



REAL-TIME DETECTION OF PROHIBITED OBJECTS IN PUBLIC SPACES BASED ON IOT AND DEEP LEARNING

**Donatien Kadima Muamba¹, David Muanza Lubukayi², Ali Akake
Nengo³**

*Corresponding author

doschelah@gmail.com

¹Faculté des Sciences Informatiques, Université Révérend KIM Congo

²Institut Supérieur Pédagogique Congo

³Faculté des Sciences Informatiques, Université Islamique au Congo

E-Mail : doschelah@gmail.com¹, dmuanza05@gmail.com²,

akakenengo@gmail.com³

Article history:

Received : March 22, 2026

Revised : March 29, 2026

Accepted : April 15, 2026

Keywords

IoT, computer vision, deep learning, intelligent monitoring, ESP32-CAM

Abstract

Security in public spaces is a major challenge in the face of diversifying threats. Traditional video surveillance systems, relying on continuous human supervision, have limitations in terms of responsiveness and reliability. This article proposes an intelligent system for detecting prohibited objects, combining the Internet of Things (IoT) and deep learning. The architecture is based on an embedded ESP32-CAM module for image acquisition and a backend server using a deep learning model for analysis. Experimental results show an overall accuracy of 92.8%, demonstrating the suitability of this approach for real-time automated surveillance applications.

1.0. INTRODUCTION

The rapid increase in security threats across public environments such as airports, educational institutions, transportation hubs, and shopping centers has intensified the demand for more effective surveillance technologies. Conventional surveillance systems primarily depend on human operators to monitor video streams and identify suspicious activities or prohibited objects. While these systems have been widely adopted, their effectiveness is often constrained by the limitations of human observation, including fatigue, distraction, and reduced vigilance during prolonged monitoring periods. As a result, critical incidents may go unnoticed, potentially compromising public safety and security [1], [2], [3].

Advancements in Artificial Intelligence (AI), particularly in Deep Learning, have created new opportunities for the development of intelligent surveillance systems capable of automatically identifying threats in real time. Deep Learning models can learn complex visual patterns from large-scale datasets and achieve high levels of accuracy in object detection and classification tasks. These capabilities make AI-based surveillance systems more reliable and responsive than traditional approaches, enabling automated monitoring that reduces dependence on human intervention while improving detection performance [1], [2].

Despite these advancements, existing intelligent surveillance solutions still face significant challenges. Many current systems rely on computationally intensive Convolutional Neural Networks (CNNs) that require substantial processing power, memory resources, and energy consumption. Such requirements limit their applicability in edge-computing environments and

embedded Internet of Things (IoT) devices, where computational resources are often constrained. Consequently, there is a growing need for lightweight and efficient models that can deliver high detection accuracy while maintaining low computational overhead[4], [5][6]

To address these challenges, this study proposes an integrated intelligent surveillance framework that combines IoT technology with Deep Learning-based object detection. The proposed system is designed to operate on low-cost embedded hardware platforms, enabling practical deployment in various real-world environments. By leveraging lightweight processing capabilities and efficient communication mechanisms, the system aims to provide an affordable and scalable solution for automated monitoring without sacrificing detection performance [7]

A key component of the proposed framework is the implementation of advanced object detection models capable of recognizing prohibited or potentially dangerous objects in near real time. The architecture integrates multiple system layers, including sensing devices, embedded processing units, backend services, databases, and web-based monitoring interfaces. This comprehensive design facilitates seamless data acquisition, processing, storage, and visualization, thereby supporting continuous surveillance and rapid decision-making processes in distributed environments [6], [8] [9] [10].

Compared with previous studies that primarily focus on standalone object detection models, the proposed approach emphasizes the integration of lightweight Deep Learning techniques within a distributed IoT ecosystem. This combination enhances system efficiency, reduces deployment costs, and improves scalability for practical applications. Furthermore, the system is evaluated using realistic image datasets to assess its effectiveness under diverse operational conditions, ensuring that the proposed solution can meet the requirements of modern intelligent surveillance systems while maintaining near real-time performance [11][12][4].

2.0. SYSTEM ARCHITECTURE

Overview

The system is based on a client-server architecture composed of:

- **ESP32-CAM:** The acquisition system is based on the ESP32-CAM module, A microcontroller integrating a camera and Wi-Fi connectivity. It is responsible for capturing images in real time and transmitting them to the server via HTTP requests. Its low cost, compact size, and wireless communication capabilities make it a suitable solution for embedded surveillance systems.



Figure 1. ESP32-CAM Board

- **Flask Server:** The backend server is implemented with Flask and ensures the reception of images, their processing by the AI model, as well as the storage and provision of results via APIs.
- **AI Model:** The model used is a convolutional neural network (CNN) for classifying images into three categories (firearm, knife, normal). It combines convolutional, pooling, and fully connected layers. Training is performed using the Adam optimizer and the categorical cross-entropy loss function.
- **Database:** The database allows for the storage of detection results and ensures historical and statistical monitoring of the system.
- **Web interface:** The user interface is developed using web technologies (HTML, CSS, JavaScript) and allows for real-time visualization of results. It displays captured images, detected objects, and associated statistics. It interacts with the Flask server via APIs to dynamically retrieve data.

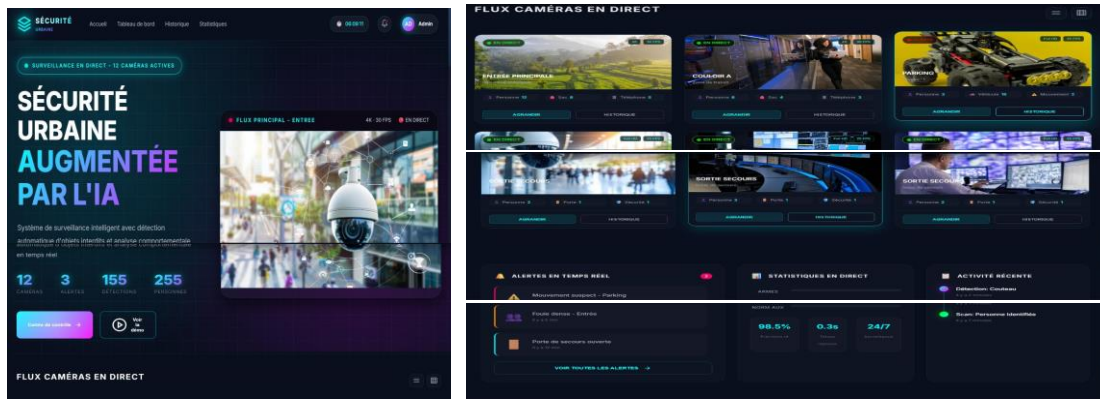


Figure 2. Home page

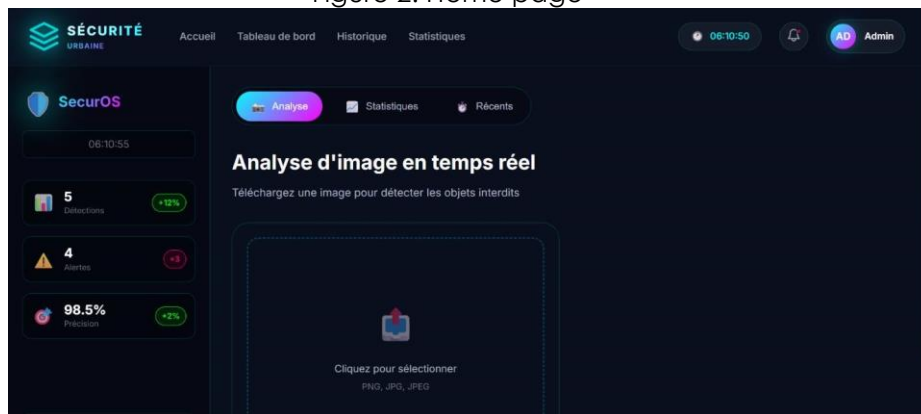


Figure 3. Analysis page



Figure 4. Detection page

Data Flow

The system's data flow is structured in several successive stages: image acquisition by the ESP32-CAM module, transmission via HTTP protocol, data preprocessing, inference by the deep learning model, results storage, and visualization via a user interface. This architecture ensures efficient and near real-time data processing.

3.0. METHODOLOGY

In order to implement the proposed architecture, a structured methodological approach was adopted based on the integration of complementary hardware and software technologies. The acquisition system is based on the ESP32-CAM module, which captures images in real time and transmits them to the server via HTTP requests. This device was chosen because of its low cost, wireless connectivity, and ease of integration into IoT environments.

Image analysis is performed using a deep learning model based on a convolutional neural network. The images are first resized to fit the model's input format and then processed to

automatically extract relevant visual features. The model then assigns each image to a class within the defined categories, along with a confidence score.

Data processing is handled by a backend server developed using the Flask framework. This server receives the images, performs the necessary preprocessing operations, executes the inference using the artificial intelligence model, and makes the results available via programming interfaces. Finally, the detection results are stored in a structured database, allowing for the maintenance of a history of events as well as statistical information. This organization facilitates data retrieval and system performance analysis.

Experimental setup and evaluation metrics

Dataset Description

The system was evaluated on 1200 images divided into three balanced classes, from the Open Images V6 dataset and a manual collection (Google, 2023).

Training configuration

The dataset was divided according to a standard supervised learning protocol: 70% of the images for training, 15% for validation and 15% for testing, in order to ensure a reliable assessment of the model's generalization ability. The model was optimized using the Adam optimizer, chosen for its adaptive learning rate and its efficiency in problems with non-stationary objectives. Training was performed over 25 epochs, offering a suitable compromise between convergence and computational cost, while minimizing the risk of overfitting.

Evaluation Metrics

The model's performance was evaluated using standard metrics widely used in classification:

- Accuracy: measures the proportion of correctly classified instances within the dataset: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$
- Precision: indicates the proportion of instances correctly predicted as positive among all positive predictions: $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$
- Recall: evaluates the model's ability to correctly identify all genuinely positive instances: $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$
- F1-score: corresponds to the harmonic mean of accuracy and recall, providing a balanced evaluation, particularly in the presence of classes of different sizes: $\text{F1} = 2 \times ((\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}))$

In these formulas, TP, TN, FP and FN respectively denote true positives, true negatives, false positives and false negatives.

4.0. RESULTS AND DISCUSSION

Results

The model's performance was evaluated on the test set using the metrics defined in section 5.3. The results obtained are presented in the table below:

Metric	Value (%)
Exactness	92.8
Precision	93.5
Reminder	92.1
F1-score	92.8

These results show high and balanced performance, with a slight superiority in accuracy, indicating that the model minimizes false alarms [5], [12]. Qualitative analysis confirms reliable detection of hazardous objects, a low false negative rate and satisfactory robustness under varied conditions [4], [6].

Discussion

The experimental results demonstrate that the proposed intelligent surveillance system achieves a high level of detection performance, characterized by strong overall accuracy and a balanced trade-off between precision and recall. These findings indicate that the Deep Learning model is capable of reliably identifying prohibited objects while minimizing false

positives and false negatives. Such performance is particularly important in security-critical environments where accurate and timely detection directly influences operational effectiveness. The results also confirm that the integration of lightweight Deep Learning techniques with IoT technology can provide effective monitoring capabilities without requiring extensive computational resources[12].

One of the main strengths of the proposed system lies in its ability to achieve an optimal balance between cost, processing speed, and detection accuracy. Unlike conventional surveillance systems that depend heavily on human supervision, or sophisticated CNN-based approaches that require powerful hardware, the proposed framework is specifically designed for deployment on resource-constrained IoT devices. The use of low-cost embedded platforms reduces implementation expenses while maintaining near real-time performance, making the solution suitable for widespread deployment in public and private security environments. Furthermore, the optimized architecture ensures rapid data processing and efficient communication between system components, enabling immediate responses to detected threats [6].

Another significant advantage of the system is its scalability and adaptability. The modular architecture allows the integration of additional sensors, software services, or analytical components without requiring substantial modifications to the existing infrastructure. This flexibility enables the system to evolve according to future operational requirements and technological advancements. Organizations can expand the surveillance network, incorporate new object categories, or integrate advanced analytics features while preserving the core architecture. Consequently, the proposed solution provides a sustainable foundation for the development of next-generation intelligent surveillance applications.

Despite its promising performance, several limitations remain. The detection accuracy can be influenced by environmental factors, particularly variations in lighting conditions that may affect image quality and object visibility. In addition, the training dataset used in this study is relatively limited, which may restrict the model's ability to generalize effectively across highly diverse real-world scenarios. Another limitation is that the current implementation focuses on single-object detection, reducing its effectiveness in complex environments where multiple prohibited objects may appear simultaneously. These limitations highlight important directions for future research, including the development of more robust illumination-invariant models, the expansion of training datasets, and the implementation of multi-object detection capabilities to enhance the system's overall adaptability and performance.

5.0. CONCLUSION

This study demonstrates that the integration of Internet of Things (IoT) technology and Deep Learning provides an effective and practical solution for intelligent surveillance applications. By combining embedded IoT devices with a lightweight object detection model, the proposed system is capable of identifying prohibited objects with high accuracy while maintaining near real-time performance. The experimental results indicate that the system achieves a favorable balance between precision, recall, processing speed, and computational efficiency, making it suitable for deployment in environments where rapid threat detection is essential. One of the main contributions of this research is the development of a low-cost and scalable surveillance architecture that can operate on resource-constrained hardware platforms. Unlike conventional surveillance systems that rely heavily on continuous human supervision, the proposed framework automates the detection process, reducing response time and minimizing the risk of human error. Furthermore, the modular design enables future expansion through the integration of additional sensors, detection models, and communication services, thereby increasing its applicability in diverse security scenarios. Although the proposed system demonstrates promising performance, several limitations remain, including sensitivity to lighting variations, the relatively limited size of the training dataset, and the current focus on single-object detection. These challenges provide valuable opportunities for future research. Future work should focus on enhancing model robustness under varying environmental conditions, expanding the dataset to improve generalization capabilities, and implementing multi-object detection mechanisms. Overall, the results confirm that the combination of IoT and Deep Learning represents a reliable, scalable, and cost-efficient approach for next-generation intelligent surveillance systems [8][6][3].

REFERENCES

- [1] F. Chollet, *Deep Learning with Python*. Manning Publications, 2018.
- [2] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition. O'Reilly Media, 2019.
- [3] A. Bourehack, "Edge-AI in IoT: Leveraging Cloud Computing and Big Data for Intelligent Decision-Making," *Sensors*, 2023.
- [4] A. C. Cob-Parro, C. Losada-Gutiérrez, M. Marrón-Romera, and A. Gardel-Vicente, "Smart Video Surveillance System Based on Edge Computing," *Sensors*, 2021.
- [5] P. Mittal, "A Comprehensive Survey of Deep Learning-Based Lightweight Object Detection Models for Edge Devices," *Artif. Intell. Rev.*, 2024.
- [6] A. G. Howard, M. Zhu, and B. Chen, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017. [Online]. Available: <https://arxiv.org/abs/1704.04861>
- [7] E. Systems, "ESP32-CAM Documentation," 2023. [Online]. Available: <https://www.espressif.com/en/products/socs/esp32/resources>
- [8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [9] M. Grinberg, *Flask Web Development*. O'Reilly Media, 2018.
- [10] O. Corporation, "MySQL Reference Manual," 2023. [Online]. Available: <https://dev.mysql.com/doc/>
- [11] Google, "Open Images Dataset V6," 2023. [Online]. Available: <https://storage.googleapis.com/openimages/web/index.html>
- [12] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as Points," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.