RESEARCH ARTICLE OPEN ACCESS

Abnormal Activity Recognition in Private Places UsingDeep Learning

CH.VIJAYAJYOTHI[1]
Asst.Prof.,
Holy Mary Institute of Technology and Science
DR S.SRINIVAS[2]
Associate Prof.,
Holy Mary Institute of technology and Science
Dr.NARASIMHA CHARY CH[3]
Associate professorDept
of CSE
Sri indu college of engineering and technology(Autonomous).

Abstract:

applying computer and engine delusion technology, the process of analyzing natural shift is known as"natural exercise recognition," or HAR. Anomaly discovery in cover systems is one of the situations in which natural conditioning recognition is useful. As the demand for screen growing, charge cameras have been generally instated as the foundation for tape anatomy. connecting anomalous actionsdemands aggressive mortal exertion, which is one of the main obstacles in charge tape anatomizing. It's necessary to establish tape recording recording in order to automatically catch anomalous exercise. operating deep education hows, our alert vid care system can identify an anomaly in a vid. Real- time unearthing of the bearing is also achievable, and these videotape frames will be latterly saved as pics in the system for the doper to examine. The indicated atypical exertion Recognition system was created with the aim of connecting and detecting irregularities through a live feed in the banking section, more specially in an ATM contexture. The original phase of the study focuses on the employment of image deep education fashions to fete chromatic points and spot unusual actions applying ATM covering 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 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 Convolution 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 suspiciousactivity—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. Conventional 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 data set 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

II.LITERATURE REVIEW

In this section, we present the related work and research undergone in developing video based 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[1] 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[3] 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 recognizing human actions even in large- scale films.

III.PROPOSED SYSTEM

With the literature review been conducted, it was revealed that the Deep Learning Models have been widelyused resulting better scales of accuracy and to serve the Human Activity Recognition process revealed that the Deep Learning Models have been widely used resulting better scales of accuracy and to serve the Human Activity Recognition process. Data set ATM Image[6] (ATM) comprises 1491 images that cover most of the angles in which an ATM box can be viewed in an ATM vestibule. Images in the are augmented with blur (up to 2. 25px) and noise (up to 6% of pixels) effects. Augmentation is done to expand the data set and increase model performance. The image dateset has been created where each image is bounding box annotated for the ATM and person class. Second freely available dateset is ATM Anomaly Video (ATMA-V)[9] Dateset from Kaggle. The video comprises 65 videos that consist of both anomalous and normal video segments.As part of our abnormal behavior classification[11], a dateset 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[7] the that activates are normal or abnormal.

CNN Architecture

The CNN model was defined as having two CNN hidden layers.Each of the mare followed by two dropout layer so connected layer is used to interpret the features extracted by the CNNhidden layers. Finally,plate layer with the soft m was add as the final layer to make predictions(TableI).The sparse categorical cross entropy loss function will be used as the loss function and the efficient Adam version of stochastic gradient descent was used to optimize the network with a learning rate of 0. 001. CNN model was trained for 50epochs and a batch size of 64 samples were used. After the model is fit, it was evaluated on the test dateset and the accuracy of the CNN[14] model was obtained

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

Table 2. The Dimensional Sructure Of The Adopted Immodest.

Python programming language. With the four model architectures described in the previous section,altogether models were compiled together with the sparse categorical entropy loss function and the Adam optimizer with nonthreatening 0. 001. All the NN models was fitted for the training data and- test data with a batch size of 64 and run for 50 epochs.The training accuracy was then plotted together With the varying the iterations for performance evaluation related to the two With respect to the CNN model, a training accuracy of0. 995%was achieved.

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 procession, MOTA (Multi Object Tracking Accuracy) and MOTP (Multi Object Tracking Procession) 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. Convolution Neural Network Design

Exploratory Data Analysis of Data-set

First, ATM Image (ATM-I) data set was loaded into Jnote book environment. Here, several python open source libraries have been employed in the EDA analysis, including Pandas

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, 358Trainableparams:162,358Non-trainableparams:0		

Table1. The Dimensional Structure Of The Adopted Cnn Model.

IV Result Analysis

Architecture

The LSTM model was defined as having a single LSTM[12] hidden layer. Dropout layer valuing0. 5followsthis. Serenade fully connected layer is used to interpret the features extracted by the single LSTM hidden layer. Finally, a dense layer was added as the final layer to make predictions.

Performance Metrics

The choice of performance metrics [13] will influence the analysis of the algorithms. This helps in identifying the reasons form is-classifications so that It can be corrected by taking necessary y measures.

	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

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

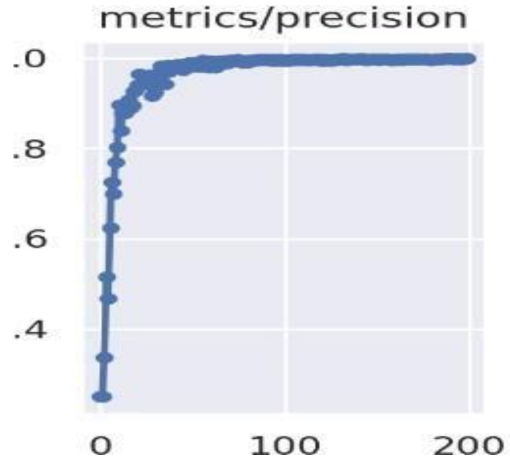
$$Recall = \frac{TP}{TP+FN}$$

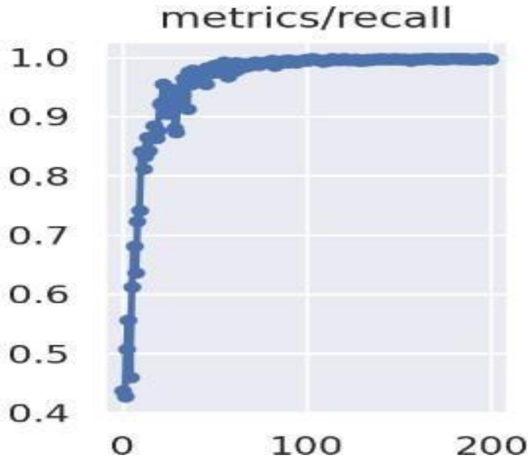
$$F-1\ Score = 2 * \frac{Precison*Recall}{Precision+Recall}$$

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

Results of object detection and Tracking

Object classification is performed on the state-of –the- art network called CNN. The network designed consists of 3,697,188 tunable parameters. The Accuracy of the network is gradually increasing, and loss curve is gradually decreasing with increase in number of epochs.





Confusion matrix

The performance of the classification model is measured using confusion matrix. Results Of Object Detection And Classifier Cnn Results And Analysis[16] Better accuracy and loss values achieved on large datasets a n d t h e model is more generalized when trained On Large Dateset. Non depreciating recall scores are comparable in both cases

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

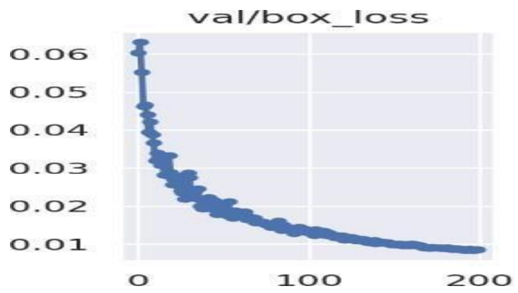
Table 6. Cnn Output Details

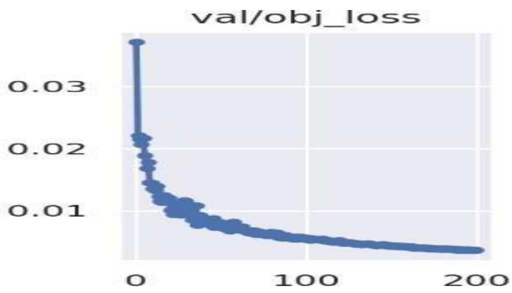
The activity column which is categorical variable In the data set was then converted into then numerical format.

CNN Architecture

The CNN model was defined as having two CNN hidden layers.Each of the mare followed by two dropout layers of

Then a dense fully connected layer is used to interpret the features extracted by the CNN hidden layers.Finally,plate layer with the soft max activation function was added as the final laye rto make predictions(TableI).The sparse categorical cross entropy loss function will be used as the loss function and the efficient adam version of-stochastic gradient descent was used to optimize the network[15] with a learning rate of 0. 001. CNN





model was trained for 50epochs and a batch size of 64 samples were used. After the model is fit, it was evaluated on the test dateset and the accuracy of the CNN model was obtained For the purpose of compiling and training the same values for the loss function, optimizer, batch size and the number of epochs, which we used, in compiling and training the CNN model we reused. After the model is fit, it was evaluated on the test dateset and the accuracy was obtained.Multipurpose, the Labe l Encoder function from the Sk learn library was used for prepossessing. In the process of feature scaling, all the features were scaled to be within the same range, which would guarantee the value manipulations of every features equivalent and reweigh naturally the prediction model by real-dependency of the corresponding relevance of the features .Here the Sklearn's Standard Scaler function, which scale each feature by its maximum absolute value, was used for the scaling. in real time. Bounding boxes, which functioned as classes in this case, are utilized to detect tagged items. This is then used to categories 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 and actions. 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 dateset has a validation accuracy of 99. 5%.It will be a perfect task if we can generate with the use of appropriate sensors and applications for a defined number of frequent activitie speople are performing in day to day lives. This research are a seems having multiple advanced applications with Deep Learning[16] 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.

V.REFERENCES

[1]. Vikas Tripathi; Hindawi Publishing Corporation, "Robust Abnormal Event Recognition viaMotion 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 detectionof 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 ofReal-Time Image Processing, pp. 1-13, 21.
[4]. R. Shobarani, R. Sharmila, M. N. Kathiravan, A. A. Pandian, C. Narasimha Chary and K. Vigneshwaran,"Melanoma Malignancy Prognosis Using Deep Transfer Learning," 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169528..
[5]. T. S. Bora, "HUMAN SUSPICIOUS ACTIVITY DETECTION SYSTEM USING CNN MODELFOR 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," InternationalJournal of Innovative Research in Electrica Electronics, Instrumentation and Control Engineering, pp. 1-6, 2021.
[9].Chary, Ch Narasimha, et al. "An Efficient Survey on various Data Mining ClassificationAlgorithms in Bioinformatics." International Journal of Engineering and Techniques 4 (2018).
[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].gupta, k. Gurnadha, ch narasimha chary, and a. Krishna. "Study onhealth care life log by the level of care required using keygraph technology in text data mining."
[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"LSTMNetworksforMobileHumanActivityRecognition,"presentedatthe2016International Conference on Artificial Intelligence: Technologies and Applications ,Bangkok,Thailand,2016,doi:10.2991/icaita-16. 2016.
[14]. BHUSHAN, P. V., NITESH, V., CHARY, C. N., & GUPTA, K. G. Novel Approach for Multi Cancers Prediction system using Various Data Mining Techniques.
.[15]. https://towardsdatascience. com/convolutional-neural-networks-explained-9cc5188c4939
[16].C. Jobanputra,J.Bavishi,andN. Doshi,"HumanActivityRecognition:A Survey," ProcediaComputer 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