ANOMALY RECOGNITION FROM SURVEILLANCE VIDEOS USING CONVOLUTION NEURAL NETWORK

S. YESWANTH, D. MOHANA MURALI KRISHNA

ABSTRACT: *Video Surveillance assumes a vital part in this day and age. The innovation has been created to much in each area of all divisions much when man-made consciousness, AI and profound learning pitched into the framework. Utilizing the above mixes, various frameworks are set up which assists with separating different dubious ways of behaving from the live following of recordings. The most erratic one is human way of behaving and it is extremely challenging to track down whether it is dubious or typical in light of the fact that it happens consistently in our everyday life. Profound learning approach is utilized to identify dubious or typical movement in a scholarly climate, and which sends an alarm message to the relating authority, in the event of foreseeing a dubious action. Checking is in many cases performed through continuous edges which are extricated from the video. The whole system is separated into two sections. In the initial segment, the elements are processed from video outlines and in second part, in view of the acquired highlights classifier foresee the class as dubious or typical.*

I. INTRODUCTION:

Video Surveillance assumes a vital part in this day and age. The innovation has been created to much in each area of all divisions much when man-made brainpower, AI and profound learning pitched into the framework. Utilizing the above blends, various frameworks are set up which assists with separating different dubious ways of behaving from the live following of recordings. The most erratic one is human way of behaving and it is undeniably challenging to track down whether it is dubious or typical on the grounds that it happens consistently in our everyday life. Profound learning approach is utilized to recognize dubious or typical movement in a scholarly climate, and which sends an alarm message to the relating authority, in the event of foreseeing a dubious action. Observing is many times performed through back to back outlines which are removed 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:

Grouping of Anomaly Activity acknowledgment can be founded on numerous boundaries. In light of the gadgets utilized in the framework, Activity Recognition is named sensor-based action acknowledgment and vision-based movement acknowledgment. Vision based movement acknowledgment is a camera-based framework that catches the video that can be handled and used to distinguish the exercises in the given climate. These frameworks typically utilize computerized picture handling to separate significant data from the video, which is considered as grouping of pictures. Activities are single-individual exercises that might be made out of different motions coordinated transiently, for example, "strolling," "waving," and "punching, and so on" Types of Activity Recognition in view of gadgets utilized: Based on the gadgets utilized in the framework, Activity Recognition is delegated sensor-based movement acknowledgment and vision-based action acknowledgment. 1. Sensor based action acknowledgment utilizes organization of sensors to screen the way of behaving of an entertainer, and some screen the environmental elements. Such information gathered from different sensors might be accumulated and handled to get some fundamental data from them. They are additionally utilized for preparing the model utilizing various information investigation, AI and profound learning strategies. 2. Vision based movement acknowledgment is a camera-based framework that catches the video that can be handled and used to recognize the exercises in the given climate. These frameworks typically utilize advanced picture handling to extricate significant data from the video, which is considered as grouping of pictures.

III. Activity Recognition Using CNN RESNET:

Movement acknowledgment has been an arising field of exploration since the beyond couple of many years. People can perceive exercises from various perceptions in their environmental factors. ▪ These perceptions are utilized in a few regions like video observation, wellbeing areas, motion location, energy protection, fall identification frameworks and some more. ▪ The Resnet-18 model is utilized as a convolutional brain network that is 18 layers profound. You can stack a pretrained form of the organization prepared on in excess of 1,000,000 pictures from the ImageNet data set. ▪ It gives alternate way associations which settle the issue of disappearing inclination. The model is prepared and tried effectively giving a good outcome by perceiving more than 400 human activities. At long last, a few open issues are introduced which ought to be tended to in future exploration.

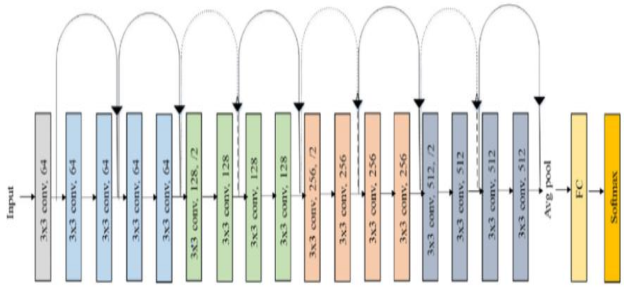


Fig 1. Original Resnet 18 CNN Architecture

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

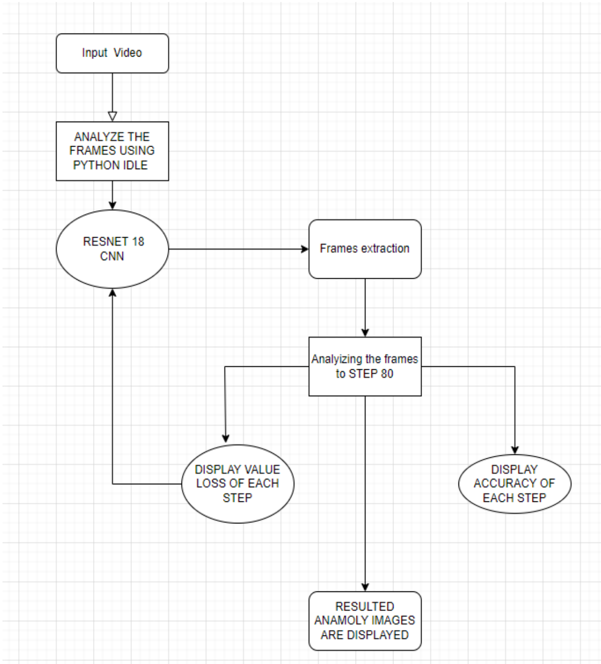


Fig 2.1. Flow chart diagram of anomaly detection using Resnet 18 CNN

Action outlines are graphical portrayals of work processes of stepwise exercises and activities with help for decision, emphasis and simultaneousness. 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:

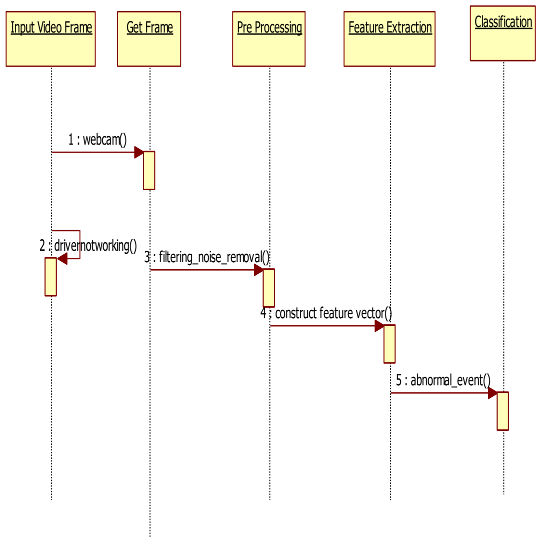


Fig 2.2. Sequence diagram for Anomaly Detection using RESNET 18 CNN

A Sequence graph in Unified Modelling Language (UML) is a sort of collaboration chart that shows how cycles work with each other and in what request. It is a build of a Message Sequence Chart. Grouping charts are now and then called occasion graphs, occasion situations, and timing outline. 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.

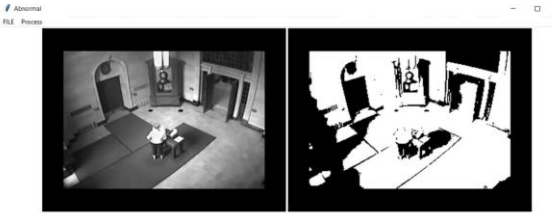


Fig1.1 Input the Anomaly video of Abuse

```
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....
```

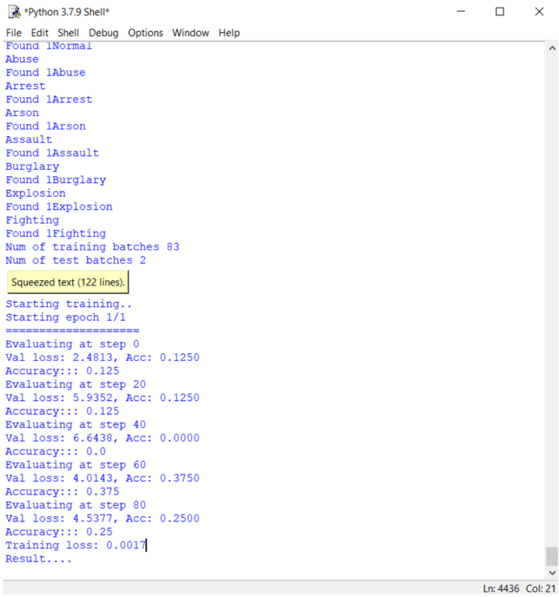
Fig 1.2 After every frame was analysed, we get the Accuracy and the result of the Abuse video at which frames the anomaly occurs. Accuracy may changes some time when you input the video

