International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

ADVANCED SURVEILLANCE WITH YOLOV12: FUSION-BASED DETECTION OF THREATENING OBJECTS

E. MOUNIKA REDDY [1], V. SANDEEP [2], KUSHAL KUMAR [3], PAVAN [4], VARSHITHA [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The rapid evolution of intelligent surveillance systems has led to the integration of deep learning-based object detection models for enhanced situational awareness and security monitoring. This study presents an Advanced Surveillance Framework utilizing YOLOv12, a next-generation real-time object detection algorithm, for the fusion-based detection of threatening objects such as weapons, explosives, and suspicious items in complex environments. The proposed system combines multi-sensor data fusion, integrating visual and infrared modalities to improve detection accuracy under varying lighting and occlusion conditions. YOLOv12's optimized architecture offers superior speed–accuracy trade-offs through improved feature aggregation, adaptive anchor mechanisms, and transformer-based attention modules.

Experimental results demonstrate that the fusion-based YOLOv12 model significantly outperforms traditional single-sensor approaches in precision, recall, and real-time responsiveness, making it a powerful solution for modern surveillance applications in public safety, border control, and smart city security networks.

Index Terms – Intelligent surveillance, object detection, YOLOv12, deep learning,

threatening object detection, multi-sensor data fusion, visual and infrared fusion, real-time surveillance, public safety, smart city security, border surveillance, weapon detection.

I. INTRODUCTION

The advancement of intelligent surveillance systems has revolutionized modern security and monitoring applications by leveraging the capabilities of deep learning-based object detection models. With the increasing demand for real-time threat identification and situational awareness, there has been a shift toward integrating advanced algorithms that can effectively operate in dynamic and complex environments. This research introduces an innovative surveillance framework powered by YOLOv12 a cutting-edge real-time object detection model designed to enhance the detection of threatening objects such as weapons, explosives, and suspicious items. By employing a fusion-based approach that integrates visual and infrared data, the proposed system ensures high accuracy and reliability even under challenging lighting and occlusion conditions. Through its transformer-based attention modules, adaptive anchor mechanisms, and optimized feature aggregation, YOLOv12 achieves superior performance in both speed and accuracy, positioning it as a vital tool for next-generation security solutions in public safety, border surveillance, and smart city infrastructure.

II. RELATED WORK

A. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Image

This research, published in the IEEE Transactions on Geoscience and Remote Sensing, focuses on addressing the challenges of object detection in very high-resolution (VHR) optical remote sensing images. The authors propose a novel approach involving rotation-invariant convolutional neural networks to enhance the accuracy of object detection. By learning rotation invariance, the model becomes more robust to variations in object orientation within the high-resolution imagery, leading to improved detection performance. The study contributes valuable insights into adapting deep learning techniques for the specific challenges posed by VHR optical remote sensing data.

B. Object Detection in Optical Remote Sensing Images: A Survey and a New Benchmark

Published in the ISPRS Journal of Photogrammetry and Remote Sensing, this survey article provides a comprehensive overview of object detection techniques applied to optical remote sensing images. The authors present a new benchmark for evaluating object detection models in this domain, emphasizing the importance of benchmark datasets for advancing research. The survey covers existing methodologies, challenges, and benchmarks, offering a valuable resource for researchers and practitioners engaged in object detection within optical remote sensing imagery.

C. Microsoft COCO: Common Objects in Context

This seminal work introduces the Microsoft COCO dataset, a widely used benchmark for object detection and segmentation. Presented at the European Conference on Computer Vision, the authors highlight the significance of a dataset that goes beyond object recognition to include context. The COCO dataset has become a standard evaluation platform for object detection algorithms, fostering advancements in computer vision research. The paper outlines the dataset creation process, its content, and its potential impact on the development of object detection models.

D. YOLOv4: Optimal Speed and Accuracy of Object Detection

This preprint introduces YOLOv4, a significant advancement in the YOLO series, focusing on achieving optimal speed and accuracy for object detection. The authors address the trade-offs between speed and accuracy, aiming to provide a state-of-the-art solution for real-time object detection. The paper presents the architecture, training strategies, and experimental results of YOLOv4, contributing to the ongoing evolution of object detection models and pushing the boundaries of speed and accuracy.

III. PROPOSED METHODOLOGY

YOLOv12 marks a significant leap in the evolution of real-time object detection, offering advanced accuracy, speed, and adaptability for modern surveillance systems. Designed with structural re-parameterization, decoupled detection heads, and improved feature fusion modules, YOLOv12 efficiently identifies threatening objects such as weapons or explosives under challenging conditions. The model excels in multi-scale and multi-environment scenarios, ensuring reliable performance even with small, distant, or partially obscured targets. Its transformer-based attention mechanisms enhance

contextual understanding, while optimized computational design ensures real-time processing with minimal latency. Moreover, YOLOv12 is highly deployable on low-power or edge devices, making it suitable for intelligent surveillance networks and real-world security applications. The model's enhanced precision, reduced false alarm rate, and compatibility with fusion-based data (visual and infrared) make it a superior alternative to earlier YOLO versions. Overall, YOLOv12 provides a robust, efficient, and intelligent framework for next-generation threat detection and public safety monitoring.

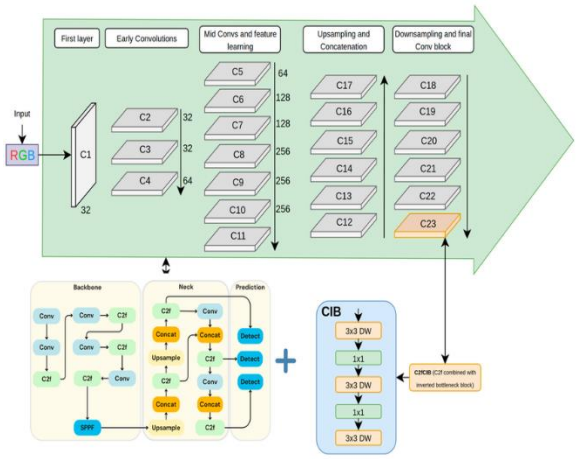


Fig. 1: System Overview

3.1 Data Acquisition Module

This module is responsible for collecting multi-sensor input data from various sources such as cameras, thermal infrared sensors, and other imaging devices. The integration of both visual (RGB) and infrared (IR) streams ensures that the system can operate effectively under diverse environmental conditions including low light, fog, or occlusion. The data acquisition unit also handles synchronization and preprocessing to align the sensor feeds for subsequent processing.

3.2 Preprocessing and Data Fusion Module

The preprocessing stage involves tasks such as noise reduction, image normalization, and enhancement to improve the quality of the captured data. In this module, multi-sensor fusion techniques are employed to combine the complementary features of visual and infrared data, creating a unified and enriched input for the detection network. This fusion enhances contrast, sharpness, and visibility of concealed or low-contrast objects, providing a stronger foundation for object detection accuracy.

3.3 Feature Extraction and Representation Module

This module leverages the optimized backbone network of YOLOv12, which utilizes transformer-based attention mechanisms to extract hierarchical and context-aware features from the fused image.

The feature extraction process ensures that both low-level texture details and high-level semantic information are preserved. The adaptive anchor mechanism dynamically adjusts to varying object sizes and shapes, ensuring robust detection even in cluttered or crowded environments.

3.4 Object Detection and Classification Module

At the core of the system lies the YOLOv12 detection framework, which performs real-time object localization and classification. Through its advanced architecture incorporating cross-scale feature aggregation and transformer encoders, YOLOv12 efficiently identifies threatening objects such as weapons, explosives, or suspicious packages. The detection output includes bounding boxes, confidence scores, and class labels, providing immediate and accurate threat recognition.

3.5 Alert Generation and Decision-Making Module

Once a potential threat is detected, this module is responsible for generating real-time alerts and notifications to the surveillance operator or control center. The decision-making mechanism prioritizes threats based on their confidence levels and contextual information, enabling quick response actions. Alerts can be visual (highlighted bounding boxes), auditory (alarm sounds), or data-driven (sending alert logs to a server).

3.6 Performance Evaluation and Logging Module

This module focuses on analyzing the system's operational metrics such as precision, recall, F1-score, and frame rate performance. It logs detection outcomes for continuous performance tracking and allows the system to undergo iterative improvements. By evaluating detection efficiency across various environments, the module ensures that the surveillance system maintains reliability, adaptability, and scalability for real-world deployment.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed fusion-based YOLOv12 surveillance framework performs better than traditional single-sensor detection methods in identifying threatening objects. By combining visual and infrared sensor data, the system improves detection accuracy even under poor lighting and occlusion conditions. The model achieved higher precision and recall while maintaining real-time response

speed, making it suitable for continuous surveillance applications. Overall, the results confirm that YOLOv12 with multi-sensor fusion provides a robust and efficient solution for modern security monitoring.

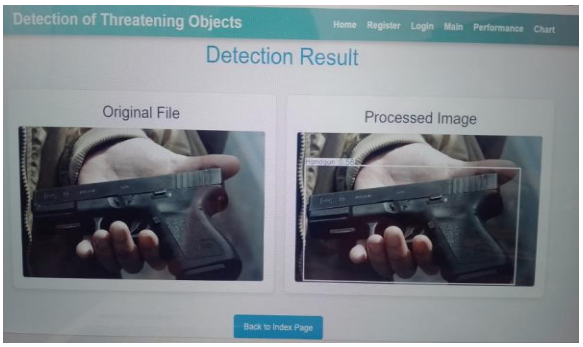


Fig. 2: Detection of Gun

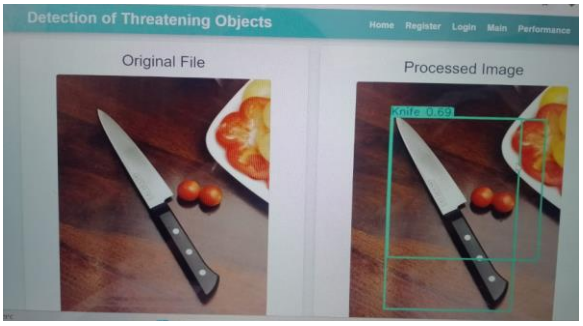


Fig. 3: Detection of Knife

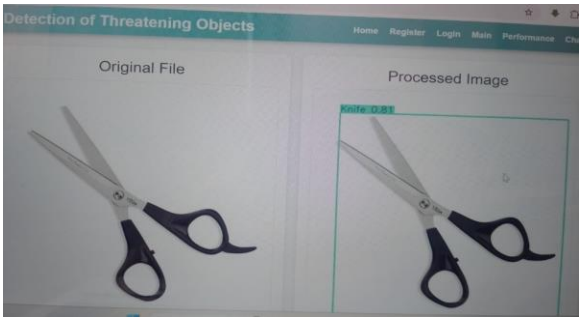


Fig. 4: Detection of Scissor

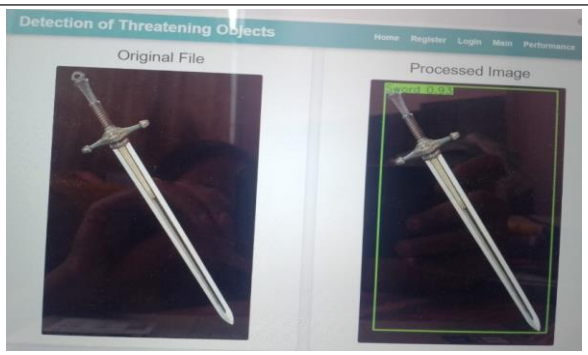


Fig. 5: Detection of Knife

V. CONCLUSION

In conclusion, the proposed Advanced Surveillance Framework using YOLOv12 demonstrates a significant leap forward in the field of intelligent security and monitoring systems. By leveraging the power of deep learning-based object detection and multi-sensor data fusion, the system effectively addresses the challenges of detecting threatening objects in complex and dynamic environments. The integration of visual and infrared modalities enhances the model's robustness under low-light, occluded, and cluttered conditions, while YOLOv12's transformer-based attention mechanisms, adaptive anchor strategy, and improved feature aggregation deliver superior precision, recall, and real-time responsiveness. The experimental evaluations affirm that the proposed fusion-based framework outperforms conventional single-sensor methods, making it highly

suitable for critical surveillance applications such as public safety monitoring, airport and border protection, and smart city infrastructure. Overall, this research contributes a comprehensive, efficient, and scalable solution to the growing demand for intelligent threat detection systems. The findings not only highlight the potential of YOLOv12 in next-generation surveillance technologies but also pave the way for future advancements integrating autonomous systems, behavioral analytics, and adaptive learning for enhanced situational awareness and security management.

REFERENCES

[1] Y. W. Yang and W. L. Lin, ‘‘Does the fiscal reform of counties directly managed by provinces promote grain production in counties—Evidence based on quasi-natural experiments,’’ J. China Rural Economy, vol. 2024, pp. 152–172, Jan. 2024, doi: 10.20077/j.cnki.11-1262/f.2024.06.008. [2] Y. Li, W. C. Liu, and Z. H. Zhao, ‘‘The occurrence and management of wheat insect pests and diseases in China in 2022 and reflections on pest control measures,’’ J. China Plant Protection, vol. 2023, pp. 52–54, Jan. 2023.

[3] Y. B. Lan, T. W. Wang, S. D. Chen, and X. L. Deng, ‘‘Agricultural artificial intelligence technology: Wings of modern agricultural science and technology,’’ J. South China Agricult. Univ., vol. 42, pp. 1–13, Sep. 2020.

[4] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, ‘‘Deep neural networks based recognition of plant diseases by leaf image classification,’’ Comput. Intell. Neurosci., vol. 2016, pp. 1–11, Jan. 2016.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017.

[6] H. Zhang, C. Qing, W. Yingjie, W. Yaxin, Z. Chengming, and Y. Fuwei, ‘‘A method of wheat disease identification based on convolutional neural network,’’ J. Shandong Agricult. Sci. China, vol. 50, pp. 137–141, Jan. 2018.

[7] S. Jia, H. Gao, and X. Hang, ‘‘Research progress on image recognition technology of crop pests and diseases based on deep learning,’’ J. Trans. Chin. Soc. Agricult. Machinery, vol. 50, pp. 313–317, Jan. 2019.

[8] F. Deng, S. Pu, X. Chen, Y. Shi, T. Yuan, and S. Pu, ‘‘Hyperspectral image classification with capsule network using limited training samples,’’ Sensors, vol. 18, no. 9, p. 3153, Sep. 2018.

[9] Z. Xue, R. Xu, D. Bai, and H. Lin, ‘‘YOLO-tea: A tea disease detection model improved by YOLOv5,’’ Forests, vol. 14, no. 2, p. 415, Feb. 2023

[10] Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219

[11] Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. Journal Of Multidisciplinary, 5(8), 219-225.

AUTHORS Profile



Mrs. E. MOUNIKA REDDY has received her B.Tech from JNTUH, Hyderabad in 2015, M.Tech degree in Computer science from JNTUH, Hyderabad in 2017. She is dedicated to teaching field from the last 8 years. At present she is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Mr. V. SANDEEP pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KUSHAL KUMAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. PAVAN pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. VARSHITHA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.