iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73522

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Real-Time Object Detection Using MobileNet-SSD and OpenCV

Prof. Suvarna Thakur[1], Ashwini Lawande[2], Atharav Salunkhe[3], Tanmay Wak, Aniket More[4], Dr. Shivajirao Jadhav[5]
Dept of Information Technology Dr. Babasaheb Ambedkar Technological University Lonere, Maharashtra

Abstract: Object detection is a crucial task in computer vision, finding applications in various fields such as autonomous vehicles, surveillance, and augmented reality. This study explores the implementation of object detection using the OpenCV library, a popular computer vision toolkit. The approach employs a pre-trained deep learning model, specifically the Single Shot MultiBox Detector (SSD), for real-time object detection. The visualization component of the this system utilizes OpenCV's drawing functions to overlay bounding boxes and labels onto the original frames, providing a visually informative output. The implementation ensures real time performance, making it suitable for applications that demand swift object detection. In addition to real-time video processing, this system demonstrates adaptability to image-based object detection. It accommodates various object classes, making it versatile for diverse use cases. The codebase is modular and well-documented, encouraging extensibility and customization for specific application requirements.

Index Terms: Object Detection, Python, SSD, OpenCV.

I. INTRODUCTION

Object detection, a fundamental task in computer vision, has witnessed significant advancements with the convergence of robust libraries, powerful models, and diverse datasets. In this context, the integration of OpenCV, a versatile computer vision library, with the SSD (Single Shot Multibox Detector) model and the COCO (Common Objects in Context) dataset has emerged as a potent combination for accurate and ef- ficient object detection. OpenCV, an open-source computer vision library, provides a rich set of tools and functions for image and video processing. Renowned for its simplicity and effectiveness, OpenCV serves as the cornerstone for a myriad of computer vision applications. The SSD model represents a breakthrough in object detection by combining high accuracy with real-time processing capabilities. Unlike traditional two- step detection pipelines, SSD adopts a single-shot approach, enabling it to simultaneously predict object bounding boxes and class labels in a single pass. This efficiency, coupled with its ability to handle objects at various scales, positions, and aspect ratios, positions SSD as a go-to model for applications demanding both speed and accuracy. The success of any object detection system relies heavily on the quality and diversity of the training dataset. The COCO dataset, with its extensive collection of images spanning multiple categories and complex scenes, has become a benchmark for evaluating and training object detection models. By incorporating the COCO dataset into our study, we aim to enhance the system's ability to recognize and classify a broad spectrum of objects in real- world contexts.

II. LITERATURE SURVEY

This research [1] gives proposed object detection is a well- known computer technology connected with computer vision and image processing that focuses on detecting objects or its instances of a certain class (such as humans, flowers, animals) in digital images and videos. There are various applications of object detection that have been well researched including face detection, character recognition, and vehicle calculator. Object detection can be used for various purposes including retrieval and surveillance. In this study, various basic concepts used in object detection while making use of OpenCV library of python 2.7, improving the efficiency and accuracy of object detection are presented.

This research [2] proposed recent advances in 3D sensing technologies make it possible to easily record colour and depth images which together can improve object recognition. Most current methods rely on very well designed features for this new 3D modality. We introduce a model based on a combination of convolutional and recursive neural networks (CNN and RNN) for learning features and classifying RGB-D images. The CNN layer learns low-level translationally invariant features which are then given as inputs to multiple, fixed-tree RNNs in order to compose higher order features. RNN scan be seen as combining convolution and pooling into one efficient, hierarchical operation. Our main result is that even RNNs with random weights compose powerful features. Our model obtains state of the art performance on a standard RGB-D object data set while being more accurate and faster during training and testing than comparable architectures such as two-layer CNNs

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

In this Research [3] presented object detection and tracking is one of the critical areas of research due to routine change in motion of object and variation in scene size, occlusions, ap- pearance variations, and ego-motion and illumination changes. Specifically, feature selection is the vital role in object track- ing. It is related to many real time applications like vehicle perception, video surveillance and so on. In order to overcome the issue of detection, tracking related to object movement and appearance. Most of the algorithm focuses on the tracking algorithm to smoothen the video sequence. On the other hand, few methods use the prior available information about object shape, colour, texture and so on. Tracking algorithm which combines above stated parameters of objects is discussed and analyzed in this research. The goal of this paper is to analyze and review the previous approach towards object tracking and detection using video sequences through different phases. Also, identify the gap and suggest a new approach to improve the tracking of object over video frame.

