Winter Internship Project Report

on

Abnormal Human Behaviour Detection in Restricted Environment

LeadingIndia.ai

Submitted To:	Submitted By: Team Number – 19
Dr. Kanad Kishore Biswas, Professor	Ashokmani S, Ramco Institute of Technology
Mr. Shreyans Sharad Jain, M. Tech scholar	Balakumaresan S, Ramco Institute of Technology
	Harsh Kataria, Bennett University
	Sparsha Mishra, KIIT University

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
SCHOOL OF ENGINEERING AND APPLIED SCIENCES
BENNETT UNIVERSITY
GREATER NOIDA
UTTAR PRADESH

BONAFIDE CERTIFICATE

This is to certify the project entitled report **"ABNORMAL HUMAN BEHAVIOUR DETECTION IN RESTRICTED ENVIRONMENT"** submitted is a bonafide work of **ASHOKMANI S, BALAKUMARESAN S, HARSH KATARIA, SPARSHA MISHRA** who carried out the project under our supervision at Department of Computer Science Engineering, School of Applied Science and Engineering, Bennett University, Greater Noida from 20 December 2018 to 18 January 2019.

SIGNATURE
Dr. Kanad Kishore Biswas
Professor (Mentor)
Department of Computer Science Engineering
School of Applied Science and Engineering
Bennett University, Greater Noida
Uttar Pradesh

SIGNATURE
Mr. Shreyans Sharad Jain
M. Tech Scholar (Mentor)
Department of Computer Science Engineering
School of Applied Science and Engineering
Bennett University, Greater Noida
Uttar Pradesh

ACKNOWLEDGEMENT

I would like to thank respected principal **Dr. N.Jawahar**, Ramco Institute of Technology, Rajapalayam, for his kind words and enthusiastic motivation which is inspired me a lot in completing this project.

I wish to express sincere thanks to **Dr. S.Rajakarunakaran**, Professor and Head, Department of Mechanical Engineering, Ramco Institute of Technology, Rajapalayam, for his support and encouragement for the completion of the project.

I wish to express my deep sense of gratitude to **Dr. Kanad Kishore Biswas**, Professor, Department of Computer Science Engineering, School of Applied Science and Engineering, Bennett University, Greater Noida, Uttar Pradesh, for his support during the course of my training.

I am very much thanked to respectable mentor **Mr. Shreyans Sharad Jain,** M. Tech Scholar, Department of Computer Science Engineering, School of Applied Science and Engineering, Bennett University, Greater Noida, Uttar Pradesh, who gave me timely advices, sufficient technical and moral support throughout the project tenure.

I also extend my gratitude to the Team members for their cooperation and valuable assistance during the internship training programme.

My special thanks to Department of Computer Science Engineering, School of Applied Science and Engineering, Bennett University, Greater Noida, Uttar Pradesh, for their help in coordinating the internship.

ABSTRACT

Abnormal detection refers to infrequent data instances that come from a diverse cluster or distribution than the majority normal instances. Owing to the increasing demand for safety and security, discovery abnormalities from video streams has attracted significant research interest during recent years. The current advancements in computer vision and deep learning have a remarkable role in enabling such intelligent frameworks. Different algorithms that are specially designed for building smart vision frameworks seek to scene understanding and building correct semantic inference from observed dynamic motions caused by moving targets. In this project the used models is Fast R-CNN for human detection and SSD for pose detection. Unfortunately, although there are many algorithms have been proposed in this interesting topic, the research in this area still lacks strongly to two important things: comparative general assessment and public-accessible datasets. This project is mainly focus on the detection of abnormal human behaviour in restricted environment.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 DETECTION OF HUMAN BEHAVIOUR USING DEEP LEARNING:

There has been a recent discussion to monitor human behaviors and activity patterns for significant research purposes. Deep Learning and human action are a complex, diverse and also a challenging area which has received a lot of attention in the past few years. The main interest was in the discovery of abnormal actions and not learning the different types of activities. It is a critical task to instantly detect abnormal behavior in real-time.

Computer vision is an interdisciplinary field that has been gaining huge amounts of attraction in the recent years (since CNN) and self-driving cars have taken the center stage. Another integral part of computer vision is object detection. Object detection aids in pose estimation, vehicle detection, surveillance etc. The difference between object detection algorithms and classification algorithms is that in detection algorithms, we try to draw a bounding box around the object of interest to locate it within the image. Also, you might not necessarily draw just one bounding box in an object detection case, there could be many bounding boxes representing different objects of interest within the image and you would not know how many beforehand.

The main reason to why it cannot proceed with this problem by building a standard convolutional network followed by a fully connected layer is that, the length of the output layer is variable and not constant, this is because the number of occurrences of the objects of interest is not fixed. A naive approach to solve this problem would be to take different regions of interest from the image, and use a CNN to classify the presence of the object within that region. The problem with this approach is that the objects of interest might have different spatial locations within the image and different aspect ratios. Hence, you would have to select a huge number of regions and this could computationally blow up. Therefore, algorithms like R-CNN, YOLO etc. have been developed to find these occurrences and find them fast.

