# Intelligent Video Surveillance using Deep Learning

**Vijay Bhanudas Gujar, Arbaaz Shaikh, Alim Bagwan, Pooja Dixit, Nidhi Todkar**

***Abstract:*** ***Now days, Big data applications are having most of the importance and space in industry and research area. Surveillance videos are a major contribution to unstructured big data.***

***The main objective of this paper is to give brief about video analysis using deep learning techniques in order to detect suspicious activities. Our main focus is on applications of deep learning techniques in detection the count, no of involved persons and the activity going on in a crowd considering all conditions [9]. This video analysis helps us to achieve security. Security can be defined in different terms like identification of theft, detecting violence etc.***

***Suspicious Human Activity Detection is simply the process of detection of unusual (abnormal)l human activities . For this we need to convert the video into frames and processing these frames helps us to analyze the persons and their activities. There are two modules in this system first one Object Detection Module and Second one is Activity Detection Module .Object detection module detects whether the object is present or not. After detecting the object the next module is going to check whether the activity is suspicious or not.***

***Keywords: Big data, Video surveillance, Deep learning, Crowd analysis, Machine Learning; Violent Activities Detection; Convolutional Neural Network; Recurrent Neural Network; Long Short-Term Memory.***

## I. INTRODUCTION

The interest and use of image processing and video analysis has been increased now a day and it has been unprecedented due to its importance in finding out and summarization and recognizing of actions [7].

This paper explains about hoe to process the videos and images in order to find difference between them.

We developed a system which classifies a video into three classes:

1. illegal or aggressive activity
2. Potentially doubtful
3. secure

Our plan to resolve this difficulty is an structural design base on convolutional and recurrent neural networks Throughout this paper, it will easy to find general description and solution, it also tells about the method that we agree to the dataset we used, how we implement it, and the outcome that we achieve.

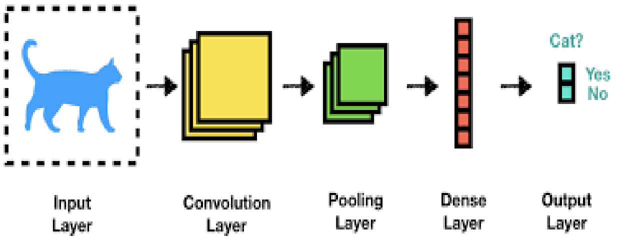


**Figure 1: Convolutional Neural Network**

## II. LITERATURE REVIEW

The primary neural network is also known as convolutional network. The reason of this arrangement is extract high-level facial appearance of the metaphors and tumbling the difficulty of the contribution [8]. We are using a labeled form called inception which is created by Google. Its v3 model is taught on the ImageNet dataset. It is the most typical task, this layout used to categorize whole imagery into 1,000 classes.

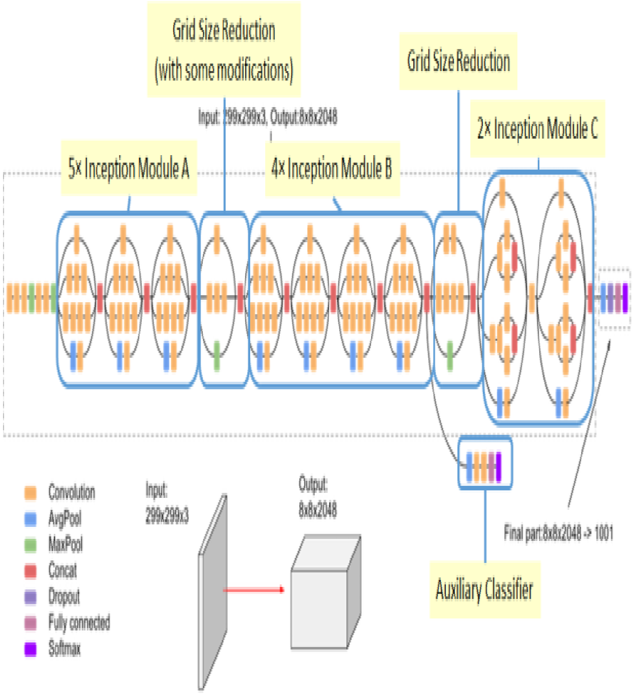


**Figure 2: Inception Model**

**Revised Manuscript Received on February 06, 2020.**

**Mr. Vijay Bhanudas Gujar,** Department of CSE, Dnyanshree Institute of Engineering and Technology, Satara, India.
E-mail: gujar.vijay@gmail.com

