ANOMALY RECOGNITION FROM SURVEILLANCE VIDEOS USING CONVOLUTION NEURAL NETWORK

S. YESWANTH, D. MOHANA MURALI KRISHNA

ABSTRACT: *Video Surveillance plays a pivotal role in today's world. The technology has been developed to much in every sector of all departments much when artificial intelligence, machine learning and deep learning pitched into the system. Using the above combinations, different systems are in place which helps to differentiate various suspicious behaviours from the live tracking of footages. The most unpredictable one is human behaviour and it is very difficult to find whether it is suspicious or normal because it occurs every day in our day-to-day life. Deep learning approach is used to detect suspicious or normal activity in an academic environment, and which sends an alert message to the corresponding authority, in case of predicting a suspicious activity. Monitoring is often performed through consecutive frames which are extracted from the video. The entire framework is divided into two parts. In the first part, the features are computed from video frames and in second part, based on the obtained features classifier predict the class as suspicious or normal.*

I. INTRODUCTION:

Video Surveillance plays a pivotal role in today's world. The technology has been developed to much in every sector of all departments much when artificial intelligence, machine learning and deep learning pitched into the system. Using the above combinations, different systems are in place which helps to differentiate various suspicious behaviours from the live tracking of footages. The most unpredictable one is human behaviour and it is very difficult to find whether it is suspicious or normal because it occurs every day in our day-to-day life. Deep learning approach is used to detect suspicious or normal activity in an academic environment, and which sends an alert message to the corresponding authority, in case of predicting a suspicious activity. Monitoring is often performed through consecutive frames which are extracted from the video. Action Recognition, a sub space of vision related applications, is the capacity to distinguish and perceive the activities or objectives of the specialist, the specialist can be any item or substance that performs activity, which has ultimate objectives. Video Surveillance assumes a vital part in this day and age. The innovations have been progressed an excess of when man-made brainpower, AI and profound learning pitched into the framework. The most capricious one is human way of behaving and it is truly challenging to track down whether it is dubious or ordinary. Observing is much of the time performed through sequential edges which are extricated from the video. The whole system is isolated into two sections. In the initial segment, the elements are registered from video outlines and in second part, in light of the got highlights classifier foresee the class as dubious or ordinary.

II. Literature Review:

Classification of Anomaly Activity recognition can be based on multiple parameters. Based on the devices used in the system, Activity Recognition is classified as sensor-based activity recognition and vision-based activity recognition. Vision based activity recognition is a camera- based system that captures the video that can be processed and used to identify the activities in the given environment. These systems normally use digital image processing to extract meaningful information from the video, which is considered as sequence of images. Actions are single-person activities that may be composed of multiple gestures organized temporally, such as “walking,” “waving,” and “punching, etc.” Types of Activity Recognition based on devices used: Based on the devices used in the system, Activity Recognition is classified as sensor-based activity recognition and vision-based activity recognition. 1. Sensor based activity recognition uses network of sensors to monitor the behaviour of an actor, and some monitor the surroundings. Such data collected from various sensors may be aggregated and processed to derive some essential information from them. They are further used for training the model using different data analytics, machine learning and deep learning techniques. 2. Vision based activity recognition is a camera- based system that captures the video that can be processed and used to identify the activities in the given environment. These systems normally use digital image processing to extract meaningful information from the video, which is considered as sequence of images.

III. Activity Recognition Using CNN RESNET:

Activity recognition has been an emerging field of research since the past few decades. Humans have the ability to recognize activities from a number of observations in their surroundings. ▪ These observations are used in several areas like video surveillance, health sectors, gesture detection, energy conservation, fall detection systems and many more. ▪ The Resnet-18 model is used as a a convolutional neural network that is 18 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. ▪ It provides shortcut connections which resolves the problem of vanishing gradient. The model is trained and tested successfully giving a satisfactory result by recognizing over 400 human actions. Finally, some open problems are presented which should be addressed in future research.

