Assignment 4 - Part 2: Object Detection Model Comparison

Jana Adel 8273 Shahd Sherif 8145

24th of June 2025

Model Comparison and Analysis

This section presents an in-depth comparison of three prominent object detection models: **Faster R-CNN**, **SSD**, and **YOLOv5**. The comparison considers both the underlying architecture and practical inference results on COCO validation images. Each model was run on a fixed subset of images, and results were evaluated visually and analytically.

1. Faster R-CNN

Faster R-CNN is a two-stage object detection model proposed by Ren and others in 2015. It significantly improved detection accuracy compared to earlier methods.

- **Backbone:** ResNet-50 with Feature Pyramid Network (FPN).
- **Stage 1:** A Region Proposal Network (RPN) scans the image to propose candidate object regions.
- **Stage 2:** These regions are fed into a classification and bounding box regression head.
- **Strengths:**
  - Excellent detection performance on small and overlapping objects.
  - High accuracy, especially in cluttered or complex scenes.
- **Weaknesses:**
  - Slow inference time due to its two-stage nature.
  - Requires significant computational resources.
- **Best For:** Applications where accuracy is more important than speed, e.g., medical imaging, satellite imagery.

2. SSD (Single Shot MultiBox Detector)

SSD is a one-stage object detection network introduced by Liu and others in 2016. Unlike Faster R-CNN, SSD performs detection in a single pass.

- **Backbone:** VGG-16.
- **Core Idea:** SSD predicts bounding boxes and class probabilities directly from multiple feature maps of different resolutions.
- **Strengths:**
  - Simpler and faster than two-stage models.
  - Performs well on medium-to-large objects.
- **Weaknesses:**
  - Less accurate with small and densely packed objects.
  - The fixed set of default boxes limits localization precision.
- **Best For:** Real-time applications with moderate accuracy needs, such as embedded systems.

3. YOLOv5

YOLOv5 is the latest iteration of the YOLO (You Only Look Once) series, developed by Ultralytics. It provides an optimal balance between speed and accuracy using modern backbone and training strategies.

- **Backbone:** CSPDarknet.
- **Architecture:**
  - Uses a single CNN to predict bounding boxes and class probabilities across the entire image.
  - Incorporates modern techniques like spatial pyramid pooling, path aggregation, and anchor-free prediction.
- **Strengths:**
  - Real-time inference speed with competitive accuracy.
  - High performance in detecting objects at multiple scales.
- **Weaknesses:**
  - Occasionally over-predicts (false positives).
  - Slightly lower accuracy than Faster R-CNN on very small objects.
- **Best For:** Real-time object detection in videos, surveillance, robotics.

Quantitative Results

Table 1: Performance Comparison Summary

Model	Speed (s/image)	mAP (estimated)	Strengths	Weaknesses
Faster R-CNN	0.5 – 1.0	0.37	High accuracy, small object detection	Very slow inference, needs GPU
SSD	0.2 – 0.4	0.31	Fast and lightweight	Poor with small or overlapping objects
YOLOv5	0.04 – 0.1	0.36	Fast and accurate balance	False positives in background

Qualitative Observations

- **Faster R-CNN** correctly detected overlapping persons in crowded scenes, and small distant objects (e.g., bottles, animals) were localized precisely.
- **SSD** performed well in detecting large cars and bikes, but often missed small objects like birds or remote people.
- **YOLOv5** showed balanced behavior: it was fast and detected most major objects correctly, though it occasionally produced boxes in background areas.

Conclusion

Each model has specific strengths and is suited to different deployment scenarios:

- Use **Faster R-CNN** when detection quality is the priority and computational cost is not a concern.
- Choose **SSD** for fast inference in constrained environments but avoid it for complex scenes.
- **YOLOv5** is a great middle-ground with state-of-the-art speed and solid accuracy for real-time use.