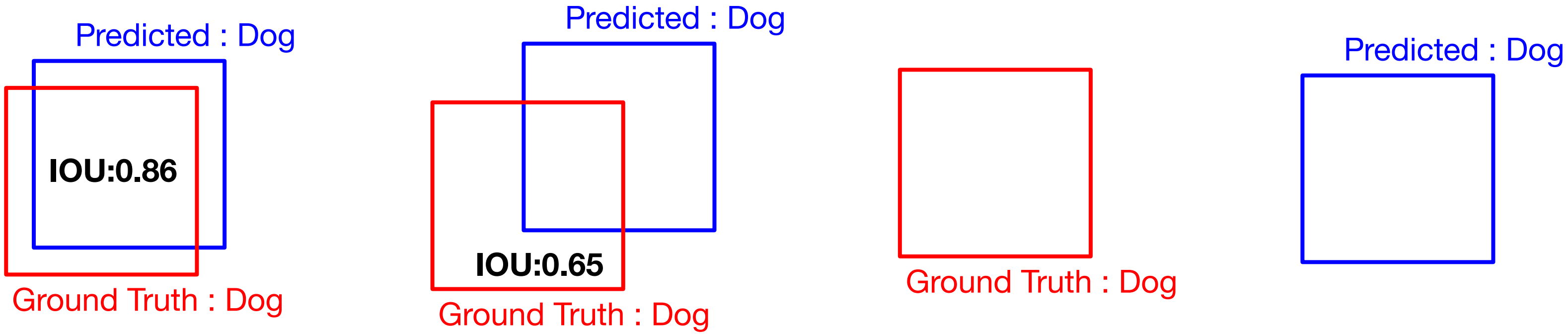
True Positive (TP)

False Negative (FN)

False Positive (FP)

Object Detection Metrics

Precision and Recall are predicted per image given an **IOU threshold = 0.5**.

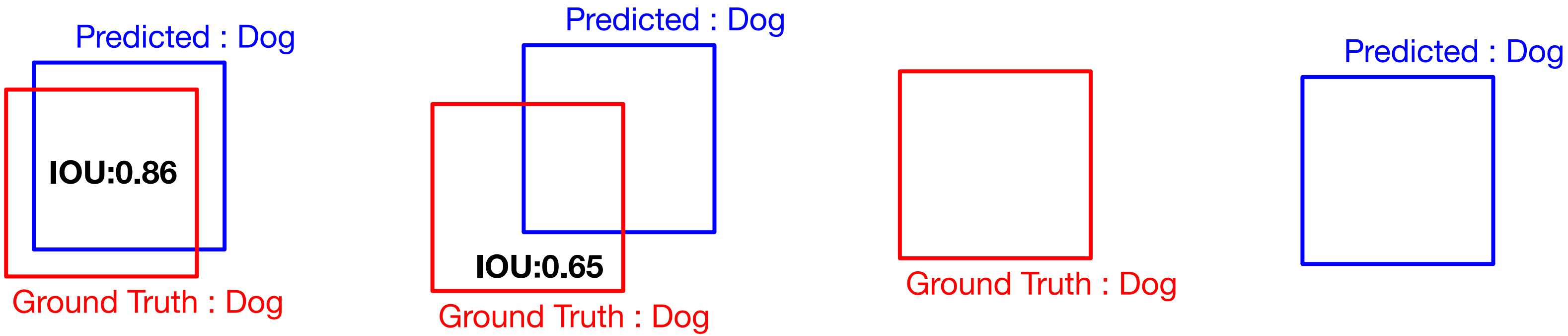


True Positive (TP) | **True Positive (TP)** | **False Negative (FN)** | **False Positive (FP)**

Precision $= TP/(TP + FP) = 2/3$ Recall $= TP/(TP + FN) = 2/3$

Object Detection Metrics

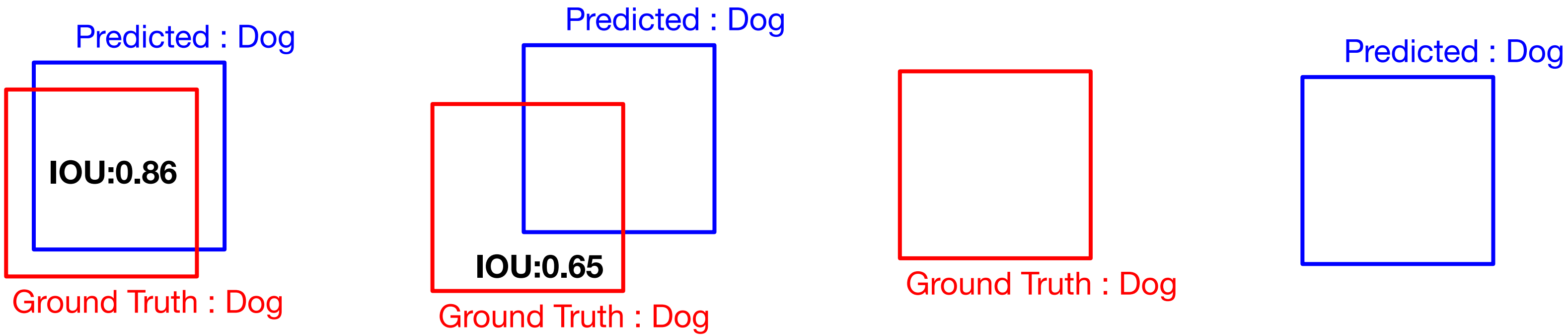
Precision and Recall are predicted per image given an **IOU threshold = 0.75**.



