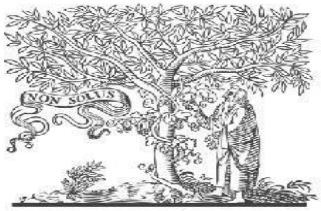
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/17

Title **ANALYSIS OF ABNORMAL ACTIVITY DETECTION IN OFFLINE SURVEILLANCE FOOTAGE**

Volume 12, ISSUE 04, Pages: 123-132

Paper Authors

K. Mohan Krishna, Karasani Mani Sai Lakshmi, Manduri Bhanu Harshitha, Kukkadapu Amitha, Konagalla Tarun

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Analysis of Abnormal Activity Detection In Offline Surveillance Footage

K. Mohan Krishna[1], M. Tech, Associate Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
Karasani Mani Sai Lakshmi[2], **Manduri Bhanu Harshitha**[3], **Kukkadapu Amitha**[4], **Konagalla Tarun**[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1,2,3,4,5] mohankrishnakotha@gmail.com[1], manislk.20@gmail.com[2],
bhanuharshitha02@gmail.com[3], amitha30k@gmail.com[4], konagallatarun@gmail.com[5]

Abstract
In public places, Abnormal activities are happening more frequently. The surveillance footages capture those actions. If an incident occurs, it will take a long time to watch the entire video in order to spot any unusual activity. So, by presenting a system to recognise activity within a fraction of time, one can save the time to detect an abnormal activity which is held at a specific moment. The proposed system keeps track of instances that show patterns of different human activities. This research focuses on a deep learning approach for detecting abnormal human activities in videos which include robbery, car accidents, and fighting. The system is implemented using pre-trained models such as VGG-16, ResNet50 and 2D-convolutional neural network (CNN). By merely providing a video as an input to the proposed system, it quickly recognises the abnormal activities present in the video.

Keywords: Identifying suspicious activity - robbery, road accident and fighting, Video surveillance system, VGG16, ResNet50, CNN.

Introduction

Video surveillance systems (VSS) play a significant role in contemporary industrial and urban contexts. Today, surveillance cameras are everywhere, from private homes for burglary protection to security-sensitive locations like borders and military posts for tracking terrorist activity. Urban environment becomes more crowded, with the extensive use of multistorey buildings and the increase of vehicle, pedestrian, and crowd flow.

The goal of moving object detection is to identify moving items in video sequences that are of interest. The difficulties with changing ambient conditions make detecting moving objects particularly difficult. Background subtraction, temporal frame differencing, and optical flow are methods that are frequently employed to detect moving objects. Object tracking comes next in the video analysis process. The establishment of temporal connection between identified objects from frame to frame is another approach to look at object tracking. Surveillance cameras are used to monitor public and private spaces and to identify people, as it is becoming both more pervasive and more invasive.

Abnormal Activities

Abnormal activities are very broad and different. The authors define “abnormal

activities" as "activities that occur rarely and have not been expected in advance". This definition can vary depending on the field or context being studied. The categories of aberrant activity are similar to those of normal activity and include gestures, simple acts, events, interactions, behaviours, and group actions.

Fig.1. Abnormal human activity- fighting

Fig.2. Abnormal human activity – robbery

Fig.3. Abnormal human activity – road accident

Literature Survey

There are several ways to detect abnormality but some techniques fail due to less performance and some are gained more usage due to its better accuracy. We looked at a wide range of research and conference publications on the identification of anomalies. we had gone through many methodologies and approaches those scientists had employed in their research.

Chao Huang, Zhihao Wu, Jie Wen, Yong X, Qiuping Jiang and Yaowei Wang proposed TAC-Net a novel approach method for Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System [1].

Kshitija Deshmukh, Aishwarya Kokane, Shweta Konde, Guide-Gauri Virka put forward Abnormal Activity Detection from Video using SVM Algorithm [2].

