iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45921

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Irregular Events Detection in Videos Using Machine Learning Techniques

Rahul C[1], Merin Meleet[2]
[1, 2]*Information Science and Engineering, RV College of Engineering, Bangalore, India*

Abstract: ***Video analytics for detecting events using machine learning is designed and developed to analyze and detect patterns in the videos. Especially in the field of criminal forensics where a video needs to be analyzed to find out what abnormal events are happening in it and who caused it and how it was caused. This is an easy task for a human as they can recognize criminal events easily but not machines. The objective is to automatically detect the irregular events in videos like burglary, fighting, arson and explosion using a CNN model by preprocessing the videos into frames and extracting information from these frames and also to train the model so that it can also detect normal events. The results of this work shows that the model detects irregular events in videos with high accuracy.***

Keywords: ***Convolutional Neural Network, Adam optimizer, Exploratory Data Analysis, Video preprocessing, Normalization.***

I. INTRODUCTION

In intelligent video surveillance, irregular event detection and localization is a very recent idea and also a difficult research challenge. It is made to detect unusual occurrences while monitoring videos automatically. The fundamental challenge with this work is that in training video sequences, there is only one class termed normal event which is used to detect abnormal situations happening in the video. However, numerous powerful algorithms based on hand-crafted characteristics or features have been presented in recent years. Only a few algorithms use high-level features, but practically all of them employ complicated multistage learning to arrive at the solution needed. It is these systems that are used to analyze the videos to detect paranormal activities. The idea of these video analytics may have been around for a long time but it is with the recent innovations in the field of analytics that progress has been made. The scope of event detection is its sheer versatility in implementing it through various industries for getting a safe and secure environment for people to work and live in. Especially in the areas of remote monitoring and surveillance these systems are highly appreciated. And in the field of criminal surveillance like arson, robbery, fighting etc. Analysis of these results will tell us which events have been detected in the model. Solutions for analyzing videos in one stop where the system provides a high level, monitoring and early alarms for having a safe secure environment. Recording these transactions/events for future deep analysis. Hence the scope of these surveillance is mainly intended for economical, easier and faster detection services for both industries as well for the public. By implementing machine learning models like CNN where both normal and abnormal events are detected. Criminal activities like arson, burglary, bomb explosion and fighting are termed as irregular events or crime events which are caught by surveillance cameras an be analysed and detected.

This work mainly helps in video analytics in the forensics field where there is a constant need to analyze every frame for criminal patterns which is helpful in identifying the crime and the culprits. This also eliminates the long work put in by manual analysis by humans and replacing it with machine learning models.

II. RELATED WORK

R. Samet and E. Sengonul indicated in their work [1] that before being utilized to train the model in the Convolution Autoencoder architecture, which is based on the Long Short-Term Memory Network, extracted video frames passed through a series of processes (LSTM). If there are no labels on the data and when a semi-supervised learning model was created utilizing typical video images as training. The input data is utilized to structure the trained model, and the model is tested with new frames. The computed regularity score was used to assess the tested frames adherence to the model. The outputs of the background removed frames-trained model were compared to the outputs of the conventional frames-trained model. According to the findings, removing the background enhances the accuracy of recognizing abnormal events.

J. Yu and Y. Lee propose a method [2] that forces AEP to learn representations for predicting future events while limiting representation learning for previous occurrences. AEP may generate a discriminative model to identify an anomaly of events without complementing information, such as optical flow and explicit abnormal event samples in the training stage, by applying the suggested adversarial learning.

©IJRASET: All Rights are Reserved

They use criminal data sets to show how effective AEP is at spotting abnormalities in occurrences. Performance analysis based on hyperparameter settings and comparisons with existing state-of-the-art approaches are among the experiments. The experimental findings suggest that the proposed adversarial learning may aid in the development of a better model for normal events on AEP, and that AEP trained using the proposed adversarial learning can outperform current approaches.

