Abnormal Event Detection Using CCTV Camera

[1]Andrew Joemon
Student, Computer Science and Engineering Department
Sahrdaya College of Engineering & Technology
Thrissur, India

[2]Ashwin Prakash V
Student, Computer Science and Engineering Department
Sahrdaya College of Engineering & Technology
Thrissur, India

[3]Awin Chris Joy
Student, Computer Science and Engineering Department
Sahrdaya College of Engineering & Technology
Thrissur, India

[4]Bibin Thomas
Student, Computer Science and Engineering Department
Sahrdaya College of Engineering & Technology
Thrissur, India

[5]Divya R
Asst. Professor, Computer Science and Engineering Department
Sahrdaya College of Engineering & Technology
Thrissur, India

Abstract:- Abnormal event detection, human behavior detection, as well as object recognition plays a vital role in the creation of a smart CCTV system. These systems make it possible to detect abnormal events in an environment, abnormal behaviors by humans and the state of alert in the environment. Machine Vision property along with Machine Learning are used in these systems to detect as well as identify the particular anomalies that arise in the video feed from the CCTV. Frame by frame processing is commonly used and Supervised Learning is the commonly used training method for these systems. However, since the anomalies are of many different kinds and also because it is not feasible to pre-detect and train all types of anomalies, supervised learning is being replaced by unsupervised learning and semi - supervised learning for training the system. This system provides a means of minimising or removing the human workload that has to be put on to manually detect and create an alert on detection of an abnormality in the live feed provided by the CCTV. Also the system increases the storage efficiency by storing only the abnormal events in original quality and storing the normal scenarios in low quality for archiving. Also this system provides an extension of creating a distributed abnormality classification system, where only the abnormal events are sent on to different dedicated systems to classify the abnormality.

Keywords:- *Convolutional Neural Network; Anomaly detection; Long Short-Term Memory;*

I. INTRODUCTION

In the present day world, CCTV cameras are seen in every nook and corners of our surroundings. The primary objective of a CCTV camera is in the post scenario analysis as the CCTV records everything and these recordings are used only after an event has occurred in order to determine its aspects. As the world today demands the system to be more active than passive, technologies such as machine vision along with sophisticated machine learning algorithms are being incorporated to develop new systems and thereby send alerts to the respective authorities as soon as the anomalies are detected. The analysis of crowd behavior and object detection can be deployed in many applications such as theft detection in crowded environments. As it is quite likely for people to be positioned at varying locations in the crowd and may move in diverse directions, it becomes a challenging task to find the effective features of the crowd and as a result, the higher level analysis of crowd behavior becomes a tedious task.

Anomaly detection is of considerable significance for video surveillance systems. Most of the systems which are proposed use methods like Convolutional Neural Network (CNN) and LSTM (Long Short Term Memory) networks to effectively train the system in order to detect the anomalies in both supervised as well as unsupervised manners. Supervised learning method emphasizes on using the existing knowledge about a particular anomaly to train a system while Unsupervised learning method on the other hand, tries to learn normality rather than learning abnormality. This implies that if a large deflection is seen from normal behavior, it provokes abnormality. A highlighting aspect of our system is that we can store original quality video snippets of the abnormal events while the normal recordings will be stored in low quality for archiving. Also this system provides an extension of creating a distributed abnormality classification

ISSN No:-2456-2165

system, where only the abnormal events are sent on to different dedicated systems to classify the abnormality.

II. RELATED WORK

A. Practical Automated Video Analytics for Crowd Monitoring and Counting

Kang Hao Cheong, Sandra Poeschmann, Joel Weijia Lai, Jin Ming Koh, U. Rajendra Acharya, Simon Ching Man Yu and Kenneth Jian Wei Tang [1], "Practical automated video analytics for crowd monitoring and counting", In this paper, video surveillance is integrated with computational analytics which in turn enables itself to greatly expand its functionality. A few significant methods are used in this system which includes a video processing back-end that encompasses recognition of human subjects and tracking, as well as a front-end graphical interface which is used for operators that use classical and CNN based object recognition techniques. One of the highlighting aspects of this system is that it has a high counting accuracy for idealized single-subject and is appropriate for multiple-subject scenarios which are more realistic. In this system, facial identification is not contained on top of the current object recognition as well as tracking for more enhanced surveillance capabilities. Both controlled and non-controlled tests were used to carry out the real-world validation of their solution and considerable accuracy was strongly indicated in the results,even in outdoor conditions. This system is useful for reducing human workload and is also able to accept multiple video streams from a centralized storage location. Also,it was possible to perform the data collection of crowd density and movement with better consistency and accuracy.