ABID MEHMOOD presented Pre-Trained2D Convolutional Neural Networks for Efficient Anomaly Detection in Crowd Videos [3].

Sathyajit Loganathan, Gayashan Kariyawasam, Prasanna Sumathipala used Faster R-CNN to identify Suspicious Activity Detection in Surveillance Footage [4].

These scientific papers assisted us in developing a distinct perspective and organising our analysis of the working process for our research project.

Methodology

We are using three separate models in the method to identify the videos of anomaly. The models are convolutional neural networks (CNN), architecture VGG16 and ResNet50. In this project, we use videos of real-time anomalies to train the pattern. We construct the model CNN, VGG16 and ResNet50 independently but we use the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

same data for all models. So that we can able to evaluate each model's performance independently and use the model to predict anomalies with the highest degree of accuracy.

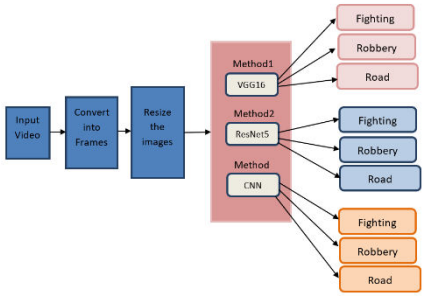


Fig.4. Block diagram of recognition of human suspicious activity using 2D-CNN, VGG16 and ResNet50.

The Proposed Approach

Visual Geometry Group (VGG-16)

It is a 16-layer convolution neural network (CNN) model. This model was proposed by K. Simonyan and A. Zisserman from Oxford University and presented in the paper called very Deep Convolutional Networks for Large-Scale Picture Recognition. 2014's ILSVR (Imagenet) competition was won using the convolution neural net (CNN) architecture VGG16. It is one of the best vision model architectures to date, according to many.

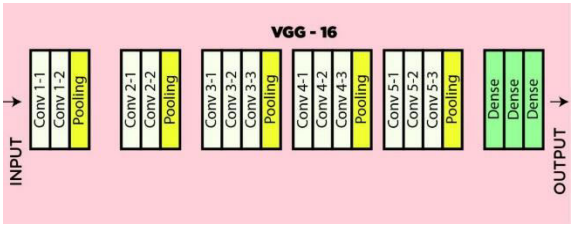


Fig.5. VGG16 Architecture

Convolutional Neural Network (CNN)

A CNN is a particular type of network design for deep learning algorithms that is utilised for tasks like image recognition and pixel data processing. For categorising time-series, signal, and audio data, they can also be highly useful. 1D convolution or just convolution is the term used to describe convolution involving one-dimensional signals. A tensor of outputs is produced by a 2D convolution layer by creating a convolution kernel that is convolved with the layer input. The usage of 2D convolution filtering can be applied to a wide range of image processing goals, examples of some of these include edge detection, texture analysis, image sharpening, and image smoothing.

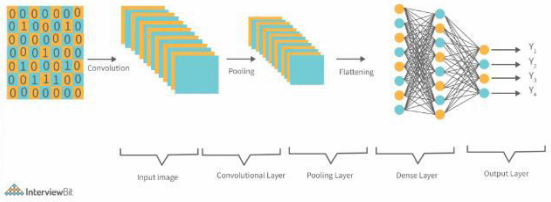


Fig.6. CNN Architecture

ResNet50

Convolutional neural network ResNet-50 has 50 layers total. ResNet, which stands for Residual Networks, is a well-known neural network that serves as the backbone for many computer vision applications. A pretrained version of the network that has been trained on more than a million photos is available for loading from the ImageNet database. The trained network is able to categorise photos into 1000 different object categories, including several different

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

animals, a mouse, a keyboard, and a pencil.