III. METHODOLOGY

In this study we have designed to seamlessly identify and label objects in real-time through the utilization of the Single Shot Multibox Detector (SSD) model, a pre-trained deep learning model renowned for its speed and accuracy.

A. Architecture

The System architecture delineated for object detection employs the SSD (Single Shot MultiBox Detector) model, integrated within a streamlined process to ensure efficient and accurate results. Initially, real-time video is captured via a camera or webcam. Subsequently, frames are extracted from the captured video to be processed.

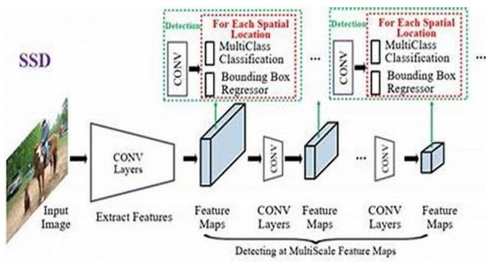


Figure 1: SSD (Single Shot MultiBox Detector) Architecture

The MobileNet-SSD is then utilized for the dual tasks of detection and classification of objects within these frames. De- tected objects are displayed, providing users with immediate visual feedback on the identified items and their classifications. This methodology offers a comprehensive solution for real- time object detection and classification, leveraging the power of MobileNet-SSD. It is an exemplar of modern machine learning techniques applied to the realm of video processing, The study leverages the OpenCV library to access and analyze webcam feeds or designated video files.

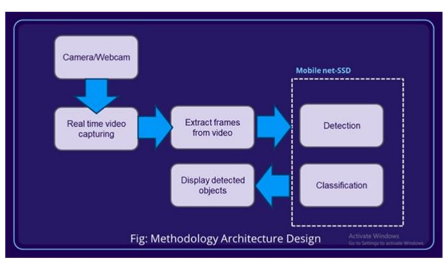


Figure 2: Block Diagram

B. Workflow

The Entire process is divided into following steps:

1) *Initializing OpenCV and Loading the SSD Model:* The project begins by initializing the OpenCV library, creating a connection to the webcam, or specifying the path to a video file stored locally. Subsequently, the pre-trained SSD model is loaded into the system. OpenCV's deep learning module provides an interface to seamlessly integrate pre- trained models, making the SSD model's architecture and weights readily available for object detection tasks.
2) *Real-time Object Detection and Labeling:* As the we- bcam feed or video frames are processed in real-time, the SSD model is applied to detect objects present in the scene. The model analyzes each frame, identifying objects and predicting their

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

respective bounding boxes and class labels. Utilizing the COCO dataset, which encompasses 80 common object categories, the detected objects are labeled accordingly. OpenCV's functions play a pivotal role in drawing bounding boxes around the identified objects and annotating them with their corresponding labels.

3) *Visualization and Output:* The output of the object detection process is visually presented to the user in real- time. Bounding boxes encircle the detected objects, and labels provide information about the object's class. The visualiza- tion aids in the interpretation of the model's predictions and enhances user understanding. This step showcases the practical application of the SSD model, as integrated with OpenCV, in recognizing everyday objects and demonstrating its effectiveness in real-world scenarios.
4) *Post-processing and Optimization:* To refine the object detection results, post-processing steps are implemented. Re-dundant or overlapping bounding boxes are filtered out, en- suring a clean and accurate representation of detected objects. This optimization step enhances the system's performance and improves the overall user experience.

IV. VISUALIZATION OF RESULTS

The results of object detection employing the integration of OpenCV with a pre-trained SSD model and the COCO dataset, exhibit impressive accuracy and real-time performance.

