International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

INTELLIGENT VIDEO SURVEILLANCE USING DEEP LEARNING

Kadikonda Mounika[1], Vancha Vaishnavi Reddy[2], Asma Begum[3]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract. Video surveillance, also known as CCTV (closed-circuit television), is a rapidly growing industry that has been around for more than 30 years and has seen its fair share of technological advancements. In today's world, video surveillance has become an essential component for ensuring public safety. Security can be defined in a variety of ways depending on the context, such as theft detection, violence detection, explosion risk, and so on. The term "security" in crowded public places refers to almost any type of abnormal event.Intelligent video surveillance delivers cutting-edge smart security by recording unexpected activities in homes, offices, and public locations based on the preferences of the user. In the event of an abnormal incident, the video surveillance system will actively respond to detect actions in advance through real-time monitoring and promptly communicate data. The primary focus is on the use of deep learning techniques to provide tracking of a moving target, high-definition picture quality, and night vision technology triggered by motion sensors, which means the system isn't running when nothing is happening at your location, automatic audio and visual detection, video recording initiation, and detection of suspicious activities that can trigger and alert the systems in all climate conditions.Deep learning technology will be used in the data processing model design to visualise data for abnormal activities, and this design also proposes an intelligent surveillance system to quickly and effectively detect activities by sending a video image and an alert message to the web via real-time processing. Recent improvements in computer vision, particularly deep learning approaches, have opened up new possibilities for these systems, enhancing their capabilities and launching new research areas in this field.

Keywords: Video Surveillance, Security, Abnormal event, Detection, Visualization, Processing, Intelligent Surveillance System, Normal Event, Climate Conditions.

1. Introduction

1.1 About Project

The interest and use of image processing and video analysis has been increased now a days and it has been unprecedented due to its importance in finding out and summarization and recognizing of actions. This project explains about how to process the video and image in order to find difference between them. We developed a system which classifies a video into three classes:

- Illegal or aggressive activity
- Potentially doubtful
- Secure

Our plan is to resolve this difficulty is a structured design base on convolutional and recurrent neural networks. Through this project, it is 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 outcomes that we achieve.

1.2 Objectives of the Project

The main objectives identified which illustrate the relevance of the topic are listed out below.

1. Continuous monitoring of videos is difficult and tiresome for humans.
2. Intelligent surveillance video analysis is a solution to laborious human task.
3. Intelligence should be visible in all real-world scenarios.
4. Maximum accuracy is needed in object identification and action recognition.
5. Tasks like crowd analysis still needs lot of improvement.
6. Time taken for response generation is highly important in real world situation.
7. Prediction of certain movement or action or violence is highly useful in emergency situation like stampede.
8. Availability of huge data in video forms.

The majority of papers covered for this survey give importance to object recognition and action detection. Some papers are using procedures similar to a binary classification that whether action is anomalous or not anomalous. Methods for Crowd analysis and violence detection are also included.

1.3 Scope of the Project

Intelligent Video Surveillance is a rapidly growing industry which is used in

Remote video monitoring: To protect against theft, burglaries, and dishonest employees.

Facility Protection: To protect the perimeter of the property or the perimeter of buildings.

Monitor operations: To monitor day-to-day operations and as a tool to streamline operations.

Loss prevention: To protect assets.

Employee safety: For compliance with safety regulations and also to protect the employer in civil proceedings.

Parking lots, Event video surveillance, Public Safety, Traffic monitoring and many more.

2. Literature Survey

2.1 Existing System

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 mode.

The present system is Convolution Neural Network. In deep learning, a Convolutional neural network (CNN, or ConvNet) is a class of artificial neural network, most commonly applied to analyze visual imagery. A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

A ConvNet is able to **successfully capture the Spatial and Temporal dependencies** in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

The Convolution Neural Network has the following drawbacks

- CNN do not encode the position and orientation of object.
- Lack of ability to be spatially invariant to the input data.
- Lots of training data is required.
- While CNNs are translation-invariant, they are generally bad at handling rotation and scale-invariance without explicit data augmentation.
- Its less expensive and time taking process.

2.2 Proposed System

- For Intelligent video surveillance, we will introduce a spatio temporal autoencoder, which is based on a 3D convolution network.
- Spatio-Temporal AutoEncoder (ST AutoEncoder or STAE), which utilizes deep neural networks to learn video representation automatically and extracts features from both spatial and temporal dimensions by performing 3-

dimensional convolutions.

- An autoencoder is an unsupervised learning approach that aims to learn an identity function, that is, the input and the expected output are equal. The goal is to reveal interesting structures in the data by placing constraints in the learning process.
- A video autoencoder, in which a set of N grayscale frames is arranged as an N dimensional image for input. This image is passed through an encoder, whose output is a three-dimensional image. This image is then passed to the decoder, where the output is again N dimensional and represents the reconstructed video. The purpose of this autoencoder is to shrink the video to a single image representation by learning how to reconstruct a set of frames using only a 3-channel tensor.
- The decoder is even simpler, consisting of only a 3×3 convolution layer with linear activation and N filters, where N corresponds to the same number as the frames in the input. The lack of batch normalization and a non-linear activation forces the model to concentrate most of the reconstruction capacity on the intermediate representation.
- Maintaining a simple decoder causes the encoder output to be encapsulated more clearly the structures of the input data, so the generated images still make sense and resemble the original video frames.

