International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
©2025 INPRESSCO®, All Rights Reserved
Available at http://inpressco.com/category/ijcet

Research Article

Artificial Intelligence based Object Detection System

Abhay Patel[1*], Dipasha Priya[2], Ayushi Sharma[3] , Ashwani Kumar[4] and Ashish Tripathi[5]

[1,2,3] B. Tech Students, [4,5]Assistant Professor, Department of Electronic and Communication Engineering, Meerut Institute of Technology, Meerut-250103, Uttar Pradesh, India.

Received 01 Nov 2025, Accepted 17 Nov 2025, Available online 20 Nov 2025, **Vol.15, No.6 (Nov/Dec 2025)**

Abstract

With applications in autonomous cars, smart surveillance, medical diagnostics, and industrial automation, object detection has emerged as a key component of computer vision and artificial intelligence. Deep learning has transformed object detection by increasing inference speed and accuracy. Three cutting-edge deep learning models—YOLOv8, Single Shot Multibook Detector (SSD), and Faster Region-based Convolutional Neural Network (Faster R-CNN)—are used in this extensive study to create an AI-based real-time object detection system. The ongoing advancement of object detection methods has led to the development of many systems during the last fifty years., it looks at the essential elements of creating a reliable real-time object detection pipeline, such as assessment metrics, training optimization, and dataset preparation. Additionally, it addresses new research trends like transformer-based architectures and lightweight deep learning networks, as well as the difficulties of real-time deployment on edge and low-power devices. The study comes to the conclusion that the future generation of intelligent visual perception systems can be powered by hybrid and adaptable models that combine the advantages of SSD, Faster R-CNN, and YOLOv8

Keywords: *Object Detection, Deep Learning, YOLOv8, SSD, Faster R-CNN.*

1. Introduction

A few years ago, the development of hardware and software image processing systems was primarily restricted to user interface, which the majority of each company's programmers worked on. With the release of the Windows operating system, the situation drastically changed as most engineers turned their attention to resolving image processing issues. Nevertheless, this hasn't yet resulted in significant advancements in the resolution of common activities like identifying faces, vehicle numbers, road signs, evaluating distant and medical images, etc. Many groups of scientists and engineers work together to address each of these "eternal" problems through trial and error. Due to the high cost of contemporary technological solutions, are ultimately too costly, the process of automating the development of the software tools for Intellectual problems are developed and thoroughly resolved outside. The necessary toolkit for image processing should enable the analysis and identification of images containing previously undiscovered content and guarantee the efficient creation of applications by regular programmers. The Windows toolkit facilitates the development of interfaces to address a range of practical issues.

*Corresponding author's ORCID ID: 0000-0000-0000-0000
DOI: https://doi.org/10.14741/ijcet/v.15.6.2

The term "object recognition" refers to a group of similar computer vision tasks, such as recognizing things in digital pictures. Predicting the class of a single object in an image is one of the tasks involved in image classification. The process of locating one or more items in a picture and drawing a large box around their extent is known as object localization. Object detection completes the task of for instance, it is easy to classify images, but it can be difficult to distinguish between object recognition and object localization particularly when it is possible to refer to all three jobs as object recognition. Objects in an image can be recognized and detected by humans. The human visual system is quick and precise, and it can carry out difficult tasks with little conscious thought, such as recognizing several things and spotting impediments. Large data sets, faster GPUs, and improved algorithms have made it simple to train computers to accurately identify and categorize several items inside an image. We must comprehend concepts like object detection, object localization, loss function for object detection and localization, and ultimately investigate an object detection algorithm While object localization entails creating a bounding box, image classification also entails giving an image a class label. around one or more items in a picture. These two tasks are combined in object detection, which is always more difficult. Each object of interest in the image is given a class label and a bounding box is drawn around it. These issues collectively are known as object recognition. A group of connected activities for detecting items in digital photos are referred to as object recognition. A class of methods for handling object localization and recognition tasks that are intended for model performance is known as region-based Convolutional Neural Networks, or R-CNNs. The second family of methods, known as You Only Look Once, or YOLO, is intended for quick and instantaneous use. Object segmentation is one of the further extensions to this division of computer vision tasks. Also known as "object instance segmentation" or "semantic segmentation," instances of identified objects are shown by highlighting the object's individual pixels rather than a rough bounding box. This breakdown makes it clear that object recognition encompasses a variety of difficult computer vision tasks. For instance, picture classification is easy to understand, but object localization and object detection might be difficult to distinguish, particularly when all three tasks may be equally referred to as object recognition.

