DETECTING CRIME ACTIVITIES BASED ON HUMAN BEHAVIOR USING MACHINE LEARNING

BY

MD. NAJMUL HUDA HRIDOY
ID: 152-15-513

MD. AZIZUL ISLAM
ID: 171-15-1380

MOHAMMAD ATIQUR RAHMAN
ID: 161-15-6983

AND

MD. SAROWAR ZAHAN
ID: 171-15-1221

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Dr. S. M. Aminul Haque

Associate Professor

Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
FEBRUARY 2023

APPROVAL

This Project/internship titled **"Detecting crime activities based on human behavior using machine learning"**, submitted by Md. Najmul Huda Hridoy, ID: 152-15-513, Md. Azizul Islam, ID: 171- 15-1380, Mohammad Atiqur Rahman, ID: 161-15-6983 and Md. Sarowar Zahan, ID: 171-15-1221 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 04-02-2023.

BOARD OF EXAMINERS

7/2/23

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. S. M. Aminul Haque
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dewan Mamun Raza
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

4/2/2023

Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of Dr. S. M. Aminul Haque, Associate Professor, **Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. S. M. Aminul Haque

Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Md. Najmul Huda Hridoy
ID: 152-15-513
Department of CSE
Daffodil International University

Md. Azizul Islam
ID: 171-15-1380
Department of CSE
Daffodil International University

Mohammad Atiqur Rahman
ID: 161-15-6983
Department of CSE
Daffodil International University

Md. Sarowar Zahan
ID: 171-15-1221
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to Dr. S. M. Aminul Haque Associate Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the Supervisor, co- Supervisor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Object detection is a key concept in optical-based video surveillance software/model/device and automated security system for analyzing/detecting/ identifying crime scene evidence. Photos and video footage can have an important role in the detection of criminal activity. As a massive amount of photos and video footage are collected as visual documentation or evidence of a crime scene the investigation process can be highly complex and may need an advanced technological process. In this research, we have developed a YOLO (You only look once) CNN (Convolution neural network) based real-time object detection model which can automatically detect criminal activities without human instruction. Our model can detect human movements and poses from an image and video footage and classify them into 5 different classes. By analyzing the human poses and movements our goal was to detect and classify the dangerous and suspicious human movements from a crime scene. We trained our model with 1300+ custom images and 5 classes of objects on the google colab with free GPU. We gain an average accuracy of 89% at 0.013 confidence thresholds after training on google colab.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction: In the field of digital surveillance systems [15], the process of detecting an object from a sequence of videos and photos plays a vital role. For the digital investigation of a crime, visual documentation such as video and images significantly help a Police Officer for detecting Fingerprints and identify criminals and realize the Crime scene. It is so complex to control and categorize a large amount of visual data.

In the investigation of a crime, some common questions will be: How can I find the real criminal and exact criminal activities inside from the huge amount of videos and images or visual documentation? The answer will be an object detection [16] model that can respond in real-time. In this task, we have developed an object detection model that can successfully detect and trace a pose of a moving person from a video sequence. By doing this we can clearly realize the human attitude. This task is going to be very complex because it merges some research fields like image processing, machine learning, human-computer interaction, pattern recognition etc. The object detection model is widely used in intelligent video surveillance systems, human-computer interaction, action analysis etc.

In the sector of intelligent video surveillance, when criminal activity happens the traditional Surveillance system can't play any active role in real-time. Because it just records the video scene. If the police want to understand the exact incident and find the real culprit, they will search the recorded videos. But it may be too late to arrest the real criminal. Our object detection model will totally change the tradition of surveillance systems. It can perfectly understand the environment and it also can detect human movements and identify criminal activities based on human movements and poses[17].

In the process of human-computer interaction[14], it is so tough to realize an environment based on the human voice. Because it could be so noisy. But it is easy to realize and detect an environment based on human movements and poses.

The increasing development of Technology such as updated mobile phones, self-driving, vehicles autopilot planes, and other embedded systems has been invested in the field of computer vision. Such as in the field of self-driving cars, it is required to parse an environment, able to identify an object with its movements, and should take an effective step based on its realization. In the field of Artificial Intelligence, a developing device that is refined with AI needs to correctly locate an object's movement and position in the given image or video sequence. The self-driving car system mainly depends on object detection. For system needs to detect an object accurately and be able to understand an environment for safe driving. All in all, object detection might be the most useful and effective 1 that is obtained core functional algorithm for detecting and realizing an environment in imported applications.

