Object Detection for Autonomous Vehicles

Team Members :

- Madhusudhan - SE21UARI091
- Prabhas - SE21UARI176
- Varshith - SE21UARI185
- Susmitha - SE21UARI180
- Susrith - SE21UARI164

FACULTY : Dr. MAHESH CHOWDARY KONGARA

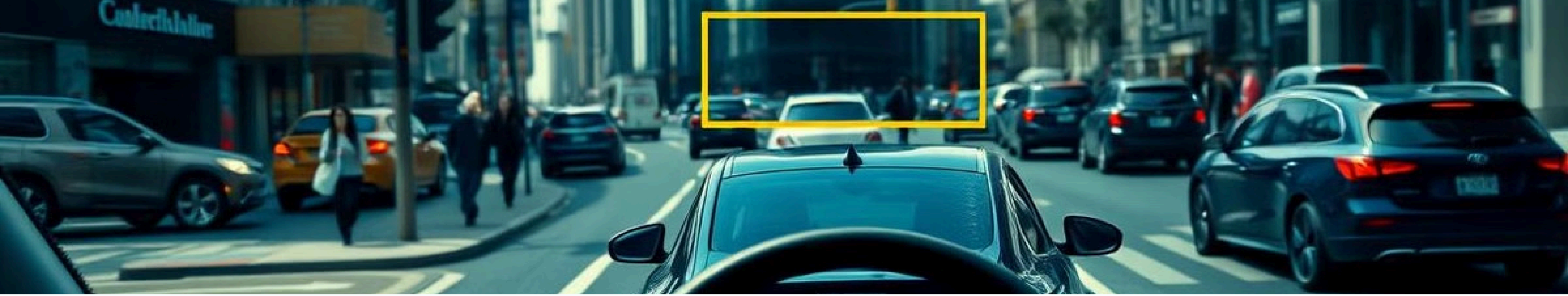
What is Object Detection?

The Task

Object detection involves identifying and locating specific objects within an image or video. This includes classifying the object type and determining its precise location and dimensions.

The Challenge

Object detection is a challenging task because of factors like varying object sizes, poses, lighting conditions, and occlusions. It requires powerful algorithms that can handle these complexities.



Objectives of Object Detection

- Accurate Object Localization : Precisely locate the boundaries of each detected object within the image, minimizing false positives and false negatives.
- Reliable Object Classificatio : Correctly identify the type of object, such as cars, pedestrians, or traffic signs, with high accuracy.
- Real-time Performance : Process images and videos at a fast enough rate to enable realtime applications, such as self-driving cars or security systems.

Real-world Applications of

1

Autonomous Driving

Object detection plays a vital role in self-driving cars by enabling them to identify and track objects like cars, pedestrians, and traffic signs.

2

Security and Surveillance

Object detection can be used to monitor security cameras for suspicious activities, such as intrusion detection or recognizing unauthorized individuals.