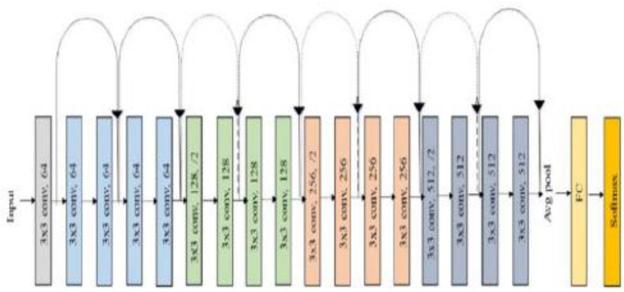


Fig 1. Original Resnet 18 CNN Architecture

IV. FLOW CHART FOR THE ANOMALY DETECTION USING RESNET 18 CNN:

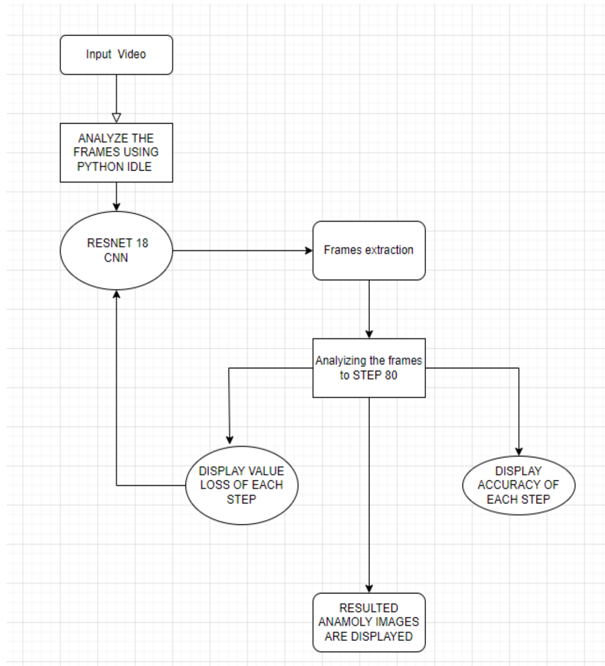


Fig 2.1. Flow chart diagram of anomaly detection using Resnet 18 CNN

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. From this Activity Diagram first we have input the video and later analyse the frames using Resnet 18 the frames should be extracted and it displays loss and accuracy of video of anomaly is displayed and later the respective anomaly frames are shown as result.

V. SEQUENCE DIAGRAM FOR ANOMALY DETECTION USING RESNET 18 CNN:

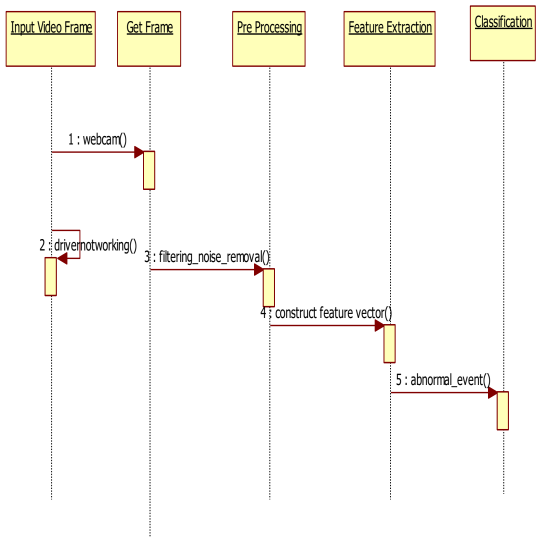


Fig 2.2. Sequence diagram for Anomaly Detection using RESNET 18 CNN

A Sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagram.

