Identification And Detection of Abnormal Activity in Atms Using Deep Learning

National Conference at Muthayammal Engineering College

[1]Sharath babu CG, [2]Dr. Anitha Devi.M.D, [3]Dr.M Z Kurian

[1]PG, [2]Associate Professor, [3]Head of Department
Electronics and Communication Engineering, Sri Siddartha Institute of Technology, Tumkur, Karnataka

Abstract: In the course of our daily lives, we are witness to a significant amount of dishonesty and theft that goes unreported. The automated teller machine is the scene of most major crimes. The purpose of this research is to present a new supervised method that can detect odd occurrences in restricted spaces such as ATM rooms, server rooms, and other such places. One of the technologies that are used to make up the technical base is abnormal behaviour detection using image processing. The purpose of the work that is being proposed is to establish a technical base that will support a social infrastructure that is both more secure and more convenient for its users. The robbery will be prevented as a result of this, and the perpetrator of the robbery will be easier to find. The use of this method will result in an end to robberies as well as a significant decrease in the number of complaints received. Because of this, the results of the suggested framework indicate that the framework is capable of delivering a high level of security to the ATM System's.

Keywords: ATM, Image processing, real time, microcontroller, crime, abnormal events

1. Introduction

The investigation of illegal activities and strange occurrences can be exceedingly challenging. The ability to anticipate the layout of a crime scene can make the work of law enforcement authorities easier. An automated teller machine, often known as an ATM, is a computerised telecommunications equipment that enables customers of a financial company to have easy access to financial transactions in a public setting without the need for the assistance of a bank teller or a cashier. The number of ATMs being installed has increased substantially in order to support the billions of transactions that are being processed each year. There has been a rise in the number of criminal acts, such as robbery, murder, and other offences, which has resulted in a heightened sense of urgency regarding the installation of a reliable system that can safeguard individuals as well as ATM installations [1, 2]. A Personal Identification Number, often known as a PIN, or a secret key is an essential part of the security system of an automated teller machine (ATM). It is common practise to utilise a personal identification number (PIN) or a secret key to protect the financial information of customers from unauthorised access. Despite this, there has been a rise in the amount of criminal activity that is associated with financial companies due to the proliferation of automation and smart technologies. From 1999 to 2003, there was a slow but steady rise in the number of these crimes; then, after a brief period of decline in 2004, there was a sharp uptick that began in 2005. As a result, there has been a significant rise in the number of robberies throughout the course of the last 12 years. As a result, we have presented a solution for this issue that makes use of the techniques for deep learning. This project sends an alert to law enforcement personnel before a theft or crime takes place within an automated teller machine. As a result, this model helps to react early to the robbery and take required actions if they are possible. This is preferable than reacting late to the heist.

Researchers are showing a significant amount of interest in the practise of analysing human behaviour based on video. The objective of the recognition of human behaviour is to automatically assess a variety of activities taken in a film that is unknown to the user. The identification of the motion pattern and the production of a high-level summary of the activities are both necessary steps in the process of conducting an analysis of a variety of behaviours. The recognition of a face and the details associated with it are both significant. When the criminal accesses the ATM while presenting authorised credentials, it is necessary to verify their identity. Multiple approaches, spatiotemporal interest of feature points, images of motion from the past, cumulative motion pictures, and a model of bag of words are just some of the different methodologies that have recently been used by a number of researchers for the successful recognition and visualisation of human behaviour. Other methodologies include: In this article, a framework is presented that has the potential to transform the existing patterns of the surveillance system. If the proposed system is implemented, there is a greater possibility that the perpetrator of the crime will be apprehended by the authorities because they will be informed about the crime as soon as it occurs. In point of fact, the research that was conducted can also be used to generate automatic alerts that can warn security personnel at the ATM site to gain urgent security, in addition to other individuals who are in the vicinity of the premises.

2. Related works

The Random forest classification technique was proposed by V. Tripathi et al. [3], and in this approach, the data was preprocessed by partitioning the movie into frames and then doing it on every nth frame. It was combined into a single frame, and then the HOG feature descriptor was applied to that single frame in order to extract information that is meaningful. CAVIAR, HMDB 51, and also their own dataset that they produced were the databases that were utilised. They achieved an accuracy rating of 75.83 percent on CAVIAR and a score of 50.84 percent on HMDB 51.

