Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 2 (2025)

A Review on AI-Powered Surveillance System using Computer Vision for Public Safety

Sahana B[1], Dr. R Kumar[2], Madhu Prakash S[3]
Department of Mechanical Engineering, Ramaiah Institute of Technology, Bengaluru-54

Corresponding Author: Sahana B

Abstract:- This review illuminates the rising role of AI-powered surveillance systems in enhancing public safety. Urban growth and evolving threats demand intelligent, automated monitoring solutions. Computer vision enables real-time video analytics, activity recognition, and object detection. These systems autonomously identify anomalies and aid law enforcement with alerts. Frameworks like Smart Watch and CV-Patrol use deep learning for rapid decision-making. Edge computing enhances performance across large- scale surveillance networks. Bots with multi-modal sensors offer 24/7 monitoring in public and high-risk areas. A layered approach—detection, verification, and response—boosts threat management. Advances include facial/vehicle recognition and behavior prediction models. Privacy-aware data handling ensures ethical surveillance deployment. Future goals involve fair AI models, better transparency, and agent coordination. This review outlines key innovations shaping AI-driven public safety systems.

Keywords: *AI-powered surveillance, computer vision, public safety, object detection, deep learning*.

1. Introduction

The increasing intricacy of urban landscapes, coupled with the escalation of public safety concerns, has necessitated the adoption of sophisticated and autonomous surveillance solutions. Conventional monitoring systems, predominantly reliant on human oversight, are frequently constrained by limited responsiveness, scalability, and susceptibility to error. In response, the advent of artificial intelligence (AI), particularly when integrated with computer vision, has revolutionised surveillance capabilities. These intelligent systems are designed to autonomously process live video streams, detect anomalous behaviour, perform facial and object recognition, and issue real-time alerts—thereby enhancing operational efficiency and enabling rapid, evidence-based decision-making.

Computer vision empowers machines to extract and interpret complex visual information with remarkable precision, facilitating uninterrupted monitoring of critical public domains such as transport hubs, civic spaces, and event venues. The incorporation of deep learning methodologies, edge computing frameworks, and multi-sensor data fusion has further augmented the accuracy, scalability, and robustness of such systems. As a result, AI-powered surveillance not only mitigates the limitations of traditional frameworks but also contributes substantively to the realization of intelligent urban infrastructure. This study seeks to examine the technological underpinnings, practical applications, and future prospects of computer vision-enabled surveillance systems within

the realm of public safety.

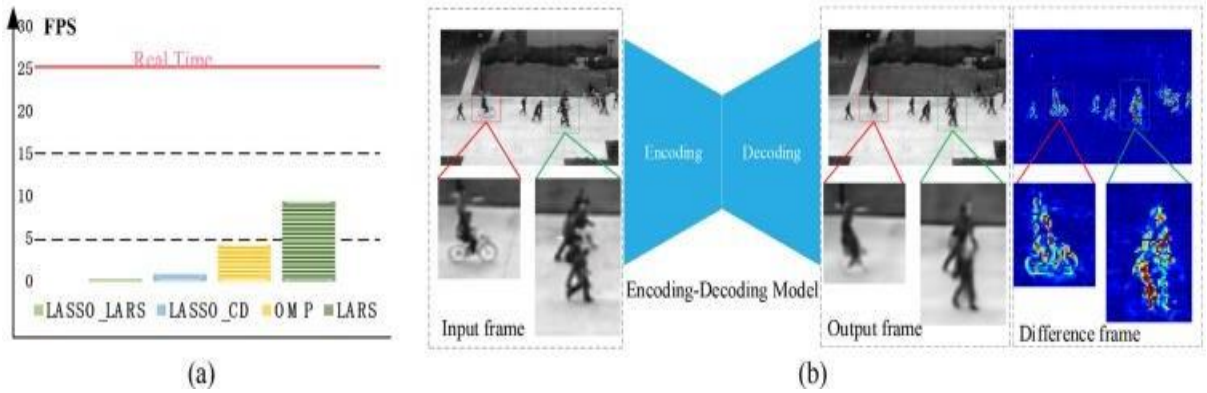

Fig1. The examples of the limitation of sparse coding based models and encoding-decoding models. (a) the frames per second (FPS) of four typical sparse coefficient solvers. (b) encoding-decoding models use frame reconstruction errors for anomaly detection. Red boxes represent the abnormal behaviors while green boxes represent the normal behaviors.

