Journal of Science, Computing and Engineering Research (JSCER)
Volume-9, Issue-4, April 2026.
DOI: ***https://doi.org/10.46379/jscer.20.***

# VIDEO BASED OBJECT RECOGNITION

**[1]Arvind Sharma,[2]Hemali Jain,[3]Ishani Agarwal**

[1]Professor, Department of CSE, Modern Institute of Technology and Research Centre, Rajasthan, India.

[2,3] UG Student, Department of CSE, Modern Institute of Technology and Research Centre, Rajasthan, India

**Article Information**

| | | |
|---|---|---|
| Received | : | Mar 31 2026 |
| Revised | : | Mar 31 2026 |
| Accepted | : | April 1 2026 |
| Published | : | April 3 2026 |

*Corresponding Author:*

*HEMALI JAIN*

***Abstract*—** Video-based object recognition is an advanced technology in the field of computer vision and artificial intelligence that enables automatic detection, classification, and tracking of objects in video streams. Unlike traditional image-based systems, this approach processes continuous video frames, allowing the system to understand motion, behaviour, and temporal changes over time. It is widely used in applications such as surveillance systems, traffic monitoring, autonomous vehicles, healthcare monitoring, and smart city solutions. One of the major goals of this project is to overcome common challenges such as varying lighting conditions, occlusion, background noise, and real-time processing limitations. To address these issues, preprocessing techniques, feature extraction methods, and optimization strategies are applied to improve system performance. The system is designed to balance accuracy and speed, making it suitable for real-world applications.

Furthermore, the system emphasizes scalability and adaptability, allowing it to be integrated with cloud platforms, edge devices, and IoT-based systems. It provides output in the form of annotated video frames with bounding boxes, labels, and confidence scores. Overall, this project demonstrates the potential of intelligent video analysis in automating tasks, improving decision-making To address these issues, preprocessing techniques, feature extraction methods, and optimization strategies are applied to improve system performance. The system is designed to balance accuracy and speed, making it suitable for real-world applications.

**Keywords**: Computer Vision, Deep Learning, Convolutional Neural Network (CNN), Object Detection, Object Tracking, Real-Time Processing, YOLO, SSD, Faster R-CNN, Image Processing, Frame Extraction, Feature Extraction



**Citation: Arvind Sharma, Hemali Jain, Ishani Agarwal,.** "VIDEO BASED OBEJCT RECOGNITION", Journal of Science, Computing and Engineering Research, 9(4), April 2026.

## I INTRODUCTION

Video-based object recognition is an important application of computer vision and artificial intelligence that focuses on identifying, classifying, and tracking objects in video streams. Unlike image-based recognition, which works on a single image, video-based systems process a sequence of frames, enabling the system to understand motion and changes over time.

This makes it highly useful for real-time applications such as surveillance, traffic monitoring, autonomous driving, and smart city systems. This project aims to create a practical, AI-driven learning assistant that helps students' study more efficiently and grasp complex material with ease. Instead of forcing students into rigid, one-size-fits all systems, the assistant adapts to each user's needs using modular GPT components.

The working of a video-based object recognition system involves capturing video input, converting it into frames, preprocessing the frames, and applying trained models to detect and classify objects. In addition, object tracking techniques are used to follow the movement of objects across frames, which helps in maintaining object identity and understanding behaviour.

## II. PROBLEM STATEMENT

Video-based object recognition systems face several challenges in achieving accurate and efficient performance in real-time environments. Traditional methods often struggle with detecting objects in dynamic scenes where factors such as
motion, occlusion, varying lighting conditions, and complex backgrounds affect accuracy. Additionally, processing

continuous video streams requires high computational power, which can lead to delays and reduced system efficiency.

Another major issue is maintaining object identity across multiple frames, especially when multiple objects are present or when objects overlap. Existing systems may produce false detections or fail to track objects consistently. Moreover, many models require large amounts of labelled data and may not perform well when exposed to new or unseen environments.

## III. PROPOSED MODEL

A. The proposed model for the video-based object recognition system is designed to efficiently detect, classify, and track objects in real-time video streams using deep learning and computer vision techniques. The model follows a structured pipeline where video input is processed step-by-step to ensure accurate and fast performance. It integrates object detection and tracking mechanisms to provide continuous monitoring and reliable results in dynamic environments.

