Seeing with Algorithms: Introduction to Object Detection

From Pixels to Predictions, and Precision to Policy

Nipun Batra and teaching staff

IIT Gandhinagar

August 30, 2025

Table of Contents

Motivation and Applications

Why Object Detection Matters

Object Detection helps machines see!

Self-Driving Cars

Self-Driving Cars

Medical Imaging

Medical Imaging

Smart Retail

Smart Checkout

Satellite Analysis

Satellite Analysis

What is Object Detection?

What is Image Classification?

Image Classification Goal

Identify what object is in the image

Image Classification Output

Single class label

Image Classification Question

"What is this?"

What is Object Detection?

Object Detection Goal

Find all objects and their locations

Object Detection Output

Class labels + bounding boxes

Object Detection Question

"What and where?"

Detection Components

What does detection give us?

Component 1: Bounding Box

$$(x_{min}, y_{min}, x_{max}, y_{max})$$

Component 2: Class Label

Dog, Cat, Car, Person

Component 3: Confidence Score

0.0 to 1.0

Detection Example

Real detection output

Class: Dog

Class: Dog

Confidence: 0.87

Confidence: 87%

Bounding Box

(120, 80, 340, 220) pixels

Our 3-Class Detection Example

3-Class Detection Problem

Class 1: Dog

Dog

Class 2: Bicycle

Bicycle

Class 3: Person

Person

Detection Pipeline

Object Detection Pipeline

Pipeline Input

Single image with unknown objects

Pipeline Processing

Computer vision algorithms

Pipeline Output

List of detected objects + locations

Step 1: Feature Extraction

Feature Extraction

Input Image

$416 \times 416 \times 3$ pixels

Backbone Network

ResNet, EfficientNet, DarkNet

Feature Maps

Rich representations

Step 2: Detection Predictions

Detection Predictions

Detection Head

YOLO, R-CNN, DETR

Raw Predictions

Bounding boxes + class scores

Step 3: Post-Processing

Post-Processing

Raw Predictions

Thousands of boxes

NMS + Filtering

Remove duplicates

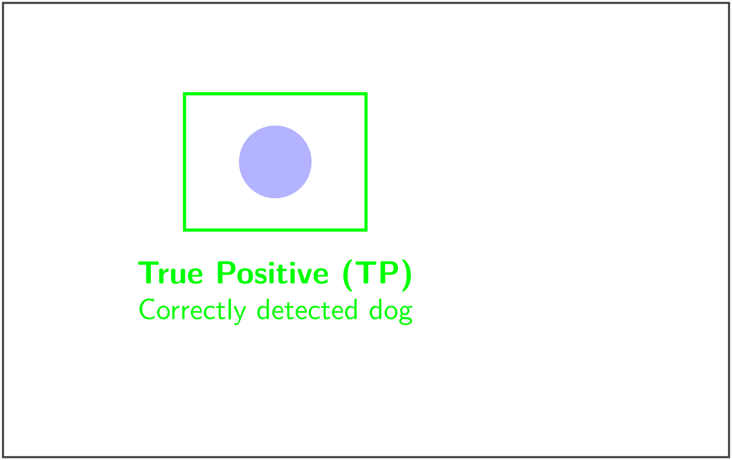
Final Detections

Clean results

Evaluation Metrics: The Foundation

Understanding Detection Outcomes

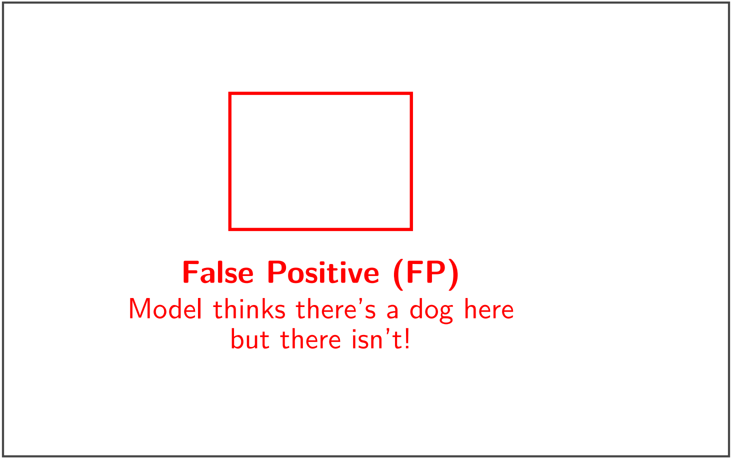
Sample Detection Results



Definition: True Positive (TP)

