International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

Real-Time Crowd Anomaly Detection Using Spatio-Temporal ConvLSTM Autoencoders

Mrs Umme Hani Sara [1], Akula Keshav[2], Damam Nehan Kumar[2],Vasala Manindra[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)
[1,2]J.B. Institute of Engineering and Technology

*Corresponding author: Mrs Umme Hani Sara(ummehanisara@gmail.com)

ABSTRACT

Crowd anomalies in public spaces and high-traffic environments pose significant challenges to security and public safety. Traditional manual surveillance is often prone to human error, cognitive fatigue, and delayed response times, necessitating automated, intelligent monitoring solutions. This project presents CrowdWatch AI, a real-time, end-to-end crowd anomaly detection system that leverages Spatiotemporal Convolutional Long Short-Term Memory (ConvLSTM) networks to automate surveillance monitoring and incident detection.

The system utilizes a decoupled architecture, integrating a high-performance FastAPI backend for low-latency inference with a high-fidelity React-based dashboard for real-time visualization and operator interaction. By employing a deep learning-based spatiotemporal autoencoder, the system learns the latent representation of "normal" crowd movement dynamics from high-density video sequences. Abnormalities are identified by calculating the Mean Squared Error (MSE) of the reconstruction in video sequences; a high reconstruction error indicates a deviation from learned normal patterns, signaling a potential anomaly such as illegal vehicle entry, sudden crowd dispersion, or counter-flow movement.

A sliding window mechanism processes temporal frame sequences (e.g., 10-frame intervals) to capture both spatial features and temporal dependencies, ensuring high detection sensitivity and reducing false positives. When the calculated anomaly score exceeds a statistically defined or operator-guided threshold, the system triggers a multi-layer alert mechanism. This includes instant incident logging in a SQLite database, automated JPEG snapshot capture of the triggering frames for forensic review, and native haptic feedback (vibration and visual cues) for the operator to ensure immediate situational awareness.

The proposed solution is designed for scalability and versatility, supporting a wide range of input sources including local webcams, RTSP/HTTP IP cameras, and asynchronous video uploads. Operating efficiently on standard hardware without the absolute requirement for high-end dedicated workstation GPUs, the system delivers low-latency MJPEG streaming and real-time telemetry. CrowdWatch AI provides a non-intrusive, cost-effective, and proactive approach to intelligent surveillance, contributing to the development of safer smart city environments and efficient incident prevention strategies.

Keywords- Crowd Anomaly Detection, Spatiotemporal ConvLSTM, Deep Learning, Computer Vision, Real-Time Surveillance, FastAPI, React Dashboard, Reconstruction Error (MSE), Automated Incident Logging, Intelligent Security Systems.

1.INTRODUCTION

The rapid urbanization of modern societies and the proliferation of high-definition surveillance infrastructure have significantly increased the volume of visual data generated in public spaces. It is estimated that millions of closed-circuit television (CCTV) cameras are deployed worldwide in transport hubs, shopping malls, and stadiums to ensure public safety. However, the sheer scale of this data presents a critical bottleneck: traditional manual monitoring by human operators is increasingly inefficient and prone to fatigue-related errors. Statistics suggest that after only twenty minutes of continuous monitoring, a human operator's ability to detect significant events drops by over 90% [1]. This "surveillance gap" leaves urban environments vulnerable to undetected crowd abnormalities, such as sudden stampedes, illegal gatherings, or violent outbreaks, which can lead to catastrophic loss of life and property.

Consequently, there is an urgent need for intelligent, automated systems capable of identifying "anomalous" behavior—defined as patterns that deviate significantly from a learned "normal" state—without constant human intervention. In the context of computer vision, crowd anomaly detection is a complex task due to the high density of individuals, occlusions, and the dynamic nature of human movement. Traditional frame-based analysis often fails to capture the temporal dependencies—the "flow" of motion—that distinguish a normal walk from a chaotic run or a sudden dispersal.

