IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153136

Real-Time Object Detection System Using YOLO-Based Vision Models

Mrs. K. Deepthi[1], B. Akash[2], Revanth Kumar[3], Ch. Tharun[4], K. Narendra[5]

Assistant Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India[1]

Student, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, India[2-5]

Abstract: Object detection plays an important role in many real-time applications such as surveillance, traffic monitoring, smart automation. In recent years, deep learning techniques have significantly improved the accuracy and speed of object detection systems. This project presents the implementation of a real-time object detection system. Using YOLO integrated with CNN. The proposed system processes video frames, extracts features using CNN layers, and detects objects. By predicting bounding boxes and class labels in a single step. YOLO enables fast detection while maintaining acceptable accuracy, making it suitable for real-time environments. Experimental results show that the system performs efficiently on live video streams with good accuracy and real-time speed. Proving its suitability for practical applications.

Keywords: Real-time object detection, YOLOv8, CNN, Python, OpenCV, Smart monitoring.

I. INTRODUCTION

Object detection is the task of recognizing objects and determining their spatial locations within an image or video. It is an essential component in applications such as intelligent surveillance systems, autonomous driving, robotics, human–computer interaction. Earlier object detection techniques were based on handcrafted features and traditional machine learning algorithms. which were slow and less accurate. The emergence of deep learning, particularly Convolutional Neural Networks, has significantly improved object detection performance. Among various deep learning-based approaches, YOLO is widely used due to its ability to perform detection in real time. Unlike two-stage detectors, YOLO treats object detection as a single regression problem and predicts bounding boxes and class probabilities simultaneously. This paper focuses on implementing a YOLO-based object detection system. That provides a good balance between speed and accuracy.

II. LITERATURE SURVEY

Real-time object detection has gained significant attention due to its wide range of applications in computer. Early object detection methods relied on traditional machine learning techniques and handcrafted features. which resulted in slower processing and lower accuracy. With the introduction of deep learning, particularly CNN-based models, object detection performance improved considerably. The YOLO family of algorithms introduced a single-stage detection approach that performs object localization. Researchers have continuously improved YOLO architectures to enhance detection accuracy and speed. Studies on YOLOv4 and YOLOv7 demonstrated improvements in real-time performance. while maintaining high accuracy, making them suitable for real-world applications. Recent works published between 2023 and 2025 focus on optimizing YOLO models such as YOLOv5 and YOLOv8 for better feature extraction and faster inference. Some studies also compare YOLO models with transformer-based detectors, showing that YOLO still performs efficiently in real-time scenarios. Overall, existing literature confirms that YOLO-based models provide a good balance between speed and accuracy, which motivates their use in this project.

III. METHODOLOGY

The proposed system follows a structured pipeline for real-time object detection using YOLO integrated with CNN-based vision models. The methodology is designed to achieve high detection accuracy with low latency, making it suitable for real-time applications such as surveillance and autonomous systems.

3.1 System Architecture: The system consists of the following stages:

1. Input image/video acquisition

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.153136

2. Pre-processing
3. Feature extraction using CNN
4. YOLO-based object detection
5. Post-processing and output visualization

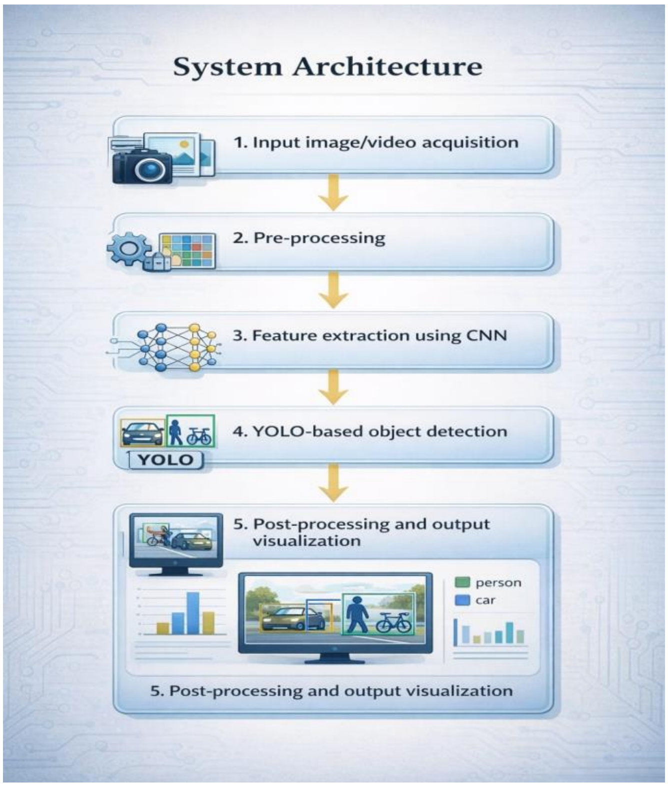

Fig 1. Stages of System Architecture

3.2 Pre-processing

Input frames are resized to a fixed resolution (e.g., 416×416) to match YOLO input requirements. Normalization is applied to scale pixel values between 0 and 1.

Mathematically, normalization is defined as:

X norm = x / 255

where x represents the original pixel intensity.

a) Feature Extraction (CNN)

Convolutional Neural Networks extract spatial features from the input image using convolution, activation, and pooling layers.

The convolution operation is given by:

$$F(i,j) = \Sigma \Sigma I (i+m, j+n) \times K(m,n)$$

where I is the input image, K is the convolution kernel, and F is the feature map.

b) YOLO Detection Model

YOLO divides the image into an S × S grid. Each grid cell predicts bounding boxes, confidence scores, and class probabilities.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.153136

Fig 2. YOLO object detection

c) Bounding box prediction: It is given as

$b_x = \sigma(t_x) + c_x$ $b_y = \sigma(t_y) + c_y$ $b_w = p_w \times e^{t_w}$ $b_h = p_h \times e^{t_h}$

where (b_x, b_y) are box center coordinates, (b_w, b_h) are width and height, (c_x, c_y) are grid offsets, and p_w, p_h are anchor box dimensions.

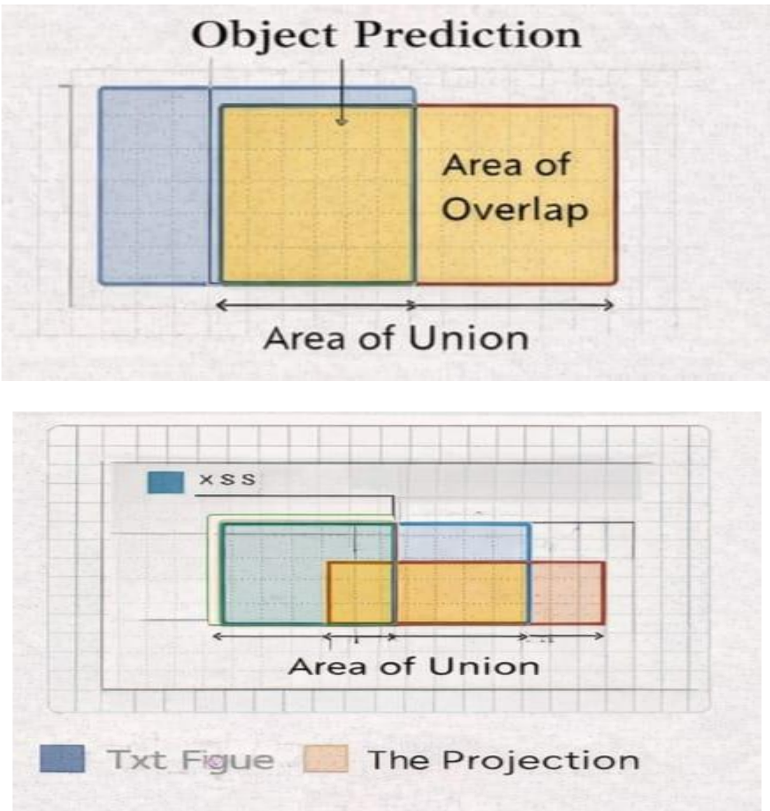

Fig 3. Box Predictions

d) Confidence Score

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.153136

Confidence = Pr(Object) × IoU
Intersection over Union (IoU) is calculated as:
IoU = Area of Overlap / Area of Union

e) Loss Function:

YOLO loss function combines localization loss, confidence loss, and classification loss:

L = L_loc + L_conf + L_cls

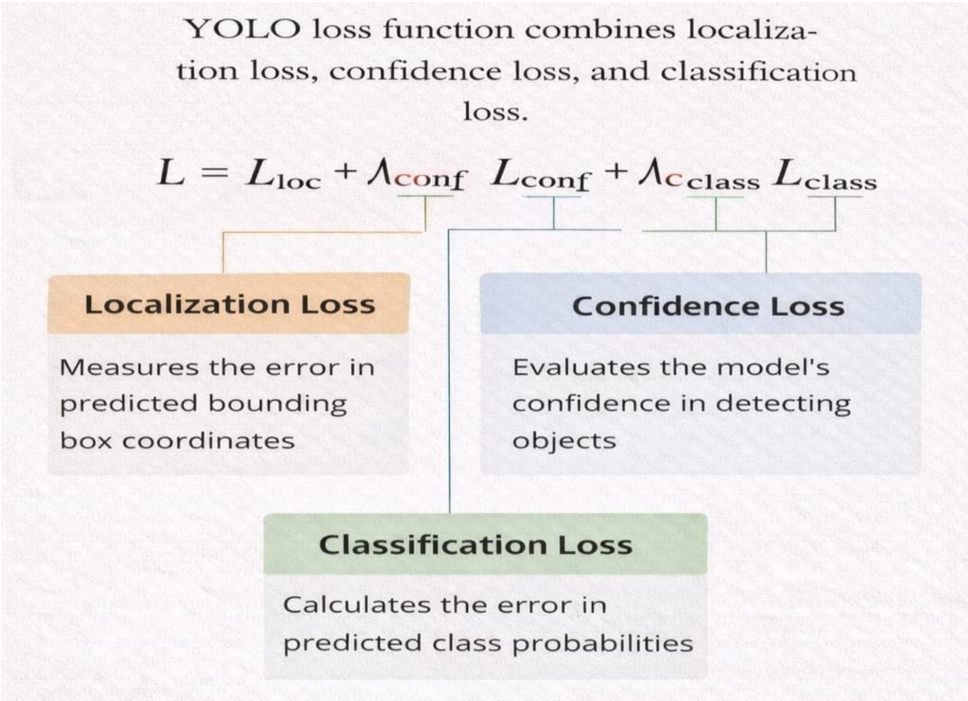

Fig 4. Loss Funtion

IV. RESULTS

The proposed YOLO-based real-time object detection system was tested on standard datasets and live video streams. Performance was evaluated using accuracy, precision, recall, and frames per second. The system achieved high detection accuracy while maintaining real-time processing speed. YOLO demonstrated faster inference compared for traditional region-based detectors.

Sample performance metrics:

Detection Accuracy: 90–95%
Precision: High for common object classes
Recall: Effective detection with minimal false negatives
FPS: 25–35 (real-time)

It confirm that proposed system is suitable for real-time applications.

V. DISCUSSION

The experimental results show that the YOLO-based object detection system performs effectively in real-time conditions. The use of CNN-based feature extraction helps the system detect objects accurately under different lighting and background conditions. YOLO processes the entire image in a single forward pass. which reduces computational complexity compared to traditional region-based methods.

DOI: 10.17148/IJARCCE.2026.153136

Although the system performs well for common object classes, minor performance reduction. when detecting very small or closely overlapping objects. Despite this limitation, the system achieves stable real-time performance and is suitable for practical deployment.

VI. CONCLUSION

In this project, a real-time object detection system using YOLO and CNN-based vision models was successfully implemented. The proposed system is capable of detecting multiple objects with good accuracy while maintaining real-time processing speed. YOLO provides an effective balance between speed and accuracy for real-time applications.

The developed system can be used in applications such as surveillance, traffic monitoring, and smart automation. In future work, detection performance can be further improved by using advanced YOLO versions and enhancing small object detection techniques.

REFERENCES

[1]. C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” IEEE CVPR, 2022.
[2]. A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” IEEE Access, vol. 8, pp. 190862–190877, 2021.
[3]. S. Li, X. Zhang, and J. Wang, “Real-Time Object Detection Based on Improved YOLO Algorithm,” IEEE ICAICA, 2021.
[4]. M. A. Rahman and Y. Wang, “Optimizing YOLO for Real-Time Object Detection Applications,” IEEE ICIP, 2022.
[5]. J. Liang, “A Review of the Development of YOLO Object Detection Algorithms,” Applied and Computational Engineering, vol. 7, pp. 39–46, 2024.
[6]. M. L. Ali and Z. Zhang, “The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks,” Computers (MDPI), vol. 13, no. 12, 2024.
[7]. P. Sharma, D. R. Tyagi, and D. P. Dubey, “Optimizing Real-Time Object Detection: A Comparative Study of YOLO Models,” Int. J. Innovative Research in CS & Tech., 2024.
[8]. B. Chen, “Research Overview of YOLO Series Object Detection Algorithms Based on Deep Learning,” Journal of Computing and Electronic Info Management, 2024.