CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090623
http://magellanes.com/

PUBLIC SAFETY AND UNUSUAL BEHAVIOUR MONITORING IN PUBLIC AREAS

**Amit Madan, Akshat Sharma, Anubhav Tyagi, Aniket Chauhan,
Mr.Jagbeer Singh**

[#]Computer Science and Engineering, Meerut Institute of Engineering and Technology (MIET), Meerut, India

amit.madan.cse.2020@miet.ac.in , akshat.sharma.cse.2020@miet.ac.in ,
anubhav.tyagi.cse.2020@miet.ac.in , aniket.chauhan.cse.2020@miet.ac.in ,
jagbeer.singh@miet.ac.in

Abstract -

Unusual Behavior Real-time in-demand research is being done on recognition from surveillance images. Quick abnormal event detection meets the growing demand for processing a vast amount of surveillance footage. Video surveillance represents a burgeoning field where artificial intelligence (AI), machine learning (ML), and deep learning technologies are actively employed. Computers possess the ability to emulate human thought processes, thanks to artificial intelligence. Machine learning, a crucial facet of this, involves acquiring knowledge based on training of information and making estimations determined by forthcoming information. Utilization of deep learning has become feasible because the accessibility of substantial datasets and powerful GPU processors (Graphics Processing Units).
Unusual Behavior Real-time in-demand research is being done on recognition from surveillance images.

To prevent theft instances, Detecting potentially suspicious human activity through video monitoring is of paramount importance, terrorists from using abandoned goods as explosives, Vandalism, altercations, and personal attacks, and fire in many extremely sensitive places like as well as airports, refineries, nuclear power plants, bus and train terminals, shopping centres, banks, hospitals, universities, borders, and so on. Campuses of universities and other academic institutions can utilise video surveillance to keep an eye on students' activities to protect property from theft and destruction. Additionally, it will assist in reducing student fights and improper behaviour. For the protection of the faculty and students, it will keep an eye on the outside of the university campus and other academic buildings. When exams are being given, video surveillance may be utilised to keep an eye on any suspicious behaviour among the students in the exam room.

Introduction -

Currently, the most effective security measure that a facility may have is CCTV surveillance. It is something we can probably find anywhere, in clinics, shopping centers, colleges, and so on. But consider a campus where more than a hundred closed-circuit television cameras are dispersed across different buildings, dorms, classrooms, canteens, sport fields, auditoriums, etc. It is neither practical nor effective to monitor all actions manually. Searching for past occurrences of an activity in a recorded dataset manually would be highly inefficient and time-consuming

Deep neural networks are notable for their effectiveness in addressing intricate learning tasks. Models built on deep learning exhibit automated feature extraction and the creation of high-level representations from visual input. This process is more generic due to the fully automated nature of feature extraction. Convolutional neural networks (CNNs) excel in directly extracting visual patterns from image pixels. In video stream scenarios, long-term dependencies can be identified using long short-term memory (LSTM) models. LSTM networks possess the capability to store data and capture sequential information effectively..

The envisioned system aims to oversee campus activities by analyzing CCTV camera footage. Intelligent video surveillance is contingent on adept human behavior recognition and event detection. The comprehensive training process for a system to monitor things can be installed. delineated into some distinct phases:

The the development of data, pattern learning, and the conclusion process constitute three key phases in the development and implementation of the surveillance system.

Detection of human behavior finds applications in various real-world scenarios, as an example of intelligent footage monitoring and the evaluation of consumer purchase patterns. Video surveillance is extensively utilized across public, outdoor, and indoor areas, playing a crucial role in monitoring activities and ensuring security.Security cameras are integral components of ensuring safety and security in today's environment. With the prevalence of Closed Circuit Television (CCTV) cameras, manually reviewing every incident has become a daunting task. Even when incidents have already occurred, the manual search for them within recorded videos is time-consuming. Analysis of unusual occurrences in video data represents a burgeoning area within domain of automated footage monitoring systems. Detecting human conduct in footage monitoring systems involves autonomous approach to identify anything which refers to activities that are suspicious. Numerous efficient algorithms are employed for this purpose.

