Survey On Real Time Multiple Object Detection Using MobileNet-SSD with Opencv

Prof. Disha Deotale
Computer department
G H Raisoni Institute of Engineering and Technology , Pune
Savitribai Phule Pune University
Maharashtra, India

Irfani Zaid Mohammad Kaif
Computer department
G H Raisoni Institute of Engineering and Technology , Pune
Savitribai Phule Pune University
Maharashtra, India

Subodh Joshi
Computer department
G H Raisoni Institute of Engineering and Technology , Pune
Savitribai Phule Pune University
Maharashtra, India

Pranav Jeurkar
Computer department
G H Raisoni Institute of Engineering and Technology , Pune
Savitribai Phule Pune University
Maharashtra, India

Ameya Bedre
Computer department
G H Raisoni Institute of Engineering and Technology , Pune
Savitribai Phule Pune University
Maharashtra, India

ABSTRACT:

The field of computer vision known as real-time object detection is vast, dynamic, and difficult. Real-time object recognition and tracking are critical and difficult problems in many computer vision applications, including video surveillance, robot navigation, and vehicle navigation. Detecting an item in a video sequence is what object detection is all about. Every tracking technique necessitates an object detection mechanism, either in each frame or whenever an item appears for the first time in the video sequence.

The practise of identifying one object or numerous objects using a static or dynamic camera is known as object tracking. The availability of powerful computers, high-quality, low-cost video cameras will enhance the need for automated video analysis. Image Localization is used when there is only one object to distinguish in an image, and Object Detection is used when there are multiple objects in an image. The most often used strategies for contemporary deep learning models to perform various tasks on embedded devices are mobile networks and binary neural networks. In this research, we propose a method for distinguishing an item based on the MobileNet deep learning pre-prepared model for Single Shot Multi-Box Detector (SSD). This technique is used to recognise objects in a video stream in real time and for webcam broadcasting. Following that, we use an object detection module to determine what is in the video stream. To complete the module, we combine the MobileNet and SSD frameworks to create a fast and efficient deep learning-based item detection technique.

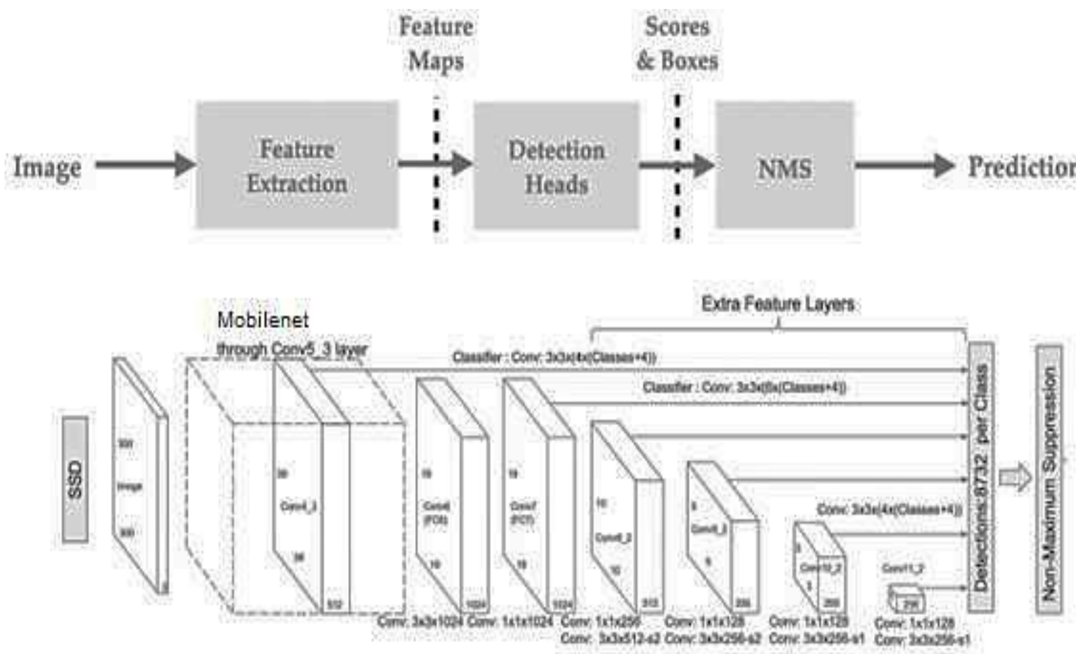
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