People identification, head-torso template extraction, tracking, and swarm cluster analysis are among the four modules proposed by A. Jalal and K. Kim [3]. To begin, the system uses an inverse transform and a median filter to extract human contours, lowering the cost of computation and managing numerous difficult monitoring circumstances. Second, because the head torso is less variable and obscured, persons are recognised by their head torso. Finally, utilizing Kalman filter techniques with Jaccard similarity and normalized cross-correlation, each individual is tracked across many frames. Finally, for normal/abnormal event identification, the template marking is applied for crowd counting with cues localisation and clustering by Gaussian mapping. The findings of two datasets show that the suggested method delivers 88.7% and 95.5 percent counting accuracy and detection rate, respectively.

L. Qing and Y. Mengqiu [4] researches the identification and localization of aberrant targets in monitoring video using a deep neural network, and the results of video content recognition are more succinct and abstract, which can better suit the development demands of information networks with a large number of videos. The goal of video anomaly detection is to quickly identify unusual occurrences in a huge number of recordings, assuring public safety and avoiding potentially harmful circumstances. An artificial intelligence system should be able to learn on its own and extract characteristics from the data it is given. Aiming to meet the demand for video monitoring systems that can analyze massive data in real time.

J. Du, H. Zhu [5] are introducing a new long short-term memory called the adaptive iterative hard-thresholding method (adaptive ISTA)., J. Du and H. Zhu address some disadvantages (e.g., nonadaptive updating) of existing sparse coding optimizers and embrace the merits of neural networks (e.g., parallel computing), we design a novel recurrent neural network to learn sparse representation and dictionary (LSTM). To our knowledge, this is one of the first studies to link the solver with LSTM, and it may bring new insights into LSTM, model-based optimization (also known as named differentiable programming), and sparse coding-based anomaly detection. Extensive studies demonstrate their method's performance in the abnormal event detection challenge.

To recognise human subjects in the image and derive the body key points, T. Gatt proposes [6] using pre-trained posture estimation algorithms. In a semi-supervised method, such data is utilized to train 2 kinds of Autoencoders centered on LSTM and CNN elements with the purpose of learning a generic representation of typical behavior. With just an aggregate F-score of 0.93, both frameworks were able to accurately identify between normal and aberrant data sequences when tested on a tough realistic video dataset. It was also discovered that posture calculated data matches sensor data quite well. This demonstrates that posture estimated data might be useful in understanding and categorizing human behaviors.

III. SYSTEM DESIGN AND ARCHITECTURE

A System Architecture is a model that identifies a system's structure, behavior, components, and other aspects. Architecture is a representation of the system and its components that will work together to realize the overall framework. Data is unprocessed information that is obtained and a dataset is formed and collected from several datasets. For training and validation, these datasets are utilized. The dataset is a collection of videos called UCF crime dataset. The architecture [11] strategy involves the implementation and training of the CNN model to accurately detect which type of event is happening in the video. Optimization of the model and the data preprocessing stages are given careful observations to get a higher accuracy score.

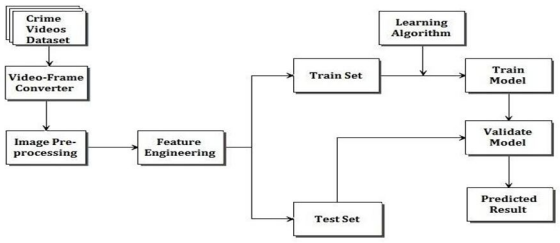


Fig 1 :System architecture

Fig 1 shows the relationships and interactions between different components within the system. where a CNN model is implemented. The classification model belongs to supervised types of machine learning methods. The neural network is trained with the processed dataset obtained after applying different pre-processing methods like image resizing, colour conversion, one hot encoding and normalization then the model is able to classify and detect the events by analyzing the frames obtained from splitting the videos in the dataset.

IV. METHODOLOGY

The different phases required are:

Data Collection: The first task is to collect the datasets that consist of surveillance videos where both normal and irregular events are consolidated in it. This dataset in particular has a large number of surveillance videos in mp4 format. And the videos are in grayscale and rgb format as well. Each of the video has a maximum length of 3 minutes under this duration criminal events like burglary, arson, bomb explosion or fighting takes place. The dataset is an open-source dataset available to download on the internet called the UCF crime dataset.