The architecture used is Fast R-CNN for object detection and SSD for human skeleton detection (Pose Detection).

1.2 PROBLEM STATEMENT:

The project is about detecting abnormal behavior done by people in an enclosed environment like offices, colleges, etc. using video clips. The project comprises of following tasks:

- Identifying classes of abnormality.
- Converting the videos to frames
- Converting each frame to xml files
- Detecting abnormal from the dataset.

1.3 OBJECTIVE:

The objective of the project is to be able to detect the abnormal behavior in a restricted environment and shows what action are taking place.

- To detect object by using Fast R-CNN and to detect human skeleton by using SSD in the video based on Tensorflow Objection Detection API. In which 10 classes have been selected for action detection, where each video was converted into frames and then to xml files.
- To collect the dataset from the HMDB database - a large human motion database which consisted of 51 action classes such as clap, wave, punch, hug, drink, sit, kiss, eat, stand, handstand etc., for our project.

1.4 MOTIVATION OF THE PROJECT:

The main inspiration behind this project is to detect the abnormal human behavior in restricted environment because there is an increasing desire and need in video surveillance applications for a proposed solution to be able to analyze human behaviors and identify subjects for standoff threat analysis and determination. The main purpose of this project is to look at current developments and capabilities of visual surveillance systems and assess the feasibility and challenges of using a visual surveillance system to automatically detect abnormal behavior, detect hostile intent, and identify human subject.

CHAPTER II

LITERATURE SURVEY

2.1 Oluwatoyin P. Popoola and Kejun Wang in [1] have presented a paper that extends previous related surveys for building intelligent vision systems which aim at scene understanding and making the correct semantic inference from the observed dynamics of the moving targets. The paper presented by them focuses on contextual abnormal human behaviour detection especially in video surveillance applications. The main purpose of the survey was to extensively identify existing methods and characterize the literature in a manner that brings key challenges to attention. The definition anomaly can have some degree of ambiguity within a domain of application. Visual behaviours are complex and have much variety in an unconstrained environment. The influence of noisy data, the choice and representation of low-level features, significantly influences the discriminative power of the classifier. Video quality, shadows, occlusion, illumination, moving camera, and complex backgrounds are challenges especially with a single-camera view. “Abnormal” activity can be maliciously adapted to appear as “normal” by the human agents. Thus, the spatiotemporal variations for the same activity can be very high even when performed by the same individual. These and many more make anomaly detection a difficult task and therefore have no “all-purpose” algorithm that works well for different contexts. Thus, the approach was to build computational action models into machines (via training/learning) to automatically determine whether a newly observed behaviour is normal or not.

2.2 Tian Wang, Jie Chen, and Hichem Snoussi in [2] have proposed an algorithm which handles the problem of detecting abnormal events. The algorithm consists of an image descriptor and online nonlinear classification method. The authors have introduced a covariance matrix of optical flow and image intensity as a descriptor encoding moving information. The nonlinear online support vector machine (SVM) firstly learns a limited set of the training frames to provide a basic reference model then updates the model and detects abnormal events in the current frame. The final task was to apply the method to detect the abnormal events on a benchmark video surveillance dataset in order to demonstrate the effectiveness of the proposed technique. The method was developed into two nonlinear one-class SVM based abnormal behaviour detection technique which is an update to the normal models of the surveillance video data in an online framework. The proposed algorithm was tested on a video dataset which yielded successful results to detect the abnormal events.

2.3 MD. Ashik Ahmed, Mushfique Ahmed Isha and Al-Amin Ahmed in [3] have made a thesis report on Dynamic image analysis for abnormal behaviour detection. Detecting abnormal behaviour has become an important area of research in computer vision which is also driven by other wide application domains. In this research, Convolutional Neural Network was used to make the process easier and efficient where the abnormal behaviour was prompted by an individual person. The proposed system detected the behaviour of the individuals in the normal scenario with a successful accuracy of 98%. It also detected any deviations from the previous data from any new scenario from different dynamic images. This system can be implemented for advance security purposes.