Objective action is an unprecedented part of the field of computer vision. The traditional object detection medal that is developed by the manual features has been removed day by day due to they are low accuracy and bad environmental adaptability with the updated Convolutional neural network(CNN) based on the object detection model. Due to the increase of the powerful computer system and updated deep learning chips and to the collection of huge scale labeled data such as (Imagenet, MS COCO) Convolutional neural network(CNN) has been widely researching because of its first and accurate learning framework.

Some object detection algorithms best on Region personal networks(RPN) like R-CNN, Fast R-CNN, Faster R-CNN and mask R-CNN are updated on a regular basis with object detection accuracy. But all those methods are depends on powerful GPU computing system because of maintaining their accuracy. So this computational cost will be very high to face the challenge. YOLOv3 maintains a low computational cost and provide the high accurate result . To overcome the challenge YOLO model will be the best option for detecting an object.

In recent time, many single-state objective detection have been developed such as SSD, DSSD, YOLO series method, RetinaNet, etc. Yolov3 might be the most effective and usable object detection model for real-time object detection applications.

In this paper, we discuss an object detection model that is trained on YOLO (You only look once) [17] with 1300+ visual data sets. All the data sets are unique and created by us manually. We have selected 5 different dangerous human poses as a class is commonly seen in a crime scene. The aim of our model is to reduce the damages caused by the crime and provide a real-time response.

The following figure 1.1.1 shows the basic procedure of YOLO for detecting an object.

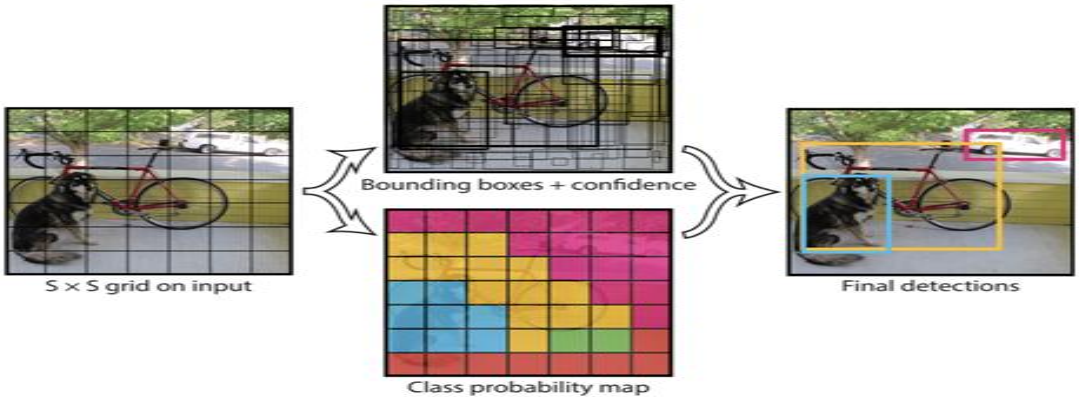

Figure 1.1.1: Basic procedures of Object detection

1.2 Objective : We are constantly victims of various types of criminal activities. Most of these activities are sudden and unexpected. But from the human attitude and movements, we can infer that something can happen. If we take proper action before something happens, we can prevent most of these crimes. From this thinking, we are trying to develop a system that can detect criminal activities based on human attitude. Our Algorithm will be able to predict any crime activities by analyzing human attitudes and movements. If it realizes that something bad is going to happen then it will take an action based on its analysis.

- To Reduce Harassment.
- To detect Terrorism Activities.
- To Prevent Hijack & Robbery.

- To detect any Abnormal or Crime Activities.

1.3 Motivation: we see many criminal activities happen in our daily life. Like hijacks, sexual harassment, bank robbery, pick pocketing, stick fighting, etc. Many of these criminal activities happen to us. Many innocent people are victims of those crimes. Some of these activities are sudden and unexpected. But many of those are per-planned. Often there is no plan to prevent these criminal activities. Since most of those activities are sudden, they cannot be avoided even if there is a pre-planned prevention. As a result, we faced a lot of damage and loss of life.