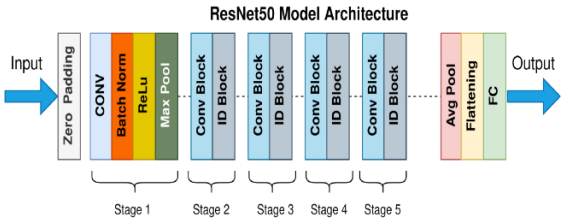


Fig.7. ResNet50 architecture

Experiment

1.Datasets

The required dataset for the project was gathered from the Kaggle datasets, an online data science and machine learning engineering community. Kaggle is a completely free dataset available in google. Over 50,000 publicly accessible datasets and 400,000 publicly accessible notebooks are listed on the Kaggle website. On Kaggle, a fresh dataset is posted each day. The collection includes films captured by surveillance cameras that fall into one of the following 13 categories: abuse, arrest, arson, assault, accident, burglary, explosion, fighting, robbery, shooting, stealing, shoplifting, and vandalism. We only took the three classes Road Accident, Fighting, and Robbery in our project.

Fighting folder has eight directories, each of which contains 32 files with the .mp4 file extension. Robbery folder has 105 directories, each of which contains 32 files with the .mp4 file extension. Road accidents has 76 directories, each of which contains 32 files with the .mp4 file extension.

	Number of files	Number of folders	Size	Size on disk
Fighting	256	8	14.7 MB	15.2 MB
Road Accidents	2432	76	82.0 MB	86.8 MB
Robbery	3359	105	163 MB	169 MB
Total	6047	189	259.7 MB	271 MB

Table.1. Datasets information

1.1 Kaggle datasets for abnormal activities

Video Frames for Fighting

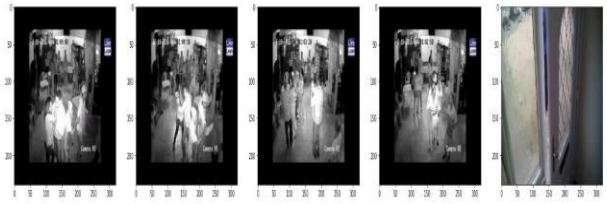


Fig.8. Fighting Frames

Video Frames for Robbery

Fig.9. Robbery Frames

Video Frames for Road Accidents

Fig.10. Road Accidents

2. Pre - Trained model

A pre-trained model is a machine learning model that has already undergone training on a large dataset for a particular job, such as speech recognition, picture

classification, or natural language processing. With the process of training, these models developed the ability to spot patterns and features in the data, and they now do the task they were designed for with a given level of accuracy.

Pre-trained models are beneficial since they can help construct new machine learning models more quickly and with fewer resources. Developers can utilise a pre-trained model as a starting point and refine it on a smaller, more focused dataset as opposed to beginning from scratch. With a process called transfer learning, programmers can take advantage of the information and experience stored in the pre-trained model to get better outcomes faster.

3. Implementation

In a model the data should be prepared is known as data pre-processing. Missing value elimination, scaling and normalisation of the data, encoding of categorical variables, and partitioning the data into training and testing sets are some of the activities that are covered. The fighting, robbery, and traffic accident scenes videos are loaded from the Kaggle DCSASS dataset. Frequently, monitoring is done by pulling out successive frames from a video. A uniform method for acquiring and utilising the train/test splits was devised to allow the network to learn the characteristics of anomalous behaviour.

Initially, the three classes have successfully loaded and the sequence frames are found. Videos are changed into image frames that are then employed by the model for detection and classification. In this study, we deployed the Vgg16, ResNet50, and CNN models.

VGG16 Model Implementation

The vgg16 model must first be loaded in order to continue with the implementation. It is a pre -trained model. The input size is (150,150,3). In the event of a theft, the VGG16 model would learn to recognise visual patterns that are common to those videos, such as people moving quickly and aggressively, raised voices, sudden movements or actions, and so on, if it had been trained on a dataset that included videos of robberies or events that are similar to robberies. It is crucial to keep in mind that the precise visual characteristics that the model learns to link with a robbery event may vary depending on the training data and methods chosen.

