

# Real-Time Video Surveillance Using ESP-32 CAM for IoT Applications

Monika Kapoor<sup>1</sup>, Aashutosh Sharma<sup>2</sup>, Jasleen Kaur Minhas<sup>3,\*</sup>,  
Kartik Bakshi<sup>4</sup>, Khushi Kumari<sup>5</sup>

## Abstract

*The rapid advancements in technology and the growing concerns for security have driven the need for innovative, cost-effective, and scalable surveillance systems. This project presents the development of a smart surveillance system using the ESP-32 CAM module, a compact and low-cost microcontroller integrated with Wi-Fi, Bluetooth, and a high-resolution OV2640 camera. The system is designed to provide real-time monitoring, motion detection, and automated alert features, catering to the security needs of residential, commercial, and industrial environments. The surveillance system harnesses the capabilities of the ESP-32 CAM to enable live video streaming and image capture, which can be accessed remotely via web browsers or mobile applications. The system is equipped with intelligent motion detection algorithms, ensuring efficient monitoring by triggering event-based video recordings or snapshots. This approach optimizes storage and reduces the consumption of bandwidth, making the system resource-efficient. Data captured by the system can be stored locally on a microSD card or uploaded to cloud platforms such as Firebase, AWS (Amazon Web Services), or Google Drive for remote access and backup. Cloud integration enhances flexibility and allows users to review footage or monitor live feeds from anywhere, ensuring a comprehensive security solution. The system also sends instant notifications via email or push notifications upon detecting suspicious activities, providing users with real-time alerts for prompt action. In addition to its advanced functionalities, the ESP-32 CAM-based surveillance system is designed with energy efficiency and scalability in mind. Its low power consumption makes it ideal for battery-operated setups, while its modular architecture allows for easy expansion, enabling multiple units to work together seamlessly for larger monitoring networks. The use of open-source software frameworks further enhances the system's affordability and customization potential. This project underscores the potential of leveraging low-cost yet powerful microcontrollers like the ESP-32 CAM to develop modern surveillance solutions that meet the demands of an increasingly connected world. By integrating real-time monitoring, intelligent alerts, and remote access, the proposed system offers a practical, reliable, and innovative approach to enhancing security across various domains.*

### \*Author for Correspondence

Jasleen Kaur Minhas  
E-mail: minhas.jasleen3@gmail.com

<sup>1</sup>Professor, Department of Electronics and Communication Engineering, Lakshmi Narain College of Technology, Kalchuri Nagar, Bhopal, Madhya Pradesh, India

<sup>2-3</sup>Student, Department of Electronics and Communication Engineering, Lakshmi Narain College of Technology, Kalchuri Nagar, Bhopal, Madhya Pradesh, India

Received Date: March 29, 2025  
Accepted Date: April 14, 2025  
Published Date: April 22, 2025

**Citation:** Monika Kapoor, Aashutosh Sharma, Jasleen Kaur Minhas, Kartik Bakshi, Khushi Kumari. Real-Time Video Surveillance Using ESP-32 CAM for IoT Applications. Journal of Embedded System & Applications. 2025; 13(2): 26–35p.

**Keywords:** ESP-32 CAM, wireless surveillance, embedded systems, internet of things (IoT) security

## INTRODUCTION

Surveillance systems have long been integral to maintaining security and safety in various settings, including residential, commercial, and public spaces. Historically, these systems relied on traditional methods, such as manual monitoring through CCTV cameras, which, although still widely used, present notable shortcomings. They depend heavily on constant human observation, making them prone to inefficiencies and lapses in vigilance due to fatigue or oversight. Additionally,

---

these systems are limited in their capabilities, as they lack advanced analytical tools that can proactively detect or respond to emerging situations in dynamic environments. As security threats grow more complex and diverse, the limitations of traditional systems have become increasingly apparent, underscoring the urgent need for smarter surveillance solutions. Modern advancements demand systems that are not only automated but also capable of intelligent decision-making, integrating technologies like motion detection, real-time data processing, and artificial intelligence (AI)-driven analysis to enhance efficiency, reliability, and adaptability.

The advent of new technologies, particularly the internet of things (IoT) and AI, has revolutionized how surveillance systems operate. IoT has enabled the interconnection of devices like cameras, sensors, and cloud servers, creating a robust network for real-time data exchange. This connectivity ensures seamless communication between various components, allowing for better data collection, analysis, and storage. On the other hand, AI has added intelligence to surveillance systems by introducing features like motion detection, object recognition, and behavior analysis. Together, these technologies have paved the way for systems that are more efficient, responsive, and capable of automating tasks traditionally reliant on human intervention.

Motion detection is a fundamental feature of advanced surveillance systems. It allows the system to remain in an idle state until motion is detected, optimizing resource utilization and reducing unnecessary data processing. The use of passive infrared (PIR) sensors for motion detection is a cost-effective and efficient solution. PIR sensors detect infrared radiation changes caused by moving objects, triggering actions like capturing images or sending alerts. This real time responsiveness significantly enhances the utility of modern surveillance systems, making them more proactive in detecting potential security breaches.

The ESP-32 CAM module is a compact and versatile device that plays a crucial role in modern IoT-based surveillance systems. It combines image capture and wireless connectivity capabilities, making it ideal for real-time surveillance applications. Upon detecting motion via a PIR sensor, the ESP-32 CAM module captures high-resolution images of the surveillance area. These images are then uploaded to a cloud-based storage platform, such as Google Drive, for secure storage and further processing. The integration of this module not only simplifies the hardware setup but also ensures cost-effectiveness and scalability, making it a preferred choice for IoT-enabled projects.

The use of AI further enhances the proposed system's capabilities. Specifically, the YOLO (You Only Look Once) algorithm is employed for real-time object detection in captured images. YOLO processes the entire image in a single instance, identifying and classifying objects, such as individuals, with exceptional speed and accuracy. By automating the analysis of surveillance footage, YOLO reduces the dependency on human operators, making the system more efficient and reliable. This feature is particularly beneficial in identifying potential security threats promptly, ensuring a swift response to suspicious activities.

One of the defining features of this project is its integration with cloud-based storage systems. The captured images are uploaded to a secure cloud database, ensuring that the data is not only safely stored but also accessible from anywhere. Cloud storage eliminates the need for large local storage devices, making the system highly scalable and cost-effective. Additionally, the cloud facilitates seamless sharing of data with centralized computers for further analysis, enhancing the system's operational efficiency. This scalability makes the proposed solution suitable for a wide range of applications, from small residential setups to large-scale commercial deployments.

This project aims to combine the strengths of IoT and AI technologies to address the limitations of traditional surveillance systems. By integrating motion detection, real-time image capture, cloud storage, and AI-based analysis, it provides a comprehensive solution to modern security challenges.

The system is designed to be efficient, cost-effective, and easy to scale, making it suitable for diverse applications. Furthermore, its ability to automate processes such as motion detection and person identification ensures minimal human intervention while maximizing reliability and accuracy.

In conclusion, the proposed project represents a significant advancement in surveillance technology by leveraging IoT and AI. By addressing the inefficiencies of traditional systems and introducing automation and intelligence, it offers a robust and scalable solution for modern security needs. The system's ability to detect motion, capture images, and identify individuals in real-time makes it an indispensable tool for ensuring safety and security in today's dynamic environments. The detailed exploration of these technologies and their integration into the project lays a strong foundation for its successful implementation.

## LITERATURE REVIEW

The integration of PIR sensors, ESP-32 CAM, YOLO for object detection, and cloud-based storage has been explored through various studies that lay the groundwork for the development of an efficient and cost-effective surveillance solution. Srilakshmi and Padma [1] offered crucial insights into the fundamental principles of PIR sensors and their application in motion detection, which served as the starting point for the motion detection mechanism in this project. Rai and Rehman [2], building on this, enhanced PIR sensor capabilities by incorporating machine learning algorithms to reduce false positives and increase accuracy, providing a foundation for improving motion detection in dynamic environments. Paunekar et al. [3] highlighted the use of ESP-32 CAM in real-time surveillance systems, emphasizing its potential for cloud integration and scalability, which was essential for the design of the image capture and transmission component of the system. The YOLO object detection framework, as introduced and later optimized, provided the real-time object identification capabilities that are key to detecting human presence in captured images. Jadhav et al. [4] further demonstrated how combining AI with PIR sensors can elevate the intelligence of surveillance systems, reinforcing the decision to incorporate AI driven analysis in this project. Balla and Jadhao [5] finally emphasized the importance of cloud-based solutions in modern surveillance systems, influencing the integration of cloud storage for secure and scalable image management. The collective contributions of these studies have been instrumental in shaping the design and functionality of the proposed surveillance system, ensuring its reliability, efficiency, and scalability.

The collective body of literature reviewed provides substantial insights that directly influenced the design and development of the proposed surveillance system. Namdeo et al. [6] demonstrated the foundational role of PIR sensors in motion detection, establishing the basis for their use in this project to trigger the ESP-32 CAM module. Jain et al. [7] advanced the motion detection capabilities by incorporating machine learning algorithms, highlighting the potential for enhanced accuracy, which led to the decision to combine PIR sensors with AI-driven analysis for improved precision in motion detection.

The research affirmed the practical application of the ESP-32 CAM in surveillance systems, reinforcing the choice of this module for real-time image capture and cloud integration. This study, alongside the work, that introduced the YOLO architecture, paved the way for incorporating YOLO's real-time object detection capabilities, ensuring that the system could accurately identify people within the captured images [8]. Further advancements in YOLO, as explored, provided the necessary optimization for efficient real-time detection on edge devices, confirming the suitability of YOLOv4 for this project.

Kothari et al. [9] contributed valuable insights into the integration of AI with PIR sensors, reinforcing the decision to implement YOLO for automated object detection, reducing the need for manual intervention and improving system intelligence. Lastly, Sails et al. [10] emphasized the advantages of cloud-based surveillance, which inspired the incorporation of cloud storage solutions like Google Drive for secure, scalable, and remotely accessible image storage.

Together, these studies formed a comprehensive knowledge base that guided the integration of motion detection, image capture, object detection, and cloud storage, ultimately resulting in the development of a modern, efficient, and scalable smart surveillance system.

## METHODOLOGY

### ESP-32 CAM Module

The ESP-32 CAM module is the central component of the surveillance system, responsible for capturing video and images. It is a low-cost, compact microcontroller with integrated Wi-Fi and Bluetooth capabilities, making it ideal for IoT-based applications. Key features of the ESP-32 CAM include:

- *Camera Module:* The OV2640 camera supports resolutions up to  $1600 \times 1200$  (UXGA) for high-quality image and video capture.
- *Connectivity:* Built-in Wi-Fi ensures seamless communication with other devices and cloud platforms.
- *MicroSD Card Slot:* Enables local storage for images and videos.
- *Power Efficiency:* Operates at low power, making it suitable for long-term deployments.

The ESP-32 CAM captures images and streams video to a cloud database or a central computer for further analysis. Its low cost and versatility make it the backbone of the system.

### Passive Infrared Sensor

The PIR sensor is crucial for motion detection. It detects infrared radiation emitted by objects in its field of view, enabling the system to sense movement. Key functionalities include:

- *Motion Detection:* When a human or object moves, the PIR sensor detects the change in infrared radiation levels and sends a signal to the ESP-32 CAM module.
- *Energy Efficiency:* Operates only when motion is detected, reducing unnecessary power consumption.
- *Integration:* The sensor's output is used to trigger the ESP-32 CAM module to capture images or record videos.

The PIR sensor ensures that the system is event-driven, capturing only relevant activity and conserving resources such as storage and bandwidth.

### YOLO

YOLO (You Only Look Once) is a real-time object detection algorithm that plays a significant role in identifying and classifying objects within an image or video stream. It processes the entire image in a single pass through a neural network, enabling rapid and accurate detection. Key functionalities include:

- *Object Detection:* YOLO divides the image into a grid and simultaneously predicts bounding boxes and class probabilities, allowing it to detect multiple objects in a single frame.
- *Real-Time Performance:* Due to its single-pass architecture, YOLO is highly efficient and suitable for real-time applications such as live surveillance.
- *Accuracy and Speed Balance:* Offers a good trade-off between detection accuracy and processing speed, making it ideal for embedded systems with limited computational power.
- *Integration:* Can be used in conjunction with the ESP-32 CAM to analyze captured frames for the presence of specific objects, enhancing the system's intelligence.

YOLO enhances the surveillance system by adding an intelligent layer of object recognition, enabling automated responses and more detailed event logging based on detected entities.

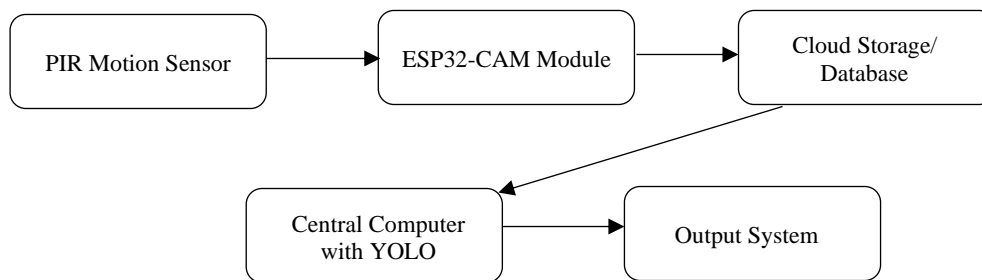
The proposed surveillance system integrates motion detection, image capturing, cloud-based storage, and AI-driven analysis for an efficient and intelligent security solution.

The block diagram in Figure 1 provides an overview of the system's architecture and data flow. It comprises the following components:

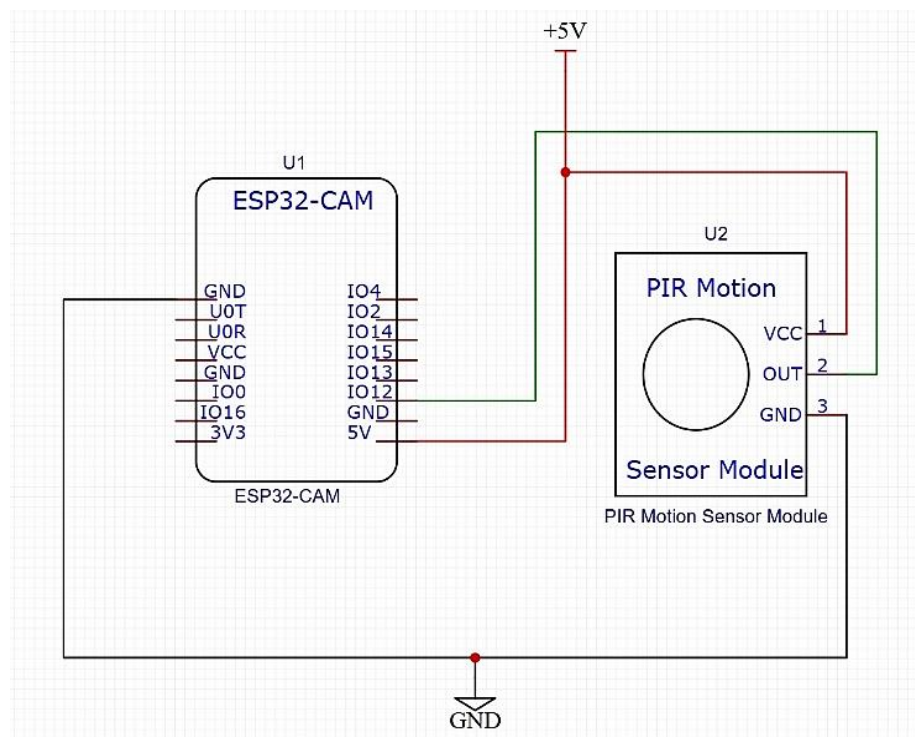
- *PIR Motion Sensor*: Detects motion and sends a trigger signal to the ESP-32 CAM module.
- *ESP-32 CAM Module*: Captures a snapshot upon receiving the trigger and uploads the image to the cloud.
- *Cloud Storage/Database*: Stores the captured images for later retrieval and analysis.
- *Central Computer with YOLO*: Downloads images from the cloud and processes them using the YOLO object detection algorithm to identify humans.
- *Output System*: Displays the results of YOLO analysis, indicating whether a person has been detected.

The circuit diagram in Figure 2 illustrates the physical connections between the components. Key connections include:

- *PIR Sensor to ESP-32 CAM*: The PIR sensor's output pin connects to a GPIO pin on the ESP-32 CAM to signal motion detection.
- *Power Supply*: Both the ESP-32 CAM and PIR sensor are powered via a 5-V source.
- *ESP-32 CAM Output*: The ESP-32 CAM communicates with the cloud through Wi-Fi, requiring minimal external circuitry for operation.



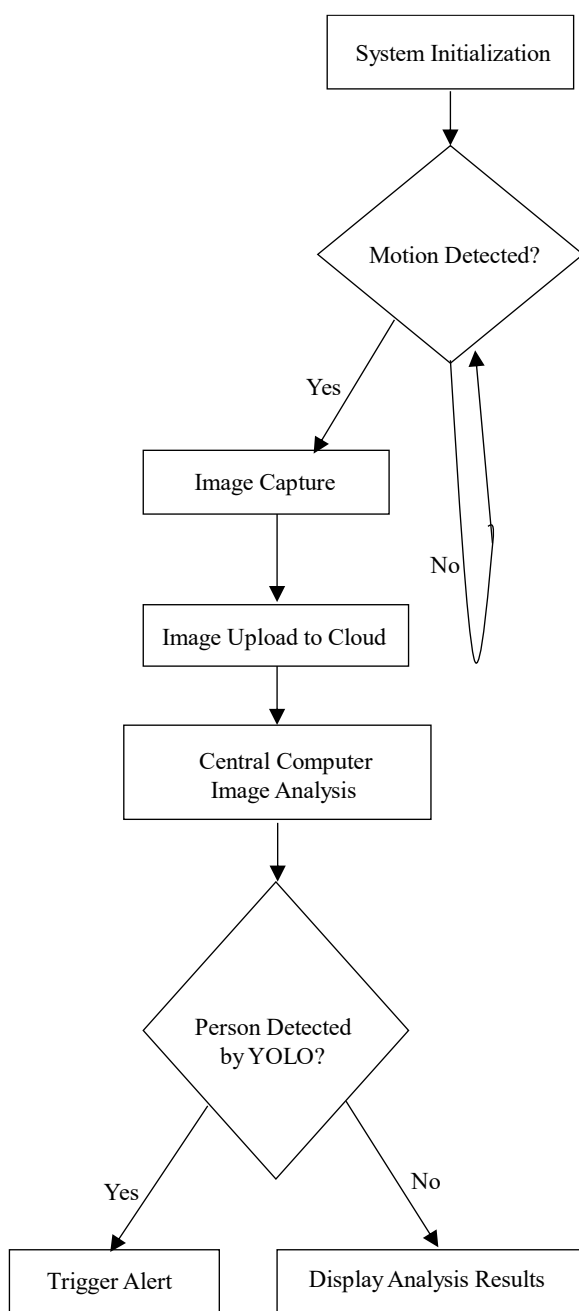
**Figure 1.** System's architecture and data flow.



**Figure 2.** Scheme of video surveillance camera.

The flow chart as shown in Figure 3, describes the sequential operation of the system:

1. *System Initialization*: Initialize the ESP-32 CAM module, PIR sensor, and establish a Wi-Fi connection.
2. *Motion Detection*: Monitor the PIR sensor for any motion.
  - a. If motion detected: Proceed to Step 3.
  - b. If no motion detected: Continue monitoring.
3. *Image Capture*: Trigger the ESP32-CAM to capture a snapshot.
4. *Image Upload*: Send the captured image to cloud storage or a database.
5. *Image Analysis*: The central computer retrieves the image from the cloud and uses the YOLO object detection algorithm to analyze it, specifically identifying the presence of a person.
6. *Output Results*: Display the YOLO analysis results on a user interface or trigger an alert if a person is detected.



**Figure 3.** Flowchart of real-time video surveillance.

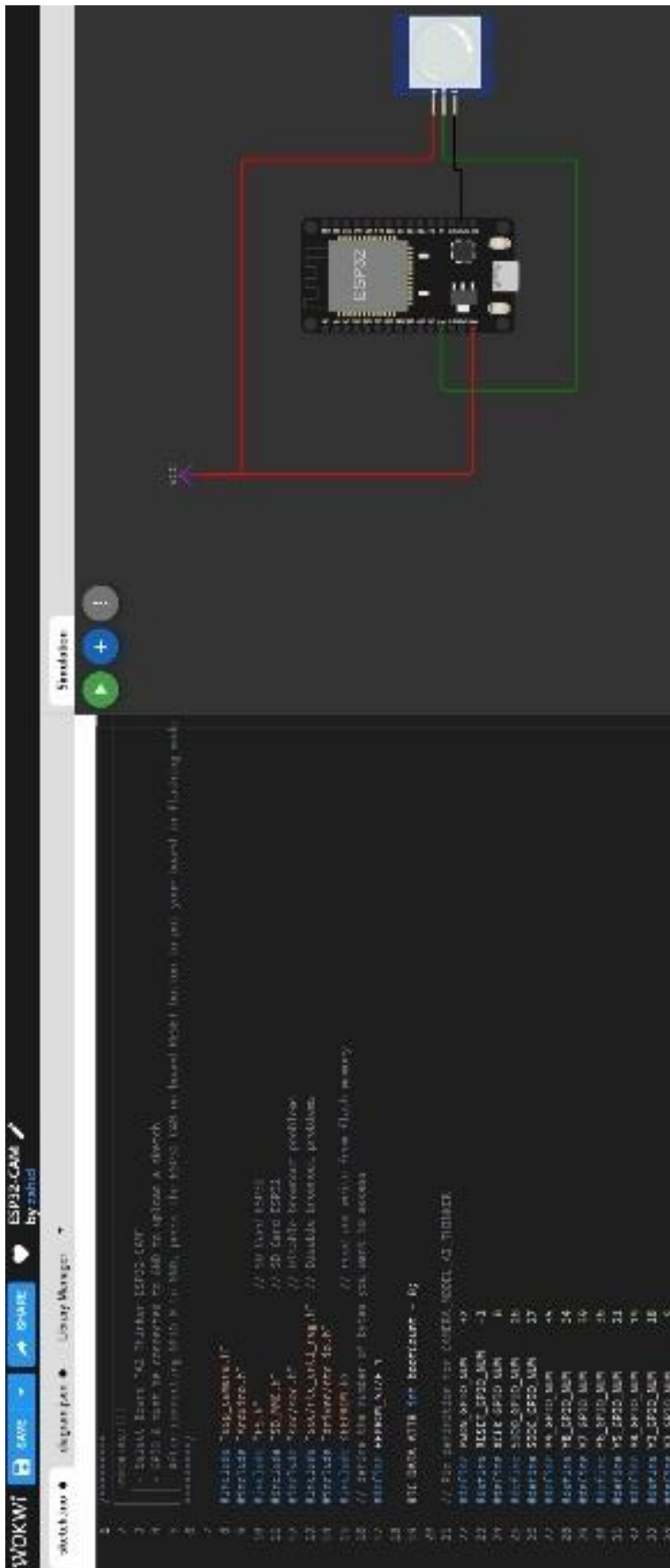


Figure 4. Wokwi simulation.

## RESULTS

The implementation of the proposed surveillance system has been carried out using Wokwi, an advanced online simulation platform for IoT and embedded systems development. Wokwi provides an intuitive, real-time environment for testing hardware and software designs virtually, eliminating the need for physical components during the initial development stages (Figure 4).

The results obtained from the Wokwi simulation and subsequent YOLO analysis are detailed as follows:

### Simulation Results

- *Motion Detection:* The PIR motion sensor successfully detected movement and sent a trigger signal to the ESP-32 CAM module.
- *Image Capture and Upload:* The ESP-32 CAM captured snapshots upon motion detection and uploaded them to cloud storage without significant latency.
- *System Response Time:* The average time between motion detection and image upload was approximately 1.2 seconds, ensuring real-time responsiveness.

### YOLO Analysis Results

The captured images were processed using YOLO, an advanced real-time object detection framework (Table 1). Performance highlights include:

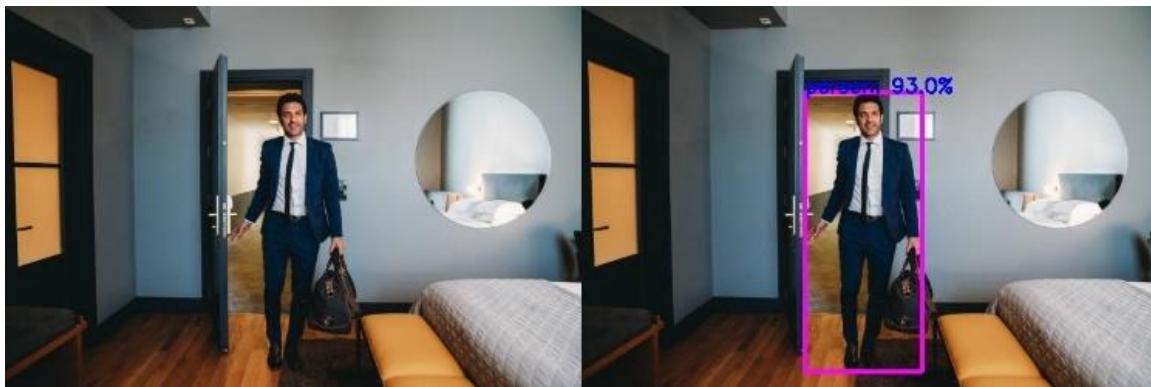
- *Accuracy:* YOLO demonstrated a detection accuracy of ~90% in normal lighting conditions and ~80% in low-light scenarios.
- *Processing Time:* Each image was processed in ~0.4 seconds, ensuring timely threat identification.
- *False Positives:* Occasional false positives were observed, primarily due to environmental artifacts, which were reduced with parameter tuning.

### Captured Image Analysis Results

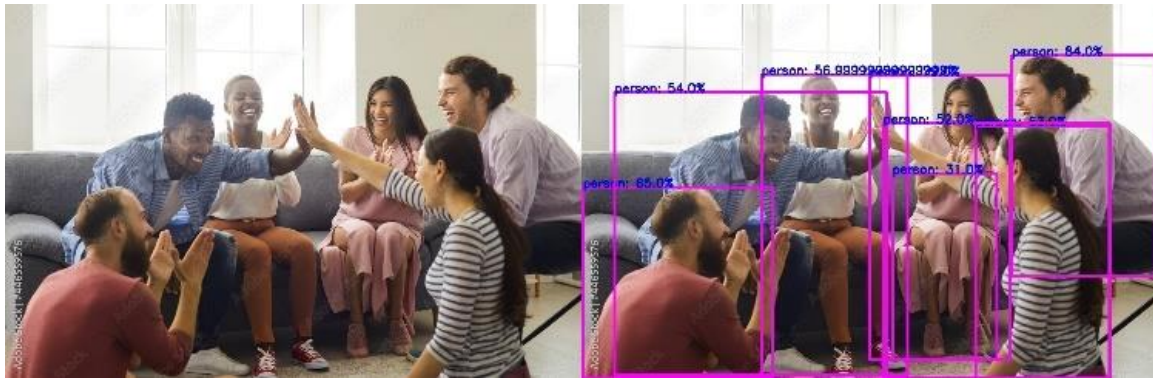
A clear daytime image where the system successfully identified a single individual in the frame with a high confidence score of 93% (Figure 5). The bounding box accurately outlines the person, showcasing the system's effectiveness in optimal lighting conditions.

**Table 1.** Performance metrics.

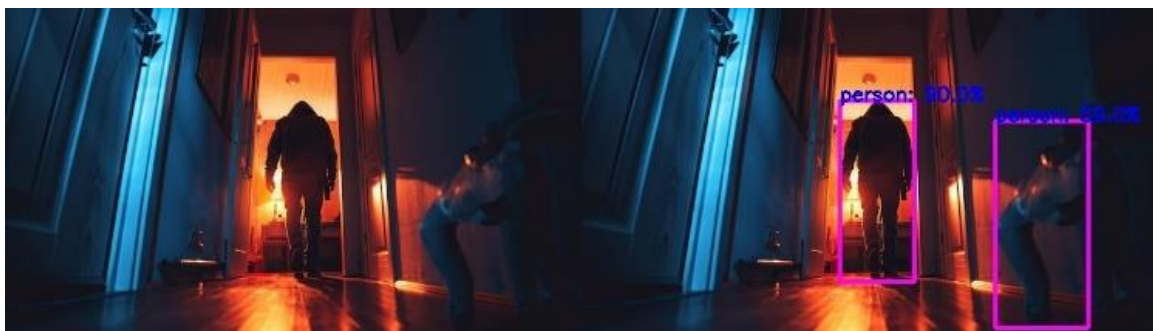
S.N.	Parameter	Value
1	Motion detection accuracy	~95%
2	Image upload latency	~1.2 seconds
3	Detection accuracy	~90%
4	Detection time	~0.4 seconds/image



**Figure 5.** Test image 1.



**Figure 6.** Test image 2.



**Figure 7.** Test image 3.

A crowded scene where the model detected multiple individuals, each assigned a confidence score ranging from 55% to 90% (Figure 6). The bounding boxes were almost properly placed around each person, demonstrating the system's capability to handle complex scenarios with multiple objects of interest with adequate accuracy.

A low-light scenario tested the system's robustness, where one individual was identified with a confidence score of 90%, along with a false positive detection (Figure 7). Although the accuracy was slightly reduced due to challenging lighting, the detection remained effective, confirming the system's ability to adapt to suboptimal environments.

## CONCLUSION

The ESP-32-based DIY (do-it-yourself) wireless CCTV camera project represents a significant step toward accessible and affordable home and small-business security systems. This solution leverages the power and versatility of the ESP-32 microcontroller, which is equipped with built-in Wi-Fi and Bluetooth capabilities, enabling seamless wireless communication and remote monitoring. By utilizing readily available components, such as an ESP-32 board, a compatible camera module, and open-source libraries, this project demonstrates how individuals can create a functional, efficient, and customizable surveillance system without requiring high technical expertise or a substantial financial investment.

The project underscores the adaptability of the ESP-32 for handling multimedia tasks, including real-time video streaming and image capture. This makes it suitable for diverse applications ranging from basic security monitoring to more complex tasks like motion detection and live alerts. Additionally, it showcases how open-source platforms and collaborative development efforts can empower users to design and implement solutions tailored to their specific security needs.

Beyond its practical application, this project serves as an excellent educational tool, providing an opportunity for enthusiasts, students, and professionals to learn about embedded systems, IoT

technologies, and electronics. It bridges the gap between theoretical knowledge and hands-on implementation, fostering innovation and creativity in the tech community.

The system's modular nature ensures that it can be enhanced with additional features, such as cloud integration, AI-driven object detection, or smart home compatibility. This flexibility means it is not only a valuable security solution today but also a platform that can evolve to meet future demands.

In a world increasingly focused on connectivity and automation, the ESP-32 wireless CCTV camera exemplifies how technology can be harnessed to improve security, reduce costs, and promote self-reliance. By empowering users to create their own security solutions, this project contributes to a growing movement of DIY technology enthusiasts who seek to control and customize their environments. It also highlights the potential of IoT devices to make advanced technologies accessible to a broader audience, paving the way for further innovation in the field of affordable and sustainable security systems.

Overall, this project demonstrates how a small, cost-effective microcontroller like the ESP-32 can be at the heart of a robust and reliable surveillance solution. With its adaptability, scalability, and ability to integrate with other technologies, the ESP-32-based wireless CCTV camera is not just a project—it is a foundation for exploring the exciting future of IoT-enabled security.

## REFERENCES

1. Srilakshmi CM, Padma DM. IoT based smart surveillance system. *Int Res J Eng Technol.* 2017; 4 (5): 2222–2224.
2. Rai P, Rehman M. ESP32 based smart surveillance system. In: 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, January 30–31, 2019. pp. 1–3.
3. Paunikar R, Thakare S, Anuse U, Balkhande BW. Action recognition using surveillance system. *Int J Eng Appl Sci Technol.* 2020; 4 (12): 658–660.
4. Jadhav MP, Suryawanshi MS, Jadhav MD. Automated video surveillance. *Int Res J Eng Technol.* 2017; 4 (5): 3564–3567.
5. Balla PB, Jadhao KT. IoT based facial recognition security system. In: 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, January 5, 2018. pp. 1–4.
6. Namdeo R, Sharma S, Anand V, Lohi C. Smart automated surveillance system using Raspberry Pi. *Int J Recent Technol Eng.* 2020; 9 (2): 308–311.
7. Jain A, Basantwani S, Kazi O, Bang Y. Smart surveillance monitoring system. In: 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), Pune, India, February 24–26, pp. 269–273.
8. Telkar AK, Gadgay B. IoT based smart multi application surveillance robot. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, July 15–17, 2020. pp. 931–935.
9. Kothari SB, Ahirrao V, Pawar M, Kapale U, Arjun M. Survey on smart security surveillance system. *Int J Res Appl Sci Eng Technol.* 2021; 9 (12): 219–222.
10. Sails VE, Chakrasali RS, Chimkode DR. A survey on human action recognition. *Int Res J Eng Technol.* 2019; 6 (5): 82–85.