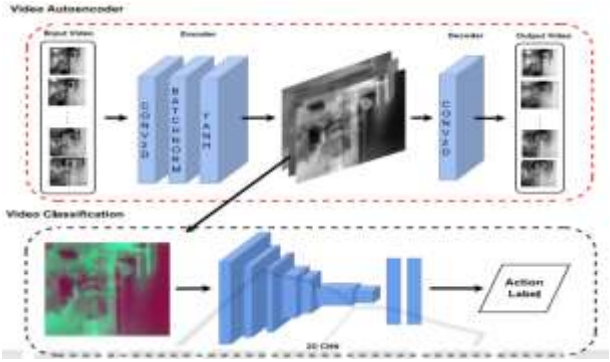

Fig.1. Proposed System

3. Proposed Architecture

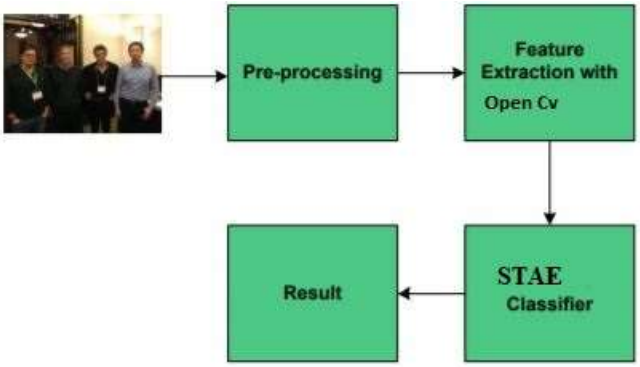

Fig.2. System Architecture

DATA FLOW DIAGRAM

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

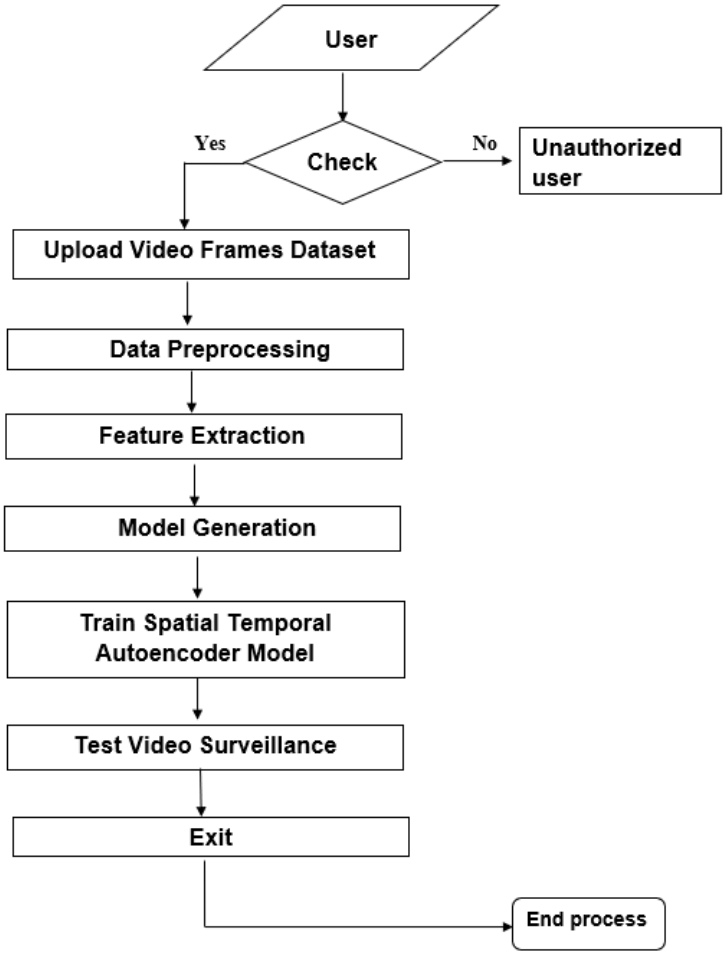

Fig.3. Data Flow Diagram

4. Implementation

4.1 Algorithm

- Spatio Temporal autoencoder is a deep learning algorithm which is based on a 3D convolution network. It is used to learn the regular patterns in the training videos.
- Consists of two parts — spatial autoencoder for learning spatial structures of each video frame, and temporal encoder-decoder for learning temporal patterns of the encoded spatial structures.

- The encoder part extracts the spatial and temporal information, and then the decoder reconstructs the frames.
- We train an autoencoder for abnormal event detection. We train the autoencoder on normal videos. We identify the abnormal events based on the Euclidean distance of the custom video feed and the frames predicted by the autoencoder.
- We set a threshold value for abnormal events. We can vary this threshold to experiment getting better results.

We will use above frames to train STAE model and we have designed following modules to complete this project

1) Upload Video Frames Dataset: using this module we can upload dataset video frames to application
2) Dataset Preprocessing: using this module we will read each image and then extract each pixel and then normalize image pixel values between 0 and 1
3) Train Spatial Temporal AutoEncoder Model: in this module we will input process and normalize images to encoder model to generate STAE model
4) Test Video Surveillance: using this module we will upload test image and then extract each frame from video and then apply STAE model on frame to predict event and this event will be compare with test frame using Euclidean distance and if this distance cross normal behaviour threshold then application will display alert message.

4.2 Code Implementation

Tensorflow. TensorFlow is an amazing information stream in machine learning library made by the Brain Team of Google and made open source in 2015.It is intended to ease the use and broadly relevant to both numeric and neural system issues just as different spaces. Fundamentally, TensorFlow is a low level tool for doing entangled math and it targets specialists who recognize what they're doing to construct exploratory learning structures, to play around with them and to transform them into running programs.

Python 3.7. Python is broadly utilized universally and is a high-level programming language. It was primarily introduced for prominence on code, and its language structure enables software engineers to express ideas in fewer lines of code. Python is a programming language that gives you a chance to work rapidly and coordinate frameworks more effectively.

Anaconda3 5.3.1. Anaconda is a free and open-source appropriation of the Python and R programming for logical figuring like information science, AI applications, large-scale information preparing, prescient investigation, and so forth. Anaconda accompanies in excess of 1,400 packages just as the Conda package and virtual environment director, called Anaconda Navigator, so it takes out the need to figure out how to introduce every library freely. to Anaconda appropriation that enables clients to dispatch applications and oversee conda packages, conditions and channels without utilizing command line directions.

5. Result

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

After implementing the algorithm we have determined whether an event is normal or an abnormal event as shown below.