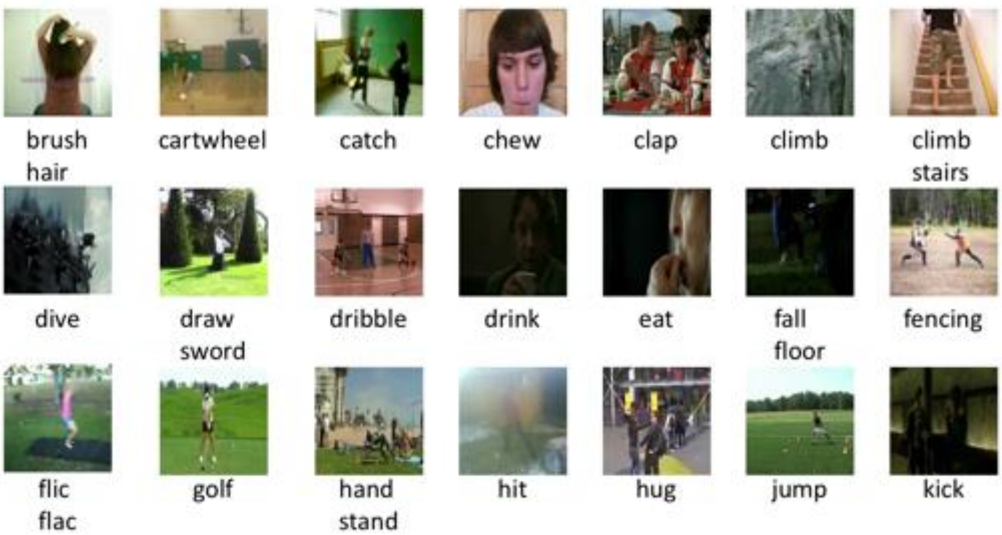
CHAPTER III

PROPOSED METHODOLOGY

3.1 DATASET:

For the proposed project, the dataset have been selected from the HMDB Dataset [7]. With nearly one billion online videos viewed every day, an emerging new frontier in computer vision research is recognition and search in video. The HMDB dataset is collected from various sources, mostly from movies, and a small proportion from public databases such as the Prelinger archive, YouTube and Google videos. The dataset contains 6849 clips divided into 51 action categories, each containing a minimum of 101 clips. The actions categories can be grouped in five types:

- General facial actions smile, laugh, chew, talk.
- Facial actions with object manipulation: smoke, eat, drink.
- General body movements: cartwheel, clap hands, climb, climb stairs, dive, fall on the floor, backhand flip, handstand, jump, pull up, push up, run, sit down, sit up, somersault, stand up, turn, walk, wave.
- Body movements with object interaction: brush hair, catch, draw sword, dribble, golf, hit something, kick ball, pick, pour, push something, ride bike, ride horse, shoot ball, shoot bow, shoot gun, swing baseball bat, sword exercise, throw.
- Body movements for human interaction: fencing, hug, kick someone, kiss, punch, shake hands, sword fight.



