ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

UNVEILING ANOMALIES IN CROWDS THROUGH ADVANCED DEEP LEARNING FOR UNUSUAL ACTIVITY DETECTION

MUTHURASU N[1], RAJASEKAR V[2]

[1,2]Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani campus, Chennai, India.

E-mail: [1]muthurasun@gmail.com , [2] brahmaasthra@gmail.com

ABSTRACT

Abnormal activity in the modern environment suggests risks and threats to other people. Anomaly refers to anything that differs from what is typical, anticipated, or normal. Given the challenges of consistently monitoring public areas, the implementation of intelligent video surveillance is imperative. Detecting unusual crowd activities is a complex subject that has spurred research advancements in the field of surveillance video applications. The main objective of this research is to identify atypical gatherings, instances of anomalous crowd behavior. Various techniques, including histogram representation, optical flow calculation, and deep learning-based algorithms, have been employed to address these issues. Nevertheless, there is a deficiency in effectively addressing this issue due to blockage, noise, and congestion. The introduction of AI techniques resulted in significant technological advancements. During the real-time monitoring of video material, the system employs various techniques to differentiate between different suspicious activities. The unpredictability of human behavior makes it challenging to discern whether it is suspicious or typical. Conducting monitoring commonly involves pulling consecutive frames from a video. There are two components in the framework. During the initial stage, the framework computes the features from the video frames. In the subsequent step, the classifier utilizes these features to determine if the class is panic or normal. The suggested methodology is evaluated using three available datasets, namely PETS 2009, MED and UMN dataset. The suggested method is compared with existing techniques to assess its efficiency.

Keywords: *Abnormal Activity, Anomalous, Video Surveillance, Deep Learning, Crowd Behavior, Video Frames, PETS, UMN Dataset.*

1. INTRODUCTION

Surveillance video analysis involves the crucial duty of detecting and tracking various moving objects. This process is essential for identifying unexpected events, abnormal crowd behavior, and human behaviors. Typically, the conventional video surveillance system (VSS) requires a human operator to monitor and identify relevant information to notify the public or themselves if necessary. Therefore, it is necessary to always notify a human operator, without interruption. In practice, determining whether an event is normal or abnormal solely based on the history of moving objects in a frame sequence is unattainable and time-consuming. Hence, the implementation of an intelligent VSS is crucial to efficiently capture and promptly respond to various scenarios based on the current condition in real time [1-3].

The increasing difficulties in urban surveillance and crowd management have led to a demand for advanced technologies that can identify abnormal actions in crowded areas. The emergence of deep learning has completely transformed the field of anomaly detection, providing unparalleled ability to detect and distinguish small deviations from typical behavior. Deep Guard, an innovative method, utilizes sophisticated deep learning techniques to detect abnormalities in crowds. A resilient anomaly detection system is required to account for the ever-changing dynamics of crowd behavior shaped with various factors, including social events, public gatherings, and urban activities. Traditional approaches frequently fail to accurately capture the complexities of unusual

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

behavior because they depend on pre-established norms and the creation of certain characteristics. Deep Guard utilizes deep neural networks to independently acquire knowledge about patterns and anomalies, enabling it to effectively manage intricate and dynamic situations. To identify uncommon crowd activities, this study suggests a strategy that takes a holistic approach based on odd crowd events.

1.1 Key terms and Concepts

Based on the context specific, the key term crowd anomaly detection defines the process of identifying unusual or abnormal activities within a crowd, using various computational techniques, with a focus on deep learning approaches. The context of the research is to develop methods that can effectively unveil anomalies in crowded settings. The other key term represents the deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or Transformers, capable of learning intricate patterns and representations from complex data. These models are employed to process and analyze crowd data for anomaly detection, leveraging their ability to capture hierarchical features. Feature extraction is the process of selecting or transforming relevant information from raw data, emphasizing important aspects for subsequent analysis. In the context of crowd anomaly detection, feature extraction methods play a crucial role in identifying distinctive patterns that can signify unusual activities.

The ensuing sections of the paper are organized in the following manner: Section II explores the pertinent literature. Section III elucidates the proposed methodology. Examine the empirical findings in Section IV. Section 5 elucidates the conclusion and outlines the future trajectory.

2. RELEVANT LITERATURE

This section presents a summary of the identification of unforeseen crowd behavior, which can be categorized into two methods: supervised learning-based techniques and threshold technique-based approaches.

