VISION AI

OBJECT DETECTION APPLICATION ON YOLOv7 MODEL

MC-15

Project Guided by :Mr Sonu P

Team members:
Anuj M
(AM.SC.P2CSC21013)
Vaishnav Babu
(AM.SC.P2CSC21064)

Content:

- Abstract
- Yolo Algorithm:
- Introduction YOLO
- Working of YOLO
- Versions
- ARCHITECTURE
- Yolov7
- Conclusion
- Reference

Abstract:

Application that can be used by blind people in delicate and confusing situations such as identify obstacles in the street or look for an object in the house etc .This also gives blind people the ability to have artificial intelligence as a guide. Identifying and locating one or more efficient targets from still images or video data is the primary goal of object detection.

Introduction to YOLO :

- Yolo stands for you only look once
- Yolo is algorithm that uses neutral networks to provide real time object detection
- The main advantage of yolo is its speed and accuracy
- It process frames at the rate of 45fps to 150fps which is better than real-time
- The network is able to generalize the image better

Why Yolo Algorithm?

1. Face detection is one of the important task of object detection
2. Object detection algorithm based on deep learning can be classified into 2 categories.
   a. 2 stage detector – R-CNN
   b. 1 stage detector – YOLO
3. R-CNN can not be used in real-time
4. YOLO outperform R-CNN in terms of speed

Three important tasks undertaken by computer vision are

- **Classification**
- **Localization**
- **Detection**

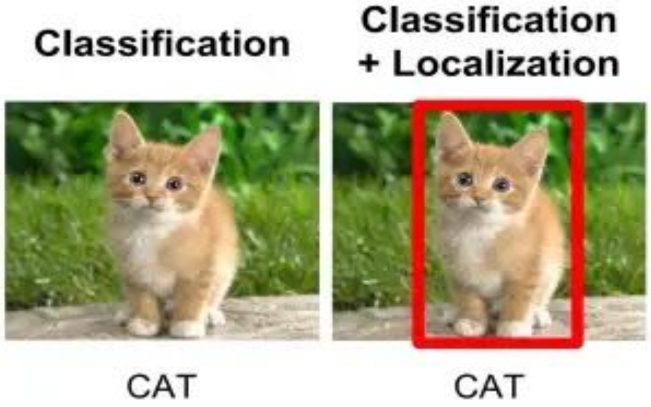
Classification

Classification is a machine learning task for determining which objects are in an image or video. It refers to training machine learning models with the intent of finding out which classes (objects) are present. Classification is useful at the yes-no level of deciding whether an image contains an object/anomaly or not.