Support Vector Machine [4, often known as SVM], was yet another method of machine learning that was applied. This was a vision-based system that worked by first extracting features, and then classifying those features. In the first stage, the video frames were first transformed into grayscale, and then the SIFT and Gabor filters were applied to those frames in order to extract features from

them. The data were separated into two categories using the SVM classifier, which took into account the characteristics of each group. Even though they used their own dataset, the accuracy that was acquired was poor.

An anomaly detection method that makes use of low-level characteristics is described in [5]. This system computes dense motion fields and statistics for each frame individually. After that, a technique called motion directed principal component analysis is performed to extract important principle features over a period of time. In conclusion, the one-class SVM differentiates abnormal occurrences from everyday occurrences.

S. Abdul Kareem et. all, [6] The use of hand signals as a natural interface serves as a driving force for research into motion scientific classifications, their representations, and recognition techniques. The writing classifies motions into two different types: static signals and dynamic signals. Calculating using the K-means starts with arbitrarily locating the nearest neighbour in the unreal space. The next step is to assign a value to each pixel in the accumulating information image. To the nearest group focus, and the areas that make up that group focus will be relocated until they are at the average of their class values. The k-nearest neighbour method is a technique for classifying items in terms of the component space's most immediate possible preparing cases.

Miwa takai [7] In this study, Motion Region is extracted from a moving person, and Motion Quantity is measured in order to determine the active state of the person. In addition, the approach of the proposal locates the point at which suspicious behaviour is detected and calculates an estimate of the level of danger posed by the suspicious activity.

Sambarta Ray et al. [8] came up with a conceptual framework that takes into consideration two odd occurrences. In the beginning, it uses a recognition calculation to recognise human faces and count the number of people who are present within the ATM booth. This helps it determine how many people are using the machine. Second, it is able to determine whether or not a man is concealing his face by wearing a veil. The Viola-Jones computation is utilised inside the recognition framework in order to differentiate the face and the various facial angles, such as the eye match, the nose, and the mouth. The system then determines whether or not an individual is covered by a veil based on the number of people that are now present within the ATM and the face component that has been identified as being used to make that determination.

CNN and LSTM were used in the method that was proposed for action recognition by S. Arif and colleagues [9]. The training of a motion map, which is a representation of videos, is what is required to accomplish the action recognition. Researchers have suggested new 3D convo-based iterative training methods to build motion maps in this study. These methods allow for the removal of redundant information and reveal discriminative data that may be represented in a good manner. The C3D network was responsible for encoding the local temporal features found within each video unit. In order to circumvent issues with vanishing gradients and get an understanding of long-term contextual information based on temporal sequences, this system makes use of LSTM. LSTM is utilised to process the fused spatio-temporal information, and as a result, it is possible to recognise complicated frame-to-frame hidden sequential patterns using this method. The UCF sports dataset was utilised, and a level of accuracy of 93.9 percent was accomplished. It has been noticed that even for movies of varying lengths, their method is rather effective, and it has demonstrated a significant improvement in motion recognition.

It was proposed by Kamarul Hawari Ghazal et. all, [10] that a unique method in include extraction might be used to arrange confined and expansive weed using SIFT key-focuses descriptor. To be more specific, we analyse the SIFT key components of weed images and sketch out a method to extract the element vectors of SIFT key-focuses according to magnitude and point heading. It has been discovered that the Scale Invariant Feature Transform, also known as SIFT, is the most effective local invariant element descriptor. Filter is a method for recognising and deleting neighbouring component descriptors that are sensibly invariant to changes in brightness, image clamour, pivot, scaling, and small changes in perspective. Filter is a technique. The SIFT calculation is typically used for question acknowledgment and discovery, as it is unaffected by shifts in lighting conditions and may be performed using relative or 3D projection.

3. Research Methodology

Because of the varying illumination, scale, position, and perspectives, as well as the complicated human body movements that occur during violent action, traditional approaches are unable to capture the distinguishing characteristics that are connected with emotions. In order to effectively identify instances of violent behaviour, it is necessary to do analysis on both spatial and temporal data. Therefore, in order to extract strong and discriminative spatial information from aggressive action, we need to make certain adjustments to MobileNetV2. Then, for spatio-temporal features, a lightweight LSTM model that can simply be deployed for devices with low resources is proposed to recognise violent action in tough contexts. Figure 1 illustrates the proposed system to be used.

