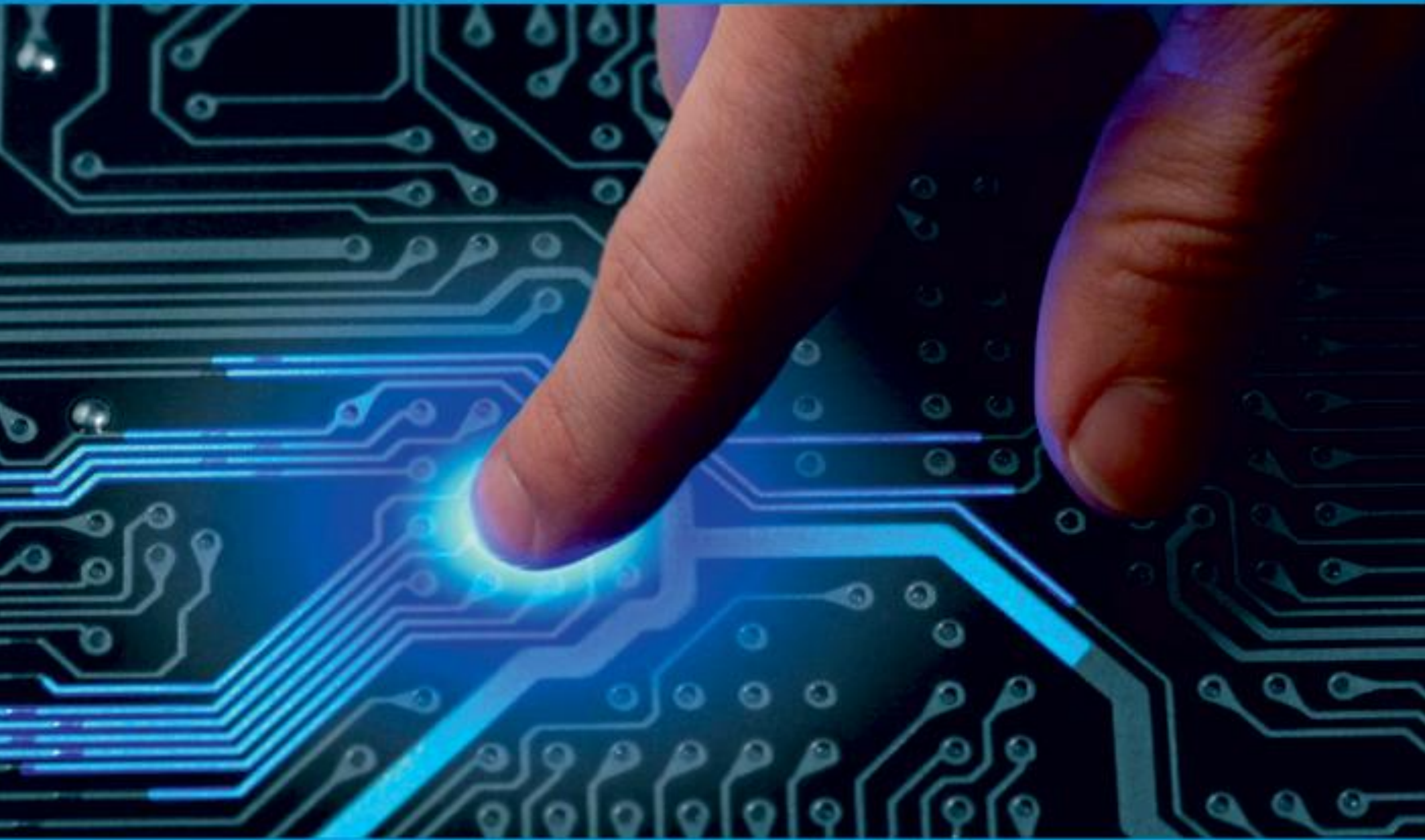
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| DOI: 10.15680/IJIRCCE.2024.1205013 |

