EasyChair Preprint

№ 9253

Object Detection Using Tensorflow and Its Methods

Vg Hemant, L Praveen, Gangula Surendar,
Karri Siva Somi Reddy, Abhishek Santra, Navjot Kaur and
Balwinder Kaur Dhaliwal

November 6, 2022

OBJECT DETECTION USING TENSORFLOW AND ITS METHODS

HEMANT VG[1], PRAVEEN L[2], GANGULA SURENDAR[3], KARRI SIVA SOMI REDDY[4], ABHISHEK SANTRA[5], Navjot Kaur[6], Dr. Balwinder Kaur Dhaliwal[7]

vghemant0101@gmail.com[1]

Student [1,2,3,4,5] Assistant Professor [6,7] Lovely Professional University India

Abstract

TensorFlow, an open-source program, is used in this study to implement object identification using several models. As we're all aware, object detection is a group of interconnected computer vision tasks that entails recognizing various things in a photograph, clip, or live broadcast. Both image classification and object identification employ object detection to identify objects in an image using multiple bounding boxes to accurately identify numerous objects in a video or image. Here, certain pre-trained models are used to recognize objects to make predictions or detect objects by determining whether the model is predictive. The different models used are Faster Region-Based Convolutional Neural Network (Faster-RCNN) is used to predict multiple objects from a digital image and with accuracy and the obtained accuracy of the model is 94%, Single Shot Detection (SSD) is used to predict the different objects from a video with accuracy and this model predicts the objects with an accuracy of 99.87%, and You Only Look Once (yolov5) is used to predicting the objects using Realtime i.e., using a webcam with accuracy this model performs the accuracy of 93%. The different accuracies that are obtained from the model are Predictable and it is working fine.

Keywords: TensorFlow, Faster Region-Based Convolutional Neural Network (Faster-RNN), Single Shot Detection (SSD), You Look Only Once (Yolo).

1. INTRODUCTION

In ancient times, hunting and protecting humans used only their bare eyes for their vision. But as time went on, technological development happens, and the threats also increased. So, humans started using digital cameras for surveillance and started fitting cameras in the car dashboard. But in this case, a human cannot surveillance for 24 hours a day with maximum accuracy. For this problem, technology gives us a solution: real-time object detection. Real-Time Object Detection is possible through Machine Learning (ML) which is the subset of Artificial Intelligence (AI). In modern society, along with the involvement of artificial intelligence, human work and working time as been reduced, and productivity has been increased. Artificial Intelligence robots, virtual or voice assistants, autonomous cars, image analysis software, speech and face recognition systems, drones, search engines, internet of things are the major examples of AI implementations. Online shopping, machine translation, web search, smart homes, cybersecurity, and digital voice assistants are some everyday life examples. There are some subsets in artificial intelligence like machine learning, deep learning, and robotics. This work is done using machine learning algorithms and models. For example, image recognition, speech recognition, traffic prediction, and medical diagnosis are some examples of projects done by machine learning. In this project Convolutional neural network (CNN) is implemented. Convolutional Neural Network (CNN), which is the subset of machine learning and the type of artificial neural network. CNN is a kind of network architecture that is most specifically used in projects which work upon pixel processing like image recognition. A CNN's architecture and human brain connectivity pattern are somehow like each other. We used Faster R-CNN (regions with convolutional neural networks) for object detection. Faster R-CNN is not like traditional CNN, it is a two-stage detection algorithm. While the Single shot

detector (SSD) is an object detection technique, which takes one shot of an image and detects the objects, meanwhile it is faster as well as more accurate. You Only Look Once (Yolo), has been especially used for real-time detection.

