DEEP SURVEILLANCE USING DEEP LEARNING-ANOMALY DETECTION

A MINI PROJECT REPORT

Submitted by

DHINESH.E-20I115

GUNALAN.A-20I117

KAAVIYASREE.S.K-20I125

SREE NILAYA.S-20I150

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

KARPAGAM COLLEGE OF ENGINEERING

Rediscover | Refine | Redefine

Coimbatore - 641 032

Autonomous | Affiliated to Anna University, Chennai

Artificial Intelligence and Data Science

JUNE 2022

KARPAGAM COLLEGE OF ENGINEERING

(AUTONOMOUS)

COIMBATORE – 641 032

DEEP SURVEILLANCE USING DEEP LEARNING ANOMALY DETECTION

BONAFIDE RECORD OF WORK DONE BY

DHINESH.E-20I115

GUNALAN.A-20I117

KAAVIYASREE.S.K-20I125

SREE NILAYA.S-20I150

Report submitted in partial fulfillment of the requirements assessment for the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Faculty guide	Head of the Department
Ms. Sathya. K M.E	**Mr. Manikandan Sundaram M. Tech, (Ph. D)**

Examiner

Mr. Surenther. I M.E (Ph. D)

ACKNOWLEDGEMENT

We express our sincere thanks to Karpagam educational and charitable trust for providing necessary facilities to bring out the project successfully. We felt greatness to record our thanks to the chairman **Dr.R.VASANTHAKUMAR, B.E.,(Hons),D.Sc.**for all his support and ray of strengthening hope extended.

It is the moment of immense pride for us to reveal our profound thanks to our respected principal, **Dr.P.VIJAYAKUMAR, M.E., Ph.D.** who happens to be a striving force in all our endeavors.

We express our sincere thanks to our **Mr. Manikandan Sundaram M. Tech, (Ph. D)** Head of the Department of Artificial Intelligence and Data Science for providing an opportunity to work on this project. His valuable suggestions helped us a lot to do this project.

A word of thanks would not be sufficient for the work of our project guide **Ms. Sathya. K M.E** Department of Artificial Intelligence and Data Science whose efforts and inspiration lead us through every trying circumstance.

We would also like to recollect the courage and enthusiasm that was inculcated in us by our project coordinator, **Mr. Surenther. I M.E (Ph. D)** Department of Artificial Intelligence and Data Science for valuable guidance and support through the tenure of our project.

We deeply express our gratitude to all the members of the faculty of the department of Artificial Intelligence and Data Science for the encouragement, which we received throughout the semester.

TABLE OF CONTENTS

ABSTRACT

Abnormal event detection (AED) in urban surveillance videos has multiple challenges. Unlike other computer vision problems, the AED is not solely dependent on the content of frames. It also depends on the appearance of the objects and their movements in the scene.Various methods have been proposed to address the AED problem. Among those, deep learning based methods show the best results. This paper is based on deep learning methods and provides an effective way to detect and locate abnormal events in videos by handling spatio temporal data. Big data applications are consuming most of the space in industry and research areas. Among the widespread examples of big data, the role of video streams from CCTV cameras is equally important as other sources like social media data, sensor data, agriculture data, medical data and data evolved from space research.Surveillance videos have a major contribution in unstructured big data. CCTV cameras are implemented in all places where security has much importance. Manual surveillance seems tedious and time consuming. Security can be defined in different terms in different contexts like theft identification, violence detection, chances of explosion etc. In crowded public places the term of security covers almost all types of abnormal events. Among them violence detection is difficult to handle since it involves group activity. The anomalous or abnormal activity analysis in a crowd video scene is very difficult due to several real world constraints. The paper includes a deep rooted survey which starts from object recognition, action recognition, crowd analysis and finally violence detection in a crowd environment.Majority of the papers reviewed in this survey are based on deep learning techniques. Various deep learning methods are compared in terms of their algorithms and models. The main focus of this survey is application of deep learning techniques in detecting the exact count, involved persons and the activity in a large crowd at all climate conditions. Paper discusses the underlying deep learning implementation technology involved in various crowd video analysis methods. Real time processing, an important issue which is yet to be explored more in this field is also considered. Not many methods are there in handling all these issues simultaneously.The issues recognized in existing methods are identified and summarized. Also future direction is given to reduce the obstacles identified. The survey provides a bibliographic summary of papers from ScienceDirect, IEEE Xplore and ACM digital library.It is crucial to ensure proper surveillance for the safety and security of people and their assets. The development of an aerial surveillance system might be very effective in catering to the challenges in surveillance systems. Current systems are expensive and complex. A cost-effective and efficient solution is required, which is easily accessible to anyone with a moderate budget. In aerial surveillance, quadcopters are equipped with state-of-the-art image processing technology that captures detailed photographs of every object underneath. A quadcopter-based solution is proposed to monitor desired premises for any unusual activities, like the movement of persons with weapons and face detection to achieve the desired surveillance. After detection of any unusual activity, the proposed system generates an alert for security personals.The proposed solution is based on quadcopter surveillance and video streaming for anomaly detection in the

received video streams through deep learning models. A well-known FasterRCNN algorithm is modified for fast learning with feature reduction in the initial feature extraction step. Five different kinds of CNNs were evaluated for their ability to identify objects of interest in surveillance images. ResNet-50–based FasterRCNN with the highest average precision performed as an excellent solution for threat detection. The average precision of the system is 79% across all categories achieved.

Surveillance systems research area these days has become more approachable duo to the increment of crimes and terrorism events, those systems ideal goal is to be able to identify any person anywhere at anytime, to prevent crimes before they even happen. This paper proposes a gait recognition scheme for person identification based on spatiotemporal pose by using skeletal data that was collected from an RGB-D sensor. Concretely, the human gait over a period of time is described by 3D geometrical features with considering both distances and orientations of joints. The extracted features are then fed to an identification model which is fine-tuned from a pre-trained Convolutional Neural Network (CNN). According to the experiment on the benchmark dataset UPVC Gait, the system achieves the person identification accuracy of 92.88%, which is surprisingly good for using a dataset such as UPVC Gait with a fine-tuned pre-trained CNN.Each object is represented by its bounding box, namely its location. The tracker has a global state that monitors all objects in a single condensed state. At each frame, a set of new detections is produced to track objects. We need to update the tracker's state by associating each track to every detection, if possible, with unassociated detections starting new tracks. We do not use the object class to perform associations for two reasons. First, we want to keep track consistency, even in cases where the detector makes mistakes for a few frames. We can recover from this error by exploiting the estimated distribution of classes over the entire track. Second, there are few cases in which the detection of a different class spawns on top of another, and often, the geometry of the new detection is not consistent with the other oneThe second functionality is the ability to alert the operator of an anomalous presence in the video stream. We consider a fully supervised approach to tackle anomaly detection and localization as a one-class classification problem

LIST OF FIGURES

Table 1.1 List of figures

LIST OF ABBREVIATIONS

ACM	Association for Computing Machinery
AED	Automated external defibrillators
CAE	computer-aided engineering
CCTV	closed-circuit television
CNN	Cable News Network
COTS	commercial-off-the-shelf
DNN	deep neural network
FPGA	field programmable gate array
GPS	Global Positioning System
IDP	integrated data processing
IDS	Intrusion detection system
IEEE	Institute of Electrical and Electronics Engineers
LSTM	Long short-term memory
RNN	recurrent neural network
SVAS	Surveillance Video Analysis System

Table 1.2 List of Abbreviations

LIST OF TABLES

CHAPTER-1
1.1 DEEP SURVEILLANCE

1.1.1 Introduction

Surveillance security is a very tedious and time-consuming job. In this tutorial, we will build a system to automate the task of analyzing video surveillance. We will analyze the video feed in real-time and identify any abnormal activities like violence or theft.There is a lot of research going on in the industry about video surveillance among them; the role of CCTV videos has overgrown. CCTV cameras are placed all over the places for surveillance and security.In the last decade, there have been advancements in deep learning algorithms for deep surveillance. These advancements have shown an essential trend in deep surveillance and promise a drastic efficiency gain. The typical applications of deep surveillance are theft identification, violence detection, and detection of the chances of explosion.

1.1.2 Surveillance System

The proposed system has 2 main functionalities that we treat separately: (i) the detection of objects of interest, including their tracking and counting, both spatially and temporally, The detection of anomalous areas that contain unexpected objects or suspicious elements. A brief system scheme which highlights its main modules

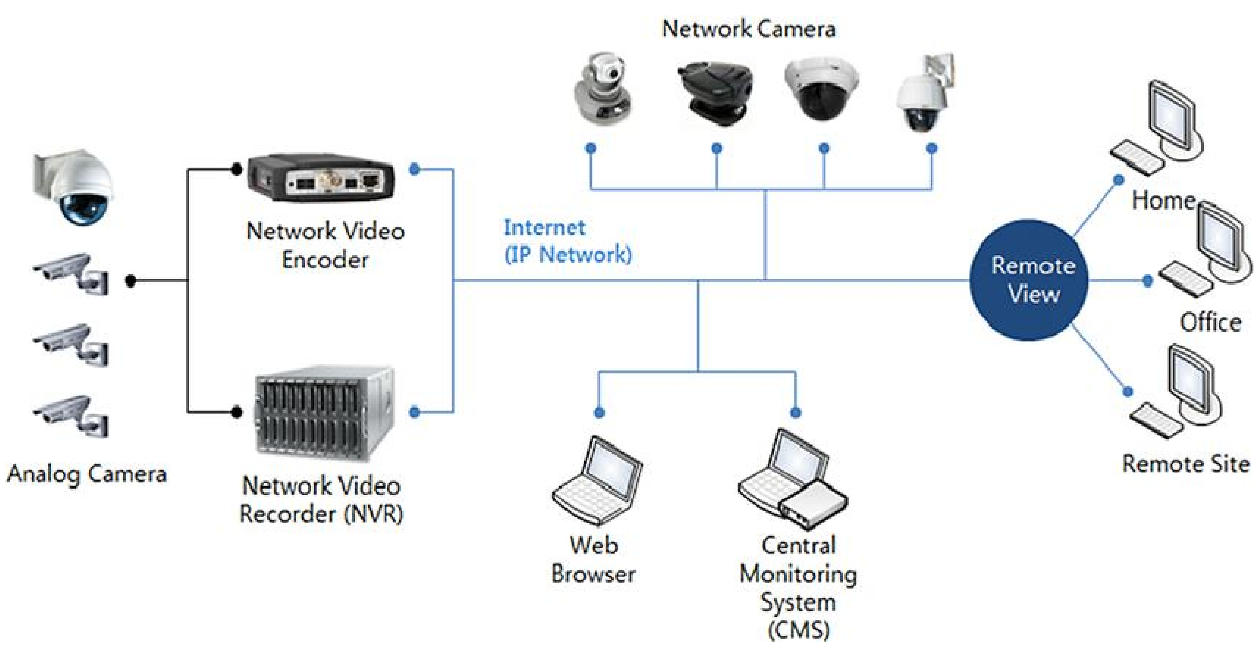

Fig.1.1 surveillance diagram

1.2 Network architecture

We have generally seen deep neural networks for computer vision, image classification, and object detection tasks. In this project, we have to extend deep neural networks to 3-dimensional for learning spatio-temporal features of the video feed.

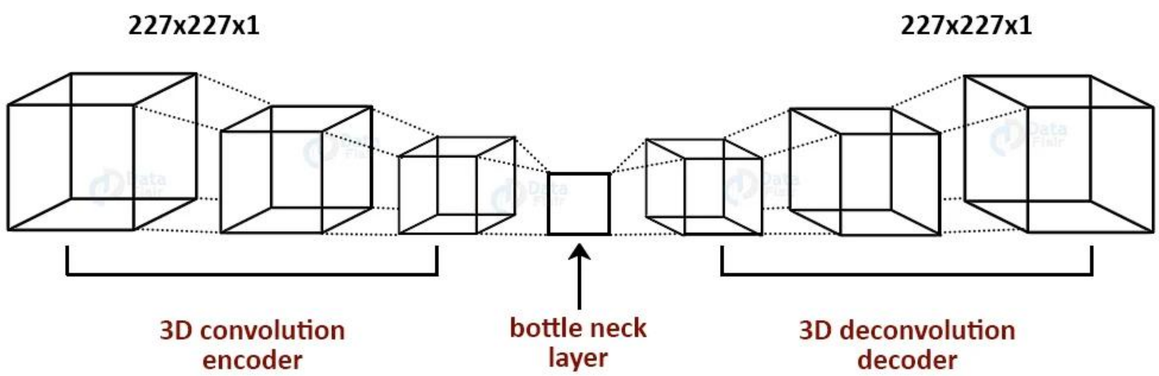

Fig 1.2 Architecture diagram

For this video surveillance project, we will introduce a spatio temporal autoencoder, which is based on a 3D convolution network. The encoder part extracts the spatial and temporal information, and then the decoder reconstructs the frames. The abnormal events are identified by computing the reconstruction loss using Euclidean distance between original and reconstructed batch. The paper includes a deep rooted survey which starts from object recognition, action recognition, crowd analysis and finally violence detection in a crowd environment. Majority of the papers reviewed in this survey are based on deep learning techniques. Various deep learning methods are compared in terms of their algorithms and models. The main focus of this survey is application of deep learning techniques in detecting the exact count, involved persons and the activity in a large crowd in all climate conditions.

Deep learning implementation technology involved in various crowd video analysis methods. Real time processing, an important issue which is yet to be explored more in this field is also considered. Not many methods are there in handling all these issues simultaneously. The issues recognized in existing methods are identified and summarized. Also future direction is given to reduce the obstacles identified. The survey provides a bibliographic summary of papers from ScienceDirect, IEEE Xplore and ACM digital library.

1.3 Application areas identified

The contexts identified are listed as application areas. Major part of existing work provides solutions specifically based on the context.

1.Inside office buildings

2.In supermarkets

3.Residential areas

4.Crowd pulling meetings

5.Traffic signals and main junctions

6.Festivals as part of religious institutions

7.Schools and colleges

Among the listed contexts, crowd analysis is the most difficult part. All types of actions, behavior and movement are needed to be identified.For example an abnormal activity that varies from the daily happening that can be detected here , if something gets fire at an instance it can be detected .

1.4Surveillance video data as Big Data

Big video data has evolved in the form of an increasing number of public cameras situated towards public places. A huge amount of networked public cameras are positioned around worldwide. A heavy data stream is generated from public surveillance cameras that are creatively exploitable for capturing behaviors. Considering the huge amount of data that can be documented over time, a vital scenario is facility for data warehousing and data analysis. Only one high definition video camera can produce around 10 GB of data per day. The space needed for storing large amounts of surveillance videos for a long time is difficult to allot. Instead of having data, it will be useful to have the analysis result. That will result in reduced storage space. Deep learning techniques are involved with two main components; training and learning. Both can be achieved with highest accuracy through huge amounts of data.

NOTE : In military bases ,many countries uses deep surveillance technique to find terrorists and it is known as Military survillance systems.

Main advantages of training with huge amounts of data are listed below. It's possible to adapt variety in data representation and also it can be divided into training and testing equally. Various data sets available for analysis are listed below. The dataset not only includes video sequences but also frames. The analysis part mainly includes analysis of frames which were extracted from videos. So a dataset including images is also useful.

Potential topics include but are not limited to the following:

- Emotion and gesture recognition
- Object tracking and segmentation
- Scene understanding and human behavior analysis
- Scene understanding and human behavior analysis
- Person re-identification
- Activity detection and recognition
- Human computer/robot interaction
- Crowd dynamics and crowd analysis
- Wildlife entity detection and tracking

1.5 Methods identified/reviewed other than deep learning

Methods identified are mainly classified into two categories which are either based on deep learning or not based on deep learning. This section is reviewing methods other than deep learning.SVAS deals with automatic recognition and deduction of complex events. The event detection procedure consists of mainly two levels, low level and high level. As a result of low level analysis people and objects are detected. The results obtained from low level are used for high level analysis that is event detection. The architecture proposed in the model includes five main modules.

The five sections are

- Event model learning
- Action model learning
- Action detection
- Complex event model learning

Interval-based spatio-temporal model (IBSTM) is the proposed model and is a hybrid event model. Other than this methods like Threshold models, Bayesian Networks, Bag of actions and Highly cohesive intervals and Markov logic networks are used.SVAS method can be improved to deal with moving cameras and multi camera data sets. Further enhancements are needed in dealing with complex events specifically in areas like calibration and noise elimination.Multiple anomalous activity detection in video is a rule based system. The features are identified as motion patterns. Detection of anomalous events are done either by training the system or by following dominant set property.The concept of a dominant set where events are detected as normal based on dominant behavior and anomalous events are decided based on less dominant behavior. The advantage of a rule based system is that it is easy to recognize new events by modifying some rules. The main steps involved in a recognition system are

- Pre processing
- Feature extraction
- Object tracking
- Behavior understanding

As a preprocessing system video segmentation is used. Background modeling is implemented through Gaussian Mixture Model (GMM). For object recognition external rules are required. The system was implemented in Matlab 2014. The areas where more concentration is needed are doubtful activities and situations where multiple object overlapping happens.

Mining anomalous events against frequent sequences in surveillance videos from commercial environments focus on abnormal events linked with frequent chain of events. The main result in identifying such events is early deployment of resources in particular areas. The implementation part is done using Matlab, Inputs are already noticed events and identified frequent series of events. The main investigation under this method is to recognize events which are implausible to chase a given sequential pattern by fulfilling the user identified parameters.The method is giving more focus on event level analysis and it will be interesting if you pay attention at entity level and action level. But at the same time going to such a granular level makes the process costly. Many trajectory based methods have been used in abundant installations. But those methods have to face problems related to occlusions. As a solution to that, feature descriptors using optical flow based methods.

NOTE : In this project we use Simple neural networks to (Basic) Deep learning, neural networks

1.6 Data Acquisition

The data acquisition phase is responsible for collecting and acquiring videos from surveillance drones to construct the required data set. The videos in the data set are obtained from different environments through a surveillance drone camera. The data set consists of videos of different persons with different poses, with and without weapons, etc. The persons include guards and common people. Later the captured videos are transformed into image frames, which are then taken from the data set for further processing. A good data set plays an important core role in the development of the proposed surveillance system. A machine learning model for the detection will be trained by giving them image frames from the developed data set. In the next phase, necessary preprocessing steps are applied to the image frames. The detail of the preprocessing step is as follows.

MOTIVATION

Now-a-days the huge population is growing and surveillance is an integral part of security and patrol. For the most part, the job entails extended periods of looking out for something undesirable to happen. It is crucial that we do this, but also it is a very mundane task.Wouldn't life be much simpler if there was something that could do the "watching and waiting" for us? Well, you're in luck. With the advancements in technology over the past few years, we could write some scripts to automate the above tasks — and that too, rather easily. But, before we dive deeper, let us ask ourselves.

CHAPTER-2

LITERATURE SURVEY

Currently, many robotic applications are being developed for doing tasks autonomously without any commands from a human. A system enabling a robot to surveillance, like detecting and tracking an object in motion, will carry us to a more advanced task. For taking the role of UAV (Unmanned Aerial Vehicle) AR , the drone comes in place as a flying robotic platform. As far as the implementation part is concerned, they developed an algorithm to detect and track an object by analyzing its shape and color. Unmanned Aerial Aircraft Systems (UAS) are being divided into different transportation engineering areas. Traffic analysis and work on the grounds are highly compatible with the systems. Such studies are helpful for understanding and working UAS in transportation.

2.1 WIRELESS SENSOR NETWORK :

In the study by Berrahal et al. , the authors proposed a cooperative border surveillance solution that relies on WSN (Wireless Sensor Network), implemented to detect and track transfer and UAVs in the form of quadcopters. A heuristic-based scheduling algorithm is used by increasing the rate of trespassers spotted by the quadcopters.

Also, techniques to localize sensors using RFID technology and optimal computing positions and relay data between isolated islands of nodes were used for evaluating the performance of the mentioned WSN-based surveillance system. In another work , the authors started their studies by defining the capabilities of aerial surveillance systems and suggested several technologies that can be used in it. They also discussed how drones could be used in aerial surveillance. The work also found that surveillance can be utilized in peace retaining activities and time monitoring of a place at any time of the day. The intention was to provide speedy and green surveillance at a low-cost charge so that it could be used extensively at the nonpublic, institutional, and governmental stage. Gadda and Patil stated that security issues of borders are increasing with rapid tensions across the countries. Terrorist attacks at the border are risking the lives of common people living near the borders. Monitoring of terrorism activities and security lapses becomes quite challenging and challenging in changing weather conditions. Designing an UAV for monitoring the border area from a long distance conveyed to the observer. In this study , the IR remote is used for controlling the quadcopter.

The Audio-Visual will be transferred to the computer through a wireless camera. The work presented in this article focused on trends of image processing for UAV (Unmanned Aerial Vehicles) and quadcopter data. Examples for totalling the total number of vehicles, edge detection, and evasion algorithms were discussed. SIFT (scale-invariant features transform) matching and the median filter are mainly used in UAVs image processing and visualization of new urban buildings and mechanisms that control the quadcopter. In this work, a simple camera-based navigation system for an autonomous quadcopter was proposed. Authors observed that additional infrastructure like artificial landmarks and GPS is not required in navigation methods and algorithms. Factors affecting the computational complexity were independent of the environment size and factors like sensing only one landmark at a time.FPGA-based embedded realization of the methods was discussed with the most computationally demanding phase. In the study by Haque et al. They proposed quadcopter as a low-weight and low-cost autonomous flight Unmanned Aerial Vehicle (UAV) for delivering parcels ordered online. Usage of Google maps to navigate to destinations are the significant functions of the quadcopter. The authors also discussed the capability of delivering orders through quadcopter using an online process and coming back to the original point or starting place.The work by Ding et al. presented that those earlier systems were expensive and limited to outdoor environments with access to Global Positioning Systems (GPS). This work aimed to develop a design for a quadcopter to create a map of an unknown space using commercial off-the-shelf (COTS) to reduce cost and for mapping SLAM (Simultaneous Localization and Mapping) for creating an automatic map.

Lastly, it was suggested that the SLAM algorithm might run efficiently on the Android phone. This work by Ding et al stated that applications in communications, photography, agriculture, and mainly surveillance make drones as mini-UAVs a bit more attractive. State-of-the-art studies for that amateur drone with a vision named Dragnet, tailoring the recent Cognitive Internet of Things framework.At the same time, the detection and classification of authorized and unauthorized drones are done, which eventually allow only authorized ones to fly over.Amato et al.discussed that Public security and safety are ensured through wireless communication technologies and support mobility. In the study by Alford et al. The authors stated that the primary cause of security and critical information could be transmitted among the entities involved, dealing with cyber issues and mainly risking privacy of UAV equipped with communication hardware and directed to specific positions for privacy and safety. For the past few decades, Iraq and Syria among the most affected countries due to terrorism by ISIS were liberated, and internationally displaced persons (IDP) were going back to their homes. Infrastructure loss in the form of destroyed buildings and operators are exposed to undefined situations.Situations like this make innovations in the search and clearance of operational technology for ensuring safety and peace. In the study by Bangare et al. focussed on ways to save homes from fire and smoke. The primary concern was home monitoring, appliances controlling, and door latches control from faraway regions.With the use of their device, people will be able to manage our home appropriately from remote places.

In the study of Munagekar researchers focused on detecting places where criminal talk happens and tracing the robber quickly and efficiently. The purpose of this work was to minimize the efforts and finances of the security forces. Canny edge detection algorithm was used to develop this system, which was further used to detect and catch the robbers and stop unusual happenings. Usage of a canny edge filter reduces the postprocessing size of objects, which saves much memory of the hard disk. The focus of Alshammari and Rawat work was on a multi camera video system for the betterment of security . The multi camera video system can make a widespread impact within the security industry. The technical aspect of this may help many people to counter the protection challenges of their everyday lives.The proposed technique may locate and understand human behavior from videos taken from cameras mounted at the wall. The proposed method consists of the detection and monitoring of any targets. In this article a live surveillance system was proposed for weapon and abnormal events detection. The system was composed of three modules.

2.2 CONVOLUTIONAL NEURAL NETWORK BY MAHALANOBIS :

The detection was done using the convolutional neural networks in the first processing module, whereas the second module manages guns and tracking. The alarm operations were done in the last processing module. In their work, shape detection and object detection techniques were experimented to find detection accuracy. The authors claimed that the proposed method substantially reduces the crime rate and the time required to trap the criminal. In this article, Mahalanobis et al. presented autodetection and monitoring strategies based on three significant aspects. Three different aspects of this work were the automatic goal detection and reputation strategies with monitoring, the handover, and smooth tracking of objects within a community and the improvement of actual-time communication and messaging protocols using COTS networking components.The comprehensive review of the various security and surveillance techniques shows that this area has the potential to work. The mechanism of surveillance can be enhanced using modern deep learning approaches. This work proposed a technique for the detection of arms and other objects using advanced deep learning models.The primary neural network is also known as convolutional network. The reason for this arrangement is to extract high-level facial appearance of the metaphors and tumbling the difficulty of the contribution .

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.This replication of video images is used to transfer learning. Recent detection of objects replicapable to recognize a number of factors 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 a completely trained model for categories like ImageNet and keeps hold on fresh classes . 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.

While many types of research have been carried out on video processing , very few studies have been systematically analyzed that focus on video processing using deep learning techniques. Instead, most of them perform surveys by targeting only specific functionality.

However, a study conducted by Nayak et al. shows the advancement in video anomaly detection using deep learning techniques. The authors present the various deep learning techniques for video processing to detect the anomalies such as abnormal activities- fighting, riots, traffic rule violations, stampede, and strange entities - weapons, abandoned luggage, etc. In another survey , researchers reported video processing for abnormal human activity recognition by leveraging the deep learning method for video processing. Borja-Borja et al. surveyed state of-the-art deep learning methods for video processing to list the group and crowd activities.

2.3 MULTIPLE OBJECT TRACKING

The main techniques of deep learning are grouped into Convolutional Neural Network (CNN), Autoencoders (AEs), and Recurrent Neural Network (RNN). Another survey-based on anomaly detection from video data by focuses on the deep learning approach where the author listed generative adversarial networks (GANs) along with other deep learning. A significant application of video processing in computer vision research is pedestrian detection. Brunetti et al.present a review on deep learning video processing methods for pedestrian detection focusing on methods CNN, Deep Neural Network(DNN), Restricted Boltzmann Machine (RBM), and Gaussian Mixture Model. In a survey by Ciaparrone et al. , authors report deep learning methods for Multiple Object Tracking (MOT) from video data. They explored the Faster R-CNN, Mask R-CNN, SSD methods of deep learning for multi-object tracking.

2.4 YOLOV2

Apart from that, the authors also listed out the YOLO series of detectors – YOLOV2. Yan et al reported a review on deep multi view learning from videos focusing on representational deep learning methods such as conventional neural networks, deep brief networks, and multi-view auto-encoders. Taskiran et al. present a taxonomy for face recognition as an image-based and video-based method. For video-based face recognition, various recent deep learning methods were discussed by grouping asset-based methods and sequence based methods. Authors of present a review of the video scene parsing application of video processing using deep learning techniques. They highlight the 2D CNN, 3D CNN, Clockwork FCN, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Spatio-temporal transformer GRU (STGRU), and GAN methods of deep learning.Wang et al. surveyed salient object detection from video data using deep-learning-based methods. They mentioned the specially designed methods for salient object detection such as fully convolutional network, Spatio-temporal cascade neural network, attentive feedback network, etc. Tong et al. investigate deep learning algorithms for video processing, specifically small object detection. The authors show the gradual improvement in CNN-based methods such as R-CNN, Fast R-CNN, Deconv-R-CNN, Improved Faster R-CNN, etc., along with other deep learning approaches. Sánchez et al. demonstrate a study on crowd behavior analysis or crowd anomaly detection by video processing using deep learning techniques. Authors listed convolutional RBM, Fast R-CNN, 3D CNN, PCANet, deep Gaussian Mixture Model, Convolutional AutoEncoder (CAE) with LSTM (CAE - LSTM), Spatio-temporal CNN, and GAN based approach. The related work suggests different approaches for detecting human behaviors from video.

2.5 ADVANCED MOTION DETECTION

Motion detection is the first essential process in the extraction of information regarding moving objects and makes use of stabilization in functional areas, such as tracking, classifica- tion, recognition, and so on.The objective of the works was to detect any abnormal or suspicious events in a video surveillance.Advance Motion Detection (AMD) algorithm was used to detect an unauthorized entry in a restricted area . In the first phase, the object was detected using background subtraction and from frame sequences the object is extracted. The second phase was detection of suspicious activity. Advantage of the system was the algorithm works on real time video processing and its computational complexity was low. But the system was limited in terms of storage service and it can also be implemented with hi-tech mode of capturing of videos in the surveillance areas. A semantic based approach was proposed in .The captured video data was processed and the foreground objects were identified using background subtraction. After subtraction, the objects are classified into living or non-living using a Haar like algorithm. Objects tracking was done using a Real-Time blob matching algorithm. Fire detection was also detected in this paper. Based on the motion features between the object, suspicious activities were detected in . Semantic approach was used to define suspicious events. The object detection and correlation technique was used to track objects . The events are classified based on motion features and temporal information. The computational complexity of the given framework was less. Abnormal events from a university were detected by dividing into zones and estimated the optical flow in each zone using state detection and anomalous detection. The trained model decides whether the student was in a suspicious state or not and the Gaussian distribution decides whether the student performs any anomalous behavior. The accuracy achieved was 97%. Intelligent video surveillance for crowd analysis was discussed in. This was a review paper which covers relevance of video surveillance analysis in today’s world, various deep learning models, algorithms and datasets used for video surveillance analysis. The majority of papers mentioned above were done with the help of computer vision using various algorithms or by neural networks for detecting human behavior analysis from videos. Computer vision methods require a lot of pre-processing to extract the trajectories or motion patterns to understand the evolution of features in a video sequence . Also, background subtraction is based on a static background hypothesis which is often not applicable in real time scenarios. In the real world, most of the issues occur in the crowd. Above discussed methods lack efficiency while handling crowds.

Based on the literature survey a deep architecture can be modeled for suspicious activity prediction using 2D CNN and LSTM, so the accuracy of the system can be improved. In the deep learning approach, most of the papers detect only suspicious activity. So an efficient mechanism is needed to alert the security in the case of any suspicious behavior.

CHAPTER-3
METHODOLOGY

3.1 XGBOOST:

XGBoost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting algorithm to a known dataset and then classifies the data accordingly. The two reasons to use XGBoost are also the two goals of a machine learning project: execution speed and model performance. XGBoost is really fast when compared to other implementations of gradient boosting. From Szilard Pafka's experiment , he proved that XGBoost is fast, memory efficient, and of high accuracy. So, we use XGBoost as a part of our anomaly detection architecture to obtain perfect performance. The dataset is split into train and test sets. Three network models,namely MLP, BPN and LSTM, obtain the training data as input. Each model performs 10-fold cross-validation with various hyperparameters. These parameters include the number of epochs, learning rates, activation function and optimisers. The validation sets which yield the best cross-validation accuracy and low MSE are detailed in the experimental section. Each model output is sent to the XGBoost in the model evaluation stage. In the model prediction stage, XGBoost is fed with new data samples to determine the class labels. Different test sets are given as input to the XGBoost, and the average metrics of all test sets are considered the final result.

3.2 MODEL:

The model that has been created using the XGBoost is being more optimized this Optimized model from keras using optimizer makes the major difference in predicting and determining the accuracy of the model, the videos that recorded are analyzed using the determined frame rate ,here the model is splitted into two i.e, Violence and Non-violence ,the violence part of the model stores the abnormal activities like the following

- Fire from nowhere
- Suspicious movements like falling,running,being stabbed etc..,
- Fall of ceiling
- Road accidents
- Military attack in border

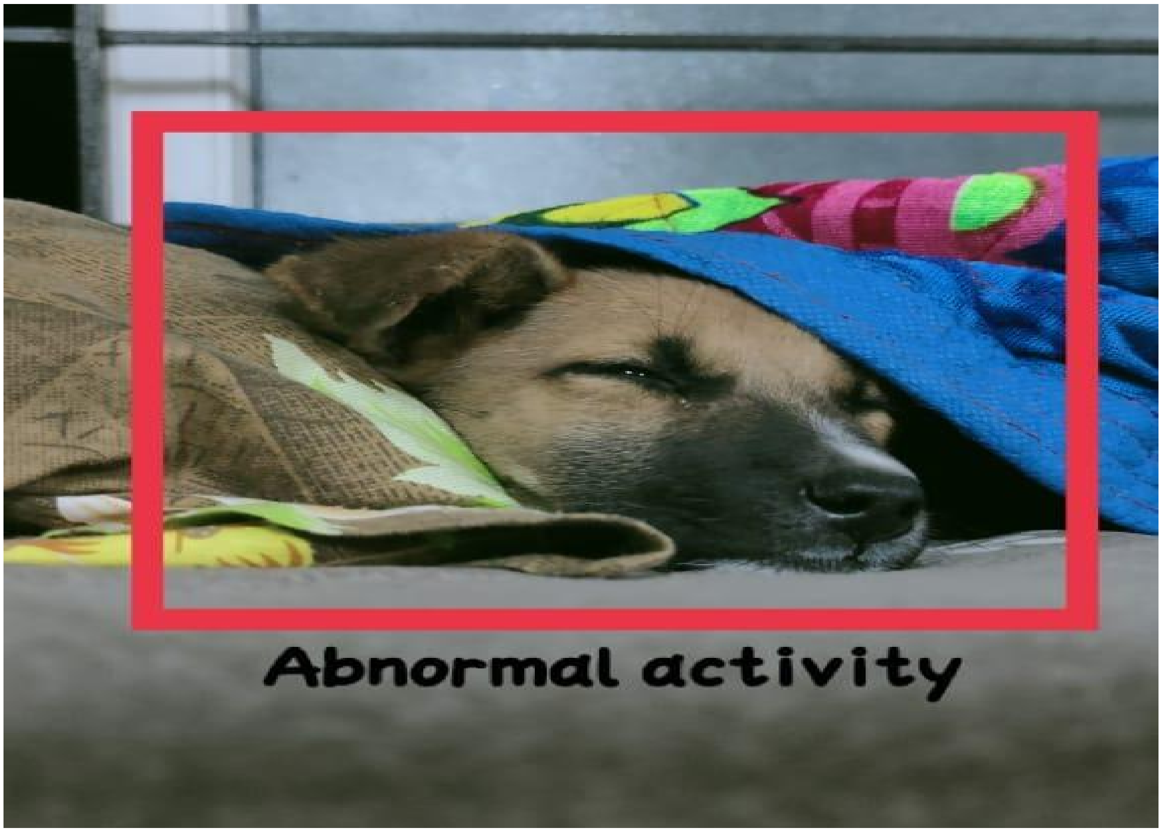

Fig 3.1 Abnormal activity

It can also detect unusual objects like knives,guns,ammo's etc., when the model is given proper training.

Here in the above image Fig 3.1 Abnormal activity of a dog sleeping in bed head resting in pillow covered with blanket is an abnormal event , so the model we built here recognizes it an abnormal event .

Similarly the non-violence model detects and and stores the normal daily activities as follows

- Walking
- Driving in proper manner
- Dusk and dawn
- Eating

OBJECTIVE OF THE PROJECT:

The main objective of anomaly detection analysis is to identify the observations that do not adhere to general patterns considered as normal behavior. Anomaly Detection could be useful in understanding data problems. There are domains where anomaly detection methods are quite effective. There are two directions in data analysis that search for anomalies: outlier detection and novelty detection. So, the outlier is the observation that differs from other data points in the train dataset. The novelty data point also differs from other observations in the dataset, but unlike outliers, novelty points appear in the test dataset and are usually absent in the train dataset.

The most common reason for the outliers are;

- data errors (measurement inaccuracies, rounding, incorrect writing, etc.);
- noise data points;
- hidden patterns in the dataset (fraud or attack requests).

So outlier processing depends on the nature of the data and the domain. Noise data points should be filtered (noise removal); data errors should be corrected. Some applications focus on anomaly selection, and we consider some applications further.

FRAMEWORK USED:

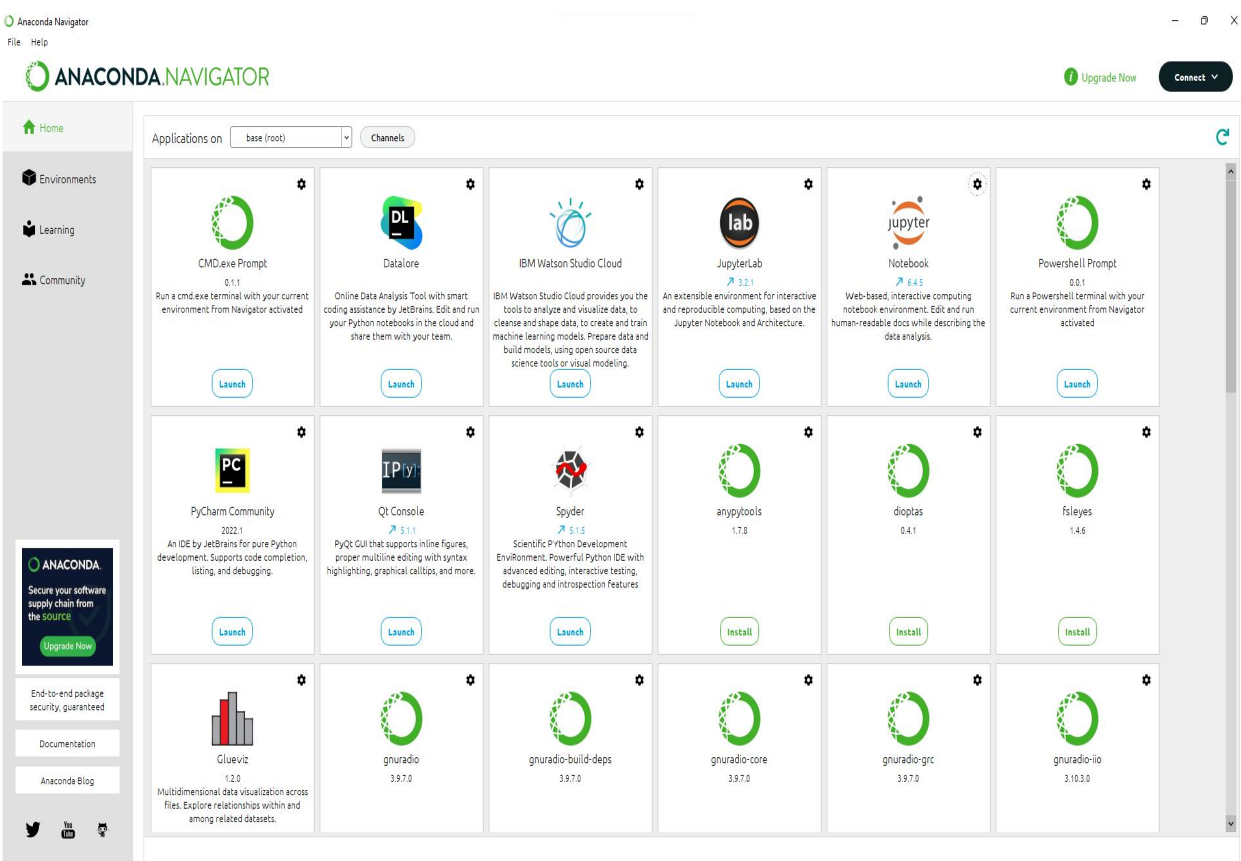

Fig 3.2 Anaconda Navigator

Anaconda framework is used here for this deep surveillance project it contains many frameworks like pycharm , R studio , spyder , Jupyter for different languages like Python, R etc.., Here only jupyter notebook is used for our anomaly detection project

CHAPTER-4
WORKING

In the network anomaly detection/network intrusion and abuse detection context, interesting events are often not rare—just unusual. For example, unexpected jumps in activity are typically notable, although such a spurt in activity may fall outside many traditional statistical anomaly detection techniques.

Many outlier detection methods, especially unsupervised techniques, do not detect this kind of sudden jump in activity as an outlier or rare object. However, these types of micro clusters can often be identified more readily by a cluster analysis algorithm.

There are three main classes of anomaly detection techniques: unsupervised, semi-supervised, and supervised. Essentially, the correct anomaly detection method depends on the available labels in the dataset.

Supervised anomaly detection techniques demand a data set with a complete set of “normal” and “abnormal” labels for a classification algorithm to work with. In this supervised learning there must be a training set for both data objects and expected anomalous objects. We have to observe that there can be more than one anomalous class.This kind of technique also involves training the classifier. This is similar to traditional pattern recognition, except that with outlier detection there is a naturally strong imbalance between the classes. Not all statistical classification algorithms are well-suited for the inherently unbalanced nature of anomaly detection.

Semi-supervised anomaly detection techniques use a normal, labeled training data set to construct a model representing normal behavior.Sometimes when there is training data with labeled normal objects and score given, but has no anomalous objects, then we can implement the semi supervised anomaly detection to find the anomalies. We use the normal objects to find the anomalies. But, the difficulty is sometimes it is not easy to find that representative set of normal objects using which we have to find anomalies. They then use that model to detect anomalies by testing how likely the model is to generate any one instance encountered.

Unsupervised methods of anomaly detection detect anomalies in an unlabeled test set of data based solely on the intrinsic properties of that data. For situations where class labels are not available. We can give a score for each object that shows the degree to which the instance is anomalous. We also can observe that if there are many anomalies present which are similar to each other, then we can group them as a normal group or the outlier score is low. So, we can say that for unsupervised anomaly detection to be successful it is necessary that anomalies are distinct. The working assumption is that, as in most cases, the large majority of the instances in the data set will be normal. The anomaly detection algorithm will then detect instances that appear to fit with the rest of the data set least congruently.

This model that we have created is more optimized than any other model that is existing till now once the target variable is set the model works to get the targeted trainee model so that after determining the frame rate it works more efficiently to predict the abnormal activity which is our main aim.

Video surveillance systems are a system of one or more video cameras on a network that send the captured video and audio information to a certain place. The images are not available to the public like television. They are live monitored or transmitted to a central location for recording and storage.

4.1 VISUALIZATION OF OUTPUT

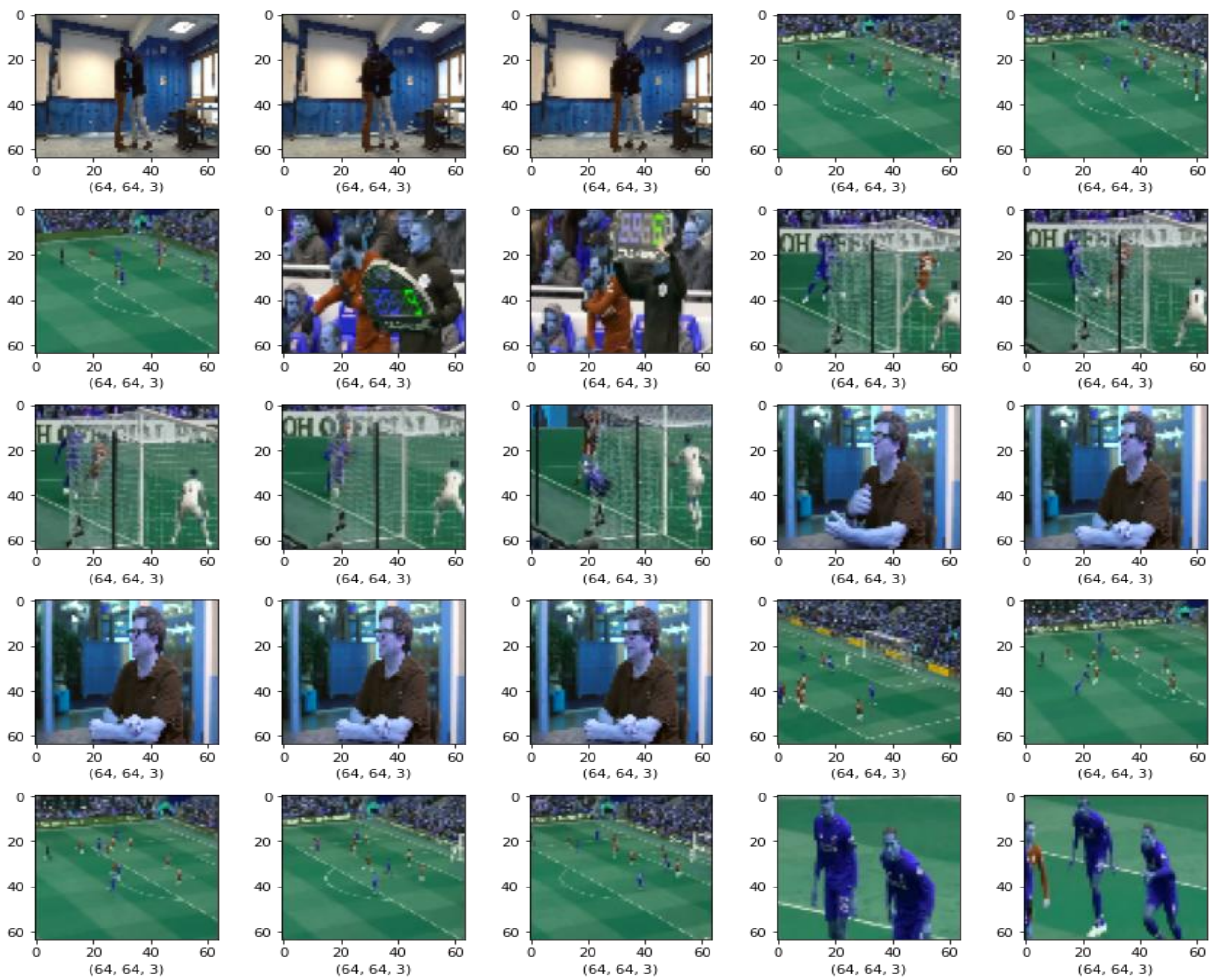

Fig 4.1 Non violence model output

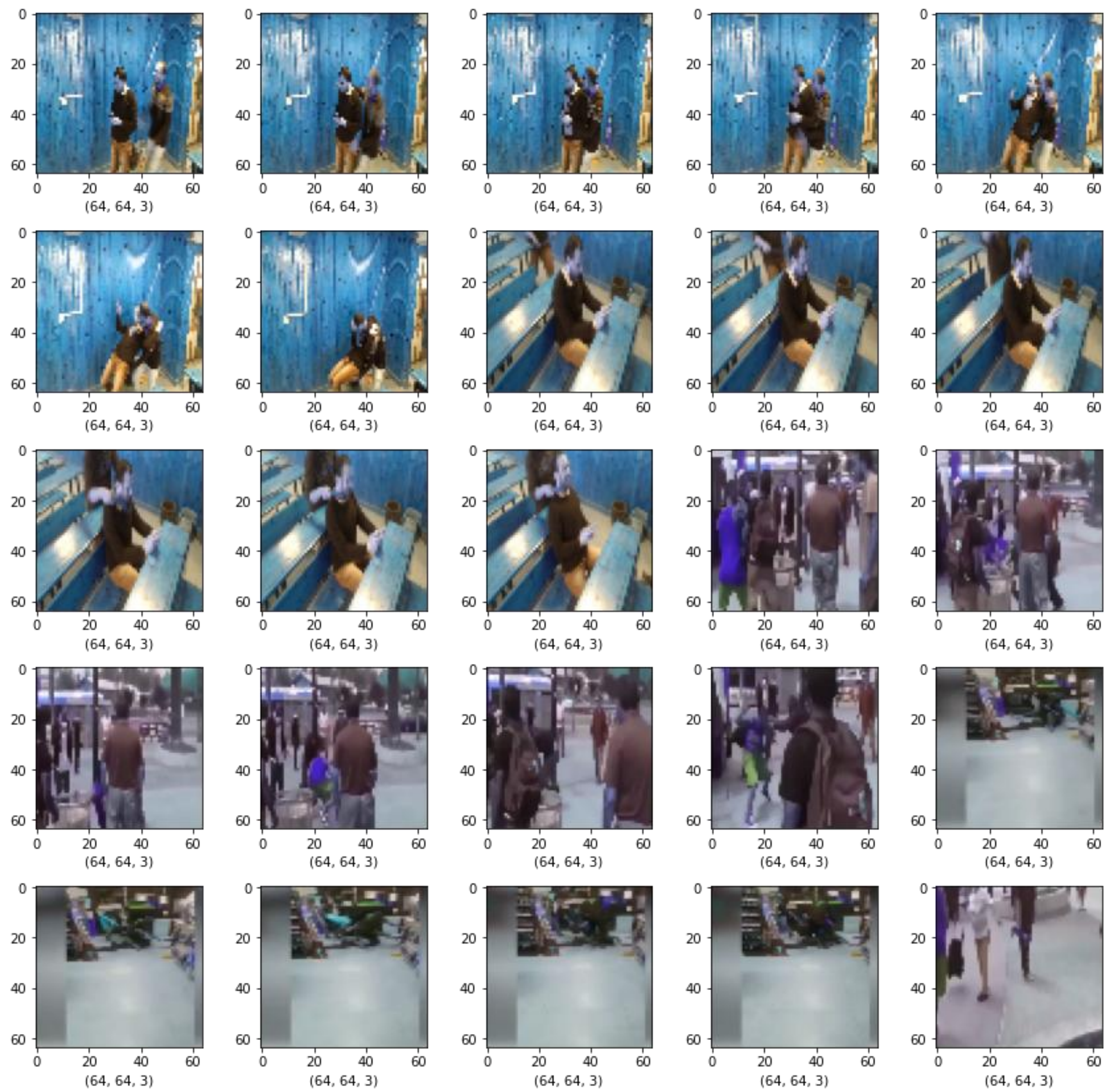

Fig 4.2 Violence model output

CHAPTER-5
FUTURE SCOPE

The model can be extended in future by handling local abnormal event detection through proposed features which are related with objectness method.Multiple Hierarchical Dirichlet processes for anomaly detection in Traffic is mainly for understanding the situation in real world traffic. The anomalies are mainly due to global patterns instead of local patterns. That includes the entire frame. Concept of super pixels is included. Super pixels are grouped into regions of interest.Optical flow based method is used for calculating motion in each super pixel. Points of interest are then taken out in active super pixels. Those points are then tracked by the Kanade–Lucas–Tomasi (KLT) tracker. The method is better to handle videos involving complex patterns with less cost. But not to mention videos taken in the rainy season and bad weather conditions.Intelligent video surveillance beyond robust background modeling handles complex environments with sudden illumination changes. Also the method will reduce false alerts. Mainly two components are there. IDS and PSD are the two components.First stage intruder detection system will detect an object. Classifier will verify the result and identify scenes causing problems. Then in the second stage problematic scene descriptors will handle positives generated from IDS. Global features are used to avoid false positives from IDS.Towards abnormal trajectory and event detection in video surveillance works like an integrated pipeline. Existing methods either use trajectory based approaches or pixel based approaches. But this proposal incorporates both methods.

Proposal include components like

- Object and group tracking
- Grid based analysis
- Trajectory filtering
- Abnormal behavior detection using actions descriptors

The method can identify abnormal behavior in both individuals and groups. The method can be enhanced by adapting it to work in a real time environment.

RIMOC: a feature to discriminate against unstructured motions: application to violence detection for video surveillance . There is no unique definition for violent behaviors. Those kinds of behaviors show large variances in body poses. The method works by taking the eigenvalues of histograms of optical flow.

CHAPTER-6
CONCLUSION

It is critical for network admins to be able to identify and react to changing operational conditions. Any nuances in the operational conditions of data centers or cloud applications can signal unacceptable levels of business risk. On the other hand, some divergences may point to positive growth.

Therefore, anomaly detection is central to extracting essential business insights and maintaining core operations. Consider these patterns—all of which demand the ability to discern between normal and abnormal behavior precisely and correctly:

- An online retail business must predict which discounts, events, or new products may trigger boosts in sales which will increase demand on their web servers.
- An IT security team must prevent hacking and needs to detect abnormal login patterns and user behaviors.
- A cloud provider has to allot traffic and services and has to assess changes to infrastructure in light of existing patterns in traffic and past resource failures.

An evidence-based, well-constructed behavioral model can not only represent data behavior, but also help users identify outliers and engage in meaningful predictive analysis. Static alerts and thresholds are not enough, because of the overwhelming scale of the operational parameters, and because it's too easy to miss anomalies in false positives or negatives.

To address these kinds of operational constraints, newer systems use smart algorithms for identifying outliers in seasonal time series data and accurately forecasting periodic data patterns.

In this paper, we described a third generation surveillance system based on deep learning that can be applied to complex surveillance scenarios. Our system is able to detect a wide range of events of interest such as vehicles moving in the wrong direction, an excessive amount of people entering a certain area, and so on. Regarding all abnormalities that can not be defined with location and direction based patterns, we rely on a statistical based anomaly detector. To the best of our knowledge, such a comprehensive integrated solution for wide area surveillance has not been presented previously. Existing methods usually either focus on statistical anomaly detection or on parking occupancy assessment. Our proposed system is able to keep track of all moving objects in a parking lot to detect mispositioned vehicles, wrong direction accesses, and spot occupancy with high reliability even in complex scenarios. Regarding generic abnormal behaviors, we presented a method reporting state-of-the-art results in the challenging anomaly localization task. The deployed systems are connected with remote interfaces that need to be updated in real-time; therefore, all algorithms were developed to abide this constraint. A few limitations may be highlighted. First, is the greedy nature of our tracker data association strategy. While, in practice, we have seen this to work very well, some situations with long-term target occlusions leading to ID switching may be solved. A main drawback of non-greedy trackers is the computational complexity which,in turn, may require extensive resources to be used in real-time scenarios. Second, at the moment, our method supports multiple cameras, but decisions are taken separately on each one. On one hand, this setting allows us to install new cameras without complex multi-camera calibration; on the other hand, a multi-camera tracker may improve system accuracy in cases where occlusions are present. **Author Contributions:** Conceptualization, L.S., T.U. and A.D.B.; Software, F.T., L.S. and T.U.; Writing—Original Draft, F.T., L.S., T.U. and A.D.B. Funding acquisition, A.D.B.

Funding: This research was partially supported by Project TESEO, funded by Regione Toscana POR-FESR 2014–2020, Grant No. 3389.30072014.068000017. This research was partially supported by Leonardo S.p.a.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

Conflicts of Interest: The authors declare no conflict of interest.

APPLICATION:

Applications of anomaly detection include;

- Fraud detection in financial transactions
- Fault detection in manufacturing
- Intrusion detection in a computer network
- Monitoring sensor readings in an aircraft
- Spotting potential risk or medical problems in health data and predictive maintenance.
- Real-time surveillance and remote monitoring are gaining large importance in a wide range of industries such as Oil and Gas, Power Grid, Industrial Automation, and Smart Buildings. The demand for high-quality real-time video streaming for surveillance applications is higher than ever and this calls for compute-intensive image processing designs and advanced encoding and decoding techniques.

- Rapid developments in Sensor technologies and embedded electronics, particularly the arrival of compute intensive SOCs, high-speed DSPs, high-speed wireless / wired interfaces, video streaming protocols and cloud technologies among others enabled cameras go beyond traditional surveillance and facilitate advanced AI-based vision applications such as safe-zone monitoring, No-go zone, object detection, etc.

- Design and implementation of a high-speed, real-time video streaming camera requires expertise in a wide array of embedded technologies. This article is an attempt to dive a little deeper into the design of a high-end camera for surveillance applications. We explore the basics of camera architecture and video streaming along with several streaming protocols, selection of sensors and processing engines, system software requirements and cloud integration among others.

- Image sensor is the eye of a camera system which contains millions of discrete photo detector sites called pixels. All imaging sensors - used in still, movie, industrial or security cameras - use either Charge-Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) sensor technology. The fundamental role of sensors is to convert light falling on the lens into electrical signals, which in turn will be converted to digital signals by the processor of the camera.

- Surveillance cameras demand higher light sensitivity, which produces better low-light images. Advanced image sensors provide high-quality images with low noise due to higher dynamic range, field of view and higher shutter speeds.

CHAPTER-7

REFERENCE

1. J. Kooij, M. Liem, J. Krijnders, T.Andringa, and D. Gavrila. Multi-modal human aggression detection. ComputeR Vision and Image Understanding, 2016.

2. S. Mohammadi, A. Perina, H. Kiani, and M. Vittorio. Angry crowds: Detecting violent events in videos. In ECCV, 2016.

3. Convolutional Neural Network (CNN) in Keras, https://towardsdatascience.com/building-a-convolutional-neural-network-cnn-in-keras329fbbadc5f5

4. Recurrent Neural Networks (RNN) in Keras, https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-forsentiment-analysis-af410fd85b47

5. W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. TPAMI, 2014.

6. X. Cui, Q. Liu, M. Gao, and D. N. Metaxas. Abnormal detection using interaction energy potentials. In CVPR, 2011.

7. T. Hospedales, S. Gong, and T. Xiang. A Markov clustering topic model for mining behavior in the video. In ICCV, 2009.

8. Y. Zhu, I. M. Nayak, and A. K. Roy-Chowdhury. Context-aware activity recognition and anomaly detection in video. In IEEE Journal of Selected Topics in Signal Processing, 2013.

9. L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using Spatio-temporal motion pattern models. In CVPR, 2009.

10.Howto Automate Surveillance Easily with Deep Learning,
https://medium.com/nanonets/how-to-automate-surveillance-easily-with-deeplearning-4eb4fa0cd68d,2018.

12. IR. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using the social force model. In CVPR, 2009.

13. I.Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. TPAMI, 31(8):1472– 1485, 2009.

14. B.Zhao, L. Fei-Fei, and E. P. Xing. Online detection of unusual events in videos via dynamic sparse coding. In CVPR, 2011.

15. Unusual crowd activity dataset of the University of Minnesota. In
http://mha.cs.umn.edu/movies/crowdactivity-all.avi.

16.A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. TPAMI, 2008.

17. Boss dataset,
http://www.multitel.be/image/researchdevelopment/research-projects/boss.php.

18. Waqas Sultani, Chen Chen, Mubarak Shah. Real-world anomaly detection in surveillance videos. In IEEE/CVF Conference, 2018.

19.DataAugmentation,https://medium.com/nanonets/how-to-use-deep-learning-when-you-have-limited-data-part-2-data-augmentationc26971dc8ced

20.Transferlearningfrompre-trainedmodels,https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751, 2018.

21.RegularizationinMachineLearning,https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a, 2017.

22. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In Google Research, 2015.

23. Gaurav Kumar Singh, Vipin Shukla, Pratik Shah. Automatic Alert of Security Threat through Video Surveillance System. In 54th Institute of Nuclear Material and Management Annual Meeting, 2013.

24. Biryukova EV, Roby-Brami A, Frolov AA, et al. Kinematics of human arm reconstructed from spatial tracking system recordings. J Biomech. 2000;33:985–995.

25. Yuan Gao, Hong Liu, Xiaohu Sun, Can Wang. Violence detection using Oriented VIolent Flows. In ResearchGate, 2016.

26. Waqas Sultani, Chen Chen, Mubarak Shah. "Real-World Anomaly Detection in Surveillance Videos", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

27. Junseok Kwon. "Rare-Event Detection by QuasiWang –Landau Monte Carlo Sampling with Approximate Bayesian Computation", Journal of Mathematical Imaging and Vision, 2019.