Anomaly Activity Detection on Real Time Surveillance Videos with Deep Learning

Sai Deshith, Yashwanth Putluru, Sai Deshik, Deekshith Koripalli

Dept Artificial Intelligence and Machine Learning BMS Institute of Technology and Management Bengaluru, India

ABSTRACT: This project addresses a critical and contemporary challenge in surveillance security systems. Era where video data is abundantly generated, the need for robust methods to detect criminal or anomalous activities is paramount. This research focuses on leveraging Convolutional Neural Networks to tackle this problem, achieving an impressive accuracy rate of 95%. The abstract provides a concise overview of the study's objectives, methods, and key findings. It explores the growing importance of video data analysis in security applications and highlights the significance of accurate anomaly detection. The use of CNNs is emphasized as a key technology for achieving high accuracy in identifying unusual behavior in video streams. This research contributes to enhancing the effectiveness of surveillance systems, offering a valuable tool for law enforcement. The 95% accuracy achieved by the CNN model demonstrates its potential to significantly improve the detection of crime and anomalous activities, thereby enhancing public safety and security.

KEYWORDS: Crime detection, Anomaly detection, Video analysis, Surveillance, Convolutional Neural Networks (CNN)

I. INTRODUCTION

A substantial surge in investments by corporations, governments, and both public and private sectors, directed towards fortifying security measures across a diverse range of settings, including workplaces, structures, shopping centers, homes, and essential infrastructure. This pattern is anticipated to pick up speed in the rapidly evolving field of automated security in the coming years. The escalating threat posed by terrorist activities underscores the paramount importance of swiftly identifying suspicious or anomalous behaviors capable of disrupting normal human activities. Anomalous events, as elucidated in reference [1], encompass a wide spectrum of deviations from typical behavior, ranging from objects appearing in unexpected locations to irregular motion patterns, unauthorized access, unconventional traffic maneuvers, violent altercations, and abandoned items, atypical actions during military training, sudden and unexpected movements, and dropped objects. Notably, what constitutes normalcy in one context may be regarded as anomalous in another, underscoring the nuanced nature of anomaly detection.

The applications of anomalous event detection systems, as expounded in reference [1], span a multitude of domains, including traffic monitoring, medical science, high-security installations like military bases and airports, crowd analysis [2], recognition of criminal activities, and the automation of forensic video retrieval processes. These systems play an indispensable role in safeguarding safety and security across diverse sectors and scenarios. Among the array of aberrant behaviors frequently encountered, acts such as killing, looting, molestation, and severe assaults stand out. Killing involves the deliberate taking of another person's life, looting entails forcibly seizing someone else's property through extreme physical force, molestation pertains to the non-consensual sexual exploitation of individuals of all genders and ages, and intense assaults involve illegal confrontations often driven by motives of gain or harm towards others [3].

In the realm of video surveillance, particular emphasis is placed on crowded scenarios for the precise detection and characterization of anomalies within complex environments. An anomaly, simply put, refers to an uncommon or irregular event. It's worth noting that the task of modeling and processing results derived from anomalous scenes can be inherently challenging, at times appearing nearly insurmountable, as indicated in references [4, 5].

This research brings new perspectives on anomalies, enhancing feature engineering. It also tackles scalability issues, vital in cybersecurity and industrial monitoring. Additionally, it aids in adapting to evolving threats, ensuring agile responses. The findings result in transparent models crucial for decision-making. Moreover, the research advances automation, improving process efficiency by reducing manual intervention.

II. RELATED WORK

Previous studies in the realm of anomalous event detection have laid the groundwork for advancements in this field. In the study by Mehrsan et al. [6], the novel strategy was usedwas introduced to identify and distinguish between anomalous and dominant behaviors. This research not only contributes to the development of effective anomaly detection methods but also sheds light on the identification of dominant patterns, which can be valuable for understanding and modeling typical behaviors within various contexts. Building upon such foundational research can further enhance our ability to create robust and comprehensive systems for anomaly detection in diverse scenarios.