Chakole et al. [4] introduced a methodology that relies on the threshold technique. The author employed the UMN dataset to validate the model. As the initial step, the model processed the supplied video by extracting each frame individually and performing background subtraction. An optical flow analysis has been conducted on consecutive frames, and the correlation coefficients have been examined. The judgement criteria have been established based on the requirement that the correlation be below the threshold value for non-anomalous events. An anomalous event is found if the system does not meet this criterion.

Luque Sánchez et.al [5] reviewed the deep learning techniques and dataset approaches for crowd anomaly detection. A novel hierarchical taxonomy has been presented to address the various steps involved in analyzing crowd behavior. Prior classifications of works assigned disparate activities equal importance, resulting in a basic listing of issues that could arise while analyzing crowd conduct. This structure was deficient, as it failed to consider the interdependence of sub-tasks. Within our newly established structure, writers now specify distinct stages within a comprehensive pipeline.

Bhuiyan et.al [6], reviewed video analytics using CNN. The technique efficiency has been verified by global regression. The abnormal event detection has been done by various CNN techniques. Feature learning based on the PCANet. Most existing approaches need manual selection of spatial and temporal properties, including intensity, colour, gradient, and optical flow. The calculation of the 3D gradient is performed for video events.

The authors, Lalit et al. [7], have conducted an analysis on the detection of aberrant crowd behaviour in video sequences using a supervised convolutional neural network.

Sonkar et al.[8], Created a system that can identify and classify typical and atypical behaviours in crowds by employing a continuous video surveillance system and utilising a Convolutional Neural Network (CNN) algorithm to analyse densely populated urban areas.

Bamaqa et al. [9], The study showcases the data preparation process, including the aggregation of data and the introduction of new crucial features, such as the amount of overcrowding and the severity of crowds. These features are valuable for the development of crowd prediction and anomaly detection models.

Sharif et al. [10], Create a comprehensive framework called rpNet to efficiently detect

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

anomalies in movies by combining RNet (reconstruction network) and PNet (prediction network) to propose a series of deep models. Within the RNet, one has the option to utilise convolutional autoencoder. The combination of RNet and PNet amplifies discrepancy in errors. Deep learning models provide varying degrees in the process of feature extraction.

3. PROBLEM STATEMENT

In densely populated public spaces, the potential for unusual or anomalous activities poses significant challenges for ensuring safety, security, and effective crowd management. Traditional methods of anomaly detection struggle to cope with the dynamic and complex nature of crowd behavior. The need for advanced surveillance systems that can reliably unveil anomalies in real-time has become increasingly critical. This research addresses the following key problems:

3.1 Ineffectiveness of Conventional Anomaly Detection

Traditional anomaly detection methods often lack the ability to discern subtle deviations in crowded scenarios, leading to a high rate of false positives or overlooking genuine anomalies. Inaccurate anomaly detection can compromise public safety and security, resulting in delayed responses to potential threats or disturbances.

3.2 Dynamic Nature of Crowds

Crowds exhibit intricate and dynamic behavior, making it challenging to establish a baseline for normal activities. Sudden changes, such as in crowd density, movement patterns, or group formations, are indicative of anomalies but are difficult to identify manually. Failure to adapt to the evolving nature of crowd dynamics hinders the detection of unusual activities, leaving public spaces susceptible to security breaches or emergent threats.

3.3 Scalability Challenges

The increasing scale of crowded environments, such as transportation hubs, public events, or urban centers, demands anomaly detection systems that can efficiently scale to process vast amounts of data in real-time. Inefficient scalability limits the deployment of effective anomaly detection solutions, hindering their applicability in large and complex crowd scenarios.

3.4 Privacy Concerns and Ethical Considerations

Advanced deep learning models capable of crowd anomaly detection may inadvertently infringe upon individual privacy rights, raising ethical concerns about the surveillance and monitoring of public spaces. Failure to address ethical considerations may lead to public resistance, legal challenges, and potential misuse of surveillance technologies, compromising the societal acceptance of anomaly detection systems.

3.5 Need **for Real-time Responses**

Timely response to unusual activities is crucial for preventing or mitigating potential threats. Many existing systems lack the ability to provide real-time insights into anomalies, leading to delayed reactions. Delayed responses reduce the effectiveness of security measures, increasing the likelihood of harm or disruptions in crowded environments.