Our plan is to develop a system that utilizes Long-Term Recurrent Convolutional Networks (LRCNs) to analyze CCTV footage and identify anomalous behaviors such as fighting, strolling, and running. If the system detects any suspicious activity, the user will be notified through the creation of an alarm.To prevent theft, attacks, snatching, and other aberrant activities, intelligent models are deployed to observe human actions at various locations, including terminals for transportation such as buses and trains, airlines, educational institutions, sports venues, medical centers, parking areas, shopping centers, and a lot of other buildings including other public spaces.

Literature Review -

The proposed related papers present various techniques for recognizing an individual activity in videos. The main focus of these efforts was to identify any unusual or suspicious activity within surveillance videos.

[1] The advanced motion detection (AMD) method turned out to be utilized in order to identify unlicensed entry to a restricted area. Initially, objects were recognized through background subtraction, and subsequently, they were extracted from sequences of frames. This innovative approach enhances security by swiftly identifying and responding to potential breaches in restricted areas, ensuring the effective protection of assets and personnel. The uncovering of dubious actions was the second stage. The system's algorithm has the benefit of processing videos in real time, while requiring little computational complexity. It should be noted that the system had storage service restrictions, but it could also be utilized to build a sophisticated system for gathering video in monitoring zones.

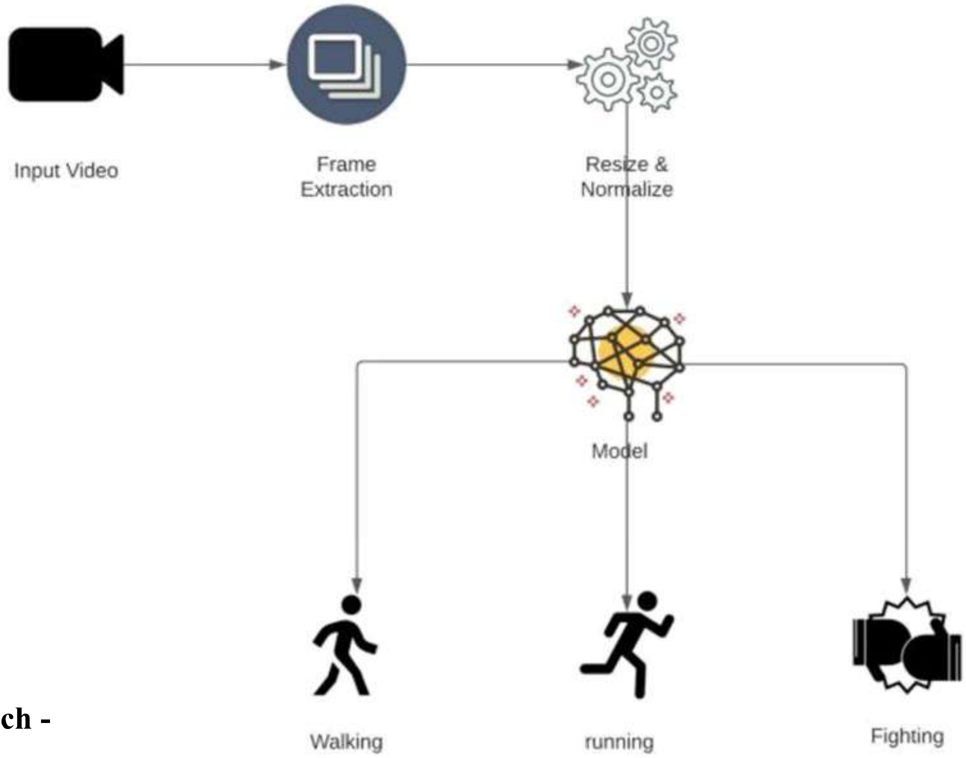
[2] A method for detecting violence in real time, based on deep learning(DL), was created in order to curb violent behavior within crowds. Frames were extracted from live videos and processed in a spark environment. Upon detecting any violent acts, the system promptly notifies the security staff. This technology enables real-time detection of actions in the video, alerting security personnel to intervene and prevent violence before it escalates.

Utilizing individual tracking can be employed to pinpoint unusual occurrences in video footage. The backdrop subtraction method is applied to identify human subjects in the video, and the features are extracted by a Convolutional Neural Network (CNN) before being fed into a Discriminative Deep Belief Network (DDBN).

Before being sent to the DDBN, videos of some dubious instances are labeled so that traits can be extracted from them. Many suspicious activities were discovered in the provided video after features retrieved from the designated sample video of classified suspicious behaviors were contrasted to features acquired using CNN employing a DDBN.

[4] Considering current time recognition of violence by deep learning was designed to proactively avoid acts of violence by spectators or athletes during sporting events, particularly in the context of football. Frames were extracted from live videos and processed in a spark ecosystem. In the event that the system detects any violence related to football, authorities are promptly notified. This technology aims to enhance security and ensure a timely response to potential incidents during sports events.

The technology actively identifies actions in real-time footage and notifies security personnel to prevent violence before it happens. Through the utilization of the VID dataset, the system demonstrated an accuracy rate of 94.5% in accurately identifying incidents of violence in soccer fields.



Approach -

The suggested approach involves utilizing CCTV footage to monitor student behavior on campus and promptly notifying the relevant authorities of any suspicious activities.

The architecture is broken down into multiple phases, such as video pre-processing, video capturing, and class prediction. Videos are split into three groups using this technique:

- Fighting on campus
- Deceptive class
- Running or strolling
- typical class

A. Video capture

Installing CCTV cameras and the acquisition of footage serves as the first procedure in a monitoring system for visual monitoring system. Many different cameras capture in different video formats, including the entire surveillance area.

B. Video Pre-processing

Thirty frames, evenly spaced in time, are picked from each of the recorded videos for pre-processing.

The 30 retrieved frames that are were adjusted to 64 × 64 and then read into a numpy array with the following aspects: 64 x 64 x 3. (RGB x Image Height x Image Width)

Python OpenCV Library.

Then, by dividing each value in the frame by 255, it is normalised.

Every footage has averaged frames are retained as a sequence in a numpy array with the following aspects: 30 by 64 by 64 by 3.

C. Class Prediction

The model predicts the class of given video after receiving the numpy array as input.

Proposed System And Design-

The LRCN is employed to identify unexpected actions. Accurately classifying strange behaviors depends on your ability to recognize the there is time-based data in video recording. CNN has been mostly used recently to extract key features from every video frame. In order to classify the input data efficiently, the features must be extracted from CNN; therefore, CNN requires being capable of identify and then locate the necessary properties derived from footage's scene.
The multiple frame sequence of the footage is revoked as well as issued to LRCN Model.

The recommended approach makes use of CCTV footage to monitor student behavior on campus and notify the proper authorities when anything questionable occurs. The architecture is broken down into multiple phases, such as video pre-processing, video capturing, and class prediction. The videos are split into three groups using this method:

–Fighting on campus
–Deceptive class
–Running or strolling
–typical class

Implementation of our Upgraded Model-

1. DATASET

Running and walking activity identification using the KTH dataset.

KTH Dataset - cvap/actions at https://www.csc.kth.se likewise the Kaggle dataset for fighting detection.

Movies-Fight-Detection Dataset on Kaggle: https://www.kaggle.com/naveenk903/

The KTH dataset is a typical dataset with combination of sequences denoting 6 ventures, with hundred sequences for every kind of operation. The filmed footage was captured at 25 frames per second, and every section has over six hundred frames in addition, the filmed footage was captured at 25 frames per second.