Fig. 1 illustrates the aforementioned limitations. From Fig. 1(a), it can be observed that most existing sparse coefficient solvers are computationally intensive and fail to produce optimal sparse representations in real time. As shown in Fig. 1(b), the disappearance of the bicycle and the distortion of individuals within the anomalous region in the output frame highlight a key distinction: the semantically consistent areas (denoted by green bounding boxes) between the input and output frames exhibit greater similarity compared to the anomalous regions.

Despite these technological advancements, the implementation of AI-based surveillance systems must also account for ethical, legal, and societal implications[1]. Concerns regarding mass surveillance, data privacy, algorithmic bias, and the potential misuse of facial recognition technology have sparked considerable debate among policymakers, technologists, and the general public[2]. Ensuring that these systems are designed and deployed in a transparent, accountable, and privacy-conscious manner is essential for fostering public trust and upholding civil liberties[3]. Therefore, the development of regulatory frameworks and privacy-preserving mechanisms must proceed in tandem with technological innovation[4].

Moreover, future advancements in AI surveillance are likely to be shaped by the ongoing convergence of technologies such as the Internet of Things (IoT), 5G connectivity, and federated learning[5]. These innovations are poised to enhance system responsiveness, interconnectivity, and real-time data processing capabilities across distributed surveillance networks. Additionally, the integration of behaviour prediction models, crowd analytics, and multi-agent coordination will further refine the effectiveness of these systems in complex environments [6]. This paper aims to provide a comprehensive overview of the current landscape of AI-powered surveillance using computer vision, while identifying key challenges and opportunities for future research and development[7].

Together, these technologies facilitate adaptive, real-time decision-making within increasingly complex urban security landscapes[8]. The confluence of computer vision and artificial intelligence markedly augments situational awareness, enabling surveillance systems to autonomously detect, interpret, and respond to potential threats with notable precision. This technological synergy proves particularly indispensable in densely populated or high-risk public settings, where the prompt identification of anomalous behaviour or hazardous incidents can substantially enhance operational efficacy while simultaneously minimising reliance on human intervention[9].

Although a considerable body of research has emerged in the domain of intelligent surveillance, much of it remains narrowly confined to discrete functionalities such as facial recognition, object tracking, or crowd analytics[10]. Such compartmentalised approaches frequently neglect the holistic integration and interoperability that are

imperative for scalable, context-sensitive public safety infrastructures. This paper seeks to redress this lacuna by adopting a comprehensive, system- level perspective on the architecture, functional capacities, and ethical challenges associated with AI-driven surveillance systems underpinned by computer vision[11]. The principal contributions of this study are delineated below:

2. Overview of Algorithms

2.1 Object Detection Algorithms

- **YOLO (You Only Look Once)**: YOLO is a real-time object detection system that is highly effective for detecting multiple objects in images or videos[12]. It divides the image into a grid and predicts bounding boxes and class probabilities directly, making it ideal for monitoring public spaces for unusual behavior[13].
- **SSD (Single Shot Multibox Detector)**: SSD is another real-time object detection algorithm that can classify objects in images at high speed with good accuracy. It's particularly useful in surveillance to identify and track people, vehicles, and other entities of interest[14].
- **Faster R-CNN**: An extension of the traditional CNN, Faster R-CNN uses a region proposal network (RPN) for generating region proposals, which makes it suitable for detecting objects in real-time with high accuracy[15].

2.2 Facial Recognition Algorithms

- **Haar Cascades**: An early but efficient object detection method, especially for facial recognition[16]. It's widely used in real-time systems for identifying individuals in security footage.
- **Deep Face Recognition (DeepFace)**: DeepFace uses deep learning techniques for highly accurate facial recognition and is capable of identifying people across different lighting conditions and facial angles[17].
- **FaceNet**: FaceNet is a deep learning algorithm that provides highly accurate facial recognition through embeddings, transforming images into vectors that can be used to compare face similarity[18].

2.3 Human Activity Recognition (HAR) Algorithms

- **LSTM (Long Short-Term Memory Networks)**: LSTM is a type of recurrent neural network (RNN) that is useful for analyzing sequential data, such as human activities over time[20]. It is particularly effective for recognizing abnormal human behavior in surveillance footage[21].
- **3D Convolutional Networks (3D CNNs)**: These networks capture both spatial and temporal features, making them well- suited for activity recognition in video sequences[22]. They are useful for detecting suspicious actions or patterns like running, fighting, or loitering[23].