2. LITERATURE REVIEW

Faster RCNN is an object identification architecture proposed in 2015 by Ross Girshick, Shaoqing Ren, Kaiming He, and Jian Sun. It is one of the well-known object detection designs that employ convolutional neural networks, such as YOLO (You Look Only Once) and SSD (Single Shot Detector) [1]. The original Faster Region-based Convolutional Neural Network (Faster R-CNN) algorithm was tested on two convolutional network architectures: the Zeiler and Fergus model, which shares 5 convolutional layers with a Fast R-CNN network, and the VGG-16 (Simonyan and Zisserman) model, which shares 13 convolutional layers [2]. The ZF model is based on an older Convolutional Network model (made by Krizhevsky, Sutskever, and Hinton). This model has eight layers, five of which were convolutional, and the remaining three were fully connected [3]. The fundamentals of fast R-CNN and faster R-CNN are the same, but faster R-CNN is quicker and more accurate than fast R-CNN. Fast R-CNN takes an input image first, then processes the entire image with various layers of convolutional neural networks, such as max-pooling for feature maps, then the Region Pooling layer will extract the unwanted size of an image, and feed into the fully connected layers, which consist of two output layers, namely SoftMax and background class, for the n- the number of objects, finally for each predicted image a bounding box representation is obtained. [4]. Wei Liu et colleagues developed a new approach for detecting objects in photos that use a single deep neural network. This method was dubbed the Single Shot Multi-Box Detector SSD. SSD is a simple approach that requires an object proposal, according to the team, because it is based on the total removal of the process that creates a proposal. It also gets rid of the pixel and resampling steps. As a result, it condenses everything into a single phase. SSD is also simple to train and integrate. This facilitates detection. SSD's main characteristic is the use of multiscale convolutional bounding box outputs that are coupled to several feature maps [5]. Research is based on a sophisticated sort of SSD, Wong A. The authors of this study suggest that Tiny SSD, a one-shot detection deep convolutional neural network, be introduced. To make real-time embedded object identification easier, TINY SSD was created. It consists of significantly improved layers made up of a stack of non-uniform SSD-based auxiliary convolutional feature layers and a non-uniform Fire subnetwork. The size of Tiny SSD, which is even smaller than Tiny YOLO at 2.3 MB, is its strongest feature. The findings of this study demonstrated that Tiny SSD is a good choice for embedded detections. For the sake of comparison, a comparable SSD model was utilized [6]. An enhanced pedestrian identification system based on the SSD model of object detection has been presented by Fan et al. The Squeeze-and-Excitation model was added as a layer to the SSD model in this work's multi-layered system. With the use of self-learning, the system's accuracy for small-scale pedestrian identification was substantially improved. Research using the INRIA dataset revealed good accuracy. To comprehend the SSD model, this study was needed [7]. One of the fastest and easiest object detection algorithms, Single Shot Multi-Box Detector (SSD), was the attention of Chengcheng Ning et al. To identify the items in a picture, it just uses one convolutional neural network. Even while the SSD object detection technique is quick, there is still a significant disparity when comparing the two. The authors provide an improved method for the algorithm that will increase classification accuracy while having no impact on speed [8]. Wei Xiang et al. focus on single-shot detection (SSD) which is considered the newest algorithm for detecting an object. Thus, it is broadly observed that this SSD algorithm has a comparatively smaller level of accuracy required to distinguish between small and large objects. This is because it does not pay heed to the context

from out of the proposal boxes. The paper presents shorthand for single-shot multi-box detectors i.e., CSSD. Two variants of CSSD have been discussed in this paper. The outcomes of the example illustrate how multi-scale context modeling significantly enhances the precision in detection [9].

3. METHODOLOGIES

FASTER-RCNN

One of the most effective ways to recognize objects using the R-CNN series is the Faster Regional Based Convolutional Neural Network (Faster-RCNN). Since Ross Girshick et al. developed both the Fast-RCNN and Faster-RCNN, the Faster-RCNN is faster than the other R-CNN because it makes use of the Region Proposal Network. This method was developed by Ross Girshick et al. in 2014. It started with the Fast-RCNN and then progressed to the Faster-RCNN (RPN). A convolutional network's RPN feature builds boundary boxes around the items in the picture to forecast the objects and their accuracy. And for detecting several objects that Fast-RCNN utilized, pre-trained models use its high-quality areas.