After any criminal activity is reported to the police, the police come and start the investigation. Most of the time, the police do not have the necessary evidence to investigate these crimes. As a result, the police have to face a lot of trouble to find the real culprit. Finding the real culprit by analyzing traditional CCTV footage is very difficult and time-consuming. Because traditional CCTV cameras can only record video, but they cannot understand and realize any environment.

But if we can develop a system that will be available everywhere like a CCTV camera, but that will not only record the video but also be able to instantly understand what is happening in the video and will respond in real-time, then we can prevent and reduce the damage of these crimes to a great extent. As a result, we will be able to detect the real culprit immediately. This will be very helpful to the police and save effort and time.

We are facing different kinds of problems in our daily life, such as: Harassment., Terrorism, Hijack/Rubbery, Kidnap and Abnormal Activities. We want to Resist all of those illegal Activities by applying our research.

From this thought, we developed our project which can understand and realize an environment by analyzing human movements. It will detect a crime in real-time and provide real-time responses.

1.4 Rational of the studies: We have completely focused on human movement in this project. We wanted to research what people's common movements at a crime scene look like. For example, we can see on figure 1.3.1 (left), how a person might act in a fight scene and how the movements are. During a bank robbery, we wanted to see how the robber gang and the people there might behave. We have tried to observe these activities closely. We emphasized the difference in human behavior in these different crime scenes. If human poses and movements in certain environments can be properly analyzed, the situation can be understood by looking at the movements of people. So we have designed our project based on this human movement analysis.

The following figure 1.4.1 shows how people behave in fight scenes (left) and how people behave during a bank robbery (right).

Figure 1.4.1: fighting scene (Left), and robbery scene (right).

1.5 Expected output: We develop our model based on human behavior and movements by applying the YOLOv3 algorithm[18]. Our developed model can successfully detect crime in any public place in real-time. It will mainly observe people's movements from a scene. By analyzing these movements, if it can detect a suspicious movement that can only be seen at a crime scene, the model will understand that a crime is taking place. This model will be able to detect criminal activity instantly and it will respond real-time by sending SMS and signals to the control box. And a security guard or police will instantly be informed about the crime.

1.6 Paper Layout: In this paper, we discuss the introduction part of our model in chapter 1. This chapter presents our motivation, rationale of studies, and expected outcomes. Chapter 2 briefly discusses the background research and related work of our research. In that part, we shortly discuss the related works, comparative analysis and challenges of our work. We introduce our research methodology in Chapter 3. In this section, we discuss the instruments, data collection procedure, and training approach step by step. After the training process, we refer to the experimental results and analysis in chapter 5. We think our research is very impactful for society. So we briefly discuss the impact of our model on society in chapter 5. At the end of our discussion, we summarize our paper at Chapter 6. We also discuss the Implication for Further Study in this Part.

CHAPTER 2

BACKGROUND

2.1 Terminologies: YOLO, Object Detection, Neural Network Architecture, Image Processing, Classification, Human Behavior, Crime Activities.

2.2 Related Works: Samson, Oladipo & Emaka [1] discovered a model that can detect a crime scene by analyzing crime evidence without human interaction. In the whole process they maintained the Cross-Industry Standard process for data mining. Koichino, Xiaoxue [2] were presented a public safety System(PSS) for developing an automated video surveillance system based on statistical characteristics for CCTV footage.There are three critical processing stage are introduced in [3] Shian-Ru-ke, Hoang Le, Yong-Jin, Neng's work. Object segmentation is the first satge.It mainly segmented an image from background. Feature extraction and representation is the second stage.In this part the effective images are extracted and represented for creation of a feature in a systematic way. [18] Joseph Rendom, Ali Farhadi introduce a new YOLOv3 algorithm that works significantly faster than other object detection models with their performance. They run fourteen different types of models. Yolov3 performed more accurately and faster than others.

Yolov4 is another object detection model that was introduced by [5] Alexey Bochkosk, Mark Lia. This is another optimal and powerful model that can be used with a 1080 Ti or 2080 Ti GPU for training a super-fast accurate object detection. In the computer vision sector, object detection plays an impact rule. There are two most effective algorithms are Single Shot Detector(SSD) [2] and YOLO Shinde, S., Kothari, A. and Gupta, [17]. [6] Zhitho, Redouane, Benoit experiments with those methods. That Research is mainly focused on the motions estimates and object detection part within real time in traffic environments and public places. They only focused on the comparison between SSD and YOLO algorithm. [7] Ashok Shows the effectiveness of the shape analysis in objective recognition, matching and registration. In their work, they describe both parametric method and Non-parametric method.

In computer vision, Chen C H [13] shows that, object detection is a prime matter. It has two main issues: detecting objects from images and estimating their class and location. Different approaches have been proposed by O'Shea, Keiron, and Ryan Nash [19] over the past few years to successfully manage these issues based on CNN (Convolutional Neural Network)[16]. These procedures are usually done in two steps. In the first step, regions are classified into specific categories by identifying the region where the object is located. In the second step, class and location are estimated. RCNN is one of them which works in two steps. Its improved version is based on two independent neural networks. Held C, Krumm J, Markel P, et al, [15] introduce a digital Intelligent Surveillance system that can detect criminal activities without human interaction. One is the classification network and the other is the regional proposal network [17]. These methods are almost similar in efficiency according to mAP. It is a criterion that measures the proportion of correct identification as the function of recall, which is the proportion of objects identified. However, two independent networks make very slow predictions in embedded computing platforms. In the one-step detection method, classification is performed on a fixed size and number of bounding boxes. Then with localized regression of the detection object. Single-stage architectures are faster than two-stage architectures and performance is almost the same. One step architectures are SSD (Single Shot Detector) and YOLO (You Only Look Once) which are very popular methods. We choose one widely applied one phase detection model: YOLO (You Only Look Once) to evaluate based on the above discussion.

The foreground detection feature extraction movement behaviour classification and recognition technique at the key components in machine learning field. All five technologies are introduced in Honghua Xu, Li Li!", "Ming Fang.Movement Human Actions Recognition Based on Machine Learning" paper. In there work they adopt Convolutional neural network (CNN) to extract feature and reduce the dimension of object based on optical flow energy. For classify and recognizing and action, the support vector machine classification was developed. After the training and testing process with human movement it count effectively make a difference the abnormal human activities.

At last they applied their model on a intelligent video surveillance system that successfully identify abnormal behaviour and alert. Such as robbery fighting car accident etc.

QI-CHAO MAO[1], HONG-MEI SUN [1], [2], YAN-BO LIU[1], AND RUI-SHENG JIA[8] Have designed a lightweight embedded structure called Mini-YOLOv3 which maintaining high detection accuracy. The built a feature extraction backbone network almost 16% size of parameter of Dark net-53. They have also used a Multi scale feature pyramid depends on a simple U shaped structure to prohibit the reduction o f accuracy. They call it Mini-YOLOv3. It is smaller model. Its parameters and floating point operation is almost similar to the origina YOLO algorithm with MS-COCO datasets.

A video surveillance and monitoring(VSAM) system is established by Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N.,Hasegawa, O., Burt, P. and Wixson, L.,[15]. In the digital Security System, automatic video surveillance system and monitoring system is a covetable area. It is also impact full research in commercial sector. By increasing technology the normal security camera is getting cheap and available day by day. But that camera only captured the video and investigator have to analyse the captured video to the understand what happened.In their paper, they developed a surveillance system that can continuously 24-hours monitoring and analyses the video data to alert and ensure an a abnormal activities to the security guard.

For the decision making and forecasting techniques determining process plate a powerful role. Adriaans,[21] Has describes some impactful power of data mining. Data mining mostly depends on machiner learning, pattern recognition and statistics to automatically filter a conception and finding a interconnection patterns from the huge collections of datasets. The data mining model mainly works based on previous data and have taken a decision from its prediction. The motive of their paper is to investigate how determining will be used in industrial and corporate areas.

Ayodele, Taiwo Oladipupo [22] describe different kinds of machine learning algorithm. Supervised learning, Unsupervised learning and semi supervised learning are three main

algorithm is proposed in their paper.The supervised learning algorithm create or generate a function that maps inputs to design outputs. The task to get the computer to learn a classification system supervised learning is mostly common. This model is perfectly work as long as the inputs are given. When the output values are missing, it is not possible to determine anything about the output figure-3(left).