B. Anomaly Detection in CCTV Using Optical Flow and Convolutional Autoencoder

Elvan Duman, Osman Erdem [2], "Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder" Ayhan In this study, a convolutional autoencoder method is used to learn the pattern of normal activities in videos. The main idea of the framework is that the frames, which contain any abnormal event, give significantly different motion patterns than the normal frame. Dense optical flow maps are used as an input to the encoder. Then the network is trained with videos in which no abnormal event is included. After the training stage is properly done, the autoencoder can model the distribution of the pattern of normal motion changes. If an input video has an abnormal event, the model is expected to give a higher reconstruction error. Besides, the model was able to reconstruct optic flow maps for corresponding normal video volumes.

This framework consists of three main stages. The first stage of the framework, called preprocessing, aims at extracting dense optical flow maps of each frame. In the second stage, the convolutional autoencoder is used to obtain the spatial structure of each of the dense optical flow map volumes. The last stage includes a convolutional long short-term memory network to learn the temporal patterns of encoded optical flow maps .

C. Unsupervised Anomaly Detection and Localization Based on Deep Spatio Temporal Translation Network

Akshara Alex, Ashi Sahu, Avni Tanwar, Nisha Rathi, Kavitha Namdev [3] "Abnormal Event Detection By Machine Vision Using Deep Learning," This paper introduces a Deep Spatiotemporal Translation Network (DSTN), which is highly effective in the field of unsupervised anomaly detection and localization method. The performance of these oddment localization in the pixel level evaluation is enhanced by proposing the Edge Wrapping to reduce the noise and conquer non-related edges of abnormal objects. Accuracy on any kind of anomalies on pixel level and Computational Resource requirement for pixel level detection remains high. This outperforms other art algorithm stages with respect to the evaluation in both frame and pixel level evaluation, and the time complexity for abnormal object detection and localization events.

D. Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks

Thittaporn Ganokratanaa, Supavadee Aramvith, Nicu Sebe [4], "Unsupervised Anomaly Detection and Localization Based on Deep Spatio Temporal Translation Network". In this study, a novel method for abnormal crowd event detection in surveillance videos is used. The proposed approach is based on a new Motion Information Image (MII) model that is formulated using optical flow. Optical Flow Vectors that can generate Motion Information Image which is then trained and tested using Convolutional Neural Network is used. It have high accuracy on abnormal motion detection. Evaluations are conducted on publicly available UMN and PETS2009 datasets. The Computational Resource usage and computational time is comparatively less in this approach. The Requirement of huge datasets to train the system perfectly and identifying anomalies other than Motion Anomalies seems difficult in this case.

III. PROPOSED SYSTEM

With our system, we can automate the process of detecting abnormal events from CCTV camera feeds. CNN and LSTM technologies are used to detect anomalies in both supervised and unsupervised manner. Alert messages can be sent to authorities on detection of events. Original Quality video snippets of the abnormal event can be stored in high quality and low quality recordings for events which are considered as normal are stored separately. By implementing this system, we reduce the time taken and human workload for detecting anomalies and also the system becomes more storage efficient.

The design of this system consists of various modules or parts that have to be integrated together to complete the system. This involves the creation of Video Compressor, Anomaly Detector, Storage Management and Alert Management modules.

Video Compressor is the first module in the system which is used to compress the original video in order to store it as low quality, low resolution video for archiving.

ISSN No:-2456-2165

Anomaly Detector detects the presence of any kind of anomalies in the live feed from the CCTV system. This module is trained by Unsupervised learning method so that the system can detect all kinds of anomalies (both pre-determined and undetermined). Also the video snippet where the anomaly is occurring is stored in the original video quality and resolution.

Storage Management module is used to manage the storage of both the low quality archived video and original quality anomalous video snippets. These are stored separately for future reference.

Alert Management module is used to manage and send the alert to respective personnel on identification of the anomaly.

IV. METHODOLOGY

It is all about the reconstruction error. We have used an autoencoder to learn the regularity in video sequences. The intuition is that the trained autoencoder will reconstruct the regular video sequences with low error but will not accurately reconstruct motions in irregular video sequences.

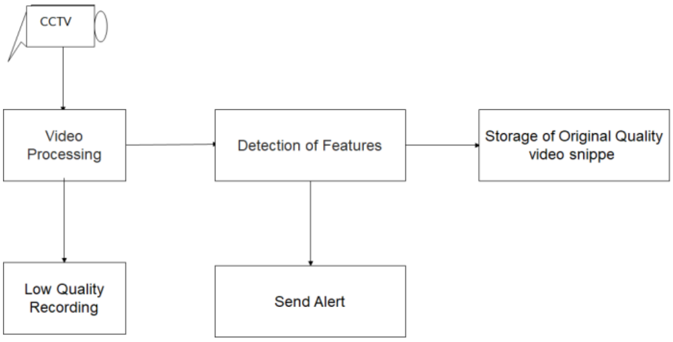

Fig.1 Architecture Diagram

We are using the UCSD anomaly detection dataset and Avenue dataset, of which UCSD dataset contains videos acquired with a camera mounted at height overlooking a pedestrian walkway. These videos mainly contain pedestrians.

Abnormal events are mainly non-pedestrian entities in the walkway which are bikers, skaters, and small carts and also include unusual pedestrian motion patterns like people walking across a walkway or at the grass surrounding it. The two parts of UCSD dataset are ped1 and ped2. We are using Ped1 Ped2 and Avenue dataset for training and Testing.

A. Preprocessing

The training set consists of sequences of normal video frames. The model will be trained to reconstruct these sequences. Initially we are taking only the 5th alternate frame from the video sequence. This is done to reduce the processing time and memory usage.Get the data ready to feed the model by following these three steps:

1. Divide the training video frames into temporal sequences with each of size 4 using the sliding window technique.
2. Resize each frame to 256 × 256 to ensure that input images have the same resolution.
3. Scale the pixel values between 0 and 1 by dividing each pixel by 256.

The number of parameters in this model is huge,Therefore a large amount of training data is required. so perform data augmentation in the temporal dimension. To generate more training sequences, and concatenate frames with various skipping strides. For example, the first stride-1 sequence is made up of frames (1, 2, 3, 4), whereas the first stride-2 sequence consists of frames (1, 3, 5, 7). We are using 2 strides to extend our data for training. Since we are only using the 5th alternating frames, our strides are (1, 6, 11, 16), and (1, 11, 21, 31).

Along with this above processing, we are reducing the resolution and quality of each and every frame, and then storing them as video. This video is the low resolution video that is stored for archiving. Since they don't contain any abnormality, they are normally not referenced much in the future.

B. Building And Training The Model

Keras is used to build the convolutional LSTM autoencoder. The below image shows the training process. Train the model to reconstruct the regular events to start discovering the model settings and architecture.

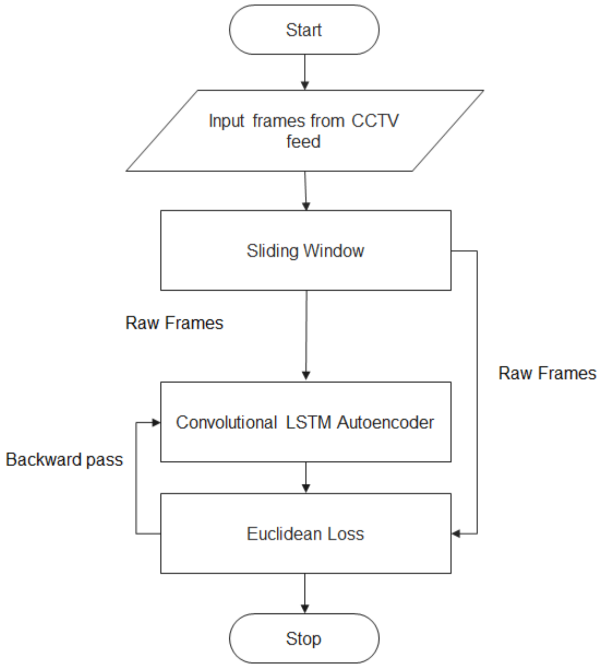

Fig.2 Training Mode Flowchart

To build the autoencoder, Encoder and the decoder should be defined. The encoder accepts as input a sequence of frames in chronological order, and it consists of two parts: The spatial encoder and the temporal encoder. The encoded features of the sequence that comes out of the spatial encoder are given to the temporal encoder for motion encoding.

The decoder will mirror the encoder for reconstructing the video sequence.