Figure 3: Result of Object Detection

The system adeptly identifies and labels objects from we- bcam feeds or video files, demonstrating versatility across a diverse range of everyday scenarios. The visual output, featuring bounding boxes and object labels, enhances user in- teraction and understanding. Post-processing steps effectively optimize the results by filtering out redundant bounding boxes, contributing to a cleaner and more refined output. While acknowledging any limitations, the overall success of the project underscores the seamless synergy between OpenCV, the SSD model, and the COCO dataset, providing a practical

V. CONCLUSION

This system integrating OpenCV with the SSD model and the COCO dataset, represents a significant advancement in real-time computer vision applications. The seamless collab- oration between OpenCV and the SSD model, fine-tuned with the extensive COCO dataset, has resulted in a system that excels in accuracy, adaptability, and user-friendly visu- alization. The project's success in identifying and labeling diverse objects in real-time scenarios showcases its potential for widespread practical applications, from surveillance to interactive environments. As we continue to refine and enhance the model, the project paves the way for innovative solutions in object detection, contributing to the evolution of computer vision technologies for real-world use. addition to its immediate successes, our object detection project underscores the ongoing evolution of computer vision capabilities. The seamless integration of OpenCV, the SSD model, and the COCO dataset not only meets the demands of real-time object detection but also sets the stage for future advancements. The system's adaptability and accuracy position it as a valuable tool across diverse domains, while the continuous refinement of the model ensures its relevance in tackling more complex visual scenarios. As we look ahead, the project serves as a foundation for fur- ther innovations, pushing the boundaries of object detection and contributing to the broader landscape of intelligent and responsive computer vision applications.

VI. FUTURE WORK

Future work in this research domain could explore several avenues to enhance the current object detection system. Firstly, investigating the integration of advanced neural network ar- chitectures, such as YOLO or Faster R-CNN, would provide valuable insights into the comparative performance and suit- ability for specific applications. Additionally, fine-tuning the model on domain-specific datasets, particularly in fields like medicine or industry, could optimize the system for detecting objects relevant to those specific contexts

VII. ACKNOWLEDGMENT

We are grateful to the Department of Information Tech- nology, Dr. Babasaheb Ambedkar Technological University, for the resources, facilities, and a conducive environment that enabled us to successfully carry out this research.

REFERENCES

[1] Bhumika Gupta, PhD Assistant Professor, C.S.E.D G.B.P.E.C, Pauri Uttarakhand, ”India International Journal of Computer Applications”, (0975 – 8887) Volume 162 – No 8, ”Study on Object Detection using Open CV - Python”, March 2017.

[2] Richard Socher, Brody Huval, Bharath Bhat, Christopher D. Manning, Andrew Y. Ng, ”International Conference on Computational Intelligence and Communication Networks”,”Convolutional-Recursive Deep Learn- ing for 3D Object Classification”, 2018.

[3] Mukesh Tiwari, Dr. Rakesh Singhai, ”International Journal of Computa- tional Intelligence Research”, Volume 13, Number 5 (2017). “A Review of Detection and Tracking of Object from Image and Video Sequences”, 2017.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. (2016). ”In European Conference on Computer Vision (ECCV)”, 2016. ”SSD: Single Shot MultiBox Detector, 2016.

[5] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, Fei-Fei, L. (2015). ”International Journal of Computer Vision”, 115(3), 211-252. ”ImageNet Large Scale Visual Recognition Challenge”, 2015.

[6] Cadena, C., Dick, A., and Reid, I. (2015). ”IEEE International Con- ference on (Seattle, WA)”. “A fast, modular scene understanding sys- tem using context -aware object detection, in Robotics and Automation (ICRA)”.