Researchers encountered a challenge when dealing with the utilization of relevant exemplars in scenarios with multiple features. In response to this issue, Zhongwen Xu.[7], and colleagues introduced an innovative concept. Their proposal involved the incorporation of multi-level relevance labels within videos to enhance event detection. This approach aimed to provide a more comprehensive understanding of the relationships between features and events, ultimately contributing to improved accuracy and efficiency in event detection processes. By assigning varying levels of relevance to different features, their method offered a promising solution to the intricate task of extracting meaningful information from complex video data.

In their research, Jifeng Ning and colleagues [8] introduced a novel method that combines active contour segmentation and registration approaches to improve object tracking precision. This method provides a robust means of tracking objects by seamlessly integrating registration and contour-based segmentation, thereby improving tracking precision, especially in challenging scenarios with heavy occlusion. Furthermore, Fan Yang and colleagues developed a discriminative appearance model to address the problem of severe occlusion in object tracking. Their innovative approach allows tracking algorithms to better adapt to situations where objects are partially or fully obscured, enabling more reliable and resilient tracking performance even under challenging conditions. These advancements collectively contribute to the evolution of object tracking methodologies.

In a research investigation as W. Sultani et al. [9], a comprehensive examination was conducted to evaluate the effectiveness of anomaly detection within urban settings. Anomalies were consolidated within a grid measuring 200 × 250 meters and observed retrospectively. To enhance anomaly identification, a novel approach utilising an ensemble model incorporating logistic regression and neural networks was developed. The findings of this study highlight a remarkable improvement in predictive accuracy when employing a fortnight-based prediction model in contrast to the conventional monthly predictions.

The research presented H.W. Kang et al. [10] introduced an approach to forecast anomalies through an analysis of historical anomaly data and their corresponding trends. K-Nearest Neighbors (KNN), Decision Trees, and other machine learning techniques were utilized in this research with Adaptive Boosting, and Random Forest, with the aim of enhancing the accuracy of anomaly predictions.

A. Rummens et al. [11], anomalous activities were discerned and monitored by analyzing 15 years' worth of anomaly data from Vancouver city. To find these unique events, decision tree algorithms and K-Nearest Neighbours (KNN) were both used. The examination of a sample comprising 560,000 entries from the anomaly dataset led to the prediction of anomalous activities with an accuracy range spanning from 39% to 44%.

In recent times, there has been a notable emergence of deep learning-based methodologies. Xu et al. [12] offered a novel method that relied on learning to acquire video features rather than manually constructed characteristics, using a machine learning framework. They then used several Single Class Support Vector Machine (SVM) models, putting the newly acquired features to use in determining the degree of abnormality for each input. The ultimate detection of abnormalities was produced by successfully combining the results from these many SVM models.

According to their study [13], Hasan and his associates developed the novel Framework for Convolutional Auto-Encoding (Conv-AE). The objective of this framework is to rebuild complicated situations. Subsequently, they leveraged the reconstructed scenes to compute reconstruction costs, a crucial step in identifying anomalies within the data. By combining the power of Conv-AE and reconstruction cost analysis, their approach provides an effective means of anomaly detection and contributes to the advancement of scene analysis and security applications.

Zhou and his team's research, detailed in reference [14], introduced a pioneering approach involving spatio-temporal Convolutional Neural Networks (CNNs). This innovative method aimed to comprehensively capture and understand the combined nuances of appearance and motion characteristics within data sequences. By leveraging spatio-temporal CNNs, their model successfully facilitated the simultaneous learning of both static and dynamic features, enabling more robust and nuanced analysis of data, particularly in domains where the temporal aspect is crucial. This advancement holds substantial promise for applications across various fields, from video analysis to action recognition and beyond. Mohammadi and colleagues [15] introduced an innovative behavioral heuristic approach designed for the classification of videos into two distinct categories: violent and non-violent. This novel method leverages behavioral cues and patterns within the video content to make precise determinations regarding the presence or absence of violent actions. By focusing on the behavioral aspects, this approach offers a promising avenue for more accurate and context-aware video classification, contributing to various applications such as content filtering, security surveillance, and content moderation in digital platforms.

Dataset Details: A new extensive dataset, known as UCF-Crime, has been introduced to assess a methodology. This dataset contains uncut surveillance videos of extended duration, covering Arson, Burglary, Explosion, Fighting, & Normal Events are five real-world abnormalities.

Dataset link:
https://www.crcv.ucf.edu/projects/real-world/

A. Deep Learning

Deep learning, a subset of machine learning, employs neural networks with three or more layers. These networks have intrinsic limitations despite their attempts to mimic real brain activity. They enable computers to 'learn' from vast datasets, with the potential for increased accuracy through additional hidden layers, even though single-layer networks can still provide approximate predictions. Deep learning underpins numerous artificial intelligence (AI) applications and services, automating mental and physical tasks without human intervention. With well-known products and services like voice-activated TV remotes, virtual assistants, as well as credit card fraud protection, this technology works seamlessly. Additionally, it supports innovative research and cutting-edge innovations like driverless vehicles. Fig. 1 depicts the deep learning architecture.

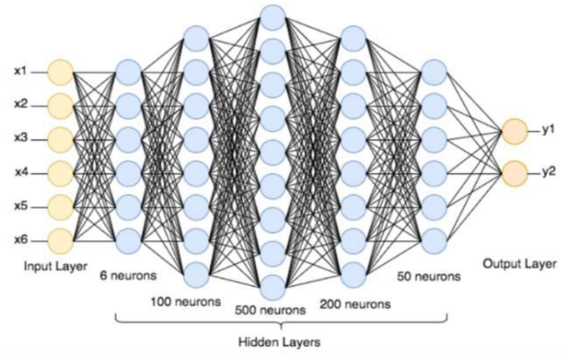


Fig 1: Architecture of Deep Learning

III. PROPOSED METHOD

This anomaly detection system's main goal is to infuse passive video surveillance with intelligence in order to transform it from what it now is. Its purpose is to capture and identify abnormal human activities within surveillance video footage. Following analysis, the system can then generate alerts, utilizing methods like alarms or messages, to preempt unusual behavior. Fig 2 below illustrates the architecture of the proposed system.

It was suggested that a deep-learning approach to crime classification be implemented to achieve this. Specifically, we employ Convolutional Neural Networks (CNNs) to enhance accuracy and precisely identify the nature of the observed activities. Among the various abnormal activities, such as Arson, Burglary, Fighting, Explosion, and normal activities in public spaces, there exists a pressing need for an intelligent surveillance system that can autonomously trigger alarms or alerts when irregular behavior is detected.

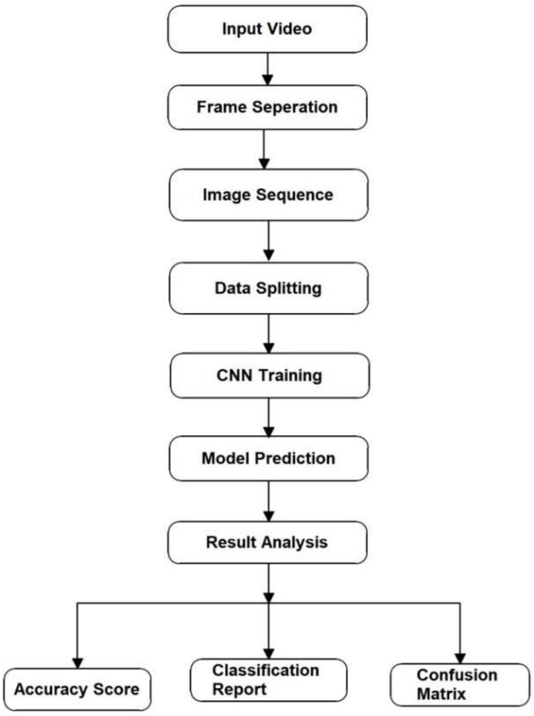


Fig 2: Proposed System Architecture

A. CNN

Convolutional Neural Networks (CNN), a deep learning method specifically designed for processing visual data, has revolutionized computer vision. Its widespread adoption spans diverse fields such as image categorization, object recognition, and image synthesis. At the heart of CNNs lie convolutional layers, fundamental for feature extraction. These layers employ filters or kernels to convolve over input data, multiplying receptive field values with filter weights to generate feature maps. This process enables the network to capture local patterns and spatial relationships among pixels. CNNs further employ pooling layers to reduce spatial dimensions and down sample feature maps. Towards the network's end, fully connected layers take center stage, culminating in classification or regression tasks using previously acquired features. Class probabilities are determined by feeding the output of all connected layers into a softmax activation function. A visual illustration of the CNN structure is shown in Fig 3 elucidating its structural components.

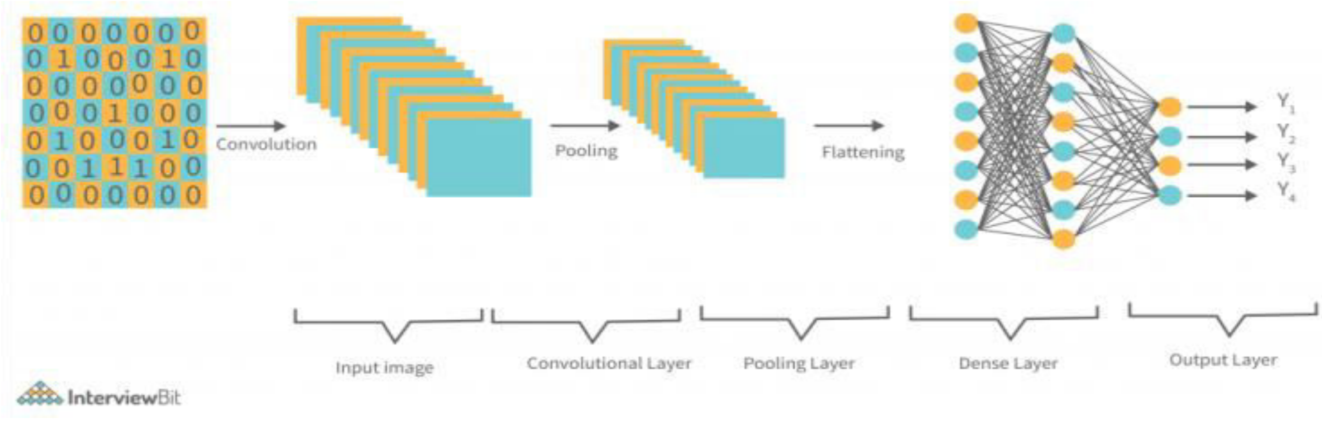


Fig 3: The CNN Architecture

The subsequent sections employ Convolutional Neural Network layers along with their corresponding mathematical formulations:
Conv2D Layer: The 2D convolution operation between an input tensor and a kernel tensor can berepresentedas:This operation is applied over all spatial locations .

$$\text{Conv2D}(X, K)_{i,j} = \sum_m \sum_n X_{i+m,j+n} \cdot K_{m,n}$$

MaxPool2D Layer (Max Pooling): Max pooling reduces the size of the input tensor by selecting themaximum value within a sliding window. If the window size is , the operation is:This operation is applied over non-overlapping windows.

$$\text{MaxPool2D}(X)_{i,j} = \max_{m,n} X_{n+i,n+j}$$

Dense (Fully Connected) Layer: The output of a dense layer with input and weights can becalculatedas:Here, represents the bias term.

$$\text{Dense}(X, W) = X \cdot W + b$$

IV. Result Analysis

Using important analytical criteria including recall, precision, and accuracy, the classification report provides a thorough review of the categorization model's performance precision, and F1-score. Additionally, it presents individual class contributions, summarizing the model's effectiveness in categorizing distinct groups. These metrics are pivotal in assessing the model's accuracy and its proficiency in correctly classifying instances across various categories. Making educated decisions about the model's applicability for particular classification jobs is made easier with the help of the categorization report, which provides relevant information about the model's performance. This report serves as a vital tool for gauging the model's ability to meet classification requirements and ensures that its predictions align with the intended objectives, thereby enhancing the quality and reliability of classification outcomes.

Precision (P): Precision, which measures how well the model predicts positive outcomes, is calculated as the ratio of the the proportion of real positives (TP) over the total number of both true and false positives (FP).

Precision = TP / (TP + FP)

The recall rate, also known as the true positive rate, gauges how many real positives & false negatives are accurately distinguished from actual positives. It evaluates the system's effectiveness at spotting good events.

Recall = TP / (TP + FN)

F1-score: The accuracy and recall components are harmoniously combined in the F1 score, which strikes a satisfactory equilibrium between the two of them.

*F1-score = 2 * (Precision * Recall) / (Precision + Recall)*

Support: The number of samples inside each target class is referred to as support.

A. *Hardware Requirements:*

The system needs a Processor Core i3/i5, 500 GB hard disc, and a 500 GB SSD. and 4 GB of RAM. Any desktop or laptop system meeting or exceeding these specifications can be employed for the setup.

B. *Software Requirements:*

For this software setup, you'll need Windows 8/10 as the operating system, Python as the programming language, Anaconda as the framework, Jupyter Notebook as the integrated development environment (IDE), and essential deep learning libraries including Numpy and Pandas.

The categorization report offers information on the model's effectiveness in each class in addition to an overall evaluation.

	Precision	Recall	F1	Support
CNN				
0	**0.99**	**0.89**	**0.94**	**142**
1	**0.93**	**0.98**	**0.95**	**89**
2	**0.95**	**0.96**	**0.96**	**141**
3	**0.94**	**0.99**	**0.97**	**129**
4	**1.00**	**1.00**	**1.00**	**2**

Table 1: Performance Metrics for CNN

The above table 1 presents performance metrics for a CNN model across five different classes. The precision values indicate the accuracy of positive predictions, with class 4 achieving a perfect score of 1.00. Class 4 once more received a flawless score on the recall metric, which assesses the model's capacity to recognize positive examples. The F1 score, which combines precision and recall, is highest for class 4 at 1.00, indicating excellent overall performance. Support represents the quantity of each class's instances, with the highest support in class 2 at 141. Overall, the CNN model demonstrates strong performance across most classes, with class 4 being particularly noteworthy for its perfect precision, recall, and F1 score.

C. *CNN-Confusion matrix*

By comparing a classification model's predictions to the dataset's actual numbers utilising the confusion matrix, also known as a square matrix, its performance may be assessed. The rows of the confusion matrix represent the real or true classes, and the columns represent the expected classes.

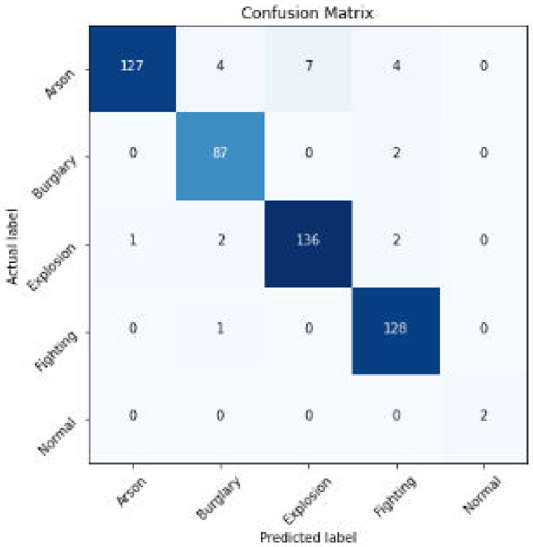


Fig 4: Confusion Matrix for CNN

Fig 4 displays the CNN model's confusion matrix, illustrating its performance effectively. The model attains an impressive accuracy rate of 95%.

