National University of Sciences and Technology

School of Electrical Engineering and Computer Science

Department of Computing

CS-405 Deep Learning

Fall 2023

# Lab 8

**Object Detection**

# Date: 2nd November 2023

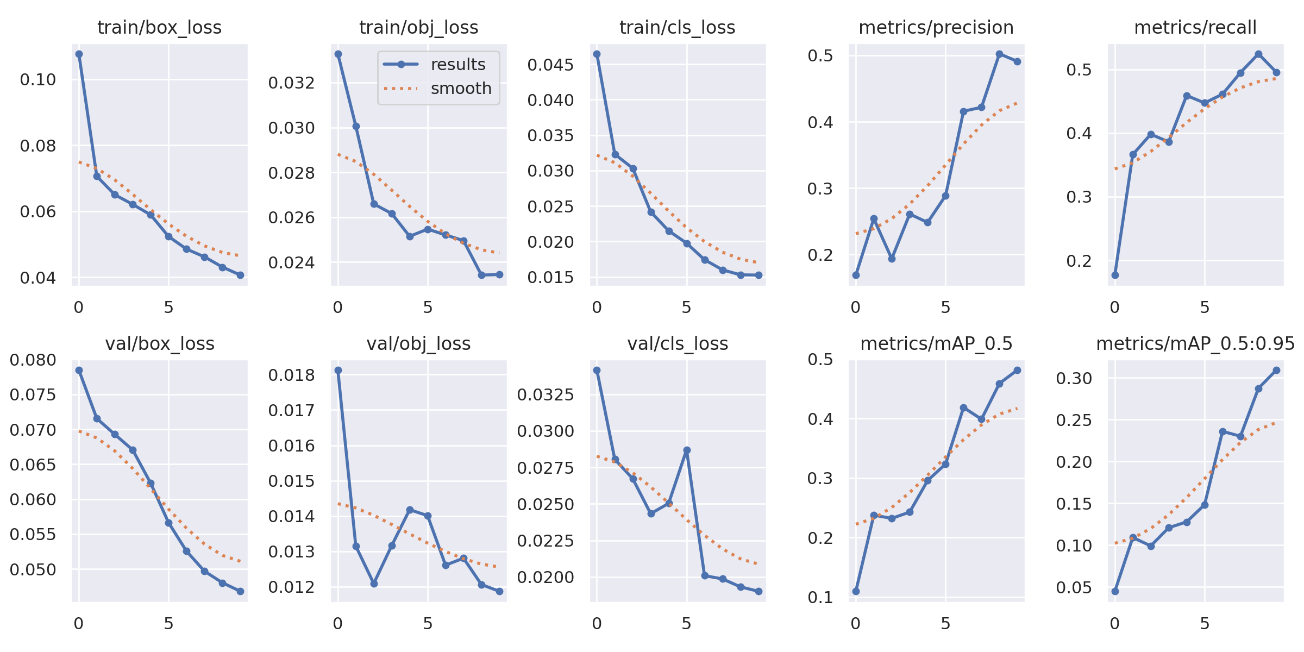
# Name: Alina Nasir

# CMS ID: 342350

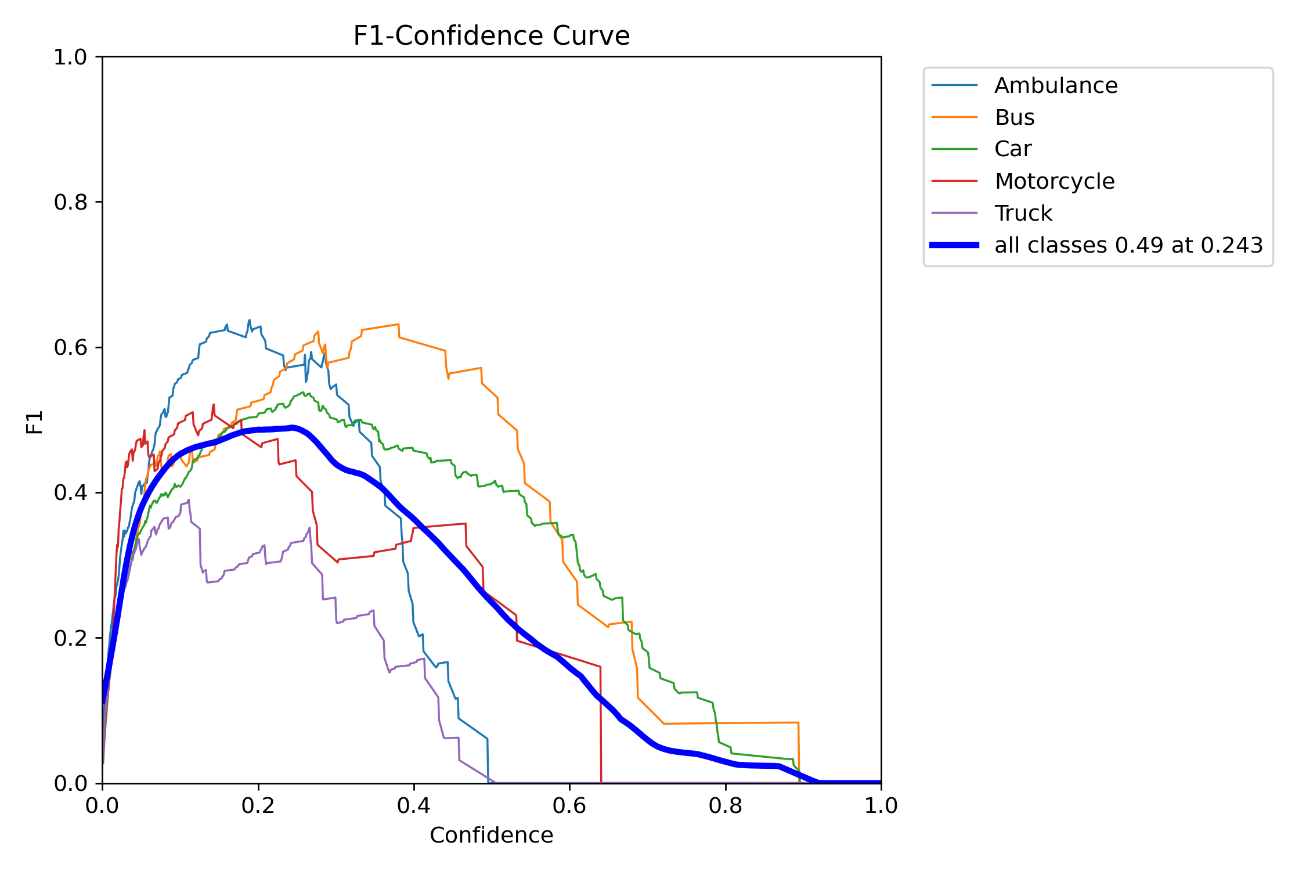
# Instructor: Dr Daud Abdullah

**Please fully train the YOLOv5s (small) and YOLOv8s (small) models on the provided dataset**

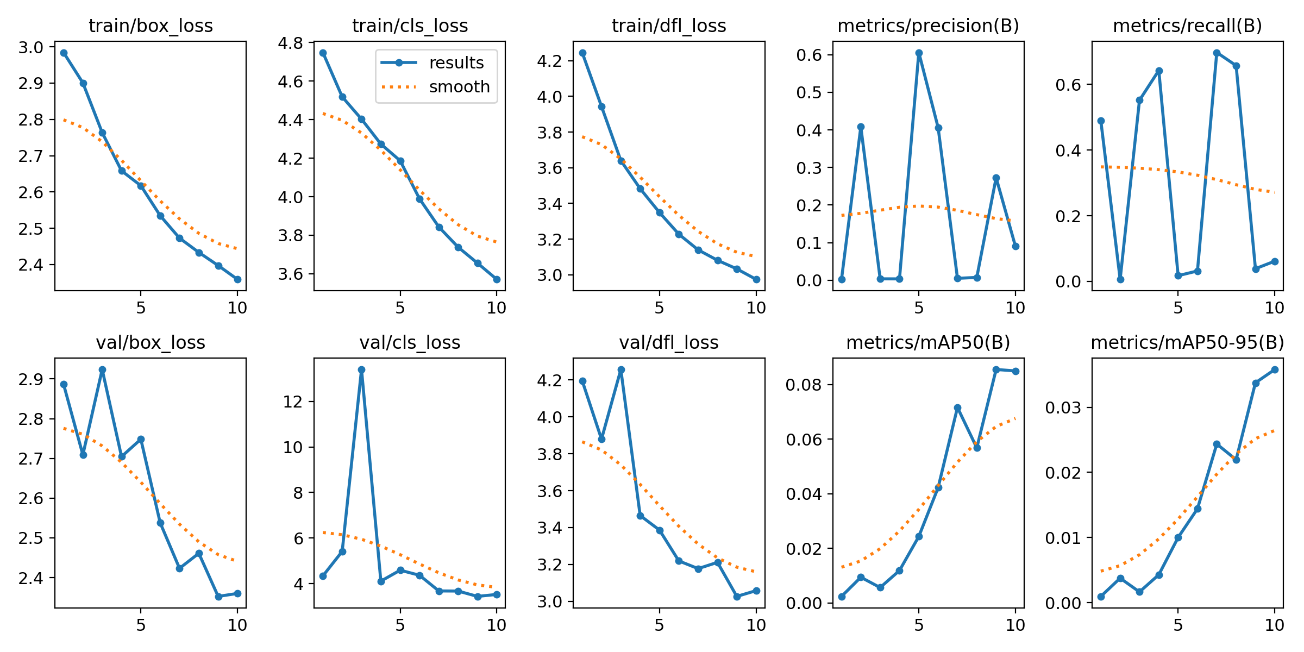
**Yolov5 training plots:**

****

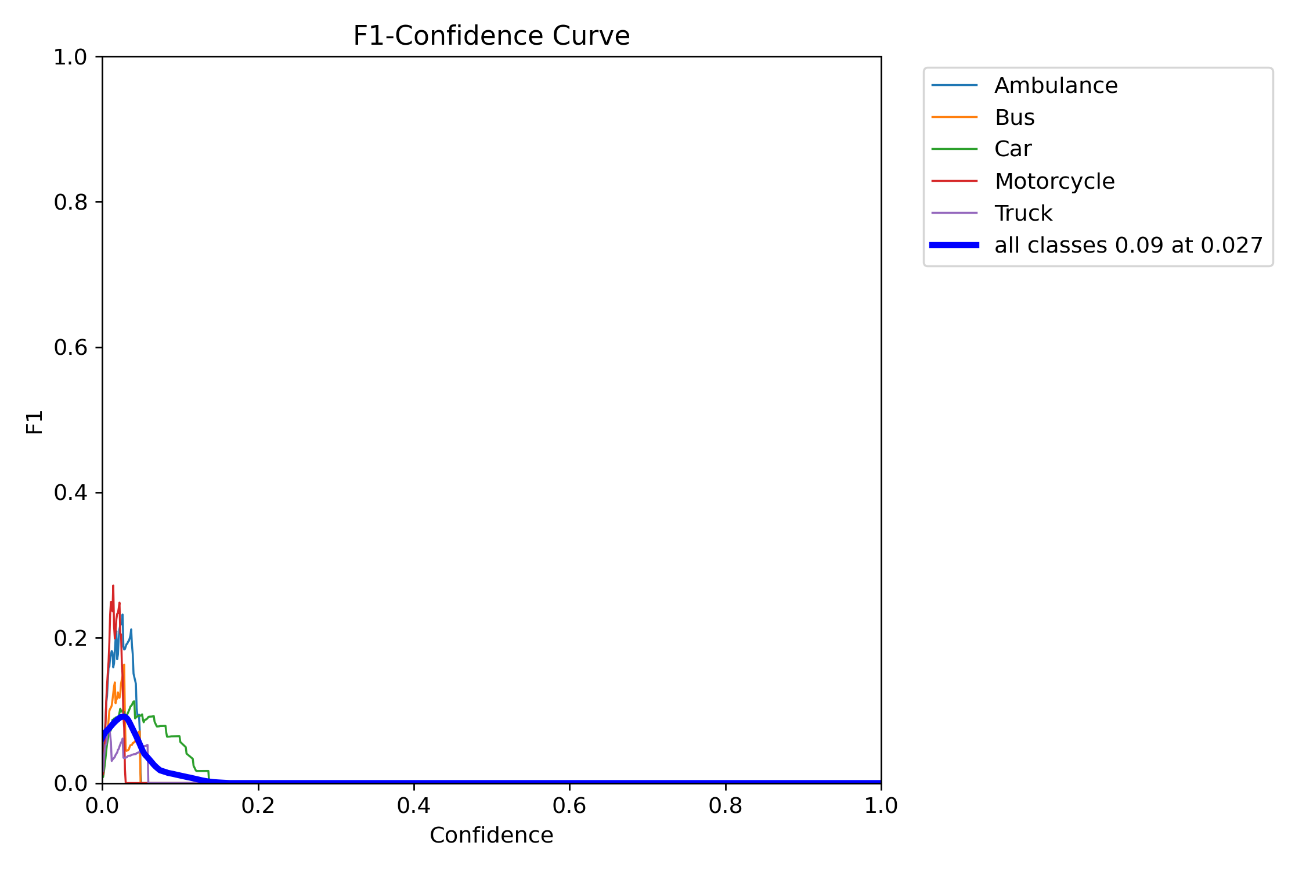
**F1 Curve:**

****

**Yolov8 training plots:**

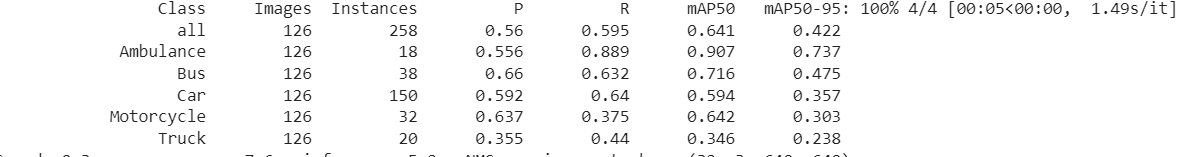
****

**F1 Curve:**

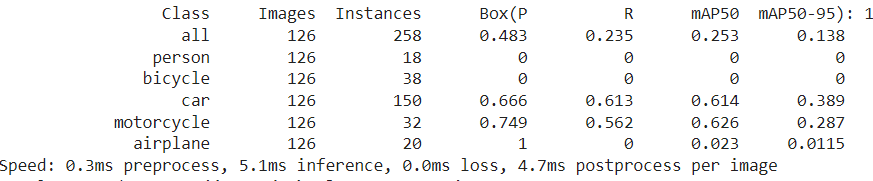
****

**Evaluate your models on the test sets and provide the mAP values.**

**Yolov5 test mAp values:**

****

**Yolov8 test mAp values:**

****

**Architecture of Yolov5:**

The architecture of Yolov5 has the following characteristics:

1. **Backbone:**

YOLOv5 uses a CSPDarknet53 architecture as its backbone network. This architecture is designed to extract feature maps from the input image and perform initial feature extraction. CSPDarknet53 is an extended version of Darknet, which is a lightweight neural network framework designed for object detection and related tasks and has been applied Cross Stage Partial (CSP) network strategy. This helps to solve the problem of vanishing gradients in the deep neural network

1. **Neck:**

YOLOv5 brought two major changes to the model neck. First a variant of Spatial Pyramid Pooling (SPP) has been used, and the Path Aggregation Network (PANet) has been modified by incorporating the BottleNeckCSP in its architecture. YOLOv5 features PANet (Path Aggregation Network) in its neck architecture, which helps improve the fusion of features at different scales. PANet allows the model to capture object details at both coarse and fine scales, making it more effective in detecting objects of various sizes. SPP block performs an aggregation of the information that receives from the inputs and returns a fixed length output. Thus it has the advantage of significantly increasing the receptive field and segregating the most relevant context features without lowering the speed of the network.

1. **Detection Head:**

The detection head of YOLOv5 consists of three YOLO layers. Each YOLO layer predicts object bounding boxes, class probabilities, and objectness scores for a specific scale. The predictions are made in a grid format, where each grid cell predicts objects within its spatial context.

1. **Anchor Boxes:**

YOLOv5 employs anchor boxes to predict object locations and sizes. Anchor boxes are predefined bounding box shapes and sizes that are used to estimate the objects' positions and dimensions. YOLOv5 adjusts these anchor boxes during training to better fit the dataset being used.