B. **1. Video Input Module**

C. This module captures video either from a live camera (webcam/CCTV) or from stored video files. It serves as the starting point of the system by providing continuous visual data for processing.

D. **2. Frame Processing Module**

E. The input video is divided into frames, and each frame is pre processed using techniques like resizing, normalization, and noise reduction. This improves the quality of input data and prepares it for detection.

**3. Object Detection Module**

In this module, deep learning models such as CNN-based algorithms (YOLO, SSD, or Faster R-CNN) are used to detect and classify objects in each frame. The system identifies objects and draws bounding boxes with labels.

**4. Object Tracking Module**

After detection, tracking algorithms are applied to follow objects across consecutive frames. This helps in maintaining object identity and analyzing movement pattern

**5. Feature Extraction Module**

Important features such as shape, colour, and motion are extracted to enhance recognition accuracy and reduce false detections.

**6. Output & Visualization Module**

The final results are displayed as video output with bounding boxes, object labels, and confidence scores. It may also include object counting and tracking paths.

**7. Performance Evaluation Module**

The system is evaluated using metrics such as accuracy, precision, recall, and FPS to measure its efficiency and reliability.

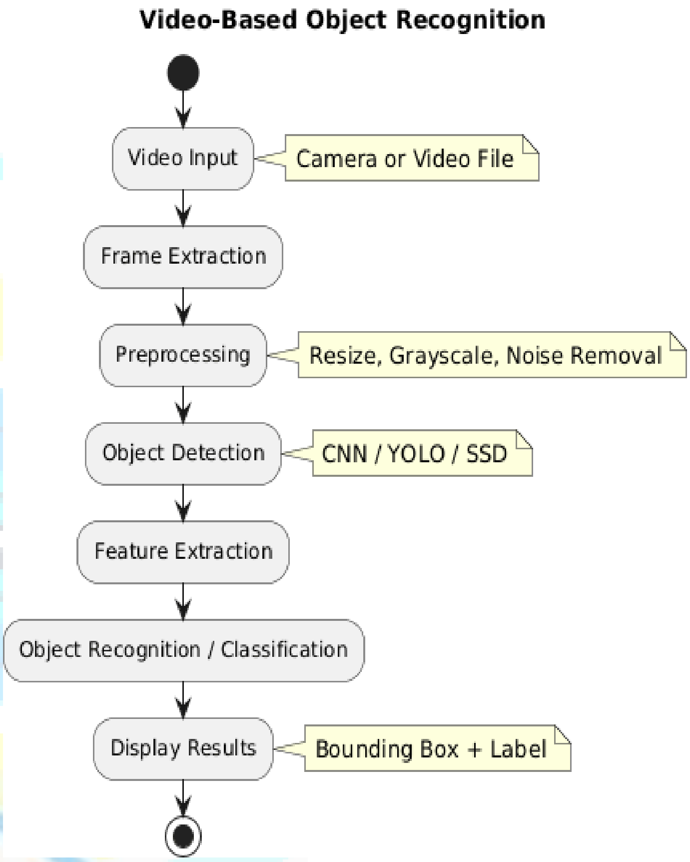


Figure 1 Block Diagram

## IV TECH STACK

The technology stack of the video-based object recognition project includes various tools, frameworks, and hardware components that work together to build an efficient and scalable system.

**Deep Learning Frameworks**

Frameworks such as **TensorFlow** and **Py Torch** are used to build and train deep learning models. These frameworks support GPU acceleration for faster processing.

**Detection Algorithms**

Algorithms like **YOLO**, **SSD**, and **Faster R-CNN** are used for real-time object detection and classification with high accuracy

**Data Processing Libraries**

Libraries such as **NumPy** and **Pandas** are used for handling data, performing numerical computations, and managing datasets.

**Visualization Tools**

Matplotlib and other visualization tools are used to display results, graphs, and performance metrics.

**Hardware Requirements**

A system with a good CPU and preferably a GPU is used for faster training and real-time processing.

**Deployment & Integration**

Technologies like Flask or Django can be used for building a user interface, while cloud platforms and IoT integration help in scaling the system.

Therefore, there is a need to develop a robust and efficient video-based object recognition system that can accurately detect, classify, and track objects in real time while minimizing computational cost. The system should be capable of handling challenging conditions and provide reliable performance for practical applications such as surveillance, traffic monitoring, and smart systems.

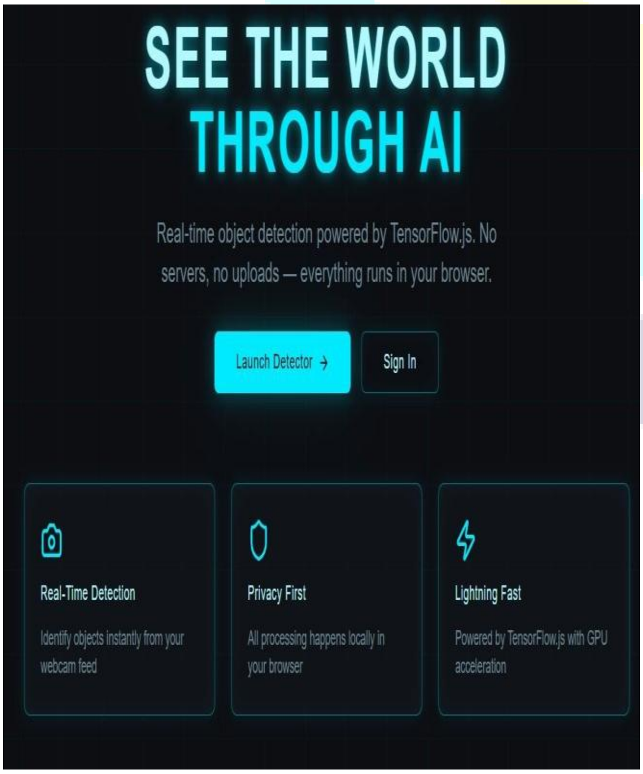


Figure 2 Home Page

## V RESULTS

The project addresses key challenges such as varying lighting conditions, motion, and complex backgrounds by applying optimization and feature extraction techniques.

It also demonstrates the importance of using advanced algorithms like CNN-based models for improving performance in real-time applications.
The system is capable of handling multiple objects simultaneously and provides clear visual outputs with bounding boxes and labels.
Overall, the proposed system proves to be efficient, scalable, and suitable for real-world applications such as surveillance, traffic monitoring, and smart systems.

It highlights the potential of artificial intelligence and computer vision in automating tasks and improving decision-making. With further improvements, the system can be enhanced to achieve even higher accuracy and broader applicability in future technologies.

The integration of object tracking helped in maintaining the identity of objects across consecutive frames, ensuring smooth and continuous tracking even when objects were in motion. The system performed well under normal lighting conditions and showed acceptable performance in moderately challenging environments such as slight occlusion and background variations.

Performance evaluation showed that the system achieved satisfactory accuracy, precision, and recall values, along with a reasonable frame processing speed (FPS), making it suitable for real-time applications. The use of optimized algorithms and preprocessing techniques helped in reducing false detections and improving overall efficiency.

## VI CONCLUSION

The implemented video-based object recognition system produced effective and reliable results in detecting, classifying, and tracking objects in real-time video streams. The system was able to successfully identify multiple objects such as people, vehicles, and other common items with good accuracy. Bounding boxes, labels, and confidence scores were correctly displayed on the video output, making the results clear and easy to understand.

In this project, a video-based object recognition system has been developed to detect, classify, and track objects in real-time video streams using computer vision and deep learning techniques.

The system successfully processes video input by converting it into frames, applying preprocessing methods, and using trained models to identify objects with good accuracy. Integration of object detection and tracking ensures continuous monitoring and better understanding of object movement.

## VII REFERENCES

1. Alex Krizhevsky et al., “ImageNet Classification with Deep Convolutional Neural Networks (Alex Net)”, 2012.
2. Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG Net)”, 2014.
3. Kaiming He et al., “Deep Residual Learning for Image Recognition (Res Net)”, 2015.
4. Tsung-Yi Lin et al., “Microsoft COCO: Common Objects in Context”, 2014.