The VGG16 model has probably learned to identify fighting situations by using some characteristics may include:

- Individual's movements and postures as seen in the video frames.
- Physical interaction between humans, including its duration and intensity.
- The existence of weapons or any other items that are usually linked with violent behavior.
- The people's body language and facial emotions in the video frames.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The VGG16 model uses its feature extraction skills to identify the motion and positions of humans in the video frames. It recognises the frames having edges, corners, and other patterns that represent human movement and positioning. By training itself to recognise specific visual elements that are frequently present in traffic accidents, VGG16 can identify road accidents in a video. These attributes could consist of:

- Having several vehicles nearby one another in close proximity.
- Unusual adjustments to a vehicle's speed or direction.
- Vehicle collisions, whether they involve hitting another car or something else.
- Roadside debris, including broken bits of things or vehicles.

The VGG16 model may be taught to recognise these traits and use them to determine whether a new video involves a road accident by training it on a huge dataset of road accident clips.

Model	VGG-16
Testing score	0.99991
Training score	0.99986
Balanced Accuracy	0.999
Training log loss	0.00047
Testing log loss	0.00045
prediction time(sec)	32.60931

Table.2. VGG16 Prediction Details

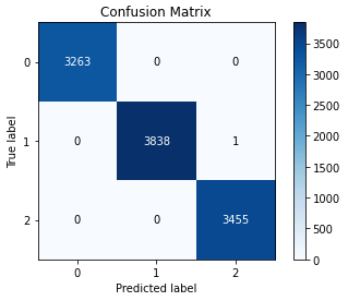


Fig.11. VGG16 Confusion Matrix

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3263
1	1.00	1.00	1.00	3839
2	1.00	1.00	1.00	3455
accuracy			1.00	10557
macro avg	1.00	1.00	1.00	10557
weighted avg	1.00	1.00	1.00	10557

Fig.12. VGG16 classification report

ResNet50 Model Implementation

ResNet50 is a deep convolutional neural network architecture that is frequently employed for image classification and recognition applications, such as the identification of robberies in pictures.

ResNet50 can distinguish different features of robbery photographs, including:

- Suspect conduct: The network is able to identify individuals displaying odd behavior, such as wearing masks or armed with weapons.
- Items that have been stolen: The network can spot objects that have been taken, including jewellery or electronics, in photos captured during or after a robbery.

ResNet50 is a deep convolutional neural network design that is used mostly. The model has been trained using a massive image dataset that contains examples of various sorts of traffic accidents, including auto accidents, accidents

involving passengers, and accidents involving bicycles. As a result, the network has developed the capability to identify numerous elements of automobile accidents, including:

- Vehicles damaged in accidents can have their level of damage determined using ResNet50.
- The network can identify debris on the street, such as broken glass or auto components, and it can also spot patterns of debris that point to particular situations.

For figuring out the reason and severity of an accident, ResNet50 can identify the presence of pedestrians and bicycles in photos.

ResNet-50 may learn to recognise parameters including limb orientation, body position, and body component motion in order to detect human movement in images. ResNet-50 is used for identifying aspects of human movement in photos.

Using a large - scale dataset of fighting-related photos and videos, ResNet-50 is taught to recognize fighting. ResNet-50 may employ various factors to predict fighting which features quick, forceful motions, that can be identified by analyzing the motion patterns in a video.

Model	ResNet50
Testing score	0.333333
Training score	0.333333
Balanced Accuracy	0.3094
Training log loss	1.09861
Testing log loss	1.09861
Prediction time(sec)	82.10928

Table.3. ResNet50 Prediction Details

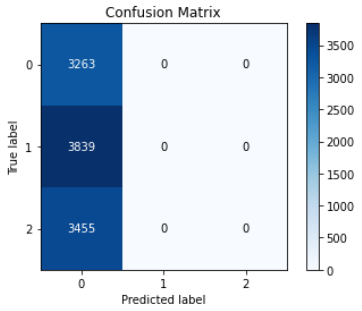


Fig.13. ResNet50 Confusion Matrix

	precision	recall	f1-score	support
0	0.31	1.00	0.47	3263
1	0.00	0.00	0.00	3839
2	0.00	0.00	0.00	3455
accuracy			0.31	10557
macro avg	0.10	0.33	0.16	10557
weighted avg	0.10	0.31	0.15	10557