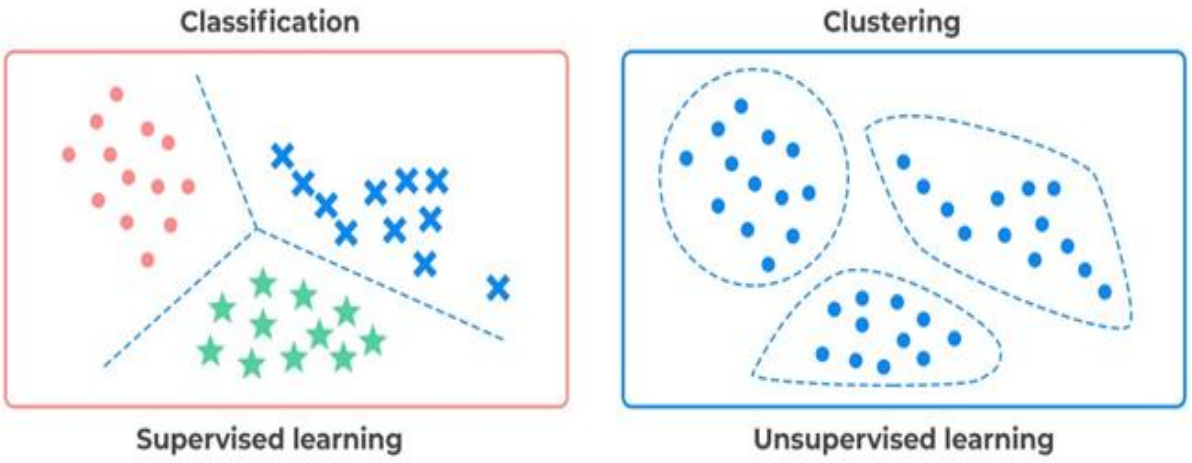

Figure-2.2.1: The pictorial representation of supervised learning(left) vs Unsupervised learning (right)

In Unsupervised learning, the main task is to have the computer learn in new thing that we don't give him. It is much difficult part. There are many approach, the popular approach is to teach a computer not by giving exact classification but by applying many sort of reward system to obtain success figure-3(right).

To develop a faster artificial neural network, Graphics Processing Unit (GPU) is used.

Oh, K.S. and Jung, K.[23] uses the matrix multiplication method of a neural network to upgrade the time performance of GPU.

2.3 Comparative Analysis: We consider one phase method instead of two stages methods based on media (video for example). So now our deal is with YOLOv3. It also trained an update classifier network that is better than others.

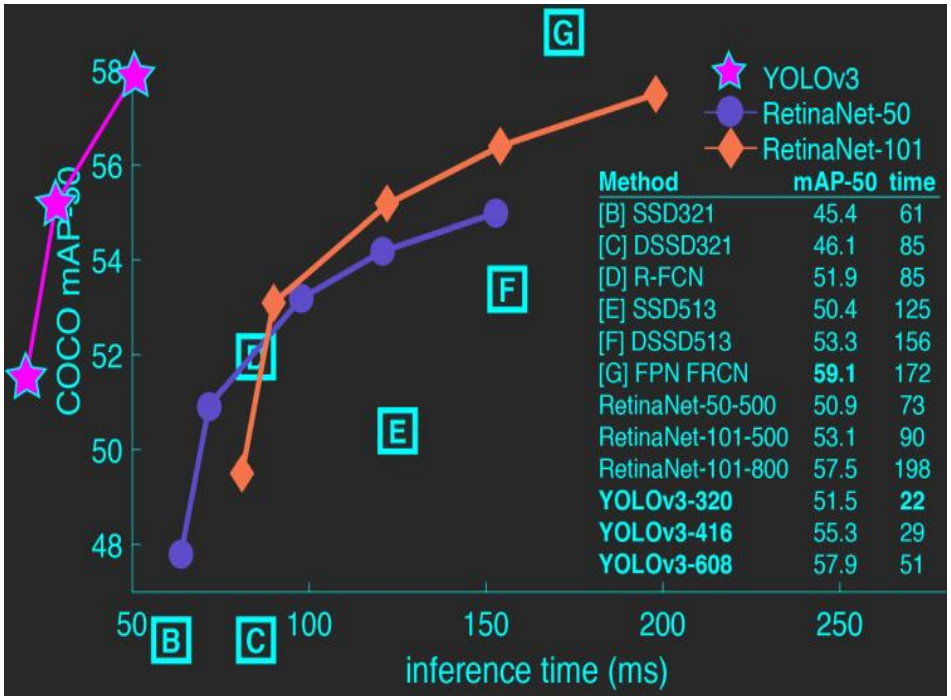

Figure 2.3.1: Performance comparison of YOLOv3 with other object detection models.