More than a hundred clips from films and YouTube videos are included in the Kaggle Dataset It could be utilized to train inappropriate conduct (fighting).

2. Data Pre-processing

a) Utilizing library of the OpenCV , movies are read via specific Class folders, and a NumPy array is employed to store the associated Class labels, enhancing the efficiency of data organization and processing in the system.

a) Making a single sequence out of a video by dividing it into Different frameworks, Library of OpenCV is used to research each footage, and only thirty frames are scanned at consistent gaps so to produce a thirty frame pattern.

b) Transforming: Image scaling is necessary when there is a need to change the total number of pixels. We reduced the width and height of each frame to 64 pixels to be able to maintain the integrity in terms of what are the inputted photographs in terms of architecture.

c) Normalization: The learning system will be able to identify important features from the photographs faster thanks to normalization. To normalize the expanded frame, we divided it as a result.

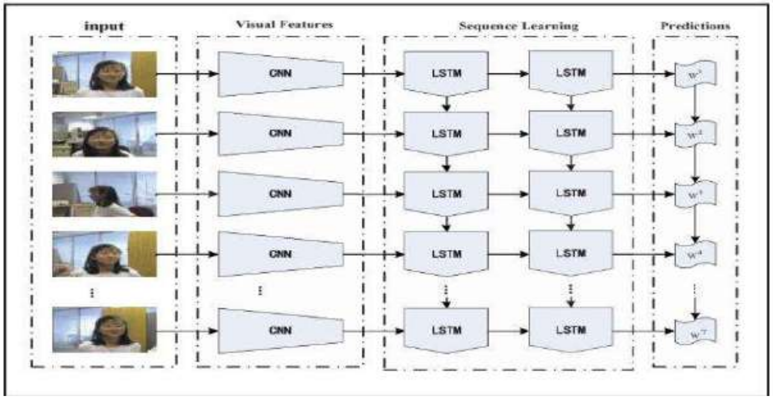
d) Implementation of numpy arrays for collection of data: The set of thirty scaled and standardized images is saved in a numpy array and used for providing guidance for the model.

3. Train Test Split Data

75 percent of the data is utilized for training, while 25 percent is used for testing.

4. Model Creation

In our suggested method for suspicious conduct authentication via observing video, LRCN is employed to consider a deep neural network which is for efficient processing.



5. Model Training

Three types of events have been taught to the programme: battling, running around, and walking.

The training set along with the following hyperparameters are used to train the model:

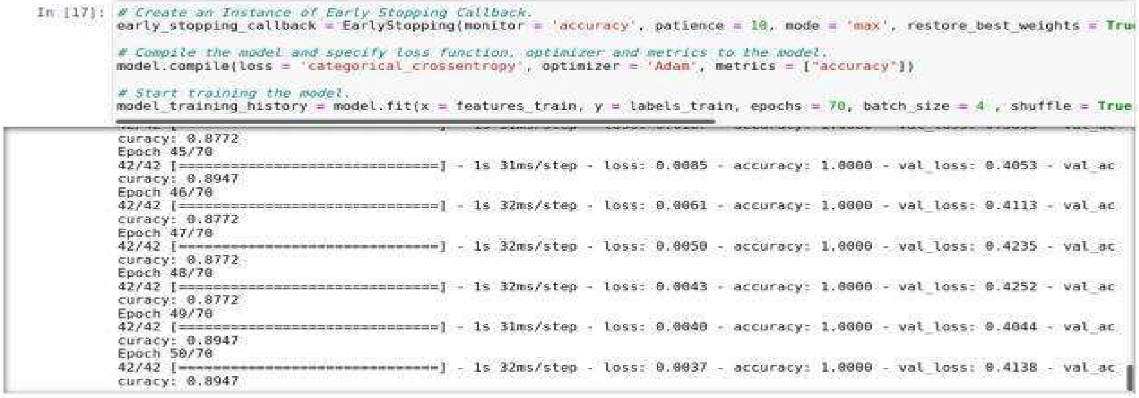
Epochs is equivalent to 70

batch size of input is 4

The split value for validation is 0.25

Output Screens -

Pattern Training :