2.4 Anomaly Detection Algorithms

- **Isolation Forest**: This algorithm is widely used for detecting outliers in data. In surveillance, it can be used to identify unusual movements or behaviors in crowds or public spaces[24].
- **Autoencoders**: These unsupervised neural networks can learn a compressed representation of data[25]. When applied to video surveillance, autoencoders can detect anomalies by measuring reconstruction error, where high error indicates unusual events or objects[26].
- **One-Class SVM (Support Vector Machine)**: One-Class SVM is a classification algorithm that learns the normal patterns in the data and classifies any deviation from this as an anomaly. It's useful for identifying uncommon events in surveillance footage[27].

2.5 Tracking Algorithms

- **Kalman Filter**: A popular algorithm for object tracking in real-time, especially for continuous tracking of moving objects such as people or vehicles[28]. It estimates the state of a moving object based on noisy observations, ensuring smooth tracking in complex environments.

- **SORT (Simple Online and Realtime Tracking)**: SORT is a simple yet effective algorithm for tracking objects in video, particularly when you need to track multiple objects simultaneously in a crowded environment[29].
- **DeepSORT**: An extension of SORT, DeepSORT incorporates deep learning features for more robust tracking, especially in dense or occluded environments[30].

2.6 Scene Understanding Algorithms

- **Semantic Segmentation**: Algorithms like **U-Net** or **DeepLab** can segment different objects or regions in a scene, helping the system understand the context (e.g., distinguishing between sidewalks, roads, and buildings) and making it easier to spot unusual behavior in specific areas[31].
- **Detectron2**: A Facebook-developed object detection framework that includes capabilities for instance segmentation, keypoint detection, and panoptic segmentation, which could be useful for segmenting specific regions in public safety surveillance[32].

2.7 Behavior Prediction and Classification

- **Recurrent Neural Networks (RNNs)**: RNNs are useful for time-series data, where understanding the sequence of past behavior can help predict future actions, such as identifying if a person is likely to engage in suspicious activity[33].
- **Transformer Networks**: These models are gaining traction for sequence prediction tasks, as they can handle long-range dependencies in data, making them useful for predicting the next move or action of individuals in a surveillance system.

2.8 Edge Computing & Real-Time Processing

- **TensorFlow Lite** or **OpenVINO**: These frameworks can optimize AI models for edge devices, allowing real-time inference on cameras and drones for low-latency, on- site decision-making in surveillance scenarios.

3. Prevailing Security Technologies

Contemporary public safety infrastructure predominantly depends on closed-circuit television (CCTV) systems deployed in high-risk zones and densely populated public spaces. These surveillance frameworks generally necessitate manual supervision, wherein security personnel continuously monitor live camera feeds to identify anomalous behaviour or emergent incidents. While certain advanced configurations integrate rule-based video analytics—featuring pre-programmed criteria such as motion detection, virtual tripwires, facial recognition, or unattended object detection—these mechanisms are inherently rigid. They exhibit limited adaptability, are prone to false alarms, and fail to evolve or enhance performance over time.

Moreover, such systems typically lack the integration of deep learning methodologies or real- time behavioural inference, thereby rendering them ineffective at discerning nuanced threats such as violent altercations, larceny, or mass panic scenarios[34]. This technological shortcoming often results in protracted response intervals, heightened reliance on human oversight, and restricted scalability in expansive urban environments. Consequently, there exists an imperative for the assimilation of intelligent surveillance infrastructures predicated upon artificial intelligence and deep neural networks.

These next-generation systems are endowed with the capability to autonomously construe intricate visual stimuli, discern anomalous behavioural patterns in real time, and adaptively recalibrate in response to shifting threat vectors. By harnessing the potential of advanced computer vision algorithms and contextual behavioural analytics, such frameworks may substantially mitigate the prevalence of spurious alerts while markedly enhancing situational cognisance. Moreover, the institution of scalable, data-centric surveillance

architectures stands to obviate excessive reliance on manual observation and engender a more anticipatory, pre-emptive approach to security management within complex metropolitan environments.

Table1: Comparison of Different Algorithms

Ref	Year	Title	Algorithm	Advantages	Drawbacks
[35]	2010	A Survey on Transfer Learning	Inductive Transfer Learning, Transductive Transfer Learning, Unsupervised Transfer Learning	Facilitates knowledge transfer across domains, reducing reliance on extensive labelled datasets.	Risk of negative transfer when source and target domains are insufficiently related.
[36]	2021	Action CLIP: A New Paradigm for Video Action Recognition	Action CLIP	Enables zero-shot action recognition by leveraging a multimodal framework that integrates video and text data.	Performance may decline when applied to domains with limited or non-representative textual annotations.
[37]	2016	Learning Temporal Regularity in Video Sequences	Fully Convolutional Autoencoder	Facilitates unsupervised learning of temporal patterns in video sequences, enhancing Anomaly detection capabilities.	May produce false positives when encountering a typical yet benign activities due to reliance on reconstruction errors.
[38]	2022	Anomaly Detection in Video Sequence with Appearance-Motion Correspondence	Joint Convolutional Autoencoder and U-Net	Integrates spatial appearance and temporal motion information to enhance the detection of anomalies in video sequences.	May encounter challenges in distinguishing atypical yet benign activities due to reliance on learned appearance-motion correspondences.
[39]	2024	Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes	Fully Convolutional Neural Network (FCN) with Cascaded Outlier Detection	Achieves real-time anomaly detection and localisation in crowded scenes by integrating spatial and temporal features.	Performance may decline in scenarios with high visual complexity or subtle anomalies due to reliance on reconstruction errors.
[40]	2023	Video Anomaly Detection Based on Attention Mechanism	Attention- Enhanced Autoencoder with Dynamic Prototype Units	Effectively captures contextual features by integrating attention mechanisms, thereby improving anomaly detection accuracy in complex scenes.	Performance may decline in scenarios with high visual complexity or subtle anomalies due to reliance on reconstruction errors.

[41]	2020	Multivariate Time-series Anomaly Detection via Graph Attention Network	MTAD-GAT (Multivariate Time-series Anomaly Detection via Graph Attention Network)	Dynamically captures intricate temporal and feature-wise dependencies through dual graph attention layers, enhancing Anomaly detection accuracy.	The model's complexity may hinder real-time deployment in resource-constrained environments due to increased computational demands.

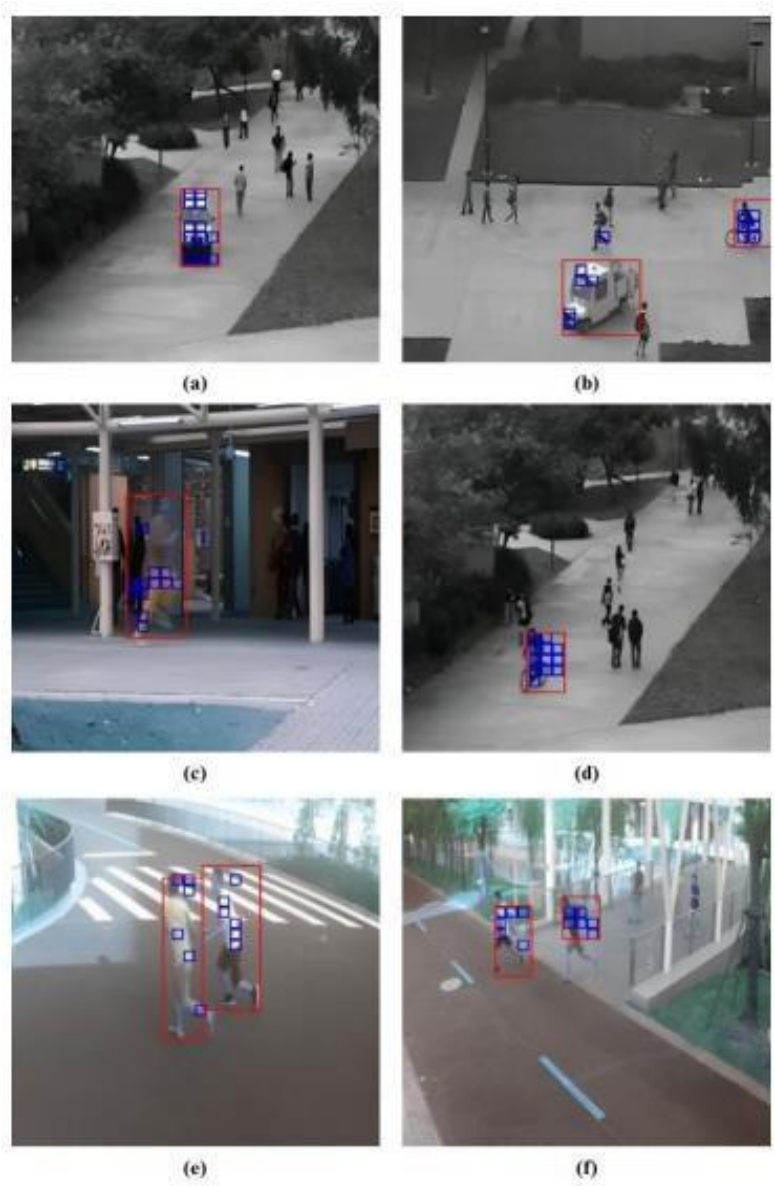

Fig 2: Visualization of video output showcasing anomaly detection in real-time.

Arpit Bajgot[42] reviews le nature in real-world scenarios. In this paper, we propose Swin Anomaly, a video anomaly detection approach based on a conditional GAN-based autoencoder with feature extractors based on Swin Transformers. Our approach encodes spatiotemporal features from a sequence of video frames using a 3D encoder and upsamples them to predict a future frame using a 2D decoder. We utilize patch-wise mean squared error and Simple Online and Real-time Tracking (SORT) for real-time anomaly detection and tracking.

4. Architecture

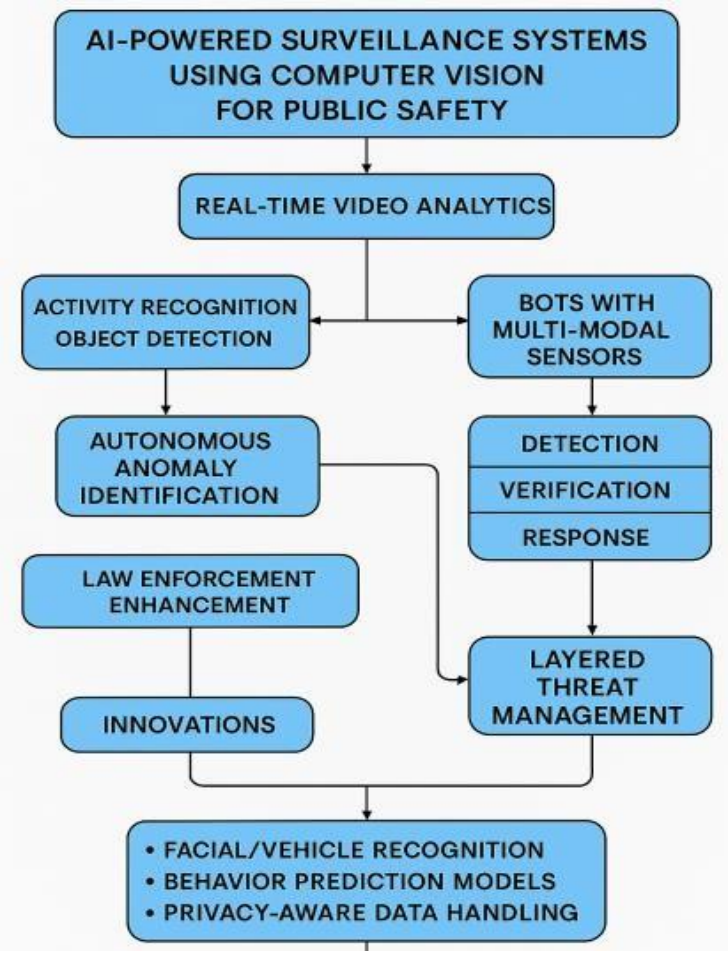

Fig 3: Architecture of AI-Powered Surveillance System Using Computer Vision for public safety.

The diagram presents a conceptual framework for AI- powered surveillance systems that utilise computer vision to bolster public safety. At the core of the system lies real-time video analytics, which facilitates the continuous assessment of visual data streams[43]. This central mechanism branches into two primary domains: activity recognition and object detection, as well as bots integrated with multi-modal sensors. These components collectively enhance the system's capability to discern human actions, identify objects of interest, and gather a multitude of environmental inputs simultaneously, thereby establishing a robust foundation for intelligent surveillance[44].

Subsequent to initial detection, the system employs autonomous anomaly identification to pinpoint irregular or potentially threatening behaviour without necessitating human intervention[45]. This capacity significantly augments law enforcement operations, enabling a more timely and effective response to incidents[46]. The bots further employ a structured approach encompassing detection, verification, and response, ensuring that threats are corroborated prior to engagement. This methodical progression contributes to a layered threat management strategy, which mitigates false positives and ensures that each level of security response is proportionate and evidence-based[47].

The culmination of these technological advancements is reflected in their practical applications, which include facial and vehicle recognition, the deployment of behaviour prediction models, and privacy-aware data handling[48]. These innovations enhance the system's predictive capabilities while maintaining ethical standards concerning data privacy and surveillance. By synthesising machine intelligence with structured threat

management, the system offers a comprehensive, anticipatory, and ethically conscious approach to modern public safety challenges[49] [50].

5. Applications

1. **Crowd Surveillance and Behavioural Monitoring:** AI-driven systems enable real-time observation of densely populated areas, facilitating the identification of abnormal crowd dynamics, potential unrest, or hazardous behaviour, thereby allowing authorities to intervene pre-emptively.

2. **Criminal Activity Detection and Deterrence:** Through advanced facial and vehicle recognition technologies, these systems assist in the identification of suspects, tracking of illicit movements, and automated anomaly detection, enhancing the efficacy of law enforcement operations.

3. **Traffic Management and Road Safety:** Computer vision algorithms analyse vehicular patterns, detect violations such as speeding or signal breaches, and provide instantaneous alerts in the event of accidents, contributing to smoother traffic regulation and reduced fatalities.

4. **Protection of Critical Infrastructure:** By integrating multi-modal sensor networks, AI surveillance ensures the continuous safeguarding of vital facilities such as airports, energy plants, and government buildings, detecting unauthorised access or security breaches with high precision.

5. **Disaster Response and Emergency Coordination:** During natural calamities or emergencies, AI-enabled systems aid in real-time situational analysis, victim localisation, and the strategic deployment of rescue teams, thereby augmenting the overall effectiveness of emergency response mechanisms.

6. Conclusion

The deployment of artificial intelligence in conjunction with computer vision within surveillance infrastructures epitomises a profound evolution in contemporary public safety mechanisms. These intelligent systems afford unparalleled capabilities in real-time situational awareness, autonomous anomaly recognition, and prognostic threat evaluation, thereby augmenting the efficacy of both preventive and reactive security measures. Their applicability extends across diverse domains, including urban crowd governance, vehicular regulation, infrastructural protection, and disaster mitigation. Crucially, the incorporation of privacy-conscious data governance frameworks ensures a harmonious equilibrium between technological vigilance and individual civil liberties. As these innovations continue to mature, AI-powered surveillance paradigms are poised to serve as pivotal instruments in the cultivation of secure, adaptive, and ethically accountable civic environments.

References

[1] ESMA DILEKAND MURAT DENER, Enhancement of Video Anomaly Detection Performance Using Transfer Learning and Fine- Tuning, 2024.

[2] Waqas Sultani1, Real-world Anomaly Detection in Surveillance Videos, 2019.

[3] Chongke Wu, Video Anomaly Detection Using Pre-Trained Deep Convolutional Neural Nets and Context Mining, 2017.

[4] Waseem Ullaha, TransCNN: Hybrid CNN and transformer mechanism for surveillance anomaly detection, 2023.

[5] Shalmiya Paulraj, Transformer-enabled weakly supervised abnormal event detection in intelligent video surveillance systems, 2025.

[6] Areej Alasiry, A Region based Salient Stacking Optimized Detector (ReSOD) For an Effective Anomaly Detection and Video Tracking in Surveillance Systems, 2024.

[7] Kai Chenga, Normality learning reinforcement for anomaly detection in surveillance videos, 2024.

[8] Javaria Amin, Detection of anomaly in surveillance videos using quantum convolutional neural networks, 2023.

[9] Pushpajit Khaire, A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments, 2022.

[10] Aniruddha Prakash Kshirsagar, YOLOv3-based human detection and heuristically modified-LSTM for abnormal human activities detection in ATM machine, 2023.

[11] Wai-Kong Lee, ArchCam: Real time expert system for suspicious behaviour detection in ATM site, 2018.

[12] Romany F. Mansour, Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model, 2021.

[13] B.S. Murugan, Region-based scalable smart system for anomaly detection in pedestrian walkways, 2013.

[14] A.A. Afiq, A review on classifying abnormal behavior in crowd scene, 2013.

[15] Zheng-ping Hu, Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes, 2020.

[16] Dongyue Chen, Anomaly detection in surveillance video based on bidirectional prediction, 2020.

[17] Xinfeng Zhang, Video anomaly detection and localization using motion-field shape description and homogeneity testing, 2020.

[18] Renzhi Wu, Improving video anomaly detection performance by mining useful data from unseen video frames, 2021.

[19] Peng Wu, Fast sparse coding networks for anomaly detection in videos, 2020.

[20] Nanjun Li, Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder, 2019.

[21] Yao Tang, Integrating prediction and reconstruction for anomaly detection, 2020.

[22] Dan Xu, Detecting anomalous events in videos by learning deep representations of appearance and motion, 2017.

[23] Mohammad Sabokrou, Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes, 2018.

[24] K Nandhini, ANAMOLY DETECTION FOR SAFETY MONITORING, 2017.

[25] KASHAF U. DUJA, Video Surveillance Anomaly Detection: A Review on Deep Learning Benchmarks, 2024.

[26] Ismael Serrano, Fight Recognition in Video Using Hough Forests and 2D Convolutional Neural Network, 2018.

[27] YUDAI WATANABE, Real-World Video Anomaly Detection by Extracting Salient Features in Videos, 2022.

[28] Xianlin Zeng, A Hierarchical Spatio-Temporal Graph Convolutional Neural Network for Anomaly Detection in Videos, 2023.

[29] Chao Huang, Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System, 2022.

[30] PASIKANTI ROHITH, Using CCTV Footage for Anamoly Detection, 2024.

[31] Waqas Sultani, Real-world Anomaly Detection in Surveillance Videos, 2022

[32] Radu Tudor Ionesc, Detecting Abnormal Events in Video Using Narrowed Normality Clusters, 2019.

[33] Zan Gao, Temporal Action Localization with Multi-temporal Scales, 2022.

[34] Du Tran, Learning Spatiotemporal Features with 3D Convolutional Networks, 2015.

[35] Sinno Jialin Pan, A Survey on Transfer Learning, 2009.

[36] Mengmeng Wang, ActionCLIP: A New Paradigm for Video Action Recognition, 2019.

[37] Mahmudul Hasan, Learning Temporal Regularity in Video Sequences, 2017.

[38] Trong-Nguyen Nguyen, Anomaly Detection in Video Sequence with Appearance-Motion Correspondence, 2017.

[39] M. Sabokrou, Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes, 2017.

[40] Qianqian Zhang, Video Anomaly Detection Based on Attention Mechanism, 2023.

[41] Hang Zhao, Multivariate Time-series Anomaly Detection via Graph Attention Network, 2020.

[42] Debo Cheng, Disentangled Representation Learning for Causal Inference with Instruments, 2024.

[43] Bharathkumar Ramachandra, Street Scene: A new dataset and evaluation protocol for video anomaly detection, 2020.

[44] Dan Xu, Learning Deep Representations of Appearance and Motion for Anomalous Event Detection, 2015.

[45] Peng Wu, VadCLIP: Adapting Vision-Language Models for Weakly Supervised Video Anomaly Detection, 2023.

[46] Oliver Rippel, Transfer Learning Gaussian Anomaly Detection by Fine-tuning Representations, 2022.

[47] Mariana-Iuliana Georgescu, Anomaly Detection in Video via Self-Supervised and Multi-Task Learning, 2021.

[48] Guangyu Sun, Anomaly Crossing: New Horizons for Video Anomaly Detection as Cross-domain Few- shot Learning, 2022.

[49] Zhewen Deng, Prior Knowledge Guided Network for Video Anomaly Detection, 2023.

[50] ARPIT BAJGOTI, SwinAnomaly: Real-Time Video Anomaly Detection Using Video Swin Transformer and SORT, 2023.