Data Pre-processing: In order to develop the system, the next task is crucial and plays an important role in getting high quality results. Here the data obtained from datasets must be pre-processed to remove the outliers, noise to improve the quality of video so that feature extraction can take place. Since the dataset consists of videos some pre-processing techniques like resizing of frames, colour conversion and video format conversion are done so that all dataset are in one format only this is done through the use of python libraries and software tools available on the internet. Also, the resolution and format of the entire video dataset must be set into a single or similar size, pixel resolution and format to make the feature extraction process easier.

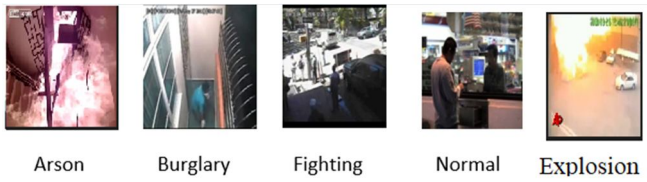


Fig 2 : Dataset format

Figure 2 gives a peek of how the dataset that is collected will look like.

Next class labels are created to store the values of the classes that are to be detected and here there are five labels (burglary, arson, bomb, explosion, fighting and normal) each of which is assigned a number from 0 to 4 to identify and differentiate between the classes.

Then the videos are split into frames and each of these frames are stored in one location so that they can be accessed easily and also each of these frames are tagged with an identity so that the frames from each video can be separated making it easier for the CNN model to access. The path of these frames is saved into a separate csv file and it is this file that is used to validate the output from the model. Next the frames themselves undergo pre-processing to resize it and to ensure that all the frames are in one jpg format which is given as input for CNN.

Model Training & Evaluation: Next a convolutional neural network will be built to analyse the data obtained from the previous phase and then in order to build a relationship between the features for analysis to take place. Extraction of the features will take place next inside the neural network. The pixels in the frames are converted into a numerical value and an array is built for each frame and these numerical values of the frames are then normalized to give more accuracy for prediction. One hot encoding is done on the array of frames so that each individual frames of different videos which are grouped together can be distinguished from each other.

Extraction of features takes place in the convolutional layer. Pooling layer reduces the dimension of the parameters captured in the CNN. Dropout layer is added to remove parameters after successive convolutions. Flatten and dense layer compiles to give output from the previous layers. Optimizers fine-tuned for the model reduces the loss during the training phase.

Fine Tuning: As it is a machine learning method, parameter estimates can range from hundreds to millions, therefore training it on a tiny dataset from the start will only result in overfitting. As a result, it is usually preferable to start with a pre-trained model and then fine-tune the model using a comparatively tiny dataset that is relevant to the domain. That is why the frames undergo preprocessing so that it fits into the CNN model to increase the efficiency of the model. Fine tuning for the neural network model is also done by defining the features of the network like the relu function introducing bias for the neural network and also by defining the weights for the model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Classification: Detection is done for different types of events where more than one scenario type of burglary, arson, fighting, explosion and normal can be differentiated from the normal events and are identified. This method also acts as a reference for using more complex features in classification and hybrid models for future implementations.

Results Generation: Next is the generation of results where irregularities present in the footage are detected and identified showing that the irregular events have been detected. And the categorical metrics used to validate the model gives the overall accuracy and loss values obtained at different epochs and then the confusion matrix of the model is implemented. The confusion matrix plotted against the actual events versus the predicted events gives an idea about the number of predicted events and its numbers for different events needed for detection. It also shows the accuracy of irregular events detected.

V. RESULTS

The performance of each model is analyzed in the given table .

Table 1: Performance Analysis

Epoch	Loss	Accuracy	Val_Loss	Val_Acc
1	1.46	0.51	0.75	0.77
2	0.47	0.82	0.28	0.87
3	0.22	0.89	0.16	0.92
4	0.13	0.95	0.09	0.95
5	0.08	0.97	0.07	0.98

Table 1 demonstrates the accuracy score of the model at different epochs and it is observed that the neural network model gives the accuracy of 97 percent with the loss or error in detecting events being very small.