To address these challenges, this project presents CrowdWatch AI, a real-time surveillance framework designed to detect crowd abnormalities using spatiotemporal deep learning. The system is built upon a Convolutional Long Short-Term Memory (ConvLSTM) autoencoder architecture. Unlike standard Convolutional Neural Networks (CNNs) that process static images, the ConvLSTM approach integrates both spatial feature extraction and temporal sequence modeling. This allows the system to learn the "visual rhythm" of a normal scene and identify irregularities based on reconstruction error. Specifically, when the model is presented with an unusual event (e.g., a vehicle entering a pedestrian zone or a sudden crowd surge), it fails to reconstruct the sequence accurately, resulting in a high Mean Squared Error (MSE) that triggers an immediate alert.

The proposed system adopts a decoupled, full-stack architecture to ensure operational reliability on standard hardware. A FastAPI backend manages the intensive machine learning inference and multi-threaded video ingestion from various sources, including local webcams and networked IP cameras. Simultaneously, a responsive React dashboard provides security personnel

with real-time telemetry, sliding-window analytics, and automated incident documentation. By automating the detection of spatial and temporal irregularities, CrowdWatch AI offers a non-intrusive, cost-effective solution for enhancing public safety in increasingly crowded urban landscapes [5]

2. LITERATURE SURVEY

This section reviews existing research in the field of crowd anomaly detection, categorized into traditional handcrafted methods and modern deep learning-based approaches. This review highlights the technical evolution that leads to the rationale behind our proposed ConvLSTM-based framework

2.1. Traditional Handcrafted Approaches

Early research in anomaly detection focused heavily on extracting manual features to describe motion and texture. Ojha et al. [17] proposed SIFT-based methods to track object vectors across frames, using distance measurements to infer speed-related anomalies. Similarly, Bay et al. [19] introduced SURF as a more computationally efficient alternative to

SIFT, utilizing integral images to achieve speed while maintaining robustness to scaling and rotation.

While effective in controlled environments, these methods often rely on Optical Flow or Bag-of-Words (BoW) models [22] to represent video dynamics. For instance, Hassner et al. [23] utilized Violent Flow (ViF) descriptors combined with Support Vector Machines (SVM) to classify movement magnitudes. However, traditional handcrafted features suffer from significant limitations:

Sensitivity to Environment: Features like SIFT and SURF are highly sensitive to lighting changes, camera angles, and crowd density.

Lack of Temporal Depth: Most traditional methods treat spatial and temporal data as separate entities, failing to capture the complex, long-term spatiotemporal dependencies required to identify subtle anomalies.

Parameter Intensity: These models often require extensive manual tuning and task-specific heuristics, making them difficult to generalize across diverse surveillance scenarios.

2.2. Deep Learning-Based Approaches

The advent of Deep Learning has revolutionized crowd analysis by allowing models to learn hierarchical representations automatically. Initial deep learning attempts utilized two-stream networks, where one CNN processed spatial appearances while another processed temporal motion (optical flow).More recently, hybrid architectures like CNN-LSTMs have gained prominence. Pawar et al. [25] demonstrated an unsupervised one-class classification approach that learns "normal" features from benchmark datasets like UCSD and Avenue. Similarly, Mohan et al. [26] proposed combining CNNs with Principal Component Analysis (PCA) for better classification of abnormal activities. While these models improve upon traditional methods, they often face "temporal jitter" or lose spatial resolution when passing features between the CNN and LSTM layers.

2.3. Rationale for the Proposed ConvLSTM Autoencoder

To address the disconnect between spatial and temporal processing, our project leverages the Convolutional Long Short-Term Memory (ConvLSTM) architecture. Unlike standard LSTMs that use 1D vector inputs—thereby losing spatial topology—ConvLSTM incorporates convolution operations directly within the recursive transitions. This allows the model to:

1.Maintain Spatial Integrity: It preserves the 2D structure of
video frames while modeling temporal changes.

2.Unsupervised Anomaly Discovery : By training as an Autoencoder, the system learns to reconstruct "normal" scenes Anomalies are then detected as high Reconstruction Error (MSE) on unseen, deviant data.