Fig.14. ResNet50 Classification Report

CNN Model Implementation

For the purpose of extracting features, convolution is a particular kind of linear process. In order to create a feature map, the convolution layer applies a filter or kernel to the incoming data. Convolution, pooling, and nonlinear activation layers on its architecture are used to extract features from the input image. To do this, labelled video data can be used to train a CNN, where each frame of the video is labelled with the position and motion of various body components. So that it can identify human movements in new videos, CNN can train itself to identify these patterns.

CNN can detect human movement by determining the direction and rate of motion in the video by following the movement of pixels. Convolutional Neural Networks (CNNs) can recognise robbery in a video based on a number of factors, which include:

- Recognition of items: CNNs can identify people acting suspiciously

or violently, as well as objects in a video like guns or tools frequently used in robberies.

- Motion analysis: CNNs are able to spot rapid and unexpected movements of people in the video, which might be a sign of a robbery or other criminal behavior.
- Examining the scenario on the footage will allow CNNs to spot any telltale evidence of a robbery, such as shattered windows, damaged goods, or a person under a disguise.

CNNs can precisely identify instances of robbery in a video by integrating these several forms of analysis. CNNs (Convolutional Neural Networks) can determine whether there is fighting in a video depending on a number of factors, including:

- CNNs can identify rapid and violent actions, like as punches, kicks, or grappling, which are suggestive of fighting, by analyzing the motion patterns of the people in the video.
- CNNs can identify fighting-related things like fists, feet, or weapons, and use this knowledge to spot fighting.
- In order to spot physical altercations or physical contact between people, CNNs can examine the positioning and posture of the people in the video.
- CNNs may successfully locate instances of fighting in a video by combining these several forms of analysis.

Road accidents can be predicted in videos using CNNs (Convolutional Neural Networks) based on a variety of parameters, including:

- CNNs can identify accident-related objects like vehicles, bicycles, people, or road debris and use this knowledge to identify accidents.
- CNNs can examine the movement patterns of the objects in the video to spot abrupt changes in direction or speed that might be signs of an accident.

In a video, CNNs can precisely detect instances of auto accidents by incorporating these several forms of analysis.

Model	CNN
Testing score	0.90544
Training score	0.89843
Balanced Accuracy	0.3079
Training log loss	0.64455
Testing log loss	0.62831
Prediction time(sec)	28.25491

Table.4. CNN Prediction Details

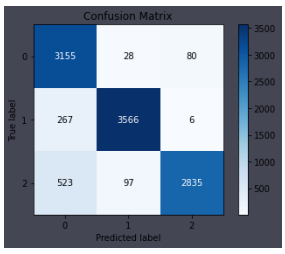


Fig.15. CNN Confusion Matrix

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

	precision	recall	f1-score	support
0	0.80	0.97	0.88	3263
1	0.97	0.93	0.95	3839
2	0.97	0.82	0.89	3455
accuracy			0.91	10557
macro avg	0.91	0.91	0.90	10557
weighted avg	0.92	0.91	0.91	10557

Fig.16. CNN Classification Report

Result and Conclusion

VGG16 model is chosen from the three models for the prediction based on its performance in training and testing. In this project, we opted to try to implement the detection of anomalous behaviours by utilising the VGG16 model because the CNN model is utilised for many activities as regular usage.

Model	Accuracy (%)
TAC-Net	98.1
SVM	96.4
2D-CNN	98.81
Faster R-CNN	89.4
CNN	>=90

Table.5. Analysis of previous models

The model is able to detect the videos which are trained. The model can recognize the three abnormalities in a video other than the datasets successfully while some other videos are misclassified. Frame to frame probability score can be observed in a video.

