Seeing with Algorithms: Deep Dive into Object Detection

From Classification to Localization and Detection Metrics

Nipun Batra and teaching staff

IIT Gandhinagar

August 30, 2025

Learning Outcomes

- Understand **classification**, **localization**, and **detection**
- Master **precision**, **recall**, **AP**, **mAP**, and **CA-mAP**
- Build strong intuition with toy examples and visual explanations
- Learn to evaluate object detectors thoroughly and effectively

"Detection is not just about finding objects, but finding them right."

Roadmap

Classification vs Localization vs Detection

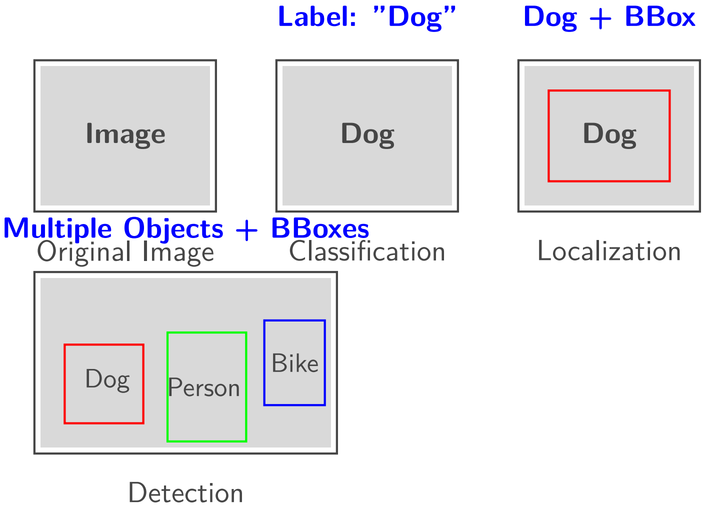
Task Definitions

Definition: Three Fundamental Computer Vision Tasks

- **Classification**: What is present in the image?
- **Localization**: Where is the object in the image?
- **Detection** = Classification + Localization (for multiple objects)

Each task builds upon the previous one, increasing in complexity and practical utility.

Visual Examples



Output Formats

Task	Output Format	Example
Classification	`label`	`"dog"`
Localization	`label, bbox`	`"dog", (30,30,100,1`

Detection	`[label, conf, bbox] × N`	`["dog", 0.95, (30,3`
		`["person", 0.87, (1`
		`["bike", 0.72, (80,`

Key Points:

Key Insight: Detection outputs include confidence scores, enabling ranking and threshold-based filtering!

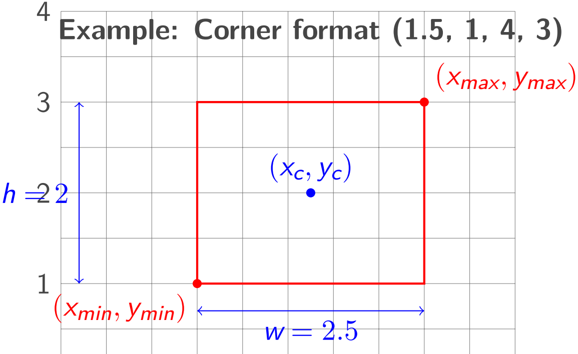
Bounding Boxes & Coordinates

What is a Bounding Box?

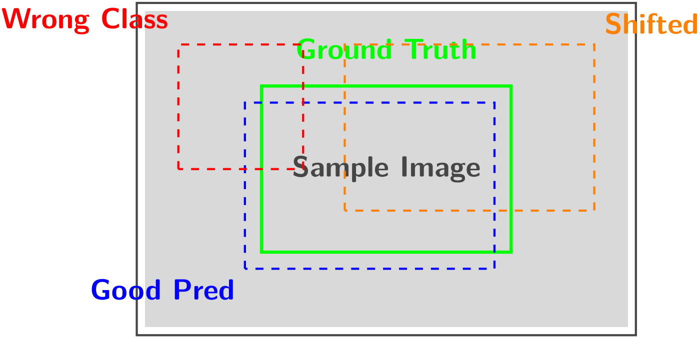
Definition: Bounding Box Formats

- **Corner format**: $(x_{min}, y_{min}, x_{max}, y_{max})$
- **Center format**: $(x_{center}, y_{center}, width, height)$

Example: Corner format (1.5, 1, 4, 3)



Ground Truth vs Predictions



Key Points:

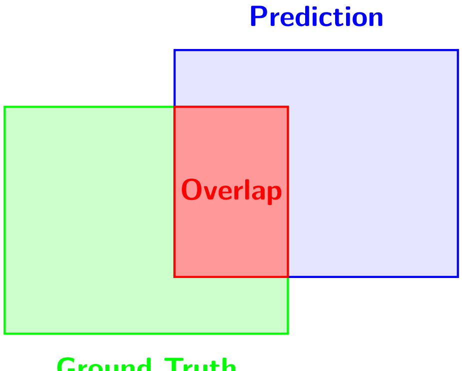
Matching Question: How do we decide which predictions correspond to which ground truth objects?

