ISSN 2063-5346

UNRAVELING COMPLEXITIES OF ABNORMAL ACTIVITY DETECTION IN REAL-TIME VIDEO STREAMS USING CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Sheriff M [1], Dinesh Kumar T R [2], Sudhakar S [3], Naveesh Kumar D [4] , Manoj K [5] , Murugesan S [6]

Article History: Received: 01.02.2023 Revised: 07.03.2023 Accepted: 10.04.2023

Abstract

Systems for video surveillance are widely used and ubiquitous in various settings. bank, airports, and the prison has seen substantial improvements by video monitoring. Nowadays, corporations, government organizations, and even schools have started to use video surveillance to improve public safety. In this paper, Convolution Neural Network (CNN) algorithms are used in conjunction with advanced deep learning techniques to monitor video data from numerous cameras and identify anomalous activities in real-time by analyzing using UCF-Crime trained datasets. The Smart Video Surveillance system examines behavioral patterns and a number of other factors, such as body posture and movement, in order to spot potential risks and foretell anomalous activity. It notifies specified people in real-time when an unexpected behavior is discovered and records strange activities in a database for further examination. A proactive approach to security is provided by the smart video surveillance system, which foresees potentially dangerous scenarios like (Abuse, Arrest, Burglary, Fighting, Etc.) and takes action to prevent them. Its sophisticated algorithms and real-time alert feature using SMTP server gives an extra layer of security by enabling people to react swiftly to any potential dangers. With the CNN deep learning algorithm, we were able to effectively predict unusual activities for both prerecorded video files and real time detection. By utilizing this technology, public safety and security are tremendously boosted, and crime rates are simultaneously reduced.

Keywords: Recognition, Convolution Neural Network, Video cameras, surveillance systems.

msheriff@velhightech.com [1], trdineshkumar@velhightech.com [2], sudhakar@velhightech.com [3], naveeshkumar.d.2001@gmail.com.[4], k.manoj03012001@gmail.com [5], murugesh15041@gmail.com[6]

[1,2,3] Assistant Professor, Department of Electronics and Communication Engineering

(Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India)

[4,5,6] UG Student, Department of Electronics and Communication Engineering

(Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India)

DOI:10.31838/ecb/2023.12.s1-B.223

1. INTRODUCTION

Security in public areas like airports, retail malls, and streets is increasingly dependent on video surveillance equipment. The detection of unusual events, which may be an indication of possible dangers or criminal activity, is one of the main issues in video surveillance. Activities that drastically depart from the usual patterns of behavior in a certain situation are referred to as abnormal events. Algorithms for abnormal event identification in videos have developed significantly during the past few years. The developments of deep learning methods, computer vision systems, and machine learning have greatly contributed to the field's advancement. With an emphasis on the most recent research and developments, we will evaluate the current state of the art in anomalous event detection in videos. Videosurveillance technology is becoming more and more important for maintaining public safety in places like airports, shopping centers, and streets. One of the biggest problems with video surveillance is the identification of odd events that can be a sign of impending danger or criminal activity. Abnormal events are actions that significantly deviate from the typical patterns of behavior in a specific circumstance. Over the past few years, algorithms for identifying odd events in videos have advanced dramatically. Its growth hasbeen spurred by the development of computer vision, machine learning techniques. This paper aims to review the current status of the art in detecting anomalous events in videos. It also explores the most recent advancements in the field. The majority of video surveillance systems have many cameras that record various angles of the scene. The essential elements of video surveillance systems, such as cameras, image and video processing algorithms, and storage systems, were evaluated by Ovsenk et al [3]. They also talked about how to cope with occlusions, shadows, and shifting lighting conditions when constructing efficient video surveillance systems. A benchmark dataset for abnormal event detection in films, dubbed UCSD Pedestrian [1], was proposed by Wan et al [4]. 34 video sequences from their collection, which includes abnormalities like running, fighting, and stealing, are categorized. They also suggested a computational model for abnormal event identification based on local binary patterns and optical flow. For the UCSD Pedestrian dataset, their methodology produced results that were competitive. A common method for detecting anomalous events in movies is sparserepresentation, which can capture the underlying data structure and spot anomalies by their departure from expected patterns. Sultani et al [5]. suggested a method for real-time abnormal event identification based on sparse representation and dictionary learning. Their system's output on benchmark datasets for aberrant event detection was cutting-edge. Deep learning has significantly impacted the detection of aberrant events in films and has transformed computer vision in recentyears. A thorough analysis of deep learning-based object identification techniques, including Faster R-CNN, YOLO, and SSD, was provided by Zhao et al [8]. Also, they talked about the difficulties and potential possibilities for object detection research. For the purpose of identifying anomalous events in movies, Yeet al [9]. introduced the feature expectation subgraph calibrating classification (FESCC) approach. Their approach models the connections between video frames and extracts discriminative features using a graph-based framework. Moreover, deep learning-based techniques have become a potential way foranomaly detection in video surveillance in recent years. Convolutional neural networks (CNNs) are used in these techniques to extract information from videoframes and categorize them as normal or abnormal. In order to categorize video frames as normal or abnormalbased on their representation in a learnt vocabulary, Sultani et al [5]. suggested a real-time abnormal event

detection system that employs sparse representation and deep neural networks. Several methods, such asfeature-based and graph-based approaches, have been employed for anomaly identification in video in addition to deep learning. Although graph-based methods describe video sequences as graphs and utilizegraph-based algorithms to find anomalies, feature-based approaches leverage handmade features likecolor, texture, and motion to detect abnormalities. Thework of Ye et al [9]., who suggested a feature expectation subgraph calibrating classification method for anomaly detection in video surveillance situations, is an illustration of a graph-based approach. Moreover, it is important to remember that the caliber of the datautilized for training and evaluating anomaly detection systems has a significant impact on how well they work. The UCSD dataset, the Shanghai Tech dataset, and the UCF-Crime dataset are just a few examples of benchmark datasets that have been created to assess the effectiveness of anomaly detection techniques. These datasets offer a common platform for testing various anomaly detection techniques and feature a variety of anomalous events, such as stealing, violence, and loitering. In conclusion, researchers have been paying a lot of attention lately to the essential and difficult topic of anomaly identification in video surveillance. The state of the art is always changing in this area as new methods and algorithms are created to increase the precision and effectiveness of anomaly detection systems. Although deep learning-based methods have produced encouraging results, additional effort has to be done to increase their robustness and generalizability as well as to investigate novel methodologies including feature-based and graph-based methods.

2. RELATED WORK:

This study is the result of extensive research onsupervised learning for the detection of deviant behavior. The creation of behavior recognizers for applications in smart building surveillance has received a variety of contributions. For monitoring and warning purposes, autonomous roaders detect and recognize human activity and behavior in order to detect human behavior. The following are some examples of anomaly types to look for in an object or behavior.

Existing methods for video surveillance using deep learning utilize computer vision techniques and deep learning models to analyze video data. Common approaches include usingconvolutional neural networks (CNNs) for object detection, recurrent neural networks (RNNs) for activity recognition, and deep learning models for abnormal event detection.

These methods have demonstrated good performance, but they also require high computational resources and large amounts of training data.

2.1 Video-based abnormal human behavior recognition

Fig. 2.1.1 Example of differencefrom walking or jogging

- As depicted in Fig.2.1.1, the Dynamic Bayesian Network Model (DBNM) and Hidden Markov Model (HMM) are used to identify suspicious activity
- This method just updates the identification ofunusual human behavior.

2.2 Motion detection, tracking, and classificationfor automated video surveillance

Fig. 2.2.1 Tracking of moving object

By detecting and tracking movements in a video stream, the system can analyze the behavior of individuals and identify potential threats or anomalies as depicted in Fig. 2.2.1, the classification of these movements can further enhance the system's ability to distinguish between normal and abnormal activities, and respond accordingly.

3. PROPOSED SYSTEM

As depicted in Fig.3.1, the system receives video input from security cameras that are strategically positioned to cover the entire area of interest. Preprocessing is used to the input video to reduce noise, improve image quality, and stabilize the video. Using motion detection method, the system detects events, such as the appearance ofpersons or objects in the video stream.

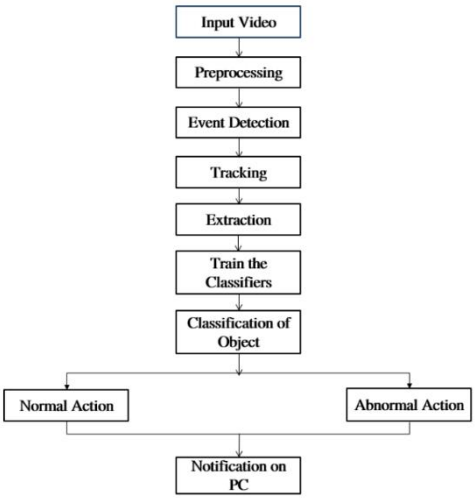

Fig 3.1: The framework for activity recognition for surveillance application

Upon the detection of an event, the system usesobject tracking to track the movement of the item within the video frame. The tracking system takes measurements of the tracked object's size, shape, and motion. A deeplearning- based classifier is trained to discriminate between normal and abnormal activities using the retrieved features. Based on the retrieved features, the trained classifier is used to categorize the behavior of the object as normal or abnormal. The system creates a notification when aberrant behavior is found, which can be relayed to security staff or law enforcement organizations for further action.

4. EXPERIMENTAL METHODOLOGY

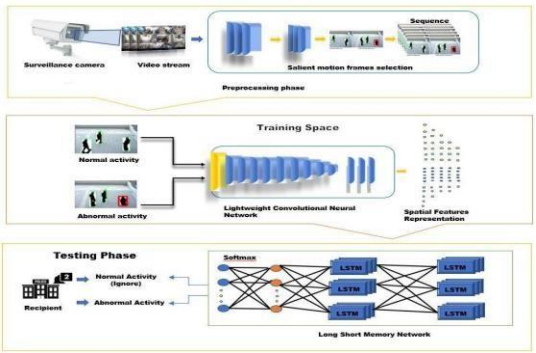

Fig 4.1: Architecture diagram

As depicted in Fig.4.1 the Lightweight Convolutional Neural Networks (CNNs) typically consist of several layers, similar to traditional CNNs. However, the layers are designed to be computationally efficient and require fewer resources. Some common layers used in lightweight CNN architectures include:

Convolutional Layers: These layers are the corebuilding blocks of CNNs and consist of a set of filters that convolve over the input data to produce a set of output feature maps. Inlightweight CNNs, depth-wise separable convolutions which apply a single filter to each input channel before combining the results are frequently used.

$$Z[i,j,k] = (W[:,:,k] * X[i:i+f, j:j+f, :]) + b[k]$$

Pooling Layers: These layers down sample the feature maps produced by the convolutional layers, reducing the spatial resolution of the data. Max pooling is commonly used in lightweight CNNs, which takes the maximum value from each pooling region.

Activation Layers: These layers apply a non-linear activation function to the output of the previous layer. Common activation functions include ReLU (rectified linear unit), which sets negative values to zero, and Swish, which is a smooth function that hasbeen shown to improve performance.

Batch Normalization Layers: In order to increase the stability and convergence of the network during training, these layers normalize the output of thepreceding layer to have a zero mean and unit variance.

Fully Connected Layers: These layers are typicallyused at the end of the network to produce the final output. In lightweight CNNs, Global average pooling, which calculates the average value of each feature map and generates a single output for each channel, may be used to replace these layers.

Overall, the layers used in lightweight CNNs are designed to be computationally efficient and require fewer resources than traditional CNNs, while still achieving high performance on tasks such as image classification and object detection.

As depicted in Fig.4.1 in the lightweight convolutional neural network (CNN), spatial features are extracted through a series of convolutional layers that apply filters (such as the Sobel or Canny filters), Gaussian filters, and high-pass filters to the input image. The filters are designed to detect specific visual patterns, such as edges, corners, and textures,and the output of each filter if the input image or feature map has a size of (224, 224, 3) and a set of filters of size (3, 3, 3, 32), and a stride of 1 and no padding is used, the output feature map would have a size of (222, 222, 32) that highlights the presence of those patterns within the image. The spatial features extracted by the lightweight CNN are then used as inputs todownstream tasks such as object detection, image classification, or abnormal activity detection in video surveillance**.**

A RNN (recurrent neural network) called an LSTM(Long Short-Term Memory) network can successfully simulate long-term associations in sequential data. In the context of video surveillance using CNNs, an LSTM network can be used toanalyze the temporal relationships between the spatial features extracted from video frames. The LSTM network can learn patterns of normal activity and detect abnormal activity by comparing the temporal features of the input video to the learned patterns. This makes LSTMs a powerful tool for activity detection in smart video surveillance systems.

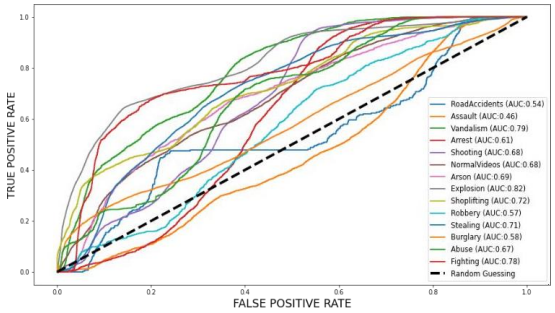

Fig 4.2: AUC Curve Diagram

The AUC curve, which is shown in Fig. 4.2, illustrates the accuracy of the classes (Normal Videos, Road Accidents, Fighting, Shooting, Assault,Abuse, Arson, Burglary, Explosion, Robbery, Shoplifting, Stealing, Vandalism, Arrest). The final ROC AUC value is **0.99**

5. SIMULATION OUTPUT

5.1 Video Processing for abnormal activitydetection:

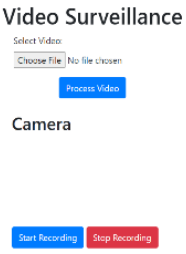

Fig 5.1.1 Web Interface

As shown in fig 5.1.1, The web page has a select file button that, whenclicked, enables visitors to explore and choose a file containing videos of unusual behavior

Fig 5.1.2 File Selection

Users can click the start process button to start the video processing after choosing a file as shown in fig 5.1.2. To find anomalous activity in the chosen video footage, the video processing system makes use of deep learning techniques.

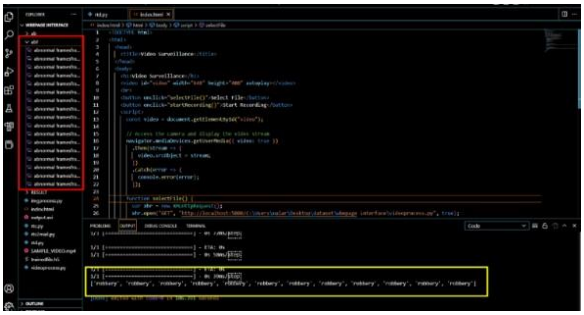

Fig 5.1.3 Prediction Result

In fig 5.1.3 highlighted red area displays the saved frames of anomalous activity in the video, while the highlighted yellow area displays the kind of activity found in the video.

The algorithm stores the suspicious video frames to the chosen folder, which the user can specify, as soon as odd behavior is recognized

5.2 Realtime abnormal activity detection

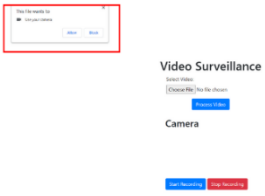

Fig 5.2.1: Access Camera Permission

In fig 5.2.1.A start recording button has been addedto the web page, and when it is clicked, the user is prompted to grant access to camera.

Fig 5.2.2: Activity Prediction

After permission is given, the camera activates and begins to immediately detect anomalous actions using the same deep learning-based algorithm as for video processing as shown in fig 5.2.2.

