DEEP LEARNING APPROACH FOR SUSPICIOUS ACTIVITY DETECTION FROM SURVEILLANCE VIDEO

[1]D Suman,[2]E Madhu Goud,[3]N Shiva Kumar,[4]G Nikhil Kumar

[1,2,3]Assistant Professor,[4]Student

Department of CSE

Christu Jyothi Institute of Technology &Science,Colombo Nagar, Telangana

ABSTRACT

The importance of video surveillance in modern society cannot be overstated. Once AI, ML, and DL were introduced, the technologies were already too far forward. Utilizing the aforementioned permutations, several methods have been developed to discern between different types of suspicious behaviour based on the live monitoring of film. Human conduct is the most erratic, and it's sometimes hard to tell whether an unusual pattern of behaviour is really typical. An alarm message is sent to the appropriate authority if the system predicts suspect behaviour in a school setting, and if the activity is normal, no alert is sent. In many monitoring applications, a series of frames taken from a video are used in rapid succession to conduct checks. The whole structure may be broken down into two halves. First, a classifier uses video frame computations to acquire features; next, using those characteristics, a prediction of whether a frame is suspicious or not may be made.

Index: Video Surveillance, deep learning,

1. INTRODUCTION

Several real-world contexts benefit from human behaviour detection applications including intelligent video surveillance and purchasing behaviour analysis. A wide variety of environments, both inside and outdoors, are ideal for installing video surveillance systems. A key component of safety is constant monitoring. The use of security cameras has grown commonplace in modern society for obvious reasons of personal safety and security. The Indian government has made electronic monitoring a central part of its growth plan. The Internet of Things in India. Video monitoring is still a key component. It's easier to keep an eye on things when there are fewer people to do it, and you can perform audits more cheaply and easily using video surveillance. So far, humans have been doing all of the tracking. Dealing with such a large volume of video data may be taxing on human resources, and errors are inevitable throughout the manual curation process. The system's efficiency suffers severely as a result. Thanks to advances in video surveillance automation, this problem has been mitigated. It is now humanly impossible to manually monitor every single event captured by a CCTV (Closed Circuit Television) camera. Manually looking for the same event in the recorded video is a time-consuming process, even if the event has already occurred. An emerging field in automated video surveillance is the examination of footage for signs of unusual activity.

Automatically and intelligently recognising any suspicious behaviour in a video surveillance system, human behaviour detection relies on the presence of a human observer. There are a variety of effective algorithms for automatically detecting human behaviour in crowded public places like airports, train stations, banks, workplaces, exam rooms, and the like. With the advent of AI, ML, and DL, video surveillance has become a promising new field of study. By using AI, we can make computers behave more rationally and creatively. Fundamental to the field of machine learning are the processes of acquiring knowledge from training data and extrapolating that knowledge to the prediction of new data. The availability of powerful GPU (Graphics Processor Unit) processors and massive datasets in recent years has led to widespread adoption of the deep learning paradigm. When used together, computer vision and video surveillance will make the streets more safer for everyone.

Computer vision involves a number of processes, including environment modelling, motion detection, object classification, tracking, behaviour description and interpretation, and synthesis of data from several cameras. More work is needed before using this approach to extract features from videos. The two types of classification methods are known as supervised and unsupervised classification. Appeared differently in relation to solo order, which is altogether PC driven and needs no human support, supervised classification relies on labelled training data annotated by hand.

When it comes to complex learning tasks, one of the most effective architectures is deep neural networks. High-level representations of images are automatically constructed by Deep Learning models, which also extract features. When feature extraction was done mechanically, this became more generalised. Convolutional neural networks (CNNs) are able to directly learn visual patterns from picture pixels. When it comes to video streams, LSTM models may learn complex relationships over time. Long short-term memory (LSTM) networks may retain information. The planned system would employ CCTV camera video to keep an eye on campus activity and sound an audible and visual alarm if anything seems out of the ordinary. Identifying events and identifying people's actions are the backbones of intelligent video surveillance systems. Human conduct is difficult to automatically grasp. Campuses often use extensive video surveillance systems to keep tabs on all the goings-on throughout their many buildings. Campus video footage has been utilised in testing. Data preprocessing, model training, and inference are the three main stages that make up the cycle of preparing a reconnaissance framework. Two neural networks, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) (RNN). It is for this reason that CNN is employed: to extract high-level characteristics from pictures and so simplify the input. Convenient for video stream processing, RNN is employed for categorization. This suggested system makes use of a VGG-16(Visual Geometry Group) model that has already been trained using the ImageNet dataset. At now, a model is being trained to infer behaviour from the available video. The algorithm is able to make predictions about whether or not human behaviour in the tape is suspicious, and hence may be utilised to improve the effectiveness of the monitoring process. Most of the current framework depends on CCTV camera video. This footage will be utilised for forensic purposes in the event of a violent or criminal incident. The most intriguing option, though, is a framework that can be utilized both inside and outside to naturally identify any strange or unexpected situations and inform the appropriate authorities. Using the offered strategy, scholars may create such a system.

II. RELATED WORK

The cited papers provide a variety of methods for identifying human actions in recorded footage. The efforts aimed to improve the ability to spot unusual or suspicious behaviour in video surveillance systems. Unauthorized entrance was identified utilizing a High level Movement Identification (AMD) technique [1]. To start with, the thing was isolated from its backdrop using frame sequences. After that, it was time to look for signs of foul play. The system's method is advantageous since it processes videos in real time and has a low computing cost. Nevertheless, the device has a low capacity for storing data and cannot be used in conjunction with a sophisticated video catch arrangement in touchy locales. In [2], a semantic-based technique is presented. With the help of background removal, the processed video data was able to pick out the foreground items. After deduction, a Haar-like calculation is utilized to determine if an item is alive or nonliving. We used a Real-Time blob matching method to follow the movement of objects. Throughout this research, a method for detecting fires was also identified.

Suspicious behaviours were identified in [3] based on motion characteristics between the objects. Suspicious occurrences were defined using a semantic approach. Objects were followed using a method based on object detection and correlation [2]. Based on motion characteristics and timing data, the occurrences are categorised. The provided framework required less computing. The optical flow at a university was approximated using a Lucas-Kanade approach, and any anomalous occurrences were found by dividing the campus into zones. After that, a histogram of optical flow vector magnitude was developed. In order to determine whether or not an occurrence is normal or abnormal, software algorithms are utilised to analyse the video's content [4].

The method was developed to identify anomalous occurrences by analysing motion data from videos. The video frame's histograms of optical flow orientations were learned using the HMM technique. It does this by comparing the obtained video casings to the standard edges already in existence, and then determining how similar they are. Many datasets, including the UMN dataset and PETS[5,] were used to test and verify the system. Keeping track of everything happening in front of a closed-circuit television (CCTV) camera physically is an unthinkable errand in the cutting edge world. Physically searching for a similar occasion in the recorded video is a period -consuming process even if the event has already occurred. A relatively new area of study in automated video surveillance is the analysis of footage for anomalous occurrences. When integrated into a video observation framework, human conduct identification might work as a programmed, wise technique for detecting any problematic way of behaving. Openly puts like air terminals, train stations, banks, working environments, diagnostic rooms, and so on, various successful calculations are accessible for consequently recognizing human behaviour. Video surveillance is a new frontier for the use of AI, ML, and DL.

By the use of AI, we can make computers behave more rationally and creatively. The ability to learn from existing sets and anticipate new data is crucial in machine learning. The accessibility of strong GPU (Designs Processor Unit) processors and enormous datasets has prompted the boundless reception of the profound learning paradigm. Understanding crowd behavior using a deep spatiotemporal approach classifies the videos into pedestrian future path prediction, destination estimation and holistic crowd behavior.es three different categories. Spatial information in the video frames was extracted using a convolutional layer. LSTM architecture

was used learn or understand the sequence of temporal motion dynamics. Data sets used in the proposed system were PWPD, ETH, UCY and CUHK. The accuracy of the system can be improved by using deeper architectures [6].

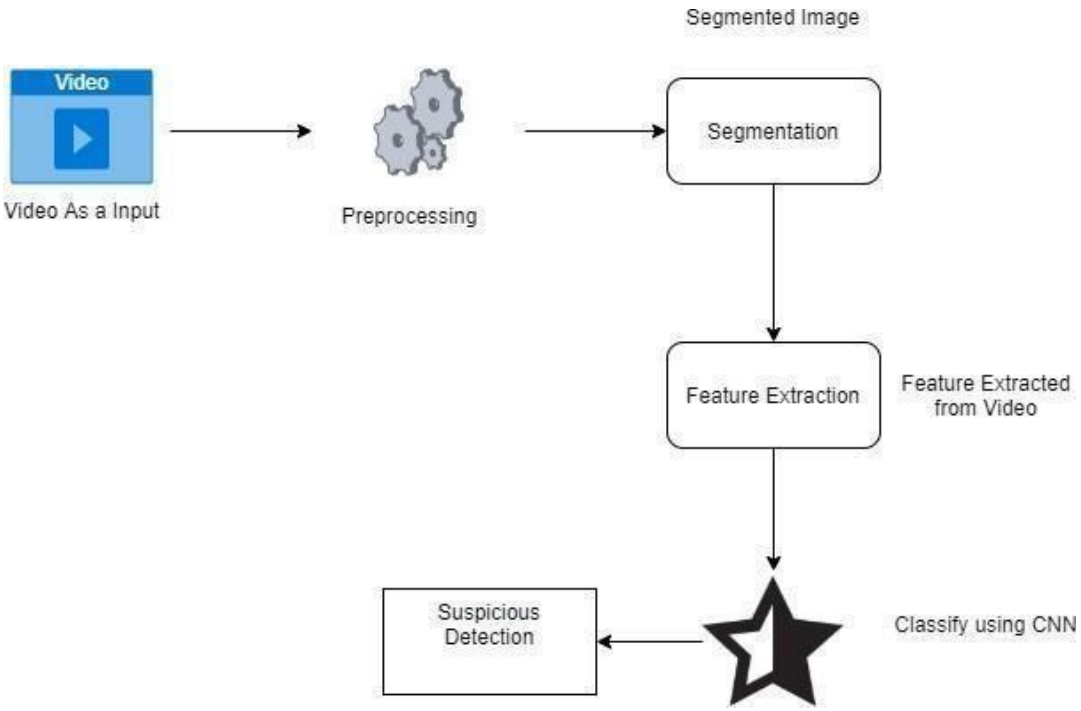
Daily human activities were captured from videos and classification of those videos in to household, work related, caring and helping. Sports related are done through deep learning. CNN was used for retrieving input features and RNN for classification purpose. They used Inception v3 model and UCF101, Activitynet as datasets. The accuracy achieved was 85.9% on UCF101 and 45.9% on Activitynet [7].

A system was developed to monitor students' behavior in examination using neural network and Gaussian distribution. It consists of three different stages: face detection, suspicious state detection and anomalous detection. The trained model decides whether the student was in suspicious state or not and Gaussian distribution decides whether the student performs any anomalous behavior [8]. The accuracy achieved was 97%.

III. PROPOSED METHODOLOGY

When it comes to solving complex learning problems, Deep Neural Networks is one of the most effective designs. Features are extracted and a high-level representation of visual data is built automatically using Deep Learning models. The fact that feature extraction is entirely robotic makes this more generalizable. Convolutional neural networks (CNNs) may acquire knowledge about visual patterns directly from picture pixels. Because of their ability to learn long-term dependencies, LSTM models are a good fit for processing video streams. Remembering is a strength of the LSTM network. The planned system would utilise CCTV camera video to observe campus activity and provide a warning if anything out of the ordinary happens. The capacity to recognise human behaviour and the detection of events are two of the most crucial features of intelligent video surveillance. It's not easy to build a system that can automatically analyse human behaviour. It is the responsibility of school administrators to keep an eye on the many goings-on throughout their sprawling campuses, many of which are under constant video surveillance. Campus-based video collected for evaluation purposes.

System Architececure



Video capture

The installation of closed-circuit television cameras and subsequent monitoring of the generated footage is the initial step in any video surveillance system. Several cameras collect a wide range of video formats over the whole monitored area. Frames are used for processing in our solution, thus movies must be transformed to frames before processing can begin.

Dataset Description

The KTH dataset is typical in that it contains six types of activities and a hundred sequences for each kind of activity. There are more than 600 frames in each sequence, at a pace of 25 frames per second [14]. Typical behaviours are taught to the model using this dataset (running and walking). Using the CAVIAR dataset, video footage from universities, and footage found on YouTube, we train for suspicious behaviour (mobile phone using inside the campus, fighting and fainting). There are 7335 still images in total, all culled from different sources. Completely manually labelled data, with 80% used for training and 20% for testing. In Fig.2 you can see the dataset's directory structure. Our system utilises video content from a variety of sources, including KTH, CAVIAR, YouTube, and campus-based recordings.

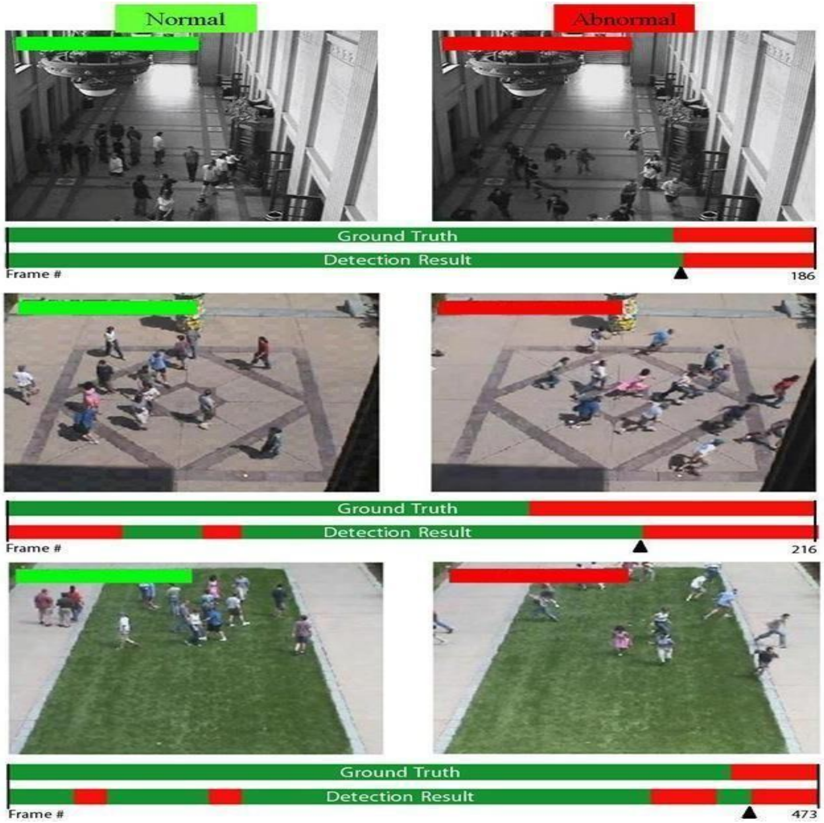


Figure: Normal vs Abnormal Activity

Video pre-processing

Our suggested approach uses a deep learning network to identify potentially malicious behaviour in surveillance footage. Using deep learning architectures may improve accuracy, especially when dealing with huge datasets. The schematic overview of the design is shown in Fig. 3. Datasets both already and newly developed serve as the source of the input videos. Frames are taken out of the recorded movies as a part of the processing step before they are used. The frames from the videos are organised into three distinct folders with descriptive names. JPG files are created from the 7035 individual frames extracted from the video. After that, we scale each frame to 224 224 so that it may be used with 2D CNNs. The testing video is additionally scaled to 224x224 pixels and converted to frames before being saved in a separate folder. To prepare videos for analysis, we utilise the Python OpenCV package.

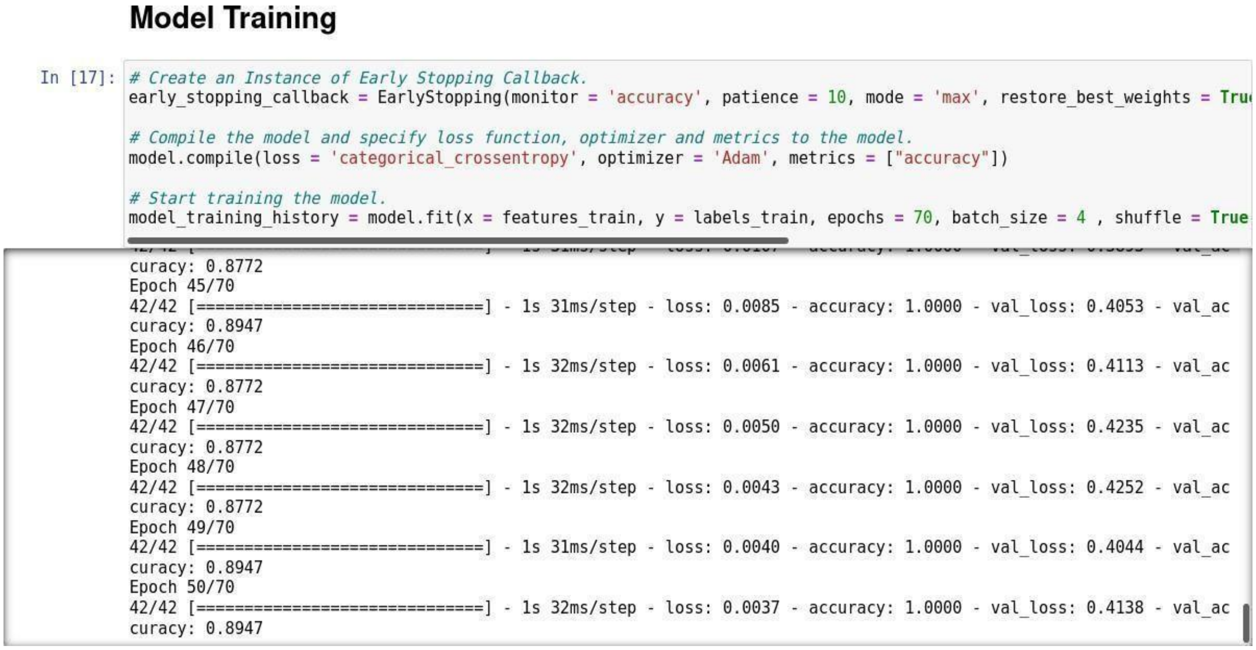
IV. EXPERIMENTAL RESULTS

The project's goal is to use surveillance video to keep an eye out for any suspicious behaviour on campus and to notify security immediately if anything out of the ordinary happens. To do this, CNN was used to draw characteristics from the images. The extracted frames are then classified using LSTM architecture to determine whether they are suspicious or not. Gathering video groupings from CCTV film, extricating outlines from films, preprocessing pictures, getting ready preparation and approval sets from datasets, preparing, and testing are vital undertakings in fostering a completely utilitarian framework. When it detects anything fishy, the system will send an SMS to the proper authorities. Python was used in the system's development, and it was built on an open source platform. An SMS sending account may be set up after introducing the twilio library in Python. Automatically settle on and get telephone decisions, send and get instant messages, and more using Twilio.

Model Training

The model is trained to predict over 3 classes – walking, running and fight. The training set is given to the model for training, with the following hyper parameters:

- epochs = 70
- batch_size = 4
- validation_split = 0.25

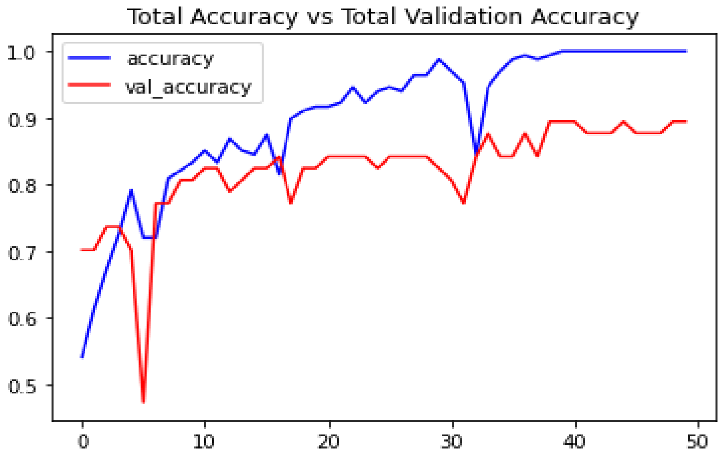


Model Training

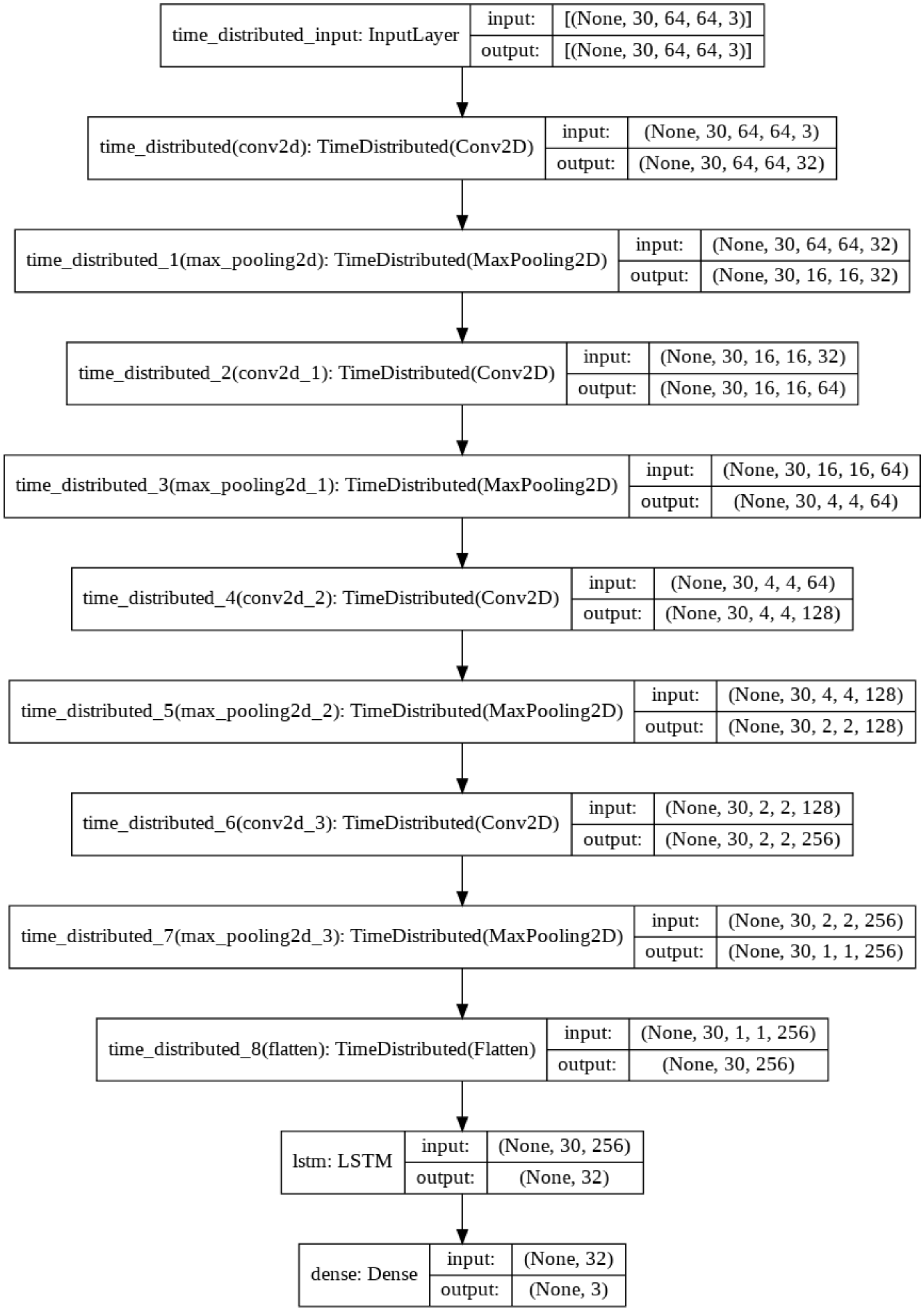
```
In [17]: # Create an Instance of Early Stopping Callback.
         early_stopping_callback = EarlyStopping(monitor = 'accuracy', patience = 10, mode = 'max', restore_best_weights = Tru

         # Compile the model and specify loss function, optimizer and metrics to the model.
         model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

         # Start training the model.
         model_training_history = model.fit(x = features_train, y = labels_train, epochs = 70, batch_size = 4 , shuffle = True
```

```
curacy: 0.8772
Epoch 45/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0085 - accuracy: 1.0000 - val_loss: 0.4053 - val_ac
curacy: 0.8947
Epoch 46/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0061 - accuracy: 1.0000 - val_loss: 0.4113 - val_ac
curacy: 0.8772
Epoch 47/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0050 - accuracy: 1.0000 - val_loss: 0.4235 - val_ac
curacy: 0.8772
Epoch 48/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0043 - accuracy: 1.0000 - val_loss: 0.4252 - val_ac
curacy: 0.8772
Epoch 49/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0040 - accuracy: 1.0000 - val_loss: 0.4044 - val_ac
curacy: 0.8947
Epoch 50/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0037 - accuracy: 1.0000 - val_loss: 0.4138 - val_ac
curacy: 0.8947
```

Accuracy vs Validation Accuracy



Model Layer Diagrams



time_distributed_input: InputLayer | input: [(None, 30, 64, 64, 3)] | output: [(None, 30, 64, 64, 3)]

time_distributed(conv2d): TimeDistributed(Conv2D) | input: (None, 30, 64, 64, 3) | output: (None, 30, 64, 64, 32)

time_distributed_1(max_pooling2d): TimeDistributed(MaxPooling2D) | input: (None, 30, 64, 64, 32) | output: (None, 30, 16, 16, 32)

time_distributed_2(conv2d_1): TimeDistributed(Conv2D) | input: (None, 30, 16, 16, 32) | output: (None, 30, 16, 16, 64)

time_distributed_3(max_pooling2d_1): TimeDistributed(MaxPooling2D) | input: (None, 30, 16, 16, 64) | output: (None, 30, 4, 4, 64)

time_distributed_4(conv2d_2): TimeDistributed(Conv2D) | input: (None, 30, 4, 4, 64) | output: (None, 30, 4, 4, 128)

time_distributed_5(max_pooling2d_2): TimeDistributed(MaxPooling2D) | input: (None, 30, 4, 4, 128) | output: (None, 30, 2, 2, 128)

time_distributed_6(conv2d_3): TimeDistributed(Conv2D) | input: (None, 30, 2, 2, 128) | output: (None, 30, 2, 2, 256)

time_distributed_7(max_pooling2d_3): TimeDistributed(MaxPooling2D) | input: (None, 30, 2, 2, 256) | output: (None, 30, 1, 1, 256)

time_distributed_8(flatten): TimeDistributed(Flatten) | input: (None, 30, 1, 1, 256) | output: (None, 30, 256)

lstm: LSTM | input: (None, 30, 256) | output: (None, 32)

dense: Dense | input: (None, 32) | output: (None, 3)

V. CONCLUSION

In today's society, almost everyone is aware of the value of CCTV video, but in most circumstances, this material is only utilised for inquiry after the act or occurrence in question has already occurred. The proposed approach has the potential to forestall the initiation of illegal behaviour, which is a significant benefit. We are monitoring the situation in real time using CCTV cameras and analysing the data as it is captured. If the analysis reveals an impending disaster, the final result is a command to the appropriate authorities to avert it. In other words, we can put an end to this. Techniques like Region-based Convolutional Neural Networks (RCNN), Faster-RCNN, Single Shot Detector (SSD), and You Only Look Once (YOLO) are often used for object recognition (YOLO). Faster-RCNN and SSD perform better when speed is more important than accuracy, but YOLO excels when precision is paramount. Deep learning combines SSD and

REFERENCE

[1] P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli 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) cameras ", International Journal for Research in Applied Science &amp; 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] Zahraa Kain, Abir Youness, Ismail El Sayad, Samih 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-Optik,vol152,January-2018.

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

[7] Dinesh Jackson Samuel R,Fenil E, Gunasekaran Manogaran, 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).

[9] Yuke Li "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior", IEEE Transactions on multimedia, Vol. 20, NO. 12, December 2018.

[10] Javier Abellan-Abenza, Alberto Garcia-Garcia, Sergiu Oprea, David Ivorra-Piqueres, Jose Garcia-Rodriguez "Classifying Behaviours in Videos with Recurrent Neural Networks", International Journal of Computer Vision and Image Processing,December 2017.

[11] Asma Al Ibrahim, Gibrael Abosamra, Mohamed Dahab "Real-Time Anomalous Behavior Detection of Students in Examination Rooms Using Neural Networks and Gaussian Distribution", International Journal of Scientific and Engineering Research, October 2018.

[12] G. Sreenu and M. A. Saleem Durai "Intelligent video surveillance: a review through deep learning techniques for crowd analysis" , Journal Big Data ,2019.

[13] Radha D. and Amudha, J., "Detection of Unauthorized Human Entity in Surveillance Video", International Journal of Engineering and Technology (IJET), 2013.

[14] K. Kavikuil and Amudha, J., "Leveraging deep learning for anomaly detection in video surveillance", Advances in Intelligent Systems and Computing,2019.

[15] Sudarshana Tamuly, C. Jyotsna, Amudha J, "Deep Learning Model for Image Classification", International Conference on Computational Vision and Bio Inspired Computing (ICCVBIC),2019.