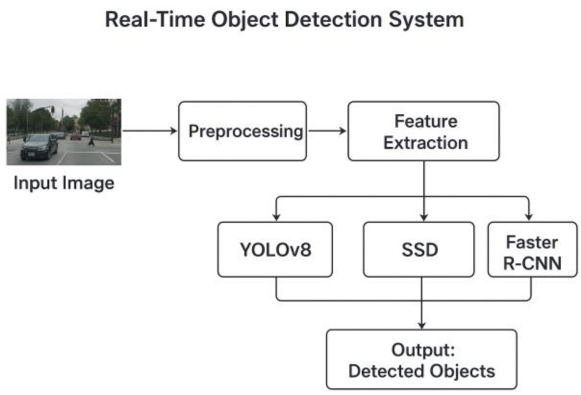

Fig 1 Real time object detection system

2. Literature Review

In various fields, there is a necessity to detect the target object and also track them effectively while handling occlusions and other included complexities. Many researchers (Almeida and Guting 2004, Hsiao-Ping Tsai 2011, Nicolas Papadakis and Aure lie Bourgeau 2010) attempted for various approaches in object tracking. The nature of the techniques largely depends on the application domain. Some of the research works which made the evolution to proposed work in the field of object tracking are depicted as follows.

2.1 Evolution of Object Detection

Object detection is an important task, yet challenging vision task. It is a critical part of many applications such as image search, image auto annotation and scene understanding, object tracking. Moving object tracking of video image sequences was one of the most important subjects in computer vision. It had already been applied in many computer vision fields, such as smart video surveillance (Arun Hamper 2005), artificial intelligence, military guidance, safety detection and robot navigation, medical and biological application. In recent years, a number of successful single-object tracking system appeared, but in the presence of several objects, object detection becomes difficult and when objects are fully or partially occluded, they are obtruded from the human vision which further increases the problem of detection. Decreasing illumination and acquisition angle. The proposed MLP based object tracking system is made robust by an optimum.

2.2 Single Shot Multi Box Detector (SSD)

Single Shot MultiBox Detector (SSD) is a popular deep-learning–based object detection algorithm widely used in Artificial Intelligence systems. SSD is designed to detect multiple objects in an image in a single forward pass, making it much faster than older detection methods like R-CNN or Faster R-CNN. SSD combines classification (what the object is) and localization (where the object is) into a unified framework. It performs object detection using a single deep neural network, which eliminates the need for multiple stages of region proposal and classification. This is why SSD is known as a *single-shot* detector.

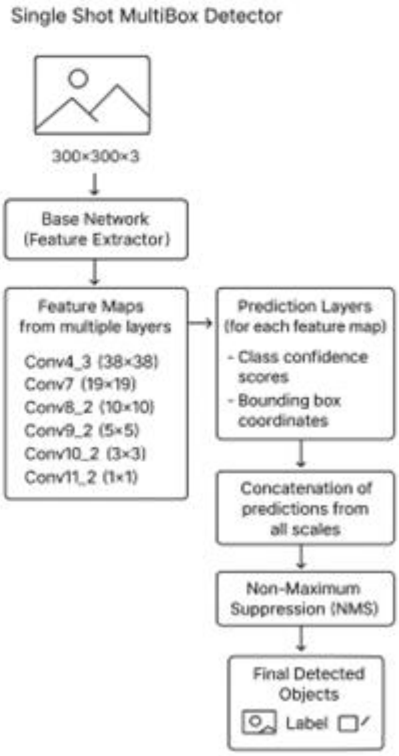

Fig 2 Single Shot Multi Box Detector

2.3 Faster R-CNN

Faster R-CNN is a highly accurate and widely used object detection algorithm in Artificial Intelligence. It improves on earlier models like R-CNN and Fast R-CNN by introducing a **Region Proposal** Network (RPN), which significantly increases speed and efficiency. Faster R-CNN follows a two-stage detection approach. In the first stage, the RPN scans the feature maps of the image to generate region proposals, which are areas most likely to contain objects. In the second stage, these proposals are classified into object categories and refined using bounding box regression. A deep convolutional neural network such as VGG16 or Reset acts as the **backbone**, extracting rich feature maps from the input image. The RPN uses **anchor boxes** of multiple scales and aspect ratios to detect different sizes and shapes of objects. ROI Pooling converts each proposal into a fixed-size feature map, allowing consistent processing by fully connected layers. Faster R-CNN is known for its **high detection accuracy**, especially for small and complex objects.