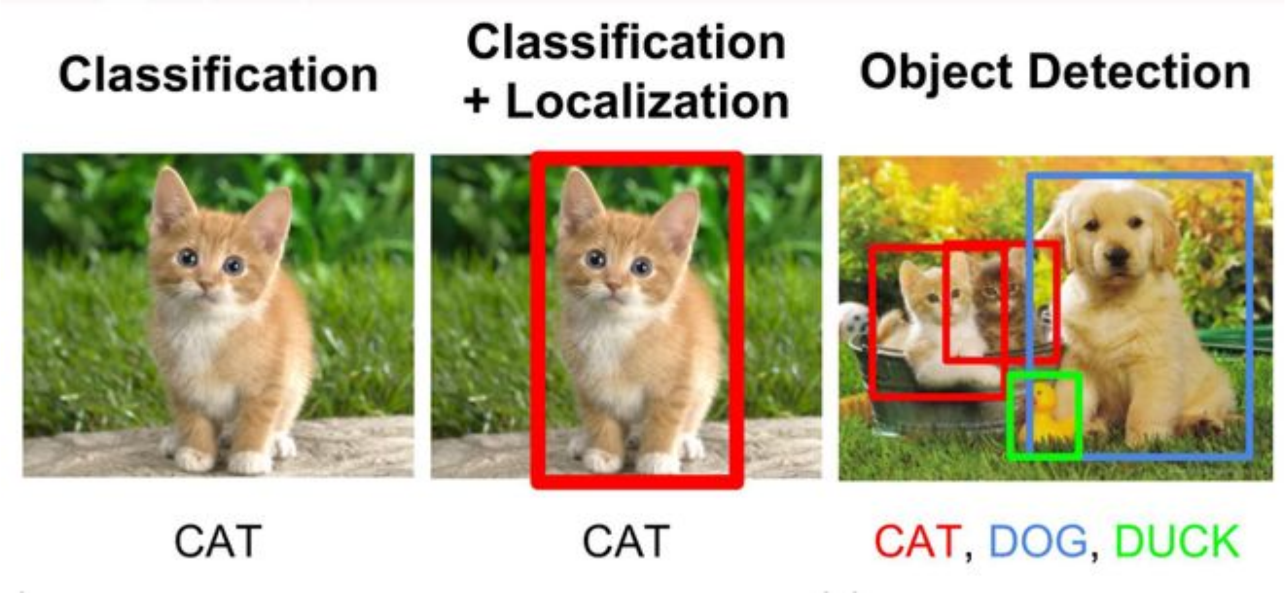
image classification

Localization:

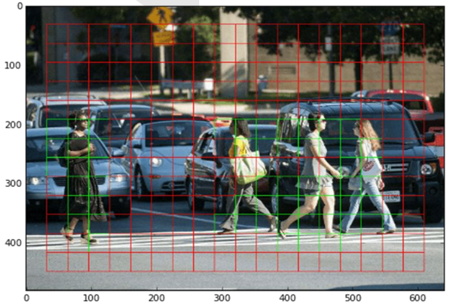
Object localization refers to **identifying the location of objects in an image and drawing abounding box around their extent**.

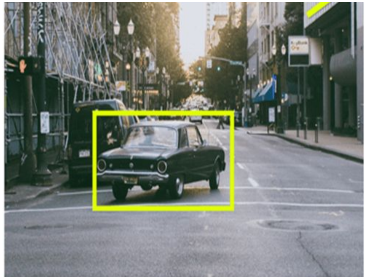


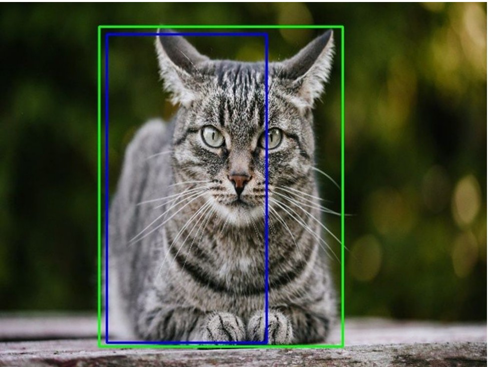
Detection : It is detecting multiple object in a single image



How Does Yolo Work?







Three techniques:

- **Residual blocks**:-Divide the image into grids of same dimensional size.Detection of object grid contains.
- **Bounding box**:-A bounding box is an outline that highlights an object in an image. Bounding box coordinates relative to grid cells.
- **Intersection Over Union (IOU)**:-IOU is used by YOLO to create an output box that properly encircles the items.This mechanism eliminates bounding boxes that are not equal to the real box.

Intersection Over Union

One of the evaluating factors which determines how well is the bounding box is predicted

It is calculated as : **IOU =Area of intersection/Area of union**

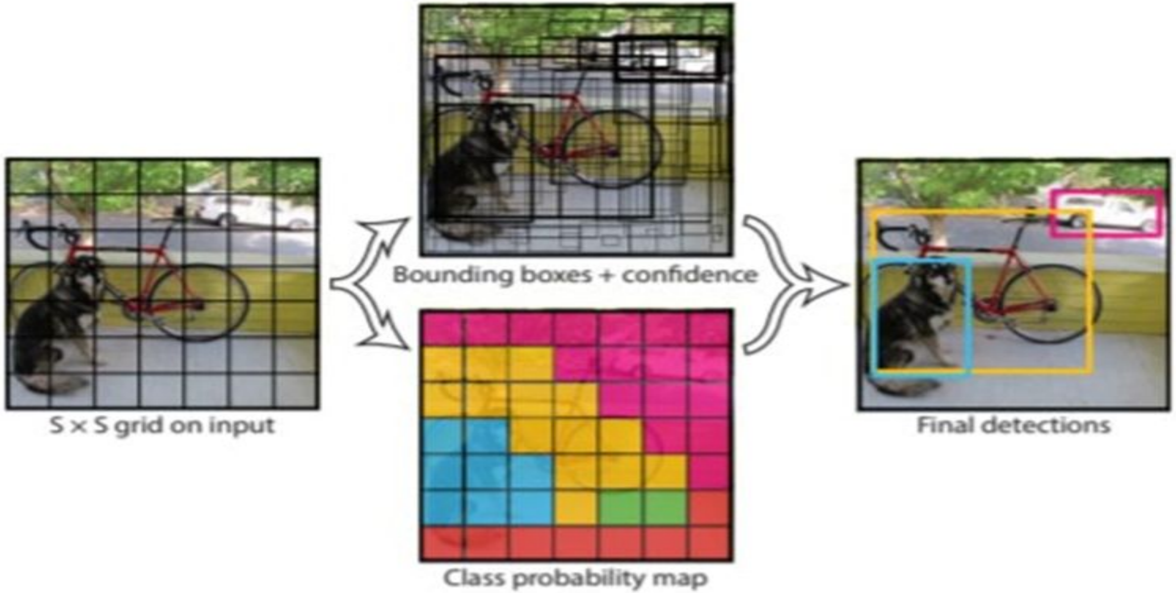
If the value of IOU >0.5 then it is considered as good prediction