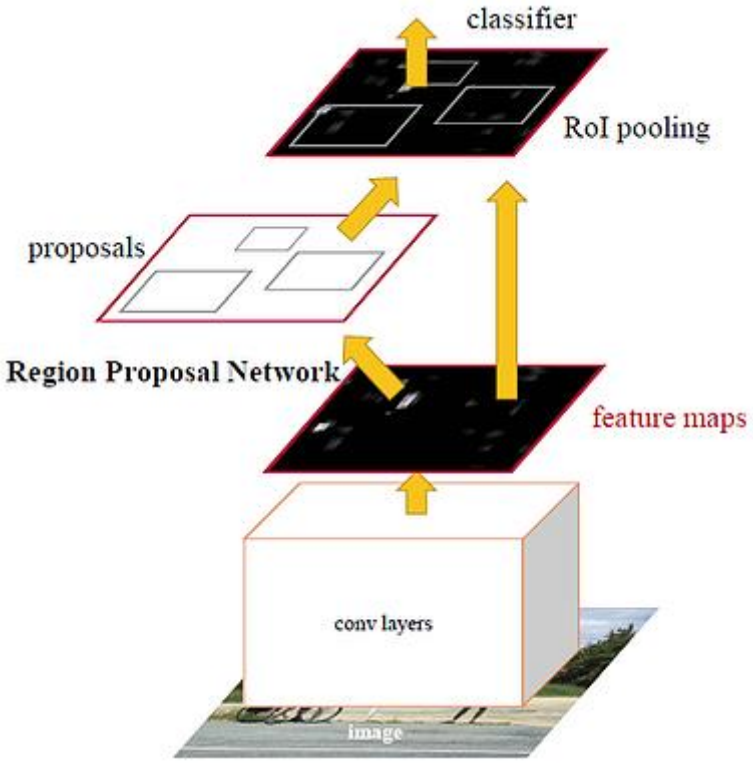

Fig1.Faster-RCNN Architecture. [12]

The first component of the F-RCNN is feature extraction, and a pre-trained Convolutional Neural Network (CNN) model is provided for the model. The F-RCNN subnetwork, however, may be trained. The Regional Proposal Network Serves as the Foundation for the Faster-Regional Based Convolutional Neural Network (F-RCNN) (RPN) It is taught to create top-notch region recommendations for object identification and is often utilized to detect the item by giving the appropriate real class. Height and width, which are hyper-parameters, will be used to build the feature maps via ROI Pooling. Because the ratios and sizes of the bounding boxes change, Anchor Boxes make it possible to identify pictures while also predicting the

positions of items in the RPN and the images. It is generated by using the height width and aspect ratio.

4. RESULT AND ANALYSIS

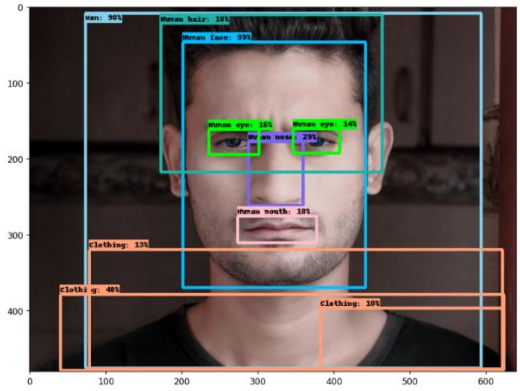

Fig2. Faster-RCNN Result. [1]

ACCURACY

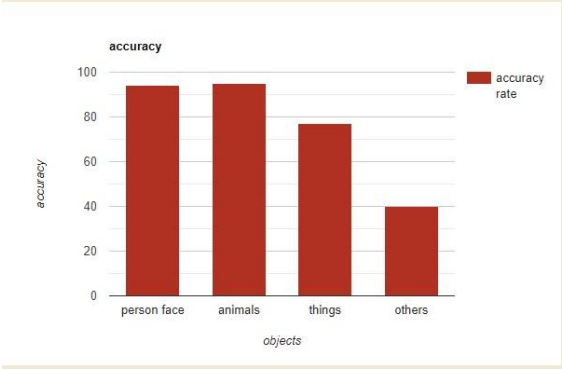

Fig3. Accuracy Representation of Faster-RCNN

SINGLE SHOT DETECTOR (SSD)

The implementation of a Single Shot Detector (SSD) for Real-time object recognition utilizing video representation for identifying the items in the movie using a Single Shot Detector (SSD), one of the most popular techniques for recognizing Real-time objects using a single deep neural network. SSD operates at 7 frames per second (FPS), which makes it quicker than Fast-RCNN and Faster-RCNN during training. SSD also generates good detection accuracy, especially for low-resolution objects. The default set of bounding boxes is generated by Single Shot Detector, and SSD generates various aspect ratios. To accommodate bigger and smaller objects, the Single Shot Detector (SSD) uses several feature map locations from the convolutional neural network to output distinct sizes. The Single Shot Detector primarily works with feature extraction and convolutional layers among the many components.