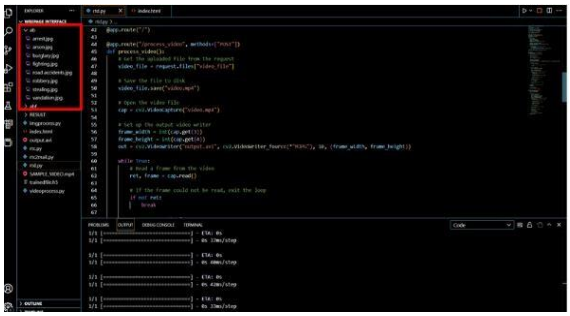

Fig 5.2.3: Saved Abnormal Images

From fig 5.2.3 The system saves the abnormal video frames with the label of the abnormal activity in the designated folder

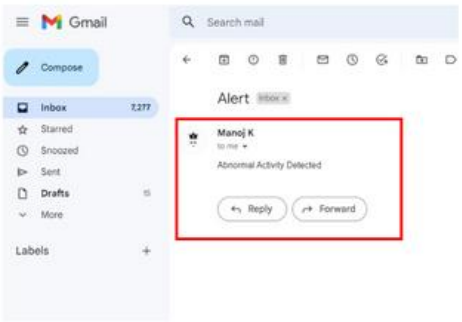

Fig 5.2.4: Alert system

Additionally, the system uses the SMTP server tosend email alerts to security staff or any other person as the fig 5.2.4 shows, the user specifies to inform them in real-time to any detected anomalous activity.

Fig 5.2.5: Process Terminate

Also, the website has a stop recording button as shown in fig 5.2.5 that when pressed, halts the system as well as the recording process.

6. CONCLUSION

In conclusion, Smart Video Surveillance to Predict Unusual Activities Using Deep Learning is a cutting-edge technology that offers a highly effective solution for detecting and analyzing unusual activities in real-time. The system leverages the power of deep learning algorithms to learn the patterns of normal behavior in a given environment and then uses this knowledge to detect deviations from these patterns, flagging them as potential anomalies. The system's ability to save and analyze unusual activities over time makes it an invaluable tool for organizations that need toimprove security, enhance customer experiences, and increase operational efficiency. With its wide range of applications, the Smart Video Surveillance system is poised to play a key part in determining the direction ofsecurity and surveillance in the future. Overall, the Smart Video Surveillance system represents a major stepforward in the development of advanced video analytics technologies, providing organizations with a powerful tool for predicting and preventing unusual activities in real-time.

REFERENCES

1. P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state ofthe art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 743–761, April2012.
2. Anti and B. Ommer, "Video parsing for anomalyidentification," ICCV, 2011.
3. U. Ovsenak, A. Kaimrova Kolesarova, and J. Turan, "System for video surveillance,"Departmentof Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics, Technical University of Koice,Letná 9, 042 00 Koice.
4. B. Wan, W. Jiang, Y. Fang, Z. Luo, and G. Ding, "Benchmark and Computational Model forAnomaly Detection in Video Sequences," Jiangxi University of Finance and Economics, School of Information Management.
5. M. Shah, S. Chen, and M. Sultani, "Sparse representation for real-time anomalous event detection in videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 12, pp. 3395-3405, 2018.
6. P. Zheng, S. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," IEEE, vol. 44,pp. 151-162.
7. Q. Ma, "Deep-learning-based abnormal event detection in videos," Scientific Programming, Hindawi, vol. 2021, Article ID 6412608.
8. X. Chen, J. Chen, H. Hao, and P. Liu, "BMAN: bidirectional multi-scale Aggregation networks for anomalous event detection,"IEEE Transactions on Image Processing, vol. 29, pp. 2395-2408, 2019.
9. O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, "Abnormal event identification using feature expectation subgraph calibrating categorization in video surveillance scenes," IEEE Access, vol. 8, pp. 97564-97575, 2020.
10. Y.-L. Hsueh, W.-N. Lie, and G.-Y. Guo, "Human behaviour recognitionfrom Multiview films," Inf. Sci., vol. 517, pp. 275-296, May 2020.