Suspicious Activity Detection from Surveillance Video using Deep Learning

Shashank Reddy Nallu[1], Vamshi Krishna Kunuru[2], Harshavardhan Reddy[3], Praveen H[4]

[1]Shashank Reddy Nallu, JB Institute of Engineering and Technology

[2]Vamshi Krishna Kunuru, JB Institute of Engineering and Technology

[3]Harshavardhan Reddy, JB Institute of Engineering and Technology

[4]Praveen H, JB Institute of Engineering and Technology

Abstract— **Video surveillance has been used since long to ensure security in many sensitive places, so with this great advancement in various aspects of life, traditional surveillance operations face many challenges due to the large amount of information that has to be processed manually in a limited amount. time also the possibility of losing information that may contain important things such as suspicious behavior. Thus, a large amount of research has been conducted in the field of video surveillance recently.**

We provide a system that supports intelligent monitoring to detect abnormal behavior that poses a security risk. The proposed algorithm is designed to detect two cases of human activity: walking and running. There is no limit to the number of people on stage or the direction of travel. However, video is limited to internal color movies taken from a still camera. A background subtraction algorithm is used to detect moving objects related to people in the scene. We consider the moving speed of the center of the segmented foreground region and the size change speed of the segmented region as two main features to classify the activity v. The proposed algorithm determines the types of activities with high accuracy.

Keywords— **Suspicious activity, Video Surveillance, Deep learning, LSTM, Convolutional Neural Networks.**

I. INTRODUCTION

Detecting human behavior in real-world environments has practical applications in various domains such as intelligent video surveillance and shopping behavior analysis. The use of video surveillance in both indoor and outdoor settings to ensure security has become ubiquitous. In fact, it has become an essential component of people's lives. The Digital India program launched by the Government of India has placed considerable emphasis on e-surveillance in which video surveillance is an important aspect. Video surveillance offers many advantages such as efficient monitoring, low manpower requirements, cost-effective auditing capabilities and the ability to adapt to new security trends. However, due to the large amount of video data involved, manual tracking of events can be tedious and error-prone, which can affect system performance. Automation of video surveillance has helped overcome this problem. It is not possible to manually monitor all incidents captured by CCTV cameras and finding a specific incident in the recorded video footage may take time. Consequently, the analysis of unusual events using automated video surveillance systems has become a crucial area of research in this field.

Automated detection of suspicious activity through recognition of human behavior is becoming increasingly common in video surveillance systems. A number of efficient algorithms have been developed for automatic detection of human behavior at public places such as airports, railway stations, banks, offices and examination halls. Video surveillance is an emerging field of application for artificial intelligence, machine learning and deep learning. Artificial intelligence enables computers to think like humans, while machine learning involves learning from training data and making predictions about future data. With the availability of GPU (Graphics Processing Unit) processors and large datasets, deep learning has become a popular approach.

Deep neural networks are considered to be the most effective architecture for tackling challenging learning tasks. These models can automatically extract features and create high-level representations of image data, making the process of feature extraction fully automated and more general. Convolutional neural networks (CNNs) can learn visual patterns directly from image pixels. For video streams, long short-term memory (LSTM) models are capable of learning long-term dependencies. LSTM networks have a unique ability to remember past events and information.

The proposed system aims to use CCTV camera footage to monitor human behavior on campus and gently alert authorities to any suspicious incidents. The main components of intelligent video monitoring are event detection and human behavior recognition. Automated understanding of human behavior is a complex task. In a campus setting, numerous areas are under video surveillance and a wide range of activities must be monitored. Video footage from the campus has been used to test the system. The process of training a surveillance system can be divided into three stages: data preparation, model training, and prediction. The system framework consists of two neural networks – a convolutional neural network (CNN) and a recurrent neural network (RNN). CNN extracts high-level features from images to reduce input complexity, while RNN is ideal for processing video streams and performing classification tasks. The proposed system employs a pre-trained VGG-16 (Visual Geometry Group) model, which is trained on the Imagenet dataset. Currently, models are being trained to predict behavior from video footage, enabling it to identify suspicious or normal human behavior and assist in the monitoring process.

II. LITERATURE SURVEY

Various methods have been proposed in the related literature to recognize human behavior in video footage. The primary objective of these approaches is to detect any unusual or suspicious events in video surveillance.

Advanced motion detection (AMD) algorithm was used to detect unauthorized entry into restricted area [1]. The system consists of two main phases. In the first stage, the object was detected by background subtraction and then extracted from the sequence of frames. The second phase focused on detecting suspicious activities. One of the main advantages of this system was the ability to process video in real-time with low computational complexity. However, the system had certain limitations, such as limited storage capacity and the need for high-tech video capture modes in surveillance areas.