**Mr. Arbaaz Shaikh**, Department of CSE, Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra, India.

**Miss. Pooja Dixit,** Department of CSE, Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra, India.

**Mr. Alim Bagwan,** Department of CSE, Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra, India.

**Miss. Nidhi Todkar**, Department of CSE, Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra, India.

This replication of video images is used to transfer learning. Recent detection of objects replica capable to recognize number of factor and can take many days to entirely train. To correct a lot of this work we used Transfer learning. This technique reduces the work and time needed for completely trained model for categories like ImageNet and keep hold on fresh classes [7].

The neural network is the second network we used. The main purpose of this net is to find out the logic in the series of the events. This LSTM cell present in the first layer of this network and this two hidden layers follows this layer (one which having around one thousand neurons and relu activation and the additional which having fifty neurons with a sigmoid activation), and the final production level is also known as three-neuron layer which creates software that produces the last categorization.

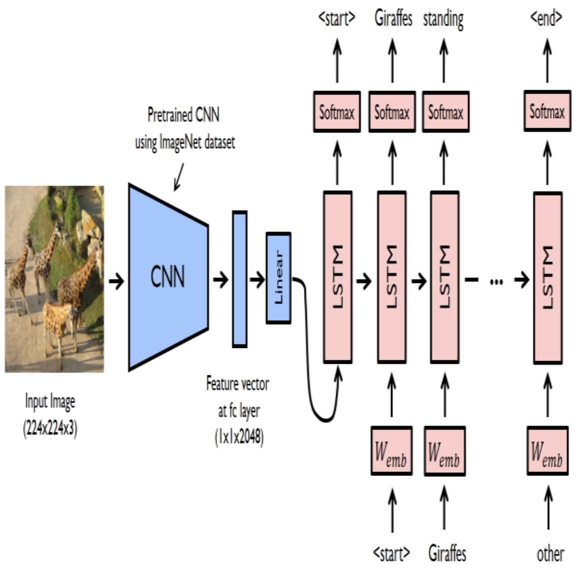
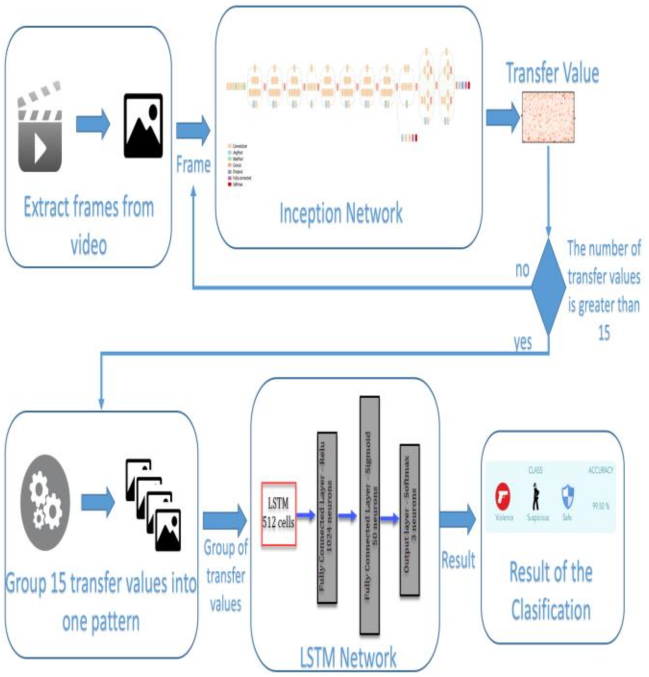


**Figure 3: LSTM Network**

## III. METHODOLOGY

Initially we are extracting the frames from the video. We take out each and every frame at every fraction of time. We use this edge to create a guess by the initiation copy [7]. By using the transfer learning technique; Instead of extracting the last categorization of the beginning model, we are going to extract the effect of the previous common layer, which is a vector of 2,048 values (high-level attribute chart). However, we had a single frame characteristic chart. We are giving our system a sense of the sequence. For this, It's not enough to consider only single frame to make our final guess this is why we take a collection of edge in arrange to categorize the section of the visual images of stationary or moving objects [8].

To make a good guess its sufficient to analyze three to four seconds of video at a time. The inception model generates fifteen feature maps. We take these frames and three seconds of video corresponding them. Now we need to form a one single pattern to join this set of characteristic. And we are ready with input of our second neural network [6].



Extract frames from video
Frame
Inception Network
Transfer Value
The number of transfer values is greater than 15
no
yes
Group 15 transfer values into one pattern
Group of transfer values
LSTM 512 cells
Fully Connected Layer - Relu 1024 neurons
Fully Connected Layer - Sigmoid 50 neurons
Output layer - Softmax 3 neurons
LSTM Network
Result
CLASS
ACCURACY
Violence
Suspicious
Safe
99.50 %
Result of the Clasification

**Figure 4: Classification Network**

As output, what we observe is going to characterize the video in real-time for where every three seconds.Carectrization factors are — secure, doubtful, or against the law activity [9].

## IV. DATASET TRAINING

We used dataset which of around two hundreds minutes of videos divided into 38 videos. We take the majority of the inputs from security cameras. We are taking each frame at 0.2 sec which makes a dataset of 45,000 edges for coaching — the equivalent of 3000 sections of film, consider so as to each a segment of a video stand for three seconds of it.

The last dataset is in fact somewhat little. We can get good quality consequences with less data Because of the transfer learning.

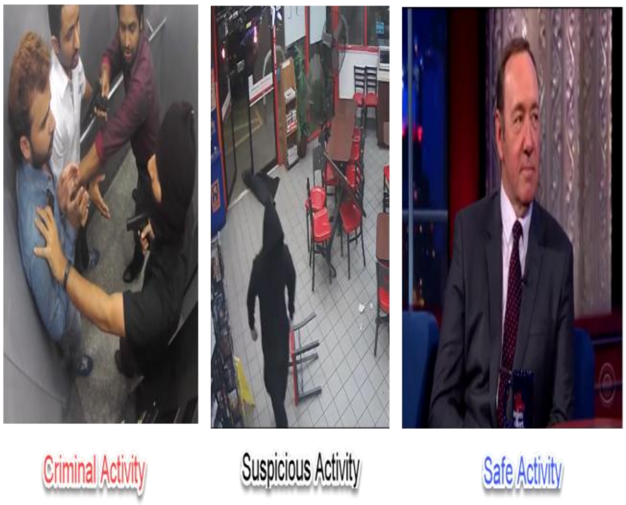


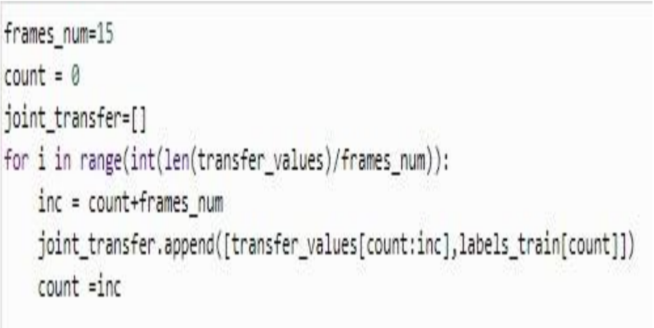
**Figure 5: Data class's example**

## V. IMPLEMENTATION

We used python 3.5 to imlepemnt this whole system.

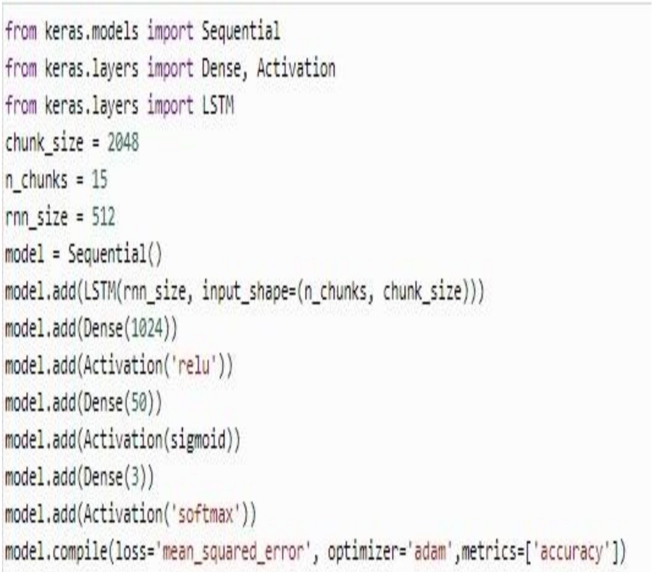
To section the video we used OpenCV for Python. Sectioning includes the conversion of videos into edges and makes the size more appropriate to them. From available, using each of them that we make a forecast on the inception model [4]. We used transfer values variable to save that into its appropriate space.

