Detection of Malfeasance through Intellectual Eye

Mylavarapu Kavya
Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
kavyamylavarapu2306@gmail.com

P.Aditya sasank
Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
adityasasank53@gmail.com

S.Vijay Simha Reddy
Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
svreddy873@gmail.

Dr.S.Venu Gopal
Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad, India
s.venugopal@vardhaman.org

***Abstract*—Observation security is extremely monotonous and tedious work. It requires a labor force and their consistent consideration to decide if the caught exercises are uncommon or dubious. In this, we will construct a framework to mechanize the undertaking of examining video reconnaissance. We will investigate the video feed continuously and distinguish any strange exercises like surprising or dubious. From previous encounters, there have been progressions in profound learning calculations for profound reconnaissance. These progressions have shown a fundamental pattern in profound reconnaissance and guarantee a radical effectiveness gain. The normal utilization of profound observation is burglary distinguishing proof, brutality discovery, and recognition of the possibilities of explosion. For this venture, we will present a spatio worldly auto-encoder, which depends on a 3D convolution brain organization. The encoder part removes the spatial and transient data, and afterward, the decoder reproduces the edges. The unusual occasions are distinguished by registering the recreation misfortune utilizing Euclidean distance among the unique and remade batches.**

***Index Terms*—Malfeasance, Surveillance, ImageAI, reconnaissance.**

I. Introduction

Surveillance cameras have become very common in public places to increase public safety. A large number of videos are created and stored for a period of time, and constantly keeping track of these surveillance videos is practically hard for authorities to evaluate whether the incidents are suspicious since it requires a large crew and continual monitoring. To make monitoring easy, automation of surveillance is required. Smart surveillance helps to increase safety in public places by automatically detecting crimes. This paper aims to detect and classify levels of high movement in the frame using various Deep Learning models. Videos are divided into segments in this project. In the event of a threat, it identifies the time duration of the threat [**?**], and the detection alert is raised, indicating the suspicious activities at a specific point in time [**?**]. The videos in this project are divided into two categories: threat (abnormal activities) and safe (normal activities) [**?**].

Deep learning techniques are used to solve existing problems, resulting in phenomenal results in the detection and categorization of activities. Two different neural networks were used in this study: CNN and RNN. CNN is a basic neural network that is primarily used to extract advanced feature maps from recorded data. The complexity of the input is reduced by extracting high-level feature maps. A pre-trained model is chosen because modern object recognition models take into account a large number of parameters and thus require a significantly more amount of time to fully train. The approach of transfer learning would improve this task by first considering the previously learned model for a set of classified inputs, such as ImageNet, which can then be re-trained using new weights assigned to various new classes. The RNN receives the output of the CNN as input. The RNN also has the ability to predict the next item in a sequence. As a result, it primarily serves as a forecasting engine. The motivation for using this neural network in this work is to provide meaning to the captured sequence of actions/movements in the recordings. The primary layer of this network contains an LSTM cell, which is followed by some hidden layers with appropriate activation functions, and the output layer provides the final classification of the video into 13 groups (12 anomalies and 1 normal). This system's output is used to perform real-time surveillance of various organizations' CCTV cameras in order to avoid and detect any suspicious activity.

The data mining method was used to analyze the crime. For preventing crime, object tracking is important if the interloper handled any weapons, the video surveillance camera is automatically detected. By using this technique, the prevention of crime is the easiest task. Machine learning and deep learning methods are used to identify crimes. Detection of an object is also a challenging task, and it plays a major role in crime identification. In object detection, there are three stages; they are differentiating the frame, optical flow, and background subtraction. The background subtraction is used to eliminate the moving forepart from the original surrounding. Facial utterance identification is the last process in crime detection, and it is used to identify a criminal face from videos or images. It takes more time to analyze the criminal, and it gives accurate detection of crime. Face detection is done by matching the human face either in a crowd or alone. The proposed method

identified the crime by using video surveillance. This paper will explain the various literature reviews of the researchers, and the survey was based on the video surveillance used for crime identification. It explains all the information about crime detection and the difficulties faced by the law of organizational agencies.

II. Related work

To maintain security at public places there are some existing methods using different methodologies.

A. Feature Selection Using Optimization Mechanism

Clustering is a kind of structure training and is an exploratory data exploratory statics finding process.The Clustering Technique was helped to grouping the data into clusters by using different techniques. It plays a major role in data mining and analysis. Rasoul Kiani presented a paper based on the clustered crime that happened during various years. For improving, outlier detection Genetic Algorithm was used by the Rapid Miner tool. As a result, to determine the effect and quality, the maximized and non maximized parameters were matched. In the past few decades, the spotting and hampering of crime required years of research and inspection. The most widely used K-mode algorithm is a failure clustering method. To address this issue, the fusion of K-modes and Elephant Herding Optimization was developed by Farhad. As a result, the proposed model shows purity inaccuracy.The comparison chart of the proposed method and K-Mode algorithm based on their precision and purity. Nowadays, in the field of crime detection, various researches are going on, and there is no advanced technology till now. To control the crimes, CCTV's are commonly used in the surroundings, but still, there is nothing improved in controlling crimes. To address this issue, Umadevi V proposed the intension Of crime detecting program. It detects the crime in cameras and alerts the organizer to take perspective action. The already trained model VGGNet-19 was used in this proposed system to detect the crime intention. An algorithm used to draw the square box over the suspecting images is Fast Regional based Convolutional Neural Network, and it is well known as Faster RCNN.

B. Crime Detection Using Facial Expression

A Facial recognition system is used for matching a human face from images, videos and it is the classification of biostatics security. Human faces contain variability in size, color, etc. In recent years, the extension in face data collection origination has been immense. To address this, C. Anitha presented a broad study of obtainable data collection used for facial utterance identification systems. Since the 1990s, Automatic Facial Utterance Identification is the analysis topic, and there are some advances in identifying face and expressions. For identifying the facial utterance, several methods were used. C. P. Sumathi present a paper about the facial specifications using the Facial Activity Model. This presented paperwork aims to manifest a clean study of identifying facial utterances. The contrary facial emotions are demonstrated in Syeda Amna Rizwan present an advance toward identifying facial utterances depend on multi landmark identifiers and local modified characteristics. The proposed method was classified into four, and the advanced facial utterance identification performs well, and the accuracy is good. As a result, it is relevant to various consumer significant arenas, and the face is identified by using YCbCr color space. Converting photos with geo-temporal documents into underlying data on spirit and then applying it for crime estimation. This method was proposed by Tiafan Zhangfor identifying facial expressions. As a result, it upgrades the operation of crime divination. Facial utterance identification plays a vital role in crime detection, and many types of research were done for this, but still, now the identification of crime is difficult in this system. To overcome this, Ashraf Abbas presented a method to identify the crime by providing facial utterances in surveillance, and it detects the person before they commit any restricted works. The result shows that it is more effective and useful for guards to seize criminals. It shows the facial identification of criminals by matching their faces and notices the person's movement in the crowd. By watching the criminal facial utterance, the prevention steps were taken to stop the crime. Moreover, it is an effective method for preventing criminal activities.

III. Methodology

The proposed model highlights a detailed specification that is used to detect suspicious activity. Archives of crime rate are increasing quickly. As a human, it is very hard to keep an eye on every place on earth for preventing these criminal activities. Hence, we tend to square measure proposing our model wherever the formula is trained for detecting Suspicious activity by deep learning technique. Pre-trained deep convolution neural network, Spatiotemporal Auto Encoders are used for initial classification and a recurrent neural network is used for the final detection of suspicious activity and is performing [1].

Firstly, a live video feed is given to the system, which is obtained from CCTV. The video is then converted into frames with a fixed and small interval of time (say 1 frame per second) [6]. These frames are passed to spatiotemporal auto-encoder encoder [1] [6], which is based on a 3D convolution network [5]. The encoder part extracts the spatial and temporal information, and then the decoder reconstructs the frames [4]. The abnormal events are identified by computing the reconstruction loss using Euclidean distance between the original and reconstructed batch.

The collection of these frames is used to classify the live CCTV feed [6]. The singly merged feature map is given as input to 3D-CNN [5]. In this methodology, we established an LSTM cell so that the training time becomes small. This 3D-CNN is trained with the UCF-Crime dataset. The UCF-Crime dataset is taken from Kaggle. The UCF-Crime dataset consists of 1900 clips each of sixty to six hundred seconds in length with variable resolution and is recorded with surveillance cameras that operate in the real world. This dataset is intended to detect 13 genuine abnormalities, such as cruelty, imprisonment, fires, attack, crash, robbery, eruption, combat,

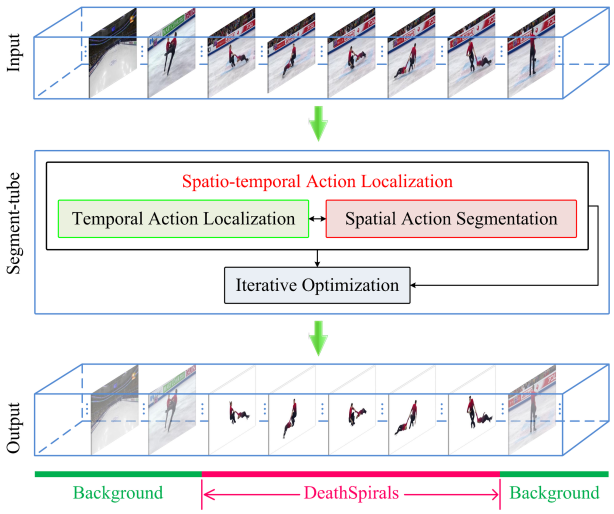

Fig. 1. Flowchart of Spatiotemporal autoencoder

theft, killing, theft, snatching, and vandalism. The last Softmax layer is used to determine the probabilistic classification. Using the trained model any event id diverging from the trained model are regarded as anomalous events [2].

A. Algorithm

1) Algorithm to generate frames::

1) void generateFrames()
2) if path not exist
3) make directory
4) else
5) remove directory
6) create new directory
7) //create a video object
8) initialize variables count and success to 0 and 1
9) while success
10) //take the values of success , count from the video object
11) if count lessthan 500
12) save the image
13) print ”image name”
14) else
15) end loop
16) increment the count

2) Algorithm to detect suspicious frames::

1) void detectActivity()
2) initialize variables count and success to 0 and 1
3) for images in the path:
4) predictions,probabilies = predictImage(images)
5) for each prediction,each probability
6) if eachprobability graterthan 80
7) increment the count
8) else if each probability lessthan 80:
9) set count 0
10) if count graterthan 10:
11) set option to 1
12) print ”The images names and its probabilities”

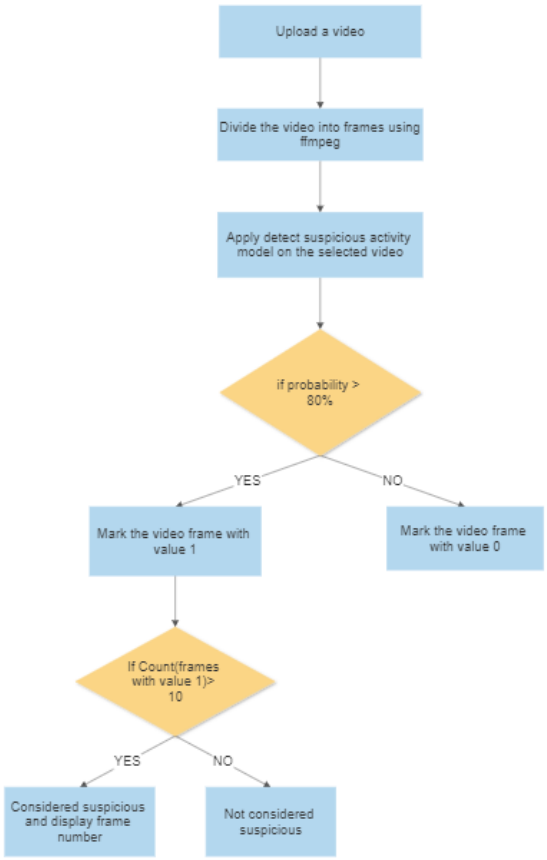

Fig. 2. Flowchart of the process

B. Video into frames

Using python the entire video is segmented into frames, which these frames are then given for Inception V3. Because videos may be thought of as a collection of individual pictures, many deep learning experts will handle video classification as if it were a series of image classifications performed N times in total, where N is the overall number of frames in the film [6]. We may often assume that consecutive frames in a video are associated with respect to their semantic contents using video classification, which is more than simply simple picture classification. We were able to enhance our real video categorization results by taking use of the temporal characters of videos.

C. Spatiotemporal AutoEncoder-Decoder

Our design comprises a transient autoencoder settled into a spatial autoencoder. At each time step, the organization takes as information a video outline Yt of size H × W, and produces a result of a similar size, addressing the anticipated next outline, Y˜t+1. In the accompanying, we portray every one of the modules exhaustively [4].

The spatial autoencoder is an exemplary convolutional encoder-decoder engineer. The encoder E contains one convolutional layer, trailed by tanh non-linearity and a spatial max-pooling with a subsampling layer. The decoder D mirrors

the encoder, aside from the non-linearity layer, and uses the closest neighbor spatial upsampling to take the result back to the size of the first information. After the forward go through the spatial encoder YtE→xt, the size of the component maps xt is d×h×w, d is the number of highlights, and h furthermore, w the level and width subsequent to downsampling, individually.

It is the lower layered secret layer where the encoding is created. The bottleneck layer has a lower number of hubs and the quantity of hubs in the bottleneck layer likewise gives the element of the encoding of the info.

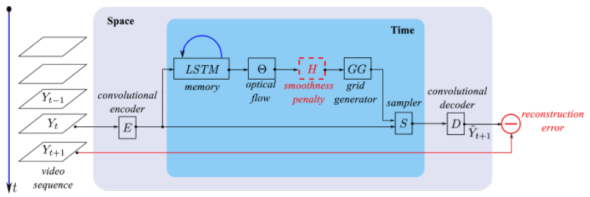

Fig. 3. Spatiotemporal video autoencoder

The transient autoencoder's goal is to identify significant changes brought on by movement (either internal self-movement or the development of the objects in the scene), allowing it to predict the visual future while being aware of the past and present. In an exemplary spatial autoencoder Masci (2011), the encoder and decoder learn exclusive component spaces that permit an ideal disintegration of the information utilizing a type of regularization to forestall learning trifling planning. The encoder chooses uninhibitedly upon deterioration in view of its ongoing element space, and the decoder obliges the learning of its own component space to fulfill this decay and to recreate the information, utilizing for the most part activities basically the same as the encoder, and having a similar number of levels of opportunity. Uniquely in contrast to this, the proposed fleeting autoencoder has a decoder with few teachable boundaries, whose job is fundamental to give prompt input to the encoder, yet without the limit of changing the encoder's missteps like in the spatial case. In improvement terms, the mistake during learning is credited primarily to the encoder, which is presently more obliged to deliver reasonable element maps.

IV. Result

A. Observation of results

Videos	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	sum
V1	0	0	0	0	0	0	0	0	0	0	0
V2	0	13	12	3	0	0	10	2	0	0	40
V3	0	0	0	0	0	0	7	3	4	1	15
V4	0	0	0	0	0	0	0	0	0	0	0
V5	0	0	0	0	4	6	3	0	0	0	13

TABLE I
COUNT OF SUSPICIOUS ACTIVITIES

For this project we taken 5 sample videos whose results are mentioned in the table mentioned above.It contains the count of frames that are found to be suspicious.To consider