VI. Resnet 18 Architecture with frame size of output:

LAYER NAME	Output Size	ResNet-18
conv1	112 x 112 x 64	7 x 7, 64, stride 2
conv2_x	56 x 56 x 64	[3 × 3, 64 3 x3, 64] x 2
conv3_x	28 x 28 x 128	[3 x 3, 128 3 x3, 128] x 2
conv4_x	14 x 14 x 256	[3 x 3,256 3 x 3,256] x 2
conv5_x	7 x 7 x 512	[3 x 3,512 3 x 3,512] x 2
average pool	1 x 1 x 512	7 x 7 average pool
fully connected	1000	512 x 1000 fully connections
Soft max	1000	

VII. EXPERIMENT SETUP AND DATA SET

UCF Crime Anomaly Detection Dataset

1,900 real-world surveillance videos of 128 hours

15 times more videos than existing datasets.

13 real-world anomalies

	# Of videos	Average frames	Dataset length	Example anomalies
UCSD Ped1[27]	70	201	5 min	Bikers, small carts, walking across walkways
UCSD Ped2[27]	28	163	5 min	Bikers, small carts
Subway Entrance [3]	1	121,749	1.5 hours	Wrong direction, No payment
Subway Exit [3]	1	64,901	1.5 hours	Wrong direction, No payment
Avenue [28]	37	839	30 min	Run, throw, new object
UMN [2]	5	1290	5 min	Run
BOSS [1]	12	4052	27 min	Harass, Disease, Panic
Ours	**1900**	**7247**	**128 hours**	**Abuse, arrest, arson, assault, accident, burglary, fighting, robbery**

ANOMALY	NUMBER OF VIDEOS
BURGLARY	100
FIGHTING	50
ROAD ACCIDENTS	150
ROBBERY	150
SHOOTING	50
SHOP LIFTING	50
STEALING	100
ABUSE	50
ARREST	50
ARSON	50
ASSAULT	50
EXPLOSION	50
VANDALISM	50
NORMAL	950

VII. RESULTS AND DISCUSSION:

1. ABUSE:

First, we have input the Anomaly video of abuse and later analyse all the frames and evaluate up to step 80 and give the accuracy and results.

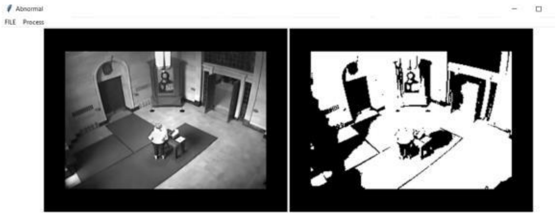


Fig1.1 Input the Anomaly video of Abuse

```
Starting training..
Starting epoch 1/1
=====================
Evaluating at step 0
Val loss: 2.4813, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 20
Val loss: 5.9352, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 40
Val loss: 6.6438, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 60
Val loss: 4.0143, Acc: 0.3750
Accuracy::: 0.375
Evaluating at step 80
Val loss: 4.5377, Acc: 0.2500
Accuracy::: 0.25
Training loss: 0.0017
Result....
```

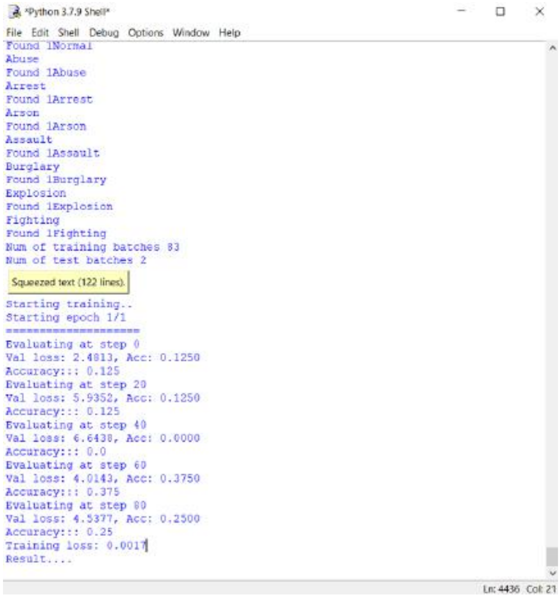
Fig 1.2 After every frame was analysed, we get the Accuracy and the result of the Abuse video at which frames the anomaly occurs. Accuracy may changes some time when you input the video