From available variables, we are going to split them into group of 15 edges and result is in the reassign to variable.



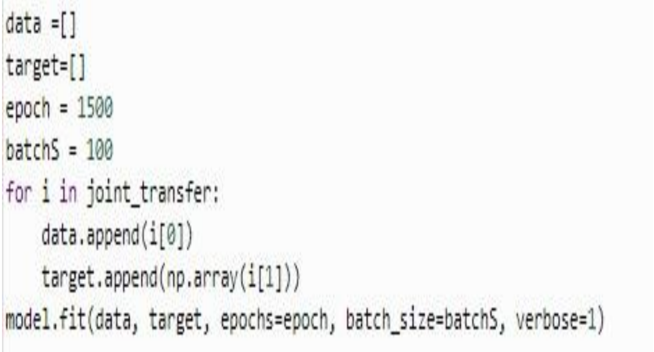
```
frames_num=15
count = 0
joint_transfer=[]
for i in range(int(len(transfer_values)/frames_num)):
    inc = count+frames_num
    joint_transfer.append([transfer_values[count:inc],labels_train[count]])
    count =inc
```

Here we are having the relocate values with respective tag; from this data we train neural network. The implementation of this network is as follows:



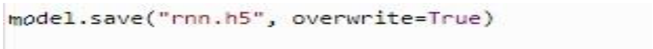
```
from keras.models import Sequential
from keras.layers import Dense, Activation
from keras.layers import LSTM
chunk_size = 2048
n_chunks = 15
rnn_size = 512
model = Sequential()
model.add(LSTM(rnn_size, input_shape=(n_chunks, chunk_size)))
model.add(Dense(1024))
model.add(Activation('relu'))
model.add(Dense(50))
model.add(Activation(sigmoid))
model.add(Dense(3))
model.add(Activation('softmax'))
model.compile(loss='mean_squared_error', optimizer='adam',metrics=['accuracy'])
```

Above code describe the structure of the representation. The next step is about training it [3]:



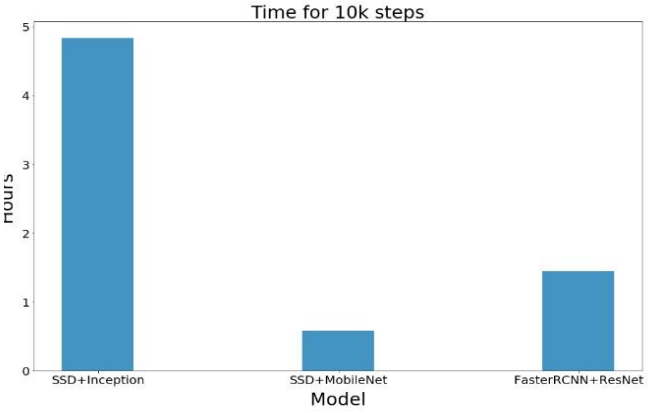
```
data =[]
target=[]
epoch = 1500
batchS = 100
for i in joint_transfer:
    data.append(i[0])
    target.append(np.array(i[1]))
model.fit(data, target, epochs=epoch, batch_size=batchS, verbose=1)
```

After training need to save it as follows:



```
model.save("rnn.h5", overwrite=True)
```

This model is fully trained now; we can start to categorize videos [1].