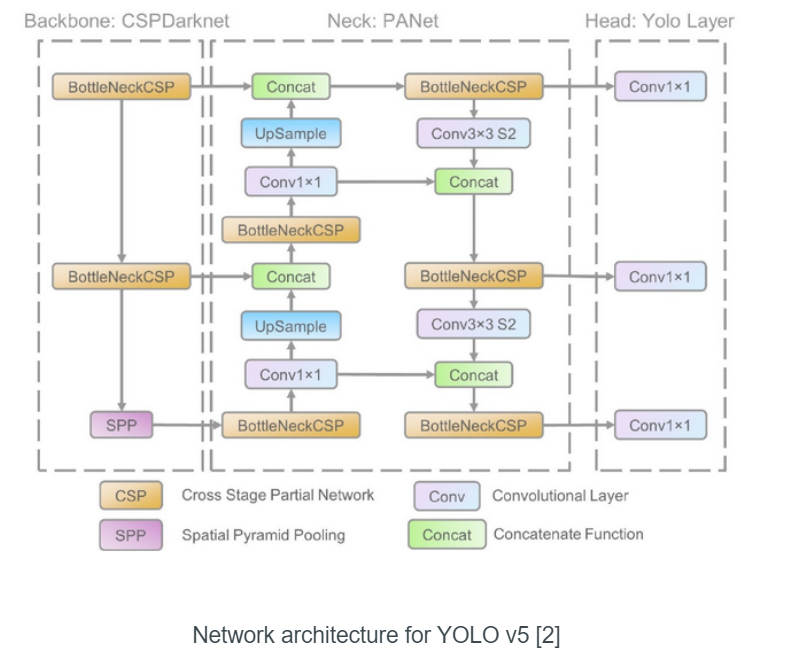
1. **Loss Function:**

YOLOv5 uses a combination of loss functions, including the localization loss, confidence loss, and class probability loss. These loss functions are used to optimize the model's ability to detect objects accurately.

1. **Post-processing:**

After the model has made predictions, post-processing is applied to filter out low-confidence detections and perform non-maximum suppression (NMS) to retain the most confident and non-overlapping detections.

This information is further displayed by the architecture:



**Architecture of Yolov8:**

The architecture of YOLOv8 builds upon the previous versions of YOLO algorithms. YOLOv8 utilizes a convolutional neural network that can be divided into two main parts: the backbone and the head.

**Backbone:**

A modified version of the CSPDarknet53 architecture forms the backbone of YOLOv8. This architecture consists of 53 convolutional layers and employs cross-stage partial connections to improve information flow between the different layers.

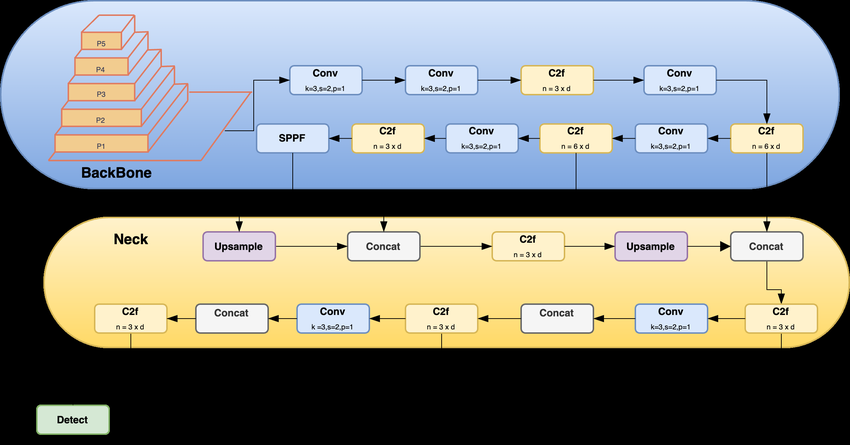
**Head:**

The head of YOLOv8 consists of multiple convolutional layers followed by a series of fully connected layers. These layers are responsible for predicting bounding boxes, objectness scores, and class probabilities for the objects detected in an image. One of the key features of YOLOv8 is the use of a self-attention mechanism in the head of the network. This mechanism allows the model to focus on different parts of the image and adjust the importance of different features based on their relevance to the task.

**Multi-Scaled Object Detection:**

The model utilizes a feature pyramid network to detect objects of different sizes and scales within an image. This feature pyramid network consists of multiple layers that detect objects at different scales, allowing the model to detect large and small objects within an image.

The following is architecture diagram of Yolov8:



**Comparison of Yolov5 and Yolov8 performance:**

The yolov8 model is faster than yolov8 model as it takes lower time to process

**In your report, discuss which model is performing better on the test set and what can be the reason? If YOLOV8 is performing better, then what are the innovations that allow YOLOV8 to perform better than YOLOV5.**

According to the mAp values of car for yolov5 is 0.594 whereas for yolov8 it is 0.614 and hence the accuracy of Yolov8 model is greater than Yolov5 model. YOLOv8 is built on the YOLOv5 framework and includes several architectural and developer experience improvements. It is faster and more accurate than YOLOv5, and it provides a unified framework for training models for performing object detection, instance segmentation, and image classification.

**Show predictions of both models on 4-5 test images in the report and compare them side by side.**

|  |  |
| --- | --- |
| **Yolov5 Predictions** | **Yolov8 Predictions** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |