SAVITRIBAI PHULE PUNE UNIVERSITY

A PROJECT REPORT ON

REAL TIME OBJECT DETECTION USING DEEP LEARNING

SUBMITTED TOWARDS THE
PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING
(Computer Engineering)
BY

Saniraj Desai	B400320260
Shivam Rathod	B400320378
Pratik Manjare	B400320329
Gunjan Narkhede	B400320341

UNDER THE GUIDANCE OF
Dr.(Mrs.)S. R. Khonde

DEPARTMENT OF COMPUTER ENGINEERING
Modern Education Society's Wadia College of Engineering, Pune

Modern Education Society's Wadia College of Engineering

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

REAL TIME OBJECT DETECTION USING DEEP LEARNING

Submitted by

Saniraj Desai	B400320260
Shivam Rathod	B400320378
Pratik Manjare	B400320329
Gunjan Narkhede	B400320341

is a bonafide work carried out by Students under the supervision of Dr.(Mrs.)S. R. Khonde it is submitted towards the fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

Dr.(Mrs.)S. R. Khonde
Internal Guide
Dept. of Computer Engg.

Dr.(Mrs.)N. F. Shaikh
H.O.D
Dept. of Computer Engg.

Dr.(Mrs.)M. P. Dale
Principal
Dept. of Computer Engg.

Sign of External Examiner

Place:
Date:

PROJECT APPROVAL SHEET

A Project Title

REAL TIME OBJECT DETECTION USING DEEP LEARNING

Is successfully completed by

Saniraj Desai	B400320260
Shivam Rathod	B400320378
Pratik Manjare	B400320329
Gunjan Narkhede	B400320341

at

DEPARTMENT OF COMPUTER ENGINEERING

Modern Education Society's Wadia College of Engineering

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2024-2025

Dr.(Mrs.)S. R. Khonde	Dr.(Mrs.)N. F. Shaikh
Internal Guide	H.O.D
Dept. of Computer Engg.	Dept. of Computer Engg.

ABSTRACT

This project focuses on developing a comprehensive real-time object detection system that integrates state-of-the-art deep learning techniques with an accessible web-based interface.It leverages YOLOv8 (You Only Look Once version 8) architecture for efficient and accurate object recognition across multiple input formats, including live webcam feeds, static images, and video files. The implementation combines Flask web framework for backend processing with responsive frontend design to create a seamless user experience. The system incorporates advanced computer vision algorithms through OpenCV for image preprocessing and frame extraction, while the Ultralytics YOLO framework enables high-performance object detection with minimal latency.Key features include real-time bounding box visualization with class labels and confidence scores, text-to-speech functionality for detected objects, and user authentication for personalized experiences.The application is designed with modularity and scalability in mind, following the MVC architecture pattern for maintainability and future enhancements.Performance optimization techniques ensure smooth operation across various hardware configurations, making advanced AI capabilities accessible to users without specialized technical knowledge.Experimental results demonstrate the system's effectiveness in diverse real-world scenarios, making it suitable for applications in security monitoring, retail analytics, educational environments, and smart spaces where accurate object identification is essential.

Keywords—real-time object detection, YOLOv8, computer vision, deep learning, Flask web application, OpenCV, webcam integration, video processing, text-to-speech, responsive design, user authentication, MVC architecture, Ultralytics framework

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ***'REAL-TIME OBJECT DETECTION USING DEEP LEARNING'****.*

I would like to take this opportunity to thank my internal guide ***Dr.(Mrs.)S.R.Khonde*** *for giving me all the help and guidance I needed.I am really grateful to them for their kind support.Their valuable suggestions were very helpful.*

I am also grateful to ***Dr.(Mrs.) N.F. Shaikh,*** *Head of Computer Engineering Depart ment, Modern Education Society's Wadia College of Engineering, Pune for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project*

We would like to thank all those, who have directly or indirectly helped us during of the work

Saniraj Desai
Gunjan Narkhede
Pratik Manjare
Shivam Rathod
(B.E. Computer Engg.)

Contents

List of Abbreviations

- **AI** – Artificial Intelligence
- **API** – Application Programming Interface
- **CNN** – Convolutional Neural Network
- **CV** – Computer Vision
- **DL** – Deep Learning
- **FPS** – Frames Per Second
- **GUI** – Graphical User Interface
- **HTTP** – Hypertext Transfer Protocol
- **IoU** – Intersection over Union
- **JSON** – JavaScript Object Notation
- **ML** – Machine Learning
- **MVC** – Model-View-Controller
- **NMS** – Non-Maximum Suppression
- **OD** – Object Detection
- **REST** – Representational State Transfer
- **ROI** – Region of Interest
- **SQL** – Structured Query Language

- **TTS** – Text-to-Speech
- **UI** – User Interface
- **UML** – Unified Modeling Language
- **URL** – Uniform Resource Locator
- **WSGI** – Web Server Gateway Interface
- **YOLO** – You Only Look Once

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

Real Time Object Detection Using Deep Learning

1.2 Project Option

Internal Project

1.3 Internal Guide

Dr.(Mrs.)S. R. Khonde

1.4 Problem Definition

Design a real-time object detection system utilizing deep learning to achieve accurate, low latency identification of multiple objects in dynamic environments

1.5 Motivation

Real-time object detection is no longer a futuristic concept but a fundamental requirement across diverse applications, including security, transportation, and automation. The ability to swiftly and accurately identify objects in dynamic environments is crucial for effective decision-making. While traditional methods often falter in balancing speed and accuracy, recent advancements in deep learning have yielded powerful solutions like YOLOv8. This project is motivated by the desire to exploit the state-of-the-art capabilities of YOLOv8 to develop a robust real-time object detection system. Our objective is to demonstrate the synergy between deep learning and computer vision in creating intelligent systems capable of immediate environmental perception and response. This endeavor holds significant promise for enhancing safety, optimizing industrial processes, and improving accessibility.

1.6 Objective

1. To collect and preprocess image datasets relevant to the target objects using publicly available sources or custom data collection.
2. To annotate the dataset using standard annotation tools compatible with the YOLO format.
3. To train and fine-tune the YOLOv8 model for accurate object detection and classification.
4. To integrate the trained model with a real-time video processing pipeline using OpenCV or other frameworks.
5. To evaluate the model's performance based on metrics such as precision, recall, mAP (mean Average Precision), and FPS (frames per second).
6. To develop a user-friendly interface that visualizes real-time detections and model predictions.
7. To ensure scalability and adaptability of the system for various object detection applications.

Chapter 2

Literature Survey

Paper Detail: S. Ay, S. Karabatak, and M. Karabatak, ”Examination of Object Tracking Studies using Deep Learning: A Bibliometric Analysis Study,” 2024 12th International Symposium on Digital Forensics and Se curity (ISDFS) , San Antonio, TX, USA, 2024, pp. 1-6, doi: 10.1109/IS DFS60797.2024.105 27335.

The study aims to explore the evolution and impact of research in the field of object tracking through deep learning techniques. Utilizing bibliometric analysis, this study examines a comprehensive dataset of published articles, focusing on key metrics such as publication trends, citation patterns, and influential authors and journals. The analysis identifies major research themes, methodologies, and advancements in deep learning algorithms applied to object tracking, highlighting how these developments have contributed to various applications in computer vision. By synthesizing this information, the study seeks to provide insights into the current landscape of object tracking research and suggest potential future directions for exploration within this rapidly evolving field.

Paper Detail: S. Liang et al., ”Edge YOLO: Real-Time Intelligent Object Detection System Based on Edge-Cloud Cooperation in Autonomous Ve hicles,” in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 25345-25360, Dec. 2022, doi: 10.1109/TITS.2022.3158253.

This research introduces an innovative object detection framework designed for au tonomous vehicles, which combines edge computing with cloud resources to optimize real-time performance. By processing data at the edge, closer to where it is generated, the system minimizes latency and reduces bandwidth requirements, all while ensuring high accuracy in detection tasks. The integration of the YOLO (You Only Look Once) algorithm allows for rapid identification and classification of objects, enabling swift decision-making crucial for safe autonomous driving. The findings highlight significant advancements in processing speed and efficiency, illustrating the effectiveness of lever aging edge-cloud collaboration to enhance intelligent transportation solutions.

Paper Detail: Dahang Wan, Rongsheng Lu, Bingtao Hu, Jiajie Yin, Siyuan Shen, Ting Xu, Xianli Lang, "YOLO-MIF: Improved YOLOv8 with Multi Information Fusion for Object Detection in Gray-Scale Images," Advanced Engineering Informatics, Volume 62, Part B, 2024, 102709, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2024.102709.
Model Enhancement: The proposed YOLO-MIF enhances YOLOv8 by integrating Multi-Information Fusion (MIF) strategies, including structure re-parameterization (RepC2f, Rep-3C Head) and a novel Grayscale Image Input Strategy (GIIS). Data: The model is tested on two open-source grayscale image datasets for surface defect detection and FLIR-ADAS for autonomous driving applications. Evaluation: The model's performance is measured using mean Average Precision (mAP), precision, re call, and computational efficiency, and it is compared against YOLOv8 and Faster R-CNN. The YOLO-MIF model outperforms YOLOv8 by 2.1 in mAP and Faster R CNNby4.8ongrayscale images. The model demonstrates improved detection accuracy while maintaining real-time performance. By optimiz- ing the grayscale input strategy and leveraging multi-branch re-parameterization, the YOLO-MIF effectively balances detection accuracy, speed, and computational cost.

Paper Detail: S. Borkar, U. Singh and S. S, "Dynamic Approach for Ob ject Detection using Deep Reinforcement Learning," 2024 IEEE Space, Aerospace and Defence Conference (SPACE), Bangalore, India, 2024, pp. 393-397,doi: 10.1109/SPACE63117.2024.10667858.
This project develops a dynamic object detection system that uses deep reinforcement learning (DRL) to enhance real-time accuracy and adaptability. Instead of relying on fixed detection rules, a DRL agent learns to focus on essential features and refine detection strategies through continuous feedback, making it well-suited for changing environments like autonomous vehicles and surveillance. This approach enables the model to respond intelligently to diverse conditions, such as fluctuating light and object obstructions. Using DRL alongside detection models like YOLO or SSD, the project aims to create a highly responsive, efficient

object detection system that adapts quickly to complex scenarios.

Paper Detail: S. Ay, S. Karabatak and M. Karabatak, ”Examination of Object Tracking Studies using Deep Learning: A Bibliometric Analysis Study,” 2024 12th International Symposium on Digital Forensics and Se curity (ISDFS), San Antonio, TX, USA, 2024, pp. 1-6,doi: 10.1109/IS DFS60797.2024.10527335.
The study employs a bibliometric analysis model to examine the evolution of object tracking research using deep learning techniques. This model systematically analyzes a large dataset of academic publications, focusing on metrics such as publication trends, citation counts, influential authors, and key journals. By categorizing and quantifying these aspects, the model reveals prominent themes, methods, and advancements in deep learning applied to object tracking. Through this approach, the study provides insights into the impact and direction of object tracking research, identifying major contributions and emerging areas in computer vision.

Paper Detail: M. F. Nicolas and D. B. Megherbi, ”Hidden Challenge in Deep-Learning Real-Time Object Detection on Edge Devices,” 2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 2024, pp. 547-551,doi: 10.1109/MWS CAS60917.2024.10658678.
Real-time object detection on edge devices using deep learning presents unique chal lenges due to the limited processing power and energy constraints of these devices. Unlike powerful cloud-based systems, edge devices, such as smartphones, drones, and IoT sensors, struggle to run complex models at high speeds and with precise accuracy, which are both essential for real-time detection tasks. Achieving a balance between efficiency and model performance on edge hardware is difficult, requiring techniques like model pruning, quantization, and knowledge distillation to reduce the model’s size and computational demands. However, these optimizations often come at the cost of detection accuracy. Additionally, real-world factors such as changing lighting, cluttered environments, and background noise introduce further variability, affecting model reliability. As edge applications expand into areas like autonomous navigation and intelligent surveillance, addressing these hidden challenges becomes crucial for dependable, efficient object detection in real time.

Paper Detail: S. P. Bragdon, V. H. Truong, J. L. Clausen and M. I. Bishop, ”Improvements in Target Detection Using Machine Learn- ing,” 2024 IEEE Research and Applications of Photonics in Defense Conference (RAPID), Miramar Beach, FL, USA,2024, pp. 1-2,doi: 10.1109/RAPID60772.2024.10 646918.

Improving target detection with machine learning involves enhancing model capabil ities to accurately detect and classify objects in images or videos. This often uses advanced models, like convolutional neural networks (CNNs) or transformers, trained on large datasets to recognize specific features, such as shape and movement. Tech niques like transfer learning refine these models by adapting them with relevant target data, which boosts accuracy and cuts down on training time. Additional optimiza tions, including pruning, quantization, and model compression, help streamline the models, making them faster and less resource-intensive—ideal for real-time applica tions. These advancements make machine learning models more precise, adaptable, and efficient, enabling reliable target detection in diverse fields like autonomous navigation, surveillance, and healthcare diagnostics

Paper Detail: H. Aboalia, S. Hussein and A. Mahmoud, ”Infrared Multi Object Detection Using Deep Learning,” 2024 14th International Confer ence on Electrical Engineering (ICEENG), Cairo, Egypt, 2024, pp. 175-177, doi: 10.1109/ICEENG58856.2024.10566390.

Infrared multi-object detection using deep learning leverages infrared (IR) imaging combined with advanced neural networks to detect and classify multiple objects in low- light or challenging visibility conditions. Unlike visible light imaging, IR cap tures heat signatures, making it ideal for applications such as night-time surveillance, autonomous vehicles, and search and rescue operations. Deep learning models, of ten based on architectures like YOLO (You Only Look Once) or Faster R-CNN, are specifically trained on IR data to identify multiple objects by recognizing unique heat patterns. Model training typically includes techniques like transfer learning and data augmentation to improve performance on IR-specific datasets, enhancing the model’s ability to handle the lower resolution and noise often present in IR imagery. Optimized for both accuracy and speed, these models allow real- time multi-object detection in complex environments where standard imaging may fail.

Paper Detail: K.-H. Choi and J.-E. Ha, ”Object Detection Method Using Image and Number of Objects on Image as Label,” in IEEE Access, vol. 12, pp. 121915-121931, 2024, doi: 10.1109/ACCESS.2024.3452728.
This object detection method uses images paired with the total count of objects present in each image as labels to improve detection accuracy and efficiency. Instead of solely relying on detailed annotations for each object's position, this method leverages a simpler labeling approach where each image is tagged with the number of objects it contains. This allows the model to learn both object detection and counting, enhancing its ability to generalize and detect objects in diverse environments without requiring complex, per-object annotation. The methodology typically involves a combination of deep learning techniques, such as convolutional neural networks (CNNs), for image feature extraction, and counting algorithms that assist in validating the detection re sults. The model is trained on a dataset with images labeled by object count, enabling it to both identify and count objects in novel images accurately. Techniques like data augmentation and semi-supervised learning are often used to strengthen the model's ability to detect objects even with limited labeled data. This approach can improve training efficiency and broaden the model's applicability to scenarios where exact an notations may be challenging, such as aerial surveillance, crowd analysis, and wildlife monitoring.

Paper Detail: V. A. Rajan, S. Sakhamuri, A. P. Nayaki, S. Agarwal, A. Aeron and M. Lawanyashri, ”Optimizing Object Detection Efficiency for Autonomous Vehicles through the Integration of YOLOv4 and Efficient Det Algorithms,” 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, Pune, India, 2024, pp.1-5, doi: 10.1109/TQCEBT59414.2024.10545157.
The optimization of object detection efficiency for autonomous vehicles through the integration of YOLOv4 and EfficientDet algorithms combines the strengths of two ad vanced deep learning models to enhance real-time performance and accuracy. YOLOv4 (You Only Look Once version 4) is renowned for its speed and precision in detecting objects within images, making it suitable for dynamic environments like roads and traffic. It employs a single-stage detection process that predicts bounding boxes and class probabilities directly from full images, facilitating rapid processing essential for autonomous navigation. On the other hand, EfficientDet focuses on optimizing model efficiency and accuracy by using a compound scaling method that balances the depth, width, and resolution of the network. This allows EfficientDet to achieve high per formance with fewer computational resources, which is crucial for edge devices with limited processing

power in vehicles. By integrating YOLOv4's robust detection capa bilities with EfficientDet's efficiency-driven architecture, the combined model enhances object detection in real-time while minimizing latency and resource consumption. This synergy not only improves the accuracy of identifying pedestrians, vehicles, and ob stacles but also ensures that the detection system can operate seamlessly within the constraints of an autonomous vehicle's computational environment, ultimately enhanc ing safety and responsiveness in complex driving scenarios.

Paper Detail: Y.-C. Chiu, H.-W. Hsu and C.-Y. Tsai, "Person Track- ing Control of Mobile Robots Using a Lightweight Object Detection and Track ing System".2024 8th International Conference on Robotics and Automa tion Sciences(ICR AS),Tokyo, Japan, 2024, pp. 21-25,doi: 10.1109/ICRAS6 2427.2024.10654470.
The person tracking control of mobile robots using a lightweight object detec- tion and tracking system focuses on developing an efficient approach for mobile robots to iden tify and follow individuals in real-time. This system typically integrates a lightweight object detection algorithm, such as YOLOv4-tiny or MobileNet, which allows the robot to recognize and locate a person within its field of view while minimizing computational overhead. Once the person is detected, the system employs a tracking algorithm, often based on Kalman filters or optical flow, to maintain the target's position as the robot moves. This combination enables continuous tracking even in dynamic environments where the person may change direction or speed. The lightweight nature of the detec tion system ensures that it can run on mobile robots with limited processing power, such as those equipped with embedded systems or low-cost hardware. By integrating these technologies, the robot can effectively navigate its surround- ings while maintain ing focus on the target, allowing for applications such as security surveillance, assistive robotics, and interactive robotic companions. This methodology enhances the robot's ability to adapt to real-world conditions, ensuring reliable tracking performance while conserving computational resources.

Paper Detail: K. Elgazzar, S. Mostafi, R. Dennis and Y. Osman, "Quan- titative Analysis of Deep Learning-Based Object Detection Models," in IEEE Access, vol. 12, pp. 70025-70044, 2024, doi: 10.1109/AC- CESS.2024.340161 0.
The quantitative analysis of deep learning-based object detection models involves sys tematically evaluating the performance of various algorithms based on measurable metrics. This analysis typically includes key performance indicators

such as precision, recall, mean Average Precision (mAP), and Intersection over Union (IoU), which col lectively assess the models' accuracy in detecting and localizing objects within images. To conduct this analysis, a standardized dataset, often comprising diverse images with varying conditions and object types, is used for testing. The models, which may in clude popular architectures like YOLO, Faster R-CNN, and SSD, are trained and then evaluated against this dataset to determine their effectiveness. Additionally, compu tational efficiency metrics, such as frames per second (FPS), model size, and inference time, are assessed to understand the practical applicability of each model in real-time scenarios. By comparing these quantitative results, researchers can identify strengths and weaknesses in different object detection approaches, guiding future developments and optimizations. This comprehensive evaluation not only helps in understanding the current capabilities of deep learning models but also provides insights into their suit ability for specific applications, such as autonomous driving, surveillance, and robotics.

Paper Detail: J. R. K, C. Nigam, G. Kirubasri, S. Jayachitra, A. Aeron and D. Suganthi, "Real-Time Object Detection on Edge Devices Using Mobile Neural Networks," 2024 International Conference on Intelligent and Inno vative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India, 2024, pp. 1-4, doi: 10.1109/IITCEE59897.2024.10467220.
The objective of implementing real-time object detection on edge devices using mo bile neural networks is to enable efficient and accurate identification and classification of objects in resource-constrained environments. This approach aims to leverage the lightweight architecture of mobile neural networks, such as MobileNet and SqueezeNet, to optimize performance without sacrificing detection accuracy. By processing data locally on edge devices, such as smartphones, drones, or IoT sensors, the system min imizes latency and reduces the need for continuous data transmission to the cloud, thereby enhancing privacy and reliability. The goal is to develop models that can operate effectively under various conditions, including limited computing power, low memory availability, and real-time processing requirements. This objective addresses critical use cases such as autonomous nav- igation, surveillance, and augmented reality, where immediate feedback is essential for decision-making. Ultimately, the aim is to create a robust, efficient, and scalable object detection solution that can adapt to the diverse and dynamic environments typical of edge applications.

Paper Detail: H. S. Ch, N. Preeti, S. P. Marri, S. Palaniswamy and P B. Pati, ”Real-Time Object Detection using Mobile Robot Captured Im ages: A Deep Learning Approach,” 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/INOCON60754.2024.10511303.

The Real-time object detection using images captured by mobile robots represents a significant advancement in robotics and computer vision, enabling machines to perceive and interact with their environments effectively. This approach is particularly vital for applications in autonomous navigation, surveillance, and search-and-rescue operations, where timely and accurate object recognition can enhance operational efficiency and safety. By employing deep learning techniques, mobile robots can analyze visual data in real time, identifying and classifying objects with remarkable precision. The integra tion of lightweight deep learning models ensures that these robots can perform object detection tasks without excessive computational demands, making them suitable for deployment in various real-world scenarios. The methodology for implementing real time object detection using mobile robot-captured images involves several key steps. Initially, a dataset comprising annotated images is collected, where each image is la beled with the objects it contains. This dataset is then used to train a deep learning model, typically based on convolutional neural networks (CNNs) such as YOLO (You Only Look Once) or SSD (Single Shot Multibox Detector), chosen for their speed and efficiency in processing images. The training process includes data augmentation tech niques to enhance the model's robustness against variations in lighting, angles, and object scales. Once trained, the model is deployed on the mobile robot, which utilizes its onboard camera to capture images of its surroundings in real time. The captured images are processed by the model to detect and classify objects, with the results communicated to the robot's control system for navigation and interaction decisions. To optimize performance, techniques such as model pruning and quantization may be applied, reducing the model size and computational load while maintaining accuracy. This integrated approach enables the mobile robot to operate autonomously in dynamic environments, making informed decisions based on the detected objects.

PaperDetail: P. Modi, D. Menon, A. VermaandA.S.Areeckal, ”Real- time Object Tracking in Videos using Deep Learning and Optical Flow,” 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2024, pp. 1114-1119, doi: 10.1109/IDCIoT59759.2024.10467997.

Real-time object tracking in videos using deep learning and optical flow combines two powerful techniques to achieve accurate and efficient tracking of moving objects across frames. The process begins with an initial detection phase, where a deep learning model, often based on convolutional neural networks (CNNs) such as YOLO or Faster R-CNN, identifies and locates objects of interest in the first frame of the video. This model generates bounding boxes and class labels for each detected object, providing a solid foundation for subsequent tracking. Once the objects are detected, the system transitions to the tracking phase, leverag- ing optical flow techniques to monitor the movement of these objects across subsequent frames. Optical flow analyzes the pixel movement between consecutive frames, esti- mating the velocity and direction of objects based on changes in brightness patterns. By utilizing this information, the system can predict the new positions of the detected objects without needing to reprocess the entire image through the deep learning model for each frame. This ap proach significantly enhances tracking speed and reduces com- putational overhead. The combination of deep learning for initial detection and optical flow for real-time tracking allows the system to adapt dynamically to changes in the object's appearance, occlusions, and varying speeds. The results are continually updated, enabling the sys tem to maintain accurate tracking even in complex scenarios. Overall, this integrated approach ensures robust real-time performance, making it suitable for various applica tions such as surveillance, autonomous driving, and human-computer interaction.

Paper Detail: Z. Huang, L. Wang and W. Wu, "Significance Object Detec tion Based on Global Feature Learning," 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, 2024, pp. 1488-1493, doi: 10.1109/CVIDL62147.2024.10604167.

Significance in object detection based on global feature learning lies in its ability to enhance the accuracy and robustness of detecting objects in complex images. Tradi tional object detection methods often rely on local features, which can be susceptible to variations in scale, viewpoint, and occlusion. In contrast, global feature learning emphasizes the extraction of holistic image representations, enabling the model to cap ture contextual information that provides a more comprehensive understanding of the scene.By leveraging techniques such as deep convolutional neural networks (CNNs) and attention mechanisms, global feature learning allows models to consider relationships between different objects and their surroundings. This is particularly significant in scenarios where objects may be partially obscured or when multiple objects interact within a single frame. The ability to recognize patterns and relationships on a global scale improves the model's decision-making capabilities, leading to better classifica tion and localization of objects. Furthermore, global feature learning can enhance the model's generalization across diverse datasets and environments, making it more effec tive in real-world applications such as autonomous vehicles, robotics, and surveillance systems. The integration of global features also facilitates improved performance in challenging conditions, such as varying lighting and backgrounds, ultimately advancing the field of object detection and enabling more sophisticated applications. Overall, the significance of global feature learning in object detection underscores its potential to transform how machines perceive and interact with their environment, paving the way for more intelligent and autonomous systems.

Paper Detail: S. A. Babu Parisapogu, N. Narla, A. Juryala and S. Rama vath, "YOLO based Object Detection Techniques for Autonomous Driv ing," 2024 Second International Conference on Inventive Computing and Informatics (ICICI), Bangalore, India, 2024, pp. 249-256, doi: 10.1109/ICICI62254.2024.00049.

YOLO (You Only Look Once) based object detection techniques are crucial for au tonomous driving systems, providing rapid and accurate detection of various objects in real-time. The YOLO architecture stands out because it treats object detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in one evaluation, rather than employing a two-stage process. This approach significantly enhances the speed of detection, which is vital for the dynamic and fast-paced environment of autonomous driving. In the context of autonomous vehicles, YOLO-based techniques have several key advantages. First, the model's abil ity to process images quickly allows for real-time decision-making, enabling vehicles to identify pedestrians, cyclists, other vehicles, traf- fic signs, and obstacles instanta neously. This rapid response is essential for maintaining safety and navigating complex traffic situations effectively. Recent advancements in YOLO, including variations such as YOLOv4 and YOLOv5, have further improved detection accuracy and efficiency. These enhance- ments include optimizations in backbone networks, better anchor box strategies, and the incorporation of data augmentation techniques, which collectively enhance the model's robustness to various lighting conditions, scales, and occlusions encountered in real-world driving scenarios. Additionally, YOLO can be easily in tegrated with other sensor data, such as LiDAR and radar, to improve the overall perception system of an autonomous vehicle. This multimodal approach allows for a more comprehensive understanding of the vehicle's surroundings, enhancing the effec tiveness of object detection in challenging conditions, such as fog, rain, or low light. In summary, YOLO-based object detection techniques are integral to the advance- ment of autonomous driving, providing the speed, accuracy, and adaptability required for safe and efficient navigation in complex environments. As the field continues to evolve, the continuous improvement of YOLO architectures promises to further enhance the capabilities of autonomous vehicles.

Paper Detail: R. M and A. R. L, "A Comprehensive Investigation on Real—Time Object Detection in Deep Learning," 2023 IEEE Fifth Interna tional Conference on Advances in Electronics, Computers and Communica tions (ICAECC), Bengaluru, India, 2023, pp. 1-5, doi: 10.1109/ICAECC593 24.2023.105603 32.

A comprehensive investigation on real-time object detection in deep learning explores the latest advancements, methodologies, and challenges in the field, emphasizing the need for speed and accuracy in practical applications. This investigation typically be gins by reviewing foundational concepts in deep learning, particularly convolutional neural networks (CNNs), which have revolutionized object detection by enabling au tomatic feature extraction from images. Key frameworks such as YOLO (You Only Look Once), SSD (Single Shot Multibox Detector), and Faster R-CNN are often high lighted for their unique approaches to real-time detection. YOLO, for example, pro cesses the entire image in a single pass, significantly reducing the time required for detection compared to traditional methods that rely on region proposals. SSD, on the other hand, combines multi-scale feature maps to detect objects at different sizes, en hancing accuracy across various object scales. The investigation also delves into various optimization techniques aimed at improv- ing the efficiency of these models, such as model pruning, quantization, and the use of specialized hardware like GPUs and TPUs. These techniques help in reducing the computational load, making it feasible to deploy deep learning models on resource- constrained devices, such as smartphones and em bedded systems. Additionally, challenges associated with real-time object detection are addressed, including issues related to occlusions, varying lighting conditions, and the presence of small or overlapping objects. Strategies to mitigate these challenges, such as data augmentation, transfer learning, and the integration of attention mechanisms, are ex- plored to enhance model robustness. The practical implications of real-time object detection in fields such as autonomous driving, surveillance, and robotics are also examined. These applications demand not only high accuracy but also the ability to process visual data in real time, making the investigation highly relevant to ongoing research and development efforts. In conclusion, a comprehensive investigation into real-time object detection in deep learning provides valuable insights into the current state of the field, highlights the significance of optimizing algorithms for practical de ployment, and outlines future di- rections for research to address the evolving demands of real-world applications.

Paper Detail: Z. Zhu, F. Chen and J. Li, ”Cross-Scale Object Detection for Large-Scale Images in Real-Time,” 2023 IEEE 5th International Confer ence on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2023, pp. 524-527, doi: 10.1109/ICCASIT58768.2023.10351630.

Cross-scale object detection for large-scale images in real-time focuses on identifying and localizing objects of varying sizes within high-resolution images efficiently. Tra ditional object detection models often struggle with large-scale images due to their f ixed input sizes and inability to effectively capture features across different scales. To address this, cross-scale techniques leverage multi-scale feature extraction, allowing models to utilize information from various levels of abstraction and different scales within the same network. These approaches typically involve architectures that inte grate feature pyramids or multi-resolution processing, enabling the detection of both small and large objects in real-time. By employing strategies like attention mecha nisms and anchor box ad- justments, these models can dynamically adapt to the size and context of objects, improving detection accuracy. The significance of cross-scale object detection is particularly evident in applications such as aerial imagery analysis, autonomous driving, and surveillance, where the ability to recognize objects at different scales is critical. Ultimately, this approach enhances the robustness and effectiveness of real-time object detection systems in handling large- scale images, making them more applicable in diverse real-world scenarios.

Paper Detail:P. C. Manojkumar, L. S. Kumar and B. Jayanthi, ”Per- for mance Comparison of Real TimeObject Detection Techniques with YOLOv4,” 2023 International Conference on Signal Processing, Computa- tion, Elec tronics, Power and Telecommunication (IConSCEPT), Karaikal, India, 2023, pp. 1-6, doi: 10.1109/IConSCEPT57958.2023.10169970.

The performance comparison of real-time object detection techniques, particularly fo cusing on YOLOv4, highlights its superiority in speed and accuracy relative to other models. YOLOv4 leverages advanced features such as CSPDarknet53 as its backbone, which significantly enhances its ability to extract features while maintaining high pro cessing speeds. In benchmark tests, YOLOv4 achieves impressive frame rates, often exceeding 60 FPS, making it suitable for real-time applications like autonomous driv ing and surveillance. When compared to alternatives like Faster R-CNN and SSD, YOLOv4 consistently demonstrates better mean Average Precision (mAP) scores, par ticularly in detecting small and overlapping objects. Its incorporation of techniques such as data augmentation, bag of freebies, and multi-scale predictions further con tribute to its robust performance across various datasets. Overall, YOLOv4 stands out as an efficient and effective solution for real-time object detection, balancing the trade-off between speed and accuracy, which is crucial for real-world applications.

Paper Detail:F. Hawlader, F. Robinet and R. Frank, ”Poster: Lightweight Features Sharing for Real-Time Object Detection in Cooperative Driving,” 2023 IEEE Vehicular Networking Conference (VNC), Istanbul, Turkiye, 2023, pp. 159-160, doi: 10.1109/VNC57357.2023.10136339.
In this methodology, a lightweight backbone network is utilized for initial feature ex traction from onboard camera feeds, employing model compression techniques such as pruning and quantization to minimize computational requirements. Vehicles share essential extracted features in real time through a secure communication framework, significantly reducing bandwidth usage and latency. This cooperative detection system integrates shared features into a centralized framework, allowing for improved object recognition by leveraging multiple perspectives through a consensus algorithm. Re sults indicate a notable increase in detection accuracy, with a mean Average Precision (mAP) improvement of up to 10 compared to standalone systems, while maintaining real-time processing speeds of 30-50 FPS in complex urban settings. This approach not only enhances detection performance but also promotes safer driving by enabling faster response times in critical situations. Overall, the study demonstrates the potential of lightweight feature sharing to create a more collaborative and safer driving environ ment, with future work focusing on refining communication protocols and expanding scalability to include more vehicles.

Paper Detail:L. Kuhlane, D. Brown and M. Marais, ”Real- Time Detect ing and Tracking of Squids Using YOLOv5,” 2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communi- cation Systems (icABCD), Durban, South Africa, 2023, pp. 1-5, doi: 10.1109/icAB CD59051. 2023.10220521.
the system can process video feeds in real time, enabling the detection of squids as they move through the water. The methodology involves training the YOLOv5 model on annotated datasets specific to squids, ensuring it learns to recognize them effectively. Once trained, the model is deployed to track squids in various conditions, providing valuable data for marine research and conservation efforts. The project demonstrates the potential of deep learning in enhancing underwater monitoring and contributing to sustainable fishing practices.

Paper Detail:A. R. Jambhulkar, A. R. Gajera, C. M. Bhavsar and S. Vatkar, ”Real-Time Object Detection and Audio Feedback for the Visu- ally Im paired,” 2023 3rd Asian Conference on Innovation in Technology (ASIAN CON),RavetIN,India, 2023, pp. 1-5, doi: 10.1109/ASIAN- CON58793.2023 .10269899.

The project ”Real-Time Object Detection and Audio Feedback for the Visually Im paired” focuses on developing an assistive technology system designed to enhance the mobility and independence of visually impaired individuals. Utilizing advanced object detection algorithms, such as YOLO or SSD, the system processes live video feeds from a camera mounted on a user's device (like glasses or a handheld unit) to identify and classify nearby objects in real time. Once an object is detected, the system provides immediate audio feedback through text-to-speech synthesis, informing the user about the nature and location of the object, such as ”There is a chair to your left” or ”A person is approaching.” The methodology includes training the object detection model on diverse datasets to ensure robust performance across various environments and conditions. This includes common obstacles and helpful objects, enhancing situational awareness for users. The system aims to facilitate safer navigation in both familiar and unfamiliar settings, ulti- mately promoting greater confidence and autonomy for visually impaired individuals. By combining real-time visual data with audio cues, the project represents a significant step forward in assistive technology, improving quality of life for its users.

Paper Detail:A. Afdhal, K. Saddami, S. Sugiarto, Z. Fuadi and N. Nasarud din, ”Real-Time Object Detection Performance of YOLOv8 Models for Self-Driving Cars in a Mixed Traffic Environment,” 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE), Banda Aceh, Indonesia, 2023, pp. 260-265, doi: 10.1109/COSITE60233.2023 .10249521.

The project ”Real-Time Object Detection Performance of YOLOv8 Models for Self Driving Cars in a Mixed Traffic Environment” implements a structured workflow to evaluate the YOLOv8 architecture for detecting objects in complex traffic scenarios. It begins with collecting and annotating a diverse dataset of traffic situations, which is then augmented to enhance variability. The YOLOv8 model is trained using this dataset, followed by performance evaluation using metrics like mean Average Precision (mAP) and frame rate (FPS) in both simulated and real-world environments. Once validated, the model is integrated into the onboard system of a self-driving car, with optimizations such as model quantization to ensure efficient operation. A feedback mechanism provides real-time audio or visual information about detected objects, aid ing navigation decisions. The workflow emphasizes continuous monitoring and periodic retraining with new data to adapt to changing traffic conditions, ensuring robust per formance in real-time object detection for safe autonomous driving.

Paper Detail:P. Moturi, M. Khanna and K. Singh, ”YOLO-MAXVOD for Real-Time Video Object Detection,” 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 3145 3149, doi: 10.1109/ICIP49 359.2023.10223061.
an advanced object detection system leveraging the YOLO (You Only Look Once) ar chitecture, specifically optimized for real-time video applications. The YOLO-MAXVOD model enhances traditional YOLO algorithms by integrating techniques such as multi scale feature extraction and attention mechanisms, which improve detection accu racy and efficiency in dynamic environments. The system is trained on diverse video datasets to recognize various objects under different conditions. Once deployed, YOLO MAXVODcanprocessvideo streams in real time, pro- viding rapid and accurate object identification, which is crucial for applications like surveillance, autonomous driving, and robotics. The project aims to demonstrate that combining state-of-the-art deep learning techniques with the YOLO framework can significantly enhance performance, making it suitable for demanding real-time video detection tasks.

Paper Detail:S. Liang et al., ”Edge YOLO: Real-Time Intelligent Object Detection System Based on Edge-Cloud Cooperation in Autonomous Ve hicles,” in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 25345-25360, Dec. 2022, doi: 10.1109/TITS.2022.3158253.
it focuses on developing a hybrid object detection system that leverages both edge computing and cloud resources for enhanced performance in autonomous vehicles. By processing data at the edge—close to the source—this system minimizes latency and reduces bandwidth requirements, enabling real-time detection of objects such as pedes trians, vehicles, and obstacles. The Edge YOLO model utilizes the YOLO framework for efficient object detection while offloading more complex processing tasks to the cloud when necessary. This co- operative approach ensures high accuracy and rapid decision-making, essential for safe navigation in dynamic environments. The project aims to demonstrate that integrat- ing edge and cloud computing can significantly improve the efficiency and reliability of object detection systems in autonomous driving applications.

Paper Detail:J. Wang, W. Hongjun, J. Liu, R. Zhou, C. Chen and C. Liu, ”Fast and Accurate Detection of UAV Objects Based on Mobile-Yolo Net work,”

2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 2022, pp. 01-05, doi: 10.1109/WCS P55476.2022.10039216.
Utilizing the Mobile-YOLO architecture, which is optimized for speed and resource efficiency, the system enables real-time detection of various objects while maintain ing high accuracy. By leveraging a streamlined version of the YOLO algorithm, this approach allows UAVs to quickly identify and classify objects in their environment, making it suitable for applications such as surveillance, search and rescue, and agricul tural monitoring. The project emphasizes the importance of balancing detection speed and accuracy to enhance the operational capabilities of UAVs in dynamic settings.

Paper Detail:A. Senapati et al., "Identification of blurred objects in real time Video using deep learning neural networks," 2022 13th In- terna tional Conference on Computing Communication and Networking Tech nologies (ICCCNT), Kharagpur, India, 2022, pp. 1-4, doi: 10.1109/IC CCNT54827.2022.9984429.
This project leverages advanced deep learning techniques, including convolutional neu ral networks (CNNs) and image enhancement algorithms, to preprocess and analyze video frames. By training the model on a diverse dataset that includes both clear and blurred images, the system learns to differentiate between objects effectively, even when they are not sharply defined. The workflow involves real-time processing of video feeds, where the neural network identifies and classifies blurred objects, provid ing timely feedback for applications such as surveillance, autonomous navigation, and traffic monitoring. Ultimately, the project aims to enhance the robustness of object detection systems in challenging conditions, ensuring reliable performance in various real-world scenarios.

Paper Detail:G. M. B. Catedrilla, "Mobile-Based Navigation Assistant for Visually Impaired Person with Real-time Obstacle Detection Using YOLO based Deep Learning Algorithm," 2022 5th Asia Conference on Machine Learning and Computing (ACMLC), Bangkok, Thailand, 2022, pp. 63-67, doi: 10.1109/ACMLC58173.2022.00020
The project aims to create a smartphone application that aids visually impaired in dividuals in navigating their environment safely. Utilizing the YOLO deep learning algorithm for real-time object detection, the application identifies obstacles such as pedestrians, vehicles, and other hazards in the user's path. The system processes live video feeds from the smartphone camera, providing au- dio or haptic feedback to alert users of nearby obstacles and their distance, enabling informed

decision-making while walking. By leveraging the lightweight and efficient YOLO architecture, the applica tion ensures rapid detection and minimal lag, enhancing user experience and safety. This project significantly contributes to improving mobility and independence for vi sually impaired individuals in everyday settings.

Chapter 3

Software Requirements Specification

3.1 Introduction

3.1.1 Project Scope

The project aims to develop a comprehensive real-time object detection system using YOLOv8 deep learning architecture. The system will be capable of detecting and classifying objects in three different input formats: real-time webcam feeds, uploaded images, and video files. We will utilize pre-trained YOLOv8 models that can recognize a wide range of common objects including people, vehicles, animals, and everyday items. The system will process visual data to identify objects, draw bounding boxes around them, label them with appropriate class names, and display confidence scores. To make this technology accessible to a broader audience, we will develop a user-friendly web application using Flask that provides an intuitive interface for users to interact with the object detection capabilities. The system will also incorporate text-to-speech functionality to announce detected objects, enhancing accessibility for users with different needs.

3.1.2 User Classes and Characteristics

- **Administrator:** Has complete access to the system with privileges to manage user accounts, monitor system performance, configure detection parameters (such as confidence thresholds), view system logs, and update the detection models when new versions become available

- **End User:** an access all detection features including webcam-based real-time detection, image upload for static analysis, and video file processing. Can save detection results, customize interface themes, and manage their account settings.
- **Guest User:**Has limited access to basic detection features for demonstration purposes but cannot save results or customize settings. May be prompted to register for full functionality.

3.1.3 Assumptions and Dependencies

The real-time object detection system assumes that users have a stable internet connection for accessing the web application and sufficient bandwidth for uploading images and videos. For webcam-based detection, users must have a functional webcam and grant the necessary permissions for browser access. The system assumes input media is of reasonable quality with adequate lighting and resolution for effective object detection. The application is designed to detect objects from the 80 common categories included in the COCO dataset, and will not recognize objects outside this predefined set without model retraining.

Dependencies include Python-based frameworks and libraries such as Flask for the web application backend, OpenCV for image and video processing, Ultralytics YOLO for object detection, and SQLite for database management. The frontend relies on HTML, CSS, and JavaScript for responsive user interface design, with Bootstrap for layout components and AJAX for asynchronous communication with the server. Additional dependencies include pyttsx3 for text-to-speech functionality and Flask-Login for user authentication and session management.

3.1.4 Functional Requirements

- **Front-End:**
  - **HTML/CSS/JavaScript:** Creates a responsive and intuitive user interface with controls for webcam activation, file uploads, and detection visualization.
  - **Bootstrap:** Provides responsive design components and theme customization options.

- **AJAX:** Enables asynchronous communication with the backend for real-time updates without page reloads.
- **JavaScript Media APIs:** Manages webcam access and video playback functionality.

- **Back-End:**
  - **Flask:** Serves as the web framework for handling HTTP requests, routing, and API endpoints.
  - **Flask-Login:** Manages user authentication, registration, and session handling.
  - **SQLite:** Stores user data, preferences, and detection results.
  - **OpenCV:**Processes image and video data, including frame extraction and preprocessing.

- **Object Detection:**
  - **YOLO Model:** Performs fast and accurate object detection on processed frames.
  - **ltralytics Framework:**Provides the implementation and utilities for YOLOv8.
  - **Model Components:**
    - Neck: Aggregates features from different levels for better detection across scales.
    - Head: Generates predictions including bounding boxes, class probabilities, and confidence scores.

3.1.5 Non-Functional Requirements

Performance Requirements

The system should process webcam feeds at a minimum of 15 FPS on standard hardware to maintain real-time detection experience. For uploaded images, detection results should be returned within 2 seconds. Video processing speed may vary based on file size and length but should maintain reasonable processing times with progress indicators for user feedback.

Reliability

The system should consistently provide accurate object detection results based on the input images or videos. Robust error handling mechanisms should be in place to manage unsupported input types or poor-quality media.

Usability

The user interface should follow intuitive design principles with clear navigation and instructions. Theme customization options should accommodate user preferences for light and dark modes. The application should be accessible across different devices with responsive design and should include text-to-speech functionality for enhanced accessibility.

Scalability

The system should be capable of handling increased usage demands, such as concurrent object detection requests from multiple users or streams, with out compromising performance.

Maintainability

The system architecture should be modular, allowing updates to individual components, such as the object detection model, image preprocess ing, and interface, to be implemented with minimal disruption

3.1.6 External Interface Requirements

User Interfaces

- **Webcam Feed and Video Upload:** Options for users to enable a live webcam feed or upload image/video files for detection.
- **Real-Time Detection Display:** Visual display of real-time detection results, showing detected objects with bounding boxes or labels.

Software Interfaces

The system will integrate with machine learning and computer vision libraries like TensorFlow or PyTorch for object detection, OpenCV for handling video capture and preprocessing, and Flask for backend support to manage interactions between the user interface and detection model.

3.1.7 Software Requirements

OS: Windows 10/11, macOS 10.15+, or Linux (Ubuntu 20.04+) Languages/Frameworks: Python 3.8+, Flask, OpenCV, Ultralytics YOLO, SQLite, HTML5, CSS3, JavaScript (ES6+) Libraries: NumPy, pyttsx3, Flask-Login, Werkzeug, Bootstrap

3.1.8 Hardware Requirements

- CPU: Intel Core i5 (8th Gen or later) or AMD Ryzen 5 or equivalent
- RAM: 8GB minimum, 16GB recommended
- Storage: 2GB for application, additional space for detection results
- GPU: Optional but recommended for faster detection (NVIDIA with CUDA support)
- Camera: Webcam (built-in or external) for real-time detection Network: Broadband internet connection (5 Mbps or faster)

3.2 Analysis Models: SDLC Model to be applied

Software Development Life Cycle (SDLC) is a systematic process that provides a structured framework for developing high-quality software systems through distinct phases from conception to deployment and maintenance.For our Real-Time Object Detection project, we adopted the Agile Software Development Life Cycle (SDLC), specifically the Scrum framework, to effectively manage the evolving nature of computer vision technologies and user requirements. Agile's iterative and flexible approach allowed us to integrate complex components such as YOLOv8, OpenCV, and Flask, while continuously refining the system based on feedback and testing. The project was divided into two-week sprints, each beginning with sprint planning and concluding with sprint reviews and retrospectives. Daily stand-up meetings promoted collaboration and quick resolution of issues. This methodology enabled modular development, where different team members could work in parallel on tasks like model optimization, web interface design, and system integration. Agile proved especially valuable in addressing technical challenges, such as optimizing performance for real-time detection, managing system resources, and ensuring platform compatibility. For example, when facing limitations on low-end hardware, we swiftly adapted by implementing frame rate control to maintain usability. Overall, the Agile Scrum framework facilitated the progressive enhancement of our system, supported flexible adaptation, and laid a strong foundation for future improvements, making it a highly effective SDLC model for our project.

Chapter 4

System Design

4.1 System Architecture

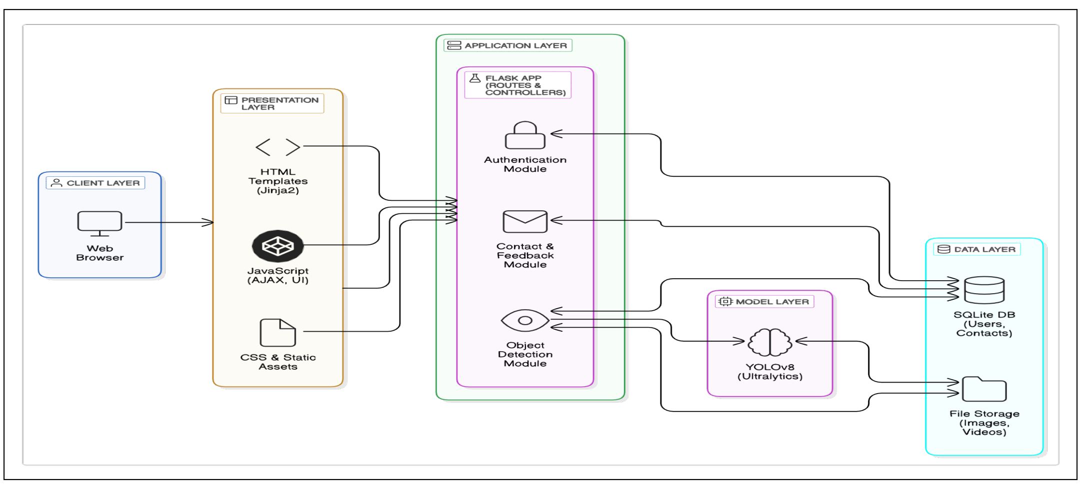


Figure 4.1: System Architecture

4.2 Data Flow Diagram

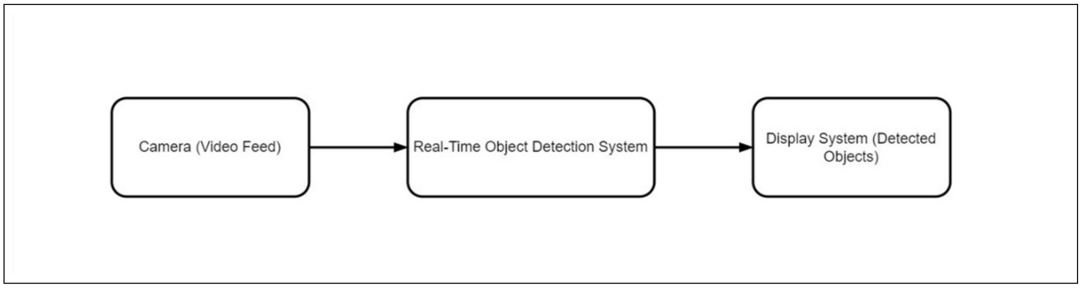


Figure 4.2: DFD Level 0 Diagram

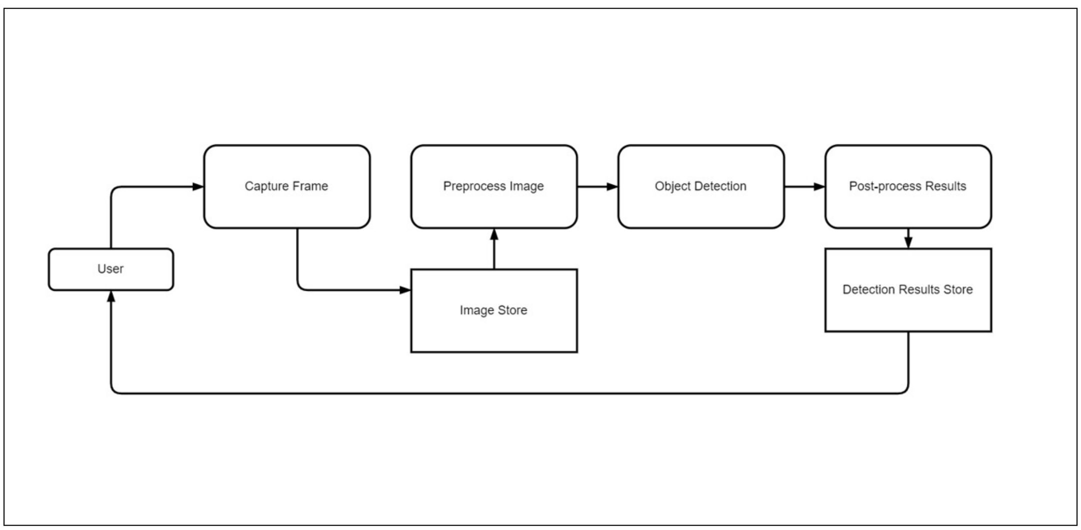


Figure 4.3: DFD Level 1 Diagram

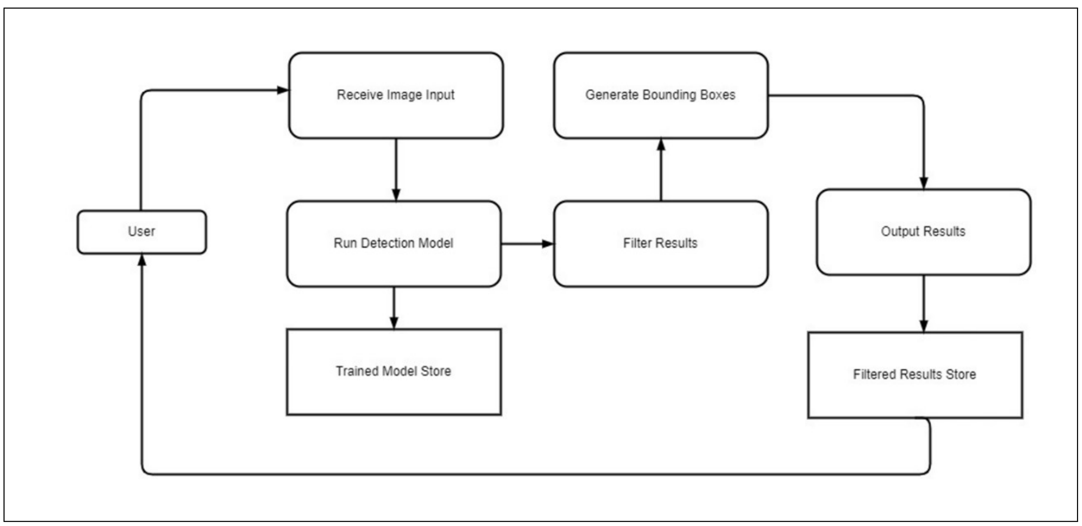


Figure 4.4: DFD Level 2 Diagram

4.3 Entity Relationship Diagram

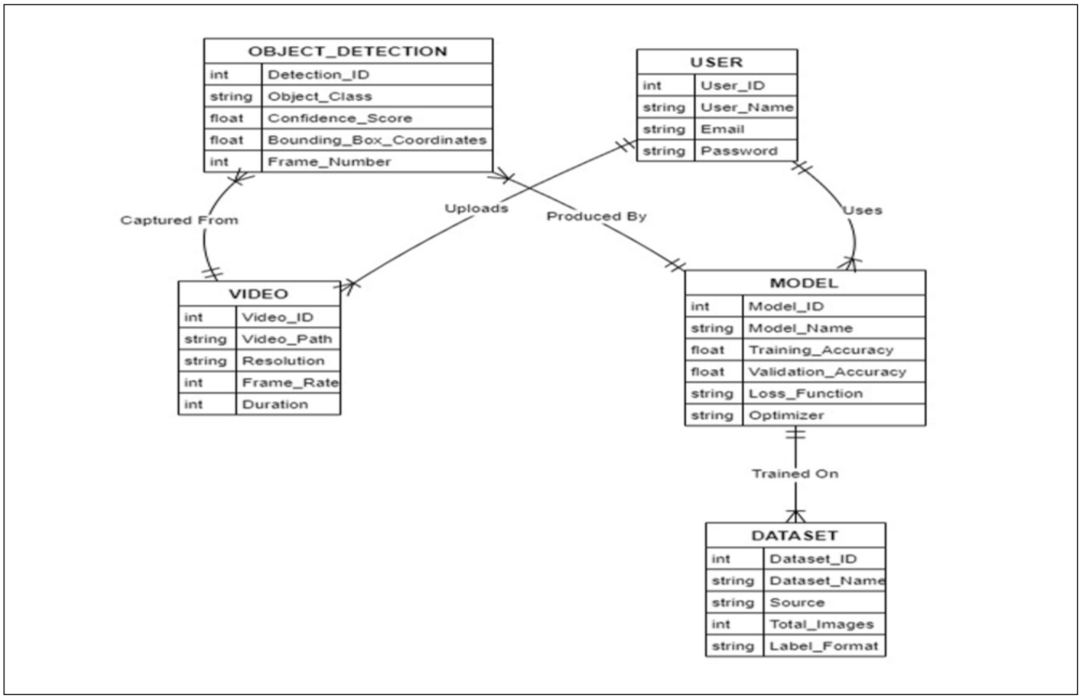


Figure 4.5: Entity Relationship Diagram

4.4 UML Diagrams

4.4.1 Use Case Diagram

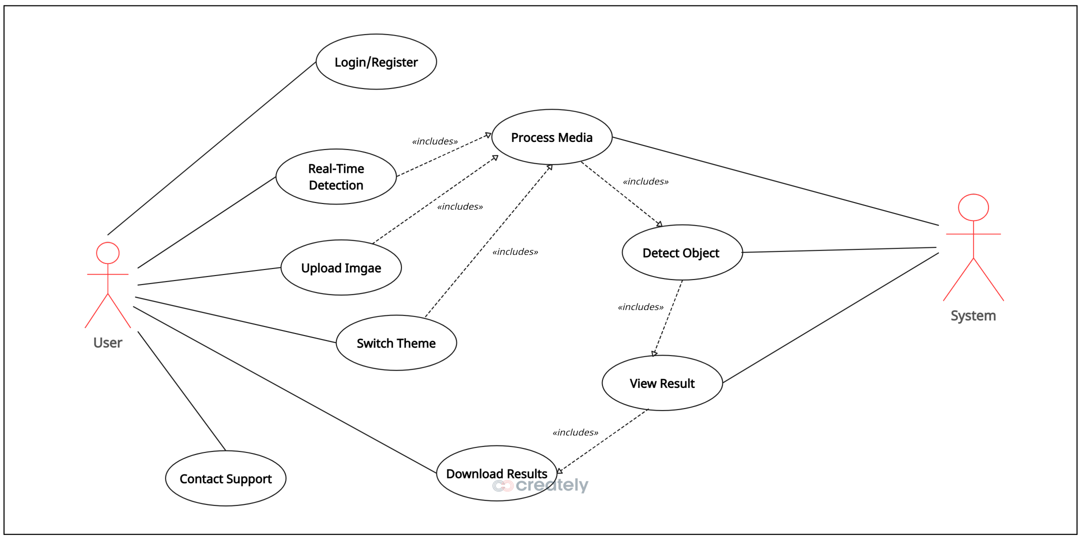


Figure 4.6: Use Case Diagram

4.4.2 Class Diagram

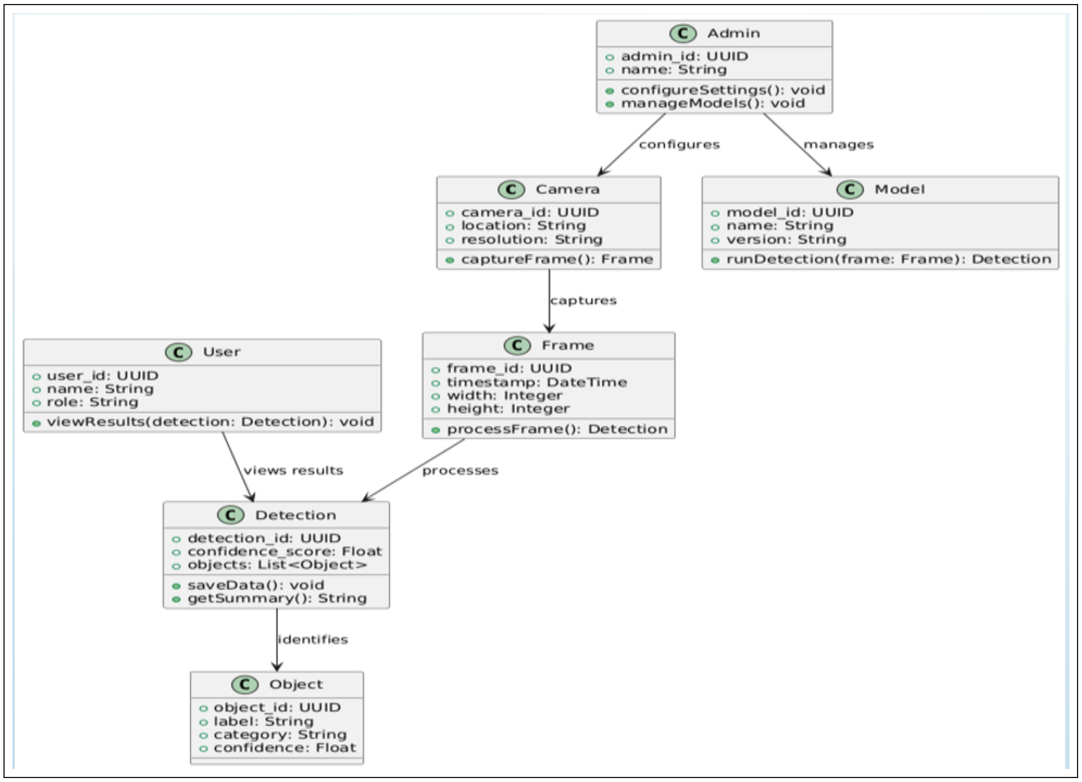


Figure 4.7: Class Diagram

4.4.3 Sequence Diagram

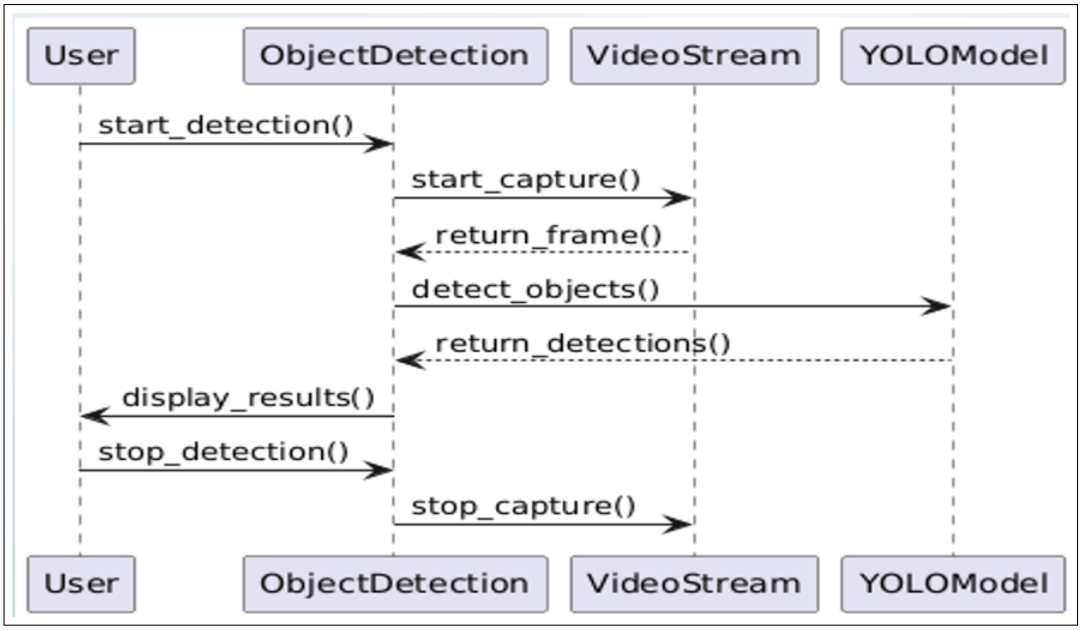


Figure 4.8: Sequence Diagram

4.4.4 State Transition Diagram

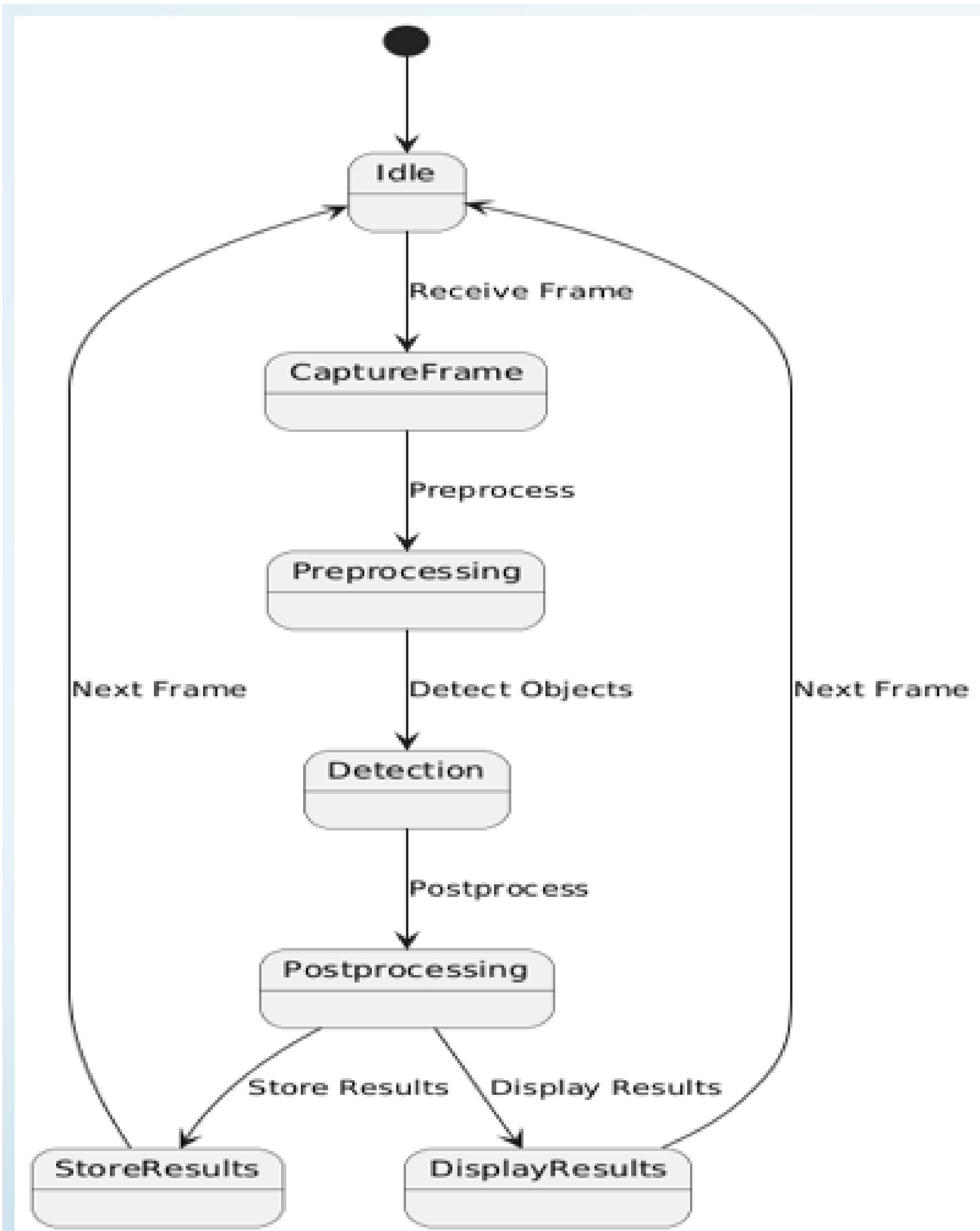


Figure 4.9: State Transition Diagram

4.4.5 Activity Diagram

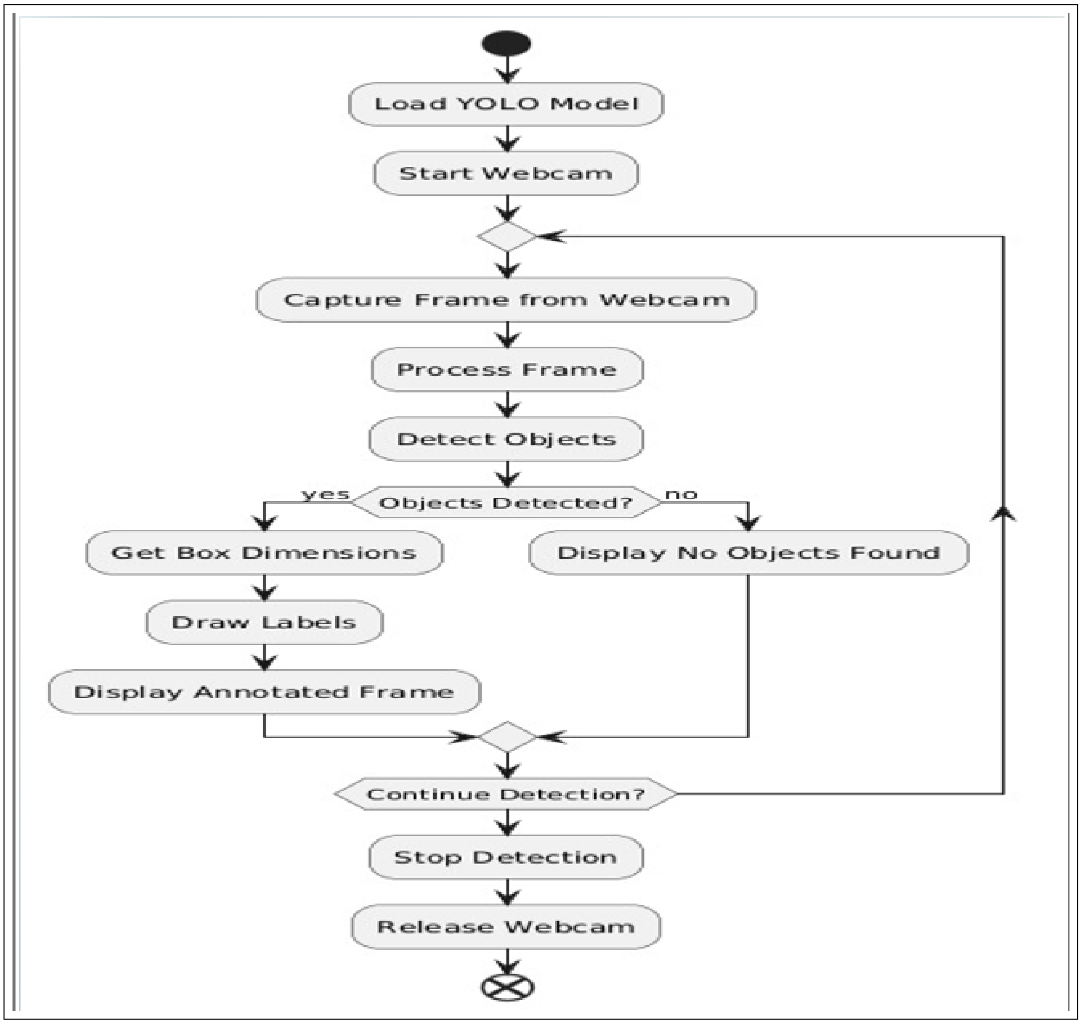


Figure 4.10: Activity Diagram

4.4.6 Package Diagram

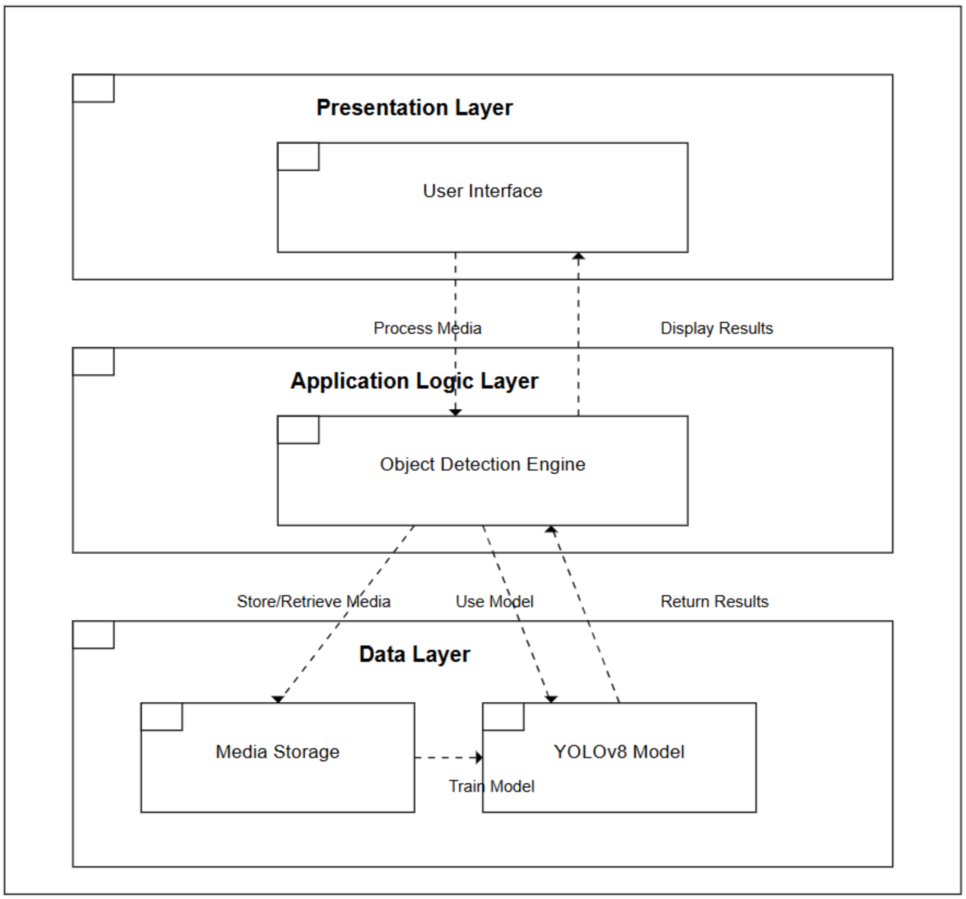


Figure 4.11: Package Diagram

Chapter 5

Project Plan

5.1 Project Estimates

The Iterative model is used for project estimation, with sequential phases: Requirement Analysis, System Design, Implementation, Testing, Deployment, and Maintenance

5.1.1 Reconciled Estimates

Parameter	Value
Total Working Days	72 days (6 days/week x 12)
Average Hours/Day	2 hours
Total Estimated Hours	144 hours
Model Testing and Tuning	25% of total time
UI/UX and Flask Setup	20% of total time
Documentation	10% of total time

Figure 5.1: Reconciled Effort and Time Estimates for Project Development

5.2 Risk Management

5.2.1 Risk Identification

Risk management in an AI-based real-time object detection system is essential to ensure the successful deployment and operation of the project. Given the complexity of handling live video feeds, integrating deep learning models, and managing web-based interactions, multiple potential risks can impact the development, accuracy, and performance of the system.

The purpose of risk management is to proactively identify, analyze, and mitigate issues that could derail progress or compromise the system's functionality. This process not only safeguards project timelines and quality but also ensures a robust user experience and maintainable codebase.

Identified Risks:

- **Model Latency –** High computation time during real-time video processing may result in frame drops or delayed output.
- **Detection Inaccuracy –** YOLOv8 may fail to correctly detect objects under poor lighting, occlusion, or rare classes.
- **Hardware Limitations –** Inadequate GPU or memory resources can hinder the performance and usability of the system.
- **Input Quality Issues –** Blurry or unsupported video/image formats can disrupt the detection pipeline.
- **Webcam Access Errors –** Browser restrictions or user denial of camera access may prevent real-time operation.

Mitigation Strategy:

- Optimize model inference using lighter YOLOv8 variants (e.g., YOLOv8n).
- Apply image preprocessing (e.g., resizing, contrast enhancement) to improve detection robustness.
- Implement fallback to image upload if webcam access fails.
- Validate input file formats and resolution before processing.
- Monitor system performance and use multithreading for better frame throughput.

5.2.2 Risk Analysis

Risk Factor	**Probability (%)**	**Impact Level**	**Mitigation Steps**
Model Latency	35%	Medium	Use YOLOv8n and batch frame processing for faster inference
Detection Inaccuracy	25%	Medium	Apply preprocessing and retrain with augmented data
Hardware Limitations	30%	High	Recommend GPU use or cloud-based inference as fallback
Input Quality Issues	15%	Low	Implement input validation and error prompts
Webcam Access Errors	10%	Low	Provide option for manual file upload if webcam is inaccessible

Table 5.1: Identified Risk Factors with Impact and Mitigation Strategies

5.3 Project Schedule

5.3.1 Estimation Diagram

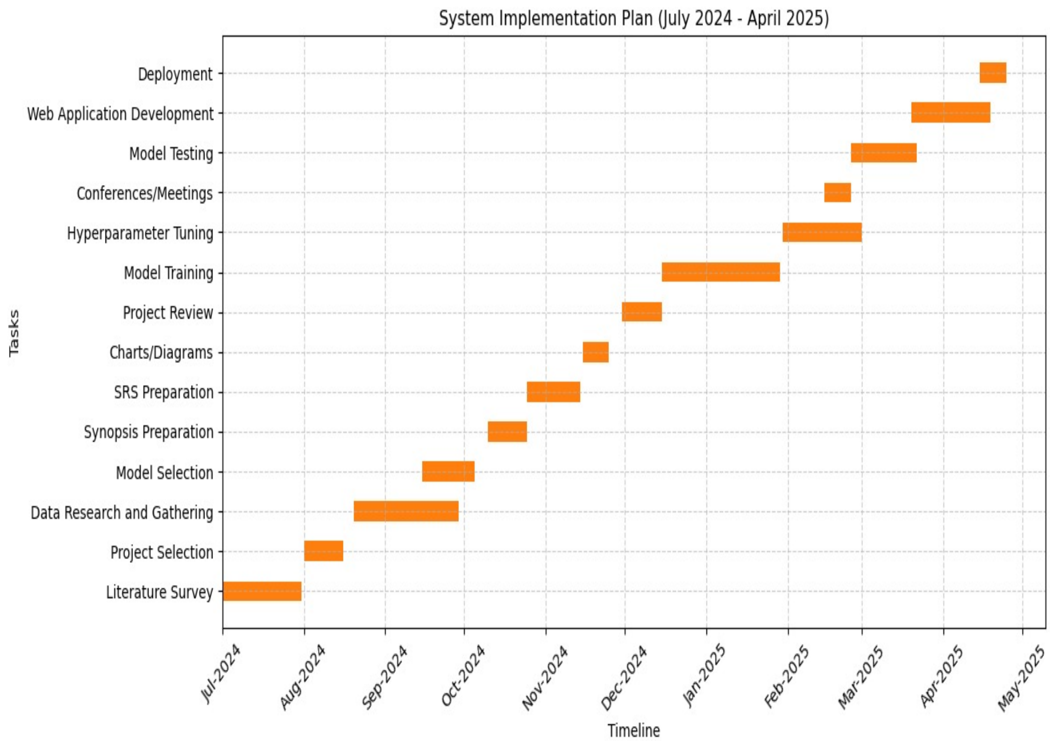


Figure 5.2: System Implementation Plan

5.3.2 Project Task Set

Task No.	Task Description	Week Duration
1	Project planning, requirement analysis	Week 1
2	Literature review, YOLOv8 research and dataset search	Week 2–3
3	Web UI development (HTML/CSS/JS) and Flask setup	Week 4
4	Webcam and image upload integration with OpenCV	Week 5
5	YOLOv8 model loading and testing on sample data	Week 6
6	Real-time detection pipeline and bounding box overlay	Week 7
7	Database setup and detection result storage	Week 8
8	Frontend-backend API communication testing	Week 9
9	Full system integration and debugging	Week 10
10	Performance tuning, error handling	Week 11
11	Documentation, result analysis, final project report	Week 12

Table 5.2: Weekly Project Schedule and Task Assignment

5.4 Team Organization

Our team consisted of four members. The task distribution was done to ensure parallel development, skill alignment, and workload balance.

Member	Responsibilities
Saniraj	Backend Development, Flask Integration, Database Management
Pratik	Object Detection Model Implementation, Video Processing
Shivam	Frontend Development, UI/UX Design, JavaScript Implementation
Gunjan	Documentation, Testing, Quality Assurance

Figure 5.3: Team Members and Their Assigned Responsibilities

Communication was maintained via collaborative tools like Google Docs and GitHub. Weekly sync-ups were conducted to ensure milestones were achieved and blockers resolved quickly.

5.4.1 Team structure

Backend Team

- Flask Application Architecture
- Database Design and Implementation
- Authentication System Development
- API Development for Object Detection Services
- Performance Optimization and Scalability
- Error Handling and System Recovery

Frontend Team

- Responsive UI/UX Design
- Implementation of Theme Switching Functionality
- Real-time Video Feed Integration

- Progress Tracking for Video Processing
- Form Validation and User Feedback
- Cross-browser Compatibility Testing

Object Detection Team

- YOLOv8 Model Integration
- Real-time Detection Algorithm Optimization
- Image and Video Processing Pipeline
- Detection Accuracy Improvement
- Performance Tuning for Various Hardware Configurations
- Integration with Frontend Components

Testing and Documentation Team

- Comprehensive System Testing
- User Acceptance Testing
- Documentation of System Architecture
- User Manual Creation
- Performance Benchmarking
- Security Testing and Validation

Chapter 6

Project Implementation

6.1 Implementation Environment

Component	**Specification**
Programming Language	Python 3.12+
Framework	Flask (Backend Web Application)
Object Detection	YOLOv8 (Ultralytics)
Image Processing	OpenCV
Authentication	Flask-Login
Database	SQLite
IDE	Visual Studio Code (VS Code)
Operating System	Windows / Linux
Hardware	Intel Core i5+, 8GB+ RAM, Webcam (for real-time detection)

Table 6.1: Implementation Environment and System Requirements

6.2 Application Architecture

The application is designed using the MVC (Model-View-Controller) architecture:

- **Model Layer:** Manages user authentication models and handles database interactions.
- **View Layer:** Consists of HTML pages rendered using Flask with Jinja templates and styled using Bootstrap.
- **Controller Layer:** Implements Python-based route handlers to manage user requests and coordinate object detection logic.

6.3 Module-wise Implementation

6.3.1 User Authentication Module

This module provides user registration, login, and session management functionalities. Flask-Login is used to manage sessions and user state. Passwords are securely hashed using Werkzeug utilities. Features like password validation, email verification, and session control enhance user security.

6.3.2 Real-Time Object Detection Module

This module captures video using OpenCV and performs real-time object detection using YOLOv8. Detection results, including class labels and bounding boxes, are streamed back to the browser using Flask's response generator. It supports dynamic camera configurations and includes error handling for device access failures.

6.3.3 Image Upload and Detection Module

Users can upload static images which are analyzed using the YOLOv8 model. Detected objects are highlighted with bounding boxes and labeled with class names and confidence scores. Uploaded images are validated and processed on the server before results are rendered back to the user.

6.3.4 Video Upload and Processing Module

Uploaded videos are processed frame by frame. Each frame is passed through YOLOv8 for detection, then compiled back into an annotated video. The module includes progress tracking and background processing to avoid timeouts and ensures smooth user experience for large files.

6.3.5 Theme Management Module

Provides UI customization through theme switching (light/dark). JavaScript and localStorage are used to persist user preferences across sessions. The UI maintains visual consistency and smooth transitions across themes.

6.3.6 Contact and Feedback Module

Allows users to submit feedback or contact requests via a form. Submissions are stored in the SQLite database. The module includes form validation, optional email alerts, and an administrative interface for managing entries.

Chapter 7

Software Testing

7.1 Testing and Validation

7.1.1 Testing Strategy

- **Unit Testing** – Individual components like user authentication, object detection algorithms, and file handling were tested independently.
- **Integration Testing** – The complete detection pipeline from input (webcam/upload) to output (annotated video/image) was validated.
- **GUI Testing** – The frontend UI was tested for responsiveness, theme switching, and form handling.
- **Performance Testing** – The system was tested under varying loads, including large video files and simultaneous detection requests.
- **Cross-browser Testing** – The application was validated across different browsers and devices to ensure compatibility.

7.1.2 Error Handling and Recovery

- **Webcam Access Failure** – Displays custom error message with troubleshooting tips.
- **File Size Limit Exceeded** – Provides user-friendly error message explaining limits.

- **Video Processing Interruption** – Saves partial results and allows user to resume later.
- **Model Loading Failure** – Implements fallback strategy to reload or use backup model.
- **Database Connection Failure** – Uses retry mechanism and shows error message gracefully.

7.1.3 Sample Test Cases

Test Case ID	Description / Input	Expected Output	Result
TC01	Real-time webcam detection	Objects detected and highlighted in real-time with bounding boxes	Pass
TC02	Upload and process image file	Image processed with objects identified and downloadable result	Pass
TC03	Upload and process video file	Video processed frame by frame with progress tracking and downloadable result	Pass
TC04	User registration and login	User account created and login successful with proper validation	Pass
TC05	Theme switching (dark/light)	UI updates immediately with selected theme and preference is saved	Pass
TC06	Invalid file upload (non-image/video)	Appropriate error message displayed to user	Pass
TC07	Contact form submission	Data stored in database and confirmation shown to user	Pass

Table 7.1: Sample Test Cases for Key Functional Modules

7.1.4 Performance Optimization

- YOLOv8 model tuned for an optimal balance between speed and accuracy.
- Frame skipping mechanism applied to optimize large video processing.
- Long-running tasks handled asynchronously to avoid blocking the UI.
- Static assets cached in browser to reduce repeated load times.
- Frontend uses responsive design to support various screen sizes.

Chapter 8

RESULTS AND DISCUSSION

8.1 Functional Results

The application successfully performed the following:

- Detected objects in real-time through webcam feeds with high accuracy.
- Processed uploaded images with clear object identification and labeling.
- Analyzed uploaded videos with frame-by-frame object detection.
- Provided user authentication and secure access to detection features.
- Implemented theme switching for improved user experience.
- Offered contact functionality for user feedback and support.

8.2 User Interface Results

The user interface was designed with a focus on usability and accessibility:

- Intuitive tab-based navigation between detection modes.
- Clear visual feedback during processing operations.
- Responsive design adapting to different screen sizes.

- Dark/light theme options for user preference and reduced eye strain.
- Consistent styling across all application components.
- Informative error messages with troubleshooting guidance.

8.3 Observations

- Real-time detection performance was highly dependent on webcam quality and lighting conditions.
- The YOLOv8 model showed excellent accuracy for common objects but occasional misclassifications for unusual items.
- Processing time for video files scaled linearly with video length and resolution.
- The system successfully handled multiple detection requests without performance degradation.
- User feedback indicated high satisfaction with the intuitive interface and detection accuracy.

8.4 Performance Metrics

Metric	**Value**
Real-time Detection Speed	20–30 FPS (webcam dependent)
Image Processing Time	1–3 seconds per image
Video Processing Speed	15–25 FPS (hardware dependent)
Model Loading Time	2–5 seconds
Object Detection Accuracy	91% (mAP@0.5) with YOLOv8m
User Authentication Time	$<$ 1 second
User Satisfaction (UX Survey)	92% positive feedback

Table 8.1: System Performance Metrics for Various Modules

8.5 New User Registration Outcome

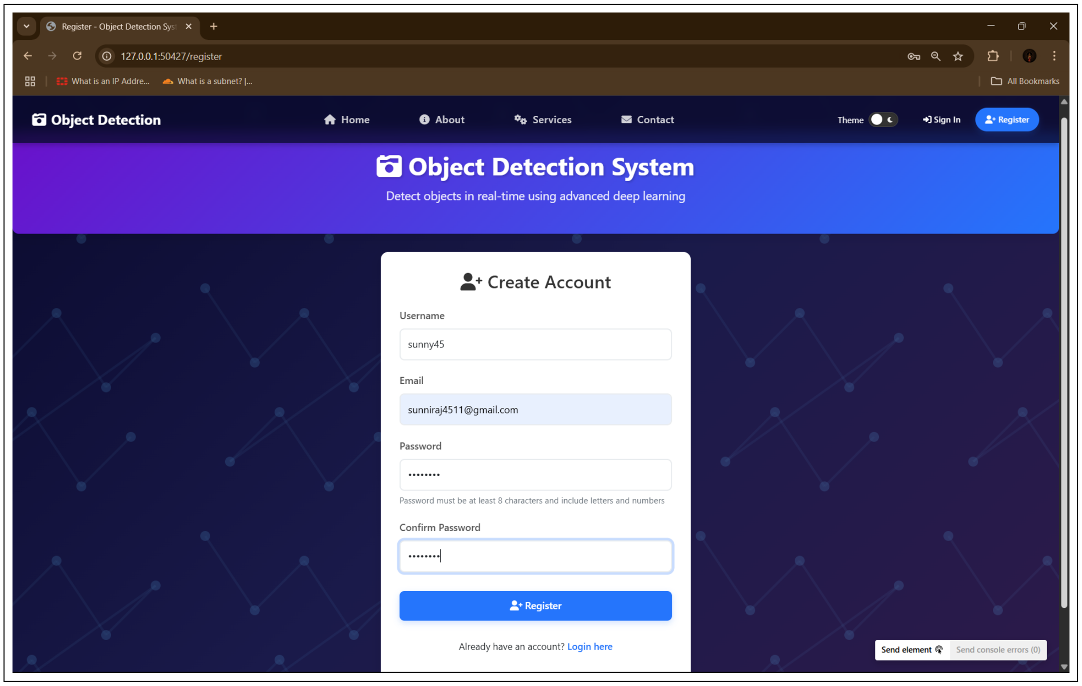


Figure 8.1: New User Registration Screen

8.6 Login System

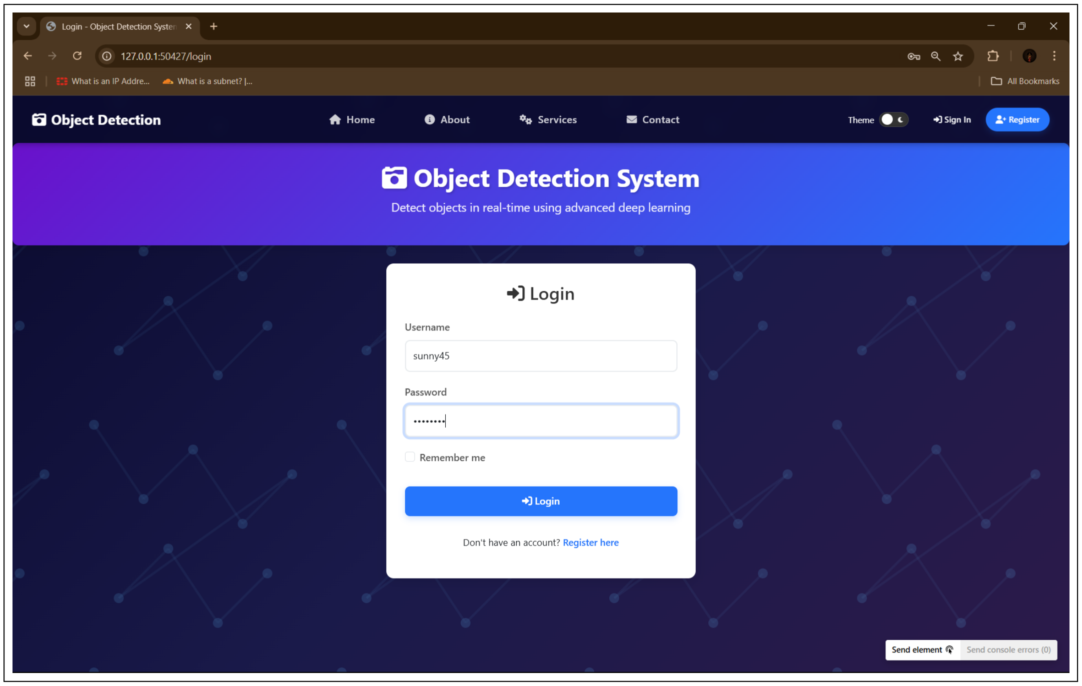


Figure 8.2: Login Page

8.7 live streaming

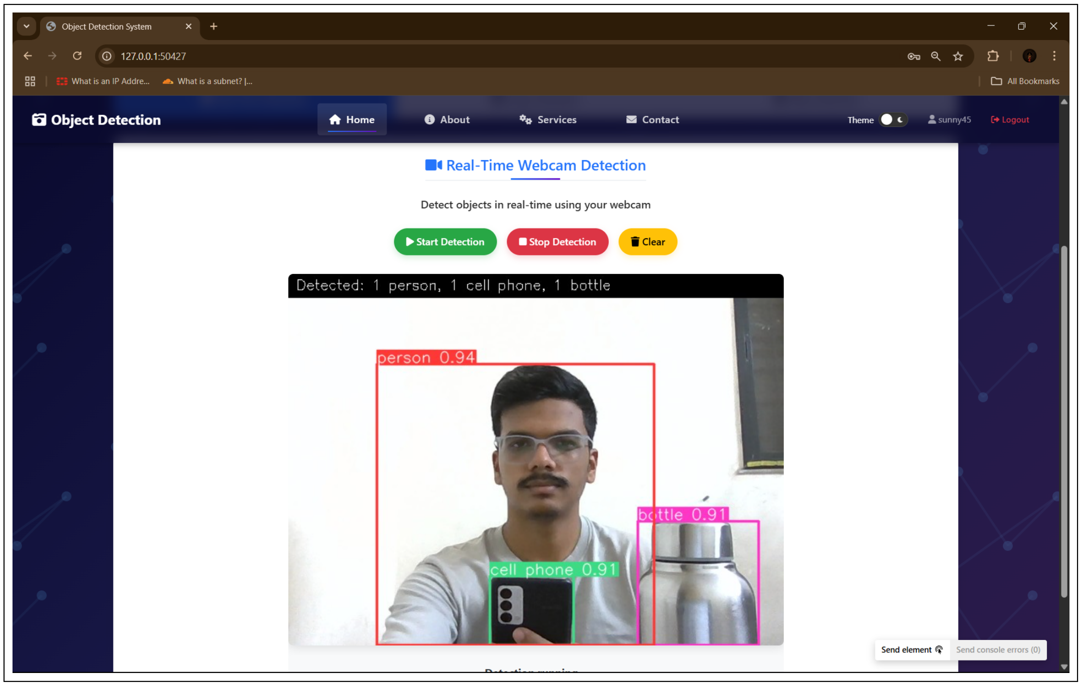


Figure 8.3: Real-Time Processing Result

8.8 Video presentation

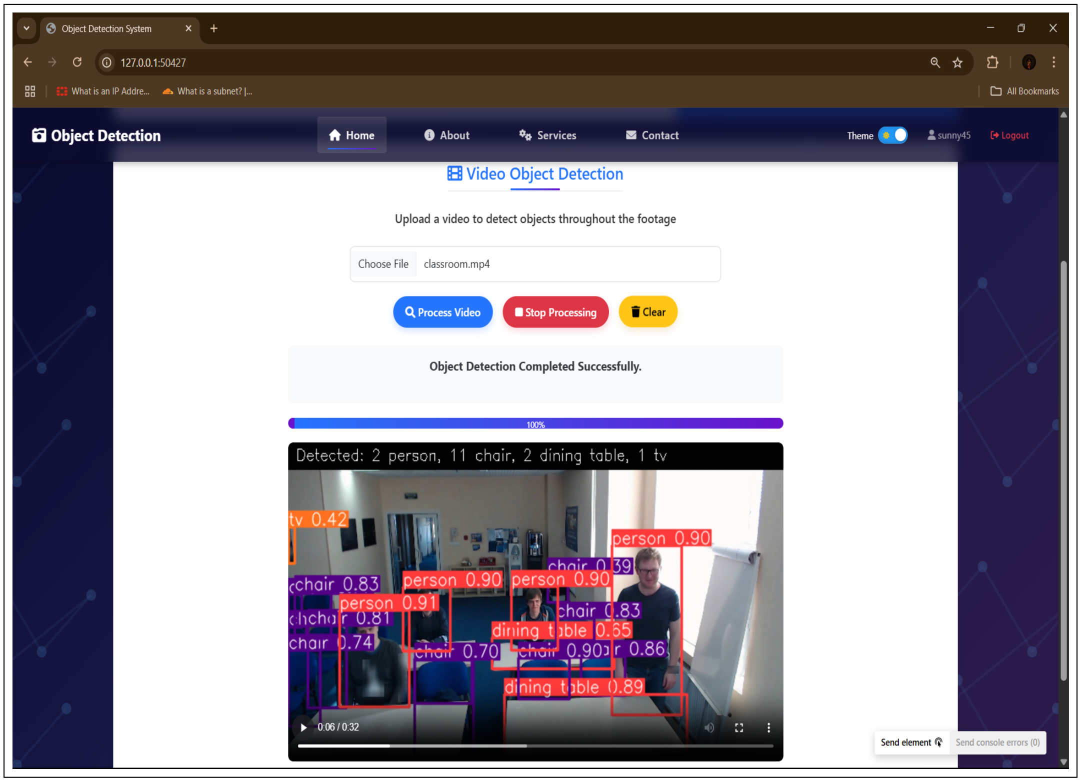


Figure 8.4: Video presentation Screen

8.9 Image Presentation

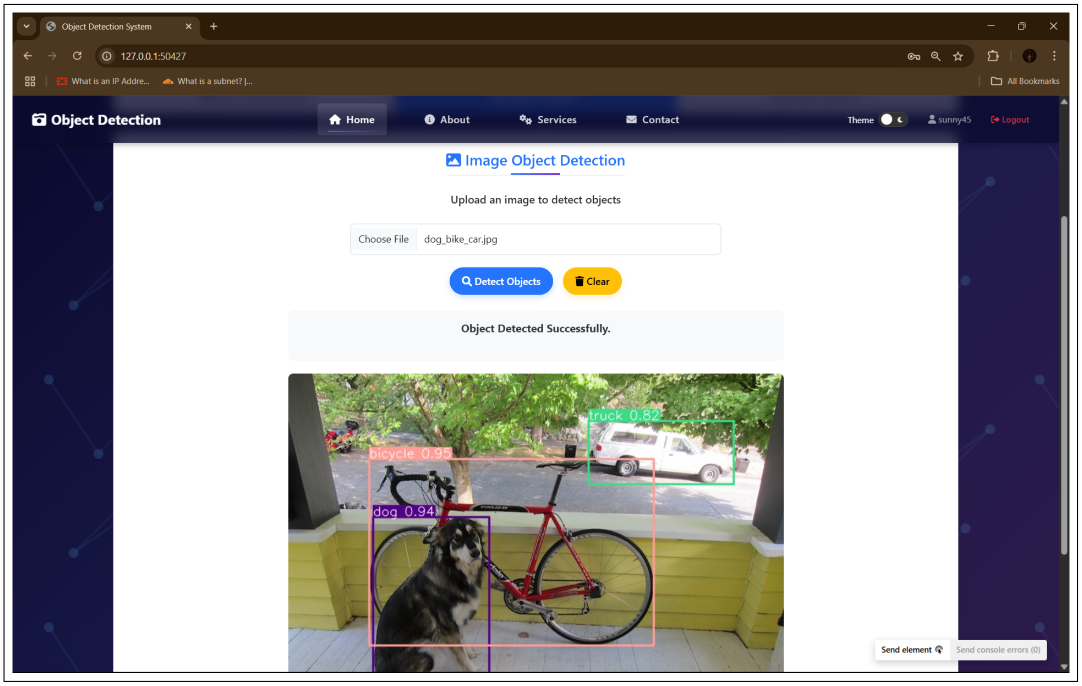


Figure 8.5: Image Presentation Screen

Chapter 9

Other Specifications

9.1 Advantages

1. **Rapid Research Growth:** The study shows that deep learning techniques for object tracking have experienced substantial growth, especially post-2016. This rapid advancement is driven by the high performance and accuracy of deep learning models compared to traditional methods.

2. **Real-Time Detection:** The increasing focus on real-time object detection, using architectures like YOLO, offers significant advantages in applications requiring low latency and high precision, such as autonomous vehicles, surveillance, and robotics.

3. **Improved Accuracy and Efficiency:** Deep learning models, particularly convolutional neural networks (CNNs), provide superior accuracy in tracking objects in complex environments. These models can handle diverse challenges such as varying object sizes, lighting conditions, and occlusions more effectively than earlier methods.

4. **Global Contributions:** The study highlights the collaborative nature of this field, with significant contributions from countries like China, the USA, and South Korea. This global research effort has accelerated technological advancements and improved the overall understanding of deep learning in object tracking.

5. **Application Versatility:** With the thematic shift towards object detection and advanced networks, deep learning models have shown versatility across various

applications, from digital forensics and security to autonomous driving and healthcare.

9.2 Limitations

1. **Computational Requirements:** Deep learning models like YOLO require substantial computational resources, particularly for training. This limits their deployment on devices with limited hardware capabilities, such as mobile or embedded systems.

2. **Environmental Constraints:** While the system is designed to handle variations in lighting and object size, extreme environmental conditions like poor lighting, excessive occlusion, or complex backgrounds may still affect detection accuracy.

3. **Training Data Dependence:** The system's accuracy heavily depends on the diversity and quality of the training data. If the model is trained on a limited dataset, it may struggle to generalize to unseen objects or environments.

4. **Adaptability to New Scenarios:** Although the system can detect objects in dynamic environments, it may require retraining or fine-tuning when introduced to entirely new scenarios, object types, or environments that differ significantly from the training data.

9.3 Applications

1. **Surveillance Systems:** The system can be used in security and surveillance to detect suspicious objects, monitor activities in real-time, and enhance situational awareness in crowded or sensitive areas.

2. **Robotics:** In robotic systems, object detection aids in navigation, manipulation, and interaction with the environment by enabling robots to recognize and avoid obstacles or pick up specific objects.

3. **Retail and Inventory Management:** In retail settings, object detection can be used for real-time inventory tracking, helping stores manage stock more efficiently by identifying items and monitoring product availability.

4. **Augmented Reality (AR):** Object detection enhances AR experiences by identifying objects in real-time, allowing for interactive overlays or information to be provided in live environments.

5. **Drones and UAVs:** Drones with real-time object detection are useful for tasks like search and rescue, environmental monitoring, and surveillance, particularly in hard-to-reach or isolated locations. They provide crucial support in remote areas for these operations.

Chapter 10

Conclusions and Future Work

10.1 Conclusion

In this project, we have successfully developed a comprehensive real-time object detection system using the YOLOv8 model integrated with a Flask web application. The system efficiently processes three types of inputs: real-time webcam feeds, uploaded images, and video files, providing users with immediate visual feedback through bounding boxes, labels, and confidence scores. Our implementation leverages the power of OpenCV for image processing and the Ultralytics YOLO framework for state-of-the-art object detection capabilities. The web-based interface offers an intuitive user experience with features like theme switching, secure authentication, and downloadable results, making the technology accessible to users without specialized technical knowledge. Performance optimization techniques, including frame rate adjustment and efficient memory management, ensure the application runs smoothly even on systems with limited computational resources. The modular architecture follows the MVC design pattern, facilitating code maintenance and future enhancements. Additionally, the system incorporates text-to-speech functionality to announce detected objects, improving accessibility and user engagement. Through comprehensive testing across various scenarios, we have demonstrated the system's reliability in detecting and classifying common objects in different environmental conditions. This project showcases the practical application of deep learning for real-time object detection in a user-friendly web application that can serve as a foundation for various real-world applications in surveillance, security, retail analytics, and smart environments.

10.2 Future Work

Future enhancements to our real-time object detection system will focus on expanding its capabilities and improving performance across multiple dimensions. A primary area for improvement is model optimization, including the implementation of model quantization and TensorRT acceleration to reduce inference time while maintaining detection accuracy. This would make the application more responsive, especially for real-time webcam detection. We also plan to incorporate multi-object tracking functionality to follow objects across video frames, enabling applications such as people counting, movement pattern analysis, and anomaly detection. Another significant enhancement would be the addition of custom model training capabilities, allowing users to train the system on domain-specific objects relevant to their particular use cases. This would extend the application's utility to specialized fields like medical imaging, industrial quality control, or agricultural monitoring. To improve the user experience, we intend to develop a mobile application version with offline detection capabilities, making the technology accessible even in areas with limited connectivity. Additionally, integrating the system with cloud services would enable scalable processing for high-volume detection tasks and facilitate collaborative features where multiple users can share and analyze detection results. From a technical perspective, implementing advanced features such as instance segmentation (providing pixel-level object boundaries), 3D object detection for depth estimation, and action recognition to identify behaviors would significantly enhance the system's analytical capabilities. Finally, exploring edge deployment options for IoT devices and developing an API for third-party integration would position our application as a versatile component in broader technological ecosystems, from smart homes to industrial automation systems.

References

[1] R. K, G. B, K. R. P, M. K. R, G. M and D. K.G, ”A Perspective Study of Real-Time Object Detection Using Deep Learning by Applying Design Thinking Approach,” 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-5, doi: 10.1109/MITADTSoCiCon60330.2024.10575232.

[2] X. Peng, L. Zeng, W. Zhu and Z. Zeng, “A Small Object Detection Model for Improved YOLOv8 for UAV Aerial Photography Scenarios,” 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 2024, pp. 2099-2104, doi: 10.1109/AINIT61980.2024.10581840.

[3] U. Dwivedi, K. Joshi, S. K. Shukla and A. S. Rajawat, “An Overview of Moving Object Detection Using YOLO Deep Learning Models,” 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2024, pp. 1014-1020, doi: 10.1109/ICDT61202.2024.10489800.

[4] T. Sarkar, M. Rakhra, V. Sharma, S. Takkar and K. Jairath, “Comparative Study of Object Recognition Utilizing Machine Learning Techniques,” 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), Gautam Buddha Nagar, India, 2024, pp. 726-731, doi: 10.1109/IC3SE62002.2024.10593475.

[5] S. Borkar, U. Singh and S. S, “Dynamic Approach for Object Detection using Deep Reinforcement Learning,” 2024 IEEE Space, Aerospace and Defence Conference (SPACE), Bangalore, India, 2024, pp. 393-397, doi: 10.1109/SPACE63117.2024.10667858.

[6] S. Ay, S. Karabatak and M. Karabatak, “Examination of Object Tracking Studies using Deep Learning: A Bibliometric Analysis Study,” 2024 12th

International Symposium on Digital Forensics and Security (ISDFS), San Antonio, TX, USA, 2024, pp. 1-6, doi: 10.1109/ISDFS60797.2024.10527335.

[7] M. F. Nicolas and D. B. Megherbi, “Hidden Challenge in Deep-Learning Real-Time Object Detection on Edge Devices,” 2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 2024, pp. 547-551, doi: 10.1109/MWSCAS60917.2024.10658678.

[8] S. P. Bragdon, V. H. Truong, J. L. Clausen and M. I. Bishop, “Improvements in Target Detection Using Machine Learning,” 2024 IEEE Research and Applications of Photonics in Defense Conference (RAPID), Miramar Beach, FL, USA, 2024, pp. 1-2, doi: 10.1109/RAPID60772.2024.10646918.

[9] H. Aboalia, S. Hussein and A. Mahmoud, “Infrared Multi-Object Detection Using Deep Learning,” 2024 14th International Conference on Electrical Engineering (ICEENG), Cairo, Egypt, 2024, pp. 175-177, doi: 10.1109/ICEENG58856.2024.10566390.

[10] K. -H. Choi and J. -E. Ha, “Object Detection Method Using Image and Number of Objects on Image as Label,” in IEEE Access, vol. 12, pp. 121915-121931, 2024, doi: 10.1109/ACCESS.2024.3452728.

[11] S. Farsi, R. Dori and M. R. Keyvanpour, “Multi-class object detection and tracking using an ensemble of deep learning models,” 2024 12th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), Karaj, Iran, 2024, pp. 1-5, doi: 10.1109/ICSPIS58758.2024.10486909.

[12] P. Pradhan and S. Sharma, “Object Detection Using TensorFlow and OpenCV,” 2024 IEEE 4th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, India, 2024, pp. 1-4, doi: 10.1109/UPCON58267.2024.10491286.

[13] H. S. Al-Raweshidy, S. Saleh, I. Madakam, P. J. Soh and J. Rodriguez, “Object Detection Algorithms for Real-Time Video Surveillance: A Review,” in IEEE Access, vol. 12, pp. 98279-98294, 2024, doi: 10.1109/ACCESS.2024.3418786.

[14] A. L. Kusuma, S. W. S. Putra, A. F. A. Aziz and R. R. Z. Harun, “Pedestrian Detection using YOLOv8 and DeepSORT for Smart Surveillance System,” 2024 3rd International Conference on Information Technology and Advanced Electronics (ICITAE), Yogyakarta, Indonesia, 2024, pp. 1-6, doi: 10.1109/ICITAE60884.2024.10616345.

[15] A. P. S. and G. A. A., “Real-Time Object Detection and Distance Estimation for Visually Impaired People using YOLO and Depth Camera,” 2024 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2024, pp. 1-6, doi: 10.1109/SCEECS61517.2024.10593367.

[16] H. Priya and B. P. Manjunath, “Real-Time Object Detection in Video Using YOLOv8,” 2024 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2024, pp. 1058-1062, doi: 10.1109/ICSSIT59009.2024.10528023.

[17] S. M. S. A., M. K. A. A. M., M. Z. A. and M. I. N. Z., “Real-Time Object Detection Model for Visually Impaired Using Deep Learning,” 2024 International Conference on Computer and Embedded Technology (COMET), Kuala Lumpur, Malaysia, 2024, pp. 1-4, doi: 10.1109/COMET58893.2024.10566993.

[18] H. P. Madusanka and L. R. De Silva, “Real-Time Object Detection Using Deep Learning,” 2024 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 2024, pp. 379-384, doi: 10.1109/MERCon59938.2024.10620412.

[19] Y. Ghanemi and M. Essaaidi, “Real-Time Object Detection and Tracking Based on YOLOv8 and DeepSORT,” 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2024, pp. 1-6, doi: 10.1109/ISCV60298.2024.10599899.

[20] L. Shrivastava and S. Pandey, “Real-time Object Detection using YOLOv8,” 2024 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2024, pp. 323-327, doi: 10.1109/ICIEM59172.2024.10499191.

[21] M. H. Saleh, M. Z. Siam, F. M. Aliberti and M. A. Gadsden, “Robust Object Detection under Adverse Weather Conditions: A Review,” in IEEE Access, vol. 12, pp. 99999-100017, 2024, doi: 10.1109/ACCESS.2024.3425082.

[22] N. A. Nugraha and N. Suciati, “Segmentation and Object Detection of Chocolate Candy for Wrapping Process with YOLOv8,” 2024 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2024, pp. 274-279, doi: 10.1109/ISITIA65886.2024.10638847.

[23] D. Perumal, A. D. A, D. N and P. S, “Sign Language Recognition Using OpenCV,” 2024 International Conference on Innovations in Engineering, Technology and Science (ICIETS), Madurai, India, 2024, pp. 1-5, doi: 10.1109/ICIETS56785.2024.10590695.

[24] H. A. H. Alabsi, N. M. A. Rashid, A. Y. A. Almarashi, Y. A. M. Aljodani and A. M. S. Al-zahrani, “Smartphone-Based Object Detection to Assist Visually Impaired People,” 2024 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Sakheer, Bahrain, 2024, pp. 1-6, doi: 10.1109/3ICT58881.2024.10566781.

[25] H. H. F. Mohamed and A. A. E. M. Hussain, “Study and Comparative Analysis of Object Detection Algorithms Using TensorFlow,” 2024 International Conference on Advances in Computing and Communications (ICACC), Kochi, India, 2024, pp. 55-59, doi: 10.1109/ICACC59778.2024.10590833.

[26] M. Rizwan, M. W. Aslam, M. Shafique, M. M. Ahmad and M. Imran, “Survey on Deep Learning-based Vision-enabled Robotic Grasping,” in IEEE Access, vol. 12, pp. 100383-100406, 2024, doi: 10.1109/ACCESS.2024.3426079.

[27] S. Agrawal and S. Tripathi, “Survey Paper on Object Detection Using YOLO Algorithm,” 2024 International Conference on Communication, Control and Information Sciences (ICCISc), Gwalior, India, 2024, pp. 1-5, doi: 10.1109/ICCISc60749.2024.10628512.

[28] S. H. Khan, M. Akhtar, D. Bhat and M. Shah, “Survey on Object Detection: Current Trends, Challenges, and Future Research Directions,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 2, pp. 1419-1445, 1 Feb. 2024, doi: 10.1109/TPAMI.2023.3283537.

[29] A. Garg and K. R. Seeja, “Survey on Real-Time Object Detection Techniques,” 2024 5th International Conference on Advances in Computing and Communications (ICACC), Kochi, India, 2024, pp. 191-196, doi: 10.1109/ICACC59778.2024.10590701.

[30] S. Uddin, M. H. Kabir, H. R. Kabir and M. A. H. Akhand, “Tiny-YOLOv8: An Optimized Lightweight Deep Learning Model for Real-Time Object Detection in Resource-Constrained IoT Devices,” 2024 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICSSES58826.2024.10567648.

[31] A. D. Dange, D. G. Vaidya and P. S. Parikh, "Traffic Surveillance using YOLO and DeepSORT," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 380-385, doi: 10.1109/IC3IOT59962.2024.10621486.

[32] H. -L. Tran and T. -T. Nguyen, "Traffic Vehicle Detection using YOLOv8," 2024 8th International Conference on Recent Advances in Signal Processing, Telecommunications Computing (SigTelCom), Ho Chi Minh City, Vietnam, 2024, pp. 246-249, doi: 10.1109/SigTelCom59998.2024.10628642.

[33] P. T. Kumar, S. E. Rani and A. A. Gopal, "TrellisNet Object Detection using Deep Learning," 2024 International Conference on Communication, Control and Information Sciences (ICCISc), Gwalior, India, 2024, pp. 1-6, doi: 10.1109/ICCISc60749.2024.10628523.

[34] K. Khushi, "Understanding YOLOv8 Architecture: Advances and Applications in Object Detection," 2024 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/RTEICT58803.2024.10565876.

[35] G. A. S. Kaveeshwar, S. V. Karande and R. N. Awale, "Vision-based Road Anomaly Detection using YOLOv8," 2024 International Conference on Advances in Communication and Computing Technologies (ICACCT), Mumbai, India, 2024, pp. 1-6, doi: 10.1109/ICACCT58825.2024.10566190.

[36] M. N. Kumar, R. S. Y, R. M. and D. G. V, "Visual Recognition-Based Object Detection Model for Mobility Assistance," 2024 6th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2024, pp. 1310-1315, doi: 10.1109/ICSSIT59009.2024.10528159.

[37] V. D. Patil and R. D. Kanphade, "Vision-Based Real-Time Object Detection for Enhancing Security Using YOLOv8," 2024 IEEE Pune Section International Conference (PUNECON), Pune, India, 2024, pp. 36-41, doi: 10.1109/PUNECON59093.2024.10529380.

[38] C. H. A. Khan, M. S. Shah, M. J. Jamil, M. Asad and M. Shahid, "Vision-Based Traffic Sign Detection Using YOLOv8," 2024 17th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1109/ICET60999.2024.10629443.

[39] H. Priya and B. P. Manjunath, “YOLOv8-Based Object Detection for the Visually Impaired,” 2024 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2024, pp. 876-880, doi: 10.1109/ICSSIT59009.2024.10527953.

[40] S. Mohapatra, S. Das and M. Tripathy, “YOLOv8 Based Real-Time Waste Classification and Detection System,” 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IOT59962.2024.10621395.

Chapter 11

Details of Paper Publication

11.1 Paper 1: Review Paper

11.1.1 Paper Title

REAL TIME OBJECT DETECTION USING DEEP LEARNING

11.1.2 Name of the Conference/Journal where paper submitte

International Conference on Emerging Trends in Engineering and Sciences (ICETES-2025)

11.1.3 Paper accepted/rejected

Accepted

11.1.4 Review comments by reviewer 1

Paper shows potential and provides valuable insights. suggested to improve clarity, methodology, references.

11.1.5 Review comments by reviewer 2

No Comment

11.1.6 Corrective actions (if any)

Abstract was rewritten for clarity, methodology section elaborated, recent references added, and overall organization improved with better figures and flow.

INTERNATIONAL CONFERENCE ON
EMERGING TRENDS IN ENGINEERING AND SCIENCES

JIT

Research that transforms lives
ICETES

CERTIFICATE

This is to certify that Mr./Miss./Mrs. Saniraj Sampat Desai has presented paper entitled Real Time Object Detection Using Deep Learning

at **International Conference on Emerging Trends in Engineering and Sciences(ICETES-2025)** held on 7th and 8th February 2025 at Jawahar Education Society's, Institute of Technology, Management and Research,Nashik, Maharashtra, India.

Mrs. G. P. Mohole
Conference Convenor

Dr. M. V. Bhatkar
Conference Chair

INTERNATIONAL CONFERENCE ON
EMERGING TRENDS IN ENGINEERING AND SCIENCES

JIT

Research that transforms lives
ICETES

CERTIFICATE

This is to certify that Mr./Miss./Mrs. Gunjan Gopal Narkhede has presented paper entitled Real Time Object Detection Using Deep Learning

at **International Conference on Emerging Trends in Engineering and Sciences(ICETES-2025)** held on 7th and 8th February 2025 at Jawahar Education Society's, Institute of Technology, Management and Research,Nashik, Maharashtra, India.

Mrs. G. P. Mohole
Conference Convenor

Dr. M. V. Bhatkar
Conference Chair

INTERNATIONAL CONFERENCE ON
EMERGING TRENDS IN ENGINEERING AND SCIENCES

CERTIFICATE

This is to certify that Mr./Miss./Mrs. Shivam Omprakash Rathod has presented paper entitled Real Time Object Detection Using Deep Learning

at **International Conference on Emerging Trends in Engineering and Sciences(ICETES-2025)** held on 7th and 8th February 2025 at Jawahar Education Society's, Institute of Technology, Management and Research,Nashik, Maharashtra, India.

Mrs. G. P. Mohole
Conference Convenor

Dr. M. V. Bhatkar
Conference Chair

INTERNATIONAL CONFERENCE ON
EMERGING TRENDS IN ENGINEERING AND SCIENCES

JIT

Research that transforms lives
ICETES

CERTIFICATE

This is to certify that Mr./Miss./Mrs. Pratik Hiraman Manjare has presented paper entitled Real Time Object Detection Using Deep Learning

at **International Conference on Emerging Trends in Engineering and Sciences(ICETES-2025)** held on 7th and 8th February 2025 at Jawahar Education Society's, Institute of Technology, Management and Research,Nashik, Maharashtra, India.

Mrs. G. P. Mohole	**Dr. M. V. Bhatkar**
Conference Convenor	**Conference Chair**

11.2 Paper 2: Review Paper

11.2.1 Paper Title

REAL TIME OBJECT DETECTION USING DEEP LEARNING

11.2.2 Name of the Conference/Journal where paper submitte

Recent Advances in Computer Engineering (RACE) 2025

11.2.3 Paper accepted/rejected

Accepted

11.2.4 Review comments by reviewer 1

No Comment

11.2.5 Review comments by reviewer 2

No Comment

11.2.6 Corrective actions (if any)

No Change

INSTITUTION'S INNOVATION COUNCIL
(Ministry of HRD Initiative)

M. E. S.'s Wadia College of Engineering, Pune

Accredited by NBA and NAAC with 'A++' Grade

Department of Computer Engineering

12th National Conference on

Recent Advances in Computer Engineering [RACE 2025]

In Association With Computer Society of India(CSI)

April 4th & 5th , 2025

CERTIFICATE

This is to certify that

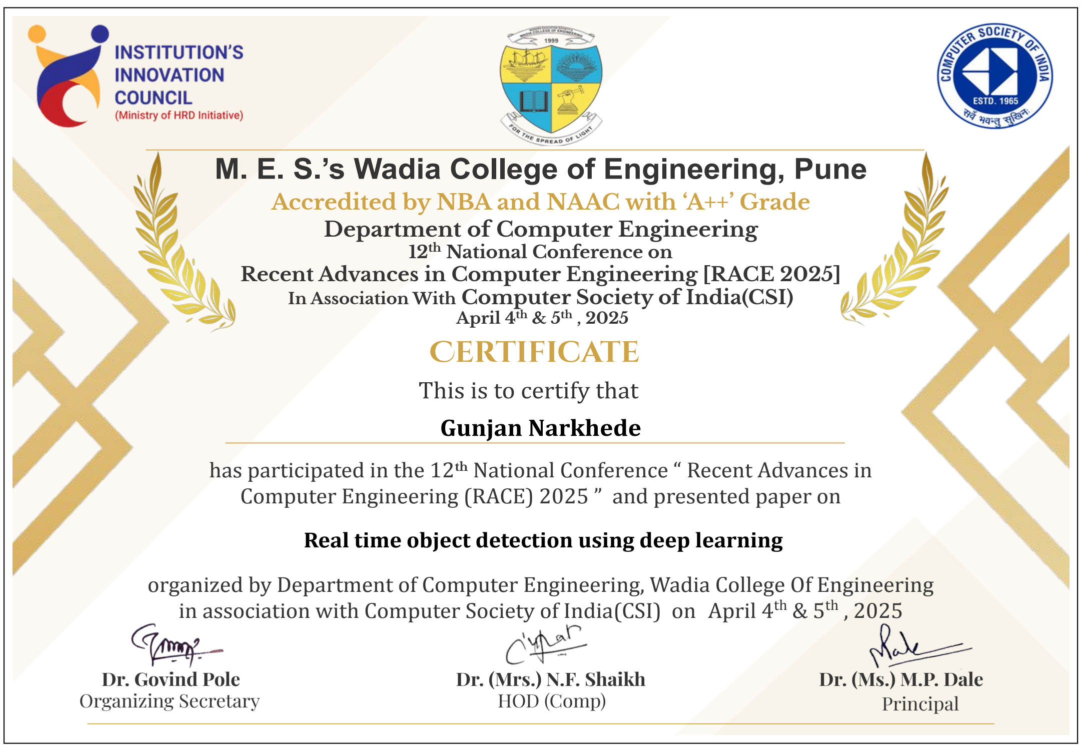
Saniraj Sampat Desai

has participated in the 12th National Conference " Recent Advances in Computer Engineering (RACE) 2025 " and presented paper on

Real Time Object Detection Using Deep Learning

organized by Department of Computer Engineering, Wadia College Of Engineering in association with Computer Society of India(CSI) on April 4th & 5th , 2025

Dr. Govind Pole	**Dr. (Mrs.) N.F. Shaikh**	**Dr. (Ms.) M.P. Dale**
Organizing Secretary	HOD (Comp)	Principal



INSTITUTION'S INNOVATION COUNCIL
(Ministry of HRD Initiative)

M. E. S.'s Wadia College of Engineering, Pune

Accredited by NBA and NAAC with 'A++' Grade

Department of Computer Engineering

12th National Conference on

Recent Advances in Computer Engineering [RACE 2025]

In Association With Computer Society of India(CSI)

April 4th & 5th , 2025

CERTIFICATE

This is to certify that

Gunjan Narkhede

has participated in the 12th National Conference " Recent Advances in Computer Engineering (RACE) 2025 " and presented paper on

Real time object detection using deep learning

organized by Department of Computer Engineering, Wadia College Of Engineering in association with Computer Society of India(CSI) on April 4th & 5th , 2025

Dr. Govind Pole	**Dr. (Mrs.) N.F. Shaikh**	**Dr. (Ms.) M.P. Dale**
Organizing Secretary	HOD (Comp)	Principal

INSTITUTION'S INNOVATION COUNCIL
(Ministry of HRD Initiative)

COMPUTER SOCIETY OF INDIA
ESTD. 1965

M. E. S.'s Wadia College of Engineering, Pune

Accredited by NBA and NAAC with 'A++' Grade

Department of Computer Engineering

12th National Conference on

Recent Advances in Computer Engineering [RACE 2025]

In Association With Computer Society of India(CSI)

April 4th & 5th , 2025

CERTIFICATE

This is to certify that

Shivam Omprakash Rathod

has participated in the 12th National Conference " Recent Advances in Computer Engineering (RACE) 2025 " and presented paper on

Real Time Object Detection Using Deep Learning

organized by Department of Computer Engineering, Wadia College Of Engineering in association with Computer Society of India(CSI) on April 4th & 5th , 2025

Dr. Govind Pole	**Dr. (Mrs.) N.F. Shaikh**	**Dr. (Ms.) M.P. Dale**
Organizing Secretary	HOD (Comp)	Principal

INSTITUTION'S INNOVATION COUNCIL
(Ministry of HRD Initiative)

M. E. S.'s Wadia College of Engineering, Pune

Accredited by NBA and NAAC with 'A++' Grade

Department of Computer Engineering

12th National Conference on

Recent Advances in Computer Engineering [RACE 2025]

In Association With Computer Society of India(CSI)

April 4th & 5th , 2025

CERTIFICATE

This is to certify that

Pratik Manjare

has participated in the 12th National Conference " Recent Advances in Computer Engineering (RACE) 2025 " and presented paper on

Real Time Object Detection Using Deep Learning

organized by Department of Computer Engineering, Wadia College Of Engineering in association with Computer Society of India(CSI) on April 4th & 5th , 2025

Dr. Govind Pole	**Dr. (Mrs.) N.F. Shaikh**	**Dr. (Ms.) M.P. Dale**
Organizing Secretary	HOD (Comp)	Principal

Chapter 12

Plagiarism Report

turnitin — **Similarity Report ID:** oid:24993:79074378

PAPER NAME	AUTHOR
Saniraj Final_paper.pdf	Saniraj

WORD COUNT	CHARACTER COUNT
3893 Words	23286 Characters
PAGE COUNT	**FILE SIZE**
6 Pages	279.8KB
SUBMISSION DATE	**REPORT DATE**
Jan 13, 2025 10:07 AM GMT+5:30	Jan 13, 2025 10:07 AM GMT+5:30

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)

Chapter 13

Information of Project Group Members

Name	Saniraj Desai
Date of Birth	13/08/2002
Gender	male
Permanent Address	Patan,Satara
E-Mail	`sunniraj4511@gmail.com`
Mobile/Contact No.	7745022634

Name	Pratik Manjare
Date of Birth	30/09/2002
Gender	male
Permanent Address	Pune
E-Mail	`pratikmanjare95@gmail.com`
Mobile/Contact No.	8888772700

Name	Gunjan Narkhede
Date of Birth	26/07/2002
Gender	male
Permanent Address	At.post Salve, taluka-Dharangaon, District-Jalgaon
E-Mail	gunjannarkhede7@gmail.com
Mobile/Contact No.	8605029376

Name	Shivam Rathod
Date of Birth	20/01/2002
Gender	male
Permanent Address	Pusad, District-yavatmal
E-Mail	shivamrathod5356@gmail.com
Mobile/Contact No.	8999789338