Introduction to Computer Vision

Mohammad Azam

Object detection is a computer vision technique that identifies and locates objects within images or videos.

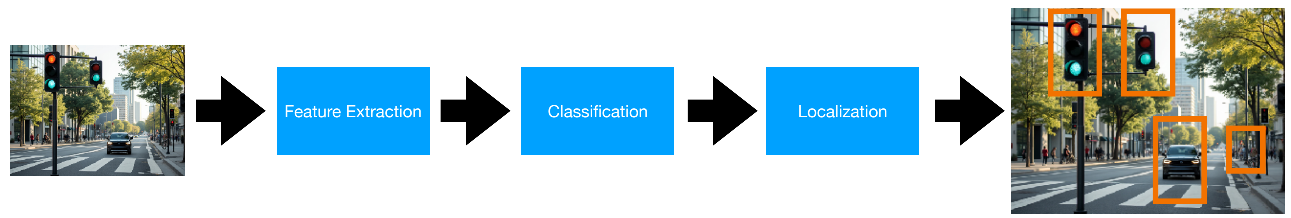
Start with a simple analogy: "Imagine you're looking at a crowded room, and you can easily point out where your friend is standing, where the door is, or where the snacks are. That's because your brain is really good at recognizing objects in your surroundings."

Transition to object detection: "Now, think about how a computer might do the same thing. That's what object detection is all about—teaching machines to not only see but also understand what they're seeing, and where those things are in an image or video."

Highlight applications: "Object detection is used in many cool technologies we interact with daily—like the face recognition on your phone, or the self-driving cars that need to recognize other vehicles, pedestrians, and traffic signs to navigate safely."

How Object Detection Works

How Object Detection Works



Introduce the process: "Let’s break down the process that a computer follows to detect objects. It’s not too different from how our brains work, but a bit more technical."
Step 1 - Input Image: "First, we start with an image. This could be a picture from a camera, a frame from a video, or any digital image."
Step 2 - Feature Extraction: "The computer then scans the image to find important features—like edges, shapes, and textures—that help it recognize what’s in the picture. Think of this as the computer’s way of 'looking for clues' about what might be in the image."
Step 3 - Classification: "Next, the computer tries to identify what these features belong to. Is that shape part of a car? A person? A cat? This step is called classification, where the computer assigns labels to different parts of the image."
Step 4 - Localization: "Once it knows what the objects are, the computer then figures out where exactly they are in the image. This step is called localization."
Step 5 - Bounding Boxes: "Finally, the computer draws a box around each identified object to show us where it is in the image. These are called bounding boxes, and they’re crucial for us to visually understand the computer’s detection."

Key Concepts in Object Detection

- **Classes:** Categories of objects (e.g., cars, people, animals).
- **Bounding Box:** Rectangular border around the detected object.
- **Confidence Score:** A number indicating how sure the model is about its detection.
- **Intersection over Union (IoU):** A measure of the overlap between predicted and actual bounding boxes.

Classes: "When we talk about classes, we're referring to the different types of objects the model is trained to detect. For example, one class could be 'car,' another could be 'bicycle,' and another could be 'person.' Each object detected is assigned a class."
Bounding Box: "The bounding box is a simple rectangular border drawn around the detected object. It shows us where the object is located in the image. The size and position of this box are crucial for accurate detection."
Confidence Score: "Now, detection isn't always perfect. The confidence score tells us how confident the model is about its detection. For example, if the model detects a dog and assigns it a confidence score of 0.95, it's 95% sure that what it detected is indeed a dog."
Intersection over Union (IoU): "IoU is a metric used to evaluate how well the predicted bounding box matches the actual location of the object. It's calculated by dividing the area of overlap between the predicted and ground truth boxes by the area of their union. The closer this value is to 1, the better the detection."

Popular Object Detection Models

- **YOLO (You Only Look Once):** Real-time object detection.
- SSD (Single Shot Multibox Detector): Fast and efficient detection.
- **Faster R-CNN:** Highly accurate detection, but slower.