SUSTAINABILITY LAB

Object Detection Metrics

Precision and Recall are predicted per image given an **<u>IOU threshold = 0.75</u>**.



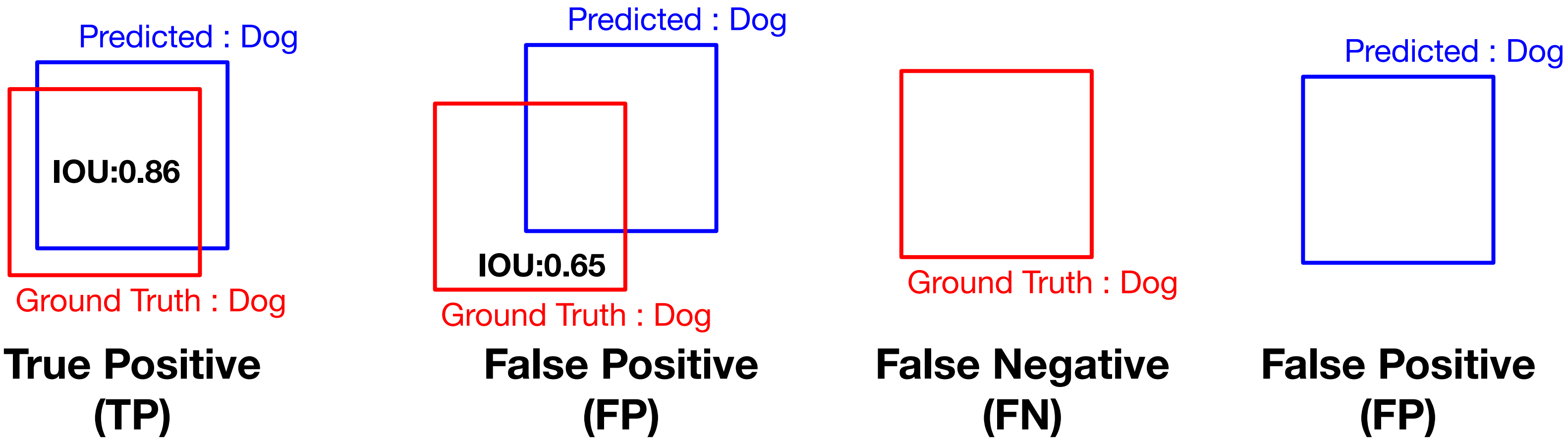
True Positive (TP) | **False Positive (FP)** | **False Negative (FN)** | **False Positive (FP)**

Precision = TP/(TP + FP) = ⅓ Recall = TP/(TP + FN) = ½

SUSTAINABILITY LAB

Object Detection Metrics

Precision and Recall are predicted per image given an **<u>IOU threshold = 0.75</u>**.



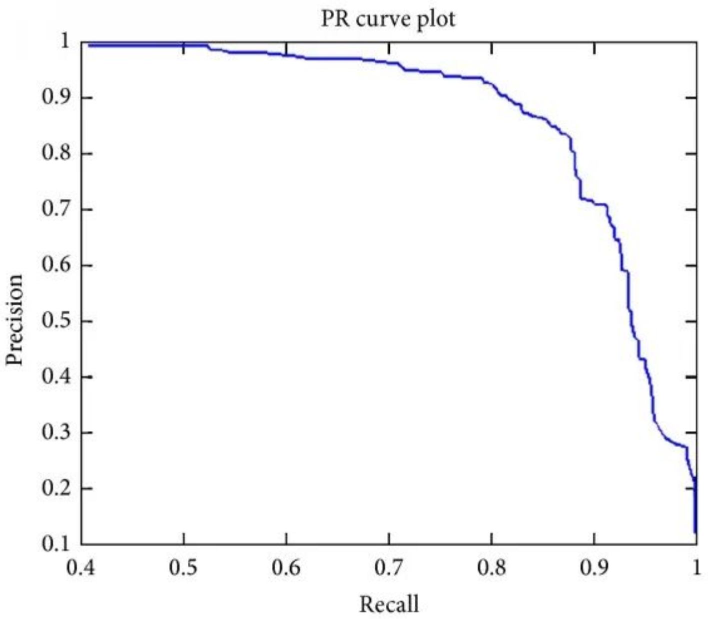
Precision = TP/(TP + FP) = ⅓ Recall = TP/(TP + FN) = ½