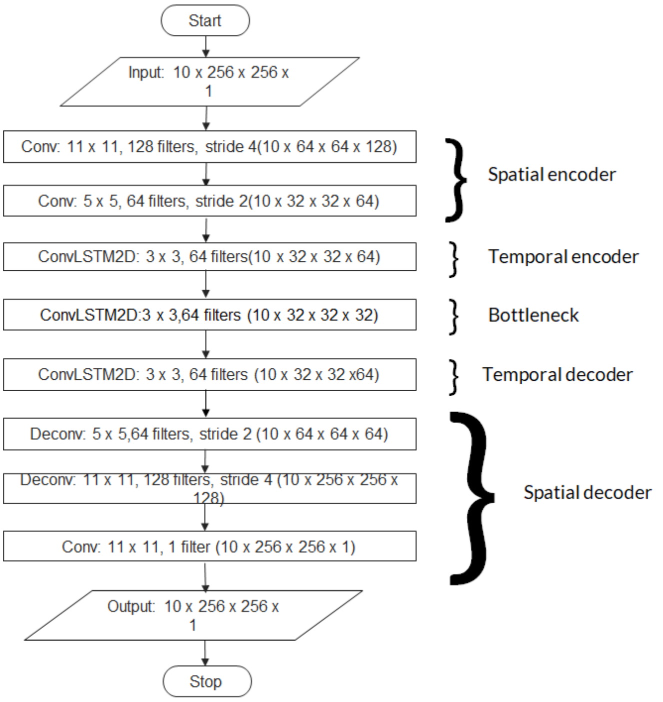

Fig.3 CNN Layers

C. Initialization and Optimization

We use Adam Optimizer with the learning rate set to 0.0001, It is reduced when training loss stops decreasing by using a decay of 0.00001, and sets the epsilon value to 0.000001. For initialization Xavier algorithm is employed , which prevents the signal from becoming too tiny or too massive to be useful because it goes through each layer.

D. Testing Phase

Each video is tested individually. UCSD Ped1 dataset provides 36 testing videos and each of these videos contains 200 frames. Since we are taking only 5th alternative frames, we get a total of 40 frames from each video in UCSD Ped1 dataset. In UCSD Ped2, we have 12 testing videos of varying numbers of frames. In Avenue Dataset, we have 21 testing videos with varying duration. Here we are only selecting 5-Alternate frames. This is done to reduce the processing time taken and memory usage. Even though we might get a much better result if we select all the frames, it is not recommended to do so as it takes a lot of time to produce the required outputs.Sliding window technique is used to get all the sequences of the 4 consecutive frames (after selecting 5-Alternating Frames). This means that for each t between 0 and 36 in UCSD Ped1 dataset, the regularity score, Sr(t) of the sequence that starts at frame(t) and ends at frame (t+3) is calculated.

When we are getting a reconstruction error value that is greater than the threshold, our system sends the alert signal. Also the system starts to store the frames of original quality and resolution from a predefined number of frames before the occurrence of the abnormality to a predefined number of frames after the occurrence of the abnormality in a video format for future reference. We find it acceptable to store 10 to 20 frames from before the occurrence of the abnormality to 100 to 120 frames after the occurrence of the abnormality, to get a clear idea of what the abnormality is, and how it is occurring.

V. RESULT

These are the test results shown while detecting the abnormal events through Avenue dataset.

It is noticed that the range above 31.5 is considered as an abnormal or unusual event and the variation in graph plot shows the same. Thus altered cases are taken separately and marked as an abnormal event.

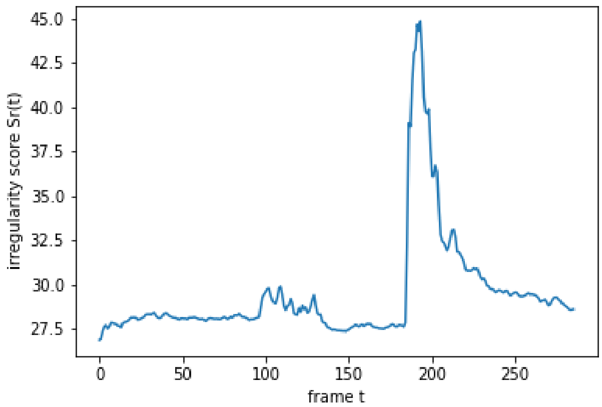

Fig.4 Plot of Second Test Video

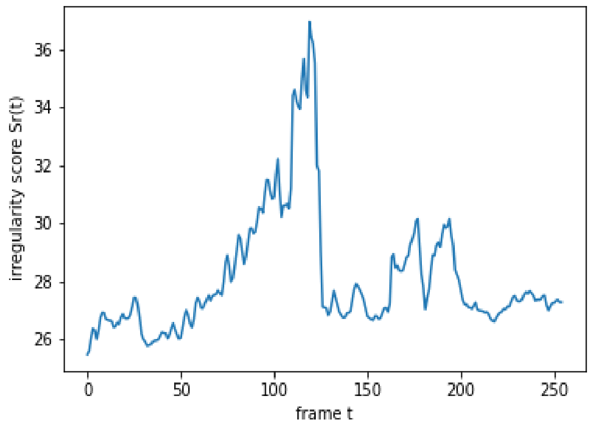

Fig.5 Plot of Second Test Video

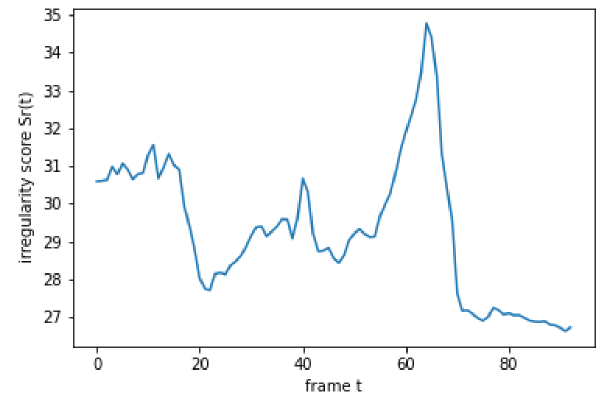

Fig.6 Plot of Third Test Video

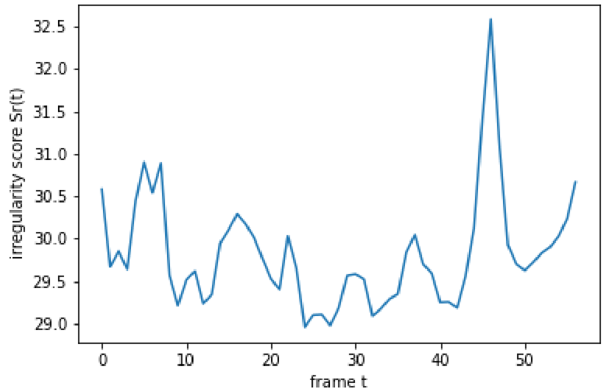

Fig.7 Plot of Fourth Test Video

VI. CONCLUSION

Abnormal event detection is a prominent feature in the creation of a smart CCTV system where it is possible to automatically detect abnormalities and create the necessary alerts. Supervised learning models are commonly used in the existing systems to detect the various anomalies along with reasonable computational resources. However, since the anomalies are of various kinds, it won't be feasible to train the system to detect all types of anomalies. For this reason, supervised learning is replaced by unsupervised learning to effectively train the system. By implementing this system, we also make a system that is storage efficient by saving only the abnormal frames in high quality while the recordings would be saved in lower quality.In the future, both supervised as well as unsupervised learning methods can be combined together to improve the system. Also anomaly identification methods could be added in the future to identify various types of anomalies as well as object detection.

FUTUREWORK

Our project aims at the future scope that has the ability to train a system that can distinguish between specific abnormalities. For example, first the system will check whether any abnormality is present or not after that the abnormal snippet will be given into other training models that each are a classification model of specific abnormality detection.

REFERENCES

[1]. Kang Hao Cheong, Sandra Poeschmann, Joel Weijia Lai, Jin Ming Koh, U. Rajendra Acharya, Simon Ching Man Yu, Kenneth Jian Wei Tang, "Practical Automated Video Analytics for Crowd Monitoring and Counting", Digital Object Identifier, 2019.

[2]. Elvan Duman, Osman Ayhan Erdem, "Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder", Digital Object Identifier, 2019.

[3]. Thittaporn Ganokratanaa, Supavadee Aramvith, Nicu Sebe, "Unsupervised Anomaly Detection and Localization Based on Deep Spatiotemporal Translation Network", Digital Object Identifier, 2020.

[4]. Akshara Alex, Ashi Sahu, Avni Tanwar, Nisha Rathi, Kavitha Namdev,"Abnormal Event Detection By Machine Vision Using Deep Learning", IJEAST, 2020.

[5]. Cem Direkoglu "Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks", Digital Object Identifier, 2020.

[6]. Sandhya Rani Sahoo, Ratnakar Dash, Ramesh Ku. Mahapatra, Baishnabi Sahu, "Unusual Event Detection in Surveillance Video using Transfer Learning", 2019 International Conference on Information Technology.

[7]. Rui Jiang, Xiaozheng Mou, Shunshun Shi, Yueyin Zhou, Qinyi Wang, Meng Dong, Shoushun Chen "Object tracking on event cameras with offline–online learning" CAAI Trans. Intell. Technol., 2020, Vol. 5, Iss. 3, pp. 165–171.

[8]. Thittaporn Ganokratanaa, Supavadee Aramvith, Nicu Sebe, "Anomaly Event Detection Using Generative Adversarial Network for Surveillance Videos", 978- 1-7281-3248-8©2019 APSIPA.

[9]. https://www.geeksforgeeks.org/introduction-convolution-neural-network/