Accuracy	0.25
Training Loss	0.0017

2. ARREST:

Input the Anomaly video of Arrest

Fig 2.1 Input the Anomaly video of Arrest



```
Found 1Normal
Abuse
Found 1Abuse
Arrest
Found 1Arrest
Arson
Found 1Arson
Assault
Found 1Assault
Burglary
Found 1Burglary
Explosion
Found 1Explosion
Fighting
Found 1Fighting
Num of training batches 83
Num of test batches 2
Squeezed text (122 lines).
Starting training..
Starting epoch 1/1
================
Evaluating at step 0
Val loss: 2.4813, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 20
Val loss: 5.9352, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 40
Val loss: 6.6438, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 60
Val loss: 4.0143, Acc: 0.3750
Accuracy::: 0.375
Evaluating at step 80
Val loss: 4.5377, Acc: 0.2500
Accuracy::: 0.25
Training loss: 0.0017
Result....
```

Fig 2.2 After every frame was analysed, we get the Accuracy and the result of the Arrest video at which frames the anomaly occurs.

Accuracy	0.3750
Training loss	0.0017

3. ARSON:

Input the Anomaly video of Arrest

Fig 3.1 Input the Anomaly video of Arrest

```
Starting training..
Starting epoch 1/1
====================
Evaluating at step 0
Val loss: 2.0477, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 20
Val loss: 2.0215, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 40
Val loss: 2.1350, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 60
Val loss: 2.0615, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 80
Val loss: 2.0156, Acc: 0.3750
Accuracy::: 0.375
Training loss: 0.0011
Result....
```

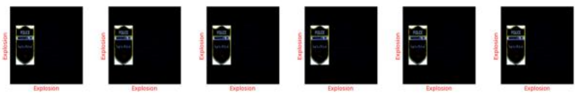


Fig 3.2 After every frame was analysed, we get the Accuracy and the result of the Arson video at which frames the anomaly occurs.

Accuracy	0.3750
Training loss	0.0011

4. ASSAULT:

Input the Anomaly video of Assault

Fig 4.1 Input the Anomaly video of Arrest

```
Starting training..
Starting epoch 1/1
====================
Evaluating at step 0
Val loss: 1.9600, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 20
Val loss: 3.7318, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 40
Val loss: 5.6654, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 60
Val loss: 5.3958, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 80
Val loss: 5.4050, Acc: 0.3750
Accuracy::: 0.375
Training loss: 0.0005
Result....
```

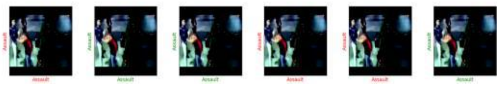


Fig 4.2 After every frame was analysed, we get the Accuracy and the result of the Assault video at which frames the anomaly occurs

Accuracy	0.375
Training loss	0.0005

5. BURGLARY:

Input the Anomaly video of Burglary

Fig 5.1 Input the Anomaly video of Burglary

```
Squeezed text (122 lines).
Starting training..
Starting epoch 1/1
====================
Evaluating at step 0
Val loss: 3.0878, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 20
Val loss: 4.3267, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 40
Val loss: 4.4866, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 60
Val loss: 6.2264, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 80
Val loss: 3.9449, Acc: 0.3750
Accuracy::: 0.375
Training loss: 0.0021
Result....
```

Fig 5.2 After every frame was analysed, we get the Accuracy and the result of the Burglary video at which frames the anomaly occurs.

Accuracy	0.375
Training loss	0.0021

6. EXPLOSION:

Input the Anomaly video of Explosion

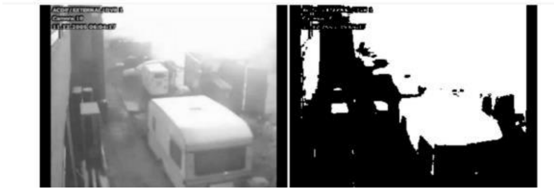


Fig 6.1 Input the Anomaly video of Explosion

```
Evaluating at step 0
Val loss: 2.3520, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 20
Val loss: 2.9448, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 40
Val loss: 4.7008, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 60
Val loss: 6.2752, Acc: 0.0000
Accuracy::: 0.0
Evaluating at step 80
Val loss: 7.3408, Acc: 0.0000
Accuracy::: 0.0
Training loss: 0.0295
Result....
```

Fig 6.2 After every frame was analysed, we get the Accuracy and the result of the Explosion video at which frames the anomaly occurs

Accuracy	0.0
Training loss	0.0295

7. FIGHTING:

Input the Anomaly video of Fighting

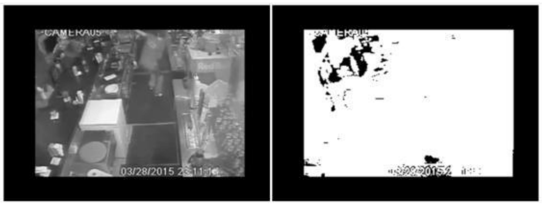
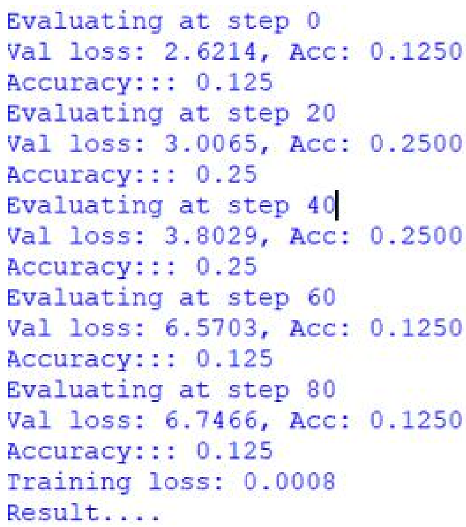


Fig 7.1 Input the Anomaly video of Fighting



```
Evaluating at step 0
Val loss: 2.6214, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 20
Val loss: 3.0065, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 40
Val loss: 3.8029, Acc: 0.2500
Accuracy::: 0.25
Evaluating at step 60
Val loss: 6.5703, Acc: 0.1250
Accuracy::: 0.125
Evaluating at step 80
Val loss: 6.7466, Acc: 0.1250
Accuracy::: 0.125
Training loss: 0.0008
Result....
```

Fig 7.2 After every frame was analysed, we get the Accuracy and the result of the Fighting video at which frames the anomaly occurs

Accuracy	0.125
Training loss	0.008

NOTE: From the Anomaly we need to understand that the Accuracy and Training loss may changes due to its runtime sometimes it shows accuracy sometimes may shows less but the anomaly frames from video are extracted as output.

VIII. CONCULSION AND FUTURE WORKS:

This exploration works basically revolves around the affirmation of different peculiarities from perception accounts to discard a lot of human intervention. This study proposes a robotized significant learning-based approach for veritable weird development affirmation. This work is coordinated considering the way that not much work has been finished as such far on the affirmation of various peculiarities and for the most part researchers address simply twofold gathering i.e., either a video is standard or containing irregularity. Our expansive composing review moreover shows that why significant learning-based approaches have pervasiveness over excellent based approaches for the extraction of components from accounts. The proposed concentrate on gives a changed, pre-arranged 3D Convents plan that outmanoeuvres on the as of late declared approaches. This 3D model is used to really remove both spatiotemporal components from surveillance accounts. The concentrate furthermore addresses the meaning of the presence of edge level checking for better learning of spatiotemporal components in a semi-oversaw manner. Besides, this work has shown the importance of spatial increment to secure better outcomes while setting up a significant plan. The proposed technique is applied to the huge degree UCF Crime dataset. The tests coordinated on this dataset show that the aligned 3DConvNets outmanoeuvres the ongoing state of-craftsmanship sporadic development affirmation approaches with respect to precision. The proposed work furthermore gives a pilot find out about different classes of the UCF Crime dataset and inspected it cut off points for the sporadic activity affirmation task that will be helpful for future work on this dataset.