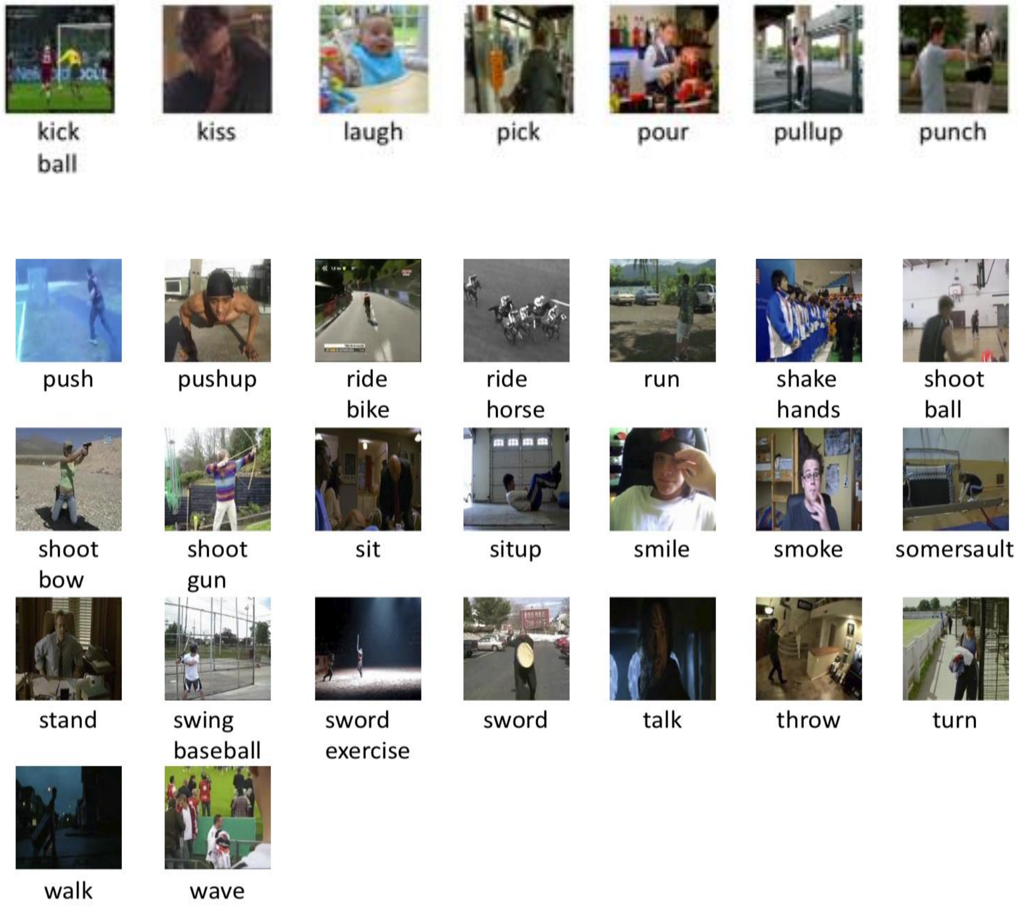


Figure 3.1 (a) Illustration of 51 actions

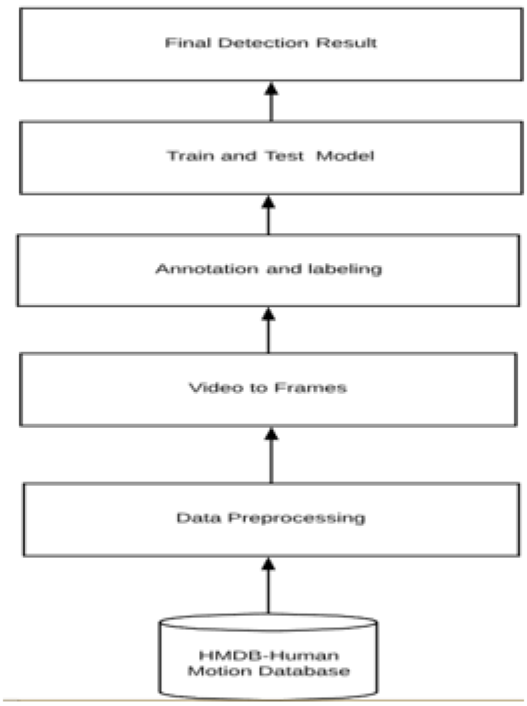


Figure 3.1 (b) Project Design Flow Diagram

3.2 FUNCTIONALITY OF THE PROJECT:

The project is based on and build using the tensorflow-gpu library and its dependencies, the major functionalities that are mentioned and depicted are object detection, human detection, pose detection, action detection, visualisers (bounding box builders, pose builders, class or action detectors and confidence counters), .xml to .csv file conversions, data pre-processors, data trainers and testers. When the project is successfully compiled and is running we get output in the form of different colored bounding boxes containing the humans and colored skeleton outlay of the pose of the human on the human. The white color bounding box identifies a normal action being performed by the human and blue boxes identifies an abnormal action.

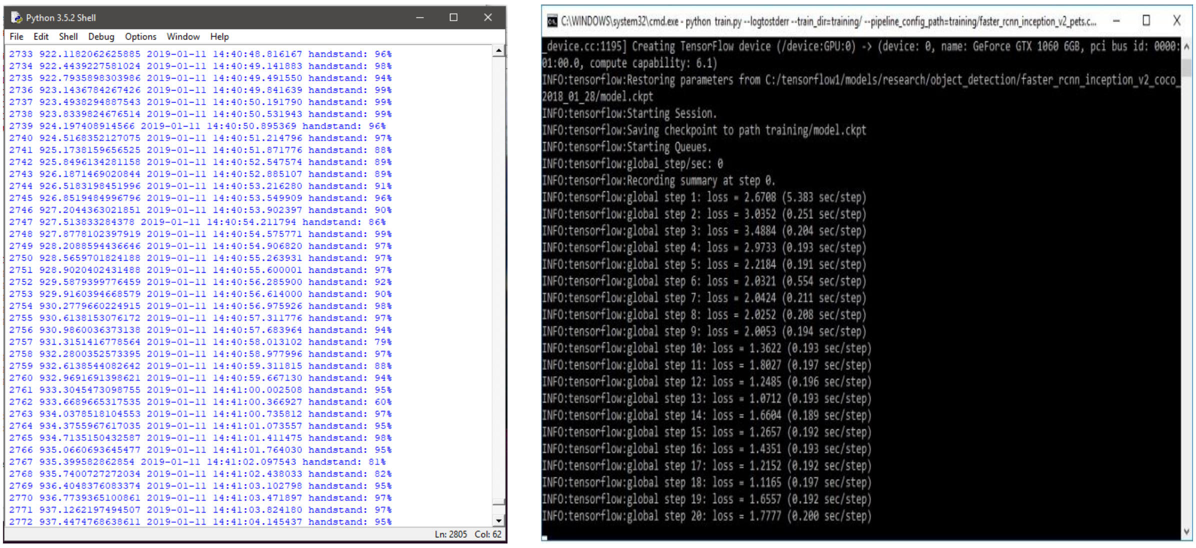
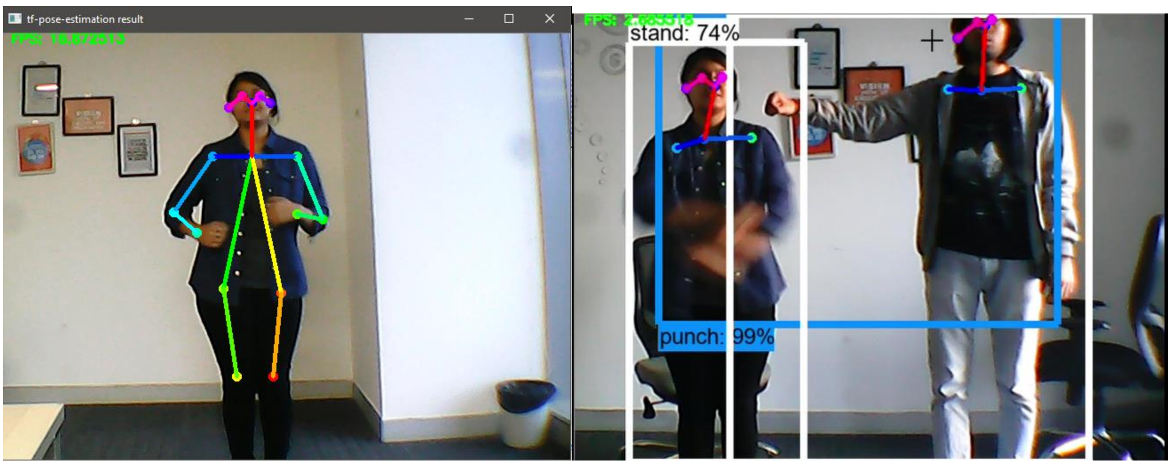
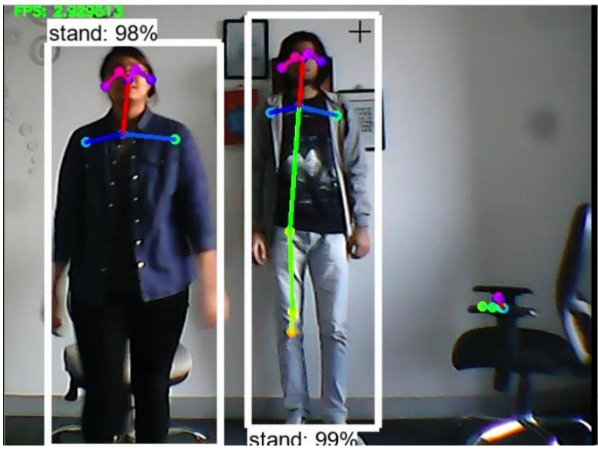
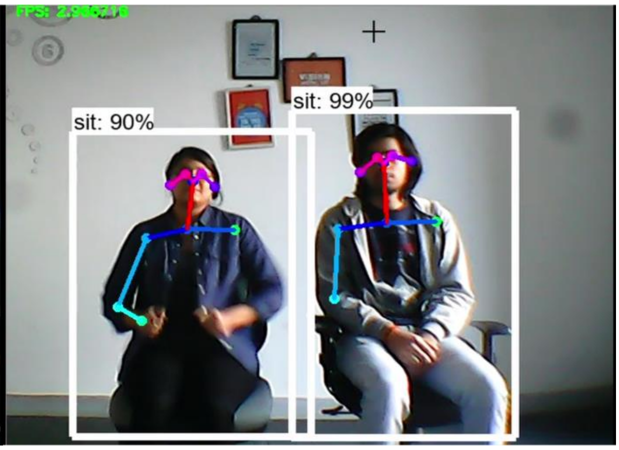