IoU (Intersection over Union)

IoU Definition

Definition: Intersection over Union (IoU)

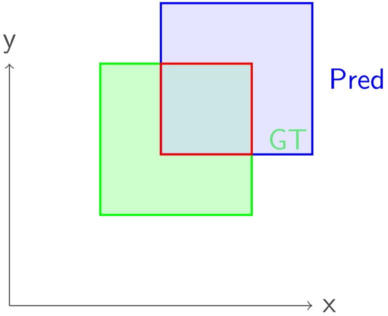
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{|A \cap B|}{|A \cup B|}$$



IoU Calculation Example

Example: Step-by-Step IoU Calculation

Ground Truth: $(30, 30, 100, 100)$
Prediction: $(50, 50, 120, 120)$



Step 1: Find intersection
$x_{min} = \max(30, 50) = 50$
$y_{min} = \max(30, 50) = 50$
$x_{max} = \min(100, 120) = 100$
$y_{max} = \min(100, 120) = 100$
Step 2: Calculate areas
Intersection: $50 \times 50 = 2500$
GT area: $70 \times 70 = 4900$
Pred area: $70 \times 70 = 4900$
Union:
$4900 + 4900 - 2500 = 7300$

Precision and Recall

Definitions

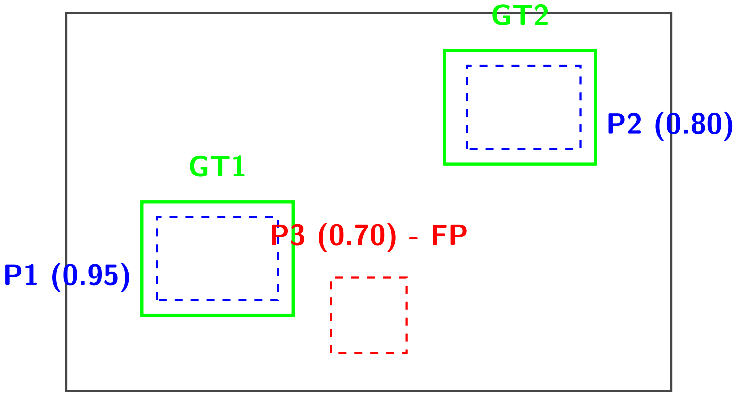
Definition: Core Metrics

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN}$$

- **Precision**: What fraction of detections are correct? (Quality)
- **Recall**: What fraction of ground truth objects are detected? (Coverage)
- **TP**: True Positive (correct detection, IoU $\geq$ threshold)
- **FP**: False Positive (incorrect detection, IoU $<$ threshold or extra detection)
- **FN**: False Negative (missed ground truth object)

Key Points:

Example: Counting TP, FP, FN



Scenario: 2 GT objects, 3 predictions

Analysis (IoU threshold = 0.5):

- P1 matches GT1: **TP**

Metrics:

- TP = 2, FP = 1, FN = 0

Ranking and PR Curve

Ranked Predictions Table

Example: Detection Results Sorted by Confidence

Given 5 predictions from our detector across the test set:

Confidence	Class	Box	TP/FP
0.95	Dog	(30,30,100,100)	TP
0.88	Bike	(150,120,200,180)	FP
0.80	Dog	(50,50,120,120)	TP
0.70	Person	(200,50,280,150)	TP
0.40	Cat	(100,100,150,150)	FP

Key Points: B

y varying the confidence threshold, we can trade off precision vs recall!