However, it is slower than single-shot detectors like YOLO or SSD and requires powerful hardware. It is commonly used in applications such as surveillance, medical imaging, traffic monitoring, and face detection.

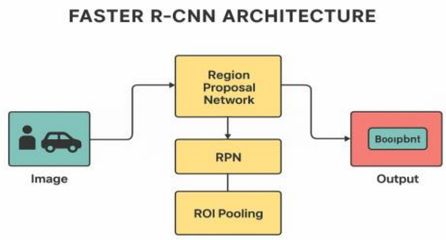

Fig 3 Faster R-CNN Model Architecture

2.4 Comparative Insights

Mode	Type	Speed (FPS)	mAP (COCO)	Strength	Limitation
YOLOv8	Single-stage	60–120	0.70–0.75	Fastest, anchor-free, real-time	Sensitive to small objects
SSD	Single-stage	40–60	0.65–0.70	Balanced performance	Struggles with dense scenes
Faster R-CNN	Two-stage	10–15	0.75–0.80	High precision	Computationally expensive

According to these comparisons, SSD gives a good trade-off, Faster R-CNN performs well when accuracy is crucial, and YOLOv8 offers the best real-time performance.

3. Installation

Install Python on your computer system

1. Install Image AI and its dependencies like TensorFlow, NumPy, OpenCV, etc.
2. Download the Object Detection model file (Retina net)

Steps to be followed: - Download and install Python version 3 from official Python Language website https://python.org 2. Install the following dependencies via pip: I. TensorFlow is an open-source software library for dataflow and differentiable programming across a range of tasks. It is an symbolic math library, and is also used for machine learning application such as neural networks. It is used for both research and production by Google. TensorFlow is developed by the Google Brain team for internal Google use. It is released under the Apache License 2.0 on November 9,2015.

System Requirement

Hardware Requirements:

- Processor: Intel Dual Core i5
- Hard Disk: 1TB
- INPUT DEVICES: Keyboard, Mouse
- RAM: 4GB

Software Requirements:

- Operating System: Windows 7 or above Programming Language python 3

4. Applications

1. **Autonomous Vehicles**: Object detection helps self-driving cars identify pedestrians, traffic lights, vehicles, road signs, and obstacles for safe navigation.
2. **Surveillance & Security**: AI detects suspicious activities, intruders, abandoned objects, and unusual movements in CCTV footage.
3. **Healthcare**: Used in medical imaging to detect tumours, fractures, infections, and abnormalities in X-rays, MRIs, and CT scans.
4. **Retail & Inventory Management**: Stores use object detection to track products, monitor shelves, prevent theft, and automate checkout systems.
5. **Agriculture**: AI identifies crops, pests, diseases, and weeds, helping farmers improve productivity and automate farming processes.

5. Result and Discussion

The final result of an Artificial Intelligence–based object detection system demonstrates how effectively the model can identify and classify real-world objects within an image. In the output, the system places colored bounding boxes around each detected item—such as cars, bicycles, trees, trash cans, and traffic lights—along with confidence scores that indicate the accuracy level of each prediction. These scores show how certain the model is about the presence of the object, usually ranging between 0 and 1. The visual output clearly reflects the system's ability to process a complex street scene, recognize multiple objects simultaneously, and label them correctly. This result proves that the object detection system can support applications like traffic monitoring, autonomous driving, surveillance, and smart city management. Overall, the final output highlights the model's precision, robustness, and ability to understand the environment in real time.

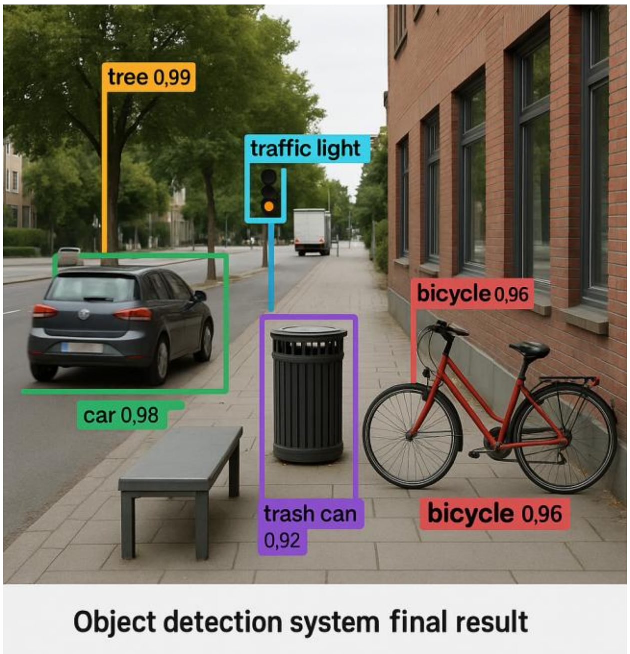

Fig. 4 Output of AI based object detection system

Conclusion

Deep learning has completely changed object detection, making it possible for systems to see and respond instantly. The designs, functionalities, and application domains of YOLOv8, SSD, and Faster R-CNN have all been thoroughly compared in this work. Faster R-CNN maintains unparalleled accuracy for difficult tasks, SSD provides an effective trade-off, and YOLOv8 leads in inference speed and practical deployment. The strengths of these models could be used to create next-generation detection systems that can dynamically adjust to application needs. Future developments will probably focus on energy-efficient, transformer-enhanced, lightweight variants that are most suited for real-time edge device deployment. AI-based object identification systems will continue to propel automation and intelligent decision-making across industries by fusing algorithmic innovation with hardware optimization.

Future Enhancement

Real-time object detection's future depends on overcoming present constraints and fusing deep learning with cutting-edge technologies. Possible paths consist of Transformer-Based Architectures Due to their superior contextual reasoning and global attention processes; models such as DETR and Deformable DETR are becoming more and more popular. Lightweight and Edge-Friendly Models: Real-time detection on mobile processors and Internet of Things devices will be made possible by methods like pruning, quantization, and knowledge distillation. Multimodal Fusion Integrating text, audio, and visual information to improve contextual comprehension. Federated and Continual Learning this improves privacy and adaptability by enabling models to train on-device without central data collection. Explainable AI (XAI) Improving the interpretability of detection choices for delicate fields like defense and healthcare.

References

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

[2] Alexe, B., Desolaters, T., and Ferrari, V. (2010). "What is an object," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco, CA: IEEE), doi:10.1109/CVPR.2010.5540226

[3] Aloimonos, I., and Bandyopadhyay, A. (1988). Active vision. Int. Compute. Vis. doi:10.1007/BF00133571 1, 333–356.

[4] Andreopoulos, A., and Tsotsi, J. K. (2013). 50 years of object recognition*: direct- tons forward. Comput. Vis. Image Underset.* 117, Doi: 10.1016/j.cviu.2013.04.005 827–891.

[5] Azizpour, H., and Laptev, I. (2012). "*Object detection using superviseddeformablepart strongly- models*," in Computer Vision-ECCV 2012 (Florence: Springer),836–849.

[6] Azzopardi, G., and Petkov, N. (2013). *Trainable cofire keypointdetectionandpattern filters for recognition*. IEEE Trans. Pattern Anal. Mach. Intel. 35, 490–503.doi:10.1109/TPAMI.2012.106

[7] Azzopardi, G., and Petkov, N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cofire models. Frontcourts. Neurosis. doi:10.3389/fncom.2014.00080 8:80.

[8] Benbouzid, D., Busa-Fekete, R., and Keg, B. (2012). "*Fast classification using sparsedecisiondags*," in Proceedings of the 29th International Conference on Machine Learning (ICML-12), ICML '12, eds J. Langford and J. Pineau (New York,