Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P) Vol.12 No.1, 2023, pp.40-44

DOI: https://doi.org/10.51983/ajeat-2023.12.1.3636

Deep Learning Driven Instinctive Surveillance

Sunil Bhutada[1], P. Srija[2], S. Sushanth[3] and A. Shireesha[4]
[1]Professor, [2,3&4]Student,
Department of Information Technology, Sreenidhi Institute of Science of Technology, Telangana, India
E-mail: sunilb@sreenidhi.edu.in, peddisrija1234@gmail.com, salpalasushanth@gmail.com, shireesha.siri1008@gmail.com
(Received 16 April 2023; Accepted 8 May 2023; Available online 19 May 2023)

***Abstract* - It is a very boring and laborious job providing observation security. In order to determine whether the exercises that were caught were unusual or suspicious, a labour force is needed. Here, we'll put together a structure to automate the task of reviewing video reconnaissance. We will regularly review the camera feed to look for any unusual activities like surprising or suspicious ones. and an automatic acknowledgment will be sent to the user with an alert email along with the suspicious frames and SMS to mobile number. Deep learning computations for deep reconnaissance have improved from earlier encounters. These developments have revealed a key pattern in thorough reconnaissance and promise a significant increase in efficacy. Deep observation is typically used for things like identifying evidence of burglary, finding violence, and recognising explosion potential. We will propose a spatio-temporal auto-encoder for this project that relies on a 3D convolutional brain structure. The decoder then reproduces the edges after the encoder section has removed the spatial and transient data. By recording the recreation misfortune using the Euclidean distance between the original and replicated batch, the odd occurrences are distinguished.**
Keywords: **Surveillance, Deep Learning, Spatio Temporal, Euclidean Distance, Auto-Encoder**

I. INTRODUCTION

Surveillance cameras are being deployed more frequently in public spaces. Videos are produced in abundance and kept in storage for a while. Since it takes a huge team and constant monitoring, it is almost impossible for authorities to maintain monitor of these surveillance films and determine whether the instances are suspicious. As a result, there is an increasing need for this procedure to be automated with great precision. Additionally, it must be stated the type of frame is being used and which areas of it include the unexpected activity. This helps determine whether the strange behaviour is abnormal or suspicious more quickly. The time and effort needed to dig through the recordings would be saved, and it will help the concerned authorities identify the underlying source of the anomalies.

By concentrating attention on a narrow area of the data while disregarding enormous volumes of irrelevant data, automated detection of anomalies is incredibly helpful in minimising the amount of information that needs to be manually analysed. Contrarily, the issue of anomaly detection is amenable to a broad variety of perspectives, and research activities are distributed not solely in terms of methodology but also in regards of how the issue is seen, assumptions are made, and goals are set. This review aims to integrate these disjointed efforts by analysing the issue formulations and solution approaches employed in anomaly detection investigation as employed in automated surveillance.

II. LITERATURE SURVEY

Some authors tried to fix up the problem with existing appointment methods. Guruh Fajar Shiidik, Edi Noersasongko, Adhitya Nugraha, Pulng Andon, Juato, and Edi Jaya Kusma [1]. They have demonstrated that the primary studies that have been used to analyze the research topics the most recently show that the visual surveillance method, intelligent and integrated video surveillance, distribution, communication, and integrated video surveillance, are the three topics and trends that are currently the focus of research on video surveillance systems. They have also talked about how it is challenging to cope with the issue of motion detection in dynamic settings due to changes in lighting and weather, as well as shadow detection. Ali Bou Nassif, Manar Abutalib, Qassim Nasir [2], investigated anomaly detection using machine learning (ML) methods. They examined ML models from four angles: the type of application for anomaly detection, the kind of ML approach, the evaluation of the correctness of the ML model, and the kind of anomaly detection (supervised, semi-supervised, and unsupervised). We found that the studies used in the anomaly detection area most frequently are intrusion detection, network anomaly detection, general anomaly detection, and data applications.