Main Screen

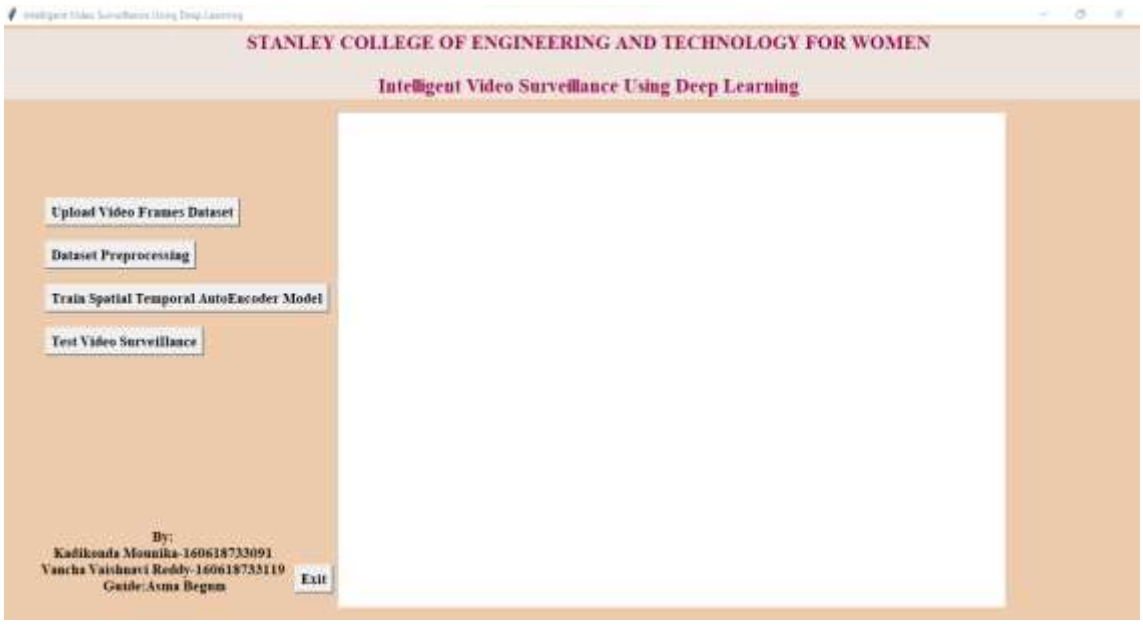

Fig.4. Home Page

#upload the dataset

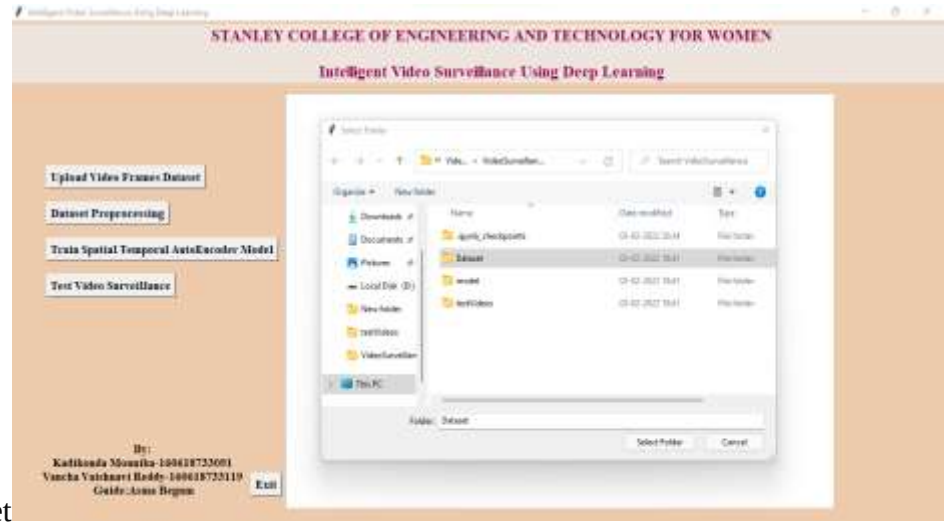

Fig.5.Dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

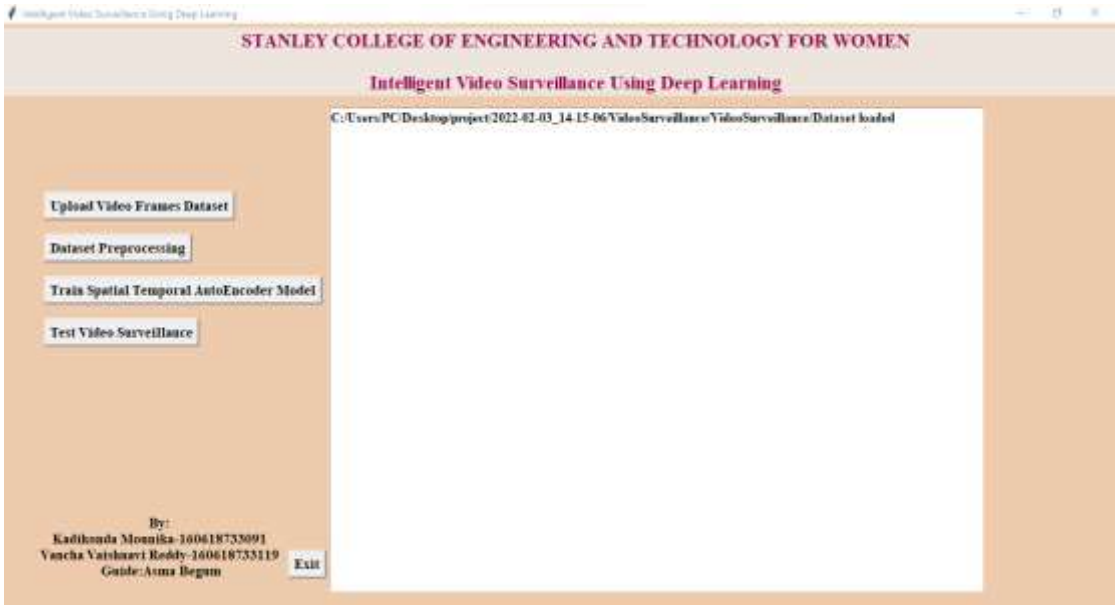

Fig.6. Uploading dataset

Processing the data

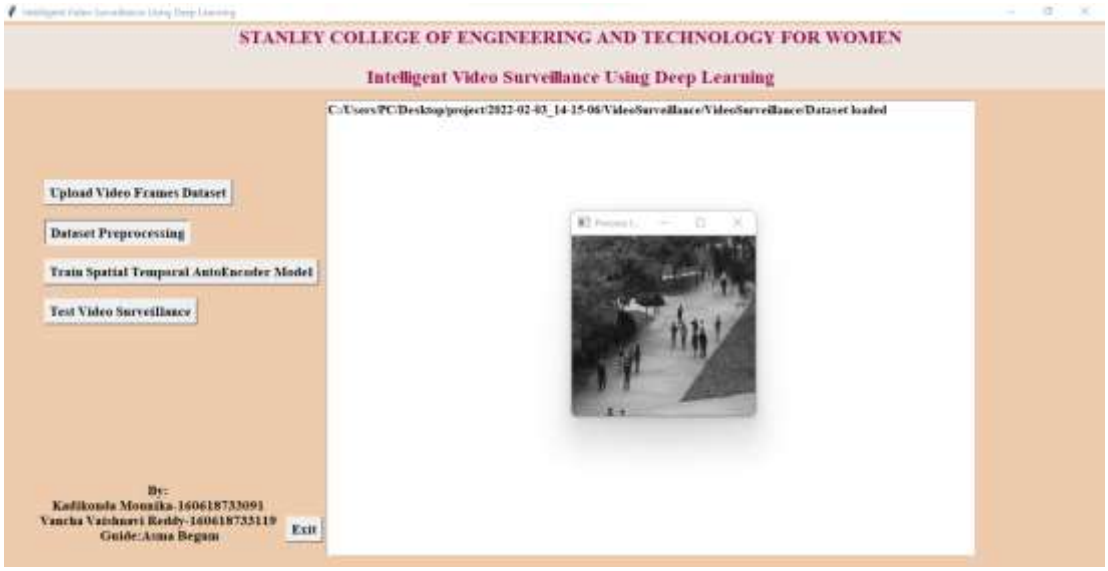

Fig.7. Dataset Processing

Number of images in dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

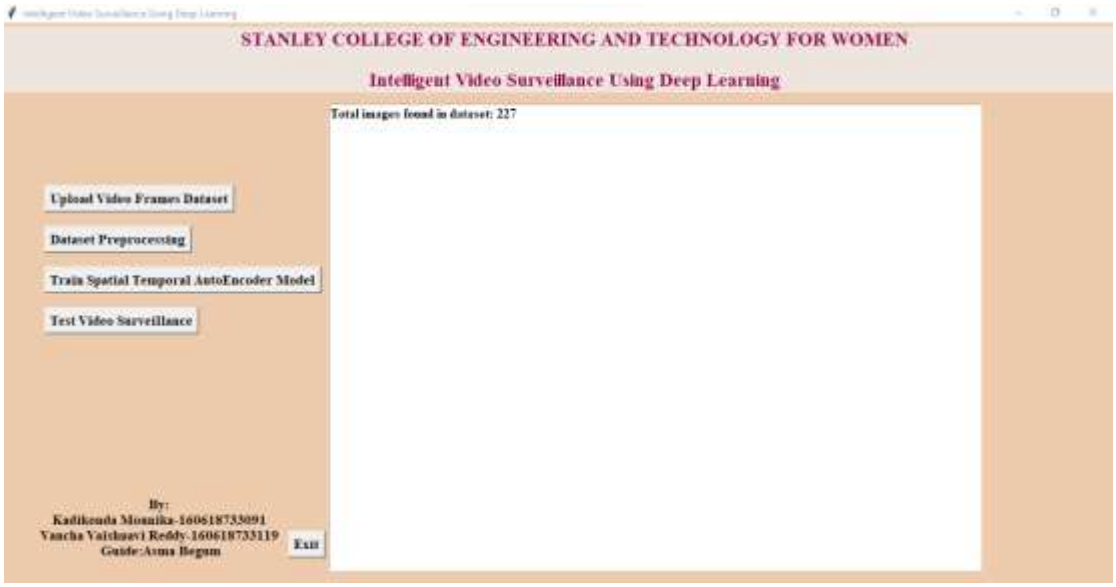

Fig.8. Total images

Training the Spatio Temporal Autoencoder

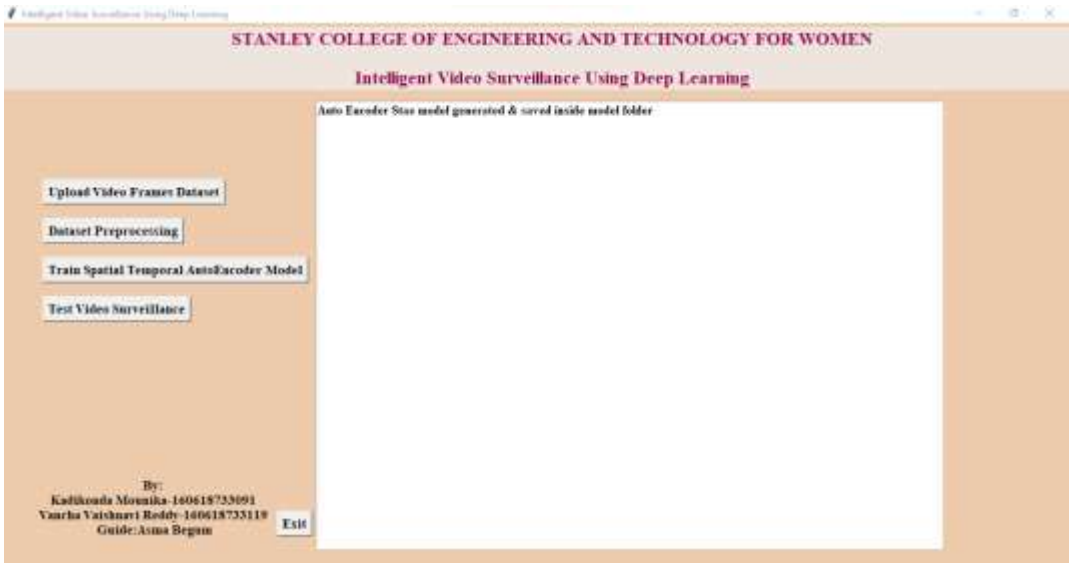

Fig.9. Training the Model

#Uploading the video and testing the video surveillance