If IOU <0.5 it is considered as bad prediction

Non Max Suppressor : A single image can have multiple boxes it takes the box with highest PC value

Combination of the three techniques:

The following shows how the three techniques are applied to produce the final detection results:-



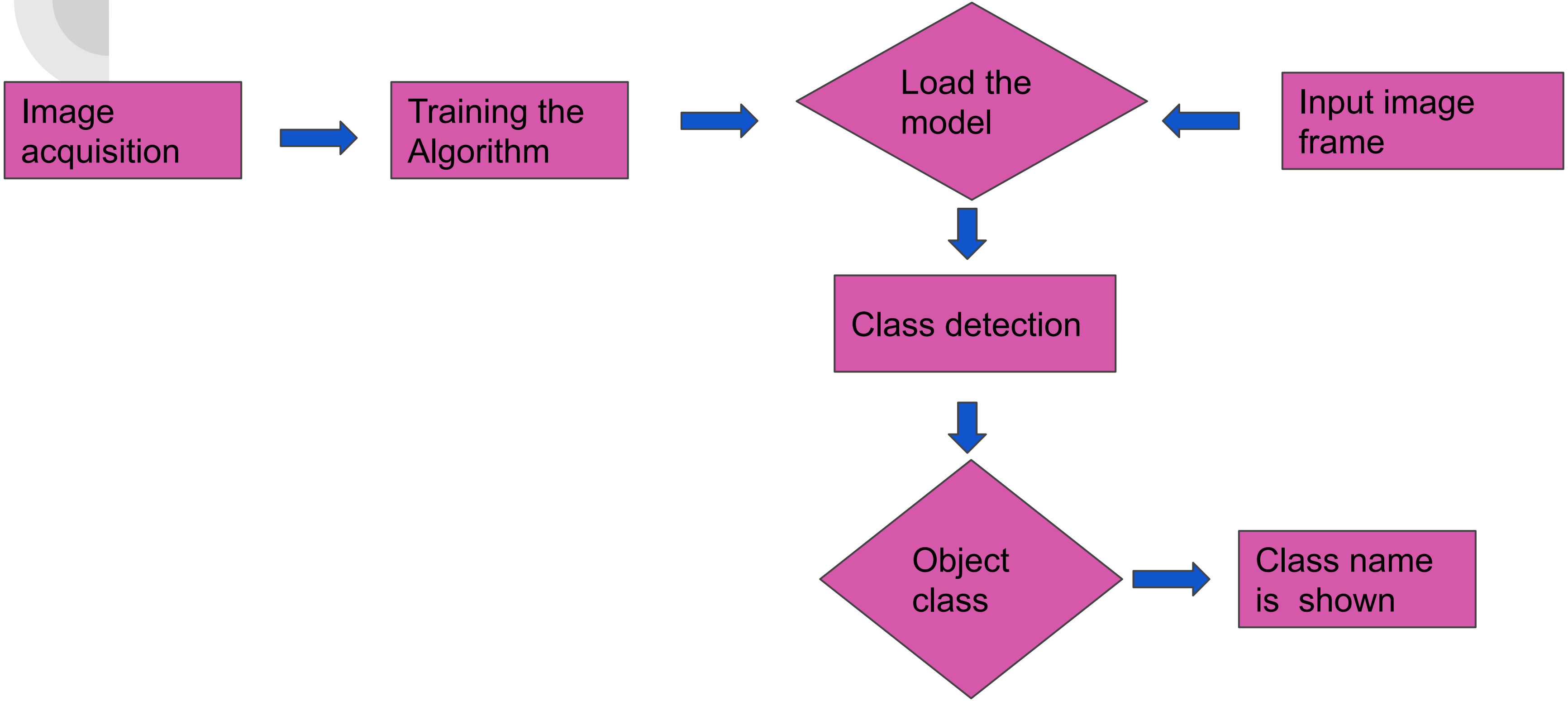


Image acquisition

Training the Algorithm

Load the model

Input image frame

Class detection

Object class

Class name is shown

Versions:

YOLO -- issues- accuracy detection of small objects in groups ie. Single object per grid cell, and localization errors.

YOLOv2 - multiple bounding boxes from a single cell, added batch normalization and higher resolution.

YOLOv3 - new network architecture ,added objectness score to bounding box prediction.

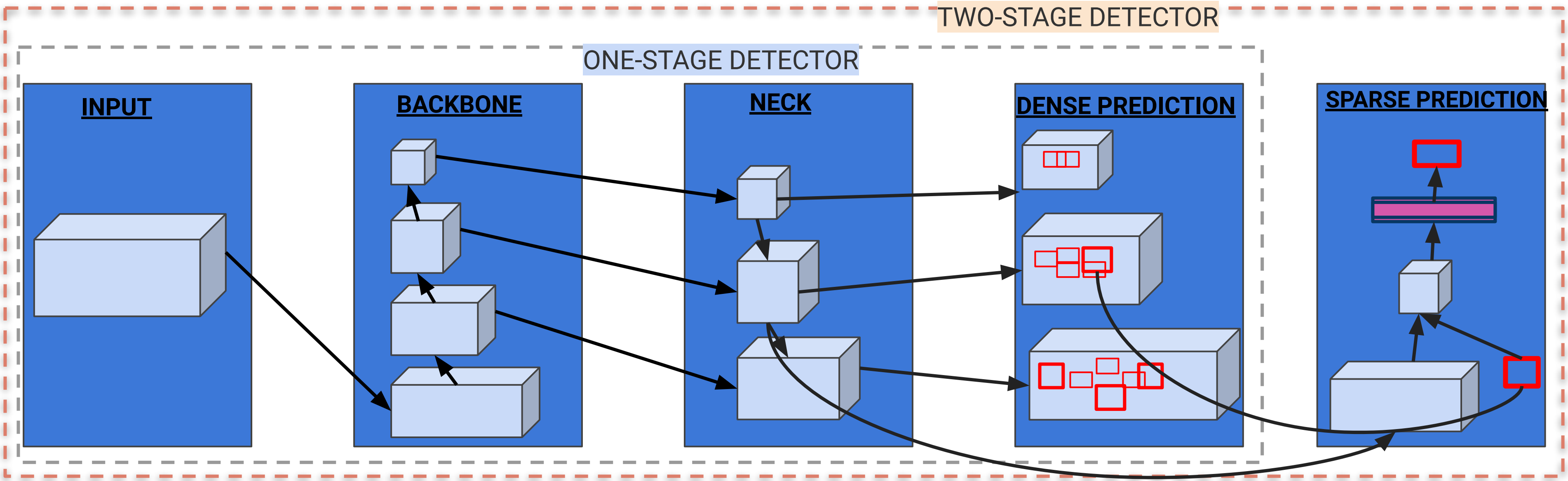
YOLOv4 - upgraded backbone,faster FPS and more accurate.

YOLOv5 - has multiple varieties of pre-trained models,detection with sufficient speed and accuracy

YOLOv6 - uses anchor-based methods for object detection,2-loss function

YOLOv7 - highest accuracy till date,trainable bag-of-freebies method

YOLO ARCHITECTURE



BACK BONE

- The back bone its a deep neural network composed mainly of convolutional layers.
- The main role/objective of the back bone is to extract the essential features the feature selection of the back bone.
- Feature selection of the back bone is the key step as it improves the performance of object detection
- Mostly pre-trained neural networks are used to train the back bone.

BACK BONE:

Layers	Yolov3	Yolov4	Yolov5	Yolov7
Backbone	Darknet	CSPDarknet	CSPVoVNET	E-LAN
Neck	FPN	SPP	PANet	FPN-RFB
Head	Bx(S+C) output	Same yolo v3	Same yolo v3	
Loss function	Binary cross entropy	Binary cross entropy	BCE and Logit Loss Funtion	BCE ,regression loss and categorical cross entropy

NECK

- Object detector models insert additional layers between backbone and head .
- Which are referred to copy of the detectors the essential role of the Neck is to collect feature maps from different stages .
- Usually a neck is composed of several bottom-up parts and several top-down part for enhancement we use FPN (Feature Pyramid Network),RFB(Receptive Field Block).
- FPN- is a feature extractor that takes a single-scale image of an arbitrary size as input, and outputs proportionally sized feature maps at multiple levels, in a fully convolutional.
- RFB-is a module for strengthening the deep features learned from lightweight CNN models so that they can contribute to fast and accurate detectors.

HEAD/DENSE Prediction

- Set the director to decouple the object localization and classification task for each module.
- once the detectors make the prediction for localization and classification at same time.
- This stage is only present in one stage detectors like - on self detection
  - YOLO
  - SSD
  - RPN
- Spares Prediction is for two stage detectors which does the class probabilities for the model input.
  - FRCNN

Training optimization

BAG OF FREEBIES:

Bag of freebies model refers to increase the model accuracy by making improvements without actually increasing the training cost.

Older version of yolo-v4 also use bag of freebies, some of the trainable bag of freebies used.

Data Augmentation: The process to increase the variability in the input images of the data, so that the designed object detection model has higher robustness to the images obtained from different environments.

Objective Function of BBox Regression: The objective function also called as loss functions are used to penalize and direct the model towards a better convergence at each training step.

Training optimization

Deep supervision:

Technique of using multiple heads -the main concepts is to add auxiliary head in middle layer with assistant loss as the guide.

Coarse -for Auxiliary heads: add multiple auxiliary heads in middle layer find the loss in the intermediate layer (this model is more stable)

Used Microsoft coco dataset to train the yolo v7 from scratch without using any data set or pre-trained weights.

Coco Dataset:

The COCO dataset contains challenging, high-quality visual datasets for computer vision.

COCO is often used to benchmark algorithms to compare the performance of real-time object detection.

The format of the COCO dataset is automatically interpreted by advanced neural network libraries.

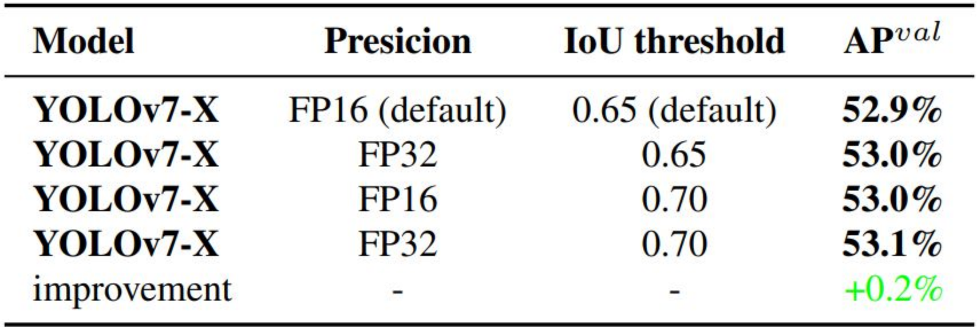
KEY Features:

- Object Segmentation
- Recognition in context
- 330K images(>200K labeled)
- 80 Object categories

Training optimization

During the research it was found that the average precision was higher when iou threshold was increased.iou-intersection over union-used to describe the extent of overlap of two boxes.

The greater the region of overlap the greater the iou,a value used in object detection to measure the overlap of a predicted versus actual bounding box for an object.



Model	**Presicion**	**IoU threshold**	**AP^{val}**
YOLOv7-X	FP16 (default)	0.65 (default)	**52.9%**
YOLOv7-X	FP32	0.65	**53.0%**
YOLOv7-X	FP16	0.70	**53.0%**
YOLOv7-X	FP32	0.70	**53.1%**
improvement	-	-	+0.2%

New in YOLOv7?

YOLOv7 improves speed and accuracy by introducing several architectural reforms.

- Architectural Reforms
    - E-ELAN (Extended Efficient Layer Aggregation Network):-Extended efficient layer aggregation networks. The proposed extended ELAN (E-ELAN) does not change the gradient transmission path of the original architecture.Increase the cardinality of the additional features by using group convolution, and mix the features of several groups in a shuffle and merge .

Layer Aggregation Networks:

The efficiency of the yolo network's convolutional layer is the backbone is essential to efficient interference speed.The shorter the gradient the more powerful the network will be to learn the final layer.

Aggregation they use E-ELAN is an Extended version of ELAN ie,Extended efficient layer aggregation networks.

E-ELAN considers designing an efficient network by controlling the shortest and the longest gradient path, so that deeper networks can converge and learn effectively.

E-ELAN uses group convolution to expand the channels and cardinality of the computational block

Model Scaling

All concatenation based models change the input width of some layers and the depth of those models as well these provide great support to the model in increasing the accuracy.

Scale up-Scale down - to meet our accuracy and speed requirements

Efficient way of doing this model scaling- compound model scaling.

In compound model scaling the width and depth are scaled in coherence for the concatenation model.

concatenation based model- width and depth are interrelated

MODEL RE-Parameterizing

Re-parameterization technique involves averaging a set of model weight to create a model that is more robust to the general pattern.

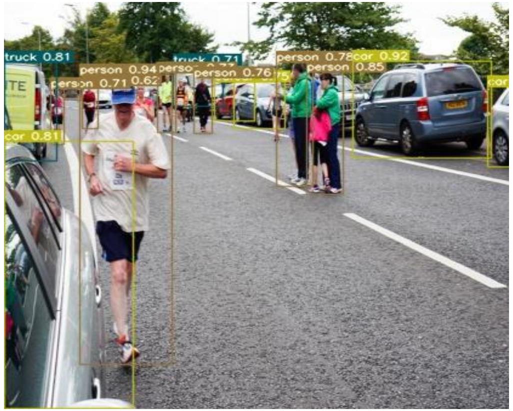
Model level and Module level Re-parameterization:

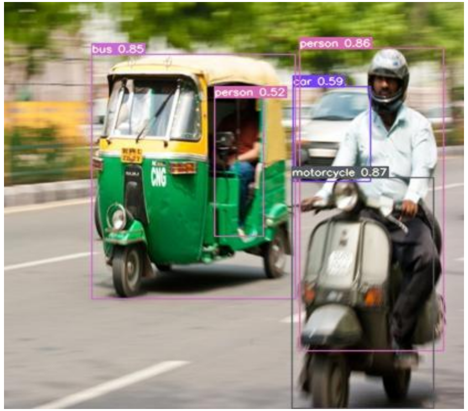
Model level: Train same model over different data set -get weights of different models get average and find the final model out of it

Module level : Split one Model into multiple modules and during inference they combine all modules into one single model

This helps in improving accuracy instead of depending on one single module, train multiple module.

Output images:





Conclusion:

Provided an overview of the YOLO algorithm and a summary of the YOLO technique and its application to object detection. Compared to previous object detection algorithms, this method yields better detection outcomes.

One stage detector architecture model of yolov7 overview of the paper.We can use this model for detection as well as tracking.The same can be used to count the objects in frames.Used the pre-trained yolov7 model and run test images, live webcam object detection.YOLOv7 does seem to be the best algorithm to use if you have an object detection problem to solve

References:

[1].Chien-Yao Wang1 , Alexey Bochkovskiy, and Hong-Yuan Mark Liao,YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors 2022.

[2]Muhammed Enes Atik, Zaide Duran , Roni Ozgunluk,Comparison of YOLO Versions for Object Detection,IJEGEO vol 9 Issue 2-june 2022.

[3].V.Neethidevan , Dr.S.Anand,A Real Time object Detection System Using a Webcam yolo algorithm- Annals of R.S.C.B vol.5, Issue 5,2021.

[4].Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun,-Towards Real-Time Object Detection, Annals of R.S.C.B vol 2017.

[5].YONGJUN LI , SHASHA LI , HAOHAO DU , LIJIA CHEN , DONGMING ZHANG , AND YAO LI-YOLO-ACN: Focusing on Small Target,Digital Object Identifier 10.1109/ACCESS.2020.3046515.

[6]https://viso.ai/computer-vision/coco-dataset/

THANK YOU.!