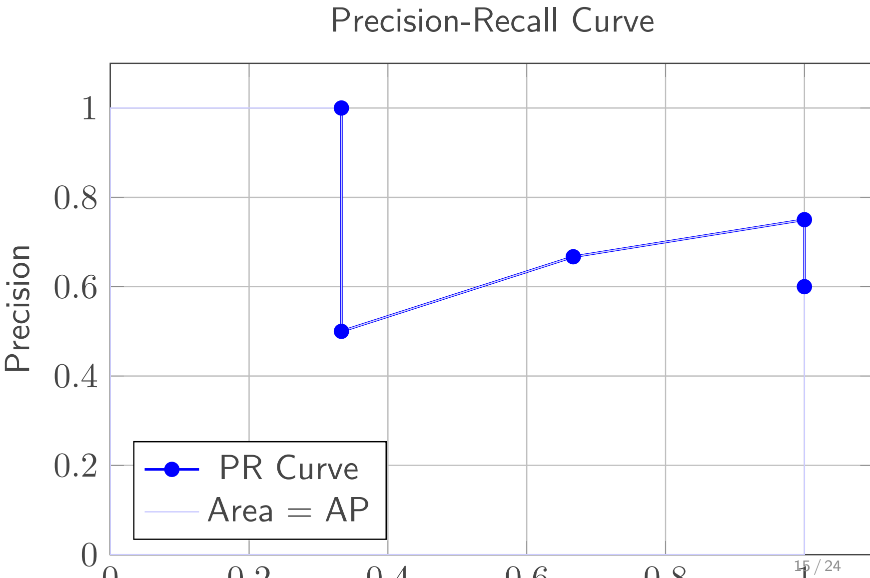
Precision-Recall Table

Threshold	Predictions	TP	FP	Precision	Recall
0.95	1	1	0	1.000	0.333
0.88	2	1	1	0.500	0.333
0.80	3	2	1	0.667	0.667
0.70	4	3	1	0.750	1.000
0.40	5	3	2	0.600	1.000

Assumptions: 3 ground truth objects total, IoU threshold $= 0.5$

- As threshold decreases $\rightarrow$ more predictions $\rightarrow$ recall increases
- But also more false positives $\rightarrow$ precision can decrease

Precision-Recall Curve



Average Precision (AP)

AP = Area under PR Curve

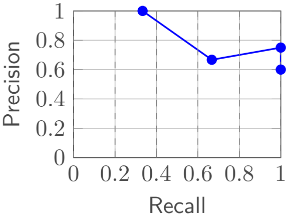
Definition: Average Precision Calculation

$$\text{AP} = \int_0^1 P(R)\, dR$$

In practice: Numerical integration or 11-point interpolation

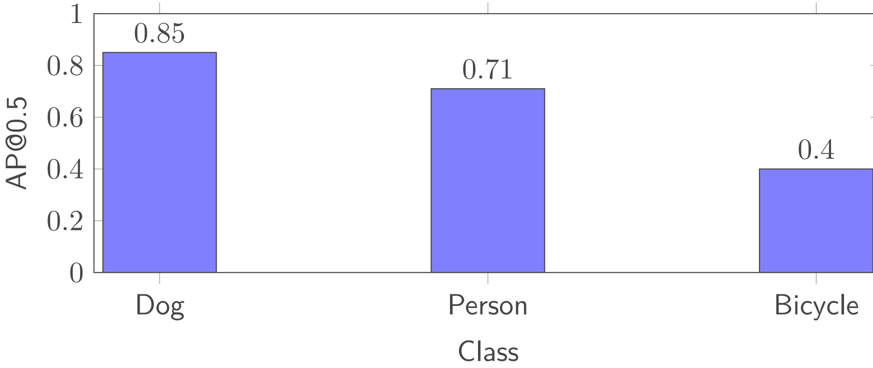
11-Point Interpolation:

- Sample at recall levels: 0, 0.1, 0.2, ..., 1.0
- For each recall r, find max precision for recall $\geq r$
- Average the 11 precision values



Class-wise AP Example

Class	AP@0.5	Visual
Dog	0.85	**Excellent**
Person	0.71	**Good**
Bicycle	0.40	**Poor**



Interpretation: Dog detection works well, but bicycle detection

Mean AP and Class-Agnostic mAP

Mean Average Precision (mAP)

Definition: mAP Calculation

$$\text{mAP} = \frac{1}{C}\sum_{c=1}^{C} \text{AP}_c$$

where C is the number of classes

Example: Our Example

$$\text{mAP} = \frac{\text{AP}_{dog} + \text{AP}_{person} + \text{AP}_{bicycle}}{3}$$

$$= \frac{0.85 + 0.71 + 0.40}{3} = \mathbf{0.653}$$

Class-Agnostic mAP (CA-mAP)

Definition: Class-Agnostic Evaluation

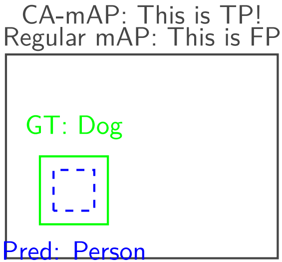
Ignore class labels when matching predictions to ground truth.
Match based on IoU overlap alone.

Standard mAP:

- Dog pred   Dog GT:
- Dog pred   Person GT:

Class-Agnostic mAP:

- Any pred   Any GT (if IoU > threshold):
- Useful for generic object detection



COCO-Style Evaluation

Strict Evaluation: COCO Metrics

Definition: COCO Evaluation Protocol

- **AP@50**: IoU threshold = 0.5 (lenient)
- **AP@75**: IoU threshold = 0.75 (strict)
- **AP@[.5:.95]**: Average over IoU thresholds 0.5, 0.55, 0.6, ..., 0.95

Metric	Value	Interpretation
mAP@50	0.71	Good localization (loose)
mAP@75	0.45	Moderate precise localization
mAP@[.5:.95]	0.42	Overall localization quality

Key Points:

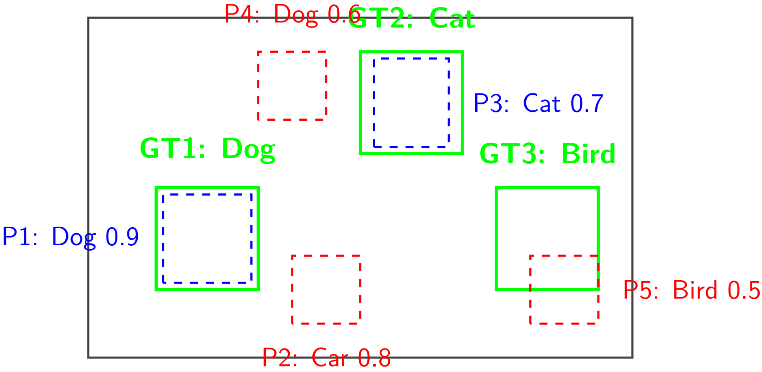
Key Insight: Higher IoU thresholds demand more precise lo

Interactive Quiz

Pop Quiz 1: Compute Precision & Recall

Answer this!

Given the detection scenario below, compute precision and recall (IoU threshold $= 0.5$):



Pop Quiz 1: Answer

Example: Solution

Analysis (with IoU > 0.5 matching):

- P1 (Dog 0.9) matches GT1 (Dog): **TP**
- P2 (Car 0.8) no GT match: **FP**
- P3 (Cat 0.7) matches GT2 (Cat): **TP**
- P4 (Dog 0.6) no GT match: **FP**
- P5 (Bird 0.5) poor overlap with GT3: **FP**
- GT3 (Bird) unmatched: **FN**

Final counts: TP $= 2$, FP $= 3$, FN $= 1$

$$\text{Precision} = \frac{2}{2+3} = 0.40 \quad \text{Recall} = \frac{2}{2+1} = 0.67$$

Summary & Takeaways

Summary Table

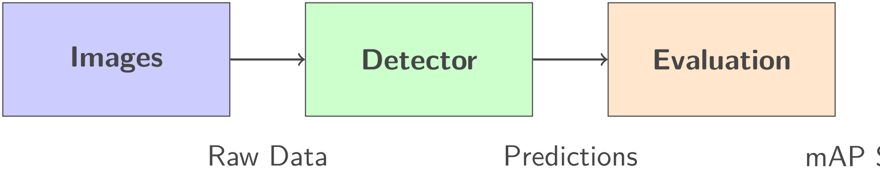
Concept	**Meaning**	**Key Insight**
IoU	Box overlap quality	Matching criterion (usually $>$ 0.
Precision	Detection quality	$\frac{TP}{TP+FP}$ (fewer false alarms)
Recall	Detection coverage	$\frac{TP}{TP+FN}$ (fewer missed objects)
AP	Area under PR curve	Single-class performance metric
mAP	Average AP over classes	Multi-class detector performanc
CA-mAP	Class-agnostic mAP	Localization-only evaluation
COCO	Multi-IoU evaluation	AP@[.5:.95] for precise localizati

Key Points:

Golden Rule **"Detection is not just about finding objects, but finding them right."**

What We've Learned

- **Task hierarchy**: Classification → Localization → Detection
- **Evaluation pipeline**: IoU matching → TP/FP counting → PR curves → AP/mAP
- **Trade-offs**: Precision vs Recall, lenient vs strict IoU thresholds
- **Practical metrics**: COCO-style evaluation for real-world deployment



Next steps: Explore modern architectures (YOLO, R-CNN, Transformers) and their mAP performance!