Literature Review On Smart Surveillance: Virtual Test Anomaly Identification Using CNN

Vishal Kumar Laxmi, Research Scholar
Dr. Anurag Aeron, Professor Department Of CSE
Meerut Institute Of Engineering and Technology
Meerut, Uttar Pradesh-250005

Abstract: ***The rapid increase in video data has necessitated smart surveillance in the contemporary security systems because it is their requirement to detect anomalies in real-time. In this survey paper, recent deep learning methods have been discussed, including the particular attention to Convolutional Neural Networks (CNNs) and their variants generated and applied in the virtual test anomaly detection. The paper shows that a variety of background subtraction, CNN-based feature extraction, autoencoders, transformer models, and hybrid deep architectures are being utilized to detect abnormal events with high precision in an automatic fashion. Along those lines, the available studies indicate a high advancement in detecting suspicious actions, enhancing the rate of detection, and lowering the role of human intervention. Nevertheless, there are still large-scale video processing, occlusions, complicated settings and generalization issues. The survey covers the state-of-the-art practices, their weaknesses and limitations, along with the most important research opportunities in order to create smarter, more reliable and live surveillance systems.***

KEYWORDS: ***Smart surveillance, Convolutional Neural Networks (CNNs), anomaly detection, video analytics, deep learning, virtual testing, autoencoders, intelligent monitoring, real-time detection, computer vision.***

I. INTRODUCTION

The implementation of smart surveillance systems is now regarded as a necessary element of a contemporary monitoring setting that allows uninterrupted monitoring and automatic scanning of suspicious behavior. As the artificial intelligence technology has advanced, especially the Convolutional Neural Networks (CNNs), the surveillance technologies have evolved into changing the manual checks to the intelligent, data-guided decisions. These systems do not only increase the level of security but also boosts efficiency in places that may not be easily monitored by humans. With the growing popularity of virtual testing environments in different industries, the capacity to automatically detect anomalies in simulated environments has grown in significance. CNN-based models are of significant importance in identifying abnormal behaviors, unexpected patterns, and technical irregularities in the most precise way.

Virtual test anomaly detection is concerned with observing anomalies in simulated conditions, where an error can have an impact on the performance of the system, safety, or its reliability. CNN models have been developed and are integrated together to enable these systems to learn large-scale simulation data of complex spatial features and patterns. As

opposed to the traditional rule-based systems, deep learning models learn to adjust to the new forms of anomalies as time progresses, and hence, they are applicable in dynamic virtual environments. Through smart surveillance in conjunction with the advanced CNN architecture, organizations are able to realize real time anomaly detection, minimize the human error and streamline testing workflows. The new discipline favors a range of applications, such as industrial control, autonomous vehicle testing, quality verification, and safety testing in virtual reality.

Cities, industries and other public spaces are ever growing; therefore, smart surveillance systems have become part of the contemporary security systems. Conventional surveillance systems are also highly dependent on human operators who have to watch multiple video feeds at the same time, which in turn makes the process exhausting and susceptible to mistake. The mounting amounts of surveillance video data produced on a daily basis have necessitated the need to incorporate intelligent automation which is able to identify abnormal activities at a speed and degree of accuracy. This change has prompted massive attention to deep learning-based solutions in particular those that can detect complex patterns in real time.

Convolutional Neural Networks (CNNs) have become one of the most compelling video analysis tools in the recent years, because of the strong feature extraction and pattern recognition abilities. Autoencoders, hybrid networks, and transformer-enhanced networks are generated CNNs that have proven to have impressive performance in detecting subtle anomalies in crowded, dynamic, or noisy settings. These models can minimize the human work being done in monitoring activities through automatic highlighting of suspicious activities, threat detection and alerts' being triggered before human intervention is necessary.

There is also the increased demand of high-precise anomaly detection in virtual testing environments due to the emergence of smart cities and autonomous systems. Virtual test surveillance allows researchers and security analysts to test the reaction of the system as well as assessment of behavior and testing of threat-detection algorithms without threatening the exposure of real users. Annanomaly detection through CNN is also critical to the process of identifying abnormal behaviors in simulating, stress testing, and real-time virtual situations.

Anomaly detection is not always an easy task regardless of the fast development because of the different human behaviors, occlusions, cluttered backgrounds, change of illumination and differences in camera views. Deep learning methods need to be able to work with massive amounts of video data, generate meaningful representations, and identify normal and abnormal events when the difference between them is subtle. Current models regularly face problems of generalization, and they need stronger training data, dynamic structures, and fine-tuned architecture that can be made to work effectively in the real world.

The increasing adoption of CNN generation models in surveillance has created an avenue of creating automated intelligent security systems. These systems

are more accurate, responsive and reliable in relation to the traditional rule based approaches. Nevertheless, the ability to provide real-time performance without sacrificing the accuracy is still a research topic. With the more sophisticated surveillance environments, more sophisticated deep learning techniques, fusion models and edge-based solution will be required in order to develop really intelligent and scalable systems.

Key components of Virtual Surveillance:

a) Monitoring

Monitoring includes live monitoring and remote based access in many cases.

b) Cameras

Very famous cameras nowadays used are thermal camera , IP cameras and CCTV camera.

c) Storage

There are below storages used commonly are local storage , cloud based storage , Video Recorders.

d) Transmission

Coaxial cable is used for wired transmission of data while 4/5 G and wifi are used for transmission of data continuously.

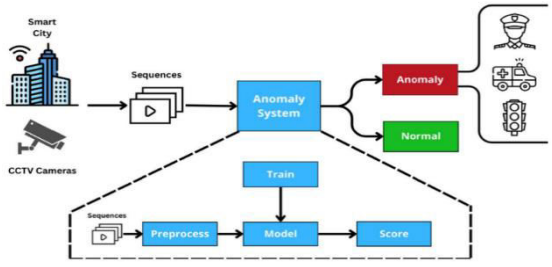

Fig. 1.1 common Structure of Smart Surveillance

The smart cities or smart places use CCTV cameras in more intelligent way in day today life by many organizations. The video is nothing but continuous sequence of frames. Anomaly system which is smart enough to identify the activity under surveillance is normal or anomaly. The anomaly system continuously gets trained on anomaly action as well as normal action. Further through real time camera continuously the video feed is taken as input and results are shown

This survey paper will examine the recent development of CNN-based anomaly detection and is the generated network as a tool of virtual test surveillance. It is a review of some of the most important techniques, their advantages and disadvantages, and any new challenges and research directions. It is aimed at giving a full picture of how deep learning is reshaping smart surveillance and facilitating the creation of more effective, precise, and flexible systems of anomaly detection.

II. LITERATURE SURVEY

The paper by Krizhevsky et al. presented one of the first and most successful Convolutional Neural Networks (CNNs) AlexNet. It was shown that deep CNNs are able to learn multi-level image features automatically and are much more effective than traditional ones on large datasets such as ImageNet. The major

contributions are ReLU activation, dropout regularization and training on GPUs, which popularized the application of deep networks to large-scale applications. The paper is the basis behind the anomaly-detection CNN models since it determined CNNs as a strong tool to create patterns and differences within the sophisticated visual data [1].

The Residual Network (ResNet) architecture suggested by He et al. addressed the problem of vanishing-gradient in deep networks by skip connections. The innovation enabled researchers to construct very deep CNNs (more than 100 layers) without any deterioration in its performance. To identify anomalies, ResNet allows more profound feature representation, which allows a model to learn fine differences between normal and abnormal test behaviors. In the industrial fault detection, virtual test inspection, and visual anomaly classification, ResNet-based models are popular because of their strength and accuracy[2].

Luo et al. suggested a CNN-based system of detecting anomalies in videos with the help of locality-sensitive convolutional autoencoders. The model acquires normal motions patterns through training videos and identifies anomalies as patterns which are not normal. They use a methodology that concentrates on frame reconstruction error to detect unusual events and this is applicable in the test anomaly detection in virtual testing like automated inspection, robotics and monitoring in industries. This study demonstrates the capability of CNNs to identify spatiotemporal anomalies across multiple frames, which is useful in detecting anomalies in real time [3].

In the given work, the authors concentrate on deep generative architecture like Variational Autoencoders (VAEs) and the GAN-based CNN architecture to detect anomalies in medical images. The model trains the distribution of normal data and indicates the samples that have a large reconstruction error as anomalies. Their article is applicable to the detection of test anomalies in virtual tests since it shows how CNN-generative systems can detect minute variations of intricate visual patterns. The generative modeling feature can be specifically applied to synthesize test samples that are used to enhance the accuracy of anomaly-detection [4].

Ruff et al. introduced a single deep learning model of outlier and anomaly detection, named Deep One-Class Classification (Deep OCC). The approach does not use labels that are supervised; rather, the decision boundaries around normal data are learned with the help of a CNN-based encoder. The method works well in industrial tests, system logs and virtual testing environments where anomalies are rare or untagged. Their study points out those one-class deep models can be used to enhance robustness because they do not rely on labeled anomalies and hence can be used in scalable virtual test anomaly identification.

The study examines multimodal variational autoencoders (MVAE), which allow combining various types of data (images, text, audio) into one anomaly-detection system. The model is able to train on missing or unfinished modalities

and hence it is very flexible in virtual testing situations where sensor data, logs and images may not be fully available. To identify anomalies, MVAEs offer a deep-learning framework that is easy to modify to integrate heterogeneous virtual test data to identify deviations with greater precision across domains.[6] In this research, there will be a complete automated surveillance system based on background subtraction, convolutional autoencoders and object detectors. A Gaussian Mixture Model is employed to extract the foreground objects after which the foregrounds are analyzed to identify abnormal events. The system draws suspicious behaviors within bounding boxes and raises real time alarms. The benchmark data experiments reveal a high performance with high AUC of 94.94%. [7].

The authors suggest a deep-learning model to detect abnormal events in large surveillance video channels. The pre-trained CNN is used to extract the spatiotemporal features and a Bi-directional LSTM model is used to classify the features. This technique saves human labor and enhances consistency of multifaceted situations. The system was more accurate on UCF-Crime datasets than some of the current methods[8].

In this paper, the video surveillance issues, including massive data, occlusions, and various human activities that complicate manual monitoring, are reviewed. It outlines the contribution of deep learning techniques, such as CNNs, transformers, YOLO, and autoencoders to automated anomaly detection. The paper makes a comparison of the state of the art methods and indicates that no one solution can support all the requirements of real-time surveillance. It offers information on design of more efficient and scalable video anomaly detection systems[9].

The paper reviews development of the anomaly detection methods in smart-city surveillance methods. It classifies various techniques, data sets and objects of interest that are involved in intelligent video surveillance. Another approach that is discussed by the authors is the edge-based anomaly detection of real-time performance. The review also covers major challenges and future opportunities, in particular, the opportunities to deploy anomaly detection on low-power edge devices[10].

III. Comparison table:

Method / Model Used	Core Techniques	Strengths	Limitations
Classical Video Anomaly Detection	Optical flow, handcrafted motion features	Low computational cost, simple to implement	Poor accuracy in complex scenes, sensitive to noise
CNN-Based Feature Extraction	Deep CNN layers, feature maps	Strong visual learning, good for object-level anomalies	Requires large labeled datasets
Autoencoder Reconstr	Encoder – decoder,	Works for unsuper	High false positives

uction Model	reconstruction error analysis	vised learning, learns normal patterns	for subtle abnormalities
Generative Adversarial Network (GAN)	GAN-based reconstruction, adversarial learning	Captures complex patterns, strong generative ability	GAN training is unstable and time-consuming
LSTM-Based Sequential Model	CNN feature extraction + LSTM temporal modeling	Good for capturing motion and sequence dynamics	Slow inference; struggles with long sequences
Transformer-Based Video Analysis	Multi-head attention, patch embeddings	Handles long-range dependencies, strong spatial–temporal modeling	Requires high compute resources
Background Subtraction + Autoencoder + Object Detectio	MoG BS, CAE, bounding-box detection	High AUC (94.94%), real-time alarms, robust feature	Less effective in heavy occlusion and low-light conditions
n		learning	
CNN + Bi-Directional LSTM	Pretrained CNN, BD-LSTM classification	High accuracy on UCF-Crime, strong temporal detection	Computationally heavy, depends on sequence length
Deep Learning Survey (CNN, YOLO, Transformers)	Review of DL architectures for surveillance	Covers state-of-the-art, highlights gaps and trade-offs	Does not propose a unified solution
Smart-City Anomaly Detection + Edge Computing	Dataset analysis, edge-device anomaly detection	Low-latency, real-time performance, suitable for IoT	Limited device memory and processing capacity

IV. Challenges faced in this study

There are multiple challenges we faced while doing survey on this topic of smart surveillance,

a) Computational cost is higher

For large video stream its very challenging to generate CNN variant such as GAN or autoencoders and apply them in real time.

b) Crowded and Complex Environment Handling

Background in the complex or more crowded videos make less accuracy of

anomaly prediction with general algorithms.

c) Different Conditions of Poort Environment

There are different conditions such as improper or more lightning, camera angles , colourful backgrounds etc may reduces the performance of general algorithms used for the anomaly detection.

d) Unavailability of Relevant data

There is need of labelled high-quality data of anomalies but such data on internet is not available for domain specific anomaly prediction.

e) False Alarm

Small changes of user activity may be predicted as anomaly which is false detection with general algorithms. So there is need of robust algorithm which reduces the false alarm.

V. Research Gap

a)One significant gap is the Edge Deployment with CNN. Whereas numerous CNN variants have demonstrated strong results in anomaly detection, little research has been conducted in order to make these models lightweight to execute effectively on edge devices. Models of real-time surveillance systems implemented on IoT cameras, drones, and embedded devices need to use low memory and low computational time. Current CNN models are largely geared towards GPU that limits their real-world usage. In order to fill this gap, researchers should pay attention to model compression, pruning, quantization, and building energy-efficient CNNs that are compatible with edge-level anomaly detection.

b) The other significant research gap is Domain Specific Working to Anomaly Detection. The majority of research concentrates solely on the surveillance settings of the masses which include streets, malls, and campuses. Nonetheless, anomaly detection is also demanded in various other fields such as testing systems in industries, virtual systems, self-driving tests, medical tests, and quality checks at the factory. General applicability of CNN-based anomaly detection systems is limited by the unavailability of domain-specific datasets and models. Cyborgering specialized datasets and customizing models to various virtual testing domains can be a great way to increase the accuracy of a model.

There is also another gap in the Fusion of Multimodal Data. The existing systems are c)primarily visual video-based, yet in real-life scenarios, deviations consist of a combination of various signals, potentially including audio notifications, sensor data, text messages, temperature measurements, or simulated data information. There is very little literature that deals with multimodal fusion which integrates these various streams of data. As virtual test environments are able to produce different sensor outputs, they can be used to produce information on anomalies significantly. To solve this gap, there is need to develop multimodal CNN structures that can handle heterogeneous data in a single framework.

d) Generalization of Cross Scene

Sometimes there is an unknown scene is present in the environment which is not

trained to the generalized model so such a scene cannot be predicted correctly using generalized models. So, there is a need of a model which is more efficient even on unseen scene.

e) Gap in Spatio-Temporal Sampling

There are many authors who worked on either spatial domain or temporal domain but we required a spatiotemporal sampling technique which can predict anomalies vary correctly.

VI. CONCLUSION

This survey paper has reviewed the developments of smart surveillance systems using CNN-based and deep learning methods in detecting virtual test anomaly. It is demonstrated in the literature that the deep learning models, including CNNs, autoencoders, transformer networks, and hybrid models, significantly enhance the ability to identify suspicious activities over traditional methods. These techniques are able to automatically detect high level features and process a large video stream, and generate a timely alert with minimum human intervention. Although most of the state-of-the-art models have demonstrated high results on benchmark data, there are still challenges related to real-world applications, such as various environments, intensive computation, and poor generalization. The future directions that can be taken include lightweight architectures, edge-based deployments, more comprehensive definitions of anomalies and stronger training data. All in all, CNN-created surveillance systems are a good move on the way to safer, smarter and more automatic security settings.

Future studies can be done to come up with the lightweight CNNs that can be used with edge devices in real-time surveillance. Systems can be generalized to real-world conditions better by using more diverse datasets and better definitions of anomalies. False alarms and enhancement of accuracy can also be achieved by integrating multimodal sensors and transformer based models.