The above figure 2.3.1 is shown the comparable performance of YOLOv3 which is remarkably faster than other algorithms.

2.4 Challenges: We encountered several obstacles during our research. Firstly, we faced challenges with data. Since the data should be based on human behavior detection from images. Therefore, according to the class of crime activities, human movements / poses need to be perfect. Otherwise, conflicts may arise when detecting specific classes correctly. There is a possibility of false detection. Secondly, if a human does not commit a crime activity but makes a movement that matches a trained movement, then our model will consider it as a crime activity. Finally, using YOLOv3 requires a high-end GPU that performs well and is expensive. We use a free GPU from Google Colab. But there is a little limitation in using it, that is, after a certain period of time, the GPU has to be purchased from Google Colab.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation: Our research subject is detecting crime activities based on human behavior using machine learning.

To get human behavior and human movement right, we need a media for our research. It can be any kind of video camera. Through which specific objects can be identified according to movement from live feeds, images, videos.

A CNN (Convolutional Neural Network)[19] is a class of ANN (Artificial Neural Network) which is applied to analyze visual imagery most of the time in deep learning. SIANN (Shift Invariant or Space Invariant Artificial Neural Networks) is also known as CNNs which is based on shared weight architecture of the convolution filters that provide translation equivariant responses and slide along input features known as feature maps. They have applications in recommender systems, image and video recognition, image classification, natural language processing, image segmentation and financial time series etc.

YOLO (You Look Only Once) is a real time object detection algorithm. It detects specific objects from live feeds, images, videos. YOLO this ML (Machine Learning) model uses features of a deep Convolutional neural network to recognize objects. Versions 1 to 3 of YOLO were developed by Joseph Redmon and Ali Farhadi, and the version three of the YOLO machine learning algorithm is a much perfect version of the actual machine learning algorithm. YOLOv3 is an elevated version of YOLO and YOLOv2. YOLO is made to use Keras or OpenCV deep learning libraries. Object classification systems are conducted by artificial intelligence programs perceiving a class of specific objects as topics of interest.

CNN is included of three types of layers. Convolutional layer, pooling layer and fully connected layers are among them. By combining those three layer CNN model has been constructed. A complete CNN architecture is explained in figure 3.1.1.