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

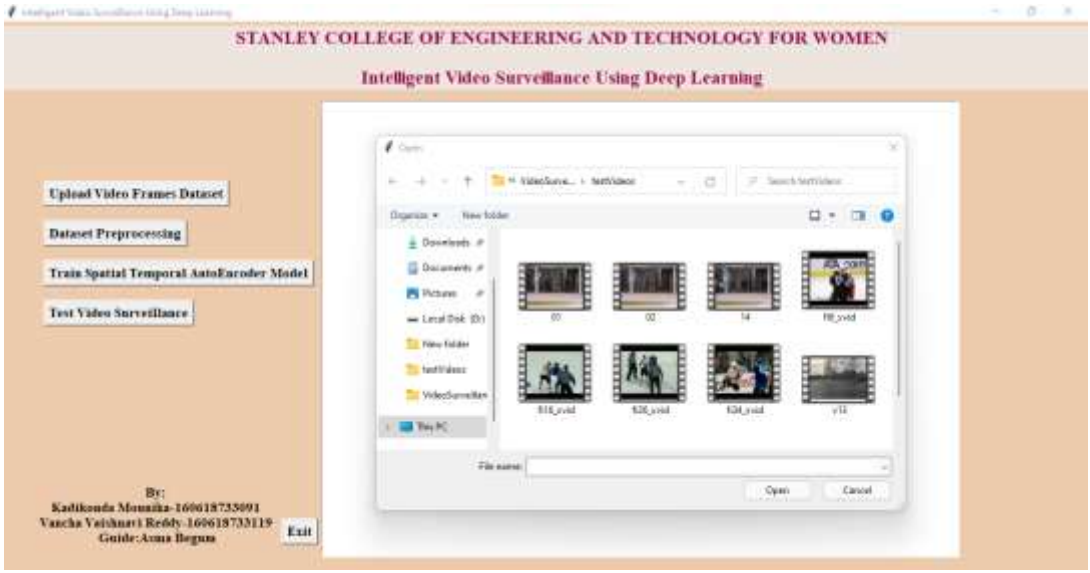

Fig.10. Uploading the video

Detecting the video whether it is normal event or abnormal event

Fig.11. Normal Event

#Abnormal Event

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fig.12.Abnormal Event

#Exiting the page

Fig.13. Exit Page

6. Conclusion

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. 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.

7. 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.

8. References

[1] Ahmed SA, Dogra DP, Kar S, Roy PP. Surveillance scene representation and trajectory abnormality detection using aggregation of multiple concepts. Expert Syst Appl. 2018;101:43–55 .

[2] Arunnehru J, Chamundeeswari G, Prasanna Bharathi S. Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos. Procedia Comput Sci. 2018;133:471–7.

[3] Guraya FF, Cheikh FA. Neural networks based visual attention model for surveillance videos. Neurocomputing. 2015;149(Part C):1348–59.

[4] Huang H, Xu Y, Huang Y, Yang Q, Zhou Z. Pedestrian tracking by learning deep features. J Vis Commun Image Represent. 2018;57:172–5.

[5] Huang W, Ding H, Chen G. A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance. Signal Process. 2018;142:104–13.

[6] Kim H, Kim T, Kim J, Kim JJ. Deep neural network optimized to resistive memory with nonlinear current–voltage characteristics. J Emerg Technol Comput Syst. 2018;14:15.

[7] Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. Procedia Comput Sci. 2018;132:1706–17.

[8] Ribeiro M, Lazzaretti AE, Lopes HS. A study of deep convolutional auto-encoders for anomaly detection in videos. Pattern Recogn Lett. 2018;105:13–22.

[9] Shao L, Cai Z, Liu L, Lu K. Performance evaluation of deep feature learning for RGB-D image/video classification. Inf Sci. 2017;385:266–83.

[10] Wu G, Lu W, Gao G, Zhao C, Liu J. Regional deep learning model for visual tracking. Neurocomputing. 2016;175:310–23.