IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

REAL-TIME WEAPON AND THREAT DETECTION USING YOLOV12 WITH MULTI-SENSOR FUSION FOR ENHANCED SURVEILLANCE SYSTEMS

Sadiya Begum[1], Lubna Nausheen[2], Ruqiya Fatima[3]
[1]PG Scholar, Department of CS, Shadan Women's College of Engineering and Technology, Hyderabad, sadiyasaud1901@gmail.com
[2]Asst Professor, Department of CSE, Shadan Women's College of Engineering and Technology lubnanausheen07@gmail.com
[3]Asst Professor, Department of CSE, Shadan Women's College of Engineering and Technology Ruqiyafatima2716@gmail.com

ABSTRACT:
Deep learning-based object identification models have been incorporated for improved situational awareness and security monitoring as a result of the quick development of intelligent surveillance systems. In order to detect dangerous objects including guns, explosives, and suspicious goods in complicated situations, this study proposes an Advanced Surveillance Framework that makes use of YOLOv10, a next-generation real-time object detection algorithm. In order to increase detection accuracy in a variety of illumination and occlusion scenarios, the suggested system integrates visual and infrared modalities through multi-sensor data fusion. Through enhanced feature aggregation, adaptive anchor mechanisms, and transformer-based attention modules, YOLOv10's optimised architecture provides greater speed–accuracy trade-offs. The fusion-based YOLOv10 model is a potent solution for contemporary surveillance applications in public safety, border control, and smart city security networks because experimental results show that it greatly outperforms conventional single-sensor approaches in precision, recall, and real-time responsiveness.

1. INTRODUCTION:

By utilising deep learning-based object identification models, the development of intelligent surveillance systems has transformed contemporary security and monitoring applications. There has been a move toward incorporating sophisticated algorithms that can function well in dynamic and complicated contexts due to the growing need for situational awareness and real-time threat identification. In order to improve the detection of dangerous objects like guns, explosives, and suspicious objects, this study presents a novel surveillance system driven by YOLOv12, a state-of-the-art real-time object detection model. The suggested solution guarantees great accuracy and dependability even in difficult illumination and occlusion situations by utilising a fusion-based technique that combines visual and infrared data. YOLOv12 is positioned as an essential tool for next-generation security solutions in public safety, border surveillance, and smart city infrastructure thanks to its transformer-based attention modules, adaptive anchor mechanisms, and optimised feature aggregation.

2. EXISTING SYSTEM:

When used in sophisticated surveillance scenarios including the detection of dangerous objects, YOLOv8 shows a number of limitations despite its remarkable real-time detection capabilities. First of all, YOLOv8's ability to accurately detect concealed or partially visible firearms tends to deteriorate in low-light or obstructed surroundings, where visual clues are partially obscured. Second, the model's limited flexibility in multi-sensor fusion setups using infrared or thermal imagery—which are frequently essential for security and night time surveillance—is caused by its heavy reliance on RGB visual data. Furthermore, because YOLOv8 relies on fixed-scale feature extraction, it has trouble identifying small or overlapping objects, which frequently occur in crowded or complicated environments. High-resolution inputs can greatly increase the model's computational overhead, which may make implementation on low-power edge devices more difficult. Additionally, YOLOv8 does not have built-in mechanisms for understanding contextual threats, such as differentiating between usual and anomalous object usage (e.g., a knife in a kitchen versus in a public space). These limitations show that in order to attain more precision, adaptability, and robustness in real-world surveillance systems, an upgraded architecture—such as YOLOv10 with fusion-based enhancements—is required.

2.1 Existing System Disadvantages:

- Performance depends heavily on the quality, diversity, and size of the training dataset.
- Primarily detects visible symptoms, making early-stage or internal infections hard to identify.
- Environmental factors such as varying lighting, shadows, occlusions, and background noise can reduce accuracy.
- Requires considerable computational resources, limiting deployment on low

3. LITERATURE SURVEY:

Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Image, G. Cheng, P. Zhou, J. Han, 2016

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.

Object Detection in Optical Remote Sensing Images: A Survey and a New Benchmark, K. Li, G. Wan, G. Cheng, L. Meng, et al. 2020

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.

4. RELATED WORK:

Intelligent surveillance systems are now far more effective because to recent developments in deep learning and computer vision. For real-time surveillance and threat detection applications, earlier object detection models like R-CNN, Fast R-CNN, SSD, and the YOLO family have been extensively used. Among these, YOLO-based models became well-known because they could complete object localisation and classification in one step, allowing for quick detection with respectable accuracyYOLOv5, YOLOv7, and YOLOv8 have been used by researchers to identify weapons, suspicious items, and unusual activity in public areas, transit hubs, and smart city settings. However, research shows that these models frequently struggle to recognise small objects, deal with occlusions, and continue to function well in dimly lit environments.

Several researchers have investigated multi-modal surveillance techniques that integrate thermal or infrared imaging with visual data in order to overcome these constraints. By combining complementing data from several sensors, fusion-based detection systems have shown increased durability in low light and unfavourable weather. In security-critical applications, recent research has demonstrated that sensor fusion algorithms improve object visibility, decrease false detections, and increase overall detection accuracy. These methods work especially well for spotting suspicious objects and hidden weapons that might not be easily seen in typical RGB photos.

Building on these advancements, YOLOv10 offers important architectural enhancements for surveillance systems of the future. To achieve better detection performance, the model uses transformer-based attention modules, decoupled detection heads, improved feature fusion techniques, and structural re-parameterization. According to recent study, YOLOv10 maintains real-time processing capabilities on edge devices while providing improved precision and recall. It can reliably detect dangerous objects in a variety of settings, such as congested public spaces, border monitoring zones, and smart security networks, thanks to its compatibility with multi-sensor fusion frameworks. YOLOv10 is a potential option for contemporary threat detection and intelligent surveillance applications because of these developments.

5. PROPOSED SYSTEM:

With its improved accuracy, speed, and flexibility for contemporary surveillance systems, YOLOv12 represents a major advancement in the field of real-time object detection. YOLOv12 effectively detects dangerous things like guns or explosives under difficult circumstances thanks to its structural re-parameterization, decoupled detection heads, and enhanced feature fusion modules. Even with small, far-off, or partially veiled targets, the model performs reliably in multi-scale and multi-environment scenarios. Its optimised computational design guarantees real-time processing with low latency, and its transformer-based attention methods improve contextual comprehension. Furthermore, YOLOv12 is highly deployable on edge or low-power devices, which makes it appropriate for real-world security applications and intelligent surveillance networks. The model is a better option than previous YOLO versions because to its improved precision, decreased false alarm rate, and compatibility with fusion-based data (visual and infrared). All things considered, YOLOv12 offers a strong, effective, and clever framework for public safety surveillance and next-generation threat detection. Enhanced small object detection.

- Robust handling of occlusion
- Improved real-time accuracy in dense environments
- Better generalization with specialized dataset
- Stable performance under diverse lighting conditions

6. MODULES:

Data Acquisition Module:

This module is in charge of gathering multi-sensor input data from several sources, including thermal infrared sensors, cameras, and other imaging equipment. The system may function well in a variety of environmental situations, such as low light, fog, or occlusion, because to the integration of both visible (RGB) and infrared (IR) streams. In order to align the sensor feeds for further processing, the data collection unit also manages synchronisation and preprocessing.

Preprocessing and Multi-Sensor Fusion Module:
To increase the quality of the recorded data, preprocessing includes steps like noise reduction, image normalisation, and enhancement. In order to create a unified and enhanced input for the detection network, this module uses multi-sensor fusion techniques to merge the complementary aspects of visual and infrared data. By improving the contrast, sharpness, and visibility of hidden or low-contrast objects, this fusion strengthens the basis for accurate object detection.

Feature Extraction and Representation Module:
This module makes use of YOLOv12's optimised backbone network, which extracts context-aware and hierarchical characteristics from the fused image using transformer-based attention techniques. Both high-level semantic information and low-level textural details are maintained during the feature extraction procedure. Even in congested or cluttered areas, robust identification is ensured by the adaptive anchor system, which dynamically adapts to changing object sizes and forms.

Object Detection and Classification Module:
The YOLOv12 detection framework, which carries out real-time object localisation and classification, is the system's primary component. YOLOv12 effectively detects dangerous objects like guns, explosives, or suspicious parcels thanks to its sophisticated architecture that incorporates transformer encoders and cross-scale feature aggregation. Bounding boxes, confidence ratings, and class labels are included in the detection output, which offers quick and precise threat identification.