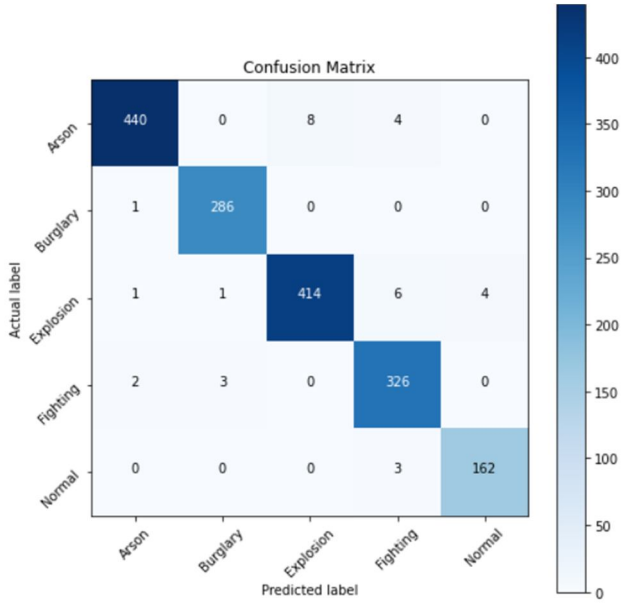


Fig 3 : Confusion matrix

The above figure shows the values of predicted events versus actual events where the accuracy achieved by the model is very high with a small negligible amount of errors or wrong event detection happening across the different classes.