4. PROPOSED METHODOLOGY

The deep convolutional framework for aberrant behavior detection in a smart surveillance system comprises three elements.

- Identification and differentiation of human subjects
- A module for classifying postures.
- A module designed to detect and identify anomalous behaviours.

Figure 1 illustrates the proposed methodology for detecting crowd behavior, specifically distinguishing between suspicious and normal behavior [11]. This is achieved by employing CNN model in input films. Video preprocessing removed frames from the input videos and cleaned the dataset using this procedure. We partitioned the dataset into three subsets: train, test, and validation. The CNN classification technique is used to categorize crowd behavior into suspect and normal based on many variables [12].

Journal of Theoretical and Applied Information Technology
15th March 2024. Vol.102. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

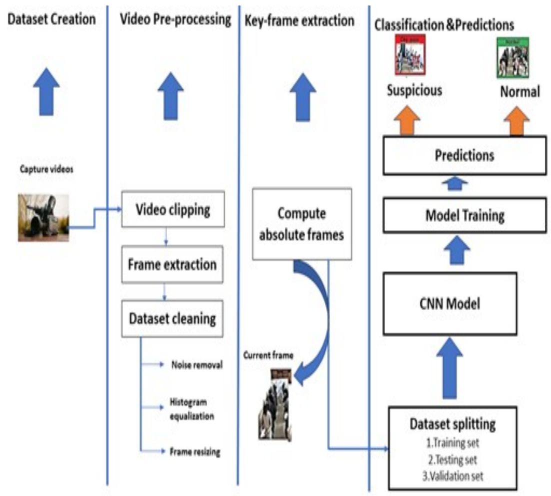

Fig.1 System flow

4.1 Network architecture

DNN (Deep Neural Networks), focused in computer vision applications. This article aims to expand the use of deep neural networks to three-dimensional structures to effectively capture and analyse spatio-temporal characteristics present in video feeds. In this video surveillance system, we will present a spatio-temporal autoencoder that utilises a 3D convolutional network and shown in Fig.2. The encoder module retrieves both spatial and temporal data, while the decoder module restores the frames. The identification of anomalous events is achieved by calculating the loss, determined by measuring the Euclidean distance between the original batch and the reconstructed batch[13].

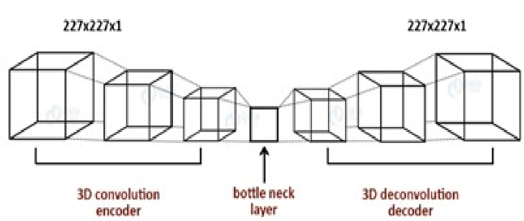

Fig.2. CNN Architecture-Spatio Temporal Auto Encoder

The paper introduces a novel model called AMDN (Anomalous Detection Network), which aims to detect anomalous events The model autonomously acquires feature representations. The model employs stacked de-noising auto encoders to learn appearance and motion aspects independently and collectively. Following the learning process, several one-class Support Vector Machines (SVMs) are trained. The Support Vector Machines (SVM) calculate the anomaly score for each input. Subsequently, these scores are aggregated to identify any anomalous occurrence. A dual fusion structure is employed. The computational burden associated with testing time is excessively large for real-time processing [14]. In Fig.3, the process of the proposed work has been visualized.

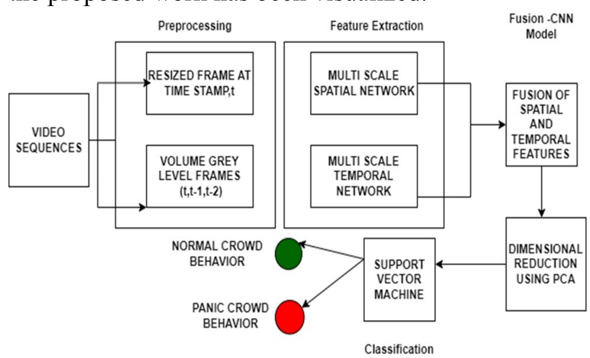

Fig.3. Block Diagram Of Proposed Model

The article presents a novel approach for detecting anomalies in movies by combining the features of auto encoders and convolutional neural networks (CNNs) in a deep convolutional auto encoder architecture. An autoencoder consists of an encoder component and a decoder component. The encoder component comprises convolutional and pooling layers, whereas the decoding component consists of deconvolutional and unpooling layers. The design enables the integration of low-level frames with high-level appearance and motion elements [15]. Reconstruction mistakes are used to indicate anomaly scores. Exploring deeper convolutions implies enhancements over conventional neural networks. Sparse architectures are implemented by replacing fully connected layers with sparse ones [16]. The paper proposes the use of dimensionality reduction techniques to mitigate the growing need for computer resources. Reductions in computing occur through the use of 1×1 convolutions prior to attaining 5×5 convolutions. The method does not provide information regarding the duration of execution [17]. The block details are given in the Table.1.