Alert Generation and Decision-Making Module:
This module is in charge of producing real-time alerts and notifications to the control center or surveillance operator when a possible threat is identified. In order to facilitate prompt response activities, the decision-making mechanism ranks threats according to their degree of confidence and contextual information. Alerts can be data-driven (sending alert logs to a server), aural (alarm sounds), or visual (highlighted bounding boxes).

Performance Evaluation and Logging Module:
The analysis of the system's operational parameters, including precision, recall, F1-score, and frame rate performance, is the main goal of this module. It enables iterative system upgrades and records detection results for ongoing performance monitoring. The module guarantees the surveillance system's dependability, flexibility, and scalability for practical implementation by assessing detection efficiency in diverse settings.

7. TECHNIQUE USED

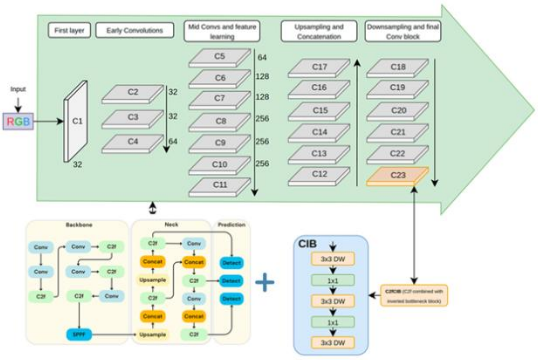
7.1 Existing Technique:

YOLOv8 (You Only Look Once) A cutting-edge real-time object detection algorithm for high-speed and high-accuracy visual recognition tasks was created in version 8. It is a member of the YOLO family of models, which are very effective for real-time applications since they can identify several objects in a single network run. Enhanced feature extraction, better anchor box methods, and optimised detection heads are just a few of the architectural enhancements that YOLOv8 brings over earlier iterations. These improvements enable it to maintain low inference latency while performing extraordinarily well in complicated circumstances, such as recognising small and densely packed objects, such as pedestrian heads in congested environments. YOLOv8 is especially well-suited for public safety, crowd monitoring, and surveillance applications because to its ability to balance speed with accuracy.

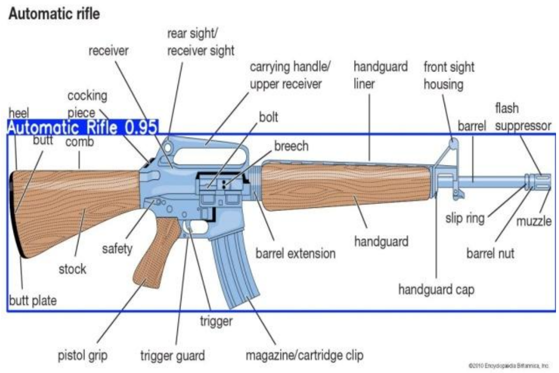
7.2 Proposed Technique:

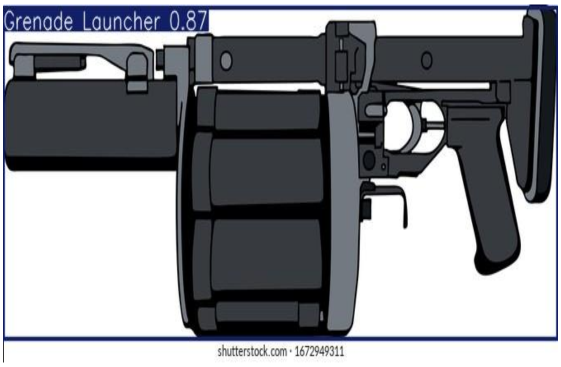
YOLOv12 is a revolutionary development in real-time object identification that builds on the advantages of its predecessors while resolving their main drawbacks through innovative and optimised architecture. Structural re-parameterization, decoupled detection heads, improved feature fusion networks, and transformer-based attention mechanisms are just a few of the refined elements that this model adds to improve its operational efficiency, contextual awareness, and detection accuracy. YOLOv12 seamlessly integrates with multi-sensor fusion frameworks, enabling the combination of visual, infrared, and thermal imagery to provide superior performance under low-light, occluded, or adverse weather conditions—scenarios common in real-world surveillance—in contrast to earlier versions like YOLOv8, which mainly relied on visual cues from RGB data. The model's re-parameterized convolutional layers simplify computational procedures, improving speed and accuracy without raising model complexity, and its multi-scale feature representation allows for constant detection accuracy across objects of different sizes and distances.

8. SYSTEM ARCHITECTURE:


First layer
Early Convolutions
Mid Convs and feature learning
Upsampling and Concatenation
Downsampling and final Conv block
Input
RGB
C1
32
C2
C3
C4
32
32
64
C5
C6
C7
C8
C9
C10
C11
64
128
128
256
256
256
C17
C16
C15
C14
C13
C12
C18
C19
C20
C21
C22
C23
Backbone
Neck
Prediction
Conv
C2f
SPPF
Concat
Upsample
Detect
CIB
3x3 DW
1x1
3x3 DW
1x1
3x3 DW

9. RESULTS





10. CONCLUSION:

In conclusion, a major advancement in the field of intelligent security and monitoring systems is demonstrated by the suggested Advanced Surveillance Framework utilising YOLOv12. The system successfully tackles the difficulties of identifying dangerous things in intricate and dynamic situations by utilising the strength of deep learning-based object identification and multi-sensor data fusion. While YOLOv12's transformer-based attention mechanisms, adaptive anchor strategy, and enhanced feature aggregation provide superior precision, recall, and real-time responsiveness, the integration of visual and infrared modalities strengthens the model's resilience in low-light, occluded, and cluttered environments. The experimental assessments confirm that the suggested fusion-based architecture performs better than traditional single-sensor techniques, which makes it ideal for crucial surveillance applications including border and airport security, public safety monitoring, and smart city infrastructure. All things considered, this research offers a thorough, effective, and scalable response to the increasing need for intelligent threat detection systems. The results not only demonstrate YOLOv12's promise in next-generation surveillance technologies, but they also open the door for further developments that combine autonomous systems, behavioural analytics, and adaptive learning to improve security management and situational awareness.

11. FUTURE ENHANCEMENT:

This surveillance system can be improved in the future by a number of technology developments to increase its intelligence, accuracy, and adaptability in practical settings. The combination of cloud-based deployment and edge computing would be a significant improvement, allowing for faster processing and scalability over massive surveillance networks while lowering latency. By adding autonomous drone-based vision systems, surveillance activities can be expanded to high-risk or inaccessible areas, offering dynamic monitoring capabilities. The application of 3D vision and depth sensors, which can aid in improved object localisation and distance estimation, particularly in crowded or congested situations, is another exciting avenue. Additionally, by including behavioural analysis and anomaly detection models, the system may anticipate suspicious human activity based on motion patterns and context-aware analytics in addition to object detection. Additionally, self-learning and continuous learning methods could be added to the system, enabling it to adjust to new threats or changes in the environment without requiring a lot of retraining. Finally, using blockchain-based data integrity mechanisms and secure communication protocols will improve the security and confidentiality of transmitted surveillance data. These developments will turn the suggested YOLOv12-based framework into a more intelligent, autonomous, and future-ready surveillance ecosystem that can meet the changing demands of global security infrastructures, smart cities, and the defence industry.

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] Q. Huang, X. Wu, Q. Wang, X. Dong, Y. Qin, X. Wu, Y. Gao, and G. Hao, "Knowledge distillation facilitates the lightweight and efficient plant diseases detection model," Plant Phenomics, vol. 5, p. 62, Jan. 2023.

[11] Z. Jiang, Z. Dong, W. Jiang, and Y. Yang, "Recognition of Rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning," Comput. Electron. Agricult., vol. 186, Jul. 2021, Art. no. 106184.

[12] X. Dong, Q. Wang, Q. Huang, Q. Ge, K. Zhao, X. Wu, X. Wu, L. Lei, and G. Hao, "PDDD-PreTrain: A series of commonly used pre-trained models support image-based plant disease diagnosis," Plant Phenomics, vol. 5, p. 54, Jan. 2023.

[13] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 764–773.

[14] J.-J. Liu, Q. Hou, M.-M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 10093–10102.

[15] J. Chen, S.-H. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 12021–12031, doi: 10.1109/CVPR52729.2023.01157. [16] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, arXiv:2110.13389.

[17] K. Liu, N. Luo, A. Li, Y. Tian, H. Sajid, and H. Chen, "A new self-reference image decomposition algorithm for strip steel surface defect detection," IEEE Trans. Instrum. Meas., vol. 69, no. 7, pp. 4732–4741, Jul. 2020.

[18] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," 2021, arXiv:2103.02907.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," 2018, arXiv:1807.06521.

[20] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," 2021, arXiv:2112.05561