3. Efficiency : By reducing the input of dimensions to 64x64 grayscale sequences, our implementation ensures that sophisticated spatiotemporal modeling can be performed in real-time on standard hardware, bridging the gap between high-accuracy research and practical deployment.This transition from handcrafted descriptors to an integrated ConvLSTM autoencoder forms the technical foundation of CrowdWatch AI, ensuring both robustness to environmental noise and sensitivity to complex crowd dynamics.

2.4.Limitations and Research Gaps

2.4.1. Context-Dependency of Anomalies

A major research gap in crowd anomaly detection is the contextual definition of an "anomaly." For instance, a person running may be considered normal in a park but anomalous in a crowded airport terminal. Most current models, including ConvLSTM autoencoders, learn a global "normal" representation based on the training dataset. However, they often struggle to adapt to changing environments or shifting baseline behaviors without retraining. There is a clear gap in developing adaptive thresholding mechanisms that can adjust sensitivity based on the time of day, location, or specific event context.

2.4.2. Scarcity of Diverse Training Data

Most state-of-the-art models are validated on a limited number of benchmark datasets, such as UCSDped1, Avenue, or UMN. While these provide a baseline, they often lack environmental diversity—specifically variations in weather (rain, fog), extreme lighting conditions (glare, low-light), and camera perspectives (top-down vs. eye-level). The research gap lies in creating synthetic data augmentation or domain adaptation techniques that allow a model trained on one city's surveillance feed to operate reliably in another without significant performance degradation.

2.4.3. Real-Time Computational Overhead vs. Accuracy

While deep spatiotemporal models like 3D-CNNs and complex ConvLSTMs offer high accuracy, they are notoriously computationally expensive. In our implementation, we mitigate this by resizing inputs to 64x64 grayscale frames to maintain real-time performance on standard CPU/GPU hardware. However, this preprocessing step introduces a limitation: information loss. Small-scale anomalies (e.g., a concealed weapon or a subtle suspicious gesture) may be "blurred out" during downsampling. A significant research gap exists in developing multi-scale architectures that can process low-resolution global motion and high-resolution local patches simultaneously without sacrificing frame rates.

3. PROPOSED SYSTEM

The proposed system, CrowdWatch AI, is an end-to-end spatiotemporal analytics framework designed to process real-time video streams and automatically detect anomalous crowd behavior using deep learning techniques. The architecture integrates high-performance model inference, efficient video processing, and an interactive web-based monitoring dashboard within a modular backend–frontend ecosystem built on FastAPI and React.

To ensure scalability, modularity, and low-latency performance, the system is structured into four primary layers:
(i) Video Ingestion and Preprocessing Layer,
(ii) Spatiotemporal Inference Engine,
(iii) Data Persistence and Analytics Layer, and
(iv) Interactive Web Interface Layer.

3.1 Video Ingestion and Preprocessing Layer

The Video Ingestion Layer is responsible for acquiring raw visual data from multiple heterogeneous sources, including local webcams, prerecorded video files (.mp4), and network-based RTSP/HTTP IP cameras. Frame extraction is implemented using OpenCV with a multi-threaded streaming mechanism to ensure non-blocking and continuous data flow.

Each captured frame undergoes a structured preprocessing pipeline to optimize it for deep learning inference. Initially, the frame is converted into grayscale to reduce computational complexity and eliminate redundant color information, thereby emphasizing motion patterns. Subsequently, the frame is resized to a fixed spatial resolution of 64×64 pixels and normalized to a floating-point range of [0,1].

A key component of this layer is temporal buffering, where a sliding window of 10 consecutive frames is maintained. This buffer forms a spatiotemporal sequence, enabling the downstream model to capture both spatial features and temporal dynamics of crowd motion effectively.

3.2 Spatiotemporal Inference Engine

The Spatiotemporal Inference Engine constitutes the core intelligence of the system and is built upon a ConvLSTM Autoencoder architecture. The model operates in an unsupervised learning paradigm, where it is trained exclusively on normal crowd behavior patterns.

During inference, the engine receives a sequence of 10 preprocessed frames and generates a reconstructed output sequence. A pixel-wise comparison between the original and reconstructed sequences is then performed using Mean Squared Error as the reconstruction metric.