False Positive: When Models Hallucinate

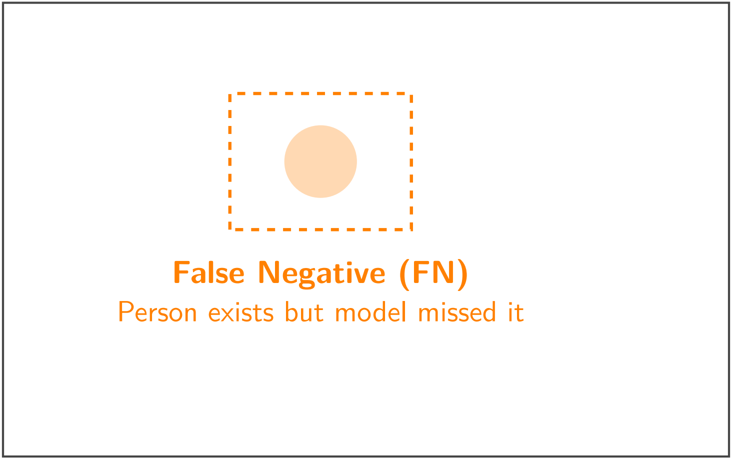
False Positive Example



Important: False Positive (FP)

False Negative: When Models Miss Objects

False Negative Example



Important: False Negative (FN)

What is Precision?

"Of my detections, how many were correct?"

Precision Formula

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Precision Meaning

Correct ÷ All detections

What is Recall?

"Of all real objects, how many did I find?"

Recall Formula

$$\frac{\text{TP}}{\text{TP} + \text{FN}}$$

Recall Meaning

Found ÷ All real objects

What is IoU?

IoU

IoU Stands For

Intersection over Union

What Does IoU Measure?

How much boxes overlap

IoU Range

0 to 1

Example Setup

Let's work through an example step by step

Ground Truth Box

Definition: Coordinates

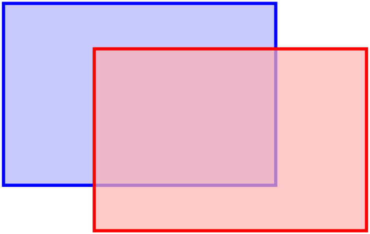
Ground Truth: (1,1) to (4,3)

Prediction Box

Definition: Coordinates

Prediction: (2,0.5) to (5,2.5)

Both Boxes Together



Definition: Question

Where do they overlap?

Finding Intersection - X Coordinates

Ground Truth X: 1 to 4

Prediction X: 2 to 5

Example: Step 1

Overlap X: from $\max(1,2) = 2$ to $\min(4,5) = 4$

Finding Intersection - Y Coordinates

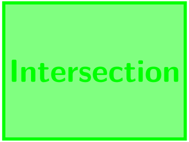
Ground Truth Y: 1 to 3

Prediction Y: 0.5 to 2.5

Example: Step 2

Overlap Y: from $\max(1,0.5) = 1$ to $\min(3,2.5) = 2.5$

Intersection Rectangle



Definition: Intersection Box

From (2,1) to (4,2.5)

