Smart Surveillance System Using AI for Real-Time Anomaly Detection

Varad Agarwal[1], Yuvraj Singh Rathore[2], Shyam Singh[3], Indra Kishor[4]
[1,2,3]*Student, Dept. of Computer Engineering, Poornima institute of engineering and Technology, Jaipur, India*
[4]*Assistant Professor, Dept. of Computer Engineering, Poornima Institute of Engineering and Technology, Jaipur, India*

Abstract: **Indeed, considering these technological advancements, smart surveillance systems are very relevant to improving security architectures and reducing the likelihood of human errors that occur in manual monitoring. This paper, premised on advanced AI technologies such as CNN and I3D, postulates a framework that shall enable the realization of an integrated smart surveillance system to detect real-time anomalies in the preventive actions of crimes and traffic incident management. Using such artificial intelligence models enables it to continuously track live video streams; and correctly detect, identify, and classify irregularities such as crimes or accidents. In case of irregularity detection, the system instantly reacts in just a minute by taking visual documentation in the form of snapshots and forwarding instant notice calls to concerned authorities for prompt action. Coupling all these AI technologies allows for really significant improvements both in terms of speed and accuracy and, more generally, dependability, which makes it a powerful resource to enhance public safety across any kind of environment. The paper presents the foundational architecture of the system, conducts a technical review of the components and equipment used, and provides a roadmap for future development and possible use in smart city infrastructures.**

Keywords: **Artificial Intelligence, Smart Surveillance, Real-Time Detection, CNN, I3D Models, Anomaly Detection, Public Safety, Crime Prevention**

1. INTRODUCTION

For a long, surveillance constituted an indispensable element to ensure security in diverse public spaces, from metropolitan complexes to critical infrastructure. Although a conventional manpower-based surveillance system can be very effective in some contexts, it is substantially constrained by the issues of human lag and attention as well as the enormous amount of data to be processed, particularly in those that involve large-scale monitoring. Such inherent handicaps often result in missed events slow response time, and inefficiency, which systematically compromises the effectiveness of traditional surveillance practices.

Within a surprisingly short period, Artificial Intelligence has gained such momentum that surveillance systems have found their new form in the way of more robust, independent, and efficient solutions. Artificial intelligence-based surveillance systems can overcome human-centered limitations as they continuously scan live feeds in real-time to detect anomalies such as criminal activity, traffic accidents, unauthorized activities, or other unusual activities attended. With such accuracy in the detection of objects and activities, far beyond the human brain's ability, these systems have become very handy in improving public safety and security.

The proposed system is a new smart surveillance system that is based on cutting-edge artificial intelligence technologies, which encompasses the capability of Convolutional Neural Networks to identify objects and Inflated 3D Convolutional Networks for real-time anomaly detection. This proposed system aims to improve effectiveness, accuracy, and reliability in surveillance by automatically performing the detection process to give an immediate response capacity. The system has been engineered to raise the proper authorities visibly through pictures documenting the incidents, thus helping them respond to such critical incidents as criminal offenses or vehicular collisions promptly.

Updating conventional security protocols can be of great utility in the building of smart cities, where data observation and analysis are speedy and prompt for proper social life regulation, optimal traffic control, and crime prevention. The systems may be expanded smoothly to provide maximum oversight with a few more instances of human judgment and discretion in cities, that are continuously expanding in their size and scale. This paper introduces the framework, techniques, and possible applications of this

intelligent surveillance system and its probable functionalities and influence on public safety.

Fig. 1(a). UCF Crime Dataset Images

2. LITERATURE REVIEW

Recent developments in the detection and real-time response related to anomalies in newly developed intelligent surveillance systems led to the introduction of innovations in anomaly detection in surveillance systems. Advancements that were more focused on the application of machine learning models, especially object detection and anomaly detection, have been the areas of the highest concern in bringing in any kind of improvement in the efficiency and reliability of surveillance systems. Undoubtedly, the most widely used framework model in object detection is the You Only Look Once model which has remarkable capabilities to identify and classify objects under difficult and high-complexity conditions. The real-time processing ability of YOLO lets it classify objects in one pass only through the neural network feature that fits places particularly well where detection must occur in real-time, such as public spaces, transit centers, and other truly trafficked locales.

At the same time, spatiotemporal analysis frameworks, such as LSTM networks and inflated 3D convolutional networks, have been quite helpful in anomaly detection where activity is time-evolving. These frameworks are not limited to static object detection since they can capture temporal interrelations among events. That would be handy for detecting dynamic or time-evolving activities, such as crime, traffic incidents, or anomalies in surveillance clips. With convolutional model integration, it becomes possible to analyze sequences of video frames, which is fundamental to detecting behavioral patterns, which may threaten or reveal less than typical activities. I3D models extended basic 2D CNNs into yet another further breaking the limitation of just anomaly detection-another step further to even more accurately depict behavior or activities in space and time, like a person suddenly leaving his bag, or erratic driving on one side of a vehicle. Such spatiotemporal models have thus been applied in cases in which correct monitoring must be realized not only in space but also in time as required for object recognition and behavioral analysis.

While the advanced functionalities of current machine learning models, the legacy monitoring systems have some severe shortcomings that ultimately work against their overall efficiency. The legacy system widely depends on alerts or human monitoring of live video streams, which encourages inefficiencies and a high possibility of human failure. Moreover, user monitoring is always a time-consuming liability to latency, particularly in huge environments where plenty of video feeds should be monitored at a given time. Classical systems cannot also satisfy real-time observation on large scales because it does not have any automated intelligence that could identify anomalies within wide scopes and diversified regions.

This intelligent surveillance framework designed for this research solves such problems through the use of models, YOLO for object recognition, and I3D for anomaly detection. Such smart surveillance is designed differently from other traditional systems where the anomalies are instantly detected in real-time, however with an immediate notification sent to concerned authorities together with the visual recording as images of the identified incident. This development reduces reliance on a human operator to a considerable extent and, at the same time, decreases response times significantly. The system can digest large amounts of video data and analyze spatiotemporal patterns, thereby enabling it to alert to a large range of anomalies that might not even be noticed by the human eye, thus raising general efficiency in surveillance.

However such advanced AI-based system inclusions in the monitoring interfaces ensure pertinent scalability in the monitoring of multiple sites at any point in time. The offered cloud computing resources allow for real-time processing of large amounts of video data so the system could function even within metropolitan landscapes or key infrastructures. Scientific literature corpus on intelligent surveillance

systems, hands down, draws attention to the fact that AI-enhanced solutions can transform even the most simplistic surveillance into predictive mechanisms, mechanized systems, and therefore significantly enhanced public safety.

With further advancement in surveillance technologies, future research will focus on adapting even more sophisticated paradigms of artificial intelligence, hybrid systems that couple object detection with behavioral analytics, and systems trained with expertise on multi-modal inputs like audio and video. All this will elevate accuracy and granular monitoring leading to the ability of such systems to identify and respond in real time to both apparent and subtle anomalies.

3. METHODOLOGY

The intelligent surveillance system is developed based on three bases of artificial intelligence models. Such models include; object detection YOLO, video recognition CNN, and spatiotemporal anomaly identification I3D. The system architecture consists of numerous cameras of high resolution at strategic points on different sites, and all of these transmit live video streams to a central, cloud-based server. Through the artificial intelligence models running in tandem with real-time video streams, irregularities are observed. Irregularities may, for instance, consist of criminal activity or traffic among others.

The YOLO model shall promptly identify the objects in the video frames. It is an extremely effective model for the recognition of entities, but these entail vehicles and people; however, the term could also be dangerous items in crowded spaces. Concurrently, CNN analyzes the behaviors and actions as seen. Such tests guarantee proper activity classification as both normal and abnormal.

The I3D model includes the temporal dimension for the classical 2D CNN that manages to spot anomalies in the video by establishing the patterns of behavior over time to catch suspicious or abnormal activities evolving. Once any abnormality is detected the system itself triggers a response by forwarding photographic or short video clip shreds of evidence to the concerned authorities. This system will forward visual evidence and also alert the concerned authorities so that they can immediately assess the situation and perform appropriate action. Such an auto-notifying, instant notice system ensures the responses are prompt, making the system effective in delivering public security. Besides this, the system is equally accompanied by high functions to increase the detection accuracy and effectiveness of the response. It can incorporate other sensors, such as motion detectors and thermal cameras with multi-modal observation possibilities in challenging environments. The system enhances the quality of images during poor lighting or adverse weather conditions through specific enhancement techniques of images.

Table 1: AI Models and Their Functions in the Surveillance System

AI Model	Function	Description
YOLO	Object Detection	Rapidly detects and identifies objects in video feeds, such as people, vehicles, and suspicious items.
CNN	Video Recognition	Analyzes behaviors and actions in the video to classify normal and abnormal activities.
I3D	Anomaly Detection	Analyzes patterns of behavior over time to detect dynamic anomalies like criminal acts or traffic incidents.

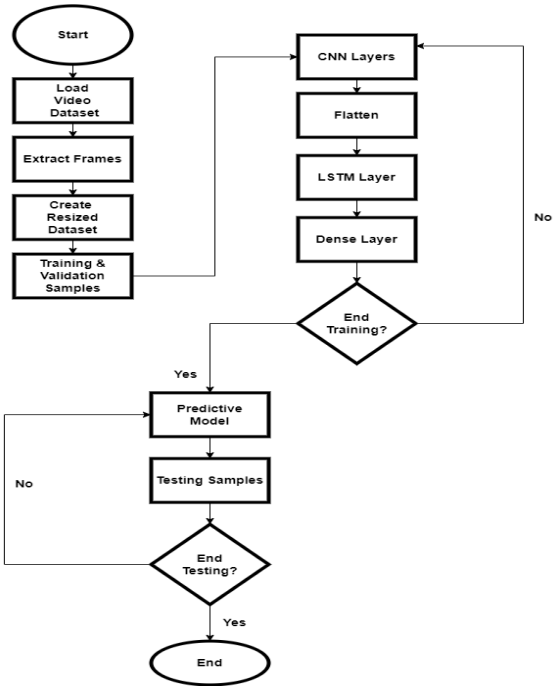

Fig 1: Criminal Activity Detection Method

4. RESULT

The tests of the intelligent surveillance system showed a lot of improvement in terms of detection time and accuracy. It continually reduced response time because it spotted anomalies in real-time, like traffic incidents, illegal activities, and crimes, with ease. This system is capable of object detection because of the use of YOLO and behavioral analysis through the CNN and I3D models, thus identifying anomalies accurately, even complex anomalies that may even be developed over time. Significantly, the system could pick up tiny behavioral cues and distinguish unnatural trends within seconds when such trends started appearing.

This with visual verification, picture or video clips included, and automatically generated alerts that greatly reduce false positives the age-old problem in traditional systems ensures that information received is correct and relevant thus enabling proper decision-making and rapid response. The system improves surveillance accuracy on the whole and is hence highly efficient in raising public safety and operational efficiency.

Table 2: Proposed Enhancements for Future Iterations

Proposed Enhancement	Description
Integration of Transformers	Incorporate advanced AI models to improve detection accuracy and efficiency.
Environmental Anomaly Detection	Expand system capabilities to detect wildfires, floods, and other natural disasters.
Edge Computing	Utilize edge computing to reduce latency and enhance real-time performance.

5. DISCUSSION

This integration of AI-powered surveillance systems is an important step in public safety with the ability to respond much faster and accurately in detecting anomalies in real-time compared to conventional surveillance systems. Often, the conventional systems show inefficiencies and slow response times; however, with AI-powered object and behavior recognition, one can greatly hasten the response times toward accidents, and crimes among others. Such key challenges, such as the computational requirements of real-time processing, have been addressed with the use of architectures from the cloud and optimized AI models like CNN and I3D. These technologies allow for the monitoring of large environments. It may start using natural disasters such as floods or fires over time, therefore making the system more useful. The precision will rise with much more advanced AI models like transformers. Eventually, the system will be far easier to adapt to a wide range of anomaly detection scenarios.

6. CONCLUSION

The structure of the article talks about how smart surveillance can complement less-than-perfect practice through its techniques that use artificial intelligence in real-time anomaly identification. The system is incorporated with models like YOLO, CNN, and I3D; therefore it achieves high public safety through automatic conditions such as a reduction of reliance on manual observation in the detection of crimes and accidents. Its cloud-based nature makes it possible to scale even massive environments in efficient ways. The system generates credible alerts and pictorial records for the authorities; therefore, it forms a basis for rapid and informed decision-making. The flexibility of the AI framework further presents the possibility for its implementation in frameworks designed for smart cities whereby real-time observation is an essential tool in the dispensation of traffic administration, crime prevention, and emergency responses. The system also has the potential of extension in the detection of other anomalies like natural calamities which would make it a significant instrument in the further improvement of urban safety and management.

7. FUTURE SCOPE

Future generations of intelligent surveillance systems will be based on complicated AI models, including transformers, in a balanced hybrid between accuracy of detection and computational efficiency. These super-advanced models will be capable of capturing much more sophisticated patterns and will further improve performance as environments change.

Furthermore, it would expand the application to detect anomalies regarding environments and disasters- the range of wildfires, floods, and many natural disasters- which would make it especially significant and useful in the frameworks of smart cities. This will open an avenue where the system can be a participant in far more extensive public safety efforts than mere crime prevention alone. Further, advancements in edge computing will be pursued so higher performances are yielded in real-time. In terms of latency and bandwidth usage, edge computing works in that data processing is located nearer to the source hastening response times. Taken overall, these efficiencies make for a much more robust, versatile, and efficient solution for challenges posed by diverse urban conditions.

8. REFERENCES

[1] S. S, H. M, D. T and S. S, "Real-time Crime Detection Using Customized CNN," *2022 1st International Conference on Computational Science and Technology (ICCST)*, CHENNAI, India, 2022, pp. 416-419, doi: 10.1109/ICCST55948.2022.10040379.

[2] K. F. Arpa *et al*., "A Machine Learning and Deep Learning Integrated Model to Detect Criminal Activities," *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, Bangkok, Thailand, 2023, pp. 1-6, doi: 10.1109/IBDAP58581.2023.10272002.

[3] V. Belsare, N. Karande, A. Keskar, S. Joshi and R. Karnavat, "Context-Based Crime Detection : A Framework Integrating Computer Vision Technologies," *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, Pune, India, 2024, pp. 1-6, doi: 10.1109/MITADTSoCiCon60330.2024.10575348

[4] R. J. Anandhi, "Edge Computing-Based Crime Scene Object Detection from Surveillance Video Using Deep Learning Algorithms," *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2023, pp. 1159-1163, doi: 10.1109/ICIRCA57980.2023.10220800.

[5] R. K. Yadav and R. Kumar, "Inflated 3D Convolution Network for Detecting Anomalies in Surveillance Videos," *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 2022, pp. 2518-2522, doi: 10.1109/ICAC3N56670.2022.10074290.

[6] S. Rishika, "OBJECT DETECTION USING ARTIFICIAL INTELLIGENCE," vol. 11, no. 5, pp. 882–887, 2022, doi: 10.17148/IJARCCE.2022.115193.

[7] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Inception recurrent convolutional neural network for object recognition," Mach. Vis. Appl., vol. 32, no. 1, 2021, doi: 10.1007/s00138-020-01157-3.

[8] Lu Tan1, Tianran Huangfu1, Liyao Wu1, Wenying Chen1 "Comparison of YOLO v3, Faster R-CNN for Real-Time Pill Identication " Department of Pharmacy, The Third Affiliated Hospital of Southern Medical University, Guangzhou 510000,China.https://doi.org/10.21203/rs.3.rs668895/v1, July 2021

[9] T. J. Nandhini and K. Thinakaran, "CNN Based Moving Object Detection from Surveillance Video in Comparison with GMM," 2022 Int. Conf. Data Sci. Agents Artif. Intell. ICDSAAI 2022, 2022, doi: 10.1109/ICDSAAI55433.2022.10028909.