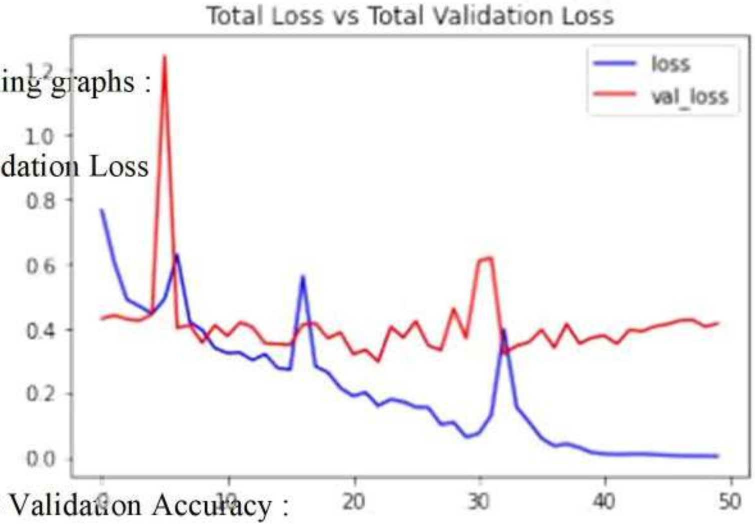


```
In [17]: # Create an Instance of Early Stopping Callback.
         early_stopping_callback = EarlyStopping(monitor = 'accuracy', patience = 10, mode = 'max', restore_best_weights = Tru

         # Compile the model and specify loss function, optimizer and metrics to the model.
         model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

         # Start training the model.
         model_training_history = model.fit(x = features_train, y = labels_train, epochs = 70, batch_size = 4 , shuffle = True
```

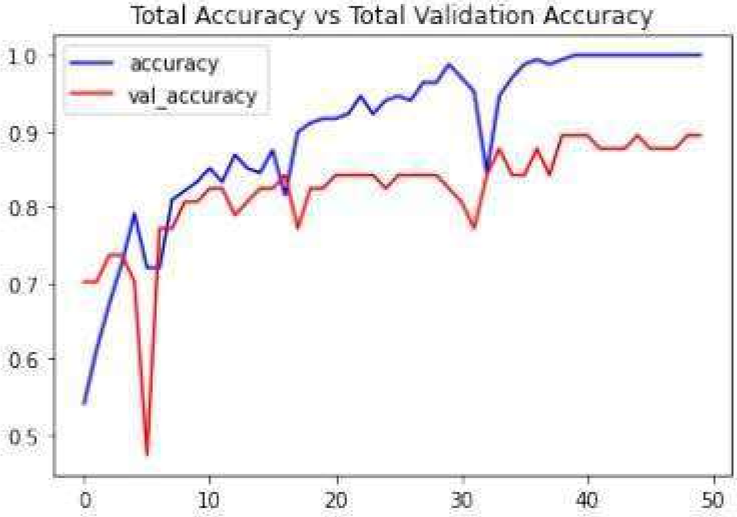
```
curacy: 0.8772
Epoch 45/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0085 - accuracy: 1.0000 - val_loss: 0.4053 - val_ac
curacy: 0.8947
Epoch 46/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0061 - accuracy: 1.0000 - val_loss: 0.4113 - val_ac
curacy: 0.8772
Epoch 47/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0050 - accuracy: 1.0000 - val_loss: 0.4235 - val_ac
curacy: 0.8772
Epoch 48/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0043 - accuracy: 1.0000 - val_loss: 0.4252 - val_ac
curacy: 0.8772
Epoch 49/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0040 - accuracy: 1.0000 - val_loss: 0.4044 - val_ac
curacy: 0.8947
Epoch 50/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0037 - accuracy: 1.0000 - val_loss: 0.4138 - val_ac
curacy: 0.8947
```

Model Training graphs :

Loss vs Validation Loss



Accuracy vs Validation Accuracy :



Experimental Result Discussion -

Based on the model we provided, the system detects abnormal behavior in the movie with an accuracy of 82% on the data set we generated. In our previous model, we employed the 16-layer VGG-16 model.

As a result, it required a lot of time and was not suitable for real-time detection. But the LRCN model made it speedier and capable of real-time detection by cutting the number of layers to 11. We reduced the size of our frames from 224 pixels to 64 pixels in order to save memory, and we added more films to our dataset in order to enhance accuracy. Videos showing odd behavior—fighting, for example—are part of the dataset for the proposed model.

The suggested model's results are shown in the photos below.

```
#@title Accuracy on test dataset


# Calculate Test Accuracy
 Dataset acc = 0

for i in range(len(features_test)):

  predicted_label = np.argmax(model.predict(np.expand_dims(features_test[i],axis =0))[0])
  actual_label = np.argmax(labels_test[i])

  if  predicted_label  ==  actual_label:
     acc += 1

acc  =  (acc  *  100)/len(labels_test)
print("Accuracy =",acc)
```

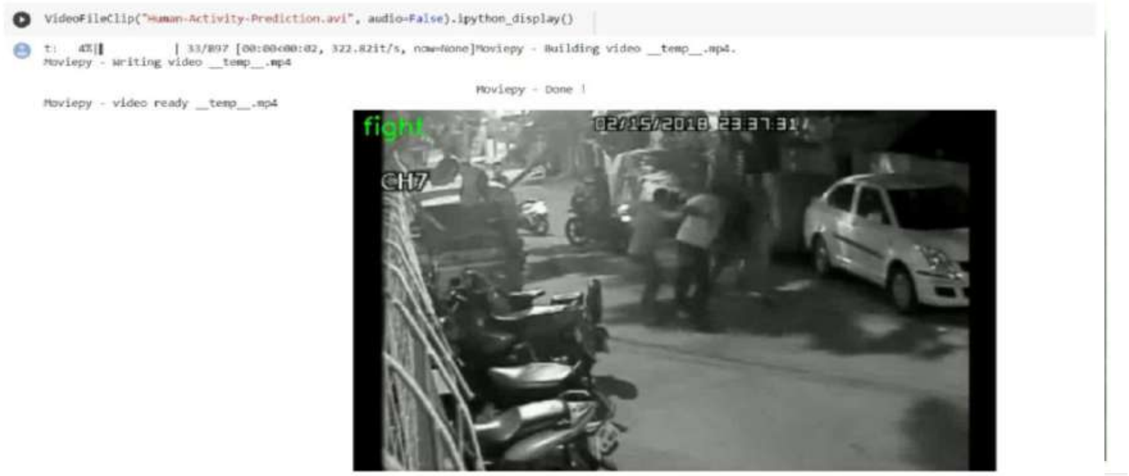
Accuracy = 86.66666666666667

```
[] single action prediction ("Predict/fight.avi",SEQUENCE LENGTH)

Action Anticipated: fight

With certainty; 0.9965279698371887

[] prediction of a particular action "Predict/running.avi," SEQUENCE LENGTH

Action Predicted: running

Assurance: 0.98820739.945123

[] foresee one action (SEQ LENGTH, "Predict/walking. avi")
```