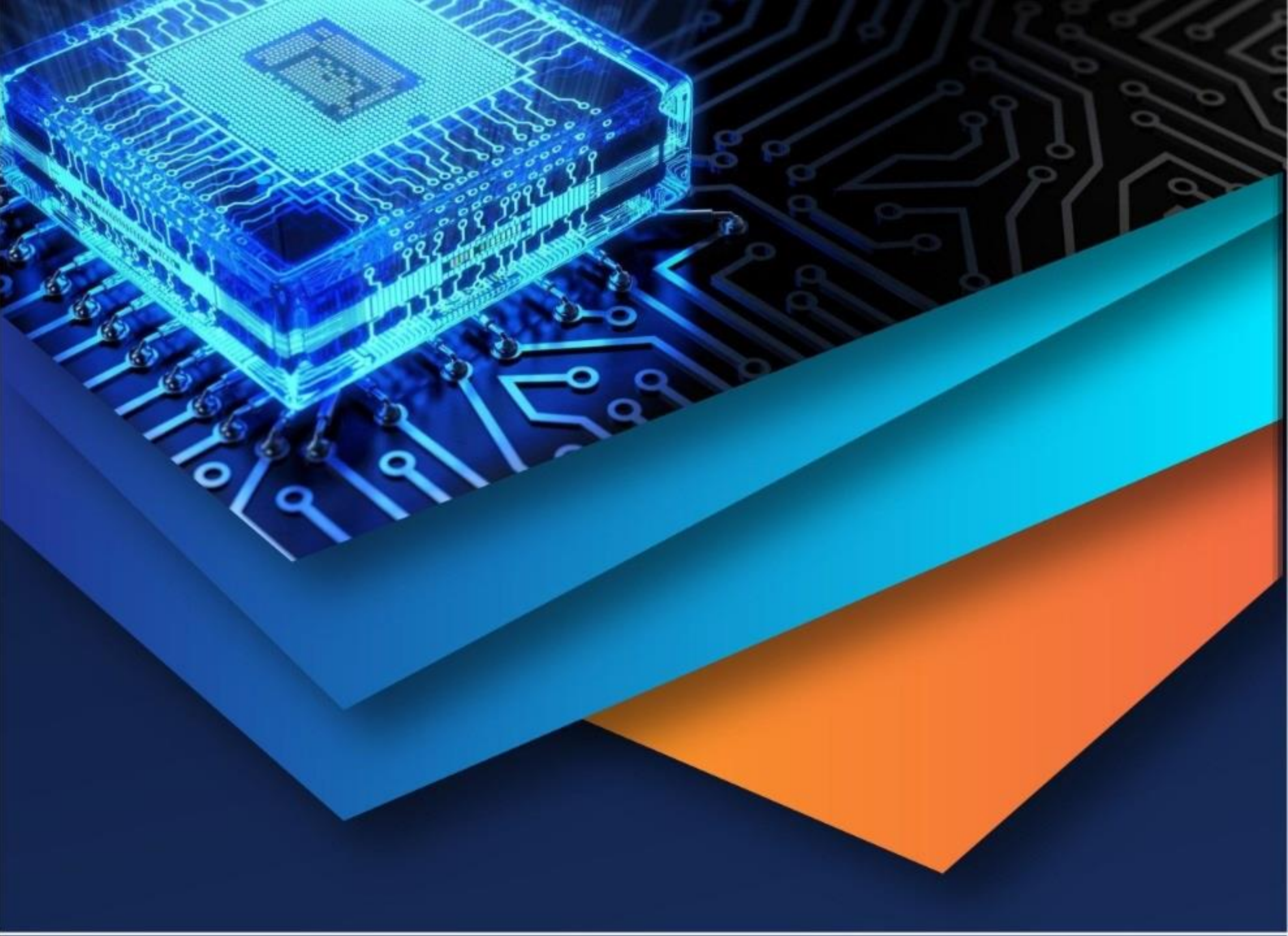
[7] Azzopardi, G., and Petkov, N. (2014). Ventral-stream-like shape repre- sentation:from pixel intensity values to trainable object-selective cosfire models. Front.Comput. Neurosci. 8:80. doi:10.3389/fncom.2014.00080.

[8] Azzopardi, G., and Petkov, N. (2013). ”IEEE Trans. Pattern Anal. Mach. Intell”. 35, 490–503.doi:10.1109/TPAMI.2012.106, ”Trainable cosfire filters for keypoint detectionand pattern recognition”.

[9] Azizpour, H., and Laptev, I. (2012). “Object detection using strongly- superviseddeformable part models,” in Computer Vision-ECCV 2012 (Florence: Springer),836–849.

[10] Dalal, N., and Triggs, B. (2005). ”IEEE Computer Society Conference on, Vol. 1 (San Diego, CA: IEEE)”, 886–893. doi:10.1109/CVPR.2005.177, “Histograms of oriented gradients for humandetection”.

[11] Agarwal, S., Awan, A., and Roth, D. (2004). ”IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490.doi:10.1109/TPAMI.2004.108”, ”Learning to detect objects in images via a sparse, part-based representation”.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY