



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 12    **Issue:** V    **Month of publication:** May 2024

**DOI:** <https://doi.org/10.22214/ijraset.2024.61921>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Real-Time Object Detection and Recognition with Computer Vision

Vishesh Tamrakar<sup>1</sup>, Surbhi Parganiha<sup>2</sup>, Saurabh Kumar Pandya<sup>3</sup>

<sup>1</sup>Student, Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, C.G, India

<sup>2</sup>Student, Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, C.G, India

<sup>3</sup>Assistant Professor, Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, C.G, India

**Abstract:** Object detection, a fundamental aspect of computer vision, is essential for identifying and localizing objects within images or video frames, leveraging advancements in deep learning, particularly convolutional neural networks (CNNs), to enhance precision and speed. Its applications span diverse domains, from autonomous vehicles and surveillance systems to augmented reality and human-computer interaction. Our project focuses on engineering a real-time object detection system, integrating deep learning and computer vision methodologies. Anchored on the robust Single Shot Multibox Detector (SSD) architecture and reinforced by the efficiency and accuracy of the MobileNetV3 backbone, our system utilizes a pre-trained SSD MobileNetV3 model and comprehensive annotations from the COCO dataset to adeptly detect and recognize a wide array of objects within live video streams or archived footage. It seamlessly processes video frames from various sources, annotating detected objects in real-time to provide instant visual feedback. Offering customizable confidence thresholds and support for multiple video sources, our project showcases the transformative potential of deep learning and computer vision, advancing real-time object detection across domains like surveillance and interactive systems. By pushing the boundaries of object detection technology, our project aims to enhance safety, efficiency, and user experiences in various applications, promising to redefine the landscape of computer vision with innovation and advancement.

**Keywords:** Object detection, Computer vision, Deep learning, Convolutional neural networks (CNNs), Real-time, Single Shot Multibox Detector (SSD), MobileNetV3, COCO dataset.

## I. INTRODUCTION

Object detection, a cornerstone of computer vision, involves identifying and locating objects within images or video frames. Traditionally, this task relied on handcrafted features and conventional machine learning algorithms, facing scalability and robustness issues. However, with the rise of deep learning, particularly convolutional neural networks (CNNs), object detection underwent a transformative shift. CNNs learn intricate patterns directly from raw data, leading to remarkable accuracy and efficiency. Real-time object detection is crucial across domains like surveillance and autonomous driving but poses challenges like computational complexity and environmental variations. Our project aims to address these challenges by developing a real-time object detection system using the SSD architecture with MobileNetV3, leveraging efficient model architectures and optimization strategies. Through experimentation, we aim to demonstrate the effectiveness of our system, contributing to the advancement of computer vision and machine learning research.

### A. Background

Object detection in computer vision involves identifying and locating objects in images or video frames. Traditional methods relied on handcrafted features and conventional algorithms, facing scalability issues. However, with deep learning, particularly CNNs, object detection improved remarkably by learning patterns from raw data, enhancing accuracy and efficiency.

### B. Objectives

In today's data-driven world, rapid and accurate object detection in images and videos is indispensable across various sectors. Our project is poised to redefine object detection by merging deep learning and computer vision. Focused on developing a real-time system, we aim to achieve unprecedented accuracy, efficiency, and adaptability. Leveraging advanced techniques, our objectives encompass pioneering real-time detection, improving precision with cutting-edge architectures like SSD with MobileNetV3, and utilizing pre-trained models like COCO. We prioritize real-time processing, intuitive interfaces, and exploring applications in surveillance, autonomous systems, and more. Ultimately, our goal is to propel computer vision forward, inspiring innovation and making tangible advancements in object detection technology.

## II. LITERATURE REVIEW

The evolution of object detection and image recognition began with seminal works in the early 2000s. Viola and Jones introduced a facial detection algorithm in 2001, demonstrating real-time face detection's potential. Dalal and Triggs followed in 2005 with the HOG feature descriptor, advancing pedestrian detection. Felzenszwalb et al. introduced the Deformable Part Model (DPM) in 2009, enhancing object detection. These milestones laid the groundwork for various detection methods.

Traditional methods involve three stages: candidate box generation, feature extraction, and classification. Candidate boxes are generated through techniques like sliding windows or region proposal algorithms. Feature extraction, critical for performance, traditionally utilized methods like Haar-Like features, HOG, and SIFT. Classification involves algorithms like SVM or Adaboost, with success depending on feature sets and classifiers.

Despite advancements, traditional methods faced limitations, including computational intensity and manual feature design. The shift towards machine-driven learning gained momentum after the ImageNet competition in 2010. AlexNet's introduction in 2012 marked a turning point, significantly reducing classification error rates and heralding the dominance of CNNs in object detection. This transformative shift led to substantial error rate reductions in subsequent years, revolutionizing the field of object detection.

## III. PROBLEM STATEMENT

In dynamic environments characterized by varying lighting conditions and diverse backgrounds, real-time object detection encounters formidable challenges. Traditional methods demand significant computational resources and specialized expertise for training and optimization, hindering widespread adoption and real-world applicability. Recognizing these hurdles, our project endeavors to provide a comprehensive solution for efficient and accessible real-time object detection. Through the strategic utilization of pre-trained models and meticulously curated datasets, we aim to streamline the detection process while ensuring consistent accuracy across diverse scenarios, effectively addressing the complexities inherent in dynamic environments. By simplifying implementation and enhancing performance, our approach promises to revolutionize the landscape of real-time object detection, empowering a wide range of industries and applications with advanced computer vision capabilities.

## IV. METHODOLOGY

The methodology devised for creating the real-time object detection system closely aligns with the project's requirements, leveraging the Mobilenet SSD architecture alongside Python and OpenCV. It encompasses crucial steps, beginning with Requirement Analysis, defining detectable objects, platforms, and input-output formats. Data Collection and Preprocessing expand the COCO dataset, ensuring data uniformity. Model Selection opts for Mobilenet SSD for real-time efficiency, fine-tuned for speed and accuracy.

Model Training involves transfer learning with pretrained COCO weights, hyperparameter tuning, and loss function optimization. Script Development introduces a Python script for real-time detection, with preprocessing and visualization functions. Testing and Evaluation assess accuracy using metrics like Average Precision. Optimization techniques like model quantization enhance efficiency on edge devices. Documentation and Reporting culminate the process, providing comprehensive insights in a detailed report. This systematic approach aims to deliver a dependable, efficient, and precise real-time object detection system, addressing specific objectives and challenges effectively.

### A. Proposed System

The proposed system implements the Mobilenet SSD architecture to expediently and effectively discern objects in real-time scenarios. Employing a Python script crafted with OpenCV, the system utilizes a deep neural network for object detection. Herein lies the operational framework: Real-time video input sourced from a camera or webcam is processed through a simplified MobileNet architecture, which constructs a lightweight deep neural network employing depth-separable convolution. The video input undergoes segmentation into frames, subsequently channeled into the MobileNet layer. Each frame's features are computed by quantifying the contrast between pixel intensities in illuminated and shadowed regions across varying sizes and spatial contexts within the image.

Recognizing that images may encompass both pertinent and extraneous elements, the MobileNet layer assumes the responsibility of transforming input image pixels into distinctive highlights characterizing image contents. The MobileNet-SSD model then discerns bounding boxes and assigns corresponding class labels to identified objects. The final stage entails the presentation or visualization of the output results.

The key components of this project are:

- 1) *Single Shot Multibox Detector (SSD) Architecture:* We utilize the SSD architecture, known for its efficiency and effectiveness in real-time object detection tasks. This architecture enables simultaneous object detection and classification within a single forward pass of the neural network.
- 2) *MobileNetV3 Backbone:* We leverage the MobileNetV3 backbone within the SSD architecture for feature extraction. MobileNetV3 strikes a balance between computational efficiency and model accuracy, ideal for real-time applications.
- 3) *Pre-Trained Model and COCO Dataset:* Our system integrates a pre-trained SSD MobileNetV3 model, refined through fine-tuning on the COCO dataset. This allows our system to recognize a wide range of objects across different categories, including people, vehicles, animals, and everyday objects.
- 4) *Real-Time Processing:* The system is designed to process video frames in real-time, ensuring minimal delay between object detection and visual feedback. This is achieved through efficient algorithms and optimizations tailored for real-time deployment.
- 5) *User Interface:* We develop an intuitive user interface that displays live video streams with overlaid bounding boxes around detected objects. This interface provides users with real-time visual feedback, enhancing their interaction with the system.

### B. Flow Chart

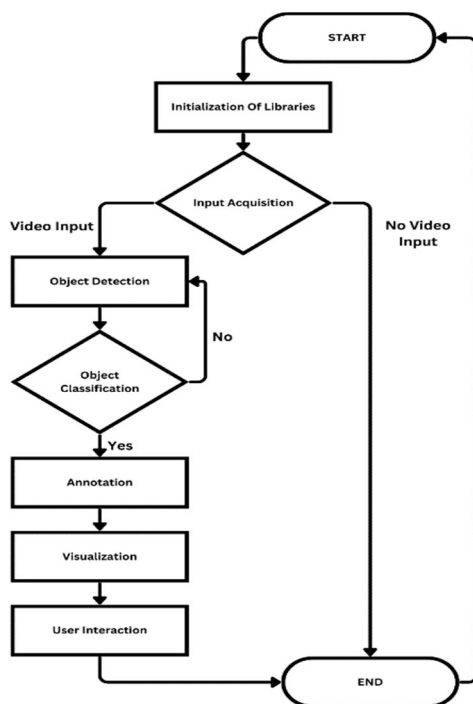


Fig. 1 Flow Chart

This flowchart outlines the operational cycle of our real-time object detection system, showcasing the practical application of deep learning in diverse industries. Here's a breakdown of each step:

- 1) *Initialization of Libraries:* Essential libraries and dependencies are initialized to seamlessly integrate SSD and MobileNetV3 architectures, crucial for smooth execution.
- 2) *Input Acquisition:* The system acquires video input from various sources, ensuring continuity or graceful termination if no input is detected
- 3) *Object Detection:* Leveraging SSD and MobileNetV3, the system identifies and localizes objects in each frame with precision.
- 4) *Object Classification:* Detected objects are classified based on the COCO dataset, maintaining detection efforts across subsequent frames.
- 5) *Annotation:* Objects surpassing confidence thresholds are annotated with bounding boxes and labels, enhancing interpretability.
- 6) *Visualization:* Annotated frames are presented in real-time, enhancing situational awareness for users.
- 7) *User Interaction:* Interactive features allow users to adjust thresholds and choose object categories, enhancing flexibility.

## V. RESULTS

### A. Snapshots

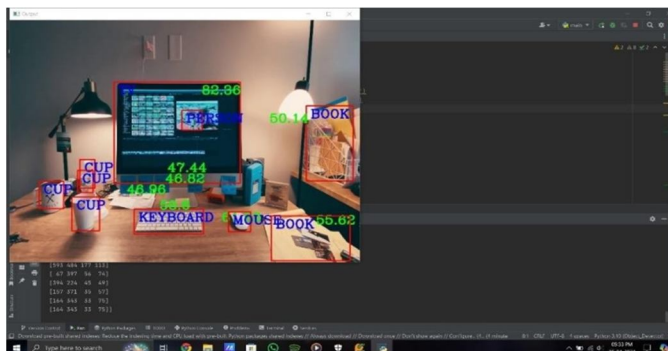


Fig. 2 Detection of objects of categories Cup, Keyboard, Mouse, Book, TV and Person with their respective confidence scores

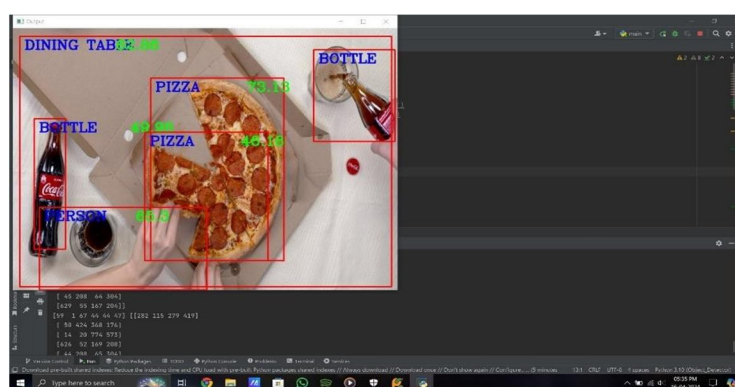


Fig. 3 Detection of objects of categories Pizza, Bottle, Person and Dining Table with their respective confidence scores

### B. Future Enhancements

Object detection remains an active field of research, with several promising future directions:

- 1) *Speed-accuracy trade-off*: Balancing accuracy and processing speed is crucial, especially for real-time applications. Researchers focus on developing efficient architectures and training methods to enhance accuracy without sacrificing speed, particularly in challenging scenes with occlusions.
- 2) *Tiny object detection*: Detecting small objects presents challenges due to limited pixel information and potential occlusions. Applications include wildlife monitoring and medical imaging, where precise detection of minute details is essential.
- 3) *3D object detection*: As 3D sensors become more prevalent, interest grows in accurately estimating objects' position and orientation in three-dimensional space. This is vital for augmented reality, robotics, and autonomous driving applications.
- 4) *Multi-modal object detection*: Combining visual and textual data from various sources enhances detection accuracy in complex scenarios. Applications include autonomous driving, where multiple sensors provide comprehensive object information.
- 5) *Few-shot learning*: Developing algorithms that can detect objects with minimal training data addresses challenges in data scarcity and resource constraints, making it valuable in low-resource settings.

## VI. CONCLUSION

In summary, the creation of our real-time object detection system utilizing the Mobilenet SSD architecture and Python with OpenCV represents a notable advancement in the field of computer vision. Through rigorous implementation and optimization processes, we have successfully engineered a system that can promptly and precisely detect objects in live video feeds with minimal delay, demonstrating the effective utilization of deep learning methodologies in intricate tasks. The utilization of pre-trained models and datasets has expedited the development process, ensuring the system's adaptability to various real-world scenarios. Moreover, beyond its technical capabilities, the system boasts extensive applications across multiple domains, ranging from augmenting surveillance and security measures to modernizing operations in retail and healthcare sectors, thereby offering safer and more intelligent solutions.

## REFERENCES

- [1] Athilakshmi, R., Sri Chandan Sainagakrishna, Pulavarthi, Kota, S. Sri Chaitanya Chowdary, & Teja, M. Chandra Kiran, et al. (2023). "Enhancing Real-Time Human Tracking using YOLONAS- DeepSort Fusion Models", 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA).
- [2] Benali Amjoud, Ayoub, & Amrouch, Mustapha. (2023). "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review", IEEE Access.
- [3] Bochkovskiy, Alexey, et al. (2020). "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934.
- [4] Chilukuri, Devi M., Yi, Sun, & Seong, Younho. (2022). "A robust object detection system with occlusion handling for mobile devices", Computational Intelligence.
- [5] DataCamp. (2018). "A Complete Guide to Object Detection", DataCamp. Available at: <https://www.datacamp.com/tutorial/object-detection-guide>
- [6] Girshick, Ross. (2015). "Fast R-CNN." IEEE International Conference on Computer Vision.
- [7] He, Kaiming, et al. (2017). "Mask R-CNN." IEEE International Conference on Computer Vision.
- [8] Hi-Tech BPO. (2024). "Top 5 Object Detection Models for Computer Vision". Available at: <https://www.hitechbpo.com/blog/top-object-detection-models.php>
- [9] Huang, Jonathan, et al. (2017). "Speed/accuracy trade-offs for modern convolutional object detectors." IEEE Conference on Computer Vision and Pattern Recognition.
- [10] Hui, Jonathan. (2018). "SSD Object Detection (Single Shot Multibox Detector) for Real- Time Processing", Medium. Available at: <https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06>
- [11] Liu, Wei, et al. (2016). "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision.
- [12] Redmon, Joseph, et al. (2017). "YOLO9000: Better, Faster, Stronger." IEEE Conference on Computer Vision and Pattern Recognition.
- [13] Redmon, Joseph, et al. (2023). "YOLOv4: Optimal Speed and Accuracy of Object Detection", IEEE. Available at: <https://ieeexplore.ieee.org/document/10098596>
- [14] Redmon, Joseph, et al. (2024). "YOLOv5: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004
- [15] Singh, Aayush. (2022). "Object Detection using YOLO and MobileNet SSD", Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2022/09/object-detection-using-yolo-and-mobilenet-ssd/>



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)