Future Scope

Semi Automated System :

In this application ai techniques are used for anomaly detection but even the human intervention to check whether false alarm occurred or real anomaly may improve the systems more reliable performance.

Addition of Neural Activity:

Some of the neural activities can be added in dataset to make application more application oriented or real time for industries. Example : running very fast in reputed industry which is under surveillance.

REFERENCES

[1] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayedd, and R. Klette, “Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes,” Computer Vision and Image Understanding, vol. 172, pp. 88–97, 2018.

[2] X. Niu, J. Li, and J. Sun, "Dynamic Detection of False Data Injection Attack in Smart Grid using Deep Learning," arXiv preprint arXiv:1808.01094, 2018.

[3] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep Anomaly Detection for Time-series Data in Industrial IoT: A Communication-Efficient On-device Federated Learning Approach," arXiv preprint arXiv:2007.09712, 2020.

[4] A. Meliboev, J. Alikhanov, and W. Kim, "1D CNN Based Network Intrusion Detection with Normalization on Imbalanced Data," arXiv preprint arXiv:2003.00476, 2020.

[5] H. T. Oğuz and A. Kalaycıoğlu, "Anomaly detection in multi-tiered cellular networks using LSTM and 1D CNN," EURASIP Journal on Wireless Communications and Networking, vol. 2022, Article no. 101, Oct. 2022.

[6] M. Priyadarsini, "A CNN-based approach for anomaly detection in smart grid systems," Electric Power Systems Research, 2025. (via ScienceDirect)

[7] Ali, M. M. (2023). Real-time video anomaly detection for smart surveillance. *IET Image Processing*, *17*(5), 1375-1388.

[8] Ullah, W., Ullah, A., Haq, I. U., Muhammad, K., Sajjad, M., & Baik, S. W. (2021). CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia tools and applications*, *80*(11), 16979-16995.

[9] Duja, K. U., Khan, I. A., & Alsuhaibani, M. (2024). Video surveillance anomaly detection: A review on deep learning benchmarks. *IEEE Access*.

[10] Patrikar, D. R., & Parate, M. R. (2022). Anomaly detection using edge computing in video surveillance system. International Journal of Multimedia Information Retrieval, 11(2), 85-110.

[11] Sultani, W., Chen, C., & Shah, M. (2018). *Real-world anomaly detection in surveillance videos*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6479–6488.

[12] Ramachandra, B., Jones, M., & Vatsavai, R. R. (2020). *A survey of single-scene video anomaly detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 44(5), 2293–2319.

[13] Zhao, Y., Xie, Z., & Xu, Y. (2021). *Video anomaly detection via deep generative models: A review*. Pattern Recognition Letters, 140, 64–71.

[14] Hasan, M., Choi, J., Neumann, J., & Davis, L. (2016). *Learning temporal regularity in video sequences*. IEEE CVPR, pp. 733–742.

[15] Ionescu, R. T., Smeureanu, S., Alexe, B., & Popescu, M. (2019). *Detecting abnormal events in video using neural networks*. International Journal of Computer Vision (IJCV), 127(9), 1239–1261.

[16] Medel, J. R., & Savakis, A. (2016). *Anomaly detection in video using predictive convolutional long short-term memory networks*. IEEE International Conference on Image Processing (ICIP), pp. 2702–2706

[17] Nguyen, T. N., & Meunier, J. (2019). *Anomaly detection in video sequence with appearance-motion correspondence*. IEEE ICCV, pp. 1273–1283.

[18] Sabokrou, M., Kumar, A., Fayyaz, M., & Pfeiffer, M. (2018). *Deep anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes*. Pattern Recognition, 105, 107107

[19] Liu, W., Luo, W., Lian, D., & Gao, S. (2018). *Future frame prediction for anomaly detection – A new baseline*. IEEE CVPR, pp. 6536–6545.

[20] Zhang, K., Wu, Y., & Yang, J. (2022). *A hybrid CNN-GAN model for real-time video anomaly detection in intelligent surveillance systems*. Computers & Security, 113, 102541.