Pre-reading Computer Vision Lecture 3

Advanced Computer Vision, Image Segmentation, Object Detection (YOLO)

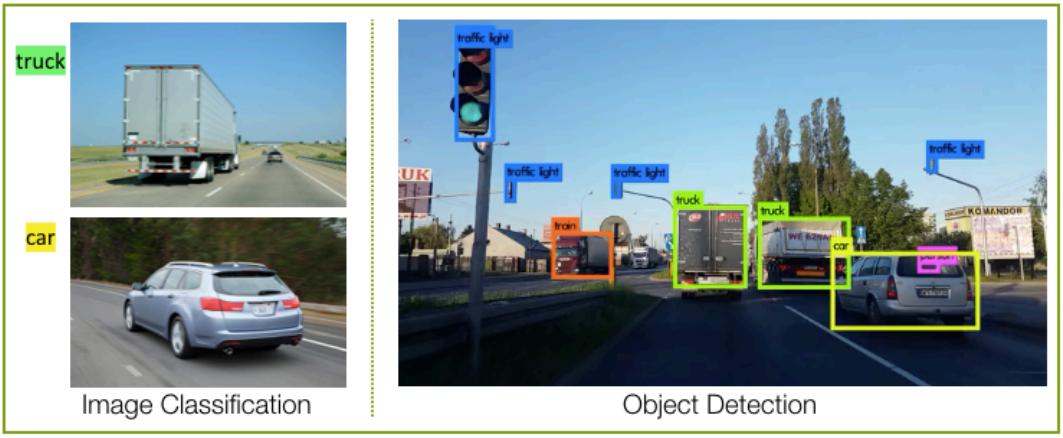
1. What is object detection, and how does it differ from image classification?

Answer:
Object detection involves identifying and locating multiple objects within an image. It not only classifies objects into predefined categories but also places bounding boxes around them to indicate their positions.
Difference:

- **Image classification** answers *"What is in the image?"* and assigns a single label (e.g., *Cat* or *Dog*).
- **Object detection** answers *"What objects are in the image and where are they?"* (e.g., *Cat at [x1, y1, x2, y2]*).

Practical Example:

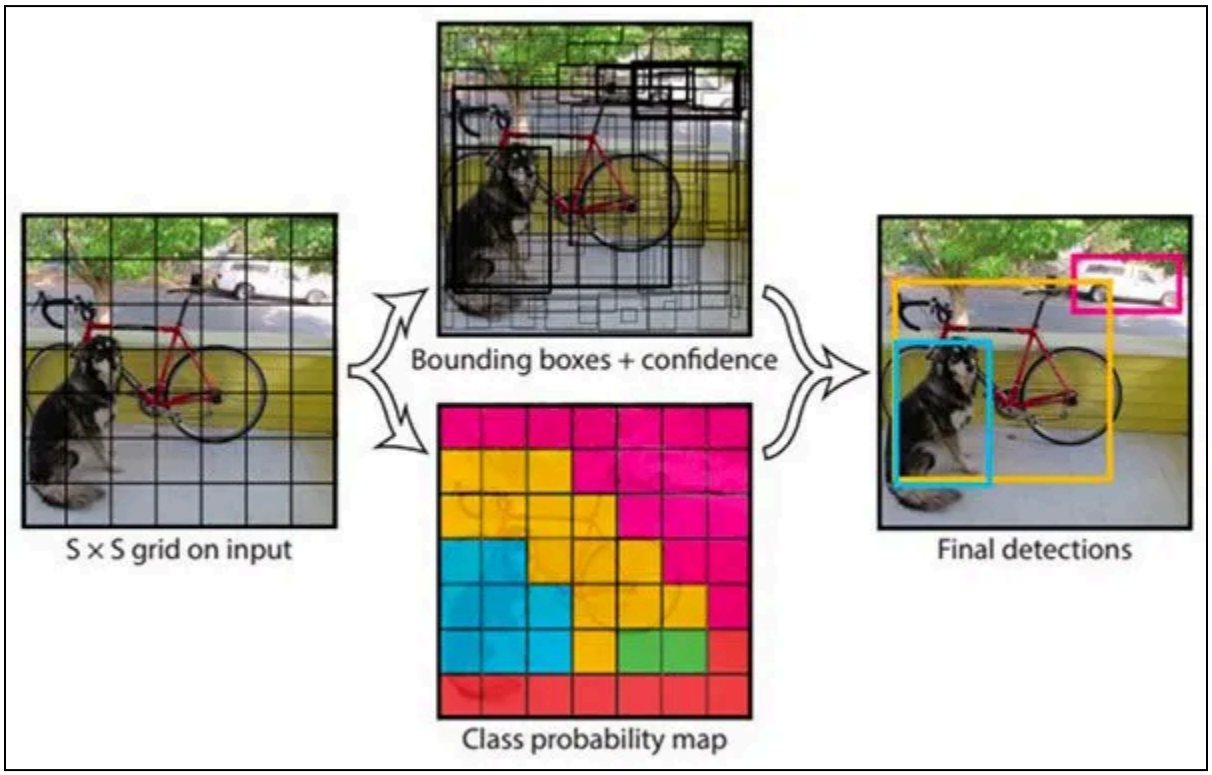
- **Image Classification**: Recognizing whether an image contains a car.
- **Object Detection**: Detecting and locating all cars in a traffic scene, which is crucial for applications like autonomous driving.