Calculate Intersection Width

$$\text{Width} = 4 \text{ - } 2 = 2$$

Example: Step 3

Right edge - Left edge = Width

Calculate Intersection Height

$$\text{Height} = 2.5 - 1 = 1.5$$

Example: Step 4

Top edge - Bottom edge = Height

Calculate Intersection Area

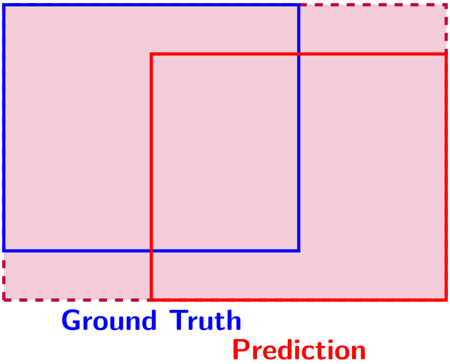
$$\text{Area} = 2 \times 1.5 = 3$$

Definition: Step 5

$\text{Width} \times \text{Height} = \text{Area}$

IoU: Calculating the Union

Union = Total Covered Area



Union = Area1 + Area2 - Intersection

Now Calculate Union

Union = Area1 + Area2 - Intersection

Definition: Why Subtract?

We subtract intersection to avoid counting it twice

Ground Truth Area

$$\text{Area1} = 3 \times 2 = 6$$

Example: Step 6

Ground Truth: Width 3, Height 2

Prediction Area

$$\text{Area2} = 3 \times 2 = 6$$

Example: Step 7

Prediction: Width 3, Height 2

Calculate Union

$$\text{Union} = 6 + 6 - 3 = 9$$

Definition: Step 8

Area1 + Area2 - Intersection = Union

IoU: The Formula

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$$

Definition: Simple Division

Take the overlapping area and divide by the total covered area

Final IoU Calculation

$$\text{IoU} = \frac{3}{9}$$

Example: Step 9

Intersection $\div$ Union

Do the Division

$$\frac{3}{9} = 0.33$$

Definition: Final Answer

IoU $= 0.33$ (33

IoU Threshold: 0.5

IoU 0.5

Definition: Standard Rule

If IoU is 0.5 or higher, we call it a True Positive

IoU Below Threshold

$$\text{IoU} < 0.5$$

Important: False Positive

If IoU is below 0.5, we call it a False Positive

Pop Quiz #1

Answer this!

Given this detection scenario:

- Ground Truth: 5 dogs in image
- Model detections: 8 boxes predicted as "dog"
- 4 detections have IoU   0.5 with ground truth

What are TP, FP, FN, Precision, and Recall?

A) TP=4, FP=4, FN=1, Precision=0.5, Recall=0.8

B) TP=5, FP=3, FN=0, Precision=0.63, Recall=1.0

C) TP=4, FP=1, FN=4, Precision=0.8, Recall=0.5

D) TP=8, FP=0, FN=0, Precision=1.0, Recall=1.0

The Answer

A)

Definition: Correct Answer

TP=4, FP=4, FN=1, Precision=0.5, Recall=0.8

Step 1: Find TP

$$\mathrm{TP} = 4$$

Example: Explanation

4 detections have IoU 0.5 with ground truth

Step 2: Find FP

$$FP = 8 - 4 = 4$$

Example: Explanation

8 total detections - 4 correct = 4 false alarms

Step 3: Find FN

$$\text{FN} = 5 - 4 = 1$$

Example: Explanation

5 ground truth dogs - 4 detected = 1 missed

Step 4: Calculate Precision

$$\frac{4}{4+4} = \frac{4}{8} = 0.5$$

Definition: Precision Formula

TP $\div$ (TP + FP)

Step 5: Calculate Recall

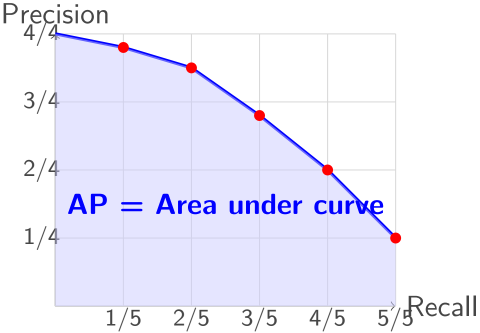
$$\frac{4}{4+1} = \frac{4}{5} = 0.8$$

Definition: Recall Formula

TP $\div$ (TP + FN)

Precision-Recall Curves and Average Precision

Precision-Recall Curve



Definition: PR Curve Interpretation

- **High precision at low recall**: Easy detections first
- **Curve drops**: As we include more detections, precision falls
- **Area Under Curve**: Average Precision (AP)