Figure 3.2 Working and Training

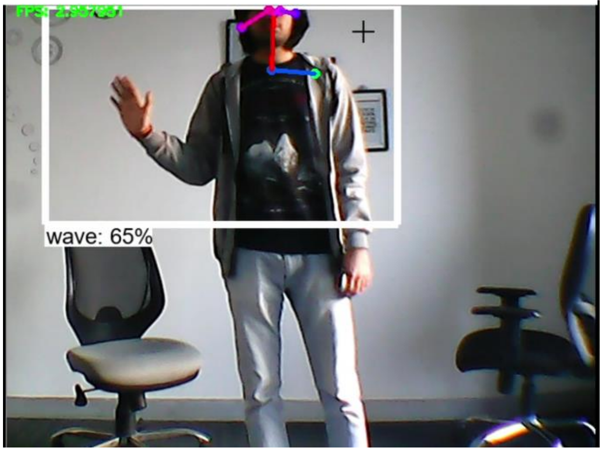


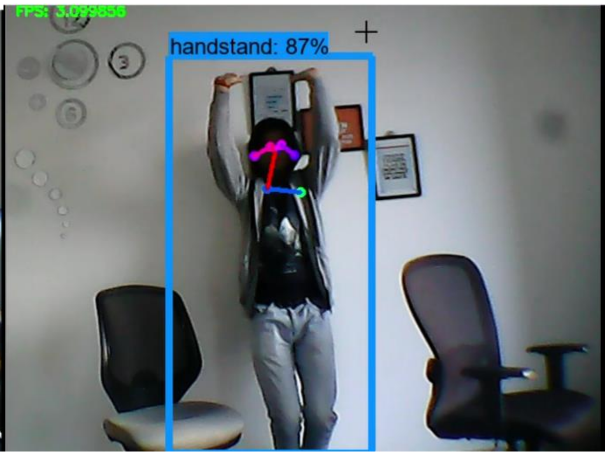
Pose detection ***Action: Punch (Abnormal)***





Action: Stand (Normal) *Action: Sit (Normal)*





Action: Wave (Normal) *Action: Handstand (Abnormal)*

3.3 TESTING PROCEDURE:

For testing purposes there are three different scripts each with their own functions to work for image, video or webcam(realtime). To be able to test the detector just select and load the model, run either of the files in your favourite IDE or using the CLI as

>python ./Object_detection_image.py


```
  <?xml version="1.0"?>
- <annotation>
     <folder>kiss train & test</folder>
     <filename>179.jpg</filename>
     <path>I:/AI/segregated datasets with videos/kiss train & test/179.jpg</path>
   - <source>
        <database>Unknown</database>
     </source>
   - <size>
        <width>960</width>
        <height>540</height>
        <depth>3</depth>
     </size>
     <segmented>0</segmented>
   - <object>
        <name>kiss</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
      - <bndbox>
           <xmin>378</xmin>
           <ymin>26</ymin>
           <xmax>884</xmax>
           <ymax>485</ymax>
        </bndbox>
     </object>
  </annotation>
```


	A	B	C	D	E	F	G	H	I	J
204	as307.jpg	960	540	sit	104	75	447	505		
205	as309.jpg	960	540	sit	498	168	824	539		
206	as31.jpg	960	540	stand	323	3	764	531		
207	as311.jpg	960	540	sit	410	254	849	540		
208	as313.jpg	960	540	sit	314	154	622	540		
209	as315.jpg	960	540	sit	256	133	607	540		
210	as317.jpg	960	540	sit	362	89	665	537		
211	as319.jpg	960	540	sit	163	54	486	540		
212	as319.jpg	960	540	sit	313	226	564	512		
213	as321.jpg	960	540	sit	89	113	449	532		
214	as321.jpg	960	540	sit	328	104	594	540		
215	as323.jpg	960	540	sit	481	39	906	539		
216	as325.jpg	960	540	sit	418	35	858	540		
217	as327.jpg	960	540	sit	53	100	321	512		
218	as329.jpg	960	540	sit	137	136	343	505		
219	as33.jpg	960	540	kiss	114	1	784	511		
220	as331.jpg	960	540	stand	592	83	860	324		
221	as333.jpg	960	540	stand	612	70	837	311		
222	as335.jpg	960	540	stand	753	10	960	461		
223	as37.jpg	960	540	sit	132	40	534	540		
224	as339.jpg	960	540	sit	141	9	480	540		
225	as341.jpg	960	540	stand	295	70	746	540		
226	as343.jpg	960	540	sit	189	70	508	532		
227	as343.jpg	960	540	sit	503	116	811	540		
228	as345.jpg	960	540	sit	183	136	500	536		
229	as345.jpg	960	540	sit	502	149	797	540		
230	as347.jpg	960	540	stand	321	58	705	540		
231	as349.jpg	960	540	sit	78	43	339	540		
232	as349.jpg	960	540	sit	636	21	941	533		
233	as349.jpg	960	540	sit	222	11	739	539		
234	as35.jpg	960	540	stand	513	50	866	531		
235	as351.jpg	960	540	sit	78	52	390	540		
236	as351.jpg	960	540	sit	279	26	718	540		
237	as351.jpg	960	540	sit	626	44	960	536		
238	as353.jpg	960	540	stand	313	38	773	540		
239	as355.jpg	960	540	sit	441	91	735	471		
240	as357.jpg	960	540	sit	385	158	742	520		

train_labels

Figure 3.3 Testing procedure

CHAPTER IV

IMPLEMENTATION DETAILS

4.1 COLLECTION OF DATASET:

Collected the video datasets (sources) from HMDB database from the link: http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/. From the HMDB database, 10 actions (normal and abnormal classes) was selected among 51 action categories. As the dataset is built on movie clips, only less classes was selected which are associated to indoor environments and appropriate for enclosed abnormal behavior detection.

4.2 DATA PRE-PROCESSING:

Data preprocessing is a technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing. Data preprocessing is used database-driven applications such as customer relationship management and rule-based applications (like neural networks).

4.3 CREATION OF MODEL:

Fast R-CNN model was used for object detection and SSD MobileNet for pose detection and load the dataset.

4.3(a) FAST R-CNN:

Fast R-CNN has two networks: region proposal network (RPN) for generating region p roposals and a network using these proposals to detect objects. The main different here with Fast R-CNN is that the later uses selective search to generate region proposals. R-CNNs for Object Detection were first presented in 2014 by Ross Girshick et al [4], and were shown to outperform previous state-of-the-art approaches on one of the major object recognition challenges in the field: Pascal VOC. Since then, two follow-up papers were published which contain significant speed improvements: Fast R-CNN [5] and Faster R-CNN [6]. The basic idea of R-CNN is to take a deep Neural Network which was originally trained for image classification using millions of annotated images and modify it for the purpose of object detection.

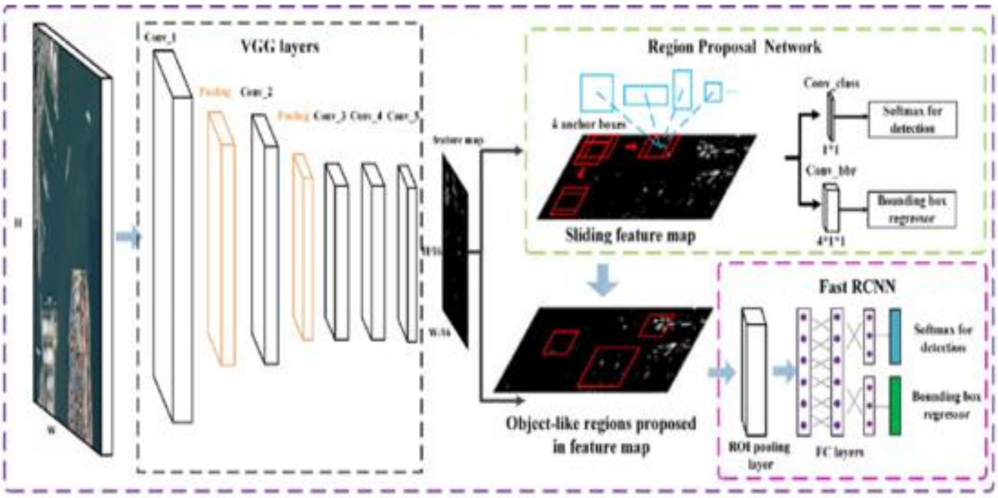


Figure 4.3(a) Fast R-CNN Architecture

4.3(b) SSD MOBILENET:

SSD, discretized the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. The fundamental improvement in speed comes from eliminating bounding box proposals and the subsequent pixel or feature resampling stage.

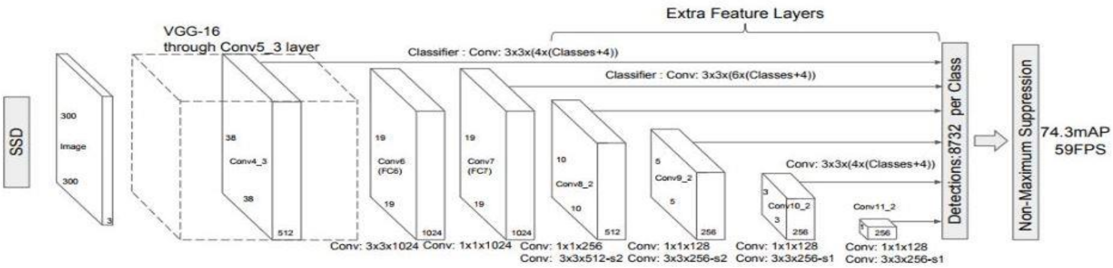


Figure 4.3(b) SSD Architecture

Single Shot: This means that the tasks of object localization and classification are done in a ***single forward pass*** of the network

MultiBox : This is the name of a technique for bounding box regression developed by Szegedy et al.

Detector : The network is an object detector that also classifies those detected objects

4.4 TRAINING DETAILS:

Once the records files are ready, it is almost ready to train the model.

- Decide the pre-trained model to be used. There's a trade off between detection speed and accuracy, higher the speed lower the accuracy and vice versa. The used model is ssd_mobilenet_v1_coco (Pose detection) and faster_rcnn_inception_v2_coco (Object detection) for demonstration purpose.
- After deciding the model to be used download the config file for the same model. In this case, ssd_mobilenet_v1_coco.config and faster_rcnn_inception_v2_coco.config was downloaded.
- Make a new file *object-detection.pbtxt* where the given class name was stored. If in case, it has multiple classes, increase id number starting from 1 and give appropriate class name.
- Change the number of classes in the file according to our requirement.
- Need to give path to both train.record file and test.record file and this two path .record file should be mentioned in the given config file. Then run the train.py

4.5 TESTING OF THE MODEL:

- Inference graph will be generated after the completion of training.
- Check the accuracy and result by feeding the input video frames.

4.6 PROJECT DEVELOPMENT TIME SCHEDULE:

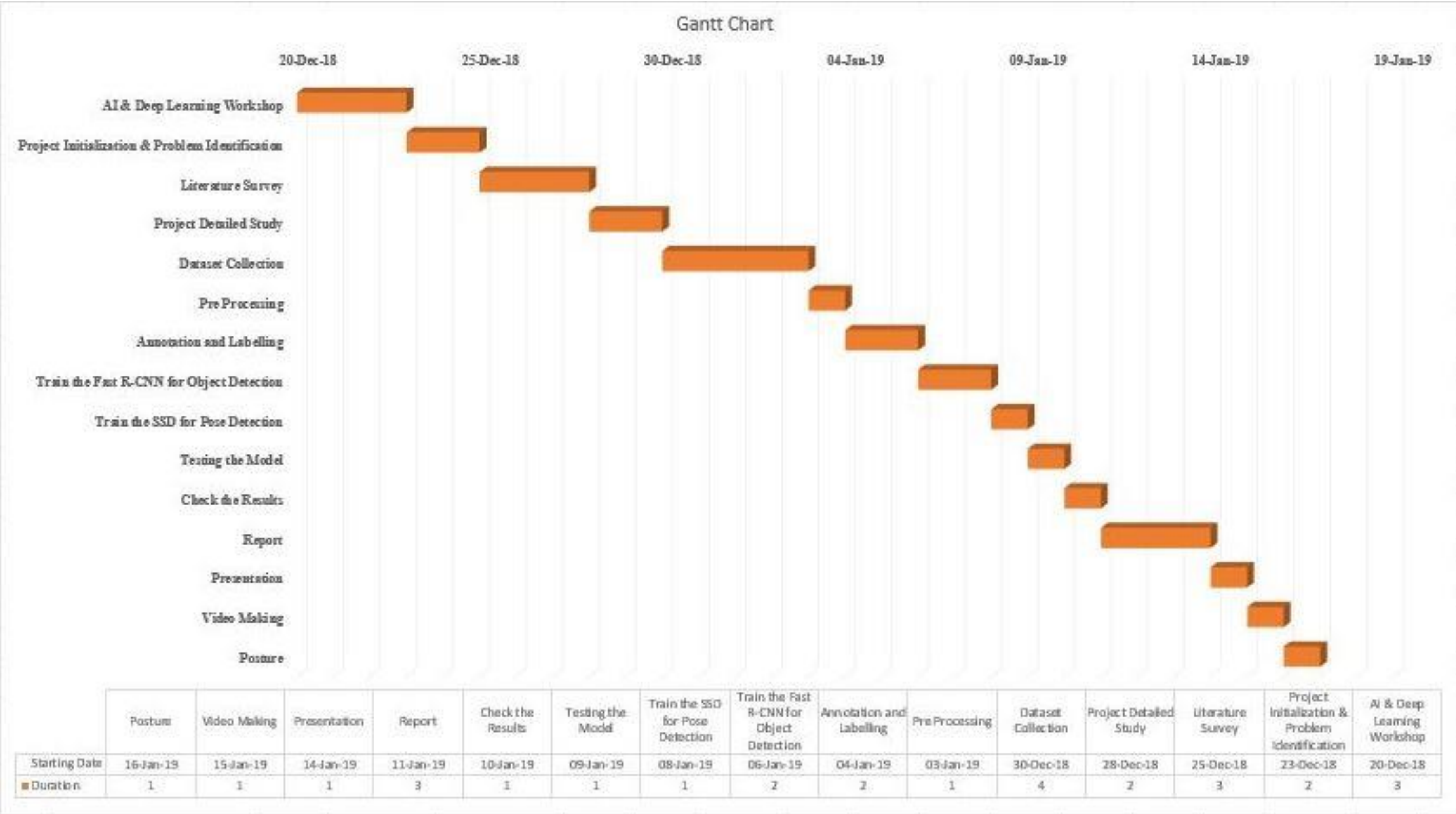


Figure 4.6 Gantt chart

CHAPTER V

LEARNING AND REFLECTION FROM THE PROJECT

We learned about the different version of anaconda and python packages and their installation procedure, how to setup a virtual environment in anaconda, came to know how to collect the required dataset according to the problem statement from the different database such HMDB etc, data pre-processing, applications of the software LabelImg and Annotation of video frames, how to use some pre-trained model for object detection and pose detection, working of Tensorflow API, Functionality of keras, open-cv, matplotlib, images,protobuf, how to train and test the model and compare the accuracy with other given models.

CHAPTER VI

LIMITATIONS AND FUTURE ENHANCEMENTS

In this project, limited number of video frames was only trained, so model was trained to certain extend and it may confuse at some human actions (classes). And the major disadvantages which is common in most frameworks is that the model may confuse if the video frames and the structure of the frames are not clear and legible. In this project, the model was only trained for 10 classes (Normal and Abnormal). The model accuracy may decreases if the abnormal behavior happens in the crowded place.

The proposed methodology can be improved by training the model on more video frames to improve the abnormal behaviour detection which then can be used in systems for future work.

CHAPTER VII

CONCLUSION

Abnormal behaviour detection through behaviour analysis is still underway. However, when traffic volume increases in a large network environment, behaviour detection becomes time consuming and increases resource inefficiency. In this study, to detect abnormal behaviour, the analysed videos get converted into single frames. The tested model is used in real-time to detect the person's action and pose. The model detects the abnormal behaviour of the subject in the frame and creates a bounding box around the same which says which action is being performed and the colour of the frame indicates whether the action is normal or abnormal that is blue indicating it is abnormal and white indicating its normal. The performance evaluation of the proposed algorithm is analysed with on real-time.

CHAPTER VIII

REFERENCES

[1] Popoola, Oluwatoyin P., and Kejun Wang. "Video-based abnormal human behavior recognition—A review." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, no. 6 (2012): 865-878.

[2] Wang, Tian, Jie Chen, and Hichem Snoussi. "Online detection of abnormal events in video streams." *Journal of Electrical and Computer Engineering* 2013 (2013): 20.

[3] Ahmed, Md, Mushfique Ahmed Isha, and Al-Amin Ahmed. "Dynamic image analysis for abnormal behavior detection." PhD diss., BRAC University, 2017.

[4] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587. 2014.

[5] Girshick, Ross. "Fast r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 1440-1448. 2015.

[6] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." In *Advances in neural information processing systems*, pp. 91-99. 2015.

[7] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A Large Video Database for Human Motion Recognition. ICCV, 2011.