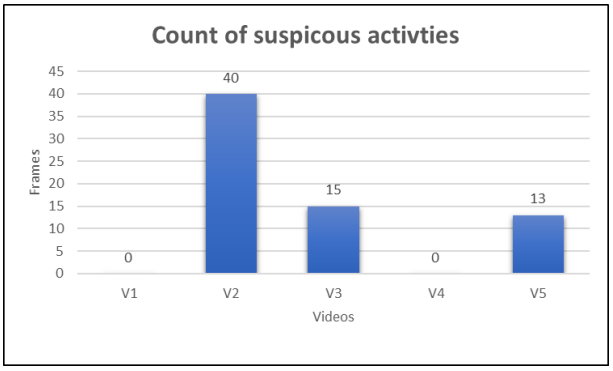

Fig. 4. Bar graph for number of suspicious frames in each video

the frame to be suspicious the probability must greater than the threshold value.

$Thresholdvalue$=80.00000%

$if probability \geq$ 80.00000% ⇒ConsideredasSuspicious

S. No	Probability
1	86.643550
2	81.815506
3	84.574853
4	81.926688
5	82.767031
6	80.778792
7	89.714965
8	86.940765
9	83.015910
10	81.221327
11	85.565089
12	89.380309
13	89.701186
14	83.363722
15	87.789232

TABLE II
PROBABILITY OF SUSPICIOUS ACTIVITIES IN V3

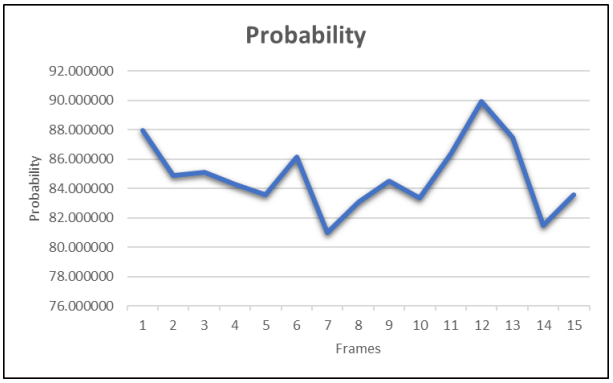

Fig. 5. Probability of the suspicious activities in V3

The above table is the sample video3 which is mentioned in the table1.It has 15 anomaly frames and their probabilities.

S. No	Probability
1	82.24384207
2	87.34415073
3	87.20511549
4	82.2536882
5	89.57724915
6	89.19642001
7	82.5877848
8	82.00354441
9	81.08333817
10	86.14629515
11	84.94310425
12	87.68037444
13	87.86786911

TABLE III
PROBABILITY OF SUSPICIOUS ACTIVITIES IN V5

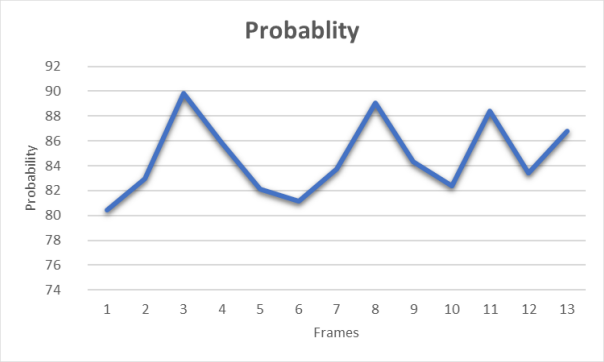

Fig. 6. Probability of the suspicious activities in V5

The above table is the sample video5 which is mentioned in the table1.It has 13 anomaly frames and their probabilities.

Here we maintain a threshold value for considering the video to be an Malfeasance which is 10.

$countofframes \geq 10 \Rightarrow Suspiciousactivityfound$

The above videos obeys the conditions so we consider it as suspicious activity.

To simplify the user interaction with the model, a simple web page is developed which provides the direct access of the overview. Whenever the model is run it automatically activates to a window as shown in the figure.

Using upload CCTV footage button the desired video or live video canbe uploaded and generate frames button start the model, which the initial step of generating frames can be done.