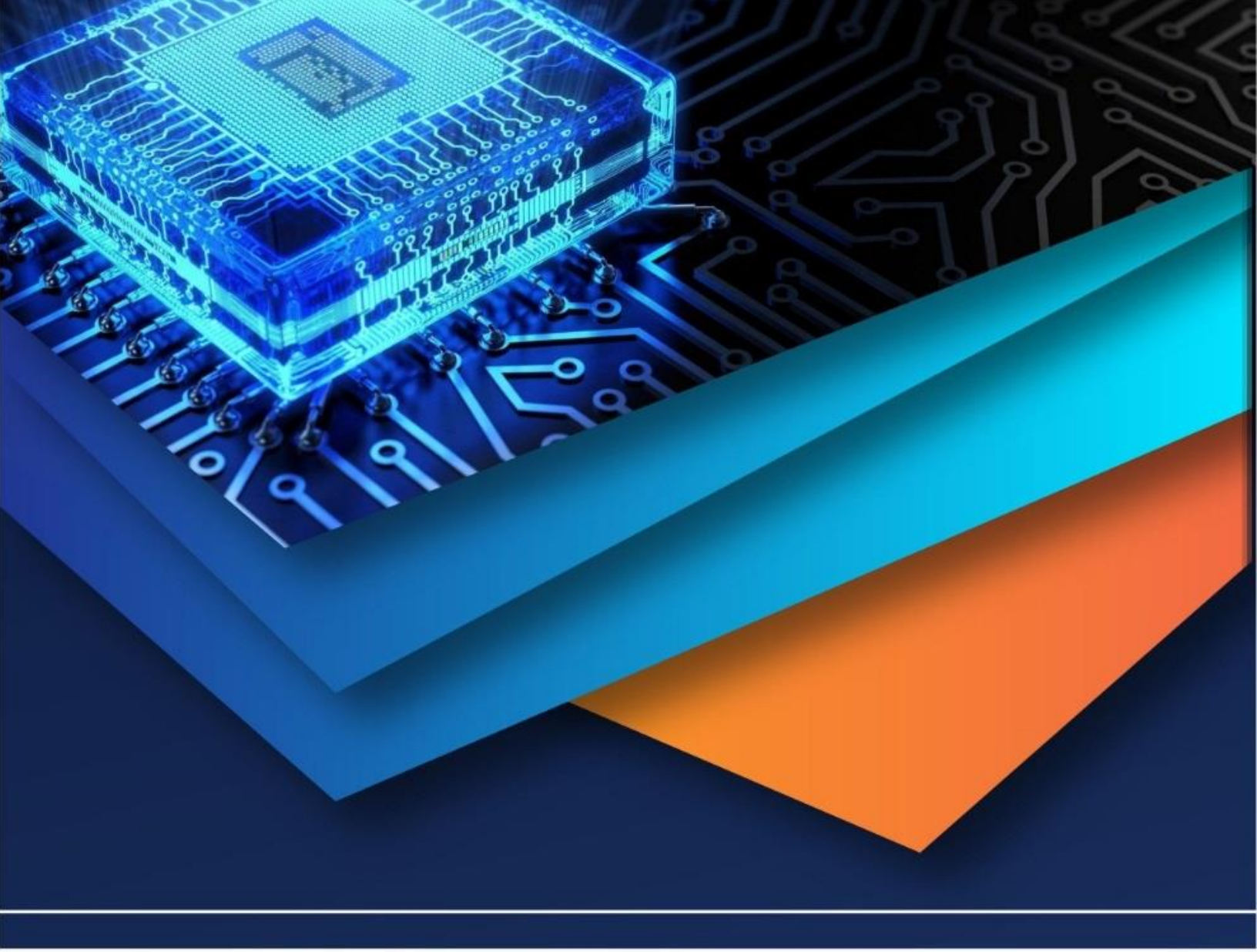
VI. CONCLUSION

The aim of this work is to build a system that is capable of identifying the irregular events that occurred in the video by utilizing machine learning techniques and neural networks to extract features and classify them so that the events can be categorized, identified and detected. This work deals with the implementation of such a system which gives a high accuracy of above 95 % percent which is more than satisfactory considering the limitations. It also explains in detail about the workings of the detection model. It also records the attributes of the model for future events analysis as well

REFERENCES

[1] R. Samet and E. Şengönül, "The Effect of Background Subtraction on Abnormal Event Detection in Video Surveillance," In proceedings of 29th Signal Processing Communications Applications Conference, 2021, pp. 1-4.
[2] L.Zhang, X.Zong,et.al, "Abnormal Event Detection in Video Based on SVDD," In proceedings of 10th IEEE International Conference on Intelligent Data Acquisition Advanced Computing System: Technology and Applications, 2019, pp. 368-371.
[3] A. Shehzed, A. Jalal and K. Kim, "Multi-Person Tracking in Smart Surveillance System for Crowd Counting and Normal/Abnormal Events Detection," In proceedings of International Conference on Applied and Engineering Mathematics, 2019, pp. 163-168.
[4] L. Qing, Y. Mengqiu, et.al, "Research on Video Automatic Feature Extraction Technology Based on Deep Neural Network," In proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 954-957.
[5] T. Gatt, D. Seychell and A. Dingli, "Detecting human abnormal behavior through a video generated model," In proceedings of 11th International Symposium on Image and Signal Processing and Analysis, 2019, pp. 264-270.
[6] C. Direkoglu, "Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks," in IEEE Transactions on Information Forensics and Security, vol. 8, pp. 80408-80416, 2020.
[7] O. Ye, J. Deng,et.al, "Abnormal Event Detection via Feature Expectation Subgraph Calibrating Classification in Video Surveillance Scenes," in IEEE Access, vol. 8, pp. 97564-97575, 2020.
[8] R. J. Franklin, Mohana and V. Dabbagul, "Anomaly Detection in Videos for Video Surveillance Applications using Neural Networks," In proceedings of Fourth International Conference on Inventive Systems and Control (ICISC), 2020, pp. 632-637.
[9] H. Jain, A. Vikram, et.al, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," In proceedings of International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 193-198.
[10] C. Wu, S. Shao, C. Tunc and S. Hariri, "Video Anomaly Detection using Pre-Trained Deep Convolutional Neural Nets and Context Mining," 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), 2020, pp. 1-8.
[11] A. M.R., M. Makker and A. Ashok, "Anomaly Detection in Surveillance Videos," In proceedings of 26th International Conference on High Performance Computing, Data and Analytics Workshop, 2019, pp. 93-98.
[12] A.M.R., M. Makker and A. Ashok, "Anomaly Detection in Surveillance Videos," In Proceedings of 26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW), 2019, pp. 93-98.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)