V. Conclusion

In conclusion, the research presented in this abstract underscore the critical relevance of "Crime/Anomaly Activity Detection through Videos" in the modern landscape of surveillance and security systems. With the exponential growth of video data generation, the demand for robust methods to identify criminal and anomalous activities has become paramount. Leveraging Convolutional Neural Networks (CNNs), this study has achieved an impressive accuracy rate of 95%, showcasing the efficacy of deep learning techniques in addressing this pressing issue. The study's objectives, methodologies, and significant findings. It highlights the increasing importance of video data analysis in security applications and emphasizes the pivotal role of accurate anomaly detection. By harnessing CNNs, a formidable technology, this research offers a substantial contribution to enhancing surveillance systems, presenting a valuable tool for law enforcement agencies and security personnel. The CNN model's outstanding 95% accuracy stands as a testament to its potential to greatly enhance the detection of criminal and anomalous activities, ultimately fortifying public safety and security in an ever-evolving world.

VI. Future work

Firstly, there's potential in integrating multi-modal data sources like audio and thermal imaging to enhance the understanding of complex environments. Real-time processing improvements are crucial for immediate threat response. Incremental learning methods can ensure adaptability over time, and privacy preservation techniques should be explored to balance security and privacy concerns. Scalability solutions, such as distributed computing, can handle large video data volumes effectively. Understanding human behavior beyond anomaly detection, benchmark datasets, and addressing adversarial attacks are areas ripe for exploration. Incorporating human feedback and ethical considerations will be integral in shaping the evolution of video-based security and surveillance systems.

Reference

[1] Hampapur, A., Brown, L., Connell, J., Pankanti, S., Senior, A., Tian, Y., 2003. Smart surveillance: applications, technologies and implications. Information, Communications and Signal Processing 2, 1133–1138.
[2] Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N., 2010. Anomaly detection in crowded scenes., in: CVPR, p. 250.
[3] M. Asif, M.I. Tiwana, U.S. Khan, M.W. Ahmad, W.S. Qureshi, J. Iqbal, Human gait recognition subject to different covariate factors in a multi-view environment, Results in Engineering 15 (2022), 100556.
[4] T. Sahar, M. Rauf, A. Murtaza, L.A. Khan, H. Ayub, S.M. Jameel, I.U. Ahad, Anomaly Detection in Laser Powder Bed Fusion Using Machine Learning: A Review. Results In Engineering, 2022, 100803.
[5] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A review of video surveillance systems, J. Vis. Commun. Image Represent. 77 (2021), 103116.
[6] Javan Shikhara, M., Levine, M.D., 2013. Online dominant and anomalous behavior detection in videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2611–2618.
[7] Xu, Z., Tsang, I.W., Yang, Y., Ma, Z., Hauptmann, A.G., 2014. Event detection using multi-level relevance labels and multiple features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 97–104.
[8] Ning, J., Zhang, L., Zhang, D., Yu, W., 2013. Joint registration and active contour segmentation for object tracking. IEEE Transactions on Circuits and Systems for Video Technology 23, 1589–1597.
[9] W. Sultani, C. Chen, M. Shah, Real-world anomaly detection in surveillance videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479–6488.
[10] H.W. Kang, H.B. Kang, Prediction of crime occurrence from multi-modal data using deep learning, PLoS One 12 (4) (2017), e0176244.
[11] A. Rummens, W. Hardyns, L. Pauwels, The use of predictive analysis in spatiotemporal crime forecasting: building and testing a model in an urban context, Appl. Geogr. 86 (2017) 255–261.
[12] Xu D, Ricci E, Yan Y, Song J, Sebe N. Learning deep representations of appearance and motion for anomalous event detection. In: BMVC, 2015.
[13] Hasan M, Choi J, Neumann J, Roy-Chowdhury AK, Davis LS. Learning temporal regularity in video sequences. In: CVPR, 2016.
[14] Zhou S, Shen W, Zeng D, Fang M, Wei Y, Zhang Z. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Proc Image Commun. 2016;47:358–68.
[15] Mohammadi S, Perina A, Kiani H, Murino V. Angry crowds: detecting violent events in videos. In: ECCV, 2016.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING