International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228688

Abnormal Activity Recognition in Private Places Using Deep Learning

Anjali Suthar[1], Prof. Jayandrath Mangrolia[2], Prof. Ravi Patel[1]
[1]M. Tech. Student, [2]Assistant Professor
[1]Department of Artificial Intelligence, Charutar Vidya Mandal University, Anand, Gujarat, India
[2]Department of Information Technology, Charutar Vidya Mandal University, Anand, Gujarat, India

ARTICLEINFO

Article History:

Accepted: 01 May 2023
Published: 20 May 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
321-328

ABSTRACT

Using computer and machine vision technology, the process of analysing human motion is known as "human activity recognition," or HAR. Anomaly detection in security systems is one of the situations in which human activity recognition is useful. As the demand for security growing, surveillance cameras have been widely installed as the foundation for video analysis. Identifying anomalous behaviour demands strenuous human effort, which is one of the main obstacles in surveillance video analysis. It is necessary to establish video recording in order to automatically catch anomalous activities. Using deep learning methods, our intelligent video surveillance system can identify an anomaly in a video. Real-time detection of the actions is also possible, and these video frames will be afterwards preserved as photographs in the system for the user to examine. The suggested Abnormal Activity Recognition system was created with the goal of identifying and detecting irregularities through a live feed in the banking sector, more specifically in an ATM setting. The initial phase of the study focuses on the application of image deep learning techniques to recognise various items and spot unusual behaviour using ATM monitoring systems.

Keywords : YOLOv5, Convolution Neural Network (CNN), Artificial Intelligence, Motion Theory

I. INTRODUCTION

The automated teller machine (ATM) is now one of the most crucial tools used by customers all over the world to withdraw cash or conduct other activities. Yet, the ATM is where the major crimes are committed. Every day, there are several locations where ATM machines are robbed, creating a security issue. Each ATM has a watchman assigned to it in order to avoid this issue. Every day, numerous such films are captured by CCTV cameras installed within the ATM. Videos that have been recorded are too long, and automated video analysis techniques [2] have not yet produced the expected outcomes. As the videos are so long, watching them all becomes difficult and tedious [4]. A system that only takes the most important information from a lengthy movie should exist. The main information in surveillance videos is any suspicious activity, such as robberies and murders. So, it is necessary to extract this crucial information from lengthy videos. It is impossible to manually monitor every incident captured on the CCTV camera. Even if the incident had already occurred, manually searching for it in the recorded video is a time-consuming process. Sadly, there are a number of reasons why the existing systems are not very effective at detecting behavior and activity.

The goal of this project is to develop an algorithm that would enable the authorities to identify suspicious frames from a lengthy surveillance video and provide them with priority information. The Convolutional Neural Networks technique with Deep Learning was utilized in this study to sample the important data from the surveillance videos. The most important information concerned any suspicious activity—such as a robbery, murder, theft, etc. —that occurred inside an ATM. The CNN model's outcomes successfully extracted suspicious activity frames from a lengthy movie, allowing users to first identify the features before extracting worrisome frames.

Intelligent solutions that can automatically provide accurate warning feedback in real time are what we need. Monitoring of the ATM that looks for unusual behaviors. It calculates their position relations and extracts features that can be utilized to study a person's behavior in an efficient manner. When the system notices an odd behavior, it notifies the ATM monitoring staff, sends a warning message, and activates an alarm in the ATM.

In this research work object detection is implemented using YOLOV5 algorithm. Convolutional Neural Network (CNN) was designed and trained on the datasets in order to evaluate the performance of CNN trained from scratch. The performance of these models are evaluated using metrics such as accuracy, loss, precision, recall and f1-score. Confusion matrix is used to evaluate the model on a test dataset

II. LITERATURE REVIEW

In this section, we present the related work and research undergone in developing videobased security system. It suggested a deep network architecture based on residual bidirectional long-term memory (LSTM). With an improvement in recognition rate, the new network was capable of avoiding gradient vanishing in temporal and spatial dimensions. To understand the complexity of activities recognition and classification, two LSTM models, the basic model and the proposed model, were used in a comparative analysis to understand the classification of the models for the classification of images of five human activities such as abuse, arrest, arson, assault, and fighting.

The suggested model is used to conduct the categorization of five distinct human activities, and its performance is excellent. The training and testing accuracies were 99. 68%. With no loss and 0. 016%, the training and classification losses are both excessively low. The findings revealed that the suggested LSTM model was extremely effective in training and comprehending human actions, as well as performing well in categorization.

Further research will focus on constructing new LSTM-based recurrent neural network models capable of recognising human actions even in large-scale films. The research is also looking at additional performance variables like as accuracy, recall, and F1-score values, which may help influence the performance of any LSTM model.

III.PROPOSED SYSTEM

With the literature review been conducted, it was revealed that the Deep Learning Models have been widely used resulting better scales of accuracy and to serve the Human Activity Recognition process.

3.1 Dataset

The dataset ATM Image (ATM-I) comprises 1491 images that cover most of the angles in which an ATM box can be viewed in an ATM vestibule. Images in the dataset are augmented with blur (up to 2. 25px) and noise (up to 6% of pixels) effects. Augmentation is done to expand the dataset and increase model performance. The image dataset has been created where each image is bounding box annotated for the ATM and person class.

Second freely available dataset is ATM Anomaly Video Dataset (ATMA-V)[9] Dataset from Kaggle. The video dataset comprises 65 videos that consist of both anomalous and normal video segments.

As part of our abnormal behavior classification, a dataset carried out those activities Such as Fight, Activity with Knife, Normal Videos, Property Damage, robbery, peeping to check the password, snatching the withdrawn money, covered face etc. and classified the that activates are normal or abnormal.

3.2 CNN Architecture

The CNN model was defined as having two CNN hiddenlayers.

Eachofthemarefollowedbytwodropoutlayersof Then a dense fully connected layer is used to interpret thefeaturesextractedbytheCNNhiddenlayers.

Finally,adenselayer with the softmax activation function was added as thefinallayertomakepredictions(TableI).

The sparse categorical cross entropy loss function will beused as the loss function and the efficient adam version ofstochastic gradient descent was used to optimize the networkwith a learning rate of 0. 001. CNN model was trained for 50epochs and a batch size of 64 samples were used. After themodel is fit, it was evaluated on the test dataset and theaccuracyoftheCNNmodelwas obtained

Layer	Output Shape	Param#
Conv2D	None,79,2,16	80
Dropout	None,79,2,16	0
Conv2D	None,78, 1,32	2,080
Dropout	None,78,1,32	0
Flatten	None,2496	0
Dense	None,64	159,808
Dropout	None,64	0
Dense	None,6	390
Total params:162, 358		
Trainableparams:162, 358		
Non-trainableparams:0		

Table1. The Dimensional Structure Of The Adopted Cnn Model.

3.2.1 LSTMArchitecture

The LSTM model was defined as having a single LSTMhiddenlayer. Adropoutlayervaluing0. 5followsthis. Thenadense fully connected layer is used to interpret the featuresextracted by the single LSTM hidden layer. Finally, a denselayer was added as the final layer to make predictions.

For the purpose of compiling and training the LSTMmodel, the same values for the loss function, optimizer, batchsize and the number of epochs, which we used, in compilingandtrainingtheCNNmodel wereused. Afterthemodelisfit,it was evaluated on the test dataset and the accuracy wasobtained.

Layer	OutputShape	Param#
LSTM	None,100	41600
Dropout	None,100	0
Dense	None,100	10100
Dense	None,6	606
Totalparams:52,306		
Trainableparams:52,306		
Non-trainableparams:0		

Table 2. TheDimensionalSructureOfTheAdopted LstmModel.

3.2.2 Resultsfrom CNNandLSTMModels

TheimplementationwasrealizedunderaJupyternoteboo k environment of Google Colaboratory® by

Pythonprogramming language. With the four model architecturesdescribedintheprevioussection,allthefour modelswere compiled together with the sparse categorical cross entropylossfunctionandtheAdamoptimizerwith alearningrateof 0. 001. All the NN models was fitted for the training data andtest data with a batch size of 64 and run for 50 epochs. Thetrainingaccuracywasthenplottedtogetherwiththevalidationaccuracy varying the iterations for performance evaluationrelatedtothetwoWith respect to the CNN model, a training accuracy of0. 995%wasachieved.

3.3 Object detection and Tracking

The frames are given as input to YOLOv5 (the best version of YOLO is considered for detection). The Bounding box output of YOLOv5 as input to the Object tracking phase. Track Identities is assigned to the detected bounding boxes, trajectory of which needs to be found. The bounding box from the object detection phase is used as reference to analyze the performance metric. Metrics such as false positive, false negative, true positive, true negative, mean average precession, MOTA (Multi Object Tracking Accuracy) and MOTP (Multi Object Tracking Precession) is analyzed to appreciate the accuracy of the detector and tracker.

Hyper Parameters	Values
Input Size	128*128*3
Filter Size	32 (3*3)
Activation	ReLU and softmax
Optimizer	SGD
Learning Rate	0.001
Batch Size	32
Epoch	10
Layers	7

Table 3. Convolutional NeuralNetworkDesign

3.4 ExploratoryDataAnalysisofDataset

First, ATM Image (ATM-I) dataset was loaded in to Jupyter notebookenvironment. Here,severalpythonopensourcelibrarieshavebeen employed in the EDA analysis, including Pandas numpy, Sklearnwithvariousdataprocessingfunctions[11]. Withthehelp of the Pandas library, records with missing probability to occurwithout any biasness. Thus, ATM-V dataset was balanced by selecting the same amount of datarowsforeachofthe2activitieswhichisgraphicallyrepresentedasthenextpiechart.

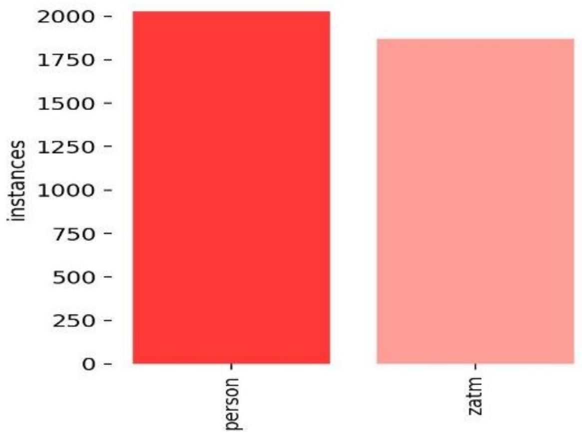

Fig 1 . Total no of classes

Fig 2. SampleImagesoftheDataset

Fig 3. Test-TrainSplit ForObjectDetection

3.4.1 PerformanceMetrics

The choice of performance metrics will influence the analysisof the algorithms. This helps in identifying the reasons formis-classificationssothatitcanbecorrectedbytakingnecessarymeasures.

	Class1 Predicted	Class 2 Predicted
Class1 Actual	True Positive (TP) Correct Decision	False Negative (FN) Type 1 error
Class 2 Actual	False Positive (FP) Type 2 error	True Negative (TN) Correct Decision

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F-1\ Score = 2 * \frac{Precison*Recall}{Precision+Recall}$$

$$mAP = \frac{1}{No.of\ divisions}\Sigma_{r\in(1.0.1.0.001)} P_{interp}(r)$$

Fig 4. Precision And Recall Calculation

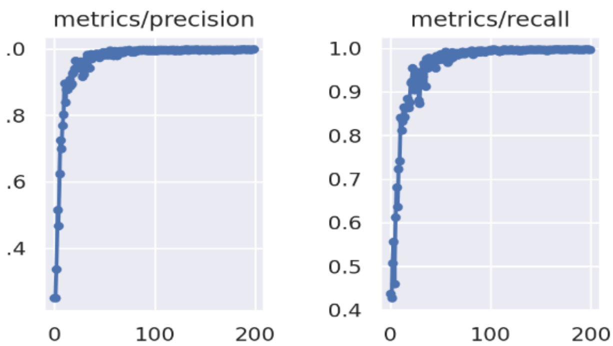

Fig 4. Precision and recall result

3.5 Resultsofobjectdetection andTracking

Object classification is performed on the state-of –the-art network called CNN. The network designed consists of3,697,188 tunable parameters. The Accuracy of the network is gradually increasing, and losscurveisgraduallydecreasingwithincreaseinnumber ofepochs.

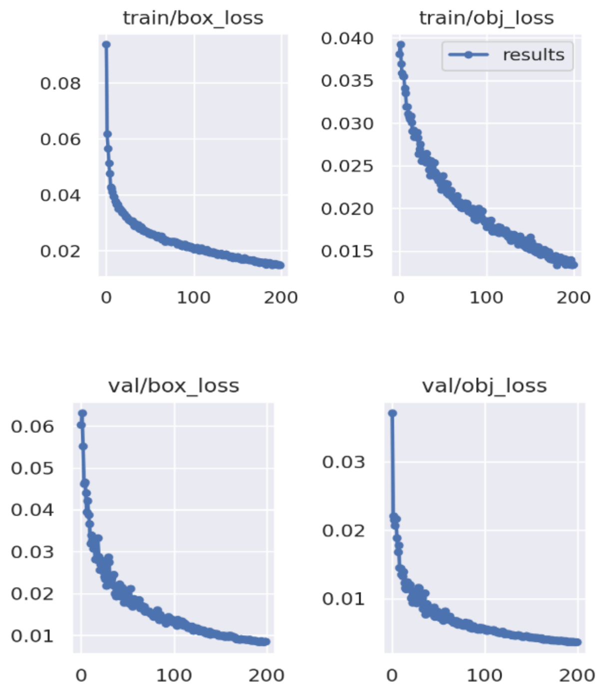

Fig 5. Test-Val Loss

3.5.1 Confusionmatrix

The performance of the classification model is measured usingconfusion matrix.

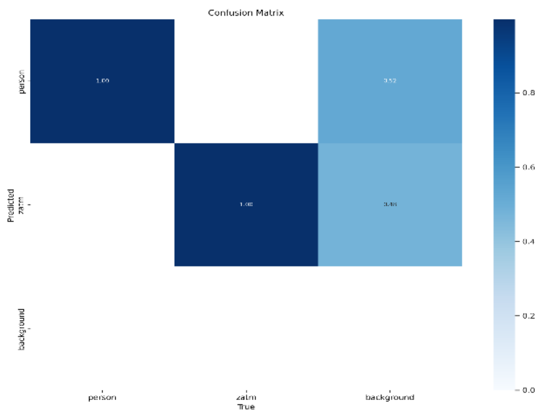

Fig 6. ResultsOfObjectDetectionAnd ClassifierCnnResultsAndAnalysis

Better accuracy and loss values achieved on large datasets andthe model is more generalized when trained on Large Dataset. Theprecisionand recallscoresarecomparableinbothcases

Measure	Value for Small Dataset	Value for Large Dataset
Time forsingleepoch	81 seconds	300 seconds
Training Accuracy	0.9896	0.9936
Validation Accuracy	0.9853	0.9812
Training Loss	0.0295	0.0193
Validation Loss	0.04803	0.06175

Table 5. CnnOutputDetails

Theactivitycolumnwhichisacategoricalvariableintheda tasetwasthenconvertedintothenumericalformat. Forthispurpose, the LabelEncoder function from the Sklearn librarywas used for preprocessing. In the process of feature scaling,allthefeatureswerescaledtobewithinthesameran ge,whichwould guarantee the value manipulations of every featuresequivalentandreweightnaturallytheprediction modelbyrealdependency of the corresponding relevance of the features. Here,theSklearn'sStandardScaler function,whichscaleeachfeature by its maximum absolute value, was used for thescaling.

IV.RESULTS

TheimplementationwasrealizedunderaJupyternoteboo k environment of Google Colaboratory `by Pythonprogramming language. With the four model architecturesdescribedintheprevioussection,allthefour modelswerecompiled together with the sparse categorical cross entropylossfunctionandtheAdamoptimizerwith alearningrateof0. 045. All the NN models was fitted for the training data andtest data with a batch size of 64 and run for 50 epochs. Thetrainingaccuracywasthenplottedtogetherwiththeva lidationaccuracy varying theiterations for performance evaluationrelatedtothetwomodels. With respect to the CNN model, a training accuracy of 99. 5% wassimultaneouslyachieved as result showninFig. 7 and 8.

Fig 7. Normal Activity Detection

Fig 8. Normal Activity Detection

V. CONCLUSION

This research proposes a deep learning-based system for detecting suspicious events in a bank-ATM context in real time. Bounding boxes, which functioned as classes in this case, are utilised to detect tagged items. This is then used to categorise labels in video and forecast whether the occurrences are normal or abnormal. that result is calculated using the Motion representation Depth data is derived from the classes' bounding boxes. Then multi-stream CNNs are used to distinguish constituents andactions. The choosing of an appropriate algorithm for a certain job. There is always a trade-off between speed and precision. The classifier trained on the Indigenous dataset has a validation accuracy of 99. 5%.

It will be a perfecttask if we can generate our owndataset with the use ofappropriate sensors and applications for a defined number offrequent activities people are performing in day to day lives. Thisresearchareaseemshavingmultipleadvancedapplications with Deep Learning applications in near future. In the future, the proposed approach can be evaluated for other real-world outdoor scenarios like railway platforms, shopping malls, etc. Also, for the detection of unwanted objects, deep learning-based object detection models can be combined with the proposed framework for further improvement.

VI. REFERENCES

[1]. Vikas Tripathi; Hindawi Publishing Corporation, "Robust Abnormal Event Recognition via Motion and Shape," Journal of Electrical and Computer Engineering, pp. 1-11, 2015.

[2]. Pushpajit A. Khaire and Praveen Kumar, "RGB+D and deep learning based real time detection of suspicious," Springer; Journal of Real-Time Image Processing, pp. 1-13, 2021.

[3]. P. A. Khaire, "RGB+D and deep learning based real time detection of suspicious," Journal of Real-Time Image Processing, pp. 1-13, 21.

[4]. C. Shiranthika, "Human Activity Recognition Using CNN & LSTM," IEEE, 2021.

[5]. T. S. Bora, "HUMAN SUSPICIOUS ACTIVITY DETECTION SYSTEM USING CNN MODEL FOR VIDEO SURVEILLANCE," IJARIIE, 2021.

[6]. R. Vrskova, "A New Approach for Abnormal Human Activities Recognition," Sensor, 2022.

[7]. S. Sabbu, "LSTM-Based Neural Network to Recognize Human Activities," Hindawi, pp. 1-8, 2022.

[8]. Rajeshwari S, Vismitha G, Sumalatha G and Safura Aliya, "Unusual Event Detection for Enhancing ATM Security," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, pp. 1-6, 2021.

[9]. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognitionusing cell phone accelerometers," SIGKDD Explor. Newsl. , vol. 12,no. 2,pp. 74–82,Mar. 2011,doi:10. 1145/1964897. 1964918.

[10]. A. Murad and J. -Y. Pyun,"Deep RecurrentNeuralNetworksforHuman Activity Recognition," Sensors, vol. 17, no. 11, p. 2556, Nov. 2017,doi:10. 3390/s17112556.

[11]. P. Kuppusamy and C. Harika, "Human Action Recognition using CNNandLSTM-RNNwithAttentionModel"International Journal odInnovativeTechnologyandExploringEngineering(IJITEE),vol. 8,Issue8,pp. 1639-1643,2019

[12]. https://www. analyticsvidhya. com/blog/2022/03/basics-of-cnn-in-deep-learning/

[13]. Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM NetworksforMobileHumanActivityRecognition, "presentedatthe2016International Conference on Artificial Intelligence: Technologies andApplications,Bangkok,Thailand,2016,doi:10. 2991/icaita-16. 2016. 13.

[14]. https://ieeexplore. ieee. org/document/9043972

[15]. https://towardsdatascience. com/convolutional-neural-networks-explained-9cc5188c4939

[16]. C. Jobanputra,J. Bavishi,andN. Doshi,"HumanActivityRecognition:A Survey," Procedia Computer Science, vol. 155, pp. 698–703, 2019,doi:10. 1016/j. procs. 2019. 08. 100.

[17]. https://deepai. org/publication/evaluating-two-stream-cnn-for-video-classification

[18]. https://www. codeproject. com/Articles/1366433/Using-Modified-Inception-V3-CNN-for-Video-Processing

[19]. https://www. kaggle. com/datasets/mehantkammakomati/atm-anomaly-video-dataset-atmav

[20]. A. Murad and J. -Y. Pyun,“Deep RecurrentNeuralNetworksforHuman Activity Recognition,” Sensors, vol. 17, no. 11, p. 2556, Nov. 2017,doi:10. 3390/s17112556.
[21]. T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, “Human activityrecognition from inertial sensor time-series using batch normalizeddeep LSTM recurrent networks,” in 2018 40th Annual InternationalConference of the IEEE Engineering in Medicine and Biology Society(EMBC),Honolulu,HI,Jul. 2018,pp. 1–4,doi:10. 1109/EMBC. 2018. 8513115.
[22]. https://github. com/pjreddie/darknet/blob/master/data/coco. names
[23]. M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, “Real-time anomaly detection and localization in crowdedness,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.
[24]. C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab ,” in Proceedings of the IEEEinternational conference on computer vision, 2013.
[25]. Lu, S. (2019). Deep learning for object detection in video Journal of Physics Conference Series, 1176.
[26]. Simonyan, K. , Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos.

Cite This Article :

Anjali Suthar, Prof. Jayandrath Mangrolia, Prof. Ravi Patel, "Abnormal Activity Recognition in Private Places Using Deep Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.321-328, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT228688
Journal URL : https://ijsrcseit.com/CSEIT228688