YOLO: "YOLO stands for 'You Only Look Once,' and it’s one of the fastest object detection models out there. It processes the entire image at once, which makes it ideal for real-time applications like live video feeds."
SSD: "SSD, or Single Shot Multibox Detector, is another model known for its speed. Unlike traditional methods that might involve multiple passes over the image, SSD does it in one go, which is why it’s both fast and efficient."
Faster R-CNN: "For situations where accuracy is more important than speed, Faster R-CNN is a popular choice. It’s slower than YOLO and SSD but tends to be more accurate, especially for complex images with many objects."

Applications of Object Detection

- **Autonomous Vehicles:** Detecting pedestrians, vehicles, and traffic signs.
- **Security Systems:** Identifying intruders or suspicious objects.
- **Healthcare:** Analyzing medical images for tumors or abnormalities.
- **Retail:** Monitoring store inventory and customer behavior.

Autonomous Vehicles: "Object detection plays a critical role in autonomous vehicles. The car's sensors and cameras detect other vehicles, pedestrians, and traffic signs, helping the car navigate safely."
Security Systems: "In security, object detection is used to monitor video feeds for intruders or suspicious objects, helping to prevent crime or respond quickly when something goes wrong."
Healthcare: "In healthcare, object detection can be used to analyze medical images, such as X-rays or MRIs, to detect tumors or other abnormalities that might require attention."
Retail: "In the retail industry, object detection is used to monitor inventory levels, track customer behavior, and even prevent shoplifting by detecting unusual movements or patterns."

Challenges in Object Detection

- **Occlusion:** Objects partially hidden by others.
- **Variation in Lighting:** Different lighting conditions affect detection accuracy.
- **Scale Variability:** Objects of different sizes can be harder to detect.
- **Real-time Processing:** The need for fast detection in real-world applications.

Occlusion: "One of the big challenges in object detection is occlusion, where objects are partially hidden by others. For example, if a person is standing behind a car, the model might only see part of the person, making detection more difficult."
Variation in Lighting: "Lighting conditions can also pose a challenge. Bright sunlight, shadows, or dim lighting can all affect how well the model detects objects. Training models to be robust to these changes is crucial."
Scale Variability: "Another challenge is scale variability. Objects can appear at different sizes in an image, depending on how close or far they are from the camera. Detecting small objects can be particularly tricky."
Real-time Processing: "Finally, processing speed is a major concern, especially in real-time applications like self-driving cars or live video feeds. The model needs to detect objects quickly enough to make timely decisions."

Behind the Scenes

YOLO (You Only Look Once)

Image Classification

Traffic Light = 1

Car = 0

Object Localization

Where is the object in the picture?

Traffic Light = 1

Car = 0

Bounding Box

We are not only classifying the object we are also finding out where in the picture is the object.

Object Localization

Where is the object in the picture?

Probability of class = 1

x = 120

y = 5

width = 133

height = 228

traffic light = 1

car = 0

Object Localization

Where is the object in the picture?

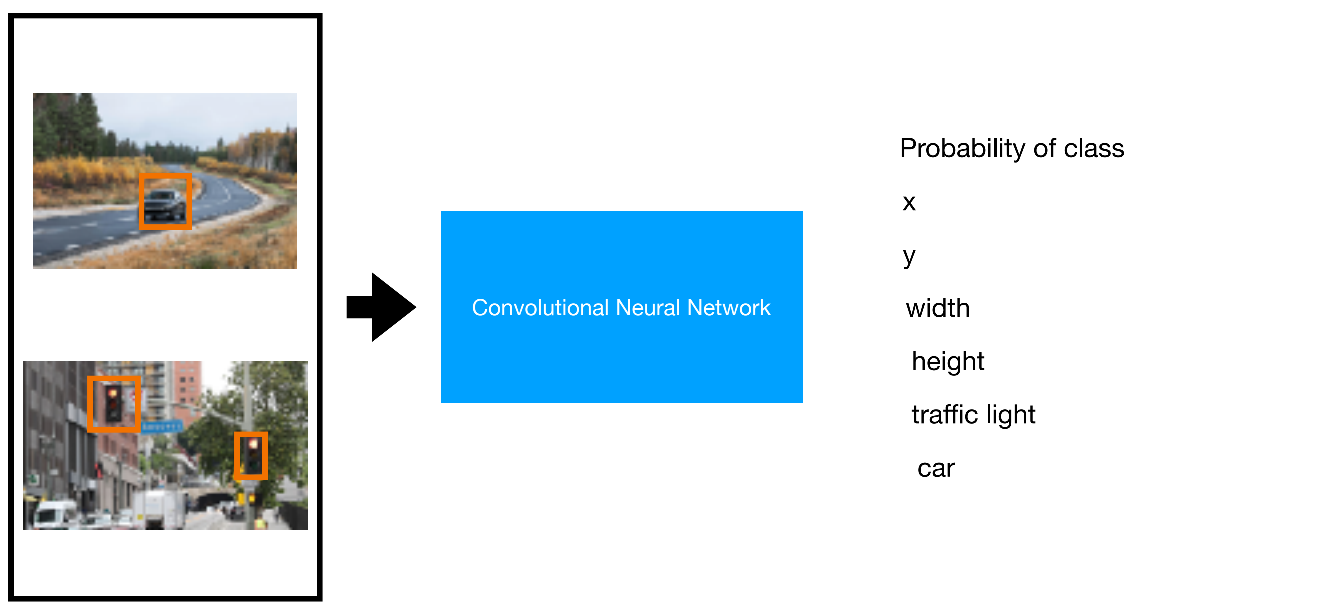
Probability of class = 1
x = 5
y = 5
width = 353
height = 189
traffic light = 0
car = 1

Object Localization

Where is the object in the picture?

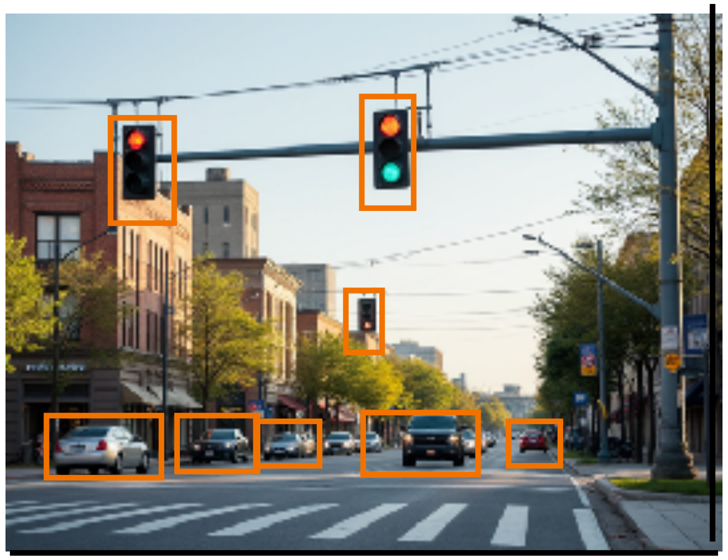
Probability of class = 1
x = 5
y = 5
width = 353
height = 189
traffic light = 0
car = 1

[1
5
5
353
189
0
1
]


Convolutional Neural Network
Probability of class
x
y
width
height
traffic light
car

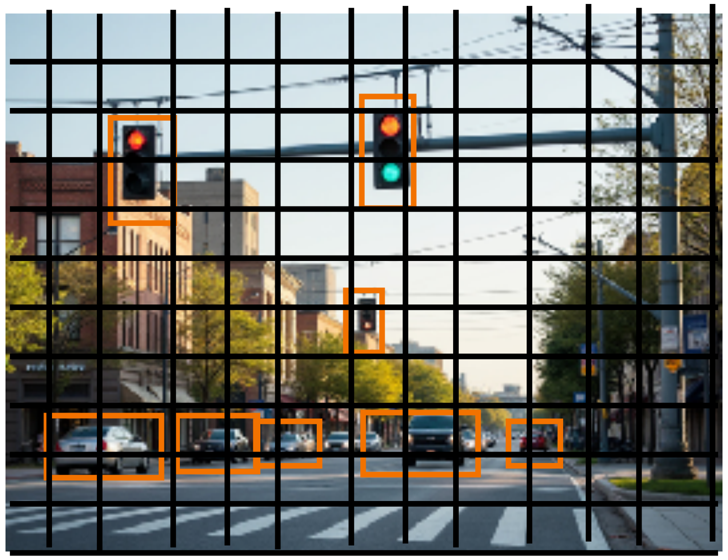
Object Localization

Multiple objects



Object Localization

Multiple objects



Object Localization

Multiple objects

Probability of class = 0

x = -

y = -

width = -

height = -

traffic light = -

car = -

Object Localization

Multiple objects

Probability of class = 1

x = 120

y = 90

width = 20

height = 60

traffic light = 1

car = 0

Object Localization

Multiple objects

Probability of class = 1

x = 20

y = 200

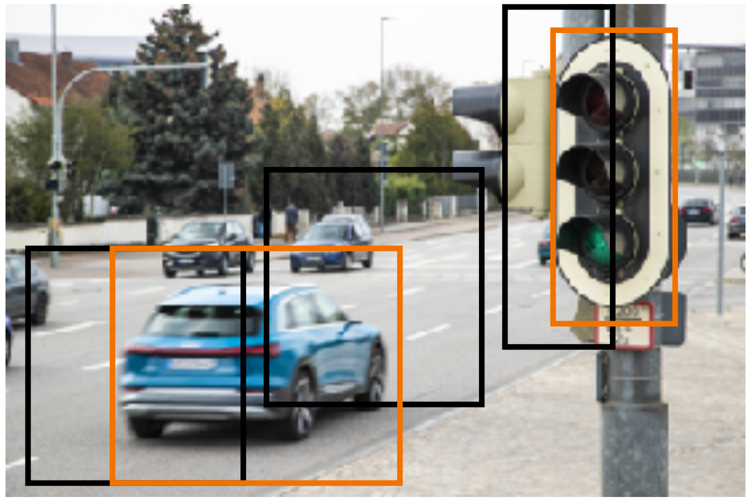
width = 100

height = 80

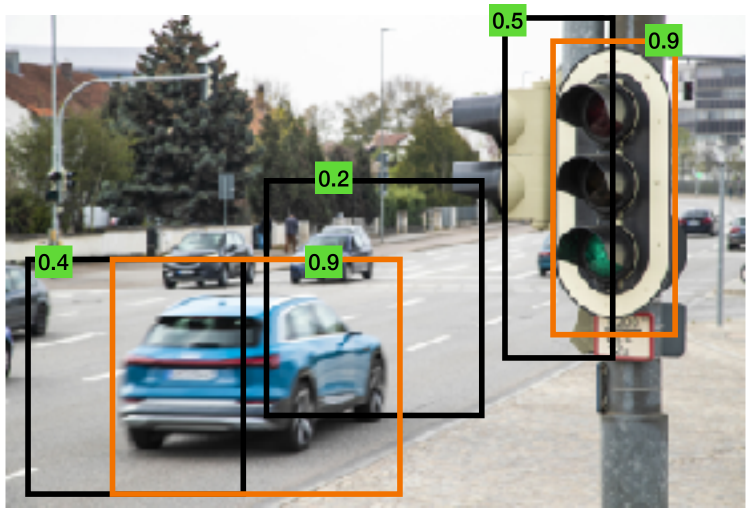
traffic light = 0

car = 1

Overlapping Bounding Boxes



Overlapping Bounding Boxes

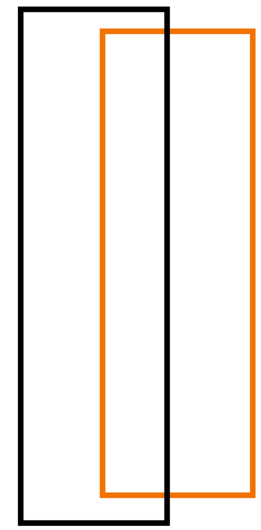


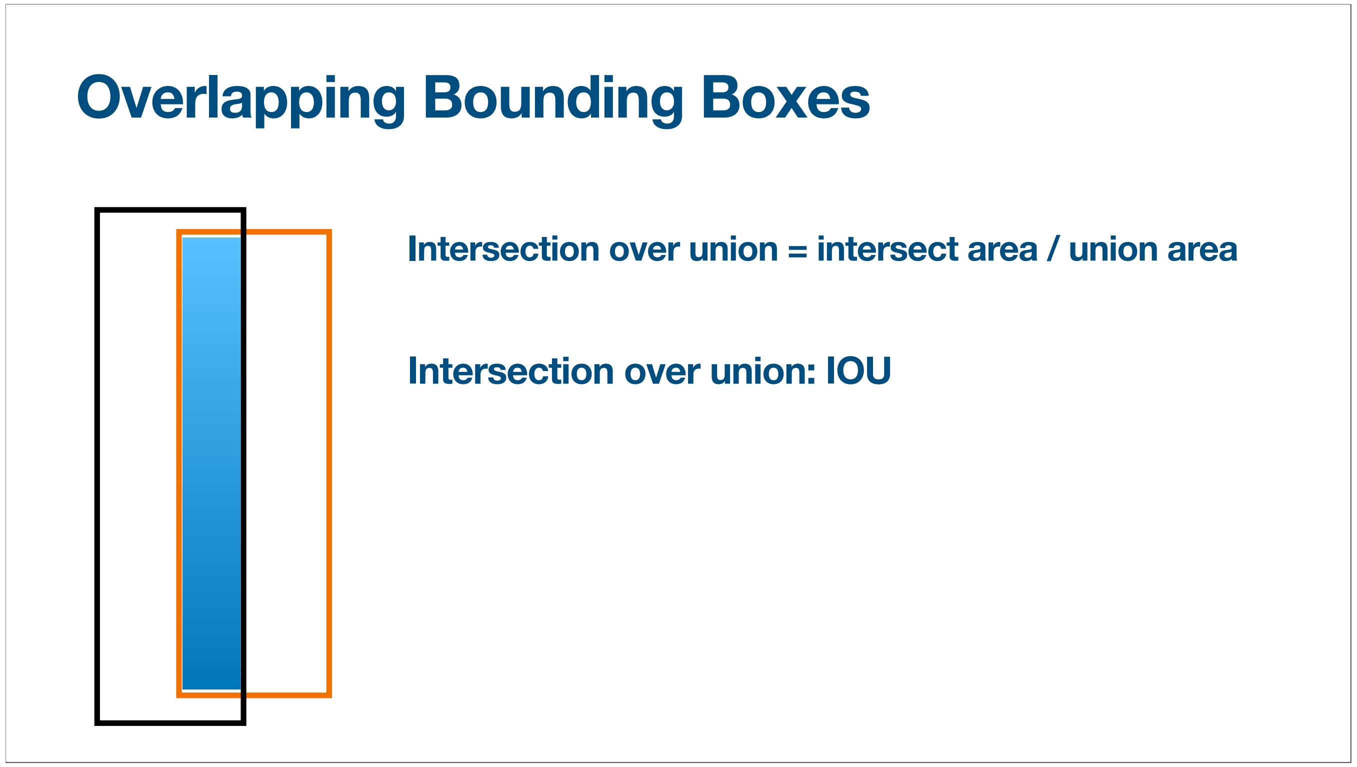
Overlapping Bounding Boxes

What if we take the max?

We cannot take the max because there can be other bounding boxes for traffic lights.

Overlapping Bounding Boxes





Intersect area value will be more if the boxes are overlapping. We discard the rectangles that are overlap. If IOU > 0.6 then overlapping. We discard those rectangles. And kept the rectangle with max probability.

Challenges and Considerations

- **Accuracy vs. Speed:** Balancing detection accuracy and processing speed.
- **Dealing with Small Objects:** Difficulty in detecting tiny objects.
- **Real-time Processing:** Importance of speed in applications like autonomous vehicles.

"Before we wrap up, it's important to consider some of the challenges in object detection. One of the biggest trade-offs is between accuracy and speed—more accurate models often take longer to process images, which can be a problem in real-time applications like autonomous driving. Another challenge is detecting small objects, which can be easily missed. Finally, for applications that require real-time detection, like security systems or live video analysis, the speed of detection is crucial."

Summary and Takeaways

- **Object detection involves multiple steps:** image acquisition, feature extraction, region proposal, classification, localization, and final output.
- Key techniques include CNNs for feature extraction and NMS for refining results.
- Applications span from autonomous vehicles to healthcare and security.

"To summarize, object detection is a complex but fascinating process that involves several key steps—from acquiring the image to outputting the final detection. Throughout this process, techniques like CNNs and Non-Maximum Suppression play crucial roles in ensuring accuracy and efficiency. Object detection is used in a wide range of applications, from helping self-driving cars navigate to improving healthcare diagnostics. Understanding these steps gives us insight into how machines are increasingly able to interpret and interact with the world around them."