[10] Y. Wang, N. Xu, A. -A. Liu, W. Li and Y. Zhang, "High-Order Interaction Learning for Image Captioning," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 7, pp. 4417- 4430, July 2022, doi: 10.1109/TCSVT.2021.3121062.

[11] S. Cao, G. An, Z. Zheng and Z. Wang, "Vision-Enhanced and Consensus-Aware Transformer for Image Captioning," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 10, pp. 7005-7018, Oct. 2022, doi: 10.1109/TCSVT.2022.3178844.

[12] Y. Tian, G. Pang, Y. Chen, R. Singh, J. W. Verjans and G. Carneiro, "Weakly-supervised Video Anomaly Detection with Robust Tempo ral Feature Magnitude Learning," 2021 IEEE/CVF International Con ference on Computer Vision (ICCV), 2021, pp. 4955-4966, doi: 10.1109/ICCV48922.2021.00493.

[13] X. Zeng, Z. Li and W. Zhang, "An adaptive learning-based weakly supervised object detection via context awareness," 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering

(ICBASE), Zhuhai, China, 2021, pp. 331-335, doi: 10.1109/ICBASE53849.2021.00068

[14] Y. M. Hemmler and D. Ifenthaler, ”Indicators of the Learning Context for Supporting Personalized and Adaptive Learning Environments,” 2022 International Conference on Advanced Learning Technologies (ICALT), Bucharest, Romania, 2022, pp. 61-65, doi: 10.1109/ICALT55010.2022.00026.

[15] R. J. Anandhi, ”Edge Computing-Based Crime Scene Object Detection from Surveillance Video Using Deep Learning Algorithms,” 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2023, pp. 1159-1163, doi: 10.1109/ICIRCA57980.2023.10220800

[16] O. Hmidani and E. M. Ismaili Alaoui, ”A comprehensive survey of the R-CNN family for object detection,” 2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, 2022. 1-6. 10.1109/CommNet56067.2022.9993862.

[17] C. Wu, P. Zhao, T. Zhang, W. Zhang, F. Meng and R. Guo, ”A Small Target Detection Algorithm of Complex Traffic Road Based on SDCYOLO,” 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT), Qingdao, China, 2023, pp. 1-4, doi: 10.1109/ICEICT57916.2023.10245967

[18] S. R. K. S. Giri, L. P, B. S. S, R. D. Praba, K. Kavitha and A. Kalaiselvi, ”Traffic Surveillance System Using YOLO Algorithm and Machine Learning,” 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICAECA56562.2023.10201021.

[19] Dodia and S. Kumar, ”A Comparison of YOLO Based Vehicle Detection Algorithms,” 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169773

[20] Q. T. Do and J. Chaudri, ”Creating Computer Vision Models for Respiratory Status Detection,” 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 1350-1353, doi: 10.1109/EMBC48229.2022.9871978.

[21] Shah, Neil & Bhagat, Nandish & Shah, Manan. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art. 4. 10.1186/s42492-021-00075-z

[22] M. Pulipalupula, S. Patlola, M. Nayaki, M. Yadlapati, J. Das and B. R. Sanjeeva Reddy, ”Object Detection using You Only Look Once (YOLO) Algorithm in Convolution Neural Network (CNN),” 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-4, doi: 10.1109/I2CT57861.2023.10126213.

[23] S. Kilic and M. Tuceryan, ”Crime Detection from Pre-crime Video Analysis with Augmented Pose Information,” 2023 IEEE International Conference on Electro Information Technology (eIT), Romeoville, IL, USA, 2023, pp. 1-6, doi: 10.1109/eIT57321.2023.10187325.

[24] V. Mandalapu, L. Elluri, P. Vyas and N. Roy, ”Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions,” in IEEE Access, vol. 11, pp. 60153-60170, 2023, doi: 10.1109/ACCESS.2023.3286344.

[25] Kwan-Loo, J. C. Ort´ız-Bayliss, S. E. Conant-Pablos, H. TerashimaMar´ın and P. Rad, ”Detection of Violent Behavior Using Neural Networks and Pose Estimation,” in IEEE Access, vol. 10, pp. 86339-86352, 2022, doi: 10.1109/ACCESS.2022.3198985.