3

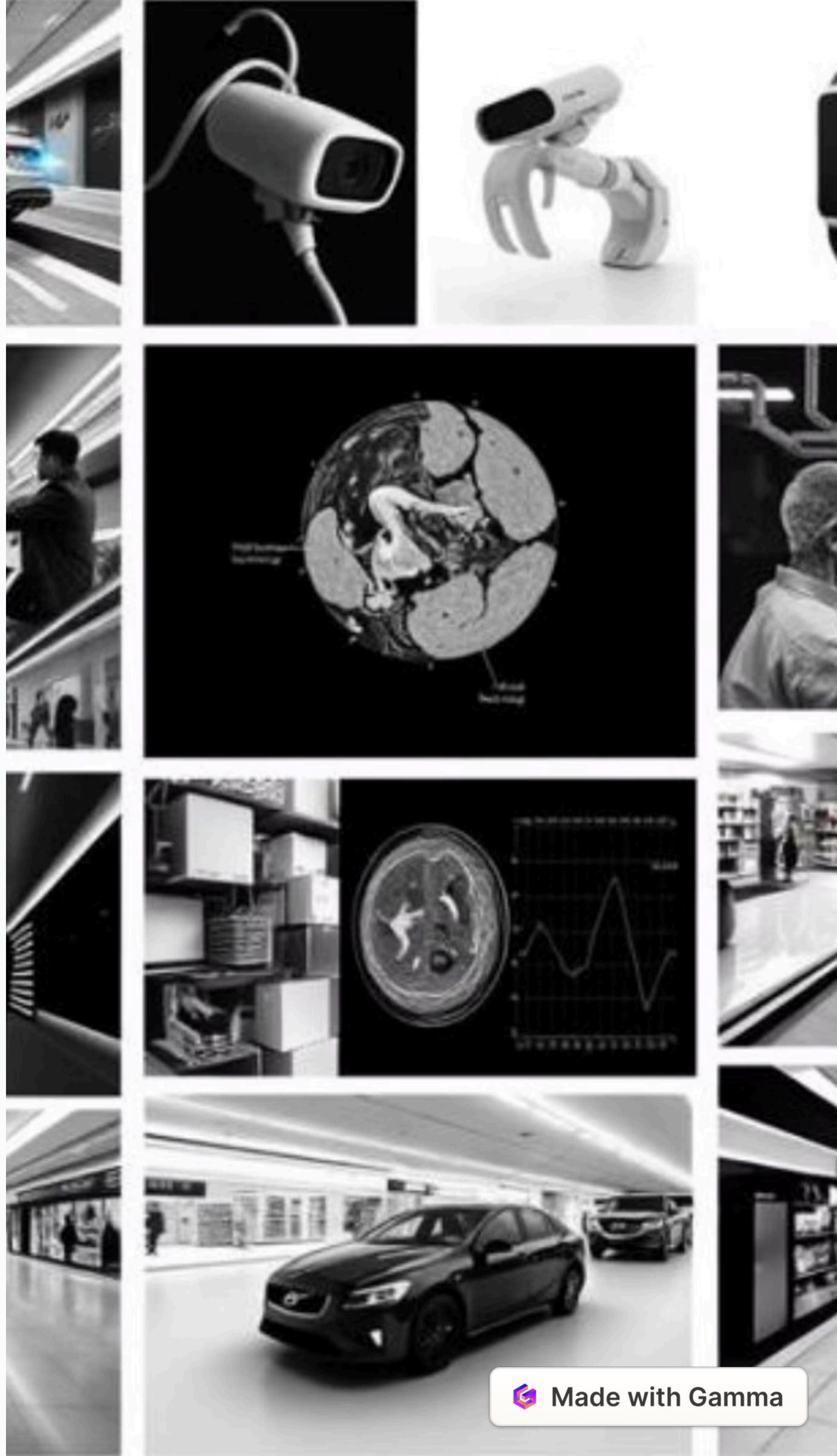
Medical Imaging

In medical imaging, object detection can help radiologists identify tumors or other abnormalities in X-rays, CT scans, or MRI images.

4

Retail Analytics

Object detection can track customer movement, analyze product popularity, and optimize store layouts to improve retail efficiency.



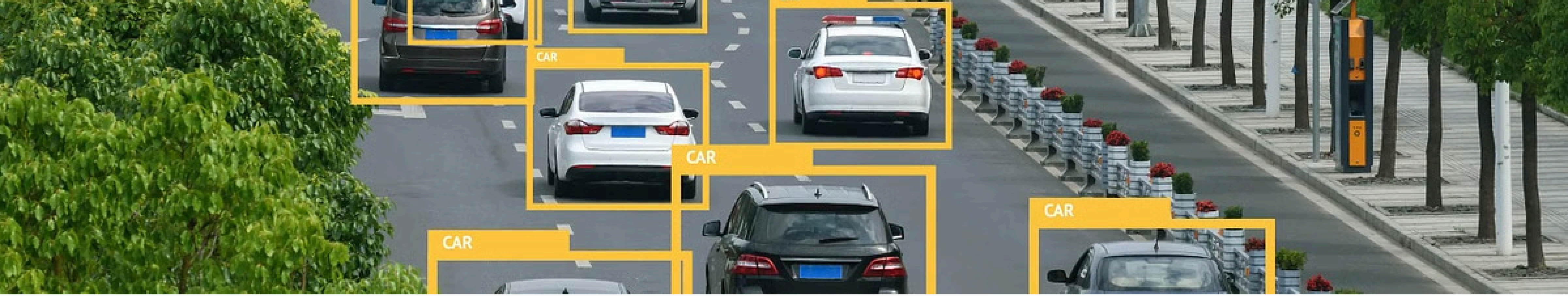
Carla & Vehicle Detection Datasets

Carla Dataset

- Open-source simulator for autonomous driving research, with realistic environments, traffic, and vehicle models.
- Image count: 1028 total images.
- 779 images are for training the mode, 249 for testing models performance.
- No. of classes : ['bike', 'motobike', 'person', 'traffic_light_green', 'traffic_light_orange', 'traffic_light_red', 'traffic_sign_30', 'traffic_sign_60', 'traffic_sign_90', 'vehicle']