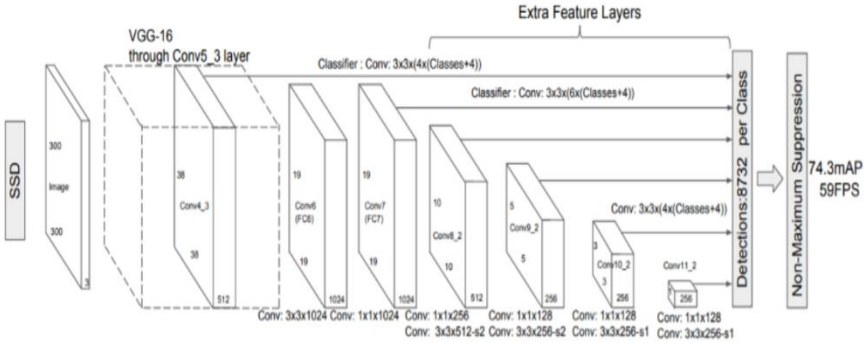

Fig4.Single Shot Detector Architecture. [13]

The SSD employs the VGG-16, a 16-layer deep neural network with several parameters, in this technique. Here, the convolutional layers of the VGG-26 are utilized to extract the feature maps before the system begins identifying objects. Bounding boxes and class scores are used to forecast the objects, and in the convolutional layer, four predictions are made for each cell of the feature map. Convolutional filters are used by the Single Shot Detector (SSD) to forecast the position and precision of the objects identified. Convolutional filters employ 3x3 for each cell to create predictions and provide related outputs, channels, and boundary boxes. The lower resolution of the feature map is used by the SSD to identify the object's various layers. As a result, the size of the item may vary from bigger to smaller. For larger objects, a 4x4 feature map is used for detection, and for smaller objects, an 8x8 feature map is used. Six more convolutional layers are added to the VGG-16 by the SSD, five of which are used for object identification and three of which are used for prediction. Using the six convolutional layers, Single Shot Detection (SSD) generates 8732 predictions.

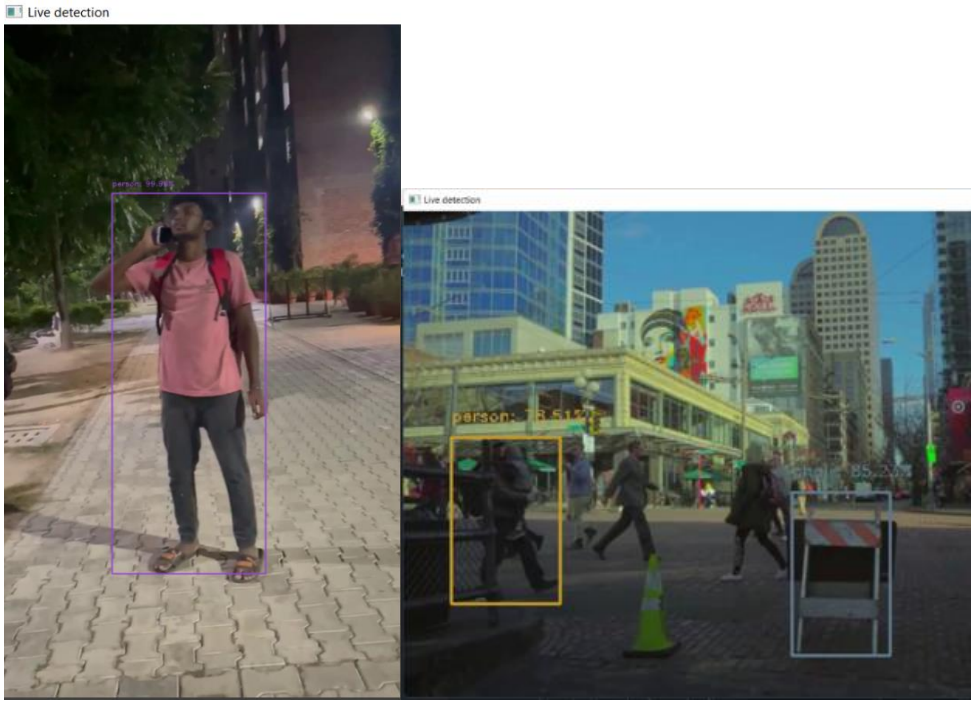

Fig5. Result Analysis of SSD

ACCURACY:

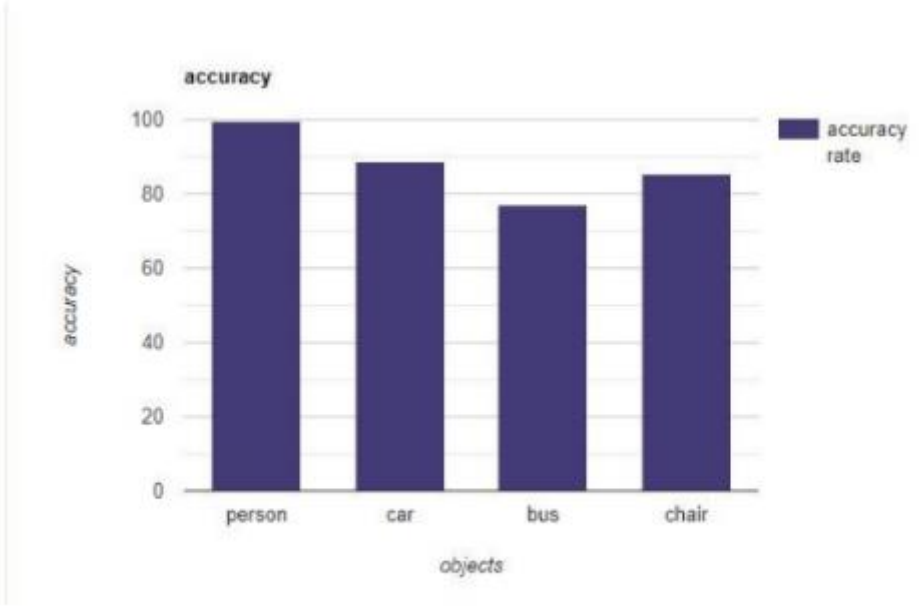

Fig6. Accuracy Representation of SSD.

YOU LOOK ONLY ONCE (YOLO)

The You Only Look Once (YOLO) algorithm is mostly used by researchers. And this algorithm is mainly used for real-time object detection using a webcam i.e., live detection of objects using a webcam. This algorithm was founded by Glenn Jocher who is popular for the implementation of PyTorch. The YOLOv5 has different models that are yolov5n (nano), yolov5s (small), yolov5m (medium), yolov5l (large), and yolov5xl (X large). The YOLO algorithm is trained on the COCO model, The YOLO can compute the real-time objection detection at 45 Frames Per Second (FPS) To 155 Frames Per Second (FPS). YOLOv5 is significantly 80% Smaller than YOLOv4, and it is 180% faster than YOLOv4, but the accuracy between the YOLOv4 and YOLOv5 slightly varies.

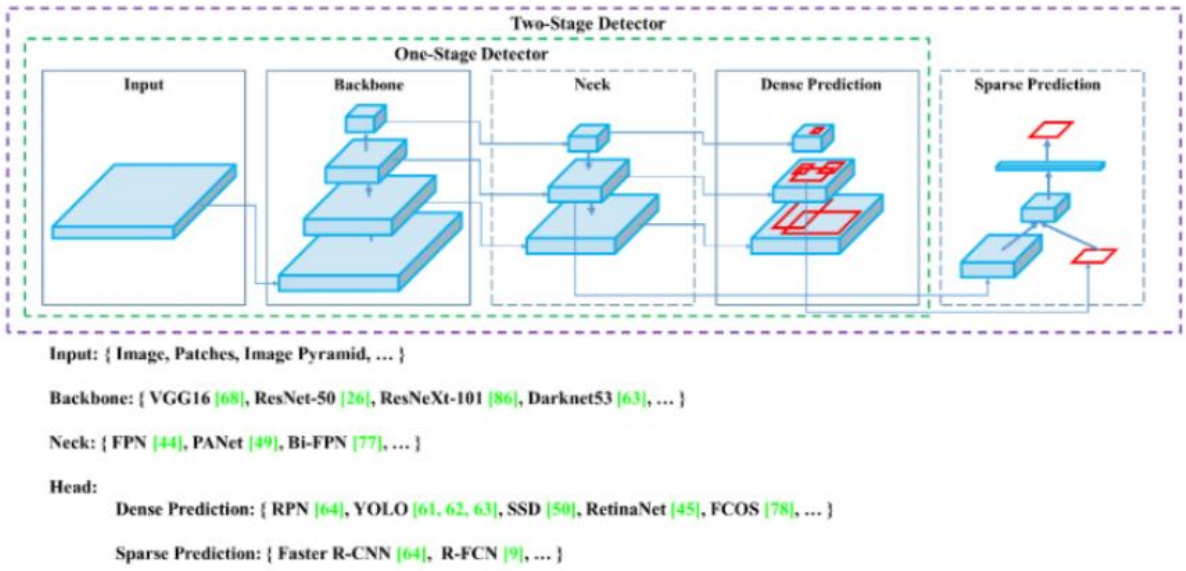

Fig7. Yolov5 Improved Architecture. [14]

1.

The YOLOv5 Architectural representations are shown in the image. Generally, it has the input, Backbone, head, ne, CK, and dense prediction, sparse prediction. The input is used to process the image for object detection and, the backbone in this model is used to pre-training the images and head to predict class and building boxes it can run on both CPUs and GPUs. In one stage predict, or it can be YOLO, SSD, or RETINANET for the dense prediction. And Faster-RCNN for the sparse predictions.

RESULT AND ANALYSIS :

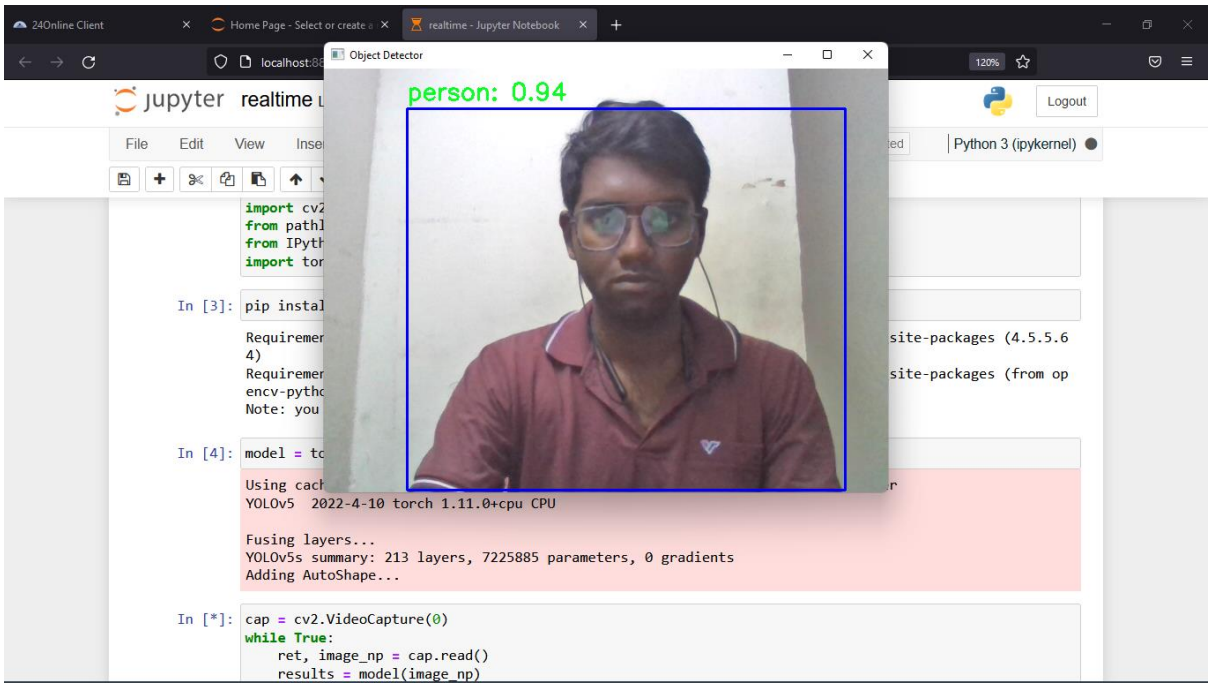

Fig8. Result Analysis of Yolov5.