Once the video is uploaded, to generate the frames click on the button which says 'Generate Frames'.All the generated frames are saved in a new folder named 'frames' in the project folder.

On clicking the button 'Detect Suspicious Activity Frame', all the frames containing suspicious activities are identified and mentioned in the box on the page. In the above detected

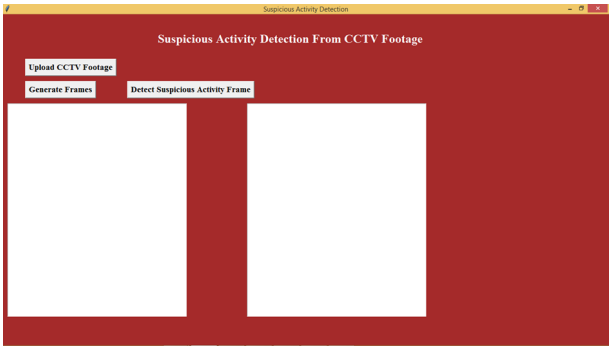

Fig. 7. Initial Webpage

Fig. 8. Frames containing Suspicious actions

frames suspicious activities can be found and can be taken care by the security. This helps the watchman to monitor public places more effectively.

V. CONCLUSION

Detection of malfeasance plays a vital role in recent years because of the increase in crimes day by day. We can conclude from this project that we can detect suspicious activities that are taking place around us by utilising Deep Learning techniques. There are plethora of methodologies we came across before proposing this project, which indeed become an high accurate model. The methodologies we came across are mentioned in detail and studied very thoroughly to conclude their advantages and disadvantages. This proposed method has a lot of advancements in future, as not all type of activities are detected. So it can be improved such a way that it can detect all types of activities from provided live CCTV footage.

REFERENCES

[1] B. Yang, J. Cao, R. Ni, and L. Zou, "Anomaly detection in moving crowds through spatiotemporal autoencoding and additional attention," Advances in Multimedia, vol. 2018, 2018.

[2] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu, and X.-S. Hua, "Spatiotemporal autoencoder for video anomaly detection," in Proc. 25th ACM Int. Conf. on Multimedia, 2017, pp. 1933–1941.

[3] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6479–6488.

[4] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 733–742.

[5] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks,” in Proc. IEEE Int. Conf. on computer vision, 2015, pp. 4489–4497

[6] Fangli, Maylor K.H Leung, Mehul Mangalvedhekar, Mark Balakrishnan, ”Automated video surveillance and alarm system”, in Proc. IEEE Conf. on Machine Learning and Cybernetics, 2008.

[7] S Venu Gopal, N Sambasiva Rao, ” An Algorithm for Simulated Routing Load while Sharing Files in Peer to Peer Systems” International Journal of Computer Mathematical Sciences (IJCMS), ISSN : 2347-8527, Volume 6, Issue 10, October 2017, Pg No: 88-93.

[8] Busaramoni Jayanth, Dr.S. Venu Gopal, ”RE-ESTIMATING PROCESS-LEVEL MICROBE ESTIMATE” Palarch’s Journal Of Archaeology Of Egypt/Egyptology 18(4). ISSN 1567-214x, pg: 1311-1317, Year:2021.

[9] S Venu Gopal, N Sambasiva Rao, S K Lokesh Naik ” Applying Load Separation Method in Structured Peer to Peer Overlay Networks” International Journal of Engineering Science and Computing (IJESC), Vol 6 Issue No:12, ISSN: 2250-1371, 2016 / Dec, pg. No: 3748 - 3750.

[10] S Venu Gopal, N Sambasiva Rao, ”Dynamic Sharing of Files from Disconnected Nodes in Peer to Peer Systems overlay Networks”, ICEEOT-2016, IEEE Conference, ISBN:978-1-4673-9940-1.

[11] S Venu Gopal, N Sambasiva Rao, ”Applying Load Separation Method in Structured Peer To Peer ”, 2016 IJESC, ISSN: 2250-1371 Volume 6 Issue No. 12.