Computing AP: Step-by-Step Example

Example: Dog Detection Results (Sorted by Confidence)

Detection	Confidence	IoU	TP/FP	Precision	Recall
1	0.95	0.8	TP	$1/1 = 1.00$	$1/3 = 0.$
2	0.89	0.3	FP	$1/2 = 0.50$	$1/3 = 0.$
3	0.76	0.7	TP	$2/3 = 0.67$	$2/3 = 0.$
4	0.65	0.6	TP	$3/4 = 0.75$	$3/3 = 1.$
5	0.43	0.2	FP	$3/5 = 0.60$	$3/3 = 1.$

Key Points:

Ground Truth: 3 dogs in image

AP Calculation (using trapezoidal rule):

$$\text{AP} = \frac{1}{2}[(1.00 + 0.67) \times 0.34 + (0.67 + 0.75) \times 0.33 + (0.75 + 0.60)$$

Mean Average Precision (mAP)

From AP to mAP: Multi-Class Evaluation

Example: 3-Class Example: Computing Individual APs

Class	Ground Truth Count	Average Precision (AP)
Dog	12 objects	AP $= 0.73$
Bicycle	8 objects	AP $= 0.65$
Person	15 objects	AP $= 0.81$

Definition: Mean Average Precision (mAP)

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

For our example:

$$\text{mAP} = \frac{1}{3}(0.73 + 0.65 + 0.81) = \frac{2.19}{3} = 0.73$$

mAP Variants: @50, @75, @[.5:.95]

mAP@50

IoU threshold = 0.5

mAP@75

IoU threshold = 0.75

mAP@[.5:.95]

Average over IoU 0.5 to 0.95

Example: Example Results Comparison

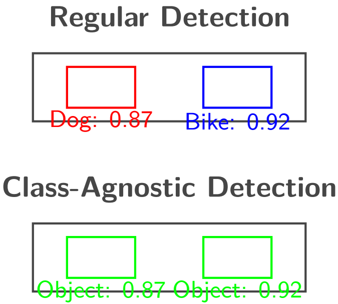
Metric	Value	Interpretation
mAP@50	0.73	Good localization (loose)
mAP@75	0.52	Moderate localization (strict)
mAP@[.5:.95]	0.61	COCO-style evaluation

Advanced Topics

Class-Agnostic mAP

Definition: What is Class-Agnostic Detection?

Instead of predicting specific classes, we just ask: **"Is there any object here?"**

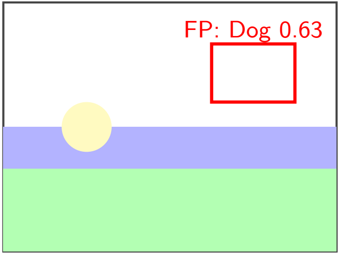


Example: Use Cases for Class-Agnostic mAP

Negative Set Evaluation

Important: Challenge: What about images with NO objects?

Negative Image (No Objects)



Results:

TP $= 0$ (no ground truth)
FP $= 1$ (false detection)
FN $= 0$ (no ground truth)

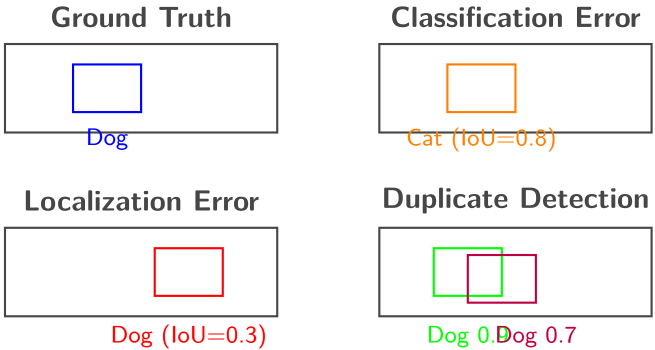
Precision $= \frac{0}{0+1} = 0$
Recall $=$ undefined

Key Points:

Negative Set Metrics:

Common Detection Errors

Localization vs Classification Errors



Definition: Error Types

- **Localization Error**: Right class, wrong location (IoU < threshold)
- **Classification Error**: Right location, wrong class
- **Duplicate Detection**: Multiple boxes for same object

Pop Quiz #2

Answer this!

You have a dataset with:

- 100 images total
- 50 images with dogs (300 dog instances total)
- 50 negative images (no objects)

Your model detects:

- 250 dogs correctly (IoU 0.5)
- 30 false positive dogs in positive images
- 20 false positive dogs in negative images

What is the Precision and Recall for the Dog class?

A) Precision=0.83, Recall=0.83

B) Precision=0.89, Recall=0.75

Pop Quiz #2 - Answer

Answer: A) Precision=0.83, Recall=0.83

Example: Step-by-Step Calculation

Given:

- TP = 250 (correctly detected dogs)
- FP = 30 + 20 = 50 (false positives in positive + negative images)
- FN = 300 - 250 = 50 (ground truth dogs - detected dogs)

Calculations:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{250}{250+50} = \frac{250}{300} = 0.83 \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{250}{250+50} = \frac{250}{300} = 0.83 \tag{4}$$

Summary and Practical Applications

What is AP?

AP

Average Precision

Definition: One Class

AP measures performance for one single class

mAP

Mean Average Precision

Definition: All Classes

mAP is the average of AP across all classes

mAP Example Setup

Let's calculate mAP for 3 classes

AP for Dogs

$$AP\_dogs = 0.8$$

Example: Given

Dog class achieved 80

AP for Cats

AP_cats = 0.6

Example: Given

Cat class achieved 60

AP for Cars

$$AP\_cars = 0.9$$

Example: Given

Car class achieved 90

Add Them Up

$$0.8 + 0.6 + 0.9 = 2.3$$

Example: Step 1

Sum all the AP values

Divide by Number of Classes

$$\frac{2.3}{3} = 0.77$$

Definition: Final Answer

mAP $= 0.77$ (77

mAP@50

Definition: Standard Evaluation

Uses IoU threshold of 0.5

Specialized mAP Variants

Class-Agnostic mAP

Ignores class labels

Just asks: "Is there an object?"

Useful for weakly supervised learning

Size-Specific mAP

Separate evaluation for

small, medium, large objects

COCO provides mAP_S, mAP_M, mAP_L

That's It!

Definition: Key Point

Object detection uses mAP to measure performance

Detection Fundamentals: Key Takeaways

Object Detection = Classification + Loca

mAP is the gold standard for model comparison

IoU thresholds matter - stricter = lower scores

Negative images crucial for real deployment

Context matters - choose metrics for your use case

Definition: Remember

Perfect metrics don't guarantee perfect real-world perfor-

Real-World Considerations

Important: Beyond the Metrics

Perfect mAP doesn't guarantee perfect real-world performance!

Example: Model Selection

- **Speed vs Accuracy**: YOLOv8 vs R-CNN
- **Memory constraints**: Mobile deployment
- **Class imbalance**: Rare vs common objects

Definition: Deployment Issues

- **Domain shift**: Training vs real data
- **Edge cases**: Unusual lighting, angles
- **Ethical considerations**: Bias, privacy

Demo Time & Further Reading

Example: Try These Demos!

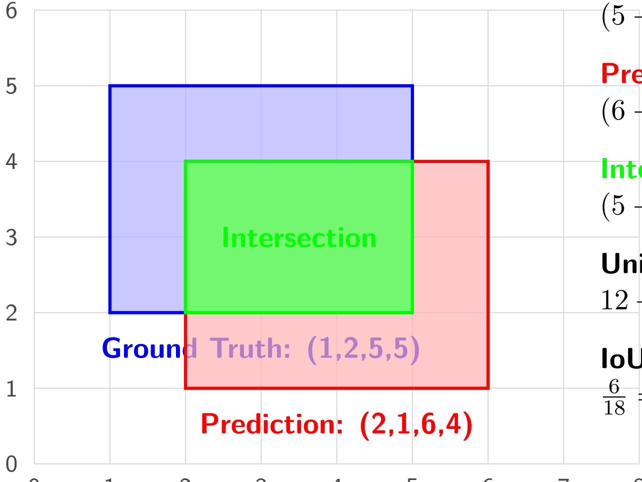
- **YOLOv8 Demo**: https://docs.ultralytics.com/
- **Roboflow Playground**: Interactive object detection
- **HuggingFace Spaces**: Search "Object Detection"

Definition: Essential Papers

- **YOLO series**: You Only Look Once (Redmon et al.)
- **Faster R-CNN**: Two-stage detection (Ren et al.)
- **COCO Dataset**: Common Objects in Context (Lin et al.)

Detailed Worked Examples

Complete IoU Calculation Example



Step-by-Step Calcula

Ground Truth Area:
$(5-1)\times(5-2)=4$

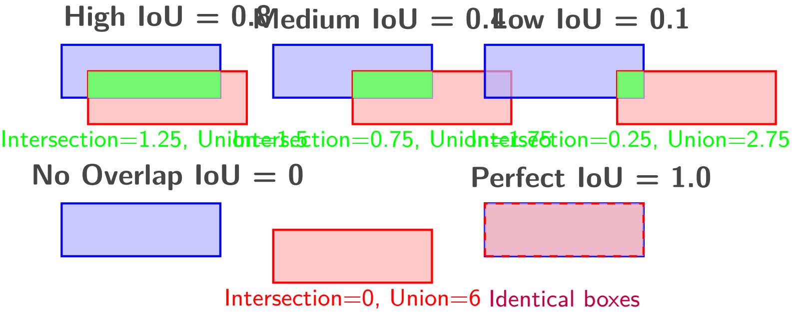
Prediction Area:
$(6-2)\times(4-1)=4$

Intersection Area:
$(5-2)\times(4-2)=3$

Union Area:
$12+12-6=18$

IoU:
$\frac{6}{18}=0.333$

Multiple IoU Examples with Different Overlaps



Key Points:

Key Insights:

- **IoU 0.5**: Generally considered good localization
- **IoU 0.7**: High-quality detection
- **IoU = 1.0**: Perfect alignment (rare in practice)

Comprehensive Precision-Recall Example

Example: Scenario: Dog Detection in 5 Images

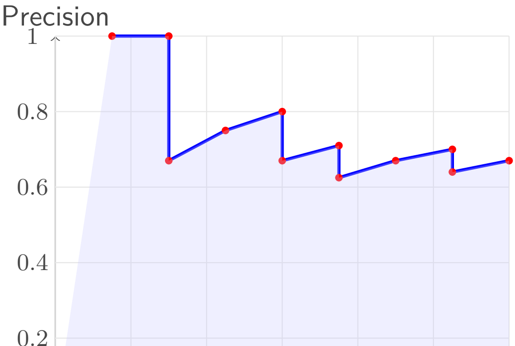
Ground Truth: 8 dogs total across all images
Model Predictions: 12 detections sorted by confidence

Det	Conf	IoU	TP/FP	Cum TP	Cum FP	Precision	Re
1	0.95	0.82	TP	1	0	1.00	0
2	0.89	0.71	TP	2	0	1.00	0
3	0.84	0.33	FP	2	1	0.67	0
4	0.79	0.65	TP	3	1	0.75	0.
5	0.73	0.58	TP	4	1	0.80	0
6	0.68	0.29	FP	4	2	0.67	0
7	0.62	0.74	TP	5	2	0.71	0.
8	0.57	0.19	FP	5	3	0.625	0.
9	0.51	0.68	TP	6	3	0.67	0
10	0.46	0.72	TP	7	3	0.70	0.
11	0.41	0.15	FP	7	4	0.64	0.
12	0.35	0.61	TP	8	4	0.67	1

Definition: Cumulative Calculations

Cumulative TP: Running count of true positives (IoU 0.5)

Plotting the Precision-Recall Curve



Definition: AP Calculation

Using trapezoidal rule:

$$AP = \sum_i \frac{1}{2}(P_i + P_{i+1}) \times \quad (5)$$

Result: AP 0.74

Key Points:

Observations:

- High precision

Multi-Class mAP Calculation Detailed Example

Example: 3-Class Detection Results

Dataset: 100 images with Dogs, Cats, and Cars

Class 1: Dogs
Ground Truth: 45 objects
AP = 0.82

Class 2: Cats
Ground Truth: 38 objects
AP = 0.76

Class 3: Car
Ground Truth: 5
AP = 0.89

mAP Calculation:

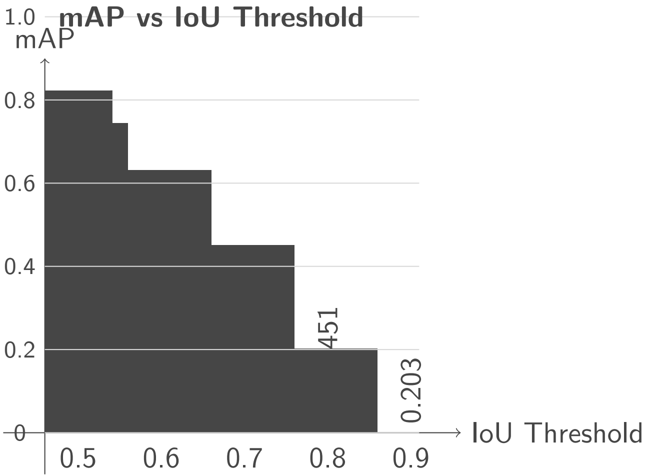
$$\text{mAP} = \frac{1}{3}(\text{AP}_{\text{Dogs}} + \text{AP}_{\text{Cats}} + \text{AP}_{\text{Cars}})$$

$$\text{mAP} = \frac{1}{3}(0.82 + 0.76 + 0.89) = \frac{2.47}{3} = 0.823$$

Class-wise Performance Analysis:

- **Cars (AP=0.89)**: Best performing class - likely larger, more distinct
- **Dogs (AP=0.82)**: Good performance - varied poses but distinct fea

mAP@Different IoU Thresholds: Complete Analysis



Definition: mAP@[.5:.95] Calculation

Important: Key Insights

Pop Quiz #3: Advanced mAP Calculation

Answer this!

You're evaluating a 2-class detector (Cat, Dog) on a dataset:

Cat Class Results:

- Ground truth: 20 cats
- Detections: 15 correct (IoU 0.5), 8 false positives
- AP@0.5 = 0.75

Dog Class Results:

- Ground truth: 30 dogs
- Detections: 25 correct (IoU 0.5), 5 false positives
- AP@0.5 = 0.83

What is the overall mAP@0.5, and which class has better precision?

Pop Quiz #3 - Answer

Answer: A) mAP = 0.79, Dog has better precision (0.83 > 0.65)

Example: Step-by-Step Solution

1. Calculate mAP:

$$\text{mAP} = \frac{\text{AP}_{\text{Cat}} + \text{AP}_{\text{Dog}}}{2} = \frac{0.75 + 0.83}{2} = 0.79$$

2. Calculate Precision for each class:

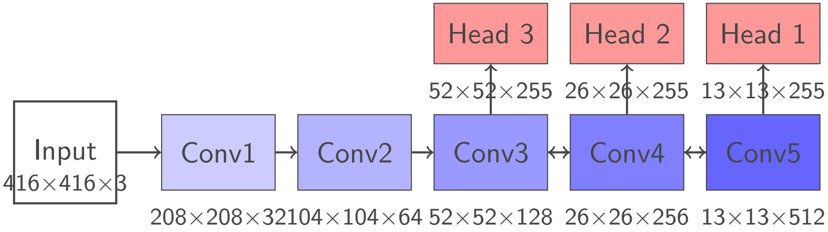
- **Cat Precision**: $\frac{15}{15+8} = \frac{15}{23} = 0.65$
- **Dog Precision**: $\frac{25}{25+5} = \frac{25}{30} = 0.83$

3. Compare: Dog class has higher precision (0.83 > 0.65)

Key Points:

Advanced Detection Architectures

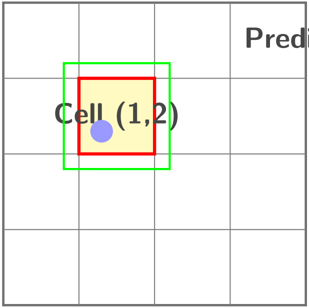
YOLO Architecture Deep Dive



Definition: YOLO Key Features

- **Single Shot**: One forward pass for detection
- **Multi-Scale**: 3 detection heads for different object sizes
- **Anchor-based**: Predefined anchor boxes for each grid cell
- **255 channels**: $(4 + 1 + 80) \times 3 = 255$ (bbox + conf + classes × anchors)

YOLO Prediction Format Explained



Prediction Vector for Cell (1,2):

Anchor 1: $[t_x, t_y, t_w, t_h, conf, p_1, p_2, ..., p_{80}]$
Anchor 2: $[t_x, t_y, t_w, t_h, conf, p_1, p_2, ..., p_{80}]$
Anchor 3: $[t_x, t_y, t_w, t_h, conf, p_1, p_2, ..., p_{80}]$

Where:

t_x, t_y: Box center offsets
t_w, t_h: Box width/height
$conf$: Objectness confidence
p_i: Class probabilities

Example: Decoding YOLO Predictions

$$b_x = \sigma(t_x) + c_x \tag{6}$$
$$b_y = \sigma(t_y) + c_y \tag{7}$$

Non-Maximum Suppression (NMS) Detailed

NMS Algorithm

Step 1: Sort by confidence
Red (0.9) > Blue (0.7) > Green (0.6) > Orange (0.4)

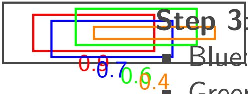
Step 2: Keep highest confidence (Red)

Step 3: Remove boxes with IoU > threshold

- Blue: IoU(Red,Blue) = 0.6 > 0.5 → Remove
- Green: IoU(Red,Green) = 0.4 < 0.5 → Keep
- Orange: IoU(Red,Orange) = 0.3 < 0.5 → Keep

Step 4: Repeat with remaining boxes

Before NMS



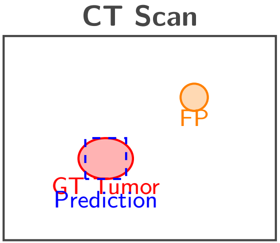
After NMS

Important: NMS Parameters

IoU Threshold: Typically 0.5 (higher = more suppression)
Confidence Threshold: Minimum confidence to consider

Real-World Case Studies

Case Study 1: Medical Imaging - Tumor Detection



Example: Results Analysis

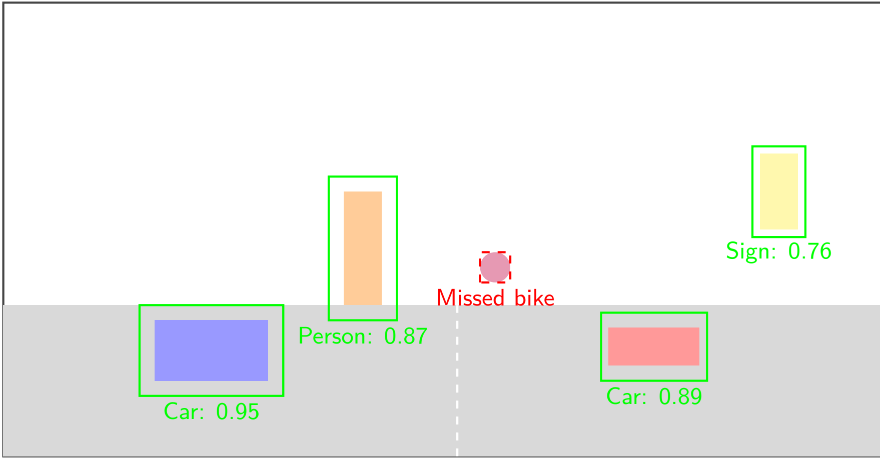
Challenge: High precision needed
IoU: 0.65 (good localization)
Issue: False positive rate too high

Important: Medical Considerations

- **High Recall** crucial (can't miss tumors)
- **False positives** create

Case Study 2: Autonomous Driving - Multi-Object Scene

Autonomous Vehicle Camera View



Example: Detection Re-