Waqas Sultani, Chen Chen, Mubarak Shah [3] studied that using quality normal data may not be the very best method for anomaly detection due to the intervention of realistic anomalies. They try to take advantage of both typical and unusual surveillance footage. They develop a universal model of anomaly identification using deep multiple instances ranking framework with weakly labelled data in order to bypass the time-taking temporal annotation of anomalous parts in training films. A large-scale anomaly dataset made up of several real-world anomalies was used to validate the suggested method. Angela A. Sodemann, Matthew P. Ross, and Brett J. Borghetti [4] have published a review that looks at research on anomaly detection in

automated surveillance from several critical perspectives, including the issue domain, strategy, and method. They presented their findings and discussed the various methodologies that have been used in the literature. However, since each solution has only one use and may not work properly when exposed to the different variety of targets and behaviour common in realistic scenarios, the lack of approaches that can be applied to a wide range of targets may be a factor in the lack of variety usage of automated anomaly detection solutions.

Ming Cheng, Kunjing Cai, Ming Li [5] proposed a Crime dataset, It is a large-scale one of different dataset of many hours of videos. It contains of huge, very long and untrimmed real-world surveillance videos, with multiple realistic anomalies including Abuse, Assault, Road Accident, Burglary, Robbery, Shooting, and Stealing. These anomalies are selected because they have a significant impact on public safety. This dataset can be used for two different tasks. First, general anomaly detection by taking consideration all anomalies in one group and all normal activities in another group. Second, for identifying each of different anomalous activities.

Shipra Ojha and Sachin Sakhare [6], began by looking at the representation of objects using fundamental feature descriptors. They have provided a thorough explanation of the tracking process, starting with detection and recognition using various techniques like background removal and subtraction, temporal differencing, and optical flow, and ending with object tracking. Additionally, it discussed other methods of object tracking, adding region-based, super active contour-based, and feature-based.

III. PROPOSED METHODS

The suggested paradigm emphasises a comprehensive specification that is utilised to identify questionable behaviour. The crime rate is rapidly rising in the archives. As a human, it is quite difficult to keep an eye out everywhere on earth to stop these criminal activities. Therefore, we come up with an idea to propose our model wherever the algorithm is trained to identify suspicious behaviour using deep learning. Deep convolutional neural networks that have already been trained in addition to a spatial-temporal auto-encoder are utilised for initial classification, and a recurrent neural network is used for the final detection of suspicious activities. This section provides a detailed explanation of the methodology we employed.

First, the system receives a live video feed from a CCTV camera. The 3D-CNN receives the single-merged feature map as input. To reduce the training time for this methodology, we created an LSTM cell. The UCF-Crime dataset was used to train this D-CNN. The Kaggle dataset for UCF-Crime is used. The UCF-Crime dataset comprises 1900 recordings, each between 60 and 600 seconds long and with a different resolution. It was recorded using real-world surveillance cameras. It is the goal of this dataset to identify actual abnormalities like cruelty, incarceration, assault, crashes, robberies, and other violent crimes, as well as theft, vandalism, and fights. To determine the probabilistic classification, the SoftMax layer is employed. If suspicious activity is discovered via email or SMS, an acknowledgment is sent to the user based on this.