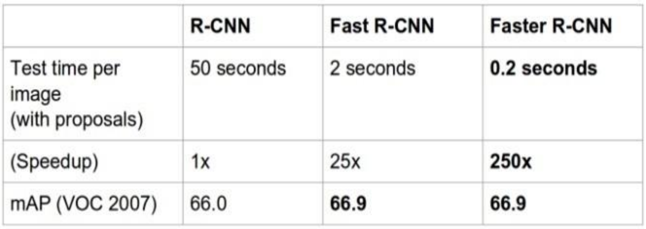


## VI. RESULTS AND OTHER POSSIBLE APPLICATIONS

Following experiment by means of dissimilar system layouts and alteration overexcited parameter, the most excellent result that we could obtain is 98% correctness.

We have the user side for to put video into server. Then classify it in instantaneous. The alteration of classes you can see now continually [2]. These values always bring up to date each three seconds waiting the video is over.

In addition what we can do is connect video classifier it to a safety astrograph to analyze the record in instantaneous, and when notice disbelieving movement scheme, it might attentive us.



| | R-CNN | Fast R-CNN | Faster R-CNN |
|---|---|---|---|
| Test time per image<br>(with proposals) | 50 seconds | 2 seconds | **0.2 seconds** |
| (Speedup) | 1x | 25x | **250x** |
| mAP (VOC 2007) | 66.0 | **66.9** | **66.9** |

## VII. CONCLUSION

Deep learning is a marvelous tool that offers excellent outcomes easily. But, to what limit can we trust our inspection structure to act on its own? There are a few instances where automation is questionable.

**A.** Adversarial Attacks
**B.** False positives
**C.** Similar faces
**D.** Lack of diversity in datasets

## VIII. FUTURE SCOPE

Effectively tracking of suspicious person's on-demand: To ensure the performance and accuracy of suspicious tracking. Suspicious tracking across multiple cameras based on correlation filters leverages entry and exit locations within the protected environment, so that a suspicious person can be tracked across cameras uniquely by a relay. Only adjacent cameras are candidates used for re-identification process, and it will reduce the computation time and cost as well as increasing tracking performance and accuracy in daily operations.

*Retrieval Number: C5233029320/2020©BEIESP*
*DOI: 10.35940/ijeat.C5233.029320*

*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

## REFERENCES

1. 1.TensorFlow Image Recognition Tutorial from https://www.tensorflow.org/tutorials/images/transfer_learning_with_hub
2. TensorFlow Image Retraining Tutorial from https://www.tensorflow.org/tutorials/images/transfer_learning
3. Python website from https://www.python.org/
4. OpenCV and Python from https://pypi.org/project/opencv-python/
5. Keras website from https://keras.io/.
6. “*Deep learning*,”, Y. LeCun, Y. Bengio, and G. Hinton, “Nature, vol. 521, no. 7553, pp. 436–444, 2015.
7. *Deep learning for recognizing human activities using motions of skeletal joints*, Cho Nilar Phyo, Student Member, IEEE, Thi Thi Zin, Member, IEEE and Pyke Ti, IEEE 2018, Vol No: 0098-3063
8. *Intelligent Transportation using Deep Learning*.”, Vijay Bhanudas Gujar,“International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020.
9. “*DEEP LEARNING: EFFECTIVE TOOL FOR BIG DATA ANALYTICS*”, 1Vijay B. Gujar, 2Omkar Nilakhe, 3Pradip Shendage, 2019 IJRAR June 2019, Volume 6, Issue 2 www.ijrar.org (E-ISSN 2348-1269, P- ISSN 2349-5138).

## AUTHORS PROFILE

**Mr. Vijay Bhanudas Gujar** ,M.Tech in Computer Engg, from DBATU lonere. Assistant Professor in CSE Department of Dnyanshree Institute of Engineering and Technology, Satara, more than 16 years Academic experience , Area of Interest: Genetic Algorithm, Big data Analytics. Deep Learning, Machine Learning.

**Mr. Arbaaz Shaikh**, BE (CSE) student of Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra. Area of Interest: Big data Analytics. Deep Learning, Machine Learning.

**Miss. Pooja Dixit,** BE (CSE) student of Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra. Area of Interest: Big data Analytics. Deep Learning, Machine Learning.

**Mr. Alim Bagwan,** BE (CSE) student of Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra. Area of Interest: Big data Analytics. Deep Learning, Machine Learning.

**Miss. Nidhi Todkar**, BE (CSE) student of Dnyanshree Institute of Engineering and Technology, Satara, Maharashtra. Area of Interest: Big data Analytics. Deep Learning, Machine Learning.