3.1. Preprocesing Phase

In this stage of the process, the objective is to harvest significant pilgrim's frames from CCTV cameras in order to make effective use of the resources provided by our suggested system. In most cases, the surveillance system is made up of a number of cameras that are connected to one another. This allows the system to effectively monitor a certain region. In most cases, processing each every frame of a movie is not required because there are just a small number of sequences that are important for comprehending human behaviour. There must be some sort of motion on the part of the pilgrims in order to facilitate the detection of the anomalous behaviour.

3.2 Spatial Features Extraction

The key benefit of having a CNN architecture is the ability to handle high-dimensional input and extract significant discriminative features more effectively because to the local connection and weight sharing characteristics. However, it has required a high level of computing, and as a result, it cannot be employed in devices that have a limited amount of resources. After doing a number of experiments on AlexNet and VGG-16, we came to the realisation that these models required a significant amount of computing. This convolution has a significant computational cost because it simultaneously pulls features from all of the channels, whereas the MobileNetV1 designs have a revolutionary method that helps lower this computational cost. It begins by extracting features from

a single channel at a time before proceeding to aggregate the features that have been extracted. The overall complexity of regular convolutional-based models can be circumvented using this method, but at the expense of some speed reduction.

3.3. Temporal Features Extraction

LSTM is an enhanced type of RNN that is capable of learning temporal patterns in time series data. These patterns can be learned from the data. The vanishing gradient problem is a limitation of a simple RNN, which means that it is unable to learn long-term sequences and loses the effect of early sequence dependencies. LSTM is the answer to this problem. The LSTM is equipped with input, forget, and output gates that collectively assist in the
acquisition of long-range sequence knowledge. The input is taken at each step by a naive RNN, but an LSTM module makes the decision about whether or not to take the input by utilising a sigmoid layer to determine whether or not each gate should be open or closed.

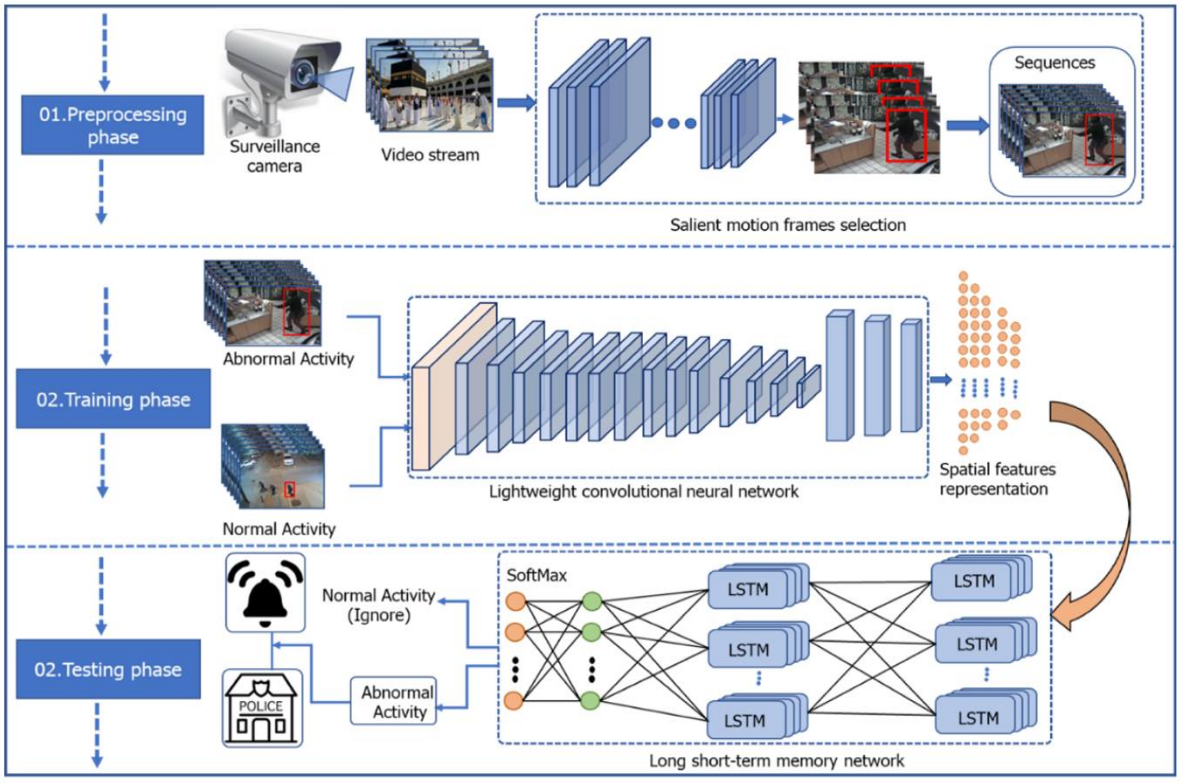

Figure 1 Proposed Block Diagram

4. Results and Discussion

Figure 2 presents an analysis of the computational difficulty of the model that has been proposed. One of the most significant benefits of machine learning, and CNN in particular, is the extraordinary performance it has on data that it has not before seen. Nevertheless, during the testing phase, it calls for a significant amount of calculations. As a result, these models have not been developed successfully in devices that have a restricted amount of resources. The transfer of the CNN application to hardware with constrained resources is something that researchers are particularly interested in doing. When performing an examination of the computational complexity, it is common practise to employ numerous parameters to analyse the performance of the models.

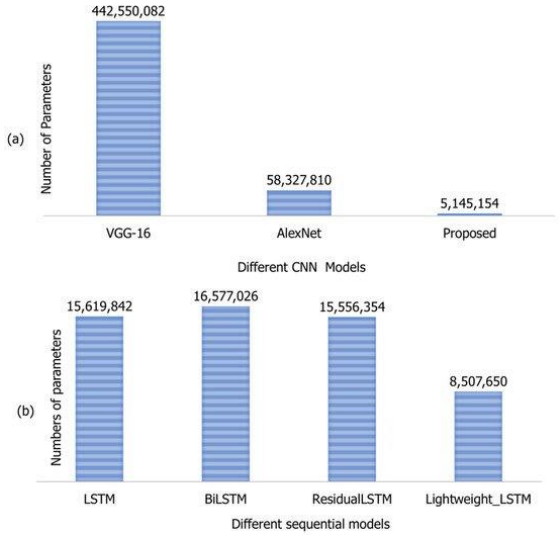

Figure 2 The computational complexity of the models

The performance of the suggested model is tested using the most recent and cutting-edge approaches that are currently available. Recognizing the presence of violent behaviour can be a difficult undertaking. Researchers from a wide variety of institutions have detailed the ways that they have developed to accurately identify instances of violent behaviour using computer vision. The abnormal activites are shown in figure 3 and figure 5.

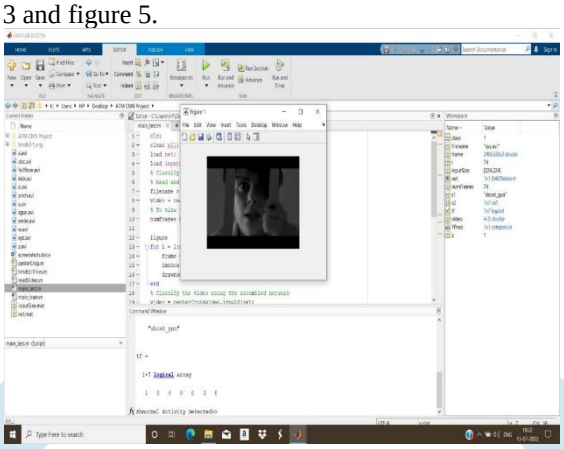

Figure 3 Abnormal activity detection

In addition to the approaches that were discussed, the CNN method that was proposed in this research has attained the highest level of accuracy possible, which is 96 percent. The normal activity is shown in figure 4. In addition to this, the

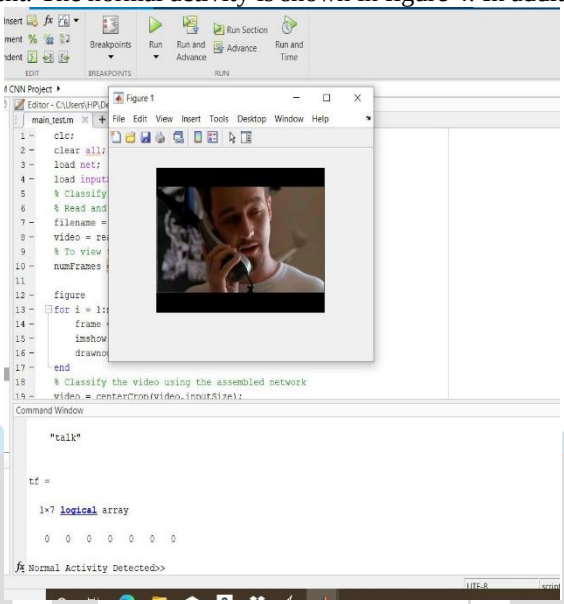

Figure 4 Normal activity detection

sequential model has been shown to reach an accuracy of 98 percent on the datasets. In addition, in comparison to the most current methods that have been described, the strategy that we have developed requires a smaller number of parameters.