IX. REFERENCES:

1. Ahmed Taha, Halazayed [1] 2021 proposed a framework for human movement acknowledgment. The proposed framework presents a human action descriptor in light of the people skeletal data of the human action is invariant to the size of the subject and the direction of camera. Secret Markov Models are utilized to perceive human exercises. For every action class, a HMM is learned. Analyse

completed on two benchmark datasets: Cornell CAD-60 and Cornell CAD-120.

2. Wei a Deng [2] 2020 researches the skeletonization issue utilizing equal diminishing strategies and propose another one pass equal lopsided diminishing calculation. Wu and Tsai introduced a one-pass equal uneven diminishing calculation that executed 4-distance, or city block distance, skeletonization. By applying 8-distance, or chessboard distance, this new calculation works on the nature of the subsequent skeletons as well as the effectiveness of the calculation. This calculation utilizes 18 examples. The proposed OTPA8 has great commotion obstruction, totally 8-associated skeleton output and a quicker speed without genuine disintegration.

3. B. Yogameena, S. Veera Lakshmi 2019[3] proposed a constant video observation framework which is fit for characterizing ordinary and unusual activity of people in swarm. The strange activity of human, for example, running, hopping, waving hand, twisting, strolling and battling with one another in a packed climate are thought of. In

4. J. Howey, P. Poupart, A. V. Bertoldi, T. Craig, C. Boutilier, and A. Mopaliids, “Automated hand washing assistance for persons with dementia using video and a partially observable Markov decision process”, Computer Vision and Image Understanding, Vol. 114, 2010, pp. 503–519.

5. F. Kruger, M. Nyolt, K. Yordanova, A. Hein, and T. Kirste, “Computational state space models for activity and intention recognition. A feasibility study”, PLoS ONE, vol.9(11),2014:e109381,doi:10.1371/journapone.0109381. K. He, X. Zhang, S. Ren and J. Sun, “DelvingDeepinto Rectifiers: Surpassing Human-Level Performance on ImageNet ClassificationK.

6. N. Tran, A. Bedagkar-Gala, I. A. Kakadiaris, and S. K. Shah, “Social cues in group formation and local interactions for collective activity analysis”, In VISAPP, 2013, pp. 539–548.

AUTHORS PROFILE:

Dr. P. KUPPUSAMY received his Bachelor’s degree in Computer Science and Engineering from Madras University, India, 2002, and Master’s degree in Computer Science and Engineering from Anna University, Chennai, India, 2007. He has completed Ph. D in Information and Communication Engineering, Anna University, Chennai, 2014. At present, he is working as Associate Professor Grade 2, School of Computer and Science Engineering, VIT-AP. He has published 30 research papers in leading international journals, 7 papers in IEEE international conferences and a book for Computer Science and Engineering. Currently, working on Machine learning, Internet of Things based research project with sensors, Raspberry Pi, Arduino for handling smart devices. He acts/acted also as a reviewer for various journals (e.g., Pervasive and Mobile Computing Elsevier, Journal of Computer Science from Science Publications) and International Conferences. He has organized research workshops, seminars based on Arduino controller with smart devices and FDPs. He is Member of IEEE, ISTE, IAENG and IACSIT. His specializations include Machine learning, Networks, Cryptography, Distributed computing and IoT. His current research interests are Machine Learning and Internet of Things.

S. Yeswanth and D. Mohana Murali krishna is Studying Integrated M.Tech Software Engineering in VIT-AP UNIVERSITY Amaravathi Vijayawada. They are currently doing Capstone Project on Deep Learning