1. INTRODUCTION:

Object detection is now one of the most significant areas of research in computer vision. It is an image classification technique whose goal is to recognise one or more kinds of items in a picture and pinpoint their existence using bounding boxes. As a result, object detection is critical in many real-world applications such as picture recovery and video monitoring. The major goal of our investigation is to expound on the accuracy of an object detection methodology SSD and the pre-trained deep learning model MobileNet, as well as to highlight some of the important features that distinguish this method.

The experiment results reveal that the algorithm's Average Precision (AP) for recognising various classes such as car, human, and chair is 99.76 percent, 97.76 percent, and 71.07 percent, respectively. This enhances the accuracy of behaviour recognition at a handling speed required for real-time location and day-to-day watching indoors and outdoors. One of the focal points of our work is the incorporation of MobileNet into the SSD system. However, MobileNet with the efficient SSD structure has been a hot investigation topic in recent years, owing to the practical limitations of running robust neural networks on low-end devices such as mobile phones, laptops to broaden the range of possible outcomes for real-time applications.

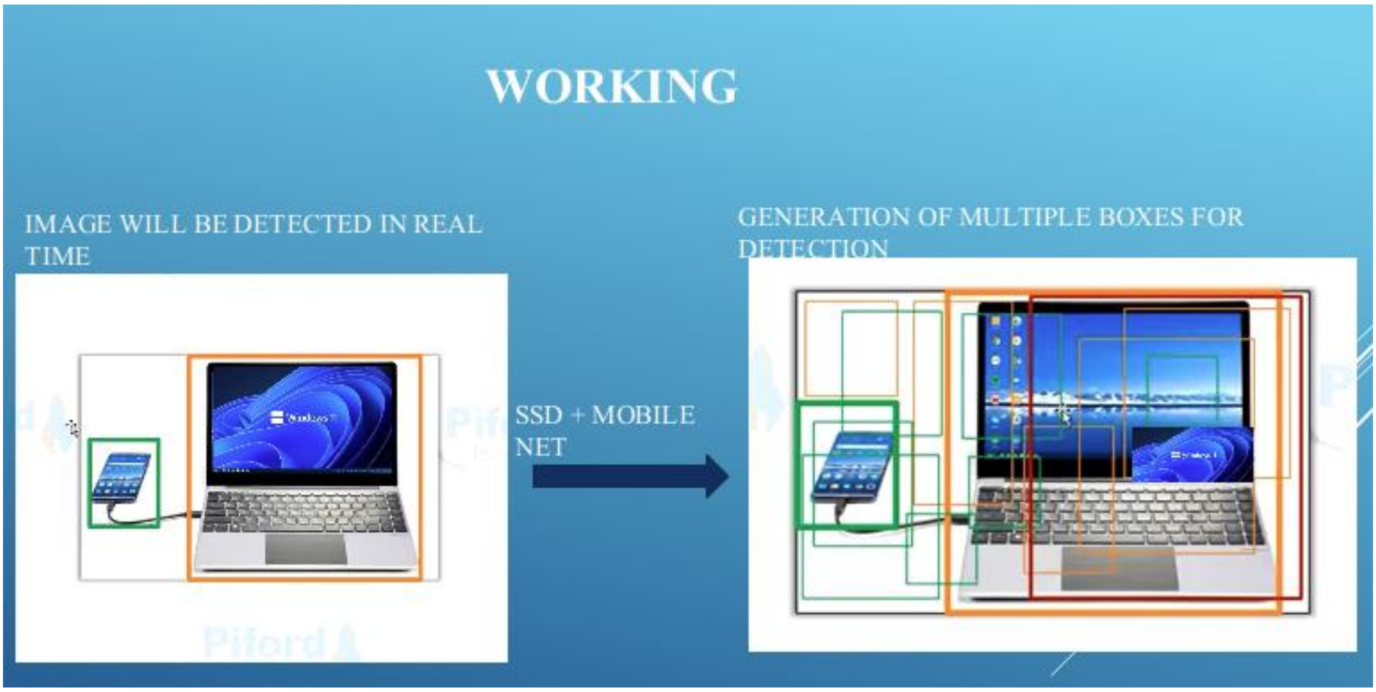
1.1 SYSTEM ARCHITECTURE

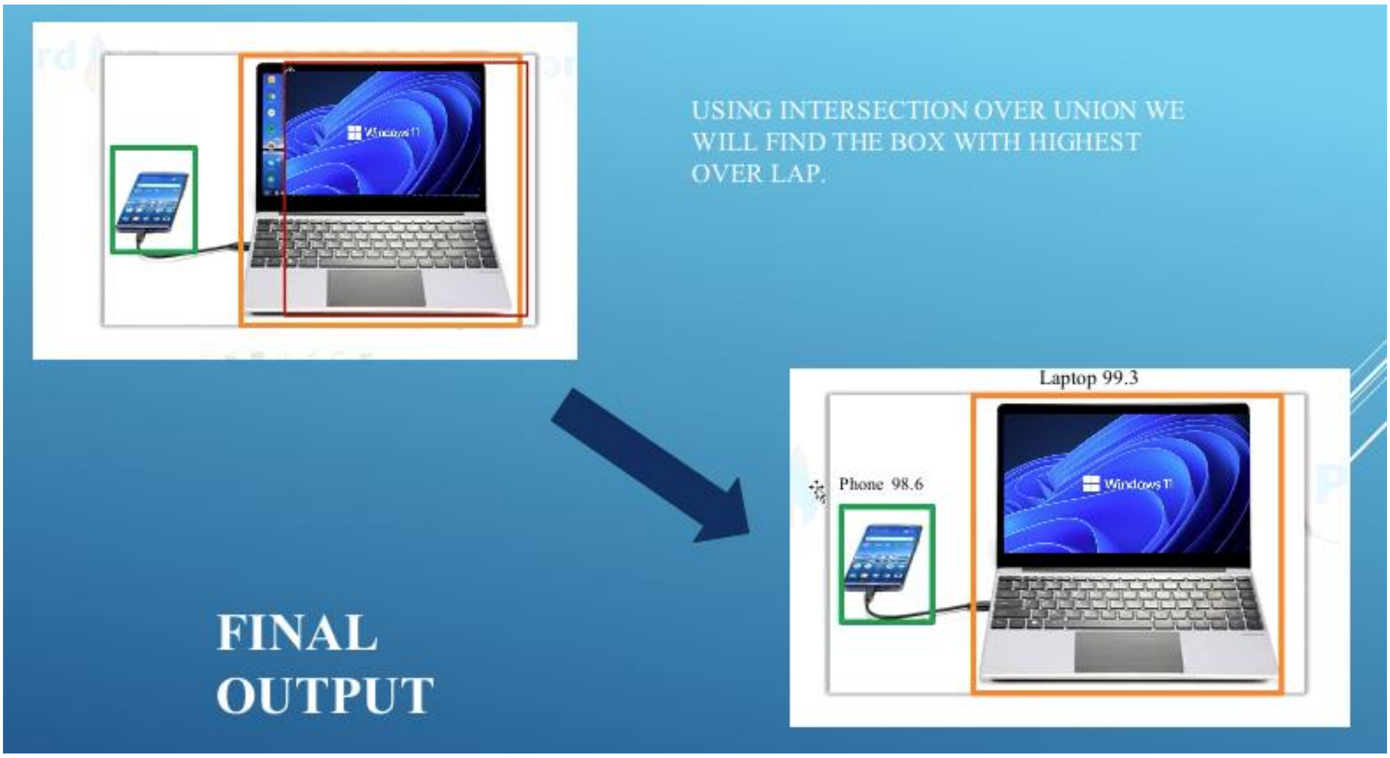


J

Fig. System Architecture

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930





2. LITERATURE SURVEY

There has already been a lot of study done on object detection and identification.

