

# Real Time Object Detection and Recognition using MobileNet-SSD with OpenCV

Mr. Harshal Honmote  
Computer Engineering  
BSIOTR, Pune

Mr. Pranav Katta  
Computer Engineering  
BSIOTR, Pune

Mr. Shreyas Gadekar  
Computer Engineering  
BSIOTR, Pune

Prof. Madhavi Kulkarni  
Assistant Professor  
BSIOTR, Pune

**Abstract--** Real time object detection is an immense, vibrant and complex area of computer vision. Assuming there is a single object to be distinguished in an image, it is known as Image Localization and in the event that there are various objects in an image, then, at that point, it is Object Detection. Mobile networks and binary neural networks are the most generally involved techniques for current deep learning models to perform different tasks on embedded systems. In this paper, we develop a method to distinguish an item thinking about the deep learning pre-prepared model MobileNet for Single Shot Multi-Box Detector (SSD). This algorithm is used for real-time detection and for webcam streaming to detect object in a video stream. Subsequently, we utilize an object detection module that can identify what is in the video stream. To carry out the module, we join the MobileNet and the SSD framework for a quick and efficient deep learning-based strategy for object identification.

**Keywords—**MobileNet, SSD (Single Shot Multi-Box Detector).

## I. INTRODUCTION

Object detection is one of the most important fields of exploration in computer vision today. It is an augmentation of image classification the objective is to identify one or more classes of objects in a picture and with the help of bounding boxes locate their presence. Consequently, object detection carries an important role in many real-world applications like image recovery and video surveillance.

The main purpose of our analysis is to elaborate the accuracy of an object detection technique SSD and the pre-trained deep learning model MobileNet and additionally feature a portion of the notable elements that make this method stand out. The trial results show that the Average

Precision (AP) of the algorithm to recognize various classes as vehicle, person and chair is 99.76%, 97.76% and 71.07%, separately. This improves the accuracy of behavior detection at a handling speed which is needed for the real-time location and the necessities of day by day observing indoor and outside. The mix of MobileNet into the SSD framework forms one of the center parts of our work.

However, MobileNet with the effective SSD framework has been a hot exploration point in recent times, to a great extent because of managing the functional limits of running strong neural nets on low-end devices like cell phones/laptops to additionally expand the horde of conceivable outcomes with respect to real-time applications.

## II. LITERATURE SURVEY

### A. MobileNet-SSD

Our proposed model depends on the MobileNet-SSD architecture. One reason why we chose this architecture is on the because that as shown in the paper [2], it gives good object detection accuracy while being quicker than different architectures, for example, YOLO. Especially, this is valid when attempting to detect object in real time in low computing devices as in our system. MobileNet-SSD permits to lessen the detection time by addressing the model utilizing 8-bit integers rather than 32-bit floats. The input of the model was set to an image with 300 by 300 pixels and the result of the model addressed the position of the bounding box as well as the detection confidences (from 0 to 1) for each identified object. A detection confidence threshold of 0.5 was utilized to decide if the detected object was valid.[3]

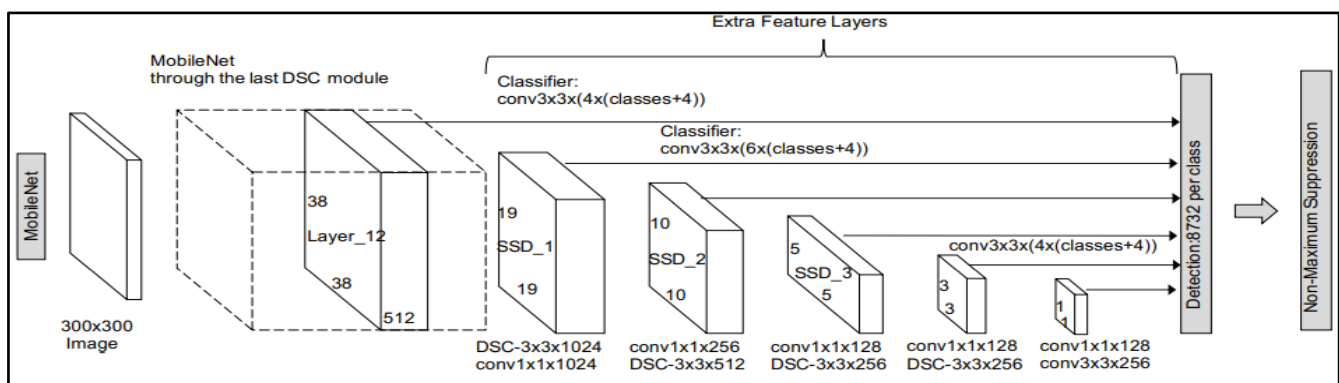


Fig. 1. SSD-based detection with MobileNet as backbone.[1]

### B. OpenCV (Open-Source computer vision)

OpenCV is a library of programming functions basically focused on real time computer vision. OpenCV is an open-source library which is useful for computer vision applications like CCTV film analysis, video analysis and image analysis. It is an incredible tool for image processing and performing computer vision tasks. OpenCV is written by C++ and has in excess of 2,500 optimized algorithms. [5] At the point when we make applications for computer vision that we do not want to make it from scratch instead we can utilize this library to begin focusing on real world problems. OpenCV has a function to read video, which is `cv2.VideoCapture()`. We can access webcam by passing 0 as function parameter. To catch CCTV film then we can pass RTSP URL in the function parameter, which is truly valuable for video analysis.

### III. PROPOSED SYSTEM

In the Proposed System, we are going to detect objects in real time with the help of Mobilenet-SSD model in fast and efficient way. We will create the Python script for object detection using deep neural network with OpenCV 3.4.

Working of the system is as follow:

Input will be given through Realtime video by camera or webcam, based on streamlined MobileNet Architecture which uses depth-wise separable convolutions to build light weight deep neural Networks. The input video divided into frames and pass it to MobileNet layers. [4] Each feature value is determined as a difference between the amount of pixel intensity under the bright region and the pixel intensity under the dark area. Every one of the possible sizes and area of the image is utilized to compute these elements. An image may contain irrelevant features and few relevant characteristics that can be used to detect the object.

The job of the MobileNet layers is to change over the pixels from the input image into highlights that describe the contents of the image. Then it passes to MobileNet-SSD model to determine the bounding boxes and corresponding class (label) of objects. After that the only last step is to show or display the Output.

### PROPOSED SYSTEM ARCHHITECTURE DIAGRAM:

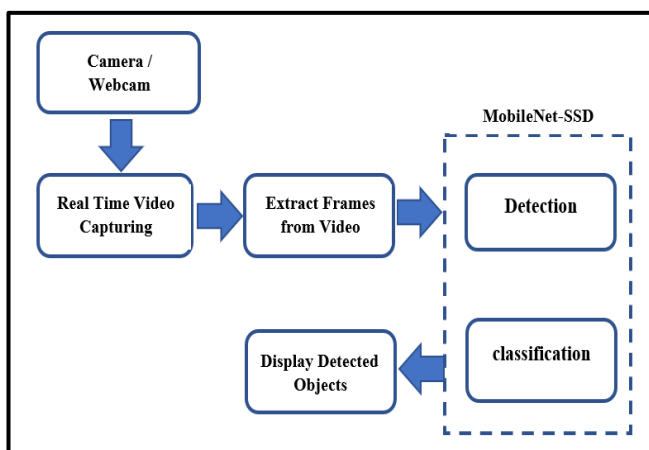


Fig.2. Proposed System Architecture Diagram

### IV. CONCLUSION

In this research, we proposed a deep learning model to identify progressively the place of the object in pictures. The framework could distinguish the item with a normal accuracy like other best in class frameworks. In this way, we utilize an object detection module that can recognize what is in the real time video stream. To carry out the module, we join the MobileNet and the SSD framework for a quick and productive deep learning-based strategy for object detection. In future work, we will keep on enhancing our detection network model, including lessening memory utilization and speeding up and additionally we will add more classes.

### V. REFERENCES

- [1] Yundong Zhang, Haomin Peng haomin and Pan Hu, "Towards Real-time Detection and Camera Triggering," CS341.
- [2] Ibai Gorordo Fernandez and Chikamune Wada, "Shoe Detection Using SSD-MobileNet Architecture," 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech 2020).
- [3] Yu-Chen Chiu, Chi-Yi Tsai, Mind-Da Ruan, Guan-Yu Shen and Tsu-Tian Lee, "Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems," ©2020 IEEE.
- [4] Andres Heredia and Gabriel Barros-Gavilanes, "Video processing inside embedded devices using SSD-Mobilenet to count mobility actors," 978-1-7281-1614-3/19 ©2019 IEEE.
- [5] G. Bradski and, A. Kaehler, "Learning OpenCV", O'Reilly Publications, 2008.
- [6] Animesh Srivastava1, Anuj Dalvi2, Cyrus Britto3, Harshit Rai4, Kavita Shelke5, "Explicit Content Detection using Faster R-CNN and SSD MobileNet v2," e-ISSN: 2395-0056 © 2020, IRJET.
- [7] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018.