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....
```

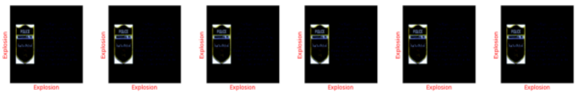


Fig 3.2 After every frame was analysed, we get the Accuracy and the result of the Arson video at which frames the anomaly occurs.

Accuracy	0.3750
Training loss	0.0011

4. ASSAULT:

Input the Anomaly video of Assault

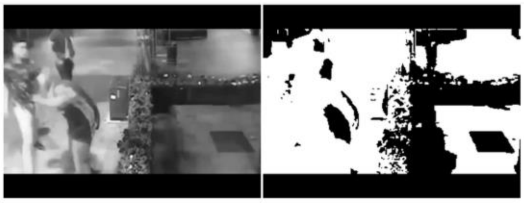


Fig 4.1 Input the Anomaly video of Arrest

```
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....
```

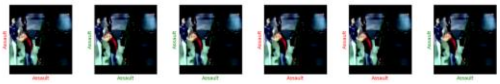


Fig 4.2 After every frame was analysed, we get the Accuracy and the result of the Assault video at which frames the anomaly occurs

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

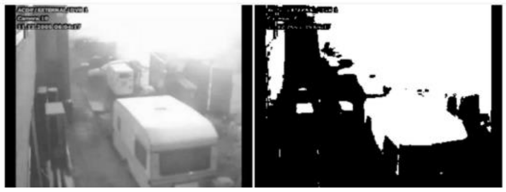


Fig 6.1 Input the Anomaly video of Explosion

```
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

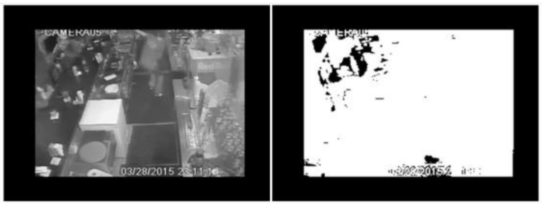
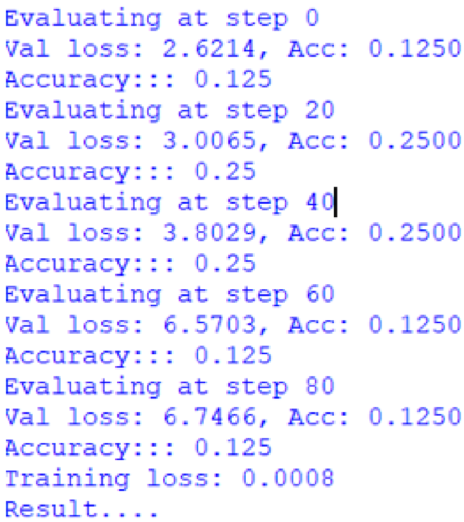


Fig 7.1 Input the Anomaly video of Fighting



```
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 intercession. This study proposes a robotized critical 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 inconsistency. Our extensive creating survey also shows that why huge learning-based approaches have inescapability over brilliant based approaches for the extraction of parts from accounts. The proposed focus on gives a changed, coordinated 3D Convents plan that outsmarts on the actually pronounced approaches. This 3D model is utilized to truly eliminate both spatiotemporal parts from reconnaissance accounts. The concentrate moreover addresses the importance of the presence of edge level checking for better learning of spatiotemporal parts in a semi-supervised way. Furthermore, this work has shown the significance of spatial augmentation to get improved results while setting up a huge arrangement. The proposed strategy is applied to the enormous degree UCF Crime dataset. The tests facilitated on this dataset show that the adjusted 3DConvNets outsmarts the continuous condition of-craftsmanship inconsistent advancement attestation approaches concerning accuracy. The proposed work moreover gives a pilot learn about various classes of the UCF Crime dataset and examined it limits for the inconsistent movement confirmation task that will be useful for future work on this dataset.

IX. REFERENCES:

1. Ahmed Taha, Halazayed [1] 2021 proposed a system for human development affirmation. The proposed structure presents a human activity descriptor considering individuals skeletal information of the human activity is invariant to the size of the subject and the bearing of camera. Secret Markov Models are used to see human activities. For each activity class, a HMM is learned. Dissect finished on two benchmark datasets: Cornell CAD-60 and Cornell CAD-120.

2. Wei a Deng [2] 2020 investigates the skeletonization issue using equivalent decreasing systems and propose another pass equivalent disproportionate reducing computation. Wu and Tsai presented a one-pass equivalent lopsided lessening computation that executed 4-distance, or city block distance, skeletonization. By applying 8-distance, or chessboard distance, this new computation deals with the idea of the ensuing skeletons as well as the viability of the estimation. This computation uses 18 models. The proposed OTPA8 has incredible upheaval impediment, absolutely 8-related skeleton yield and a faster speed without certifiable crumbling.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/journap one.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

AUTHORS PROFILE:

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