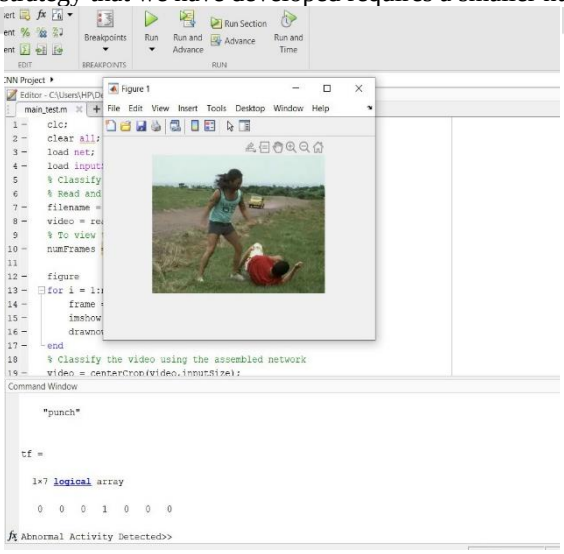

Figure 5 Abnormal Activity Detection

5. Conclusion :

The installation of CCTV is done primarily with the goal of preventing criminal activity or property damage through the identification of suspicious or unusual actions taking place within the monitoring system. There is a significant amount of interest in the creation of an intelligent surveillance system that not only decreases the amount of human involvement in monitoring but also notifies the appropriate authority in a timely manner from the occurrence of any potential future problems. Those are generally aware of the presence of surveillance cameras practically everywhere; hence, the behaviour of people who are involved in criminal activity may appear normal in most cases. However, if there are an excessive number of false alarms, this could lead to frustration or a lack of trust in the system. As a result, the development of a unique model that requires minimal training time and data set, has high accuracy, and can learn on its own over the course of time is urgently required.

References

1. M. S. Scott, Robbery at Automated Teller Machines, US Department of Justice, Office of Community Oriented Policing Services, 2001.
2. N. Sharma, " Analysis of different vulnerabilities in auto teller machine transactions," Journal of Global Research in Computer Science, pp. 38– 40, 2012.
3. Tripathi, Vikas, Durgaprasad Gangodkar, Vivek Latta, and Ankush Mittal. (2015) "Robust Abnormal Event Recognition via Motion and Shape Analysis at ATM Installations." Journal of Electrical and Computer Engineering 2015:1-10.
4. Arpitha K, Honnaraju B., "Vision Based Anomaly Detection System for ATM." International Research Journal of Engineering and Technology (IRJET), 2018, pp. 4235-4240.
5. Liu, Chang, Guijin Wang, Wenxin Ning, Xinggang Lin, Liang Li, and Zhou Liu. (2010) "Anomaly detection in surveillance video using motion direction statistics." IEEE International Conference on Image Processing 717-720.
6. Haitham Hasan and S. Abdul Kareem, "Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey." IEEE International Conference on Advanced Computer Science Applications and Technologies, 2013.
7. Miwa takai, "Detection of suspicious activity and estimate of risk from human behavior shot by surveillance camera" Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress, IEEE 2011.
8. Sambarta Ray, Souvik Das and Dr. Anindya Sen, "An Intelligent Vision System for monitoring Security and Surveillance of ATM" India Conference (INDICON), 2015 Annual IEEE. [9]. S. Arif, J. Wang, Tehseen Ul H and Z. Fei "3D-CNN-Based Fused Feature Maps with LSTM Applied to Action Recognition "Future Internet, 2019, pp. 1-17.
9. Kamarul Hawari Ghazali, Mohd. Marzuki Mustafa and Aini Hussain, "Feature Extraction Technique Using SIFT Key Point Descriptors", Proceedings of the International Conference on Electrical Engineering and Informatics Institute of Technology Bandung, Indonesia June 17-19, 2007
10. Habib, S.; Hussain, A.; Albattah, W.; Islam, M.; Khan, S.; Khan, R.U.; Khan, K. Abnormal Activity Recognition from Surveillance Videos Using Convolutional Neural Network. *Sensors* 2021, *21*, 8291. https://doi.org/10.3390/s21248291