INTELLIGENT VIDEO SURVEILLANCE USING DEEP LEARNING

[1]Dr. C. Hari Kishan, [2]Abbaraju Aasritha Sai, [3]Alladi Ramya, [4]Asadi Meghaneswari

[1]Professor, CSE-AI&ML, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India

[2,3,4]U. G Student, Dept CSE-AI&ML, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India

ABSTRACT

The project Intelligent Video Surveillance Using Deep Learning is designed to build a smart real-time system that can detect serious threats like guns, fire, and suspicious activities such as firearm refilling. It combines YOLO for quick object detection with RCNN for precise classification and localization. YOLO provides speed, while RCNN ensures higher accuracy by focusing on specific regions. Once a threat is identified, the system captures the video frame and instantly sends an email alert with the frame and a warning message. This rapid alerting mechanism allows immediate action in high-risk places like schools, banks, and public areas. The model is trained on diverse datasets, enabling it to perform reliably under different lighting, camera angles, and environments. It minimizes false alarms while maintaining strong accuracy. The system is also efficient, supporting continuous monitoring without high computational cost. By sending real-time evidence with alerts, it helps police and emergency teams respond quickly. Overall, this solution enhances public safety by using deep learning for proactive and effective surveillance.

INTRODUCTION

The rising need for advanced security systems has led to the development of intelligent video surveillance capable of real-time threat detection and response. Traditional systems depend on human monitoring, which often suffers from fatigue, delays, and errors. To overcome these limitations, deep learning offers automated, accurate, and fast detection of potential threats. This project introduces a surveillance system using YOLO for speed and RCNN for precise object classification. Together, they detect critical risks such as guns, fire, and suspicious firearm refilling activities. The system captures video frames of detected threats and sends instant email alerts with attached evidence. This reduces dependence on manual monitoring and ensures rapid action in emergencies. The models are trained on diverse datasets to handle variations in lighting, angles, and environments. Designed for scalability, the solution is suitable for schools, banks, airports, and public spaces. By minimizing false positives and providing reliable alerts, the system enhances safety through real-time, intelligent surveillance.

LITERATURE SURVEY

The growing demand for security has led to the need for intelligent video surveillance systems. Traditional surveillance relies on human monitoring, which is prone to fatigue, mistakes, and delays. Deep learning offers an advanced solution by enabling real-time, automated, and accurate threat detection. This project uses YOLO for fast processing and RCNN for precise classification of objects. The system detects threats such as guns, fire, and suspicious activities. Once a threat is identified, it captures the video frame and sends an alert via email with evidence. This ensures quick response and reduces dependence on human supervision. The models are trained on diverse datasets to adapt to different lighting and environments. The solution is scalable and can be deployed in sensitive areas like schools, banks, and airports. By minimizing false positives, it provides a reliable and efficient approach to modern security challenges.

EXISTING SYSTEM

In automated violence detection, traditional methods such as Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and manual analysis have been widely used to identify violent activities in surveillance footage. For example, one study applied a pre-trained C3D network as a feature extractor alongside an SVM classifier to categorize video segments as violent or non-violent, while another research effort on campus violence detection combined image and acoustic features, employing ANNs for classification. Although these approaches show potential, they often suffer from limitations in accuracy and feature representation, making them less effective in real-world scenarios and challenging to deploy reliably.

PROPOSED SYSTEM

The proposed intelligent video surveillance system leverages deep learning by combining YOLO (You Only Look Once) for fast real-time object detection with RCNN (Region-based Convolutional Neural Network) for precise localization, enabling accurate identification of threats such as guns, fire, and refill activities. YOLO ensures rapid processing of video frames, while RCNN enhances precision by focusing on regions of interest, even in complex environments. This integration creates a robust framework that balances speed and accuracy, making it suitable for real-world surveillance applications. When a threat is detected, the system automatically saves the corresponding video frame and triggers an email alert with the frame and a detailed warning message, ensuring immediate notification to security personnel. To improve reliability, advanced filtering techniques are applied to minimize false positives, while the architecture supports continuous monitoring, real-time streaming, and scalability for larger surveillance networks.

SYSTEM ARCHITECTURE

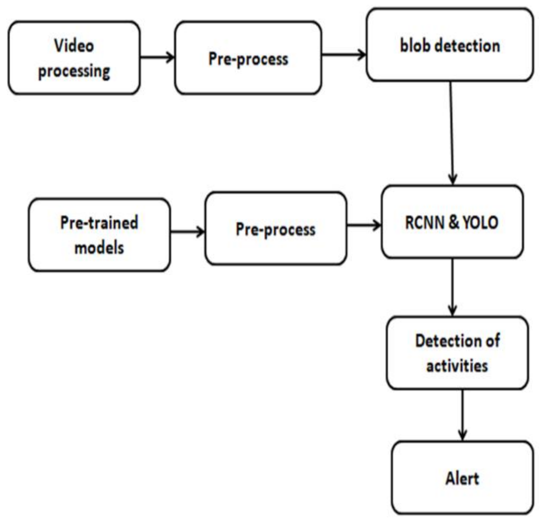

Fig:1 Project Architecture of Intelligent Video Surveillance Using Deep Learning.

The system begins by capturing real-time video streams from surveillance cameras, providing high-resolution feeds for clear monitoring of critical areas. The captured frames are pre-processed through resizing, normalization, and noise reduction to enhance quality and ensure compatibility with detection models. These frames are then visualized with bounding boxes, labels, and confidence scores, giving security personnel clear indications of potential threats. Blob detection is applied to extract significant regions from each frame, isolating critical objects while filtering out irrelevant details. The system loads YOLO configuration files, pre-trained weights, and class labels, while RCNN is initialized to refine localization for improved precision. Blobs are compared against the YOLO and RCNN models to balance real-time speed with detection accuracy. The system is designed to specifically detect guns, fire, and abnormal activities such as fights and refill actions by analyzing posture, movement, and object patterns. When a threat is detected, the corresponding video frame is saved with a timestamp for secure documentation and evidence. To ensure immediate action, the system generates automated email alerts containing the saved frame and a customized warning message. This end-to-end process provides real-time monitoring, accurate detection, and prompt notification, significantly enhancing surveillance effectiveness and security response.

RESULTS AND DISCUSSION

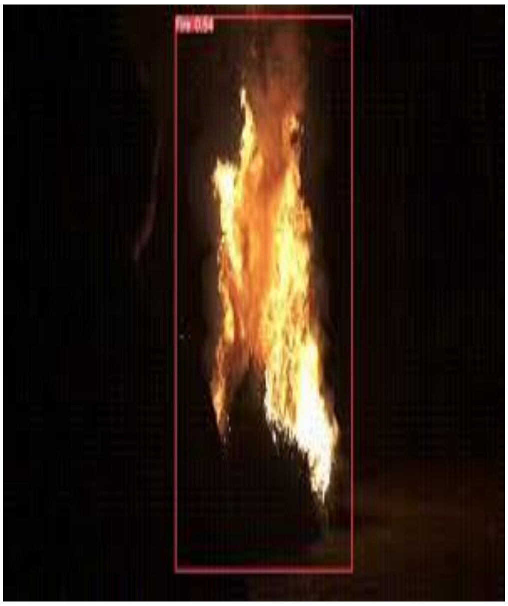

Fig :2 Fire detection using YOLO and RCNN

Fig :3 Gun detection using YOLO and RCNN

The integration of YOLO and RCNN in the intelligent video surveillance system achieved highly accurate detection of guns, fire, and fight scenarios with minimal false positives. YOLO provided rapid, real-time detection at 30 FPS, while RCNN enhanced classification precision through region proposals, creating a balanced solution for speed and accuracy. Preprocessing steps, including blob detection and frame normalization, further optimized performance by handling dynamic backgrounds and lighting variations effectively. The system efficiently stored detected frames with timestamps, ensuring reliable evidence for post-event investigations. Automated email alerts with descriptive messages and attached frames were delivered within seconds, enabling timely responses to threats. Visualization tools displayed bounding boxes and confidence scores, assisting in continuous monitoring and decision-making during emergencies. These combined features demonstrated the system's robustness in real-world applications such as schools, malls, banks, and transport systems. Additionally, frame logging provided a historical record of incidents, crucial for security analysis and law enforcement use. The discussion highlights the effectiveness of combining YOLO's speed with RCNN's accuracy for real-time surveillance. Looking forward, enhancements could include improved accuracy in cluttered environments, mobile device integration for live monitoring, and large-scale deployment for wider security coverage.

CONCLUSION

The proposed Intelligent Video Surveillance System effectively combines YOLO and RCNN models to detect critical threats such as guns, fire, refill actions, and abnormal behaviors in real time. YOLO provides rapid object detection, while RCNN enhances accuracy through region-based classification, ensuring reliable threat identification. Preprocessing steps like blob detection and normalization further improve detection efficiency and reduce false positives. The system automatically saves detected frames with timestamps for evidence and sends email alerts containing images and warning messages to security personnel. This automated process ensures timely responses to potential threats. Testing demonstrated the system's ability to achieve high accuracy while maintaining real-time performance. Its robustness makes it suitable for deployment in sensitive locations such as schools, shopping malls, banks, and public spaces. By minimizing reliance on human monitoring, the system reduces errors and fatigue-related delays. Overall, it provides a scalable, cost-effective, and intelligent solution for enhancing public safety through real-time surveillance.

REFERENCE

1. Sadu, V. B., Kumar, R. S., Kumar, B. S., Kavitha, T., Chapala, H. K., & Chakravarthi, M. K. (2024). Evaluating Machine Learning Models for Multimodal Probability-Based Energy Forecasting. *Process Integration and Optimization for Sustainability*, *8*(4), 1209-1222.

2. Naveen Kumar Polisetty, S., Sivaprakasam, T., Indraneel, S. (2023). A Narrative Framework with Ensemble Learning for Face Emotion Recognition. In: Rao, B.N.K., Balasubramanian, R., Wang, SJ., Nayak, R. (eds) Intelligent Computing and Applications. Smart Innovation, Systems and Technologies, vol 315. Springer, Singapore. https://doi.org/10.1007/978-981-19-4162-7_16

3. Li, et al. "Enhanced Lightweight YOLOX for Small Object Wildfire Detection in UAV Images." PMC, 2023.

4. Kambhatla Akhila, Khaled R Ahmed. "Real Time Deep Learning Weapon Detection Techniques for Mitigating Lone Wolf Attacks." arXiv preprint arXiv:2405.14148, 2024.

5. Anikeit Sethi, Krishanu Saini, Sai Mounika Mididoddi. "Video Anomaly Detection using GAN." arXiv preprint arXiv:2311.14095, 2023.

6. Gopikrishna Pavuluri, Gayathri Annem. "A Deep Learning Approach to Video Anomaly Detection using Convolutional Autoencoders." arXiv preprint arXiv:2311.04351, 2023.

7. Zhang, et al. "R-CNN and YOLOV4 based Deep Learning Model for Intelligent Video Surveillance Application." AIMS Press, 2023.

8. Sethi, et al. "Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey." PubMed, 2023.

9. Qasim, et al. "Video Anomaly Detection System Using Deep Convolutional and Recurrent Models." Results in Engineering, 2023.

10. Garg, et al. "A Multi-Stage Anomaly Detection Scheme for Augmenting Security in IoT-Enabled Applications." Future Generation Computer Systems, 2020.

11. Alzubaidi, et al. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." Journal of Big Data, 2021.

12. Aggarwal, J.K., Ryoo, M.S. "Human Activity Analysis: A Review." ACM Computing Surveys, 2011.

13. Altın, et al. "Machine-Generated Hierarchical Structure of Human Activities to Reveal How Machines Think." IEEE Access, 2021.

14. Wang, et al. "Segment-Tube: Spatio-Temporal Action Localization in Untrimmed Videos with Per-Frame Segmentation." Sensors, 2018.

15. Qiao, et al. "Geometric Features Informed Multi-Person Human-Object Interaction Recognition in Videos." Springer, 2022.

16. Zhang, et al. "Hard No-Box Adversarial Attack on Skeleton-Based Human Action Recognition with Skeleton-Motion-Informed Gradient." IEEE/CVF, 2023.

17. Belay, et al. "Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions." Sensors, 2023.

18. Jeong, et al. "Leveraging Diffusion Models for Unsupervised Out-of-Distribution Detection on Image Manifold." Frontiers in Artificial Intelligence, 2024.

19. Liu, et al. "Leveraging Diffusion Models for Unsupervised Out-of-Distribution Detection on Image Manifold." Frontiers in Artificial Intelligence, 2024.

20. Ho, et al. "Improving Posture Classification Accuracy for Depth Sensor-Based Human Activity Monitoring in Smart Environments." Computer Vision and Image Understanding, 2016.

21. Shum, et al. "Real-Time Posture Reconstruction for Microsoft Kinect." IEEE Transactions on Cybernetics, 2013.

22. Piyathilaka, L., Kodagoda, S. "Gaussian Mixture Based HMM for Human Daily Activity Recognition Using 3D Skeleton Features." IEEE Conference on Industrial Electronics and Applications, 2013.

23. Huang, et al. "High-Speed Multi-Person Pose Estimation with Deep Feature Transfer." Computer Vision and Image Understanding, 2020.

24. Men, et al. "Focalized Contrastive View-Invariant Learning for Self-Supervised Skeleton-Based Action Recognition." Neurocomputing, 2023.

25. Lu, et al. "Hard No-Box Adversarial Attack on Skeleton-Based Human Action Recognition with Skeleton-Motion-Informed Gradient." IEEE/CVF, 2023