A semantic based approach was proposed in [2]. The captured video data was processed to identify foreground objects by background subtraction. Objects were then classified into living and non-living entities using a Her-like algorithm. Object tracking was accomplished using a real-time blob matching algorithm, and fire detection was also a feature of the system described in the paper.

Based on the motion features between the object, suspicious activities were detected in [3]. To identify abnormal events occurring in a university setting, the area was divided into zones and optical flow was estimated in each zone via the Lucas-Kanade method. A histogram of the magnitude of optical flow vectors was subsequently created. Software algorithms were utilized to analyse video content and classify events as either normal or abnormal [4].

A system was created to distinguish between abnormal and normal events by analyzing movement data in video sequences. The system used the HMM method to learn histograms of optical flow orientation in video frames. It compares captured video frames with existing normal frames to identify similarities between them. The system was evaluated and validated using various datasets including the UMN dataset and PETS [5].

One approach to detecting unusual events in video footage involves tracking people. This is accomplished by first detecting humans in the video using a background subtraction method. The features are then extracted by CNN and input to DDBN (Discriminatory Deep Belief Network). DDBN is also fed labeled videos of suspicious events, from which features are also extracted. Features extracted from sample videos labeled with CNN and classified suspicious actions using DDBN are compared. Through this method, various suspicious activities can be detected in a given video [6].

A real-time system for detecting violent behavior using deep learning was developed to prevent crowd or player-to-player violence at sporting events. The system extracts frames from real-time video in the Spark environment, and if it detects any violence in football, it sends an alert to security personnel. By detecting video actions in real-time, the system aims to prevent violence before it happens by alerting security forces. The system was evaluated on the VID dataset and achieved a 94.5% accuracy rate in detecting violence in football stadiums [7].

Anomaly detection consists of several modules to process video data, using a deep architecture to recognize human behavior. The CNN and LSTM-based models proposed in this system were tested on the UT interaction dataset. However, one limitation of the system is its inability to differentiate between similar human behaviors, such as pointing and punching [8].

III. SYSTEM ARCHITECTURE

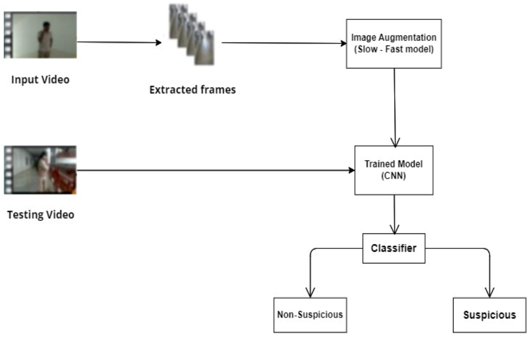

Fig.1 System Architecture

The proposed system uses footage obtained from CCTV cameras to monitor student activities on campus. When a suspicious incident occurs, the system sends a message to the appropriate authority.

The system architecture consists of several stages including video capture, video pre-processing, feature extraction, classification and prediction. The layout of the system architecture is shown in figure-1.

IV. METHODOLOGY

A. Video Capture

The first step in implementing a video surveillance system is to install CCTV cameras and monitor the footage they capture. Different types of videos are captured from different cameras covering the entire surveillance area. In our implementation, processing is carried out using frames; Thus, the video is converted into frames.

B. Dataset Description

The KTH dataset contains a standard collection of sequences representing six actions, with each action class containing 100 sequences. Each sequence consists of approximately 600 frames, and the video is shot at 25 fps. The model is trained on this dataset for common behaviors, such as walking and running. For training on suspicious behavior, such as mobile phone use inside the campus, fighting and fainting, the CAVIAR dataset, videos taken from the campus and YouTube videos are used. About 7035 frames are extracted from different videos, and the entire dataset is manually labeled and separated into 80% training set and 20% validation set. The directory structure of the dataset is shown in Fig.2. Our system uses a combination of KTH, CAVIAR, YouTube videos and videos captured from campus.

C. Video Pre-Processing

Our proposed system uses a deep learning network to detect suspicious activities in video surveillance. Using a deep learning architecture can significantly increase the accuracy of the system, especially when working with large datasets. A broad design overview of the system is presented in Figure 2.

The input videos for the proposed system are obtained from existing and created datasets. The pre-processing stage involves extracting frames from the captured video. Three labeled folders are created based on the videos, and the frames are stored in them. Captured video is converted to 7035 frames, which are then stored as images in jpg format. Each frame is resized to 224 × 224 corresponding to the 2D CNN architecture and then stored. The testing video is also converted to frames, resized to 224 × 224, and stored in a folder. OpenCV library is used for video pre-processing in Python.

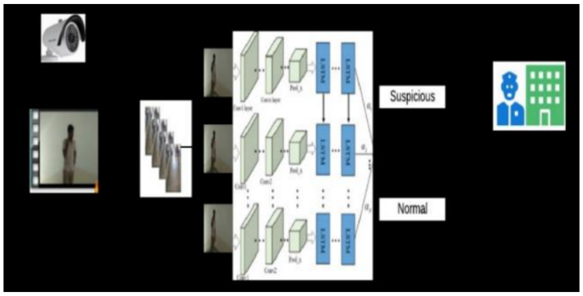

Fig.2 System Overview

For image feature extraction, we used the VGG-16 CNN model pre-trained on the ImageNet dataset. The VGG-16 architecture is shown in Fig.4, and consists of convolutional layers with a size of 3×3, maximum pooling layers with a size of 2×2, and finally fully connected layers, for a total of 16. Layers in a Deep Learning Architecture. The input image should be in RGB format with a size of 224×224×3. The model has different layer representations, including convolution layers, activation functions using ReLU, maximal pooling layers, fully connected dense layers, and normalization layers. We can fine-tune the model to our needs by removing the last layer and then training the model on the LSTM architecture. An LSTM network is a type of RNN that can learn order dependencies in inference prediction problems. Our LSTM model consists of ReLU activation functions, dropout layers, and fully connected dense layers, with the number of neurons in the last layer equal to the number of classes, which in our case is three.

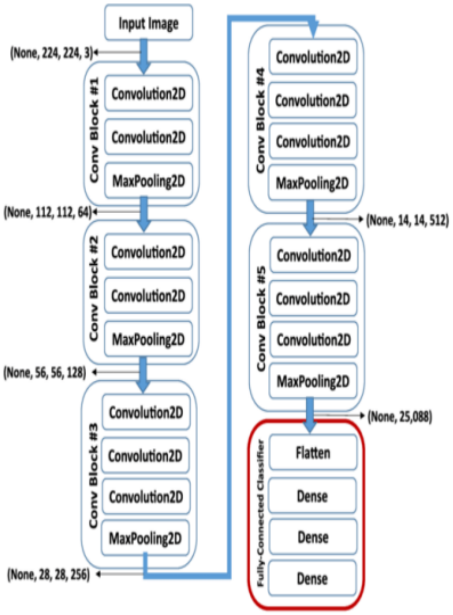

Fig.3 VGG-16 Architecture

A video surveillance system classifies captured videos as suspicious (e.g., students fighting, fainting) or normal (e.g., walking, running). If suspicious behavior is detected, the system sends an alert to the concerned authority.

V. CONCLUSION

In today's world, almost everyone recognizes the importance of CCTV footage. However, these recordings are mainly used for post-incident investigations. The proposed model offers the advantage of preventing crimes before they occur. It involves real-time monitoring and analysis of CCTV footage, which generates orders for authorities to take action if the analysis indicates an impending incident. As a result, crimes can be prevented. While the proposed system is currently limited to educational environments, it could also be used to identify suspicious behavior in public or private areas. The model can be applied to any situation by training it to identify suspicious activities related to that scenario. The effectiveness of the model can be increased by identifying suspicious individuals from suspicious activity.

VI. FUTURE ENHANCEMENT

The model should be trained against a wide range of suspicious activities to enhance security personnel's ability to identify and respond to different types of incidents. Furthermore, to ensure that the model can handle large amounts of data without any loss of information, it must be designed to handle large data sets efficiently. Finally, the model should be modified to identify suspicious activity from blurred video using various techniques and techniques to enhance video quality and extract relevant features. These enhancements will significantly improve the effectiveness of the model and enable it to detect a wider range of suspicious activities in surveillance videos.

VII. REFERENCES

[1] Bhagya Divya, Shalini, Deepa, Baddiel, Sravya Reddy, “Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras”, International Research Journal of Engineering and Technology (IRJET), December 2017.

[2] Jitendra Musale, Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, “Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed-Circuit TV (CCTV) camera”, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII December 2017.

[3] U.M. Kamthe, C.G. Patil “Suspicious Activity Recognition in Video Surveillance System”, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

[4] Zahra Kain, Abir Youness, Ismail El Sayyad, Samin Abdul-Nabi, Hussein Kassem, “Detecting Abnormal Events in University Areas”, International conference on Computer and Application,2018.

[5] Tian Wanga, Meina Qia, Yingjun Deng, Yi Zhouc, Huan Wangd, Qi Lyua, Hichem Snoussie, “Abnormal event detection based on analysis of movement information of video sequence”, Article-Opik, vol152, January-2018.

[6] Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, “Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Network”, International Journal of Control Theory and Applications Volume 10, Number 29 -2017.

[7] Dinesh Jackson Samuel R, Fenil E, Gunasekaran Manoharan, Vivekananda G.N, Thanjaivadivel T, Jeeva S, Ahilan A, “Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM”, The International Journal of Computer and Telecommunications Networking,2019.

[8] Kwang-Eun Ko, Kwee-Bo Sim “Deep convolutional framework for abnormal behavior detection in a smart surveillance system. “Engineering Applications of Artificial Intelligence ,67 (2018).