Journal of Theoretical and Applied Information Technology
15th March 2024. Vol.102. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table.1 Block details of 3D CNN Model

Layer Name	No. of Kernels	Kernel Size
Convolution3D	32	5×5×5
ReLU		
Batch Normalization		
Convolution3D	**64**	5×5×5
ReLU		
Batch Normalization		
Convolution3D	**128**	4×4×4
ReLU		
Batch Normalization		
Convolution3D		
ReLU		
Batch Normalization	**256**	3×3×3
Convolution3D	**300**	3×3×3
ReLU		
Batch Normalization		
Convolution3D	**300**	3×3×3
ReLU		
Batch Normalization		
Convolution3D	**320**	2×2×2
ReLU		
Batch Normalization		
Convolution3D	**16**	5×5×5
ReLU		
Batch Normalization		
Convolution3D	**32**	5×5×5
ReLU		
Batch Normalization		
Convolution3D	**64**	4×4×4
ReLU		
Batch Normalization		
Convolution3D	**128**	3×3×3
ReLU		
Batch Normalization		
Convolution3D	**256**	3×3×3
ReLU		
Batch Normalization		
Convolution3D	**320**	3×3×3
ReLU		
Batch Normalization		
Convolution3D	**340**	2×2×2
ReLU		
Batch Normalization		
Global Average Pooling	-	-
Average Pooling	-	2×2×2

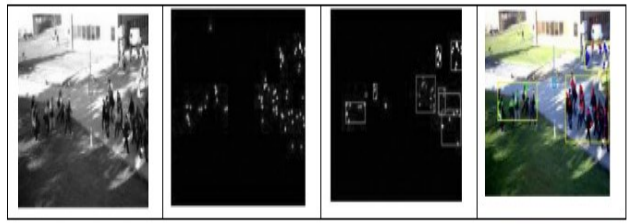

Fig.4. The Process Of Proposed Work

4.2 Pre-processing

The frames are resized with dimensions of 227 × 227 × 3. The frames after rescaling represented in set s = $\{s_1, s_2, ..., s_N\}$. Let V be a set that represents the N number of frames in terms of volume, represented as $\{v_1, v_2, ..., v_N\}$. The grayscale frames captured at t, $t-1$, and $t-2$. Every unit of volume in set V is resized to dimensions of [227 × 227 × 3].

4.3 Spatial Temporal Features extraction

There are three main methods are available for spatial temporal features extraction like CNN-RNN, 3D CNN, ConvLSTM. The CNN-RNN model initially employs a CNN network to extract spatial information, which is subsequently forwarded to an RNN network to capture temporal features [18]. It effectively extracts two distinct types of characteristics and demonstrates exceptional performance in specific applications, such as image captioning. 3D CNN are capable of feature extraction. ConvLSTM incorporates convolutional operations into the standard LSTM architecture, enabling it to process picture data. This unit extracts both geographical and temporal characteristics [19].

Utilising the multiscale characteristics enables the management of scale variation caused by perspective distortion. The purpose of employing the three-dimensional CNN, to encompass a broader region while maintaining the parameters kernel to those of the regular [20].

The activated feature maps are utilised for analysing multiscale spatial and appearance features, and they are denoted as $G_{ij}^{mas}|_{i=2,\ j=2,3,4..}$.

The chosen feature maps in various scales are inputted into Global Average Pooling (GAP) layers to extract statistical information [21]. These layers utilise the chosen data. The temporal characteristics can be denoted as $G_{ij}^{mts}|_{i=2,\ j=2,3,4...}$The spatial-temporal properties are combined as [G_{ij}^{mas} , G_{ij}^{mts}]. A solitary neuron tracks the combined feature sets to represent the regular sequences of video.

