# Deep Learning Based Human Abnormal Activity Detection In Video

A Project Report

Submitted in the partial fulfillment of the requirements for the award of the degree of

## Bachelor of Technology

## In

## Department of Computer Science and Engineering

by


| | |
|---|---|
| 180030646 | Puppala Siva |
| 180030687 | Gadde Bharath |
| 180031087 | Akula Hemanth |
| 180031117 | Sandeep Cherukuri |


under the supervision of


## Ms. B. PRABHA

## Department of Computer Science and Engineering

K L E F, Green Fields,

Vaddeswaram- 522502, Guntur(Dist), Andhra Pradesh, India.


November, 2021

# Declaration

The Project Report entitled "Deep Learning Based Human Abnormal Activity Detection in Video "is a record of bonafide work of Puppala Siva (180030646), Gadde Bharath (180030687), Akula Hemanth (180031087), Sandeep Cherukuri (180031117) submitted in partial fulfillment for the award of B.Tech in Computer Science and Science to the K L University. The results embodied in this report have not been copied from any other departments/University/Institute

Puppala Siva (180030646)
Gadde Bharath (180030687)
Akula Hemanth (180031087)
Sandeep Cherukuri (180031117)

# Certificate

This is to certify that the Project Report entitled "Deep Learning Based Human Abnormal Activity Detection in Video" is being submitted by Puppala Siva (180030646), Gadde Bharath (180030687), Akula Hemanth (180031087), Sandeep Cherukuri (180031117) submitted in partial fulfillment for the award of B.Tech in Computer Science and Science to the K L University is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/ University/Institute

**Signature of the Supervisor**

Ms. B. PRABHA

**Signature of the HOD** **Signature of the External Examiner**

# ACKNOWLEDGEMENT

First and foremost, we would like to express our sincere gratitude to our respected Chairman, **Mr. Koneru Satyanarayana** and Vice-Chairman, **Mr. Koneru Havish** for their blessings and grace in making our project great success. We would like to place a record with deep sense of gratitude to our Honorable Vice-chancellor **Dr. G. Pardha Saradhi Varma**, for having given us the opportunity to pursue B.Tech., course in this prestigious institution.

We wish to express our sincere sense of gratitude to **Prof. V HARI KIRAN**, Head of Department of Computer Science and Engineering for giving us the permission to enroll in the course. Also, we express our sincere sense of gratitude to **MS. B. Prabha**, for her kind encouragement and moral support, who has been a constant source of inspiration to us. She had enlightened us with ideas and promoted our innovation skills which enabled us to complete our project successfully. We thank our faculty who introduced us to the new standard of documentation with latex by which this documentation has been carried out efficiently, progressively, and successfully.

Puppala Siva (180030646)

Gadde Bharath (180030687)

Akula Hemanth (180031087)

Sandeep Cherukuri (180031117)

# TABLE OF CONTENTS

# CHAPTER 1: ABSTRACT

Human Abnormal activity detection is one of the important issue in video surveillance. Which predicts the activity of a person and obtain frames to characterize the abnormality. The present detector use this method, However, only motion may not be able to capture all the forms of abnormality, in particular, poses that do not amount to motion. For this approach we are using machine learning and deep learning technique to train and test the frames, to reach the human detection approach using liner SVM. For the extracting of featured design and sensitive data we use convolutional neural network (CNN), one of the supervised learning method, extracts spatial structures and we also use the different concepts of CNN. Along with ML and DL the whole idea is implemented with the help of Opencv Which is used to extract the behaviour of objects. We also used the AlexNet which has largely trained dataset which is used for feature extraction. Optimal flow method, Motion influence map are also used. A lot of research is conducted and a lot of abnormal activity detectors have been done with different methods and many techniques are developed.

# CHAPTER 2: INTRODUCTION

Human Abnormal activity detection is the one of the mechanism in which the motion and movements of a person is captured and then compared to the trained dataset which has contain different abnormal and normal activities of human. Initially the abnormal activities are detected using handcrafted features and showed that these are useful for specific tasks, this is also featured as the traditional way of detecting the abnormal activity. In this method we use Local Binary Patterns (LBP) or Scale-Invariant Feature Transform (SIFT) to extract the features for the image/video. The main two ways to extract features are supervised and unsupervised learning. Supervised learning takes care of input and output data to train the features. Unsupervised learning uses only the input data for feature training. As the detection of human moments are very difficult and we need large number of data to train and test contain normal and abnormal activity, so to save time we have AlexNet which has trained dataset, which is one of the CNN architecture. Basically, the CNN we used has five layers, max pooling layers, three fully connected layers. Apart from these methods we also use optical flow method for human activity detection in various applications and background stabilization is also obtained by this method.This abnormal activity has wide range of applications like military surveillance, ATM's etc. Recently it is used to identify the criminals from the CCTV which uses the optimal flow method. If this idea is developed to full extent, we can reduce the crime rate as well as identify them. We can also stop crime before it happened.

# CHAPTER 3: LITERATURE SURVEY

There are an ample of papers and research done on the Human Abnormal activity detection
Using DL Algorithms. Some of them are presented below –

## 3.1 Real-Time Human Detection for Aerial Captured Video Sequences via Deep Models

This paper was written and published by Nouar AlDahoul, Aznul Qalid Md Sabri, and Ali Mohammed Mansoor in 2018.

comparisons between three different deep models namely CNN supervised, CNN-trained, and HELM on display to study the features and construction of the UCFARG model online database. The optical flow model is added as the original on stage in three training and testing programs samples as inclusion in deep models.
The youth of our work is as follows:

**1)** To the best of our knowledge, this is the first project
using different in-depth UCFARG community models
online database to find people.

**2)** CNN is monitored for best results features that discriminate between the two categories of  humanity and personality. Soft-max and SVM are used
is the last layer of CNN to produce a split output.

**3)** The pre-trained AlexNetCNN model already exists trained in the ImageNet database to get visual recognition dividing 1000 different classes is shown as output element with fixed parameters in the background to remove layers that are fully connected to discriminate aspects of human segregation, non-human.

**4)** HELM is also being discussed for consideration a trade between high accuracy and low training time.

**5)** Comparison between CNN supervised feature student, CNN pre-trained as a feature detector, as well HELM as a student of unconditionally unconditional element learning speed and accuracy are assessed by five human actions.

## 3.2 Unusual Human Activity Detection Using Opencv Python With Machine Learning

This paper was written and published by Abhishek S. Mohite, Darshan K. Sangale, Prathamesh R. Oza, Tushar D. Parekar, Prof. Manisha P. Navale, in 2020.

We Recommend A Program That Uses Opencv Library Launched In Python.Opencv First Developed By Intel For Image Processing.Designed For The Performance Of Various Actions In Photography.Using The Numpy Library To Save Visual Images In Frame Stream. Basically the picture is broken into frames. Then These Images Are Stored In Stream Of Bytes. Subsequent Frames Are Compared IE If There Is A Move Of An Object And That Byte Stream That Representing That Thing Will Change. Basic Steps Are

**1) Visible Flow of Blocks:**

Frame Divided Into Blocks. Actually Each Video Is Broken Into Frames And These Frames Are Divided Into Blocks. Where The Index Can Be Given To All Blocks By {A1, A2 ... An} .Frame Size 240 * 320 Divided Into 48 Blocks When Each Block Has A Size 20 * 20.

**2) Calculate Visible Flow of All Blocks:**

One Frame Divided Into Blocks. Then Each Frame Is Considered Separately And Optical Flow Is Calculated In Each Frame. This Visual Flow Is Essential For Continuous Image Processing To Produce A Motion Effect Map. After Splitting The Frame In Blocks We Calculate The Visible Flow Of All Blocks By Calculating The Average Of All Pixels In A Block.

**3) Generate Motion Influence Map:**

Motion Influence Map Produced Using. Basically This Map Defines Block Movement. Using the Movement Impact Map The Behavior of the Object is being analyzed.This behavior is read and based on this type of behavior is determined. We can use this Movement Map to Analyze and Generate Human Activity Patterns. Similar Tasks I.E Will Have the Same Influence Map.

**4) Feature Domain:**

Using the Motion Effect Map Map Properties. These Elements Are Learned To Identify The Type Of Behavior. The Removal Phase is Important in Determining the Behavior of Objects and Identifying the Behavior.