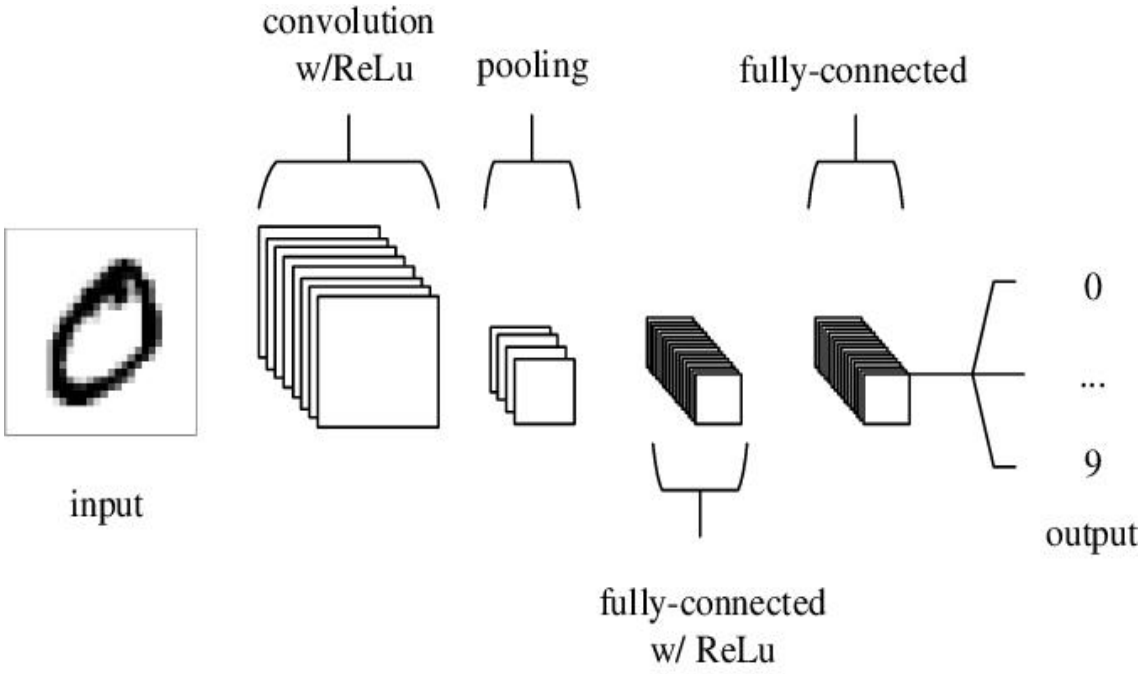

Figure 3.1.1 : A simple CNN model, combined of five layers.

Main functionality of the given figure[2] can be divided into four basic area:

1. at first the input layer will be take an image as a pixel value
2. Then the convolutional layer will find local regions connected neurones and calculate the scalar product between the weights and region connected volume
3. The pooling layer will reduce the number of parameters those are ineffective and will keep special dimensionality of the given image.

The fully connected layer will provide class level to be used for classification. It plays the similar rules as ANNs.

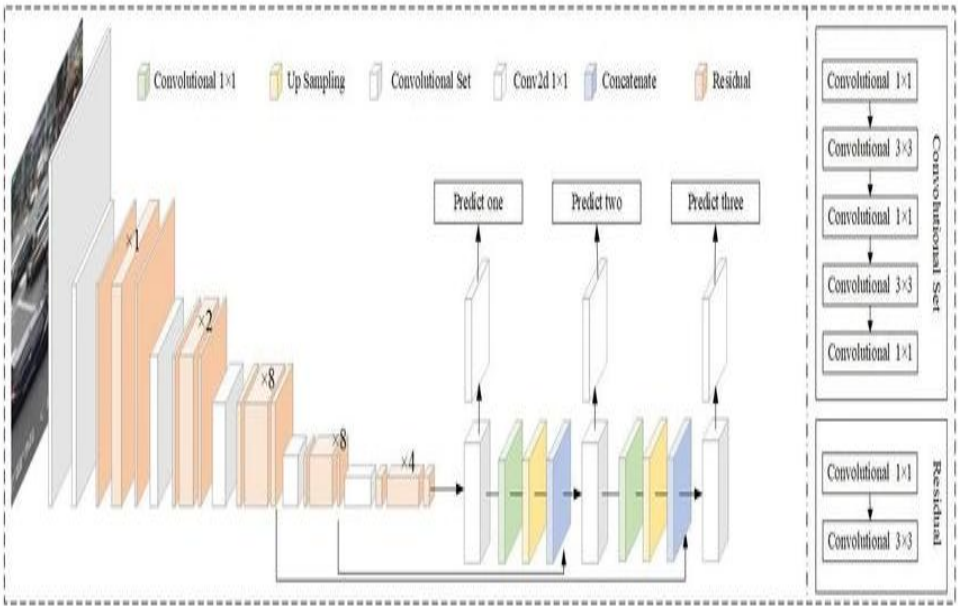

Figure 3.1.2: The visual structure of YOLO v3.

YOLO is a Convolutional Neural Network (CNN) for real-time object detection. It uses three steps prediction with the help of darknet 53. It can able to process a given image as a structural array of data by using a classification-based system. In fig 3.1.2, It shows structural details of YOLOv3 and it uses Darknet-53 as the backbone network. YOLO (You Look Only Once) has the advantage of being faster than other networks and still maintains accurate accuracy. Which passes the model to see the full picture during testing. As a result, its prediction is done correctly by the global context within the image. YOLO (You Only Look Once) and other CNN (Convolutional Neural Network) algorithms score regions according to their similarity to top redefined classes. High-scoring regions are considered positive identities for the class with which it identifies most closely. For example, in a live feed of an animal zoo, YOLO (You Only Look Once) can be used to identify different types of animals depending on which regions of the video score higher than a predefined class of animals.

