Pascal VOC 2007 Object Detection using Faster R-CNN and YOLOv8

# NAME: Ashish Dnyaneshvar Parulekar

# STUDENT ID: 1017913240

# COURSE: Advanced Computer Vision (CPS-667)

# Assignment 3

# Introduction

The goal of this project is to implement and compare two object detection models — Faster R-CNN and YOLOv8 — on the Pascal VOC 2007 dataset. The models were trained and evaluated independently, and their performance was compared in terms of detection accuracy (mAP), speed, and usability in an interactive setting.

# Steps Implemented

## Step 1: Model 1 – Faster R-CNN

* Imported from torchvision.models.detection.fasterrcnn\_resnet50\_fpn.
* Initialized with pretrained COCO weights and modified to detect 20 Pascal VOC object classes.
* Trained for 2–8 epochs using the Pascal VOC 2007 trainval split.
* Evaluated using a custom mAP implementation with IoU and score thresholding.
* Output visualizations were generated using matplotlib for bounding boxes and class labels.

## Step 2: Model 2 – YOLOv8

* Implemented using the ultralytics YOLOv8 framework.
* Pascal VOC annotations were converted to YOLO format for compatibility.
* Trained for 50 epochs on the same trainval split.
* Evaluated using YOLOv8's built-in metrics.
* Output includes bounding boxes with class scores for each detection

## Step 3: Performance Evaluation (mAP & Inference Time)

| **Metric** | **Faster R-CNN** | **YOLOv8** |
| --- | --- | --- |
| mAP@0.5 | 0.7311 | 0.8959 |
| mAP@0.5:0.95 | — | 0.6963 |
| Precision | 0.6172 | 0.8963 |
| Recall | 0.8965 | 0.8165 |
| Inference Time (100 images) | 3.07 sec | 0.96 sec |

## Step 4: Sample Visual Outputs

A dog with its mouth open

AI-generated content may be incorrect.

Figure 1: YoloV8 Prediction

A close-up of a dog

AI-generated content may be incorrect.

Figure 2: Faster RCNN Prediction

## Step 5: Interactive Detection Program

* Implemented in Model.ipynb.
* User chooses between rcnn and yolo.
* Uploads image or pastes image URL.
* The system loads the selected model and runs inference.
* Output image with bounding boxes and labels is displayed in the notebook.
* Repeat prompt allows multiple predictions.Implemented plt.
* pause() and sys.stdout.flush() to support interaction in Colab.

# Conclusion

This project successfully compares the performance of two state-of-the-art object detection models. YOLOv8 outperformed Faster R-CNN in terms of speed and detection accuracy, while Faster R-CNN showed slightly better recall. The interactive detection tool allows users to test either model on arbitrary images or visually inspect the results, demonstrating practical deployment of object detection models.