ACCURACY:

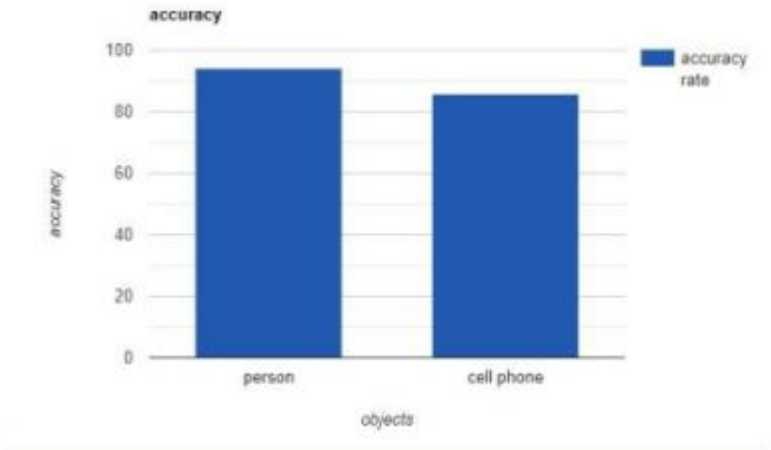

Fig9.Accuracy Representation of Yolov5.

5. CONCLUSION

Due to these demands, the speed, different scales, limited data, and class imbalance—object identification is typically thought to be far more difficult than picture categorization.

Researchers have worked hard to overcome these obstacles, frequently with astonishing outcomes; yet substantial obstacles still exist.
In general, tiny items continue to provide a challenge for all object identification systems, particularly when they are grouped together and partially occluded. Real-time detection with high classification and localization precision is still difficult to achieve, and practitioners frequently must choose between the two when making design decisions. If some continuity between frames is expected in the future rather than processing each frame separately, video tracking could get better. Additionally, extending the existing two-dimensional bounding boxes into three-dimensional bounding cubes would be an intriguing improvement that may receive additional research. Even while many object detection challenges have been overcome in novel ways, these additional factors and a great deal more indicate that object detection research is far from complete. And the various models that we have employed have a great degree of accuracy in predicting the items.

REFERENCES

[1] https://tfhub.dev/google/faster_rcnn/openimages_v4/inception_resnet_v2/1
[2] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097–105.
[3] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Semantic segmentation with second-order pooling. In ECCV,2012.
[4] R. Caruana. Multitask learning. Machine learning, 28(1),1997.
[5] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman.Return of the devil in the details: Delving deep into convolutional nets. In BMVC, 2014
[6] Ding S, Zhao K. Research on daily object detection based on deep neural network. IOP Conf Ser Mater Sci Eng. 2018;322(6):062024.
[7] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: single shot multibox detector. In: European conference on computer vision. Cham: Springer; 2016, p. 21–37.
[8] Wongomg A, Shafiee MJ, Li F, Cheryl B. Tiny SSD: a tiny single shot detection deep convolutional neural network for real-time embedded object detection. In: 2018 15th conference on computer and robot vision (CRV). IEEE; 2018, p. 95101.
[9] Fan D, Liu D, Chi W, Liu X, Li Y. Improved SSD-based multi-scale pedestrian detection algorithm. In: Advances in 3D image and graphics representation, analysis, computing, and information technology. Springer, Singapore; 2020, p. 109–118.
[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed (2015) SSD: Single shot multibox detector.
[11] Wei Xiang Dong-Qing Zhang Heather Yu Vassilis Athitsos (2018) Context-AwareSingle-ShotDetector. 2018 IEEE Winter Conference on Applications of Computer Vision, pp. 1784-1789.
[12]https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e.
[13] https://medium.com/towards-data-science/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab.
[14]https://medium.com/analytics-vidhya/object-detection-algorithm-yolo-v5-architecture-89e0a35472ef.