P represent expected result, where P represents a set including elements $p1$, $p2$, ..., pN. Similarly, let F represent the ground truth labels, where F is a set containing elements f, f2, ..., fN. Due to the limited availability of crowd scenes, we were only able to obtain a single class. For this specific purpose, the linear activation at the output layer has been implemented. The equation provided defines the MSE, the loss between P and F.

$$loss = \frac{1}{N} \sum_{k=1}^{N} (p_k - f_k)^2 \qquad (1)$$

The optimisation challenge mentioned above has been solved using Adam. The combined spatial-temporal features (f) are vectors with one dimension as shown in Fig.4. The SVM can be provided with the identical feature vector for crowd panic detection, but this may result in an escalation of computational complexity overheads for the SVM. Therefore, taking this fact into account, we have employed Principal Component Analysis to decrease the dimensionality [22].

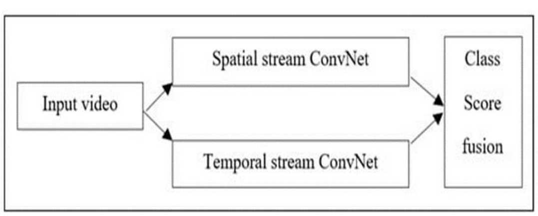

Fig.5 Dual Stream CNN [Spatial and Temporal] [23]

4.4 Crowd Classification using SVM

The SVM is used to model the usual crowd situations by reduced feature set. To train, we possess feature maps exclusively from a single class. In general, the SVM aims to maximise the distance between the hyperplane and its origin in the dimensionally reduced feature space. By employing a binary function, we can acquire knowledge of a specific area that encompasses the input data of a typical crowd scene. This function outputs a value of one if the data points fall inside this area and -1 if they are considered outliers or indicative of panic. The equation optimize the basic problem of quadratic programming and it has been shown below.

$$\min_{\omega,\rho} 0.5\|\omega\|_2 + \frac{1}{\upsilon N}\sum_{i=1}^{N}\xi_i - \rho \text{ such that } \{\omega.\varphi(k_i) \geq \rho - \xi_i \text{ and } \xi_i \geq 0 \qquad (2)$$

Here, ki represents the training data that corresponds to the ith sequence in normal sequences. The function φ () translates original to a higher-dimensional space. ω vector, perpendicular to the hyperplane. The variable υ is an upper bound for the outlier. ξ represents the relaxation variable.

4.5 Experimental analysis

The batch size for all datasets was configured as 8. This paper employed three strategies to mitigate overfitting: firstly, utilised the L2 norm to regularise the kernel weights; secondly, early stopping to terminate the network; and thirdly, it performed data augmentation during training. The learning rate is assigned a value of 0.001. The coefficient for the regularised parameter (L2) is assigned a value of 0.01. The momentum parameter of batch normalisation is set to 0.95, whereas the alpha parameter of ReLU is set to 0.1. Data augmentation has been employed during training to mitigate overfitting on the limited datasets. During data augmentation, patches of scale [227×227×3] are retrieved from the source frames. These patches are specifically taken from frames that contain crowd scenes. Random rotations of 250, 300, and 450 are applied to one-third of the retrieved patches. Data augmentation encompasses 70% of the original training samples.

Journal of Theoretical and Applied Information Technology
15th March 2024. Vol.102. No 5
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. RESULTS AND DISCUSSION

5.1 UMN dataset

The UMN dataset consists of eleven sequences, totalling around 7739 frames. These frames capture both regular and panic crowd activity [24]. The proposed model has been compared with the state of art methods. In Table.2, the comparative analysis between proposed model and existing study.

Table.2 UMN Dataset-Comparison with state of art method

Sequences	Evaluation Metrics			
	Ammar et.al[25]		Proposed model [CNN-SVM]	
	Error rate	Accuracy	Error rate	Accuracy
S1	0.90	99.0	0.48	99.52
S2	0.80	99.1	0.48	99.52
S3	0.00	100.0	0.48	99.52
S4	0.10	99.80	0.48	99.52
S5	0.10	99.90	0.48	99.52
S6	0.50	99.50	0.48	99.52
S7	0.30	99.60	0.48	99.52
S8	1.30	98.60	0.48	99.52
S9	0.10	99.80	0.48	99.52
S10	1.00	99.00	0.48	99.52
S11	0.30	99.60	0.48	99.52
Avg	0.60	99.40	0.48	99.52