Average Precision (AP)

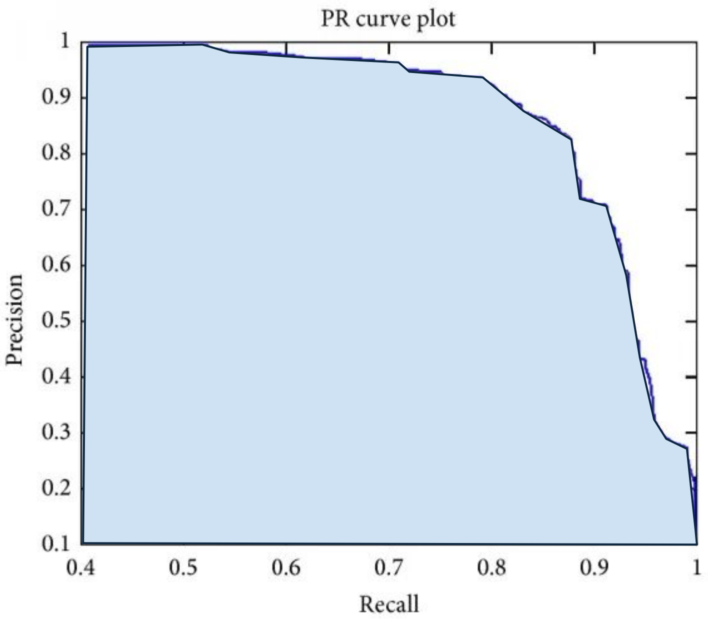
Class = Dog		
Image	**Precision**	**Recall**
img_1.png	1.0	0.2
img_2.png	1.0	0.2
img_3.png	0.67	0.4
….	…..	….
img_97.png	0.4	0.6
img_98.png	0.5	0.8
img_99.png	0.67	1.0

SUSTAINABILITY LAB

Average Precision (AP)

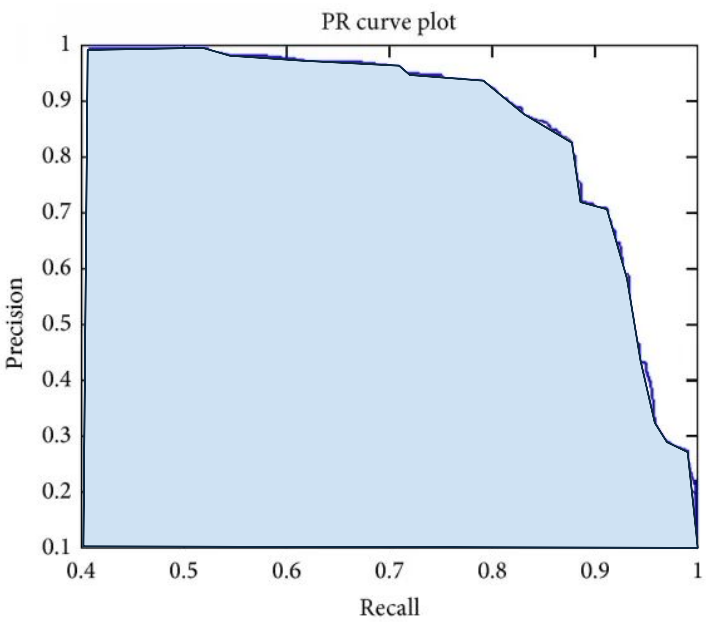


Average Precision (AP)



Area under the PR Curve = **Average Precision**

Mean Average Precision (mAP)



Area under the PR Curve = Average Precision

Class	**Avg Precision**
Dog	0.85
Cat	0.43
….	
Bird	0.76

Mean Average Precision (mAP) : 0.64

Mean Average Precision (mAP)

mAP50 :
Mean Average Precision calculated at IoU threshold of 0.5

mAP50-95 :
The average of the mean average precision calculated at varying IoU thresholds, ranging from 0.50 to 0.95. It gives a comprehensive view of the model's performance across different levels of detection difficulty.

Time for some Handson !

Open this Colab file https://tinyurl.com/IITGN-Tut1

Open https://roboflow.com/ and click on get started

Thank you!