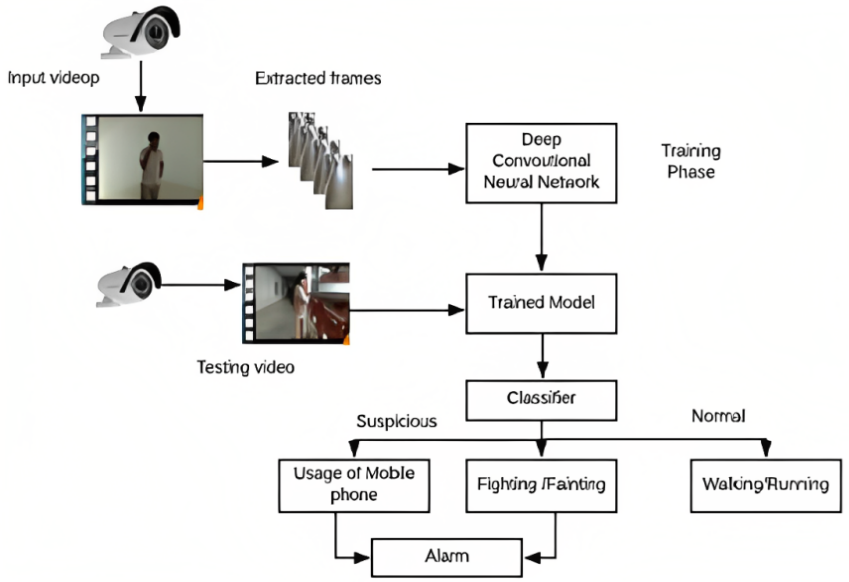


Fig. 1 Architectural Diagram

A. Video into Frames

The entire video is divided into frames using Python, which are then provided to Inception V3 as needed. Many deep learning specialists would approach video classification as if it were a series of image classifications conducted N times in total, where N is the total number of frames in the movie. This is because movies may be thought of as a collection of

individual images. With video classification, which goes beyond just simple picture classification, we can frequently infer that adjacent frames in a movie are linked in terms of their semantic contents. By utilising the temporal characteristics of videos, we were able to improve the outcomes of our real-time video categorization.

B. Convolution Neural Network

A false neural organisation does incredibly well when it comes to machine learning. Processing images, sounds, and texts are only a few of the activities that artificial neural networks are used for. For instance, we employ convolutional neural networks to forecast the arrangement of photographs and repeated neural networks, more specifically an LSTM, to forecast word grouping.

C. 3D-Convolution Neural Network

Recently, the grouping of video data has occasionally entailed using brain network engineering with a focus on 3D convolution layers. The 3D-CNN architecture is commonly used for moving 3D images, particularly clinical images because it can explore the locations of articles in time. During the convolution step, the 3D-CNN creates a 3D enactment map that is necessary for information inspection as well as for time and volumetric settings. To compute the representation of components at a low level using the dataset's 3D convolution, a three-layered channel is used. The component moves along three axes (x, y, and z). The following condition determines the value at each location of the element map in the layer: where pqrijm is the value of the portion associated with the component map in the previous layer as well, and R i is the size of the 3D piece.

A three-layered volume space is the end outcome. We create 3D convolution by wrapping around a block's central point and piling adjacent layers on top of one another. It is possible to get mobility data by connecting useful guides. Despite this, only one type of component may be separated using the convolution bit. The overall structure resembles a 2D convolutional brain. As a general rule, we can achieve better results by connecting a few convolution layers, just like with 2D convolution. Our results from building the 3D-CNN depend on the number of layers, the number of channels in each layer, and the size of the channels.

If pooling is used to create the brain organisation, as we are dealing with 3D data, the pooling size should be framed by three attributes. An information tensor is being nonlinearly down-sampled by the three-layered MaxPooling3D layer. This method divides the information tensor into three 3D subtensors and selects the subtensor component with the highest numeric value for each subtensor. By replacing each subtensor with its most extreme component, it finally transforms the information tensor into the result tensor. Frequently, Max Pooling 3D is used for variety pictures.

D. Spatio-Temporal Autoencoder-Decoder

In our system, a transient autoencoder is integrated with a spatial autoencoder. The organisation creates a result of a comparable size, addressing the predicted future outline, Y t+1, at each time step by using as information a video outline, Yt, of size H W. We thoroughly describe each of the modules in the following. The convolutional encoder-decoder engineering used in the spatial autoencoder is excellent. One convolutional layer is present in the encoder E, which is followed by a spatial max-pooling with subsampling layer and tanh non-linearity. With the exception of the non-linearity layer, the decoder D is an exact duplicate of the encoder and reduces the size of the output to that of the initial information using closest neighbour spatial up sampling.

The size of the component maps xt after the spatial encoder YtExt's forward pass is dhw, where d is the number of highlights and h and w are the level and width after down-sampling, respectively. The encoding is constructed in the secret layer that is lower on the stack. A smaller number of hubs make up the bottleneck layer, and the number of hubs in the bottleneck layer also determines how the information is encoded.

The transient auto-encoder's goal is to identify significant changes brought on by movement (such as inner self movement or the evolution of the objects in the picture), allowing it to predict the visual future while being aware of the past and present. According to Masci *et al.,* (2011), an excellent spatial auto-encoder uses a type of regularisation to prevent learning a trivial design and teaches the encoder and decoder to use separate component spaces that allow for an optimal information disintegration.

The decoder is required to learn about its own component space in order to fulfil the deterioration that the encoder freely chooses and recreate the information. The decoder uses largely the same activities as the encoder and has a comparable number of levels of opportunity. In contrast, the suggested ephemeral auto-encoder features a decoder with few teachable limits whose primary responsibility is to provide prompt input to the encoder, but without the restriction of correcting the encoder's errors as in the spatial example. In terms of improvement, the encoder, which is currently under more pressure to generate adequate element maps, is mostly to blame for the error made during learning.

IV. RESULTS AND DISCUSSION

A straightforward web page that offers immediate access to the overview has been built to streamline user engagement with the model. Every time the model is run, a window similar to the one in the figure automatically activates. The selected video or live video may be uploaded using the upload any video button, and the model can be started by clicking the create frames button after the first stage of producing frames has been completed.

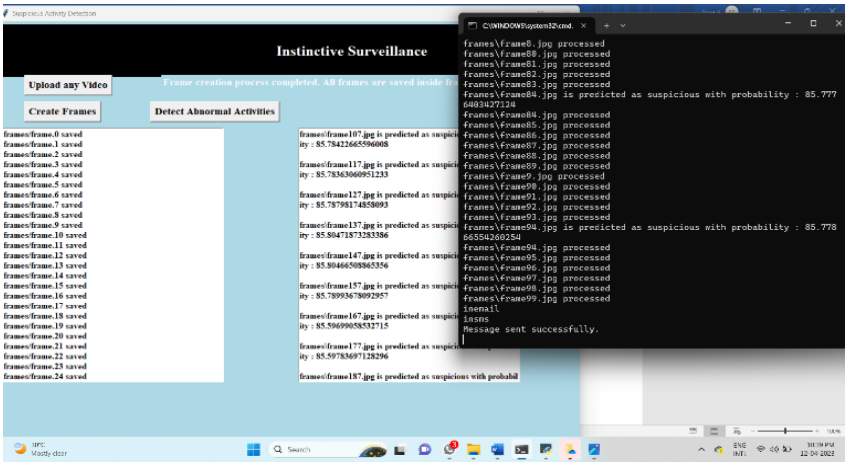
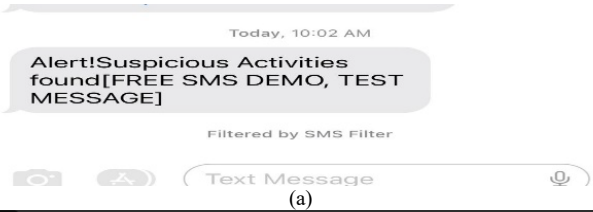


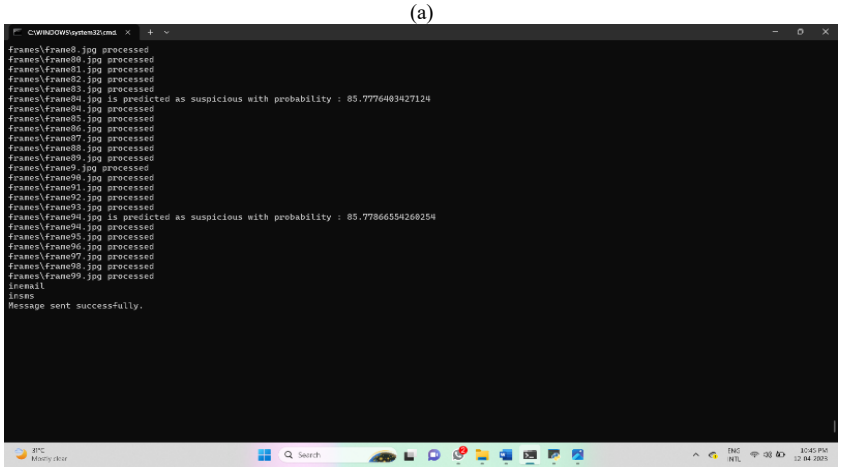
Fig. 2 Activity frames that are suspicious detected

In the given below figure is a detected suspicious activity frame by our build model and it is transparently visible that there is a human wearing a face mask who is willing to steel. Frames which are found suspicious are sent to the user through Email and an alert SMS will be sent without any human interference automatically each time when user runs the code. User can directly check the suspicious frames from anywhere.

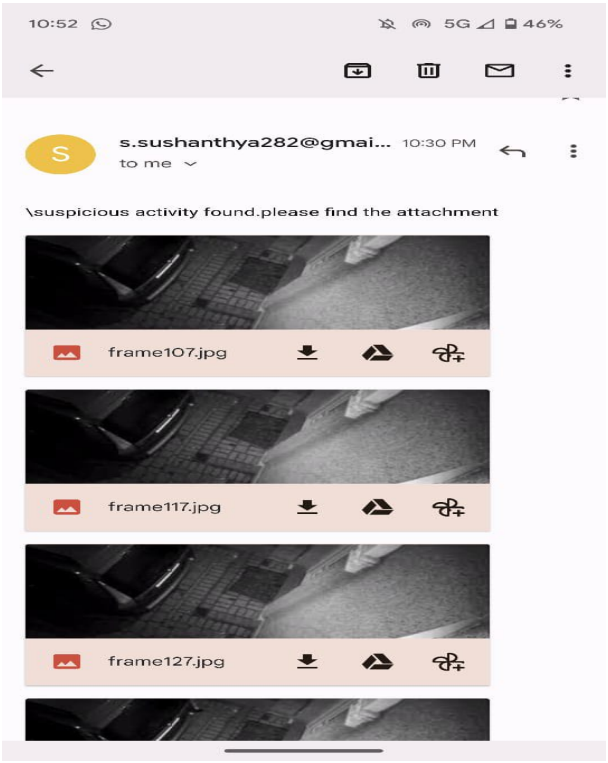
Fig 3. Suspicious activity frame



(a)



(b)



(c)

Fig. 4 An attachment of images in email and alert SMS regarding suspicious activities will be sent

V. CONCLUSION AND FUTURE SCOPE

Due to the daily rise in crime, the detection of suspicious activity has been increasingly important in recent years. We can infer from this experiment that using Deep Learning algorithms, we can identify suspicious actions that are occurring all around us. Before putting up this project, we came across a wide range of approaches that did in fact result in a highly accurate model. The approaches we found are described in detail and thoroughly examined to determine their benefits and drawbacks. Since not all types of behaviours are identified, this suggested strategy will likely develop significantly in the future. Therefore, it may be enhanced so that it can recognise any form of activity from the live CCTV footage. We also made it easier by sending an automatic acknowledgment or alert to the user with an alert email along with the suspicious frames and SMS to mobile number anytime, if any suspicious frames are found.

VI. REFERENCES

[1] Guruh Fajar Shidik, Edi Noersasongko, Adhitya Nugraha, Pulung Andono, Juanto, And Edi Jaya Kusuma, "A Systematic Review of Intelligence Video Surveillance: Trends, Techniques, Frameworks, and Datasets," *IEEE ACCESS,* Vol. 7, Dec. 2019.

[2] Ali Bou Nassif, (Member, IEEE), Manar Abu Talib , (Senior Member, IEEE), Qassim Nasir, and Fatima Mohamad Dhakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE ACCESS,* Vol. 9, June 2021.

[3] Waqas Sultani, Chen Chen and Mubarak Shah, "Real-world Anomaly Detection in Surveillance Videos," *Cornell university papers,* Feb. 2019.

[4] Angela A. Sodemann, Matthew P. Ross, and Brett J. Borghetti, "A Review of Anomaly Detection in Automated Surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* Vol. 42, No. 6, Nov. 2012.

[5] Ming Cheng, Kunjing Cai, Ming Li, "RWF-2000: An Open Large Scale Video Database for Violence Detection," *Slack,* Nov. 2019.

[6] Shipra Ojha and Sachin Sakhare. "Image processing techniques for object tracking in video surveillance-A survey," in *International Conference on Pervasive Computing (ICPC), IEEE,* 2015.

[7] [Online]. Available: https://cloud.google.com/tpu/docs/inception-v3-advanced.