---

2. How does the YOLO (You Only Look Once) algorithm work for real-time object detection?

Answer:
YOLO divides an image into a grid and predicts bounding boxes and class probabilities for objects in each grid cell in a single pass through the network.



Key features:

- It predicts multiple bounding boxes per grid cell, refining them using anchor boxes.
- YOLO is fast and efficient, making it suitable for real-time applications.

Practical Example:

- Detecting pedestrians, vehicles, and traffic signals in a live feed for autonomous vehicles.
- Real-time security surveillance to identify unauthorized objects in restricted areas.

---

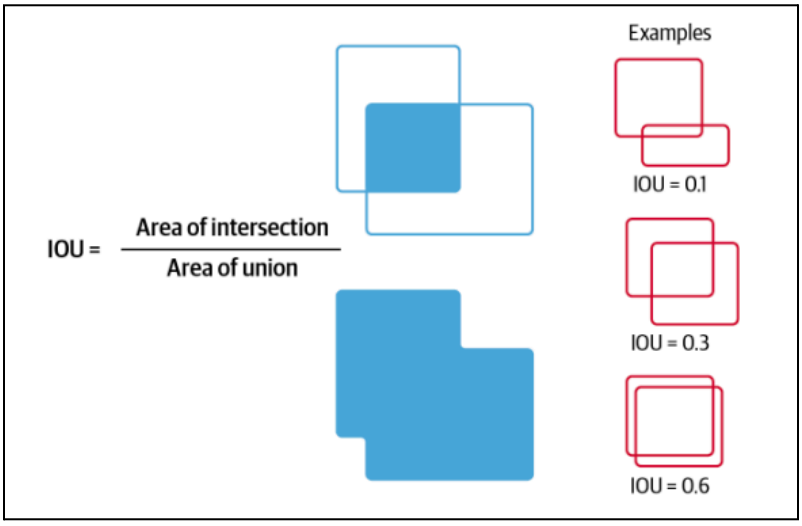
3. What is the significance of Intersection over Union (IoU) in evaluating object detection models?

Answer:
Intersection over Union (IoU) is a measure used to check how well a predicted bounding box aligns with the actual (ground truth) bounding box of an object in an image.

Imagine you're drawing a rectangle around a pedestrian in an image, and your computer model also draws a rectangle. IoU compares these two rectangles:

- **Intersection**: The overlapping area of the two rectangles.
- **Union**: The total area covered by both rectangles combined.

IoU = (**Area of Intersection**) / (**Area of Union**)



Example: Pedestrian Detection in ADAS
Advanced Driver Assistance Systems (ADAS) need to detect pedestrians to avoid accidents.

Let's see how IoU helps:

1. **Scenario**: The system detects a pedestrian and draws a bounding box. There's also a predefined "correct" bounding box (ground truth) from labeled data.

2. **IoU Calculation**:
    - The detected bounding box overlaps with the ground truth box by, say, 60%.
    - **IoU = (60% Overlap) / (Combined Area of Both Boxes)**
    - Assume this IoU comes out as 0.6 (60%).

3. **Deciding True Positive vs. False Positive**:
    - If **IoU ≥ 0.5**, the detection is considered a *True Positive*. This means the system detected the pedestrian accurately enough.
    - If **IoU < 0.5**, the detection is a *False Positive*. This means the system's prediction wasn't close enough to the actual pedestrian.

Why IoU Matters in Advanced Driver Assistance Systems (ADAS):

- **True Positive**: Helps ensure the system correctly identifies pedestrians and can initiate braking or warnings.
- **False Positive**: If the system mistakenly detects objects as pedestrians, it may trigger unnecessary braking, causing inconvenience or even accidents.

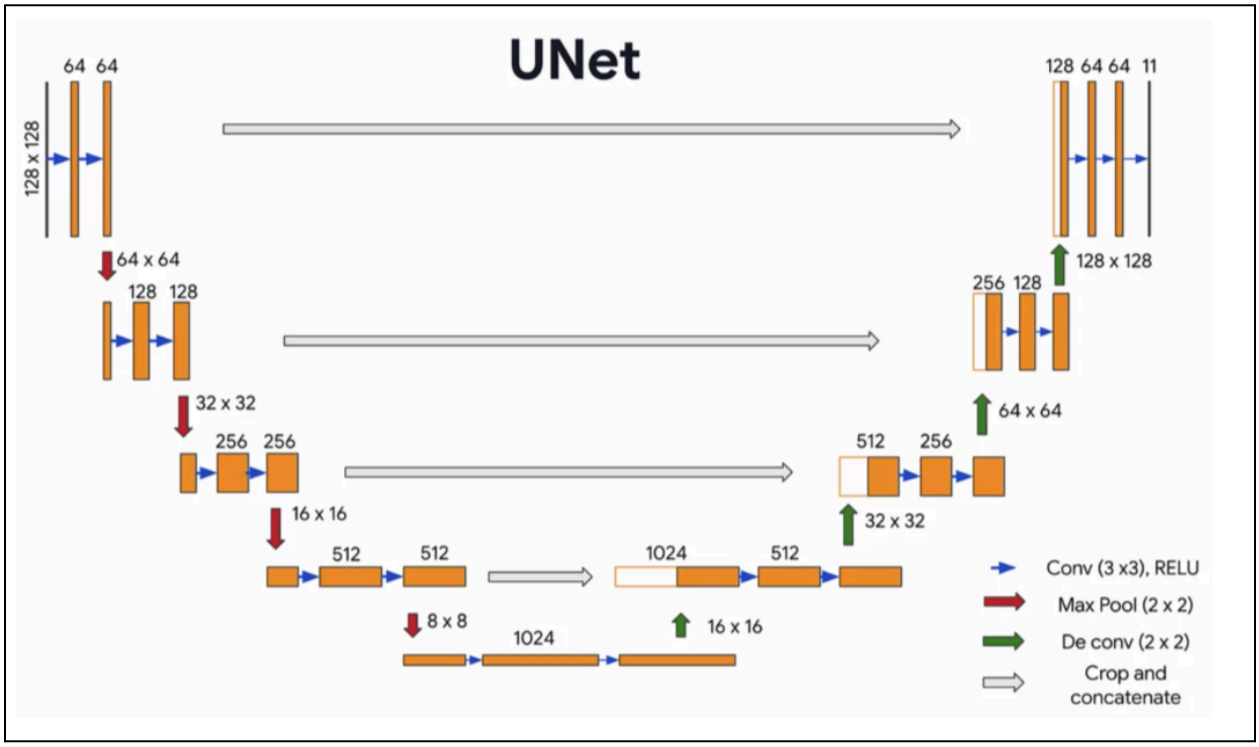
By setting a threshold like IoU ≥ 0.5, the system balances being accurate without being overly sensitive.

---

4. What role do transposed convolutions and skip connections play in segmentation models like U-Net?

Answer:

In the U-Net Architecture shown below, the green arrows represent the Transposed Convolutions (also called de-convolution) and white arrows represent Skip Connections.



- **Transposed Convolutions**: Imagine you have a small, low-resolution image, and you need to turn it into a larger, detailed one. Transposed convolutions are like a smart "zoom" tool for images. They make small, rough feature maps (patterns detected by the model) bigger and more detailed, helping the model make predictions for every pixel in the image.
- **Skip Connections**: Think of skip connections as shortcuts. When the model simplifies the image during processing, it might lose some important details. Skip connections "skip" these lost details from earlier steps and bring them back later to create sharper and more accurate boundaries.

Practical Example:
In medical imaging:

- **Transposed Convolutions** help doctors see detailed outlines of organs or tumors in blurry MRI or CT scans.
- **Skip Connections** ensure no important edges (like the exact tumor boundary) are missed during the process.

These tools make the segmentation accurate and reliable, which is crucial for medical diagnoses.

---

5. How is mean Average Precision (mAP) calculated for object detection models, and why is it important?

Answer:

What is mAP (mean Average Precision)?
mAP is a score that tells us how good an object detection model is. It looks at:

- **Precision**: How many of the detected objects are correct?
- **Recall**: How many of the actual objects were found?

It averages the best precision scores at different recall levels for every object type.

Why is it important?
mAP combines both precision and recall into a single number, making it easy to compare how well different models perform.

Practical Example:
In e-commerce, if a system detects and tags products (e.g., "shirt," "shoes") in user-uploaded photos, mAP helps measure how accurately and consistently the system identifies the products.

---

6. What are some common methods for creating and labeling datasets for computer vision tasks?

Answer:

- **Creating Datasets:**
    1. Collecting images using cameras or downloading public datasets.
    2. Simulating synthetic data using tools like Unity or Blender.

- **Labeling Datasets:**
    1. **Manual Labeling:** Annotating bounding boxes, segmenting pixels, or tagging classes using tools like CVAT or LabelImg.
    2. **AI-Assisted Labeling:** Using pretrained models to generate initial labels for review.
    3. **Crowdsourcing:** Platforms like Amazon Mechanical Turk for large-scale annotation.

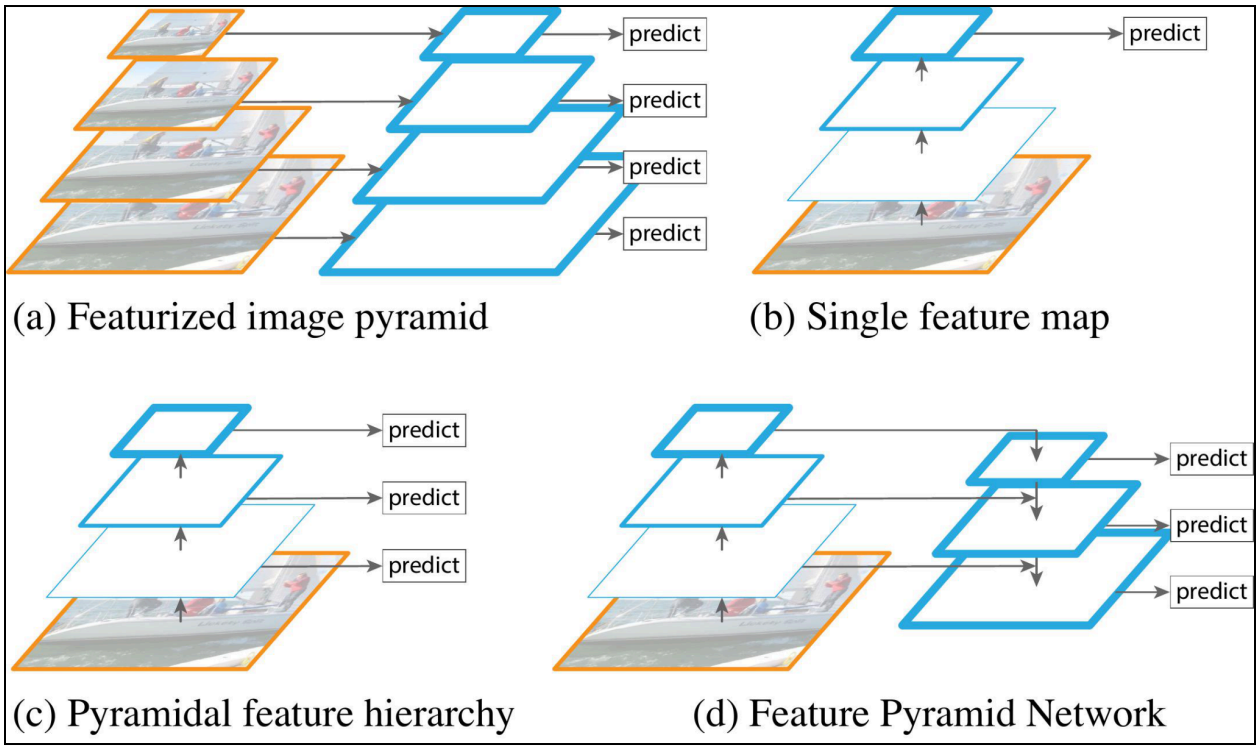
Practical Example:

- Collecting traffic images and labeling them for object detection in an autonomous driving project.

---

7. How do feature pyramid networks enhance object detection accuracy compared to using a single feature map?

Answer:
Feature Pyramid Networks (FPNs) combine feature maps from different levels of a CNN. Lower-level features provide fine details, while higher-level features capture semantic information. This enables better detection of objects at different scales.



Practical Example:

- Detecting both small objects (e.g., pedestrians far away) and large objects (e.g., nearby buses) in traffic scenarios.

