ISSN: 2581-4419 Volume 1 Issue 1

YOLOv3: Real-Time Object Detection

Atharva Mulgund[1], Angad Singh Obbi[2], Baliram Pansare[3], Tejas Phanse[4], Prof.Amarja Adgaonkar[4]
[1-3] Students
[4] Assistant Professor
Department of Information Technology
K.C. College of Engineering and Management Studies and Research, Thane
Mumbai, India

Abstract*− *More accurate and fast being the primary focus, this paper mainly focuses on object detection using the "you only look once" algorithm. YOLO is a Convolutional Neural Network (CNN) for performing object detection in real-time. CNNs are classifier-based systems that can process input images as structured arrays of data and identify patterns between them. Object detection is framed as a regression problem to spatially separated bounding boxes and associated class probabilities. The model processes images in real-time at 45 frames per second with an astounding 155 frames per second. YOLO has the advantage of being much faster than other networks and still maintains accuracy.

I. INTRODUCTION

Object detection is a computer vision technique for locating instances of objects in images or videos. Object detection algorithms typically leverage machine learning or deep learning to produce meaningful results. When humans look at images or video, we can recognize and locate objects of interest within a matter of moments. The goal of object detection is to replicate this intelligence using a computer.

It is widely used in computer vision tasks such as image annotation, vehicle counting, activity recognition, face detection, face recognition, video object co-segmentation. It is also used in tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, or tracking a person in a video.

Every object class has its own special features that helps in classifying the class – for example all circles are round. Object class detection uses these special features. For example, when looking for circles, objects that are at a particular distance from a point (i.e. the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed. A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin color and distance between eyes can be found.

Object detection is considered to be one of the most challenging tasks in the computer vision field. While there are a handful of different object detection algorithms, in this article, we will have a closer look at YOLOv3 (You Only Look Once).

Methods for object detection generally fall into either neural network-based or non-neural approaches. For non-neural approaches, it becomes necessary to first define features using one ofthe methods below, then using a technique such as support vector machine (SVM) to do the classification. On the other hand, neural techniques are able to do end-to-end object detection without specifically defining features, and are typically based on convolutional neural networks (CNN).

- Non-neural approaches:
    - o Viola–Jones object detection framework based on Haar features
    - o Scale-invariant feature transform (SIFT)
    - o Histogram of oriented gradients (HOG) features
- Neural network approaches:
    - o Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN, cascade R-CNN.)
    - o Single Shot MultiBox Detector (SSD)
    - o You Only Look Once (YOLO)
    - o Single-Shot Refinement Neural Network for Object Detection (RefineDet)
    - o Retina-Net
    - o Deformable convolutional networks.

II. LITERATURE SURVEY

"Object Detection Using Convolutional Neural Networks" published in TENCON 2018 - 2018 IEEE Region 10 Conference. This paper includes object detection using two models one with SSD with MobileNetV1 and another using Faster-RCNN with InceptionV2. we get to know from both models that faster-RCNN is slow and more accurate as compared to SSD with MobileNetV1. "Real-Time Deep Learning-Based Object Detection Framework" published in 2020 IEEE Symposium Series on Computational Intelligence (SSCI).In this paper, a nominated model based onthe YOLOV3-ResNet detection module in the Image Al deep learning library based on deep learning is proposed[2]. "Custom Face Recognition Using YOLO.V3" paper published in 2021 3rd International Conference on Signal Processing and Communication (ICPSC). In this paper,speed is taken as a constraining factor in face recognition and has implemented it using the YOLO.V3 algorithm which is a single shot algorithm that has a high processing speed compared with other algorithms. In this paper, face recognition using both R-CNN and YOLO.V3 algorithm is implemented[3]."A lightweight object detection algorithm based on YOLOv3 for vehicle and pedestrian detection" published in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC).The Lightweight-YOLOv3 has been proposed for vehicle and pedestrian detection. Channel and layer pruning is used in Lightweight-YOLOv3 to simplify the network architecture[4].

III. COMPONENTS LIST AND SPECIFICATION.

3.1 HARDWARE REQUIREMENTS

1. System: Laptop
2. Ram: 8 GB. (Minimum)
3. Minimum i5 7th gen

3.2 SOFTWARE REQUIREMENTS

1. Operating system: Windows 10.
2. Coding Language: Python 3.9.7
3. IDE: Visual Studio Code

IV. WORKING

1. YOLOv3

YOLO is a Convolutional Neural Network (CNN) for performing object detection in real-time. YOLO has the advantage of being much faster than other networks and still maintains accuracy.It allows the model to look at the whole image at test time, so its predictions are informed by the global context in the image. YOLO and other convolutional neural network algorithms “score” regions based on their similarities to predefined classes.

High-scoring regions are noted as positive detections of whatever class they most closely identify with. For example, in a live feed of traffic, YOLO can be used to detect different kinds of vehicles depending on which regions of the video score highly in comparison to predefined classes of vehicles.

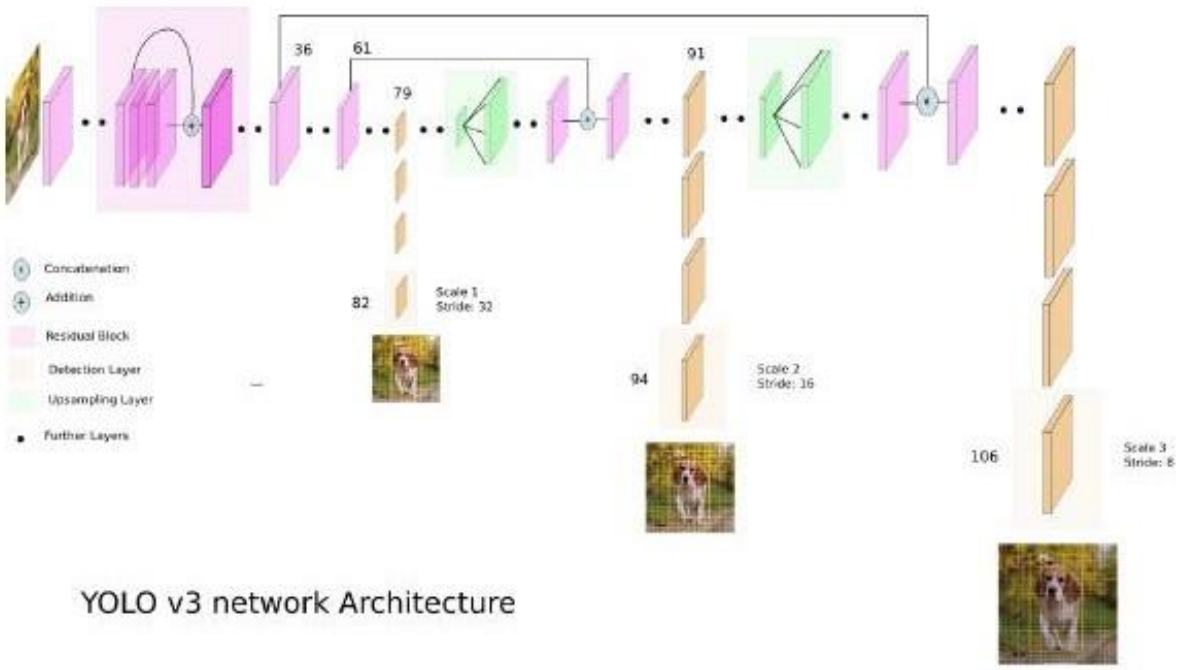

Figure 1: Architecture of YOLOv3

2. CNN

A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image isa binary representation of visual data. A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

Convolutional Neural Network a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other sorts and scans and identifies the images for different objects based on the trained data and gives out its decision on what the object is.

V. DESIGN SYSTEM

5.1BLOCK DIAGRAM

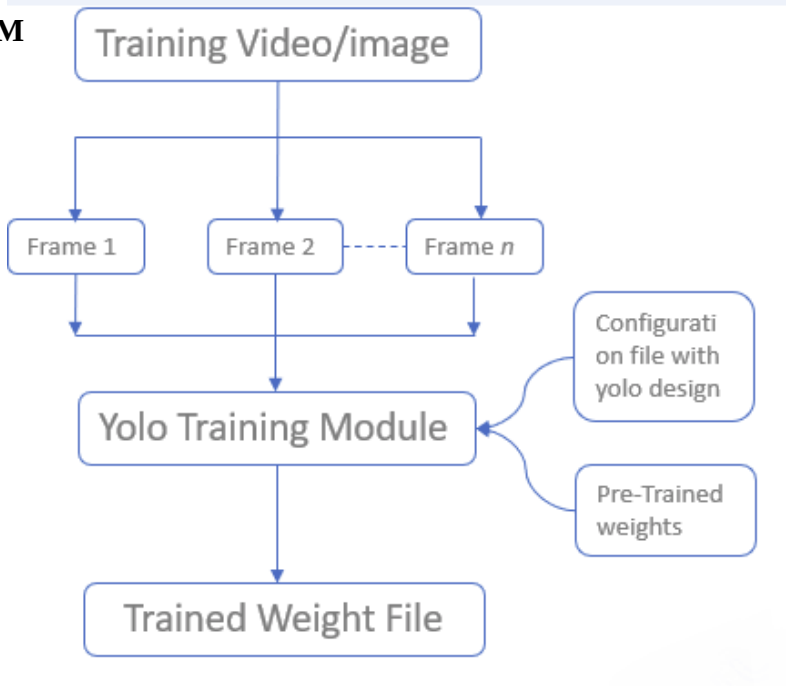

Figure 2: block diagram

5.2 USE CASE DIAGRAM

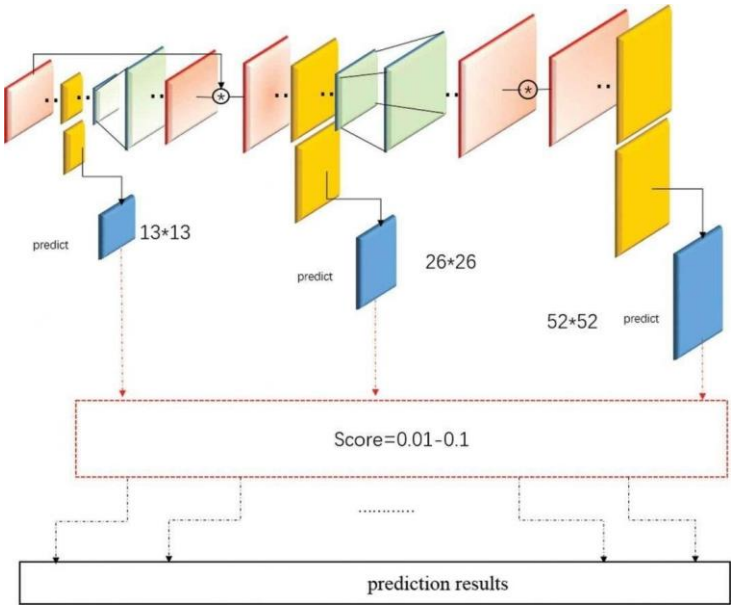

Figure 3: use case diagram

VI. RESULT AND CONCLUSION

Our project is able to detect the objects whether it be the image, video or real time camera access. Fast and accurate object detection being our main focus, we were able to achieve it using YOLOv3 with more accurate results.

Figure 4: object detection of the image

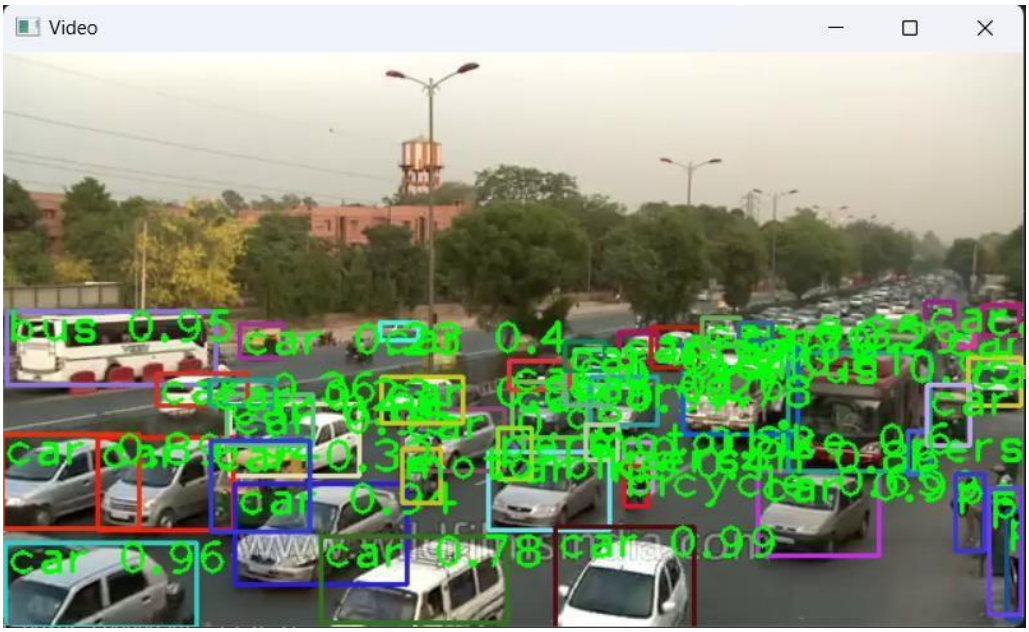

Figure 5: object detection in the video

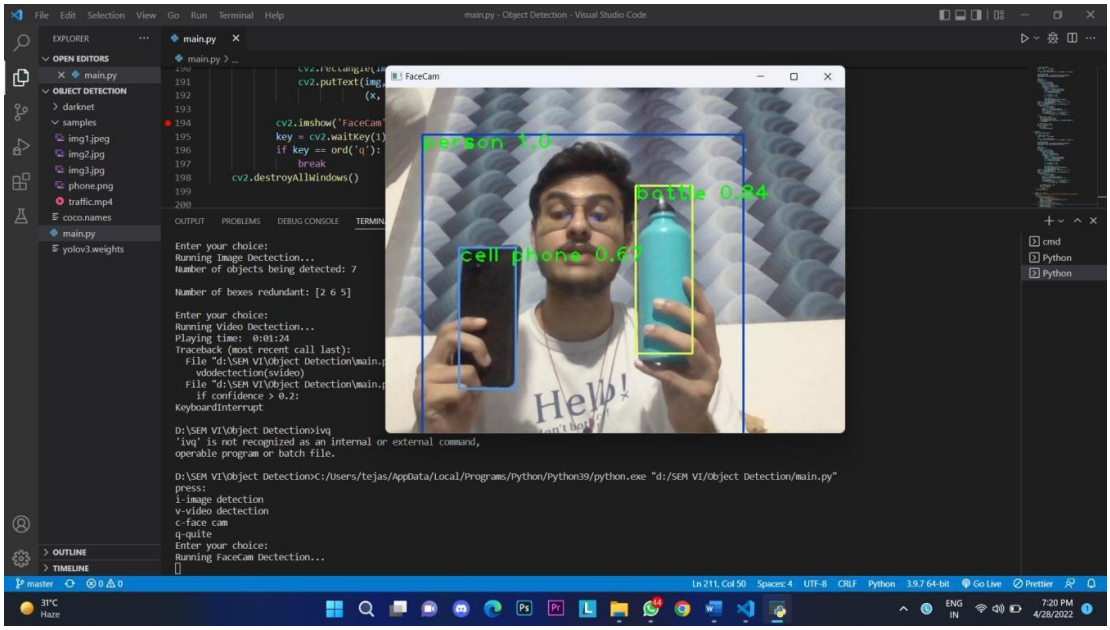

Figure 6: object detection with real time camera access

VII. FUTURE SCOPE

Object detection is increasingly important for computer vision applications in any industry. Object recognition is the core of most vision-based AI software and programs. Object detection plays an important role in scene understanding, which is popular in security, transportation, medical, and military use cases.

1. Object detection in Retail.
2. Autonomous Driving.
3. Animal detection in Agriculture.
4. People detection in Security.
5. Vehicle detection with AI in Transportation.
6. Medical feature detection in healthcare.

REFERENCES

[1] Reagan L. Galvez, Argel A. Bandala, Elmer P. Dadios, Ryan Rhay P. Vicerra, Jose Martin Z. Maningo. “Object Detection Using Convolutional Neural Networks”.TENCON 2018 - 2018 IEEE Region 10 Conference.doi: 10.1109/TENCON.2018.8650517

[2] William Tarimo, Moustafa M.Sabra, Shonan Hendre. “Real-Time Deep Learning-Based Object Detection Framework”. 2020 IEEE Symposium Series on Computational Intelligence (SSCI).doi:10.1109/SSCI47803.2020.9308493

[3] Suman Menon M, Anju Geroge, Aswathy N, Jaimy James. “Custom Face Recognition Using YOLO.V3”. 2021 3rd International Conference on Signal Processing and Communication (ICPSC). doi: 10.1109/ICSPC51351.2021.9451684

[4] Ning Zhang,Jiahao Fan. “A lightweight object detection algorithm based on YOLOv3 for vehicle and pedestrian detection”.2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). doi: 10.1109/IPEC51340.2021.9421214

[5] Maliha Khan,Sudeshna Chakraborty,Rani Astya,Shaveta Khepra.”Face Detection and Recognition Using OpenCV”. 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). doi: 10.1109/ICCCIS48478.2019.8974493

[6]YOLOv3Availabe:https://viso.ai/deep-learning/yolov3-overview