## 3.3 Vision-based human motion analysis

This paper was written and published by Ronald Poppe in 2007.

Measuring the shape from frame to frame is often called to track. Tracks are used to ensure temporary compliance between stopping over time, and providing an initial stop rate. When it is assumed that time in between subsequent the frames are smaller, the distance to the body configuration is possible to be thin again. These differences in configuration can be it is almost traced, for example using a Kalman filter. Traditional tracking was intended for attention one hypothesis over time. As this often causes to measure the loss of a track, the most recent work distributes many hypothesis over time. Generally, a sample based method taken. In some embodiments, the interim is achieved by to reduce the change in position sequence to a batch method. Related to this is 3D simulation stand from 2D points. Although this article is out of The scope of our whole view, is important and selective install it. This section discusses these methods Overcoming the problem of flooding of one hypothesis tracking methods, many ideas can be saved. Cham and Rehg [16] use a set of Kalman filters to we spread many hypotheses. This results in increased reliability movement tracking has a single Kalman filter. Testing in a challenging dance track it shows multiple hypotheses are able to track movements in single mode failed. However, because of their limited model of appearance, rotation about the axes of the legs could not be limited. Human movement is not linear due to the combined acceleration. However, Kalman filters are only suitable for tracking line movement. Sampling-based methods is able to track indirectly movement. Normally, a number of particles are distributed over time he uses a dynamics model, which includes part of the sound. Each particle has a relative weight, i.e. new depending on cost work. Configuration at lower cost they are assigned a higher weight. From all weights one summary, stop rate obtained by the total weight of all particles.

Although, in theory, sample-based methods are very numerous ready to track, high size requires I the use of multiple particles to sample the freezing point dense. All particles come with an increase in computational costs due to the distribution of particles according to dynamical model and cost function test. For each part, the model of the human body should be provided and compared with extracted image descriptors. Another problem is the fact that particles often combine in a very small area.

## 3.4 Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers

This paper was written and published by P. Lukowicz, J. Ward, H. Junker, M. Stager ,G. Troster, A. Atrash and T. Starner in 2001.

Speed or sound does not provide sufficient information complete disposal and segregation of all relevant functions; however, we estimate that their source of error may be statistically inconsistent. So, we improve a process based on the combination of both methods. Our process has three steps:

**1)** Extrusion of appropriate data segments using the stiffness difference between the arm and the chest tube. We expect this process to be different data distribution on individual actions (including multiple actions that we will liken as noise).

**2)** Independent classification of actions based on that
noise or acceleration. This step will expose the imperfections
the effects of recognition for both sound and acceleration
underlying systems.

**3)** Removal of false material. While the systems under each sound and acceleration are imperfect, they are their own
the classification is consistent, the result may be
very reliable (if there is statistical independence between error sources in audio and acceleration channels).

In order to recognize touch, a model is trained for each of them
touch to see. In our test, the touch set includes saw, ball, screw, hammer, sand, file, drawer,vice, and applause. Once the models are trained, the sequence of
features can be transferred to the calculator calculator , Opportunities for each model in terms of viewing sequence and restores the most likely touch.
For our experiments, The set of features contains readings from accelerometers placed on the wrist and elbow. This gives 6 values of whole elements are then adjusted to have their total one and collected at about 93 Hz. We found that most of the teaching activities in general they only need simple modeling HMMs. For file, sand, saw, and screw, 5 country model with 1 skip version once 1 loopback transtion is sufficient because it contains a simple repeated movements. Drill is best represented using 7 state

model. Hand clapping, drawing, and grinding are a little more complex and required 9 world models. The vice is different that it has two different movements, opening and closing. So the 9-shape model is used with two loopbacks suitable for

stands for action. These models are selected by examining data, understanding the environment for jobs, and information about HMM.

## 3.5 Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring

This paper was published by Dean M. Karantonis,Michael R. Narayanan, Merryn Mathie, Nigel H. Lovell and Branko G. Celler in 2006.

A guiding principle in building our system is profitability real-time separation of movement signals in a small, wireless area, low-power device. The analysis of TA exit is made of portable microcontroller unit and information about user movement is transferred to the local receiver before being transferred to the location computer for display and testing. With the knowledge that the memory and processing capacity of an embedded microcontroller are as limited as compared to a PC, the range of motion that can be split will face the same limit. So, in defining the system requirements, monitoring was taken with any of the categories involved complex computer tasks can be reasonably simplified during the development process without dedication the purpose of internal processing. However, the range of motion that needs to be separated by the system was identified as important in providing adequate evaluation of user activities. Central to the system is the ability to detect falls.

In this way, our device acts as an automatic personal alarm service, with the ability to alert the appropriate alert. staff where assistance may be needed. Moreover, it is desirable that any signs of recovery after a fall are indicated by the presence of the user's activity and their subsequent postural umumo. The reason here is that the level of activity as well post-fall user posture provides important information on the severity of the fall and the user's well-being. For this reason, the system is required to differentiate between times of work and rest, upright posture and lying, and lying subpostures — if the user is lying to the right or left, on his back, or face down. In addition to the discovery of the fall, the system is required to detect other types of activity, such as such as sitting and standing, and depending on the limits of the hardware, to go again. The average cost of metabolic energy will be they are also provided. Such knowledge will provide valuable insight at the user's level of physical activity and general health status. Low power consumption is a major factor to consider in the system design. This requirement will improve system usability by ensuring the battery life is in a portable unit and freeing the user from residential work recharge or replace the battery.

## 3.6 Activity Recognition from Accelerometer Data

This paper was published by Nishkam Ravi and Nikhil Dandekar and Preetham Mysore and Michael L. Littman in 2001.

The task recognition algorithm should be able to detect the accelerometer signal pattern that corresponds to everything function. It is easy to see that every job has a a different pattern. We create job recognition as a problem of classification where classes are associated with tasks as well an example of a test data set set of accelerated values over time and processed after one example
{i.e., standard deviation, power, relation}. We evaluated the performance of the following basic level dividers, available on the Set toolkit:

• Decision Tables

• Olive Trees

• K-neighbors nearby

• SVM

• Naive Bayes.

We also tested the performance of some high-level meta separators. Although the overall performance of meta-level dividers is known to be better than that of standard level dividers, basic level dividers are known for performing better than meta level separators on fewer data sets. One of the purposes of this work was to determine whether merging by divisions was the right thing to do recognition from accelerometer data

## 3.7 Human activity recognition using smartphones

This paper was published by D. Anguita, A. Ghio, L. Oneto, X. Parra, and JL Reyes - Ortiz in 2013.

We used cellular accelerometers and gyroscopes to collect linear acceleration of the three axis and angular velocity signals at a sample rate of 50 Hz. The Butterworth third-order low-pass filter filter with a frequency of 20 Hz is used to pre-process these signals to reduce noise. This measure is sufficient to record human movement because 99% of the available energy contains less than 15 Hz . That the acceleration filter also separates the Butterworth-low-pass which has part of the body movement and gravity speed and gravity. Considering that the gravitational force has only low frequency components, we have found from experiments that 0.3 Hz is the best rotating frequency of a constant gravitational signal.

By calculating Euclidean amplitude and time out (jerk da / dt and angular acceleration dw / dt) from three-axis signals, other time signals can be detected. Then, at a fixed time slide window of 2.56 seconds in signal samples, 50% of the interval between them, such as:

• normal flow rhythm [90, 130] Step / min [14] or less, that is, lower 1.5 steps per second;

• At least one complete cycle cycle (two steps) is required for each window sample;

• People with low speed (such as the elderly and the disabled) should also benefit from this approach. We assume that the low speed is equal to 50% of the normal human rhythm;

• The signal is also mapped to the frequency range by Fast Fourier Transform (FFT), the signal is designed for two vectors (2.56sec × 50Hz = 128cycles).

Thus, a total of 17 symbols were obtained in this way.

From each of the sample window described above, the element vector is obtained. The feature map uses common metrics previously used in HAR literature [15], such as definition, aggregation, signal amplitude (SMA) and autoregressive coefficient [16]. To improve learning performance, a new set of features is also adopted, which include power in different frequency bands, frequency skew and angle between vectors (such as central acceleration and vector y). Table 3 lists all time metrics that apply to the time zone and frequency signals.

subtracted a total of 561 elements to define each active window. To facilitate performance testing, the data set is also randomly divided into two independent sets, of which 70% of the data are selected for training, and the remaining 30% of the data are selected for testing. Data sets to detect human activities are already available to the public and presented in the form of green inertial sensor signals and vectors included in each mode. Posted as a set of "Personal Activity Via Data" data set.

## 3.8 HUMAN ACTIVITY RECOGNITION WITH SMART PHONE DATASET

This paper was published by Dr. C .N.Sujatha , H.Alekhya , Ch.Akhila , S.Pooja Rao in 2020.

Acknowledging a person's action is intended to distinguish the actions of the person who is inclined towards him or her being performed as often as possible.

in his daily life as a normal situation. Acknowledgment can be made for the misuse of data obtained from various sources, for example, the situation through the use of senses. In this project we collect data from 30 volunteers from the age of 19-48 and process it information using SVM calculations. Acceptance of a person’s action takes on a significant role in one-on-one communication relationship relationships. Verification of human action is a matter of compiling data collections of accelerometer engraved on certain garments or PDAs are then implemented around the defined rotation of events. Progress is a common internal test, for example, walking, talking, standing, and sitting. They may also be more physically active, for example, in the form of exercise kitchen or production area. In this project we processed data using an algorithm called SVM and used python software

and a few libraries and separate methodology. we used the tree algorithm of machine learning decision algorithm to differentiate retreat. The decision tree can be used to represent data in a transparent and transparent manner. Verification of a person's action will be a guarantee of evaluating the evolution of events and activities by foreseeing progress. It is a difficult issue, as a large amount of understanding is conveyed every second, the general thought of the commentary, as well

the lack of an effective strategy for combining accelerometer data to detect changes in events.

The recognition of human action has broad functions in medical research and the human research framework. In 2 work, we set up a cell phone
an affirmative-based framework that recognizes five human fitness: walking, standing, sitting, positioning, walking up and down first
down. The best character rating in our investigation was 84.4%, which is achieved by SVM with SFS-selected highlights. Consolidation
execution is a strong direction and status of cell phones. The results of the analysis showed the flexibility of learning to save money
marking function while performing the same function with stealth reading. Clearly, SVM is the right decision for our concerns.
Future work may consider additional tests and create a continuous framework for mobile phones.

## 3.9 Classification Accuracies of physical Activities using smartphone and motion sensors

This paper was published by wanmin wu , sanjoy Dasgupta in 2012.

Observing one's work involves a certain level of low self-esteem physical actions occur in strong emotions from fall vision, walk and stand-based knowledge, the cost of metabolic energy, up to employment physical activity.

Physical activity research basically research aimed at the ability to identify any series the body movement of the object. About human activity recognition, aimed at identifying the physical activity being performed about the person as a research object. By including smartphones, the researchers aim effective recognition of individual and user activity he is free in mind. With their beneficial features compared to the wearable sensors, the recognition of a person's work of some users are expected to be made invisible when you look at the way people behave and the environment around them places. Additionally, dress variability sensory-based systems require a separate processing unit, using a smartphone is expected to have content processing integrated into a single device from which data is collected. This you will promise a quick and reliable decision result as there is none for a long time we need some space in between processing unit and sensor. Feedback on specific services and recognition is successful and is now possible as smartphone has a variety of connectivity options. With those several chances, in the end, a smartphone-based system will do let the recognition of one's work be felt even in time robust movements of movement or movement of users.

Recognition of human activities is highly targeted the development of an intelligent health care system. Health problem in the end it is a sensitive issue that encourages research conduct research on the recognition of human activities. Research much-needed health care patient monitoring and treatment diagnostic system architecture using a physiological bodyworn and wireless sensor network. in their paper measure the daily cost of energy activities and sports activities as a basis for physical activity separation. Interestingly use the kinesthetic wearable sensor for side patients to find their extra limb the touch on which this function is followed by Tognetti. That and interestingly use clothing-based sensors as the upper limbs kinesthetic clothing and hearing gloves.

## 3.10 A Smart Home Monitoring System for Abnormal Human Activity Detection Using SVM

This paper was submitted by Ms. Apurva Landge and Prof. Sandip Kahate in 2015.

Providing an Appropriate Solution for Recognition of Unusual Personal Activity Solutions Using the Multiclass SVM Approach a very important factor in a smart home concept. In order to capture a person's activity and analyze that data, both functions they are very important in finding an unusual job. The image classification directly affects performance of the system in which this image data is used as input. Introducing a new way to improve the accuracy of the Unusual Web Site Service the division uses K-Means, a random forest and a multi-class SVM. SVM helps to separate data from it is used to separate images and its data. K-means is one of the easiest ways to learn unattended solving a well-known integration problem. Create collections of the same function. It counts the distance between the function attribute of each data and forms a separate task group. The Random Forests algorithm is correct An algorithm for using complex partition tasks. As we use K-Means and a random forest together to split up and get an unusual job successfully.

When it finds and separates work and associates certain occurrences of function, we store its relevant entries on the website. So it can create a model of awareness task and begin to analyze the feasibility of the existing and new task. But how then find a new job or part-time job that is not on the website? In that it uses Active a guide that will identify the concept of Separate reading during practice. The framework for tracking, learning and monitoring does not depend on the types of sensory data or device types, so the sensor data source is for any type of data. Selecting the appropriate set of parameters or attribute is important in improving visual accuracy. Let's say people the purpose of performing physical activities, which may include warm-ups where the combination is different small activities are performed such as raising hands, sleeping, jumping, walking, and running. Each the lower function can also be divided into finer movements of the legs, joints, and the muscles also on that basis are acquired properly. Extraordinary activities are defined using the state transformation table, which captures all possibilities he says. The program is trained to isolate and report on individual activities unusual. The program detects 9 different personal tasks using multiple categories SVM. The typical system structure is shown in Fig 1. A sliding window is used split data into size N window so that the system can see the function. Sliding window lowers the flow rate and sends small data to the system to see the work done by individual. The proposed system uses seven features of each axis, the standard deviation of each axis axis and speed. These features help to reduce the noise in the database and the impact on separating data with high accuracy.

## 3.11 Detecting Unusual Activity in Video

This paper was published by Hua Zhong Carnegie, Jianbo Shi ,Mirko Visontai in 2004.

Introducing the unattended way to find unusual activity in a large video set using many simple features. There are no complicated work models and no feature is monitored selected are used. We split the video into equal lengths
then split the extracted elements into prototypes, from which the prototype-segment co-occurrence matrix is derived counted. Encouraged by the same problem in analyzing the keyword of the document, we seek correspondence between prototypes and video segments that satisfy flexible closing limit. We show that it is important part of the family of book works can be reduced to embedding shared prototypes and segments into N-D Euclidean space. We prove it to be an effective, best global algorithm exists with the problem of co-embedding. Examining the various virtual videos has confirmed our approach. We suggest a way to use the "hard to describe" but "easy to verify" an unusual event venue without having to create public models for casual events. One can compare eachevent and all other events watched to determine how many similar events exist. If the event is regular, it should be
many similar events in this big data. If there are no similar events we look at this phenomenon: although the event is unknown, it is unique. Thus, detecting unusual events in a large data set does not require modeling common occurrences, but rather the ability to compare two events and measure their similarities. Our overall goal is to produce simpler and more reliable features

which are descriptive, i.e. can be used by an unsupervised person an algorithm for finding important image features in a large area video set to detect unusual events. Our goal is to compile documents by identifying specific links between keywords and documents. We can think of this problem as the layout of the graph. The Edges incident continues Normal and random keywords should be deleted, otherwise every graph will be heavily linked. The boundary event in the instructional keywords should be maintained, otherwise the graph will be slightly linked. Cruel power The strategy would be to search for all possible additions and the removal of the edge so that the graph separation is not too slow or too smooth.

## 3.12 Graph formulation of video activities for abnormal activity recognition

This paper was published by Dinesh Singh , C. Krishna Mohan in 2019.

The word bag method (BOW) is inclusive an unstructured histogram of the emerging vocabulary that only covers the worldwide distribution of low-level adjectives, while ignoring the local organization of the building structure (i.e. geometry) of key points and definitions associated with the lower level. However, the use of such a local structure key points and explanations that are consistent with the low standard should lead to discriminatory video representation that leads more to better recognition of video activities. In this work, we propose a unique work recognition framework that includes appearance and power and geometric relationships between the various interactions of businesses in the video industry. First, we extract the points of interest of space time and manage each point of interest as a graph node. The edges of the graph are determined for use an obscure membership function on the basis of the closeness and similarity of the organizations related to points of interest. If the two points are closer to each other, then there is a higher point .The potential for further interaction between affiliated businesses. To to keep track of things, we also integrate the look and movement of organizations using histogram of oriented gradients (HOG) and histogram of oriented optical flow (HOF).

Second, the upper part of the margin is trained on the basis of the geometric structure of graphs designed for general and informal training videos. Graph letters are used to measure similarities between two graphs. Graph letters give strength to a slight topological deterioration in comparing the two graphs because they weigh on a basis of the same routes / trips. This deformation is possible due to various factors affecting like the presence of noise in the data. The concept of constructing a video function such as a graph and the use of a graph The kernel of its similarity is a novel detection of unusual activity in surveillance videos. Finally, the integrated approach provides a solid framework for recognition of unusual activities in surveillance videos. Tests show elevation of the proposed project on existing roads based on congested roads as well word bag with various character meanings.

## 3.13 Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks

This paper was published by Damla Arifoglu, Abdelhamid Bouchachia in 2017.

We believe that the orderliness of activities and their temporary and local knowledge is essential to coding the elderly the daily processes of human life. This type of information can provide important clues to understanding everyday patterns as well thus finding any confusion in those patterns. Sequential labeling methods such as HMMs and RNNs can record temporary and spatial relationships between operations, other production methods such as SVMs can do. In this we are working, investigating the suitability of RNNs for this work. In order to identify everyday tasks, training conditions for data sets and their corresponding labels are included RNNs. Then, when a new test sequence is introduced, the trained model provides labels for each job situation of that sequence. Each model provides a certain amount of confidence about a given label for a new sequence. First, we calculate the confidence level of the training conditions given to the model. Then, when the new test sequence is introduced when the model gives you a class label with a confidence level greater than the definition, sequence is considered a normal function, otherwise it is a rare function.

With the acquisition of unconventional work, we considered only LSTM and compared it with NB, HSMM, HMM, SVM and CRF. Percentage accuracy of TPR and FPR are consistent; 40.40% and 43.50% of NB, 58.36% and 96.20% of HMM, 68.85% and 32.2% in HSMM, 66.22% and 40.50% in CRF, 72.11% and 44.0% in One-class SVM and 91.43% and 40.96% of LSTM. We used only the last element in this test. The results indicate that LSTM is the best pruning for false negatives compared to other methods. Methods like NB, One-class SVM do not capturing data order does the worst. Models ignore the frequency of work, but use it temporarily once status information to make a decision. The results show that LSTM is able to code the order of tasks. Therefore, when a project is presented in a different context or in a different format, LSTM may experience such confusion.

Our current approach may fail to detect abnormalities, when there is a gradual deterioration of health an older person. We plan to address this issue in the future as we gradually collect real-world data deterioration can be seen.

## 3.14 Untact Abnormal Heartbeat Wave Detection Using Non-Contact Sensor through Transfer Learning

This paper was published by jin-soo kim , kangyoon lee in 2018.

First, by using non-contact sensors, this study supports the sensitivity of sensory communication through a transmission learning model. Second, the unaffected sensor model was tested in a real-life environment. In contrast to existing studies that conducted study and evaluation with open heart rate data sets, in this study, experiments were performed and data collected from the test were used in a study model using an open data set and the results were displayed. The abnormal heart rate separator model is read on an open site, and bpm data measured by contact and non-contact sensors are used in the model. By using this model to feel unaffected, we ensure that possible processing is possible. To improve non-tactile sensor performance, we suggest and evaluate filtering algorithms and the complete window size, and display the optimal results. The performance of the unaffected sensor is compared to that of the contact sensor, and a filtering algorithm is developed to improve the reliability of the unaffected sensor data. There have been many studies using the factors of heart rate variability (HRV). However, previous studies have used the invalid features of HRV in learning [2]. There is a limit to the fact that only reading and testing with open data sets is done. This study has the difference between enhancing learning efficiency by minimizing scattered signals and conducting experiments in unaffected areas. The results of this study will help to promote the use of unaffected sensors in various systems. Touchless sensors provide excellent ways to monitor the progressive biomedical data of physicians and health care professionals by enabling intermediate measurements during daily activities. In addition, they will assist in the rapid implementation of AMI in the healthcare sector.

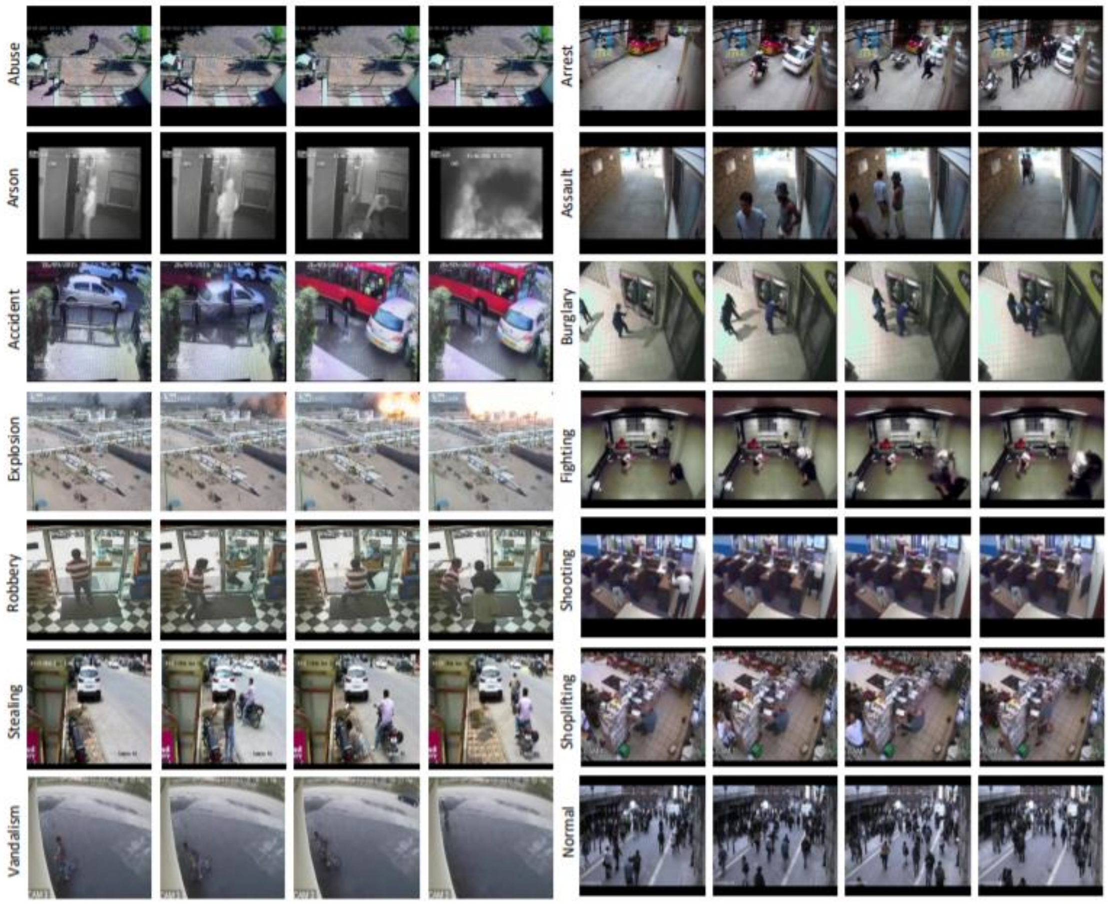
## 3.15 Multi-Modal Anomaly Detection by Using Audio and Visual Cues

This paper was published by Ata-ur rehman, hafiz sami ullah, haroon Farooq in 2008.

Several algorithms have been proposed based on the sound of confusing discovery in literature over the past decade. For example, algorithms for analyzing individual events that include speech recognition through machines introduced in The classification of hearing aids is proposed while the speaker recognition systems are introduced in . In addition, there has been a growing interest in the analysis of noise behavior in order to detect event in security systems related to public transport safety. There is a lot of research work to be done in audio classification books using hand-crafted features. These hand-made features are mainly used to train networks such as Support Vector Machine (SVM) for segmentation purposes. Manually designed features to train the SVM model to distinguish audio signals introduced from. They propose an Expectation-Maximisation (EM) algorithm to quantify the representation of feature space in the Gaussian distribution. And different handmade features are used in shouts and gunshots. In total they use 49 different features based on visual, visual, and temporary streaming.

# CHAPTER 4: DATASET

This dataset was collected from UCIF datasets which has a wide range of video sequences corresponding to the human abnormal activity detection.



# CHAPTER 5: REQUIREMENTS

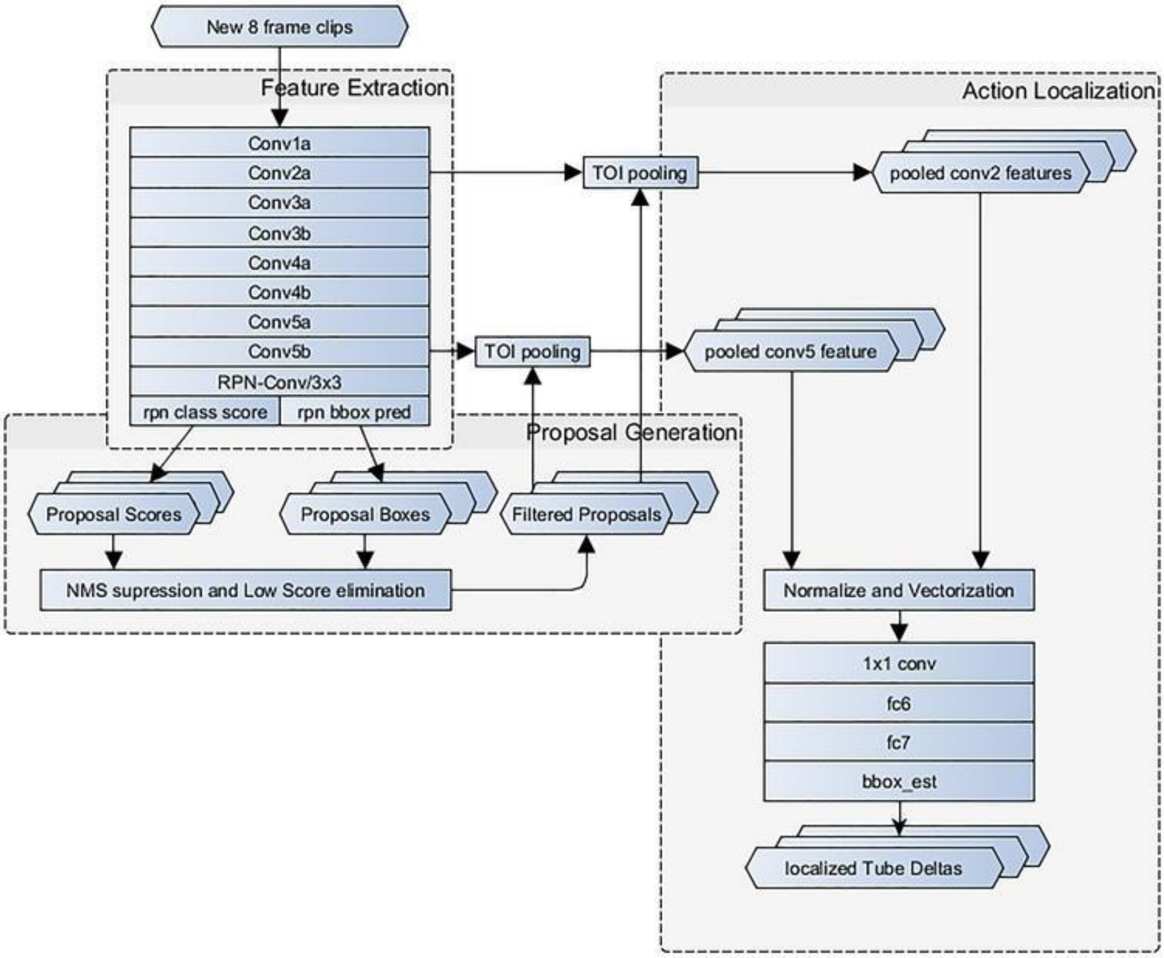
## 5.1 Software Requirements

- Spyder Python IDE
- Operating System: Windows
- Language: PYTHON

## 5.2 Hardware Requirements

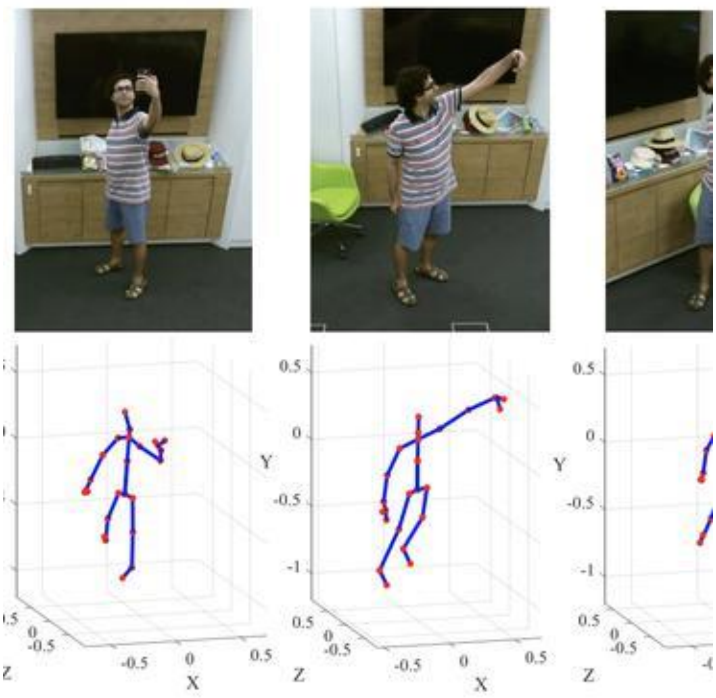
- Processor: Intel I5 or AMD Ryzen 5
- RAM: 16 GB
- ROM: 1 TB
- Graphic Card: 512 GB

# CHAPTER 6: ARCHITECTURE



# CHAPTER 7: METHODOLOGY

The dataset consists of a video sequence. The given video sequence is trimmed into images, We map layers to values to differentiate between normal and abnormal activities. To detect persons in the given video sequence we use openCV and Deep Learning. The CNN algorithm uses the detected person and traces line segments on the person and form it into a skeletal shape, which we use to identify humans. According to the trimmed images and trained dataset, it uses the confusion matrix and differentiates the actions performed by the persons and divides it into categories of Assault, theft, violence etc. And finally it displays a message on the screen indicating the corresponding action which is compared to the already determined values of categories. It uses the concepts of object detection and object tracking. We also implement some of object tracking algorithms named Simple on-line and real time tracking, Generic object tracking using regression network and Multi-Domain network.



# CHAPTER 8: ALGORITHMS

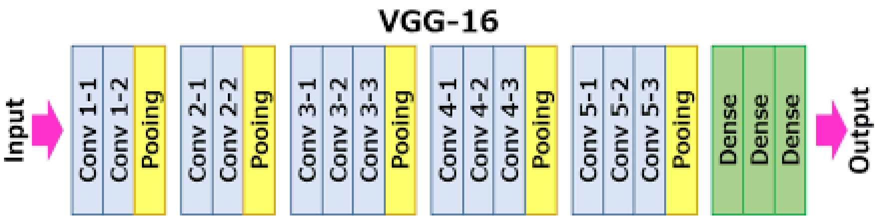
## 8.1 CNN (Convolutional Neural Networks)

In deep learning, the convolutional neural network (CNN, or ConvNet) is a phase of the deep neural network, which is widely used to analyze visual images. They are based on the distributed weight structure of convolution characters or slider filters and input features and provide the same translation responses known as feature maps. Contrary to intuitively, most convolutional neural networks are only equal, as opposed to consistent, and interpretive. They have applications for image and video recognition, complimentary programs, image classification, image classification, medical image analysis, natural language processing, brain-computer communication, and a series of financial periods. CNNs are standard versions of multilayer perceptrons. Multilayer perceptrons usually refer to fully connected networks, i.e., each neuron in a single layer is connected to all neurons in the next layer. The "full connection" of these networks makes them prone to data overload. Typical ways to do so, or to prevent excessive immersion, include: punitive restrictions during training (such as weight loss) or interconnection (skipping links, stopping out, etc.) . and incorporates growing weight patterns using small and simple patterns labeled in their filters. Therefore, in terms of connectivity and complexity, CNNs are at a very low level. Convolutional networks are promoted by biological processes in that the pattern of communication between neurons resembles the organization of the visual cortex of animals. Each cortical neurons respond to stimuli only to a limited area of the visual field known as the receptor field. The receiving fields of different neurons are so intertwined that they close the entire field of view. CNN uses less advanced processing compared to other image classification algorithms. This means that the network learns to optimize filters (or kernels) by automatic reading, and in traditional algorithms these filters are manually made.

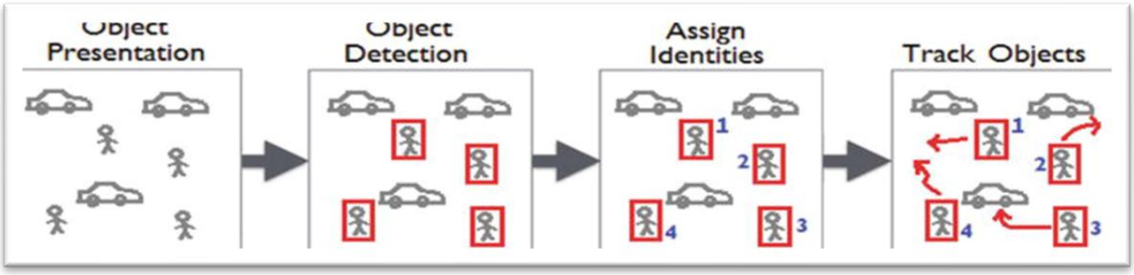
## 8.2 VGG Network (Visual Geometry Group)

VGG addresses another very important aspect of CNNs: depth.

- **Input:** VGG takes in a 224x224 pixel RGB image.
- **Convolutional Layers:** The convolutional layers in VGG use a very small receptive field. There are also 1x1 convolution filters which act as a linear transformation of the input, which is followed by a ReLU unit. The convolution stride is fixed to 1 pixel so that the spatial resolution is preserved after convolution.
- **Fully-Connected Layers:** VGG has three fully-connected layers: the first two have 4096 channels each and the third has 1000 channels, 1 for each class.
- **Hidden Layers:** All of VGG's hidden layers use ReLU. VGG does not generally use Local Response Normalization. as LRN increases memory consumption and training time with no particular increase in accuracy.



## Use of Object detection and object tracking algorithms




## challenges with object detection:

**1) Variable number of objects:**

We already mentioned the part about a variable number of objects, but we omitted why it’s a problem at all. When training machine learning models, you usually need to represent data into fixed-sized vectors. Since the number of objects in the image is not known beforehand, we would not know the correct number of outputs. Because of this, some post-processing is required, which adds complexity to the model. Historically, the variable number of outputs has been tackled using a sliding window based approach, generating the fixed-sized features of that window for all the different positions of it. After getting all predictions, some are discarded and some are merged to get the final result.

**2) Sizing:**

Another big challenge is the different conceivable sizes of objects. When doing simple classification, you expect and want to classify objects that cover most of the image. On the other hand, some of the objects you may want to find could be a small as a dozen pixels (or a small percentage of the original image). Traditionally this has been solved with using sliding windows of different sizes, which is simple but very inefficient.

## Object Tracking:

Object tracking is a field within computer vision that involves tracking objects as they move across several video frames. In this document, we’ll address the difference between object tracking and object detection, and see how with the introduction of deep learning the accuracy and analysis power of object detection vastly improved. Object detection has evolved substantially in the past two decades, with the move from traditional statistical or machine learning approaches to deep learning approaches based on Convolutional Neural Networks (CNN). The introduction of deep learning improved the accuracy and analysis power of object detection by an order of magnitude. To some, object tracking is simply an extension of object detection. The creators of a popular algorithm called Simple Online and Realtime Tracking (SORT) make the assertion that modern object detection algorithms can do most of the work of detecting objects and re-identifying in subsequent frames, and object tracking can be reduced to simple heuristics.Others have developed extensive object training algorithms that work in tandem with object detection, and apply deep learning techniques to carry over an identified object into the next video frames.

## Types of Object Tracking:

Object tracking is a field within computer vision that involves tracking objects as they move across several video frames. In this document, we'll address the difference between object tracking and object detection, and see how with the introduction of deep learning the

There are two main types of object tracking:

1. **Offline object tracking**—object tracking on a recorded video where all the frames, including future activity, are known in advance.
2. **Online object tracking**—object tracking done on a live video stream, for example, a surveillance camera. This is more challenging because the algorithm must work fast, and it is not possible to take future frames and combine them into the analysis.

## Levels of Object Tracking:

On a high level of abstraction, there are mainly two levels of object tracking.

1. Single Object Tracking (SOT)
2. Multiple Object Tracking (MOT).

Object tracking is not limited to 2D sequence data and can be applied to 3D domains as we.The strength of Deep Neural Networks (DNN) resides in their ability to learn rich representations and to extract complex and abstract features from their input. Multiple Object Tracking (MOT), also called Multi-Target Tracking (MTT), is a computer vision task that aims to analyze videos to identify and track objects belonging to one or more categories, such as pedestrians, cars, animals and inanimate objects, without any prior knowledge about the appearance and number of targets.While in Single Object Tracking (SOT) the appearance of the target is known a priori, in MOT a detection step is necessary to identify the targets that can leave or enter the scene. The main difficulty in tracking multiple targets simultaneously stems from the various occlusions and interactions between objects that can sometimes also have a similar appearance. Thus, merely applying SOT models directly to solve MOT leads to poor results, often incurring in target drift and numerous ID switch errors, as such models usually struggle in distinguishing between similar-looking intra-class objects.In recent years, due to the exponential rise in the research of deep learning methods, there have been tremendous gains in accuracy and performance of the detection and tracking approaches.Most of the state-of-the-art tracking approaches follow the ‘Tracking by Detection’ scheme where they first find objects in the scene and then find the corresponding track lets (position of it in the next frame) of the objects.Today the detectors are performing exceptionally well and can scale to different scene adaptations. Consequently, it has led to the standard input to tracking algorithms.

## Object Tracking vs Object Detection:

Object detection has evolved substantially in the past two decades, with the move from traditional statistical or machine learning approaches to deep learning approaches based on Convolutional Neural Networks (CNN). The introduction of deep learning improved the accuracy and analysis power of object detection by an order of magnitude.To some, object tracking is simply an extension of object detection. The creators of a popular algorithm called Simple Online and Realtime Tracking (SORT) make the assertion that modern object detection algorithms can do most of the work of detecting objects and re-identifying in subsequent frames, and object tracking can be reduced to simple heuristics.Others have developed extensive object training algorithms that work in tandem with object detection, and apply deep learning techniques to carry over an identified object into the next video frames.

## Challenges Faced by Object Tracking:

While solving the object tracking problem, there arises a number of issues which can lead to a poor outcome. The algorithms over the years have tried to tackle these issues but till now we have not arrived at a full proof solution keeping it an open-ended area of research.

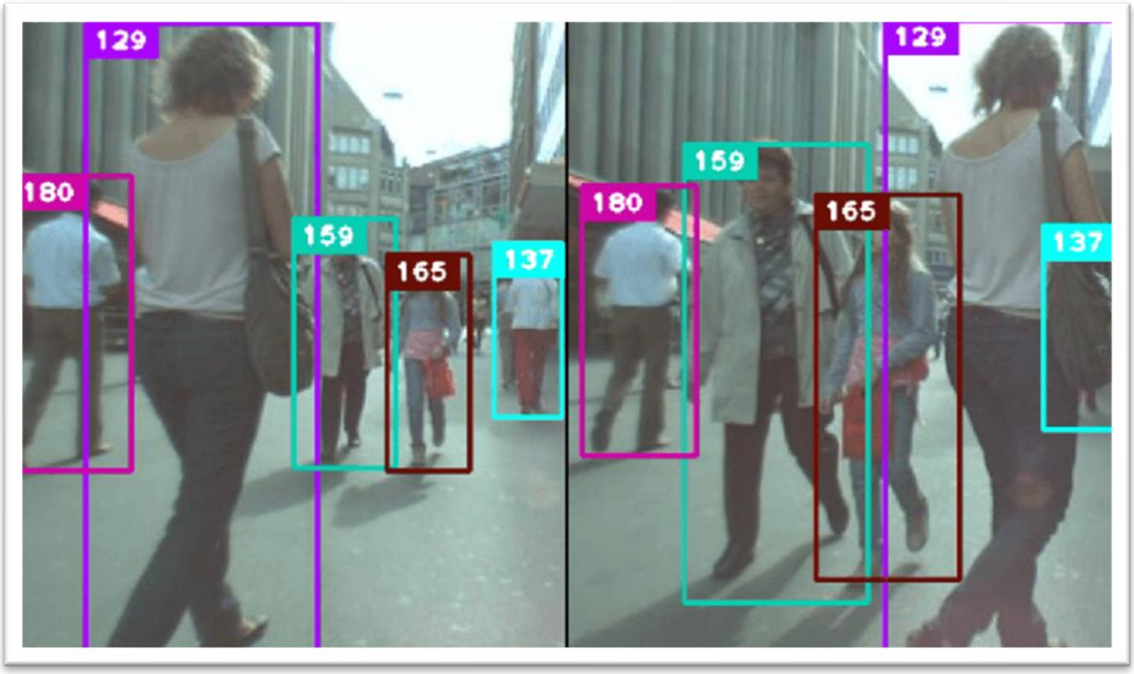
1. **Re-identification—**Connecting an object in one frame to the same object in the subsequent frames
2. **Appearance and disappearance —** Objects can move into or out of the frame unpredictably and we need to connect them to objects previously seen in the video
3. **Occlusion —** Objects are partially or completely occluded in some frames, as other objects appear in front of them and cover them up
4. **Identity switches —** When two objects cross each other, we need to discern which one is which

5. **Motion blur** — Objects may look different due to their own motion or camera motion
6. **View points**—Objects may look very different from different viewpoints, and we have to consistently identify the same object from all perspectives
7. **Scale change**—Objects in a video can change scale dramatically, due to camera zoom for example

## Object Tracking Algorithms:

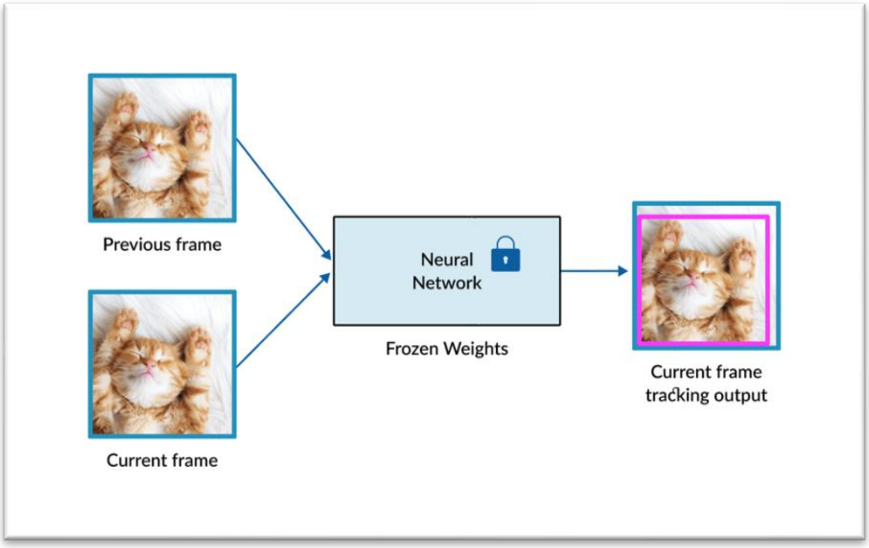
## 8.3 Simple Online and Real-Time Tracking (Sort):

Sort is an object tracking algorithm that relies mainly on the analysis of an underlying object detection engine. It can plug into any object detection algorithm. The algorithm tracks multiple objects in real time, associating the objects in each frame with those detected in previous frames using simple heuristics.



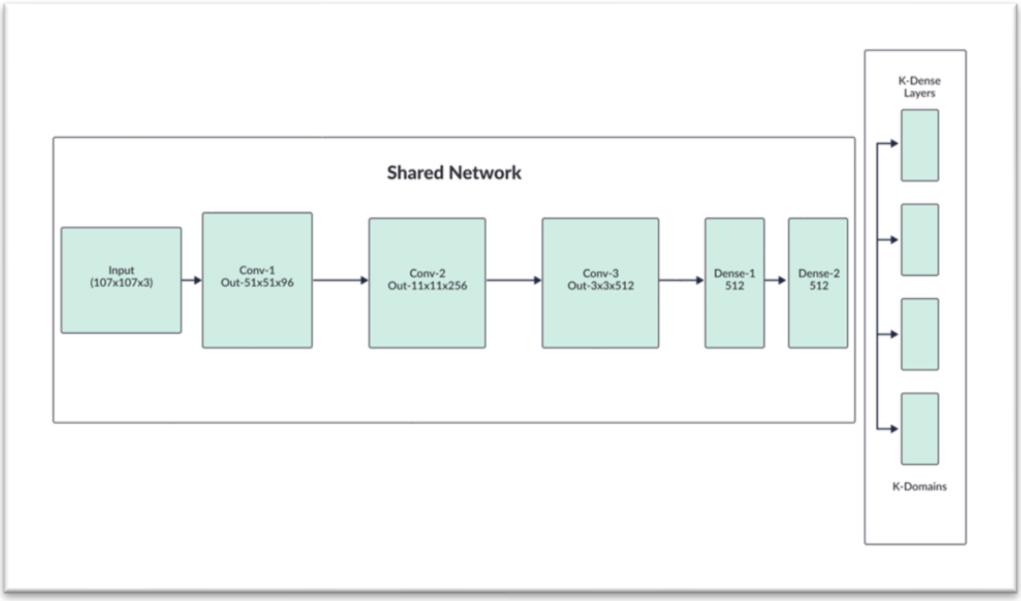
## 8.4 Generic Object Tracking Using Regression Network (Goturn):

Goturn is trained by comparing pairs of cropped frames from thousands of video sequences. When comparing two frames, in frame 1, the location of the object is known, and the frame is cropped to twice the size of the bounding box around the object, with the object centered. The algorithm then tries to predict the location of the same object in frame 2. The same double-sized bounding box is used to crop the second frame. A convolutional neural network is trained to predict the location of the bounding box in the second frame.



## 8.5 Multi-Domain Network (MDNet):

The objective of mdnet is to speed up training in order to provide real-time results. The strategy is to split the network into two parts. The first part acts as a generic feature extractor that trains over multiple training sets and learns to distinguish objects from their background. The second part is trained on a specific training set and learns to identify objects within video frames.So mdnet makes it possible to modify the weights of only the last few cnn layers during training, reducing computation time significantly.



# CHAPTER 9: RESULT ANALYSIS

As we can see in the above output video sequence, using the CNN, VGG and Object detection and object tracking algorithms, we trimmed the video into image segments and mapped them to the values which were determined in the confusion matrix, to classify the action into different categories.

# CHAPTER 10: APPLICATIONS AND FUTURE SCOPE

## 10.1 Applications:

**. Video Surveillance:**

object detection models are capable of tracking multiple people at once, in real-time, as they move through a given scene or across video frames. For any kind of stores to banks we can use this as security insights and safety.

**. Crowd Counting:**

Crowd counting is another valuable application of object detection. For densely populated areas like theme parks, malls, and city squares, object detection can help businesses and municipalities more effectively.

**. Self-driving cars:**

Real-time car detection models are key to the success of autonomous vehicle systems. These systems need to be able to identify, locate, and track objects around them in order to move through the world safely and efficiently. And while tasks like image segmentation can be applied to autonomous vehicles, object detection can be used to to this.

## 10.2 FUTURE SCOPE

As we discussed in the applications, we can enhance the current technologies and the various architectures we work on. 3D Convolutional Neural Networks can be tuned and used for better results and accuracy. There is a lot of future scope regarding this domain we can keep track of what kind abnormal activities are happening around ATM’s , Banks. We can also apply this to the Highly Classified areas like some the Military Installations or Black sites which are an issue of national defence. We can also implement this with various kinds of approaches like optical flow method, where we can indicate the persons in the frame of a video sequence given as input.

# CHAPTER 11: CONCLUSION

Through our project, we propose a deep learning model based on convolutional neural networks and VGG Networks with improved internal architecture. The proposed model has relatively fewer parameters and can effectively learn the features of violent behaviors. Experiment results on the datasets demonstrate the improvements of our model over methods. Also, the experiment on Mix dataset proves the effectiveness of our model on representation learning. At last, we evaluate the efficiency of deep learning models from theoretical and practical aspects. The proposed model is computing resource-saving, very efficient, and capable of real-time processing. For practical applications, strategies such as sliding window and voting can be adopted to achieve better recognition accuracies. By selecting appropriate sample rates, it is feasible to make trade-offs between efficiency and accuracy for violence detection tasks in various scenarios.

# CHAPTER 12: REFERENCES

[1] R. Poppe. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108(1-2):4–18, 2007.

[2] P. Lukowicz, J.A. Ward, H. Junker, M. Stager, G. Tr ¨ oster, A. Atrash, and T. Starner. Recognizing work- ¨ shop activity using body worn microphones and accelerometers. Proceedings of the 2nd Int Conference Pervasive Computing, pages 18–22, 2004.

[3] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Transactions on Information Technology in Biomedicine, 10(1):156–167, 2006.

[4] R. Nishkam, D. Nikhil, M. Preetham, and M.L. Littman. Activity recognition from accelerometer data. In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence, pages 1541–1546, 2005.

[5] L. Bao and S.S. Intille. Activity recognition from user-annotated acceleration data. In T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, M.Y. Vardi, and A. Ferscha, editors, Pervasive Computing, pages 1–17. 2004.

[6] J.R. Kwapisz, G.M. Weiss, and S.A. Moore. Activity recognition using cell phone accelerometers. SIGKDD Explorations Newsletter, 12(2):74–82, 2011.

[7] T. Brezmes, J.L. Gorricho, and J. Cotrina. Activity recognition from accelerometer data on a mobile phone. Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living, pages 796–799, 2009.

[8] Abhishek S. Mohite, Darshan K. Sangale, Prathamesh R. Oza, Tushar D. Parekar, Prof. Manisha P. Navale, Vol. 13, No. 3s, (2020), pp. 1515–1518

[9] Nouar AlDahoul , Aznul Qalid Md Sabri , and Ali Mohammed Mansoor, Volume 2018, Article ID 1639561, 14 pages.

[10] A Smart Home Monitoring System for Abnormal Human Activity Detection Using SVM Ms. Apurva Landge and Prof. Sandip Kahate IJCSMC, Vol. 4, Issue. 12, December 2015, pg.15 – 29.

[11] Detecting Unusual Activity in Video Hua Zhong, Jianbo Shi, and Mirko Visontai Volume 2, pages 819-826.

[12] Graph formulation of video activities for abnormal activity recognition Dinesh Singh, C. Krishna Mohan 2019.

[13] Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks Damla Arifoglu, Abdelhamid Bouchachia 2017.

[14] Untact Abnormal Heartbeat Wave Detection Using Non-Contact Sensor through Transfer Learning jin-soo kim , kangyoon lee 2018.

[15] Multi-Modal Anomaly Detection by Using Audio and Visual Cues Ata-ur rehman, hafiz sami ullah, haroon Farooq 2008.

DrillBit
Anti Plagiarism Software

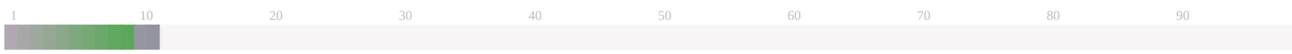
The Report is Generated by DrillBit Plagiarism Detection Software

## *Submission Information*

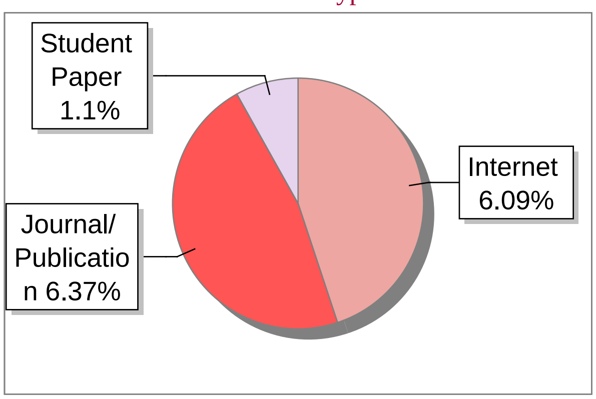
| | |
|---|---|
| Author Name | B.PRABHA |
| Title | Deep Learning Based Human Abnormal Activity Detection in Video |
| Paper/Submission ID | 413516 |
| Submission Date | 2021-11-22 13:07:12 |
| Total Pages | 50 |
| Document type | Project Work |

## *Result Information*

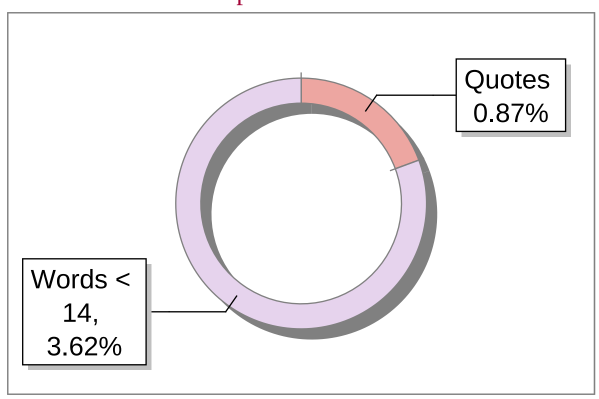
Similarity **12 %**




Sources Type




Report Content




***Alert..! Improper usage of 'Quotes' work in a report***

## *Exclude Information*

| | |
|---|---|
| References/Bibliography | Excluded |
| Quotes | Excluded |
| Sources: Less than 14 Words Similarity | Excluded |
| Excluded Source | **0 %** |

A Unique QR Code use to View/Download/Share Pdf File

DrillBit
Anti Plagiarism Software

**DrillBit Similarity Report**

| 12 | 19 | B | A-Satisfactory (0-10%)<br>B-Upgrade (11-40%)<br>C-Poor (41-60%)<br>D-Unacceptable (61-100%) |
|---|---|---|---|
| SIMILARITY % | MATCHED SOURCES | GRADE | |

| LOCATION | MATCHED DOMAIN | % | SOURCE TYPE |
|---|---|---|---|
| 1 | tryolabs.com | 2 | Internet Data |
| 2 | www.fritz.ai | 1 | Internet Data |
| 3 | PDF File Data www.cis.upenn.edu | 1 | Internet Data |
| 4 | CUSTOMER SEGMENTATION BY BATCH 167 YR-2020 | 1 | Student Paper |
| 5 | fr.scribd.com | 1 | Internet Data |
| 6 | Untact Abnormal Heartbeat Wave Detection Using Non-Contact Sensor through Transf by Kim-2020 | 1 | Publication |
| 7 | www.ijarset.com | 1 | Publication |
| 8 | Activity identification using body-mounted sensorsa review of class by Preece-2009 | <1 | Publication |
| 9 | arxiv.org | <1 | Publication |
| 10 | Detecting indicators of cognitive impairment via Graph Convolutional Networks by Arifoglu-2020 | <1 | Publication |
| 11 | Real-Time Human Detection for Aerial Captured Video Sequences via Deep by AlDahoul-2018 | <1 | Publication |
| 12 | www.iaeme.com | <1 | Publication |

| | | | |
|---|---|---|---|
| 13 | Numerical simulation of interfacial coupling phenomena in two-phase flow through by O-2004 | <1 | Publication |
| 14 | Insight about Detection, Prediction and Weather Impact of Coronavirus (Covid-19) by Haque-2020 | <1 | Publication |
| 15 | d-nb.info | <1 | Publication |
| 16 | PDF File Data www.truststc.org | <1 | Internet Data |
| 17 | downloads.hindawi.com | <1 | Publication |
| 18 | moam.info | <1 | Internet Data |
| 19 | Keeping Host Sanity for Security of the SCADA Systems by Lee-2020 | <1 | Publication |