Current work is similar on different-2 parameters, but in this case, we compare past work on the basis of techniques and datasets utilised in existing systems. In their study, all of the researchers utilised either a region-based or a regression-based approach.

Region-based algorithms are well-known for their accuracy, whereas regression-based algorithms are well-known for their speed. Most studies have employed regression-based methods for real-time object identification. In summary, regression-based techniques like as YOLO and MobileNet SSD are becoming more popular for object recognition.

Some of the existing research work related to object detection and identification have been discussed here –

Prateek Agrawal et al.[7] has suggested a technique for verifying bank checks The following information is used to verify the bank check: cheque number, bank account number, bank branch code, legal and courtesy amount, and signature. They used the IDRBT cheque dataset and deep learning-based CNN to recognise handwritten digits with a high accuracy of 99.14 percent. For signature verification in this system, they employed SIFT feature extractor and SVM (Support Vector Machine) as a classifier with a high accuracy of 98.10 percent.

Sandipan Chowdhury et al.[6] has presented a method for detecting objects using a camera. They employed a mix of Fast R-CNN (Region-based Convolutional Neural Network) and Region Proposal Network (RPN) in this technique to create high-quality region proposals, which are then utilised by Fast-RCNN for detection. These two modules work together to create the Faster R-CNN object detection system. Quicker R-CNN algorithms can recognise objects in real-time at a high speed, however there are few methods that are faster.

F Particke et al.[5] has developed a real-time object identification and localisation system using a smartphone platform. F Particke et al. employed neural networks for real-time object identification and localisation in this system. As a result, DNN detection instructions are linked with depth data from a Depth Sensor (RGB-D camera), and that RGB-D camera is staged on a smartphone platform. The YOLOv2 algorithm was chosen by researchers as an object identification technique in this system. Localization in this system can be enhanced in future research by taking spatial information into account in the clustering mechanism. Using the X-means algorithm instead of the K-means method improves item recognition and location. The drawback with the YOLO method is that it requires a GPU to run and can only identify things from 2m to 5m away, so there is room for development.

Cheng Qian et al. [11] has developed an indoor navigation system The YOLOv2 algorithm was utilised in this system to recognise indoor items such as doors, door knobs, and so on. The YOLO algorithm has the benefit of being fast. In this setup, a visually impaired person is linked to a portable camera, Bluetooth earpiece, and GPU. This system is made up of three primary components: a deep neural network, a camera, and an audio device that allows the subject to learn about the items around him. In this model, Cheng Qian et al. employed a convolutional neural network (CNN). The ConvNet employed in this model includes 22 layers, it is ideal for recognising things as soon as the input photos are sorted into the matrix in the main layer when bounding box offers are upraised. The additional hardware devices, GPU, and stereo camera required make this variant expensive, as well as an additional hardship for visually impaired persons.

Andrew G. Howard et al [1] The researchers looked into some of the key design decisions that contribute to a successful model. We then showed how to use a width multiplier and a resolution multiplier to create smaller and quicker MobileNets by trading off a respectable degree of accuracy for size and latency reduction. Then they compared several MobileNets to common models, finding that they were superior in terms of size, speed, and accuracy. Finally, MobileNet's efficacy was demonstrated across a wide range of jobs. We want to release models in Tensor Flow as a further step to aid adoption and investigation of MobileNets. Recognition tasks must be completed in a timely manner on a computationally constrained platform in many real-world applications such as robots, self-driving cars, and augmented reality.

This paper explains how to develop very tiny, low latency models using an efficient network architecture and a set of two hyper-parameters that can be readily tailored to the design needs for mobile and embedded vision applications.

Animesh Srivastava et al [2] In this study, the speed, accuracy, and model size of Faster R-CNN and SSD MobileNet v2, both object identification models for detecting explicit content from an image, are compared. Instead of utilising a selective search technique, a separate network is utilised to anticipate region suggestions on the feature map, rather than using the slower and time-consuming Fast RCNN. Faster RCNNs are perfect for real-time object identification because the projected region proposals are reshaped using a RoI pooling layer, which is then used to categorise the image within the proposed region and forecast the offset values for the bounding boxes. As a result, MobileNet v2 may be used to identify objects in real time.

Jeong-ah Kim et al [3] This paper provides a deep learning model for detecting the location of shoes in photos in real time. The system had an average precision of detecting the shoes, which was comparable to other state-of-the-art systems.

The suggested detection model was trained and evaluated using 10.000 photos featuring shoes from the Open Images V5 dataset, and it is based on the SSD-MobileNetV2 architecture provided in the paper. This research is the beginning of our efforts to design a system that combines camera and radar data to increase the accuracy of our foot position estimating system. As a result, not only must the suggested shoe detector be accurate, but it must also be able to run in real time, even on low-processing machines.

Yu-Chen Chiu et al [4] This study uses the Mobilenet-v2 backbone network to create a lightweight network design with improved feature extraction. We use a combination of Mobilenet-v2 and FPN models to increase the feature map of the input image and the back-end detection network's detection accuracy.

The paper will provide an overview of the Mobilenet-SSDv2 detector, which not only preserves the original Mobilenet-SSD detector's advantage of quick processing but also considerably enhances detection accuracy. These benefits suggest that the Mobilenet-SSDv2 detection model suggested in this paper is more suited for embedded applications.

3. CONCLUSION

In this study, we conducted a broad literature review on both object identification and tracking algorithms, as well as their varied requirements. Based on the results of the survey, we discovered the following concerns with real-time object identification and tracking: 1. The majority of known algorithms operate on grayscale images/video. When converting a colour image/video to grayscale, some information is lost. As a result, detection and tracking are difficult. 2. All tracking algorithms assume that the motion of the object is smooth and without sudden changes. 3. As the quantity of data in the video increases, object recognition and tracking algorithms use greater computing and memory resources. 4. Some detection and tracking algorithms can identify and track several objects while also dealing with occlusion. However, greater computational and memory requirements are required. A separate object tracking method can handle variable light, background clutter, camouflage, bootstrapping, and occlusion. It is difficult to incorporate all of these into a single algorithm.

Based on our findings, we concluded that real-time object recognition and tracking techniques necessitate a parallel programming environment and tools like MobileNet SSD or OpenCV, etc.

REFERENCES

[1] MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications Andrew G. Howard Menglong Zhu IEEE.

[2] Explicit Content Detection using Faster R-CNN and SSD MobileNet. Animesh Srivastava1, Anuj Dalvi2 IEEE.

[3] Shoe Detection Using SSD-MobileNet Architecture. Jeong-ah Kim, Ju-Yeong Sung IEEE

[4] Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems Yu-Chen Chiu, Chi-Yi Tsai October 18,2020 at from IEEE Xplore.

[5] Particke, F., Kolbenschlag, R., Hiller, M., Patiño-Studencki, L., & Thielecke, J. (2017). Deep learning for real-time capable object detection and localization on mobile platforms. *IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK.*

[6] Chowdhury, S., & Sinha, P. Real Time Object Detection using Deep Learning: A Webcam Based Approach

[7] Agrawal, P., Chaudhary, D., Madaan, V., Zabrovskiy, A., Prodan, R., Kimovski, D., & Timmerer, C. (2020). Automated bank cheque verification using image processing and deep learning methods. *Multimedia Tools and Applications*, 1-32.

[8] Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition. Jeong-ah Kim, Ju-Yeong Sung

[9] Ojala. T., Matti. P., Topi, M. Multiresolution gray-scale and rotation invariant texture classification with local

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

binary patterns, IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 7, pp. 971–987, 2002.

[10] Medical Waste Classification System Based on OpenCV and SSD-MobileNet for 5G Xiaoxiao Bian, Yuchao Chen, Shuyi Wang 978-1-7281-9507-0/20/$31.00 ©2021 IEE

[11] Niu, L., Qian, C., Rizzo, J. R., Hudson, T., Li, Z., Enright, S., ... & Fang, Y. (2017). A wearable assistive technology for the visually impaired with door knob detection and real-time feedback for hand-to-handle manipulation. In *Proceedings of the IEEE International*

.