For more details on Understanding Feature Pyramid Networks for object detection, refer here.

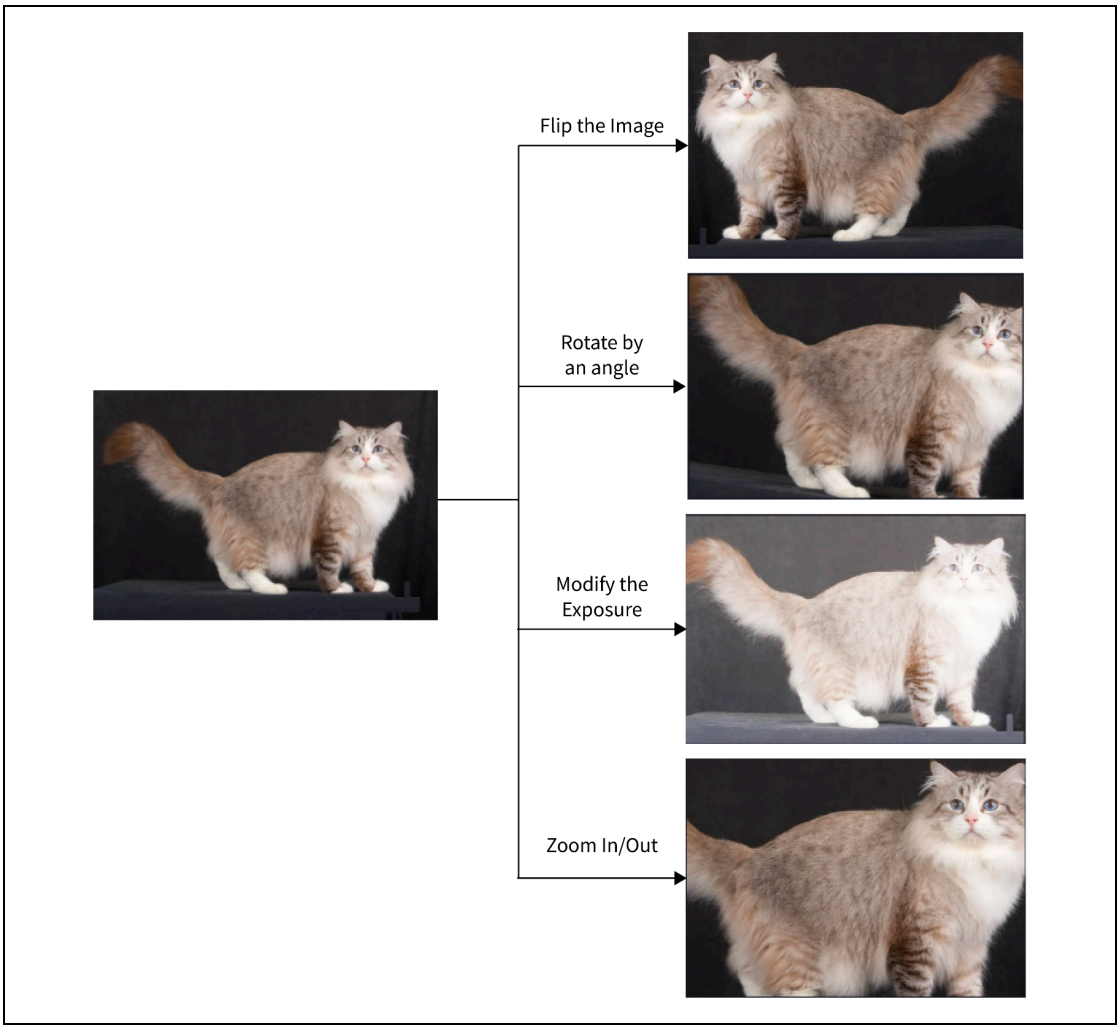
---

8. Why is data augmentation important in training computer vision models, and what are some common techniques?

Answer:
Data augmentation artificially increases the diversity of training data, improving model generalization and robustness.

Common Techniques:

1. **Flipping:** Horizontal or vertical flips.
2. **Rotation:** Random rotations to make the model invariant to orientation.
3. **Scaling and Cropping:** Changing object sizes or focusing on specific regions.
4. **Color Jittering:** Altering brightness, contrast, or saturation.
5. **Gaussian Noise:** Adding random noise to simulate real-world conditions.



Practical Example:

- In satellite image analysis, applying rotation and scaling helps detect objects from various angles and distances.

~~~END~~~