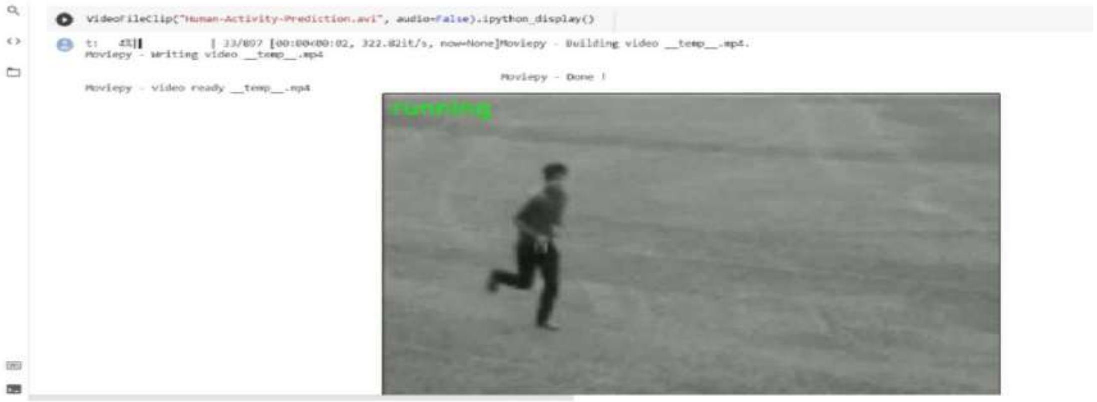
Action Predicted; walking

Confidence; 0,9890509250793457

```
VideoFileClip("Human-Activity-Prediction.avi", audio=False).ipython_display()

t:   4%|▌          | 33/897 [00:00<00:02, 322.82it/s, now=None]Moviepy - Building video __temp__.mp4.
Moviepy - Writing video __temp__.mp4

                                   Moviepy - Done !
```



```
VideoFileClip("Human-Activity-Prediction.avi", audio=False).ipython_display()

t:   4%|▌          | 33/897 [00:00<00:02, 322.82it/s, now=None]Moviepy - Building video __temp__.mp4.
Moviepy - Writing video __temp__.mp4

                                   Moviepy - Done !
Moviepy - video ready __temp__.mp4
```



```
VideoFileClip("Human-Activity-Prediction.avi", audio=False).ipython_display()

t:   4%|▌          | 33/897 [00:00<00:02, 322.82it/s, now=None]Moviepy - Building video __temp__.mp4.
Moviepy - Writing video __temp__.mp4

                                   Moviepy - Done !
Moviepy - video ready __temp__.mp4
```

Conclusion -

After training on 300 videos, we developed an LRCN model that can identify actions such as fighting, walking, running, and jumping from CCTV surveillance footage with an accuracy of roughly 83 percent.

In this paper, we introduce an automated localisation approach using strong methods for computer vision for anomaly detection. The experimental outcomes indicate that...

(1) recognizing an abnormal activity and information extracted via each section contributes to discovery of anomalies.

(2) The method we employ can produce reliable outcomes in a variety of scenarios, e.g. road collision or accidents, theft, and a dispute, running

(3) Employing strong BG reduction can aid in determining ROI to be accurately as feasible. The suggested abnormal activity detection model can achieve 82.66% accuracy. It is interesting that our training network is a weakly-supervised one. More generally, We believe that the visual focus zone extraction technique we developed will be helpful for numerous other online jobs, such as adaptation and categorization of video objects. Our work is restricted to anomalous events that include moving objects.

References -

[1] P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli Sravya Reddy,“Inspection of suspicious

human activity in the crowdsourced areas captured in surveillance cameras”,International

Research Journal of Engineering and Technology (IRJET), December 2017.

[2] Jitendra Musale,Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata

Tadge, “Suspicious Movement Detection and Tracking of Human Behavior and Object with

Fire Detection using A Closed Circuit TV (CCTV) cameras ”, International Journal for

Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII

December 2017.

[3] https://www.researchgate.net/publication/288703678_Detection_of_abno rmal_behaviors_in_crowd_scene_A_review

[4] https://www.researchgate.net/publication/318382870_Activity_Recogniti

on_and_Abnormal_Behaviour_Detection_with_Recurrent_Neural_Net works