Model	Training Accuracy	Testing Accuracy
VGG16	0.99986	0.99991
ResNet50	0.333333	0.333333
CNN	0.89843	0.90544

Table.6. Accuracy Comparison of proposed models

The vgg16 model has used because of its increased depth in layers, reduction of overfitting by pooling layers, dropout regulation, the choice of hyperparameters, optimization algorithm, and batch size used during training than ResNet50 and CNN. There are several constraints that can limit its performance such as domain shift, limited interpretability, computational resources and limited dataset diversity. The graphs represent the variation of accuracy balance over the complete execution of the model.

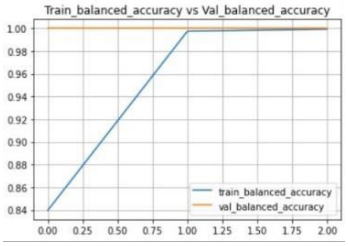


Fig.17. Accuracy graph of VGG16

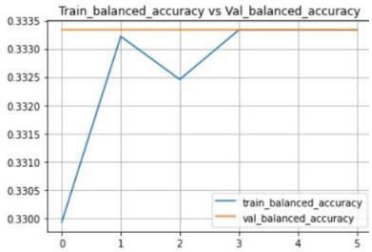


Fig.18. Accuracy graph of ResNet50

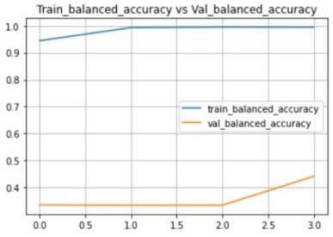


Fig.19. Accuracy graph of CNN

Future Work

In our model, we extract a number of images from the video. If any of the 3

activities are detected in an input video only the occurring activity can be extracted from the video by including the duration of the activity identified and that activity can be surrounded by the bounding box.

References

[1] Chao Huang, Zhihao Wu, Jie Wen, Yong X, Qiuping Jiang and Yaowei Wang. Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System, 2022. IEEE transactions on industrial informatics, vol. 18, no. 8, august 2022.

[2] Kshitija Deshmukh, Aishwarya Kokane, Shweta Konde, Guide-Gauri Virka. Abnormal Activity Detection from Video using SVM Algorithm, 2020. International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 06, June 2020.

[3] Abid Mehmood. Efficient Anomaly Detection in Crowd Videos Using Pre-Trained 2D Convolutional Neural Networks, 2021. IEEE access received September 2, 2021, accepted September 30, 2021, date of publication October 5, 2021, date of current version October 14, 2021.

[4] Sathyajit Loganathan, Gayashan Kariyawasam, Prasanna Sumathipala. Suspicious Activity Detection in Surveillance Footage, 2019. 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA).

[5] Digambar Kauthkar, Snehal Pingle, Vijay Bansode, Pooja Idalkanthe, prof. Sunita Vani. Suspicious Human Activity and Fight Detection using Deep Learning, 2022. Volume 7, Issue 6, June – 2022 International Journal of Innovative Science and Research Technology.

[6] Ashish Sharma, Neeraj Varshney. Identification and Detection of Abnormal Human Activities using Deep Learning Techniques, 2020. *European Journal of Molecular & Clinical Medicine*, *2020,* Volume 7, Issue 4, Pages 408-417.

[7] Shabana Habib, Altaf Hussain, Waleed Albattah, Muhammad Islam, Sheroz Khan, Rehan Ullah Khan, Khalil Khan. Abnormal Activity Recognition from Surveillance Videos Using Convolutional Neural Network, 2021. *Sensors* 2021, *21*(24), 8291; Published on 11 December 2021.

[8] Sreyan Ghosh, Sherwin Joseph Sunny, Rohan Roney. "Accident Detection Using Convolutional Neural Networks", 2019. International Conference on Data Science and Communication (IconDSC), 2019.