Vehicle Detection Datasets

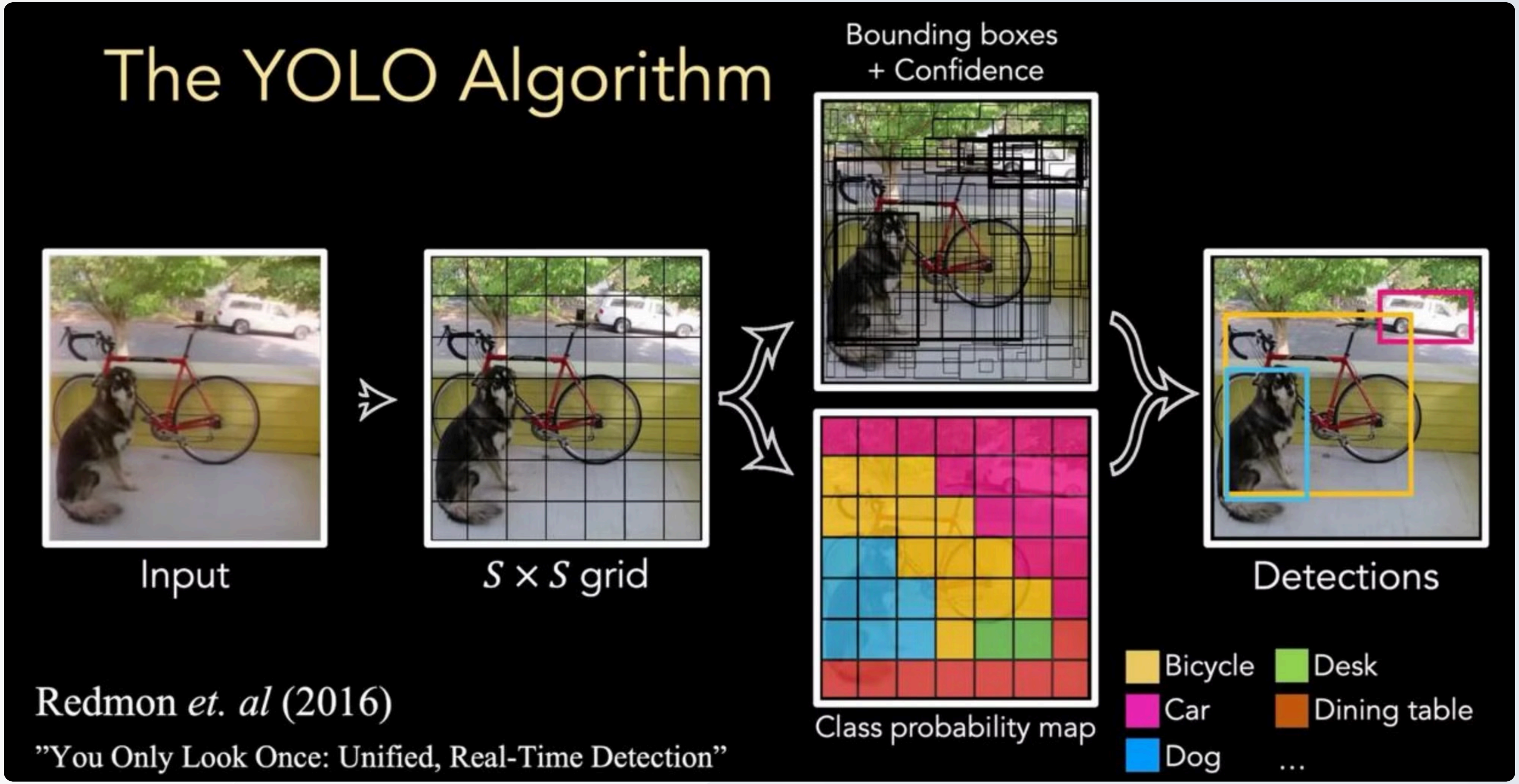
- Datasets used for training and evaluating computer vision models to detect vehicles in images and videos.
- Image count : 8219 images
- No. of classes : 8 classes
- (Motorcycle, Auto, car, Bus, LCV, Truck, Tractor Multi-Axle)



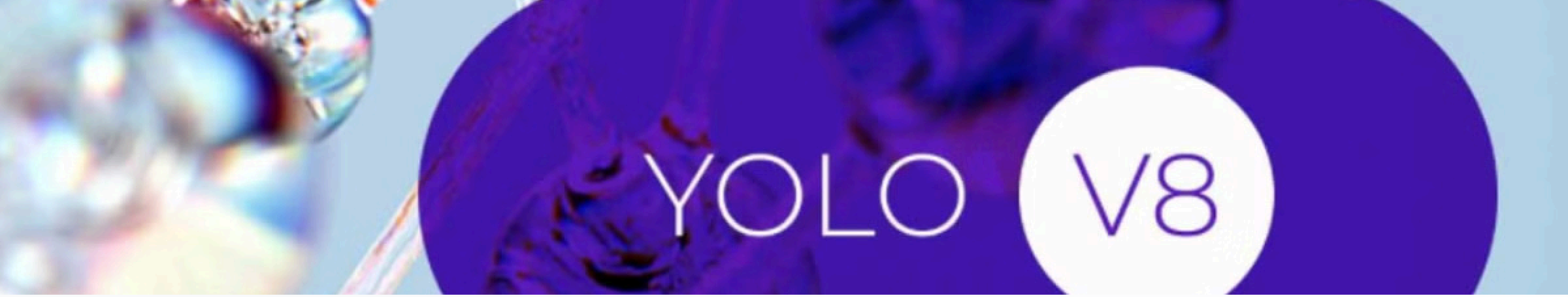
YOLO: YOU ONLY LOOK ONCE

YOLO (You Only Look Once) is a real-time object detection algorithm that processes images in a single pass through the network, rather than running the image through the network multiple times. It is highly efficient and can detect objects in images and videos with fast inference times.

YOLO divides an image into a grid and for each grid cell, it predicts bounding boxes and class probabilities for objects. This approach allows YOLO to be fast and accurate, making it popular in real-time applications like autonomous vehicles and surveillance systems.



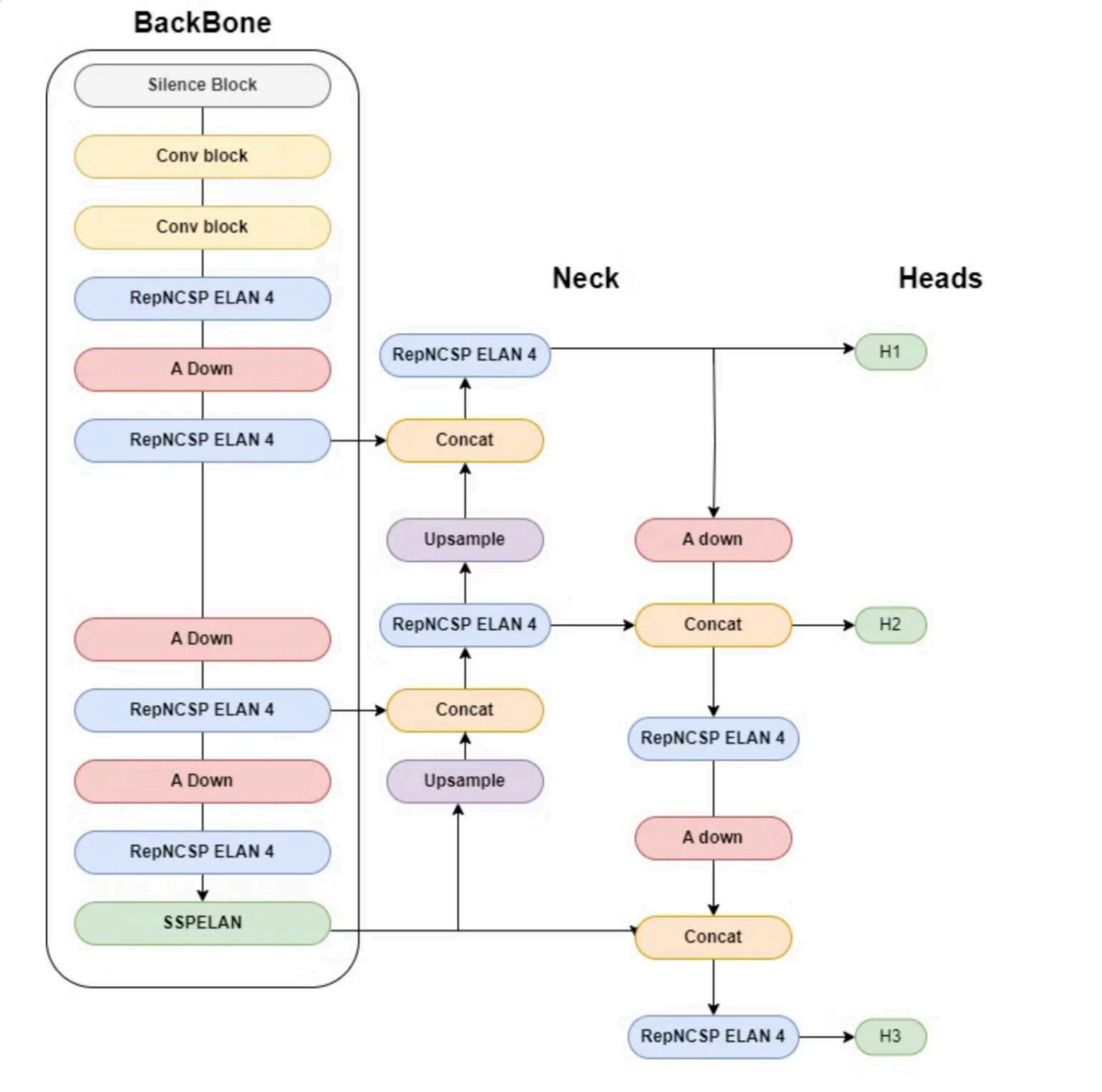
YOLO Working

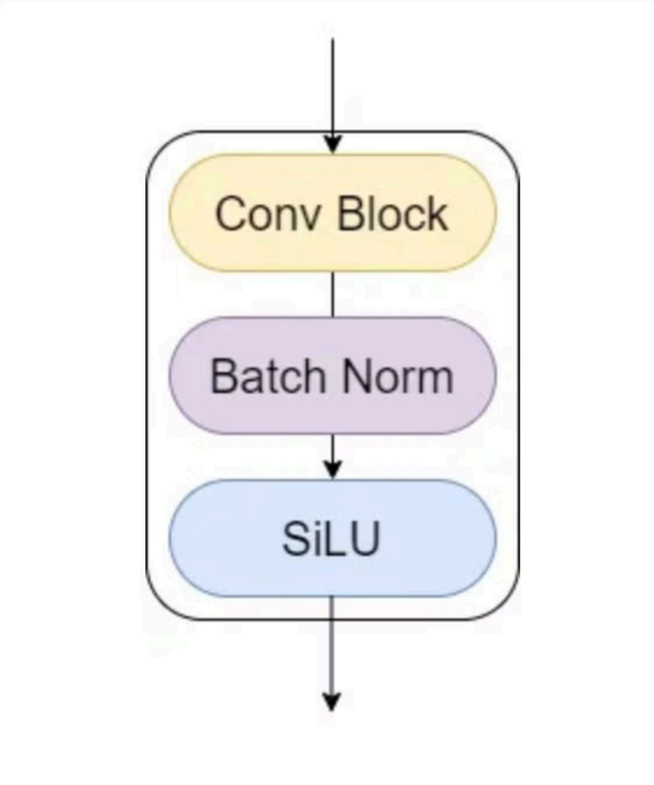


YOLO V8 -

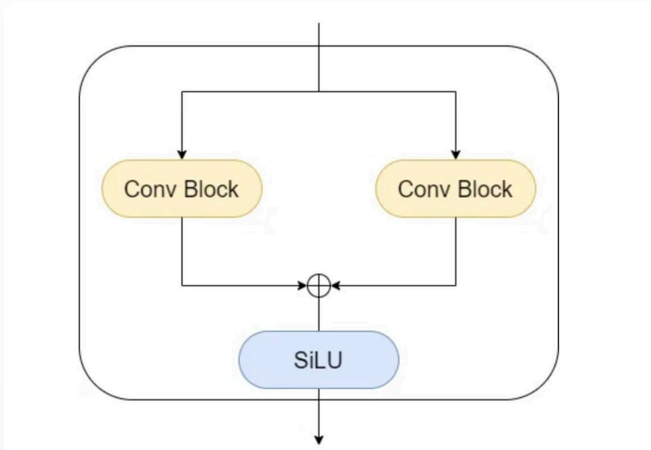
- YOLOv8 features a new backbone network with a new technique called cross stage partial (CSP) connections to improve the transmission of information and reduce information loss across the various levels of the network.
- The SPP (spatial pyramid pooling) block, contains convolution layers, and Maxpool layers to process the features of images of different sizes and where objects are at different scales within an image.
- The head of the network takes feature maps from Backbone of network.
- The new C2f block in head combines high-level features with contextual information to improve accuracy of detection.
- The head is created to be detachable, i.e it manages object scores, classification, and regression tasks in an independent manner, this allows each workflow to focus on its own task while improving the model's accuracy and precision.
- Finally, the Detection block uses a set of convolution and linear layers to map the high-dimensional features to the output bounding boxes and object classes.
- Detection block is anchor free which means it predicts directly the center of an object instead of the offset from a known anchor box. This Anchor-free detection reduces the number of bounding box predictions, which speeds up NMS during post-processing steps.

YOLO V9 ARCHITECTURE

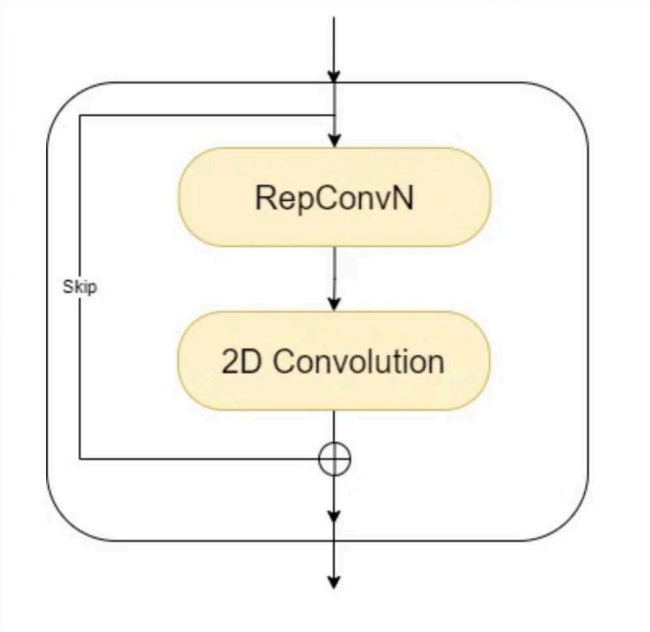




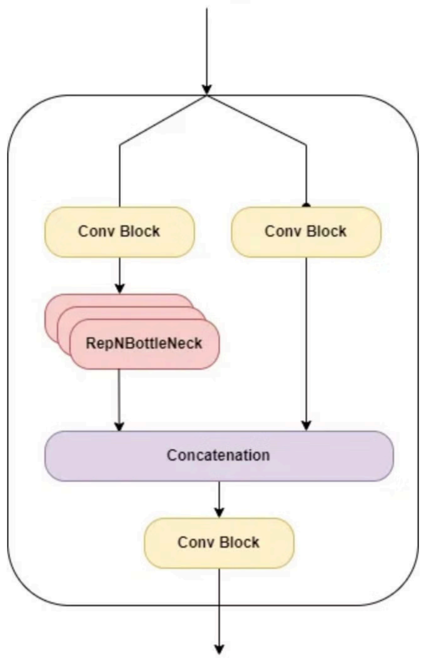
CNN Block



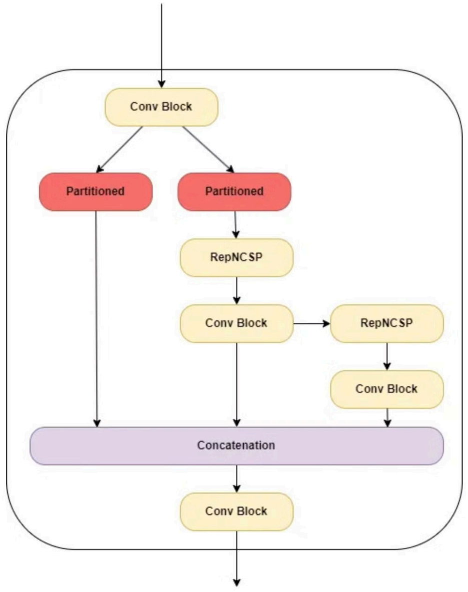
RepConvN Block



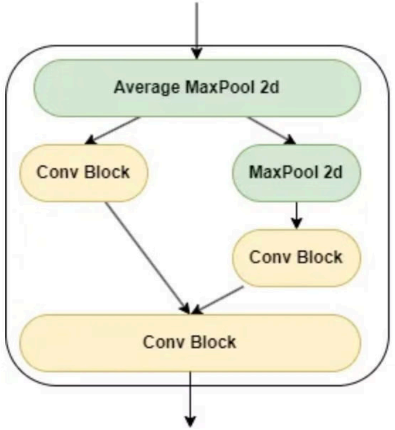
RepNBottleneck



RepNCSP Block



RepNCSP-ELAN 4

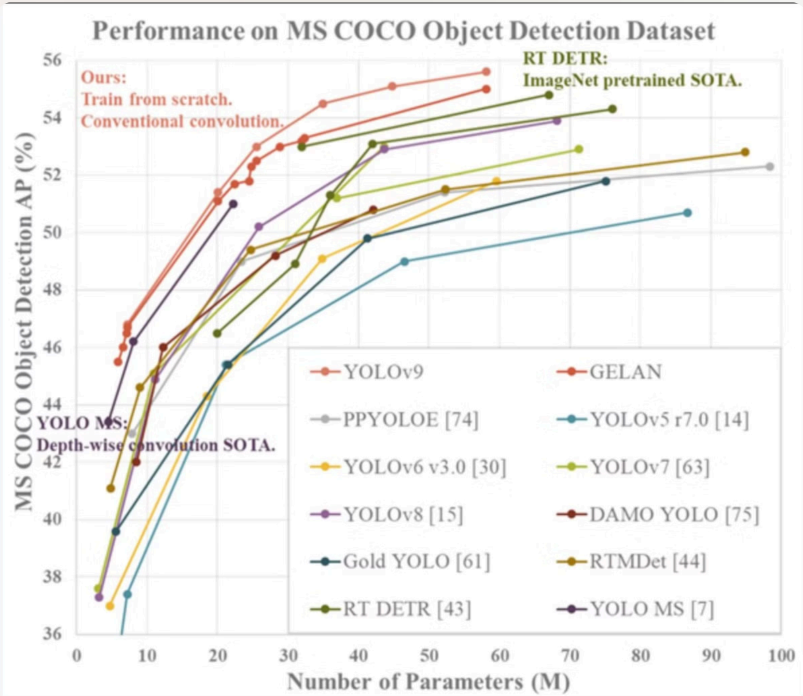


Adown

IMPROVEMENTS FROM YOLO V8 TO V9

- Architectural Enhancement
  - Programmable Gradient Information (PGI)
  - Generalized effficient layer aggregation network (GELAN)
- Improved Accuracy
  - Tomato disease dataset (92 → 93.6)
- Enhanced Feature Extraction
  - RepNCSP

AUGMENTED A.I praised that Yolo v8 for its speed , accuracy, when compared Yolo v9

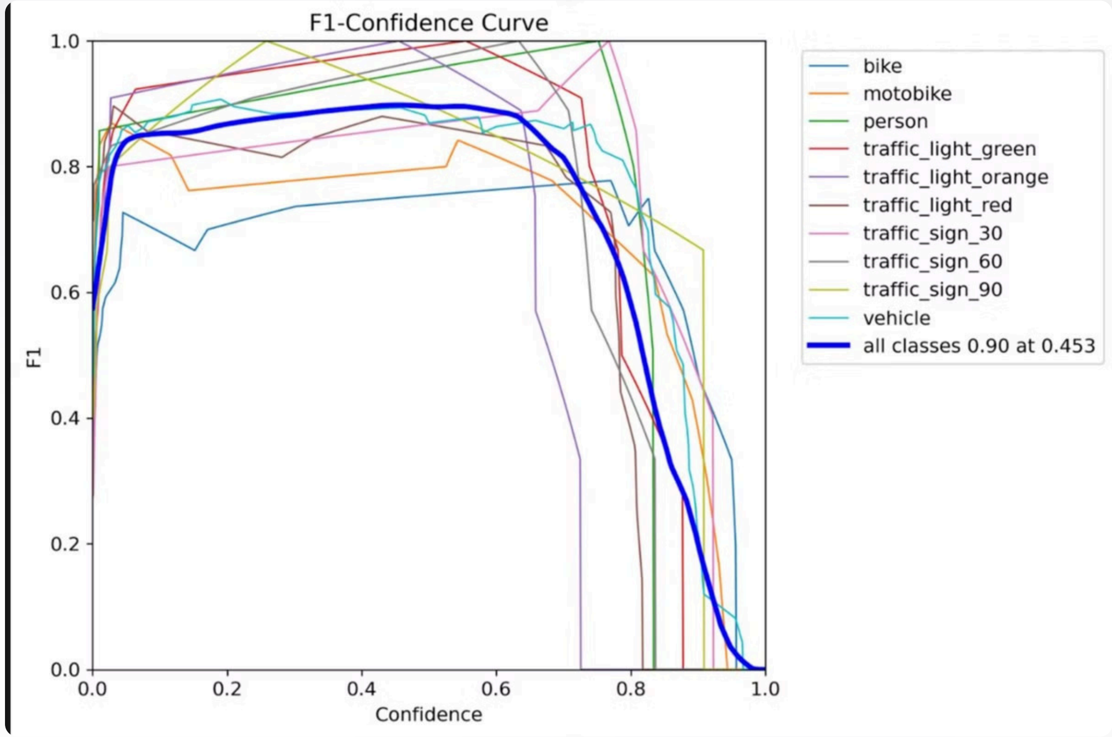


Evaluation Metrics

- mAP (Mean Average Precision): Measures detection accuracy across classes.
- Inference Time: Time taken to process an image.
- Training Time: Time for model optimization.

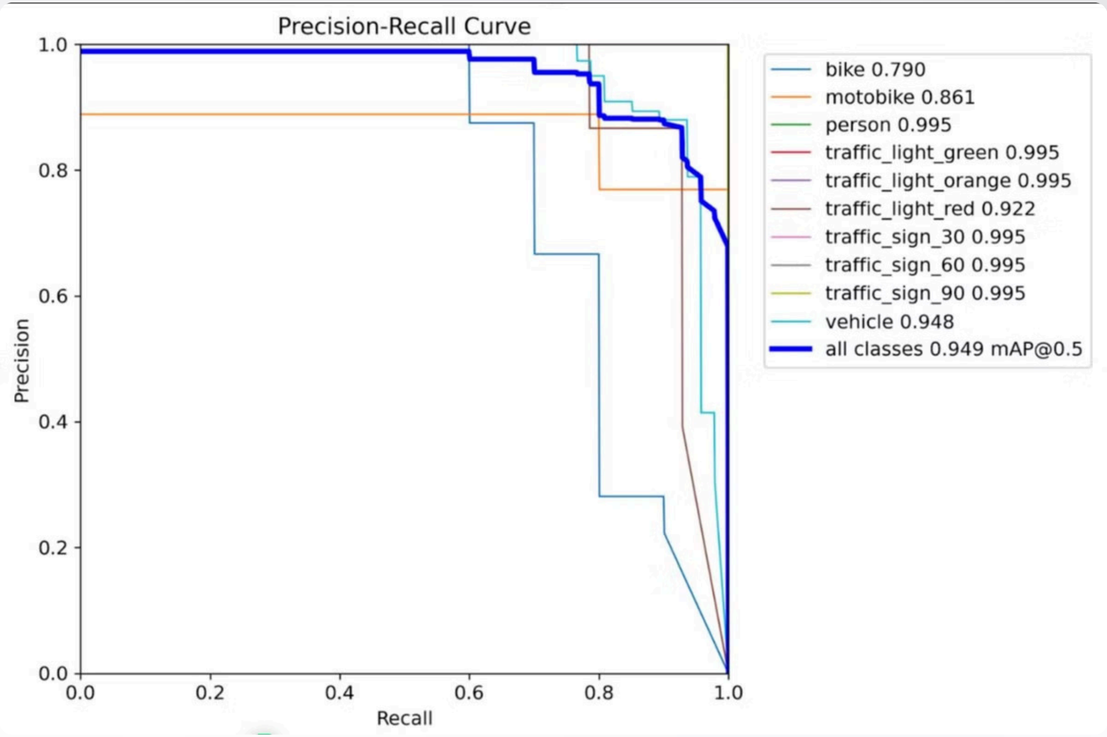
F1 CURVE :

Finds the sweet spot where the model is doing a good job at both being accurate and finding positives. Shows the best balance point for Precision and Recall.



PR CURVE:

Shows how well a model balances correct positive predictions (Precision) and finding all actual positives (Recall). Helps decide if the model is better at being accurate or finding more positives.



Made with Gamma

YOLOv8

METRIC	Carla	Real Dataset
mAP	0.949	0.937
Avg.Inference time	8.2 ms	8.7 ms
Training time	1.210 hours	2.494 hours

YOLOv9

METRIC	Carla	Real Dataset
mAP	0.957	0.829
Avg.Inference time	48.2 ms	48.9 ms
Training time	2.593 hours	3.494 hours

Comparision of YOLO V8 and V9 on both datasets

- The higher mAP scores indicate enhanced precision, which is particularly beneficial for complex object detection tasks.
- These characteristics make YOLOv8 an excellent choice for both industry and research applications where accuracy, speed, and efficiency are critical.

YOLOv8 outperforms its successor, YOLOv9, in key metrics like mAP and inference time. While YOLOv9 offers innovations that enhance detection accuracy, speed, and efficiency, its larger variants (e.g., YOLOv9c, YOLOv9e) have higher inference times than YOLOv8, making them better suited for complex tasks like medical imaging. Meanwhile, smaller YOLOv9 models (e.g., YOLOv9n, YOLOv9s) are optimized for lightweight deployment. Both YOLOv8 and YOLOv9 showcase advancements in object detection, offering flexibility for a range of applications, from low-power devices to high-end computing.

THANK YOU