Suspecting Abnormal Activities Under Survilleance

Alugubelly Varun Reddy

Computer Science and Engineering Vardhaman College of Engineering
Hyderabad,India

Dr Reddy Saisindhutheja

Thudi Arshith Reddy
Computer Science and Engineering Vardhaman College of Engineering
Hyderabad,India

Yarlagadda Pavan

Computer Science and Engineering Vardhaman College of Engineering
Hyderabad,India
Computer Science and Engineering (Associate Professor) Vardhaman College Of Engineering
Hyderabad,India

***Abstract*—The global crime rate has been increasing in terms of numbers, with an estimated hundred million offences committed each year. Surveillance cameras have been installed at every street corner to capture and record such activities. Detecting such acts in time might reduce a person's potential danger. It is difficult for human supervision of such systems to discover such behaviours on the spot for every surveillance CCTV video. In order to discover such actions as soon as possible, this study proposes an intelligence surveillance system that recognises criminal scenes and shows crime activity for a recorded surveillance video using a deep learning approach. As a result, the suggested systems employ deep learning techniques CNN and LSTM to categorise and detect odd actions for recorded data.The Convolution neural network is used to learn and train the data.**

I. INTRODUCTION

Through the human surveillance, the activities can be monitored in busy places and malls around the city in order to spot or detect abnormal activities such as Rob- bery,Shoplifting,Vandalism,Abuse,Explosions,Burglaries and other such unusual activities.Since it is exceedingly challeng- ing to continuously monitor active areas. Intelligent surveil- lance systems that can monitor people's movements and cate- gorise them as routine or unusual are required.[1]

To track the basic daily actions carried out by a person, an automated activity system is needed. High-rate accuracy for the recognition of this wide variety of human actions is difficult to attain. Development of the activity models needed for the recognition and classification[2] of human activities is based on several methods unique to the application. The activity can generally be classified into either normal or abnormal ones. Anomaly activity is defined as a as an activity that is irrelevant to day to day activities usually performed by a person that could make him possibly to be at risk. One of the most popular techniques for achieving anomaly detection is using films of typical events as training data to develop a model, followed by the detection of suspicious events.In these paper we proposed a model that uses Deep Learning techniques on the public dataset that has been provided by Kaggle.[3]

II. LITERATURE SURVEY

A study titled "Real-world Anomaly Detection in Surveil- lance Videos" [4] suggests a method for detecting anomalies that involves using both normal and anomalous videos. It uses the deep multiple instances ranking framework and trains the model using weakly labeled training videos where the anomalous or normal labels are assigned at the video level rather than the clip level to save time. This method employs Multiple Instance Learning (MIL) to create a deep anomaly ranking model that assigns high anomaly scores to anomalous video segments and treats normal and anomalous videos as "bags" and video segments as "instances." The ranking loss function is also restricted by sparsity and temporal smoothness constraints to enhance anomaly localization during training. The study also analyzed the results of several recent deep learning techniques for anomalous activity detection and found that their recognition performance was low, indicating the difficulty of the dataset.

A study titled "An Efficient Dimension Reduction based Fusion of CNN and SVM Model for Detection of Abnormal Incident in Video Surveillance" [5] outlines a method for efficiently detecting abnormal incidents in video surveillance. The main advantage of this model is its ability to effectively search for unusual events within a video stream. The pre- processing step of the model is key, as it significantly reduces the complexity of the video stream, making it easier to process. The proposed deep learning architecture reduces time complexity and is designed for continuous anomaly detection. During pre-processing, the dimensions of the collected video frames are reduced, preserving important and anomalous oc- currences in the frame. This model employs a combination of Convolutional Neural Network (CNN) and Support Vector

Machine (SVM) to identify unusual activities, treating it as an image classification problem.

Anomaly Recognition from Surveillance Videos using 3D Convolution Neural Network[6] This approach starts by break- ing each video down into a predetermined number of frames. For model training, 3D cubes are initially generated by pre-processing all of the frames. Combining the constant length of the model is given a set of consecutive frames as input. Our honed model receives original frames as well as spatially enhanced frames. In order to preserve channel information, the frames are not made grayscale. Different amounts of convolution, pooling, batch normalisation (BN), and fully Connected layers(FC) make up the suggested fine- tuned model. Each video clip's spatiotemporal properties are extracted using 3D cubes using 3D ConvNets, and a fully connected neural network is trained by utilising the multiclass classifier SoftMax, which determines the likelihood between the highest scored occurrences. in these model testing and training are both semi-supervised, meaning that only frame- level labels are visible during training and assessment. The output with the highest likelihood of the anticipated class is provided during the classification step.

III. PROPOSED METHODOLOGY

A. Image processing

Image processing is a method that processes the input data to train a model for the intended output while extracting features. image processing consists of 1)Segmentation: When an input is given to the model, it filters out extraneous background information or distracting data that are useless for the model's training.2)Feature Extraction[7]: In this case, the model takes the required features from the input signals and classifies the data based on them.3)Classification of Images: CNN is used to classify the images.We need to give the model some further information in order to categorise data, such as the number of distinct images that need to be categorised and the number of levels per pixel. The model automatically categorises the data after data assignment and generates the desired results.

B. Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a powerful class of neural networks that have proven to be capable of both region and image categorization and identification. It has been confirmed that Artificial Intelligence (AI) can recognize symptoms of faces, devices, and visitors more accurately than humans. As a result, they are trained to do so through su- pervised Convolutional Neural Networks (CNNs). This means that self-driving cars and robots can be taught to identify objects and labels through the CNN analyzing the relationships between them[8]. A hidden layer is used to extract properties, and a fully connected layer is used for processing to complete the classification task. Unlike traditional neural networks, the hidden layers of Convolutional Neural Networks (CNNs) possess a special shape of features. Each layer of a neural network typically consists of nodes that are linked to all

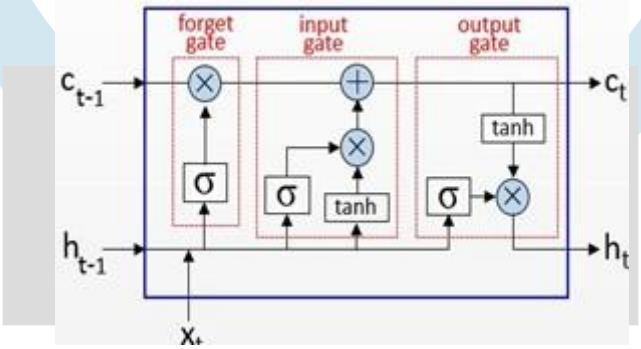

Fig. 1. LSTM ARCHITECTURE

$$f_t = \sigma (W_f x_t + U_f h_{t-1} + b_f)$$
$$c_t = c_{t-1} * f_t$$

$$i_t = \sigma (W_i x_t + U_i h_{t-1} + b_i)$$
$$a_t = \tanh (W_a x_t + U_a h_{t-1} + b_a)$$
$$c_t = c_{t-1} * f_t + i_t * a_t$$

Fig. 2. Input layer and Forget layer equations

layers above it. The neural network's nodes act as independent layers. A Convolutional Neural Network (CNN) has layers that are nearly invisible. Only a small portion of the neurons in each layer are linked to the neurons in the layer above them. Moreover, an additional pooling layer combining outputs from neighboring neurons into one value gives us a feature that is translation-

invariant. This simplifies training and decreases model complexity.

C. Long Short Term Memory

Long-Short Memory Network is a recurrent network that use gates to implement the concept of memory units[9]. Forget gate, input gate, and output gate are the three layers that make up the LSTM in order to process information. The memory units are filtered by the forget gate using the Sigmoid activation function or activation units.Input layer uses tangent activation units to remove undesirable information units that have been processed through forget gate.The last layer, the output layer, employs sigmoid function for cell input and tangent function for the current cell state and evaluates the output unit for the provided information.

D. Dataset

The dataset used was a public dataset collected from kag- gle[3].Videos from the 13 categories of abuse, arrest, arson, assault, burglary, explosion, fighting, robbery, shooting, steal- ing, shoplifting, and vandalism are included in this dataset. Depending on what is shown in each video, it is rated as either normal (zero) or aberrant (one).Finally, the dataset's

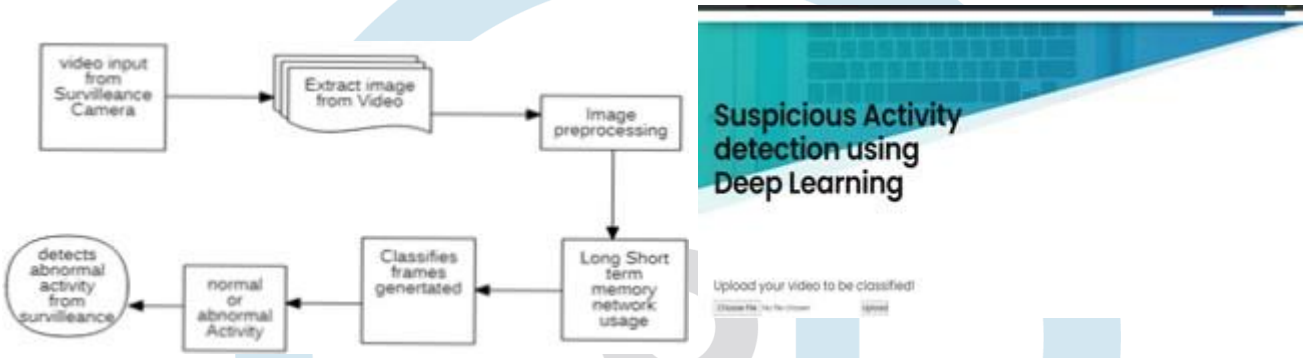

Fig. 3. System Archiecture

distribution is as follows:There are 16853 videos in all, 9676 of which are classified as normal and 7177 as abnormal.

E. System Architecture

A surveillance camera's recorded input is first sent to the suggested model, which is then trained to extract frames from the input. To resize all the frames to the same size, the incoming video is split into 60 frames, each 244*244 pixels in size. The frames are then normalised to convert the data to an o to 1 scale before being passed through an LSTM and CNN model to determine whether the input video contains normal or abnormal behaviour.

IV. EXPERIMENTAL RESULTS

It can be difficult to find suspicious behaviour in video data. There are many challenges, including the complexity of the scene, the lighting, the camera angle, and others.The model was trained using 16,853 recordings, of which 7177 were anomalous and 9676 were of normal behaviours.Extraction of frames from a particular video input constitutes the method's first phase. The analysis of the frames to determine whether frame has suspicious or aberrant activity is the primary task of the following step. The system will be created by learning the patterns of photographs and images that include questionable activity. One of the most effective structures employed is convolution neural networks, a subclass of neural networks. Figure 4 shows the user interface for uploading video clips.

Figures 5, 6, and 7 then display the outcomes for the submitted video via the interface.

CONCLUSION

Even though practically everyone in today's society is aware of the importance of CCTV footage, it is typically only used for criminal investigations after an incident or crime has occurred. The benefit of the suggested approach is that it can inform users when an accident occurs. Monitoring and analysis of CCTV footage are being done in real-time. The study's conclusion issues a directive to the relevant authorities asking them to decide whether the findings suggest that an undesirable event is likely to happen. As a result, this can

Fig. 4. System Archiecture

Fig. 5. Suspicious activity detection(1)

Fig. 6. Suspicious activity detection(2)

Fig. 7. Suspicious activity detection(3)

be prevented. Although the suggested approach is targeted toward the educational sector, it can be applied to predict more unusual behaviours in both private and public settings. Any situation where instruction is required in connection with an unusual action.

FUTURE SCOPE

The suggested model is functional for the dataset with 14 different suspicious actions. The collection includes video clips with a run time of under a minute. Since the model is simply taught to identify unusual actions for recorded clips, , it may be significantly improved to easily determine such events in future for live detection or live automated monitoring.The model can be further developed with the support of the cyber security team for developing an application such that the relatives of victims of suspicious activity can be informed with a message via bot indicating suspicious sign for the victims who are taking part in such activities.

REFERENCES

1. M. Valera and S.A. Velastin , "INTELLIGENT DISTRIBUTED SURVEILLANCE SYSTEMS," IEE Proc.-Vis. Image Signal Process., Vol. 152, No. 2, April 2005.
2. Neha sharma ,Anju mishra and Vibhor jain , "An Analysis Of Convo- lutional Neural Networks For Image Classification",Neha Sharma et al.
3. / Procedia Computer Science 132 (2018) 377–384. [3]https://www.kaggle.com/datasets/mateohervas/dcsass-dataset
4. Waqas Sultani and Mubarak Shah, "Real-world Anomaly Detection in Surveillance Videos," " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479–6488.
5. Dr. Rajesh Sharma and Dr. Akey Sungheetha, "An Efficient Dimension Reduction based Fusion of CNN and SVM Model for Detection of Abnormal Incident in Video Surveillance," ISSN: 2582-2640 (online) Journal of Soft Computing Paradigm (JSCP) (2021) Vol.03/ No.02.
6. Ramna Maqsood, . Usama Ijaz Bajwa,Gulsha Saleem and Muhammad Waqas Anwar, "Anomaly Recognition from Surveillance Videos using 3D Convolution Neural Network"
7. Theodore Bluche and Christopher Kermorvant, "Feature extraction with convolutional neural networks for handwritten word recognition", 1520- 5363/13 2013 IEEE DOI 10.1109/ICDAR.2013.64.
8. Syed OwaisAli Chishti; Sana Riaz; Muhammad BilalZaib and Moham- mad Nauman, "Self-Driving Cars Using CNN and Q-Learning", DOI 10.1109/INMIC.2018.8595684,30 December 2018.
9. Yong Yu,Changhua Hu,Jianxun Zhang and Xiaosheng Si , "A Review of Recurrent Neural Networks: LSTM Cells and Net- work Architectures",Neural Computation 31, 1235–1270 (2019) doi:10.1162/necoa01199.