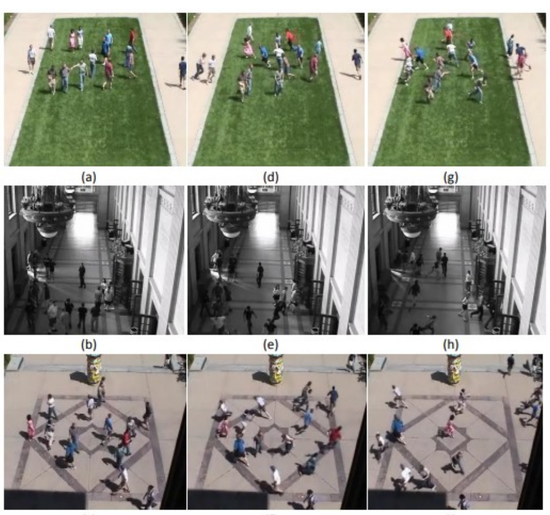

Fig.6.Sequences of model. Fig. (a), (b), (c) are normal sequences, and the Fig. (d),(e) ,(f) predicted as starting of panic behaviour. Fig. (g), (h),(i) predicted as panic.

5.2 MED Dataset

The performance evaluation of the proposed model is presented in Table.3, where it is compared with DeepROD using the MED dataset [25]. In Table.3, the comparative analysis on state of art method with proposed model.

Table.3 MED Dataset-Comparison with state of art method

Sequences	Evaluation Metrics			
	Ammar et.al[25]		Proposed model [CNN-SVM]	
	Error rate	Accuracy	Error rate	Accuracy
S1	4.00	95.50	1.54	98.46
S2	5.00	94.3	3.45	96.55
S3	2.00	97.7	2.28	97.72
S4	4.00	95.8	4.33	95.67
S5	6.00	93.4	1.36	98.64
S6	2.00	97.4	1.60	98.40
S7	0.40	99.5	5.33	94.67
S8	6.00	94.0	3.22	96.88
S9	1.70	98.3	2.49	97.51
S10	5.50	94.5	1.75	98.25
S11	9.00	91.0	1.99	98.01
Avg	4.00	95.6	2.39	97.61

The proposed model demonstrates effective utilization of multiscale spatial-temporal feature modeling for crowd panic detection, as evidenced by its performance on the MED dataset. Therefore, the model can address the problem of size caused by perspective distortion in the datasets of crowd panic videos. Figure.7 displays the results obtained from a selection of samples in the MED dataset.

Fig.7 MED dataset – The proposed model outputs. Fig. (a) normal sequence Fig. (b) starting of panic behaviour Fig.(c) Panic situation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.3 PET-2009 dataset

The performance of the proposed model is compared to various state-of-the-art models on the Pets-2009 dataset, as shown in Table.4. The suggested model has 98.37% accuracy in detecting crowd panic behavior. The model has an error rate of 1.63% on the Pets-2009 dataset. The suggested model outperforms recent approaches in terms of accuracy and false alarm rate for crowd panic detection.

Table.4 Comparison on state of art methods with proposed model- PET 2009 dataset

Sequences	Ammar et al.[25]		**Proposed Model**	
	Error rate	Accuracy	Error rate	Accuracy
Time14–16	2.70	97.30	1.77	98.23
Time14–17	0.20	97.80	1.49	98.51
Average	1.40	97.50	1.63	98.37

Fig.8 PET-2009 dataset – Fig. (a) Normal Sequence Fig.(b) Starting of panic behavior Fig. (c) Panic situation.

6. RESEARCH CONTRIBUTION

The key aspects of research are based on the deep learning models, feature representations, dataset collection and annotation, evaluation metrics, real world applications, algorithm robustness, integration with surveillance systems.

6.1 Deep Learning Models

The research likely involves the development and utilization of advanced deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), tailored for anomaly detection in crowded settings.
Model architectures might be designed to effectively capture spatial and temporal dependencies within the crowd data, allowing for more robust anomaly detection.

6.2 Feature Representation

The study may contribute to the exploration of novel feature representations that capture the intricate patterns and dynamics of crowd behavior. Feature engineering could involve the extraction of meaningful spatial and temporal features from the input data, enabling the deep learning models to better discriminate between normal and anomalous activities.

6.3 Dataset Collection and Annotation

The research may involve the creation or utilization of datasets specifically curated for studying anomalies in crowded environments. These datasets would likely be annotated to facilitate supervised learning approaches for model training. Annotations may include labels for normal and anomalous activities, providing the necessary ground truth for evaluating the model's performance.

6.4 Evaluation Metrics

The research might contribute to the development or refinement of evaluation metrics tailored to assess the efficacy of anomaly detection in crowded scenarios. Metrics focus on providing a comprehensive understanding of the model's performance.

6.5 Real-world Applications

The research likely explores the practical applications in real-world scenarios, such as public spaces, transportation hubs, or large events.The findings may have implications for enhancing public safety, security, and crowd management through the early detection of unusual activities.

6.6 Algorithm Robustness

The research may contribute insights into the resilience of deep learning models by varying environmental conditions, crowd densities, and camera perspectives.Strategies for improving model generalization and adaptability to different scenarios could be explored.

6.7 Integration with Surveillance Systems

The study may provide guidelines or insights into integrating advanced anomaly detection models with existing surveillance systems, enabling real-time monitoring and alerting.

7. CONCLUSION AND FUTURE WORK

The model, CNN-SVM, utilized the OCC technique to anticipate regular and panic crowd behaviours. The model effectively addressed the changes in human shape caused by perspective distortion in the crowd footage by utilizing scale-invariant characteristics. To demonstrate the effectiveness of the suggested model, a series of comprehensive experiments, comparative analysis of findings, and ablation studies were conducted on the benchmark datasets that are publicly accessible. The proposed model demonstrated detection accuracies of 99.40%, 97.61%, and 98.37% on the UMN, MED Panic, and Pets-2009 Panic datasets, respectively. The performance of the proposed model surpasses that of contemporary cutting-edge methods. The future research can be concentrated on multimodal data fusion by exploring the integration using video, audio, and sensor data, to create more comprehensive representations of the crowded environment. This can improve the robustness of anomaly detection systems by capturing diverse aspects of abnormal behavior. Develop methods for incremental learning that allow the model to adapt to evolving patterns of normal and anomalous behavior over time. This is particularly important in dynamic environments where crowd behavior may change over extended periods. Enhance the interpretability of deep learning models to provide more transparent and understandable insights into the decision-making process. This is crucial for gaining trust in the system and facilitating collaboration with human operators. Address privacy concerns by developing methods that allow for effective anomaly detection without compromising the privacy of individuals in the crowd. This can involve the use of privacy-preserving algorithms or techniques that prioritize data anonymization. Extend the temporal scope of anomaly detection to analyze long-term crowd behavior trends. This can contribute to understanding seasonal variations, recurring patterns, and other aspects that may impact anomaly detection performance over extended periods. Contribute to the development of benchmark datasets for evaluating anomaly detection models in crowded environments. Standardized evaluation metrics and protocols can facilitate fair comparisons between different approaches and promote advancements in the field.

REFERENCES

[1] GCR, K. (2020, July 25). Automatic Detection and Initiation for Multiple Human Tracking Using Particle Filter. Journal of Advanced Research in Dynamical and Control Systems, 12(SP7), 2607–2612. https://doi.org/10.5373/jardcs/v12sp7/20202396

[2] Neamah, K. A. (2022, January 20). Mathematical Model for Handling Unstable Time Series by Using a Linear Approximation Technique. Webology, 19(1), 2835–2852. https://doi.org/10.14704/web/v19i1/web19189

[3] Mustafa, M. (2022). Real-Time Video Anomaly Detection for Smart Surveillance. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4199296

[4] Chakole, P. D., Satpute, V. R., & Cheggoju, N. (2022, May 1). Crowd behavior anomaly detection using correlation of optical flow magnitude. *Journal of Physics: Conference Series*, *2273*(1), 012023. https://doi.org/10.1088/1742-596/2273/1/012023

[5] Luque Sánchez, F., Hupont, I., Tabik, S., & Herrera, F. (2020, December). Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. *Information Fusion*, *64*, 318–335. https://doi.org/10.1016/j.inffus.2020.07.008

[6] Bhuiyan, M. R., Abdullah, J., Hashim, N., & Al Farid, F. (2022, March 29). Video analytics using deep learning for crowd analysis: a review. *Multimedia Tools and Applications*, *81*(19), 27895–27922. https://doi.org/10.1007/s11042-022-12833-z

[7] Lalit, R., Purwar, R. K., Verma, S., & Jain, A. (2021, December 14). Crowd abnormality

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

detection in video sequences using supervised convolutional neural network. *Multimedia Tools and Applications*, *81*(4), 5259–5277. https://doi.org/10.1007/s11042-021-11781-4

[8] Sonkar, R., Rathod, S., Jadhav, R., & Patil, D. (2020). CROWD ABNORMAL BEHAVIOUR DETECTION USING DEEP LEARNING. *ITM Web of Conferences*, *32*, 03040. https://doi.org/10.1051/itmconf/20203203040

[9] Bamaqa, A., Sedky, M., Bosakowski, T., Bakhtiari Bastaki, B., & Alshammari, N. O. (2022, October). SIMCD: SIMulated crowd data for anomaly detection and prediction. *Expert Systems With Applications*, *203*, 117475. https://doi.org/10.1016/j.eswa.2022.117475

[10] Sharif, M. H., Jiao, L., & Omlin, C. W. (2023, March 23). Deep Crowd Anomaly Detection by Fusing Reconstruction and Prediction Networks. *Electronics*, *12*(7), 1517. https://doi.org/10.3390/electronics12071517

[11] Jia, D., Zhang, C., & Zhang, B. (2021, January 27). Crowd density classification method based on pixels and texture features. *Machine Vision and Applications*, *32*(2). https://doi.org/10.1007/s00138-021-01167-9

[12] Mehmood, A. (2021, April 14). Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks. Applied Sciences, 11(8), 3523. https://doi.org/10.3390/app11083523

[13] Wang, B., & Yang, C. (2022, June 20). Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder. Sensors, 22(12), 4647. https://doi.org/10.3390/s22124647

[14] Surucu, O., Gadsden, S. A., & Yawney, J. (2023, July). Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances. Expert Systems With Applications, 221, 119738. https://doi.org/10.1016/j.eswa.2023.119738

[15] Chen, S., & Guo, W. (2023, April 7). Auto-Encoders in Deep Learning—A Review with New Perspectives. Mathematics, 11(8), 1777. https://doi.org/10.3390/math11081777

[16] Smart Surveillance: A Review & Survey Through Deep Learning Techniques for Detection & Analysis. (2023, October 5). Journal of Sensor Networks and Data Communications, 3(1). https://doi.org/10.33140/jsndc.03.01.05

[17] Sreenu, G., & Saleem Durai, M. A. (2019, June 6). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0212-5

[18] Nafea, O., Abdul, W., Muhammad, G., & Alsulaiman, M. (2021, March 18). Sensor-Based Human Activity Recognition with Spatio-Temporal Deep Learning. Sensors, 21(6), 2141. https://doi.org/10.3390/s21062141

[19] Nawaz, A., Zhiqiu, H., Senzhang, W., Hussain, Y., Khan, I., & Khan, Z. (2020, April 8). Convolutional LSTM based transportation mode learning from raw GPS trajectories. IET Intelligent Transport Systems, 14(6), 570–577. https://doi.org/10.1049/iet-its.2019.0017

[20] Zou, Z., Cheng, Y., Qu, X., Ji, S., Guo, X., & Zhou, P. (2019, November). Attend to count: Crowd counting with adaptive capacity multi-scale CNNs. Neurocomputing, 367, 75–83. https://doi.org/10.1016/j.neucom.2019.08.009

[21] Shrestha, S., & Vanneschi, L. (2018, July 18). Improved Fully Convolutional Network with Conditional Random Fields for Building Extraction. Remote Sensing, 10(7), 1135. https://doi.org/10.3390/rs10071135

[22] Wan, B., Jiang, W., Fang, Y., Luo, Z., & Ding, G. (2021, May 22). Anomaly detection in video sequences: A benchmark and computational model. IET Image Processing, 15(14), 3454–3465. https://doi.org/10.1049/ipr2.12258

[23] Xia, L., Mi, S., Zhang, J., Luo, J., Shen, Z., & Cheng, Y. (2023, May 22). Dual-Stream Feature Extraction Network Based on CNN and Transformer for Building Extraction. Remote Sensing, 15(10), 2689. https://doi.org/10.3390/rs15102689

[24] Ben Mabrouk, A., & Zagrouba, E. (2018, January). Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems With Applications, 91, 480–491. https://doi.org/10.1016/j.eswa.2017.09.029Ammar, H., & Cherif, A. (2021, March 10).

[25] DeepROD: a deep learning approach for real-time and online detection of a panic behavior in human crowds. Machine Vision and Applications, 32(3). https://doi.org/10.1007/s00138-021-01182-w