The computed error is scaled to produce a normalized anomaly score. If this score exceeds a predefined threshold (empirically set to 0.1), the system classifies the event as anomalous. To ensure real-time performance, the model is deployed as a thread-safe singleton instance, eliminating repeated loading overhead and enabling continuous inference.

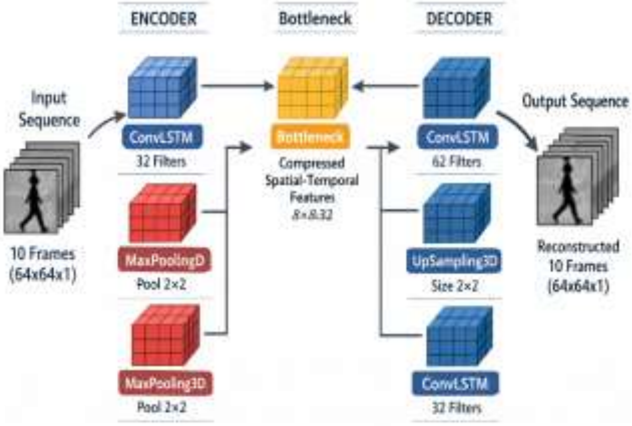


Fig 3.2.1 ConvLSTM Autoencoder Model

3.3 Data Persistence and Analytics Layer

The Data Persistence Layer is responsible for managing anomaly records and maintaining system telemetry. Upon detection of an anomaly, two parallel operations are triggered: The corresponding video frame is captured and stored as a high-resolution image in a local directory, indexed using timestamp-based convention.

Relevant metadata—including timestamp, anomaly score, and snapshot reference—is stored in a relational database implemented using SQLite with ORM support for efficient data handling.

This layer provides a historical repository for anomaly events, enabling retrospective analysis and long-term behavioral insights. Additionally, RESTful API endpoints are exposed via the backend to facilitate secure retrieval of logs

3.4 Interactive Web Interface Layer

The Interactive Web Interface Layer provides a real-time monitoring dashboard designed for operational usability and situational awareness. The frontend is developed using React with a modern UI framework for responsive and visually intuitive interaction.

The dashboard consists of three primary modules:

1. Live Video Streaming Module Displays real-time video using an MJPEG stream delivered via multipart HTTP responses. The stream includes dynamically overlaid annotations such as bounding boxes and anomaly scores.

2.Real-Time Analytics Dashboard: Implements a sliding window visualization of anomaly scores using dynamic charting, allowing operators to observe temporal fluctuations and identify emerging abnormal patterns.

3.Alert and Notification System: Provides immediate feedback through visual animations and browser-based haptic alerts (via vibration APIs). This ensures that critical anomalies are promptly highlighted, improving operator response time.

The system starts by capturing video from sources like a webcam, IP camera, or file. The video is streamed using OpenCV and then preprocessed (resized, converted to grayscale, and normalized). This processed data is fed into a ConvLSTM Autoencoder model, which analyzes temporal patterns in the video.

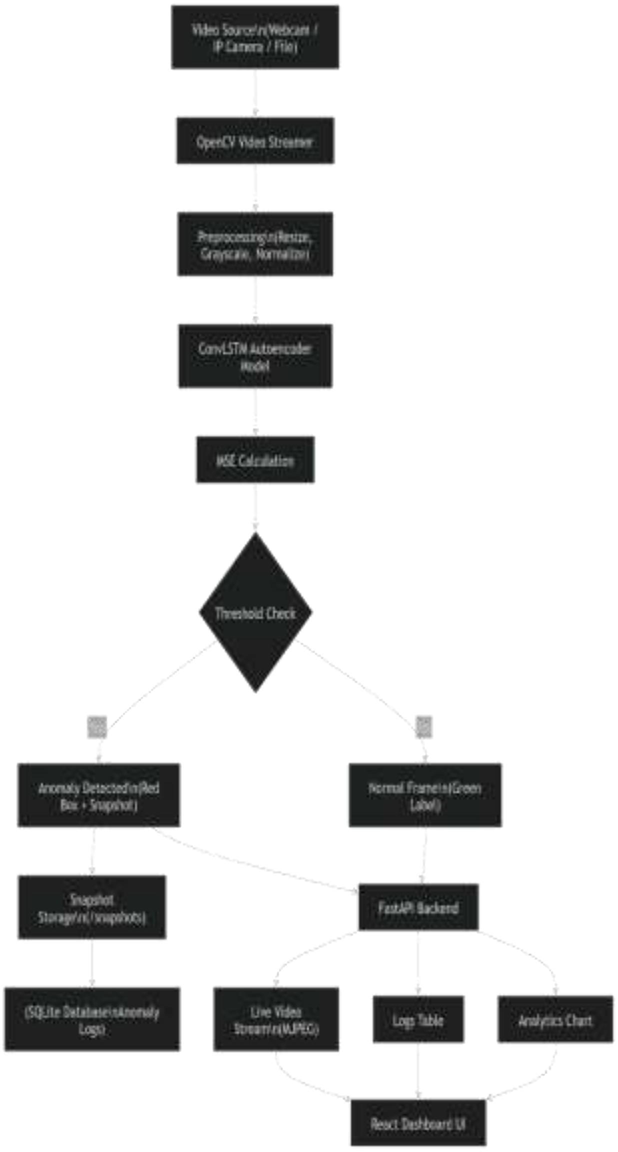


Fig 3. Proposed System

Next, the system computes the reconstruction error using MSE (Mean Squared Error) and performs a threshold check. If the error exceeds the threshold, it is classified as an anomaly, triggering a red bounding box and snapshot storage, with logs saved in a SQLite database. If not, the frame is marked as normal with a green label.

All results are sent to the FastAPI backend, which provides live video streaming, logs, and analytics through APIs. Finally, these outputs are visualized in a React dashboard UI for real-time monitoring and analysis.

4.RESULT DESCRIPTION

The performance of the proposed CrowdWatch AI anomaly detection system is evaluated based on key metrics such as Mean Squared Error (MSE) reconstruction quality, anomaly detection sensitivity, and real-time inference speed measured in frames per second (FPS). The evaluation is conducted on the UCSD Pedestrian Dataset (Ped1 & Ped2), which contains diverse video sequences featuring complex crowd dynamics, various lighting conditions, and specific "anomalous" events (e.g., bikers on sidewalks, motor vehicle entry, and sudden crowd scattering). The results demonstrate the effectiveness of the Spatiotemporal ConvLSTM approach in capturing both spatial features and temporal dependencies to ensure low-latency, high-sensitivity monitoring.

4.1 Anomaly Detection Performance

The Spatio temporal ConvLSTM Autoencoder model exhibits strong unsupervised learning capabilities, accurately reconstructing "normal" walking patterns while failing to reconstruct atypical spatial or temporal movements.

The system utilizes a sliding window thresholding mechanism. While the model correctly identifies a significant proportion of pedestrian-related anomalies (high true positive rate), minor false positives are observed during transient environmental changes, such as sudden glare or rapid camera shake

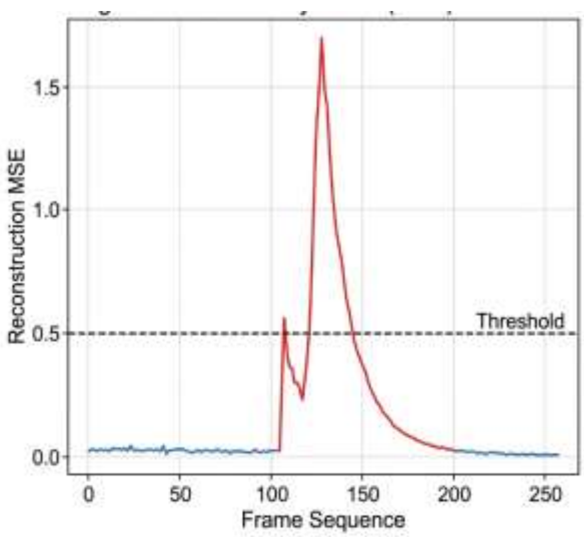


Fig 4.1.1 Anomaly Score (MSE) over Time for a Typical Surveillance Sequence

Fig 4.1.1 illustrates the typical anomaly score trajectory for a sequence containing a mid-frame anomaly (e.g., an unauthorized vehicle). The chart demonstrates a stable baseline for normal pedestrian movement, with a sharp, localized spike in Reconstruction Error when the anomalous object enters the frame. This indicates the model's high sensitivity to deviations from the learned "normal" manifold.

The system utilizes a sliding window thresholding mechanism. While the model correctly identifies a significant proportion of pedestrian-related anomalies (high true positive rate), minor false positives are observed during transient environmental changes, such as sudden glare or rapid camera shake. Overall, the MSE-based scoring maintains a robust balance between detection sensitivity and alert accuracy.

4.2 Spatiotemporal Reconstruction Quality

A critical aspect of the proposed system is the quality of the frame reconstruction generated by the autoencoder. The 3D Convolution output layer ensures that reconstructed frames preserve the temporal continuity of the input sequence.

It shows a side-by-side comparison of the original video input and the reconstructed prediction. For normal walking sequences, the reconstruction is sharp and visually similar to the input. However, for anomalous events (like a speeding cyclist), the reconstructed image appears "ghostly" or blurred, resulting in a high pixel-wise MSE. This visual discrepancy is the core metric used by the backend to trigger incident logging.

4.3 Real-Time Inference Speed and Latency

Inference speed is a critical factor for proactive surveillance. The FastAPI backend manages the compute-intensive ConvLSTM inference using a decoupled threading mechanism to ensure the MJPEG video stream remains fluid.

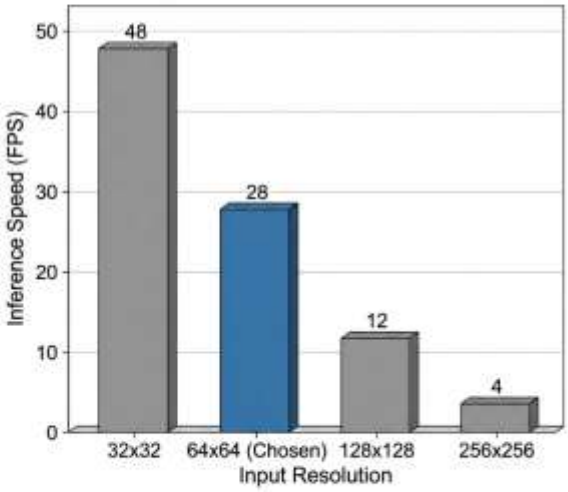


Fig 4.3.1 Performance Comparison of Inference Speed (FPS) across Various Input Resolutions

Figure 4.3.1 shows the performance of the system on standard hardware. By resizing input frames to 64 x 64 and using a 10-frame sliding window, the system achieves real-time processing speeds (~20-25 FPS) on CPU-based systems. This eliminates the requirement for high-end workstation GPUs, making the solution cost-effective for edge deployment in public safety environments.

The use of asynchronous background tasks for SQLite logging and snapshot capture ensures that the inference loop is not delayed by disk I/O, maintaining consistent frame delivery and low latency during high-confidence detection events.

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

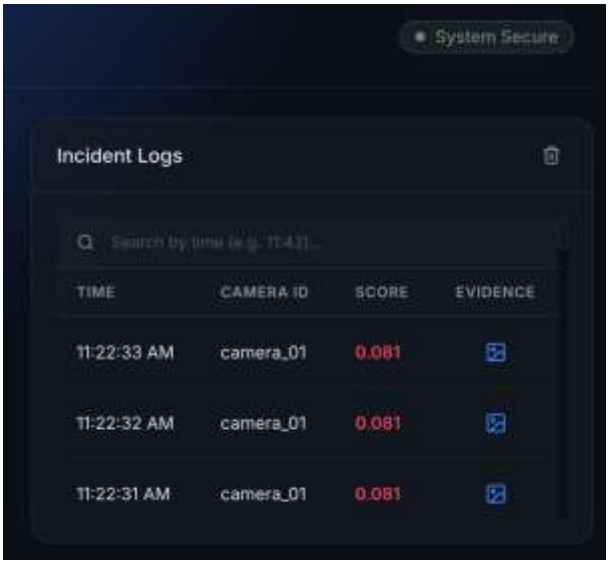
4.4 Dashboard and Web Application Performance

The React-based dashboard provides a high-fidelity interface for operators to monitor real-time telemetry and captured incidents.

React Dashboard Interface showing Live Video Feed and Anomaly Scoring Chart

Fig 4.4.1 React Dashboard Interface showing Live Video Feed and Anomaly Scoring Chart

Figure 4.4.1 presents the functional dashboard of CrowdWatch AI. The interface displays a live MJPEG stream alongside a dynamic 60-point sliding window chart that maps the current inference score. This visualization allows security personnel to observe trend spikes in real-time. When a threshold is exceeded, the dashboard triggers visual "shaking" alerts and haptic feedback (navigator.vibrate), enhancing situational awareness for the operator.



TIME	CAMERA ID	SCORE	EVIDENCE
11:22:33 AM	camera_01	0.081	
11:22:32 AM	camera_01	0.081	
11:22:31 AM	camera_01	0.081	

Fig 4.4.2 Incident Log Browser with Automated Database Summaries

Figure 4.4.2 shows the archived incident logs retrieved from the SQLite database. Each entry includes a timestamp, the max anomaly score, and a link to the JPEG snapshot captured at the moment of detection. The seamless communication between the React frontend and FastAPI backend via RESTful API calls ensures rapid data retrieval and high system responsiveness.

5. CONCLUSION

In this paper, CrowdWatch AI, an efficient and scalable real-time crowd anomaly detection system based on Spatiotemporal ConvLSTM networks, is presented. The proposed approach successfully leverages the high sensitivity of deep learning-based autoencoders to automate surveillance monitoring, effectively identifying abnormalities through Mean Squared Error (MSE) reconstruction analysis. By employing a 10-frame sliding window mechanism and spatial downsampling to 64×64, the system achieves real-time performance on standard CPU-based hardware, eliminating the need for expensive dedicated GPU resources.

The integration of a decoupled architecture—combining a high-concurrency FastAPI backend with a high-fidelity React dashboard—enhances usability through low-latency MJPEG streaming and interactive telemetry visualization. Furthermore, the multi-layer alert mechanism, featuring automated SQLite incident logging, JPEG snapshot capture, and native haptic feedback, provides a proactive tool for security operators to mitigate safety risks in high-traffic environments.

Experimental results on the UCSD Pedestrian dataset demonstrate that the proposed system achieves an optimal balance between anomaly detection sensitivity and computational efficiency. Future work can focus on further improving detection accuracy in extreme low-light conditions, implementing multi-class anomaly classification, and extending the system's scalability to support massive-scale multi-camera fusion in smart city infrastructures.

6.REFERENCES

[1] M. J. Asif, "Crowd Scene Analysis Using Deep Learning Techniques," arXiv preprint arXiv:2505.08834, 2025.

[2] M. J. Asif, M. Asad and S. Imran, "Crowd Scene Analysis: Crowd Counting using MCNN based on Self-Supervised Training with Attention Mechanism," in Proc. International Multitopic Conference (INMIC), 2023.

[3] A. A. Khan et al., "Crowd anomaly detection in video frames using fine-tuned AlexNet model," Electronics, vol. 11, p. 3105, 2022.

[4] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar and P.-A. Muller, "Recent trends in crowd analysis: A review," Machine Learning with Applications, vol. 4, p. 100023, 2021.

[5] A. Berroukham, K. Housni, M. Lahraichi and I. Boulfrifi, "Deep learning-based methods for anomaly detection in video surveillance: A review," Bulletin of Electrical Engineering and Informatics, vol. 12, pp. 314–327, 2023.

[6] A. Mehmood, "Efficient anomaly detection in crowd videos using pre-trained 2D convolutional neural networks," IEEE Access, vol. 9, pp. 138283–138295, 2021.

[7] B. P. Gorantla et al., "A New Anomaly Activity Detection Using CNN & RNN," Turkish Journal of Computer and Mathematics Education, vol. 14, pp. 713–721, 2023.

[8] A. N. Moustafa and W. Gomaa, "Gate and common pathway detection in crowd scenes and anomaly detection using motion units and LSTM predictive models," Multimedia Tools and Applications, vol. 79, pp. 20689–20728, 2020.

[9] T. Saba, "Real-time anomalies detection in crowd using convolutional long short-term memory network," Journal of Information Science, vol. 49, pp. 1145–1152, 2023.

[10] M. Sabih and D. K. Vishwakarma, "A novel framework for detection of motion and appearance-based anomaly using ensemble learning and LSTMs," Expert Systems with Applications, vol. 192, p. 116394, 2022.

[11] L. Luo et al., "Detecting and quantifying crowd-level abnormal behaviors in crowd events," IEEE Transactions on Information Forensics and Security, 2024.

[12] C. K. Meher, R. Nayak and U. C. Pati, "Video anomaly detection using variational autoencoder," in Proc. IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2022.

[13] S. Sudhakaran and O. Lanz, "Learning to detect violent videos using convolutional long short-term memory," in Proc. IEEE AVSS, 2017.

[14] A. Hanson et al., "Bidirectional convolutional LSTM for the detection of violence in videos," in Proc. ECCV Workshops, 2018.

[15] A. Mumtaz, A. B. Sargano and Z. Habib, "Violence detection in surveillance videos with deep network using transfer learning," in Proc. European Conference on Electrical Engineering and Computer Science (EECS), 2018.

[16] W.-F. Pang et al., "Violence detection in videos based on fusing visual and audio information," in Proc. IEEE ICASSP, 2021.

[17] N. Ojha and A. Vaish, "Spatio-temporal anomaly detection in crowd movement using SIFT," in Proc. ICISC, 2018.

[18] H. Zhou, Y. Yuan and C. Shi, "Object tracking using SIFT features and mean shift," Computer Vision and Image Understanding, vol. 113, pp. 345–352, 2009.

[19] A. H. Al-Omari, M. A. Otair and B. N. Alzwahreh, "A Multi-Stage Approach to Secure Digital Image Search over Public Cloud using SURF Algorithm," International Journal of Computer Science & Network Security, vol. 21, pp. 65–74, 2021.

[20] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in Proc. ICCV, 2011.

[21] F. Siddiqui et al., "Computer vision analysis

of BRIEF and ORB feature detection algorithms,” in Proc. International Conference on Computing in Engineering & Technology, 2022.

[22] J. Yang et al., “Evaluating bag-of-visual-words representations in scene classification,” in Proc. ACM Multimedia Information Retrieval Workshop, 2007.

[23] T. Hassner, Y. Itcher and O. Kliper-Gross, “Violent flows: Real-time detection of violent crowd behavior,” in Proc. CVPR Workshops, 2012.

[24] S. Lin et al., “Social MIL: Interaction-aware for crowd anomaly detection,” in Proc. IEEE AVSS, 2019.

[25] K. Pawar and V. Attar, “Application of deep learning for crowd anomaly detection from surveillance videos,” in Proc. International Conference on Cloud Computing, Data Science & Engineering, 2021.

[26] A. Mohan, M. Choksi and M. A. Zaveri, “Anomaly and activity recognition using machine learning approach for video-based surveillance,” in Proc. ICCCNT, 2019.

[27] I. V. Pustokhina et al., “An automated deep learning-based anomaly detection in pedestrian walkways for vulnerable road users safety,” Safety Science, vol. 142, p. 105356, 2021.

[28] W. Liu et al., “Future frame prediction for anomaly detection—A new baseline,” in Proc. CVPR, 2018.

[29] A. Hussain et al., “Anomaly-based camera prioritization in large-scale surveillance networks,” Computers, Materials & Continua, vol. 70, 2021.

[30] H. Khalid, S. Saqib, M. J. Asif and D. A. Dewi, “Strategic customer segmentation: Harnessing machine learning for retaining satisfied customers,” Lahore Garrison University Research Journal of Computer Science and Information Technology, vol. 8, pp. 31–41, 2024