We need a high level programming language for our machine learning algorithm to execute using code. Python is a high-level programming language, an interpreted, object oriented language with dynamic semantics. It has built in DSA (Data Structure & Algorithm) combined with dynamic binding and typing, making it very useful for our research methodology.

LabelImg is a tool for graphical image annotation. LabelImg is developed using Python. For the graphical interface, it uses Qt. PASCAL VOC format saves annotations as XML files. PASCAL VOC format used by ImageNet. Also, LabelImg supports CreateML formats and YOLO.

Google Colab is a free platform where anyone can write Python code and execute it in a browser. It is mainly used for machine learning, data analysis, and education purposes. Technologically speaking, Colab is a hosted Jupyter notebook service that requires no setup, providing free access to computing resources including free GPU.

3.2 Data Collection Procedure and Data-set Preparation: The entire process of our project followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) in Figure 3.2.1 shown as it is a completely standard process for solving a problem by using data science and machine learning algorithm [22]. It is a complete process for developing a model from business understanding to model deployment. The ten steps of CRISP-DM in figure 3.2.1 are applied in section 3.2 describing the data-set collection and model creation, and section 3.3 describing the object detecting approach of our model.

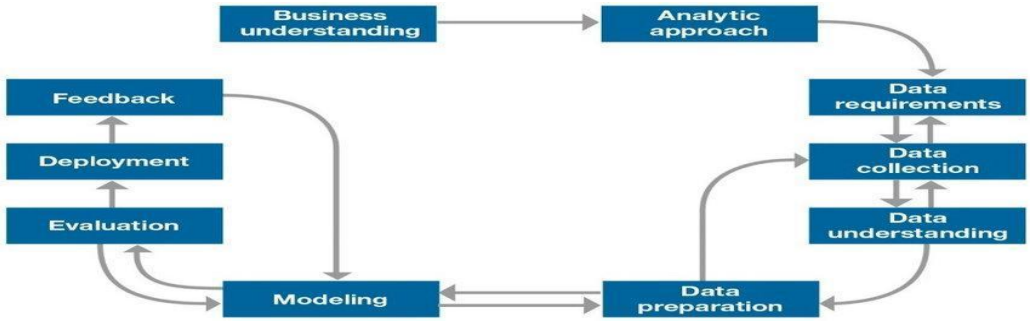

Figure 3.2.1: process for developing a model from business

The data collection procedure is the primary or initial part. But it is a one of the most significant parts of any object detection related work. At first we collect data by capturing thousands of images. We also find some an appropriate data that is not perfect for our project. Then we remove those interrelated data from our data set. we just only select the images that t are captured perfectly. After reducing the Irrelated images now we have a huge collection of optimal visual data. After that we label our optimal data with labelimg tool. We have selected 5 common classes those are seen in a crime scene. Then we classify our data with those 5 classes. After getting the labeled data divide them into two groups and named test data and trained data. after that we ready to train our model. we have selected YOLOv3 detection algorithm to train our model.It is a most popular and more effective object detection algorithm nowadays. It provides the more accurate result then other object detection algorithm. For training the model we use the google colab with free GPU. After the training process we test the model by test data. At the end of successful training our propose model would be detect and abnormal activities based on the human movement. The main motive of this work is to build a model that can play an effective role on digital security system. The digital security system might be keep an effective and significant role to reduce the crime and also reduce its damages.

In an object detection model, data plays a key role for detecting an object from the given images and video scene. If we train a model with a collection of effective datasets, then we get a more accurate result. Anyways, in our model, we fully focused on the human body movements, because our proposed model will be able to understand a crime scene and abnormal activities based on human body movement and pose. So we have to need the specific human pose related data that indicates a criminal activity. We searched related data on google or other sites but we did not find the efficient required data for our model. So, we decided to create our own data. We have selected five specific human movements that indicate and are often seen in criminal activities. Then we captured thousands of images of these selected poses ourselves. We have collected 1600+ images of different poses and classified them into five classes.

The given figure 3.2.2 is shown the whole procedure and methodology of our work in a diagram representation.

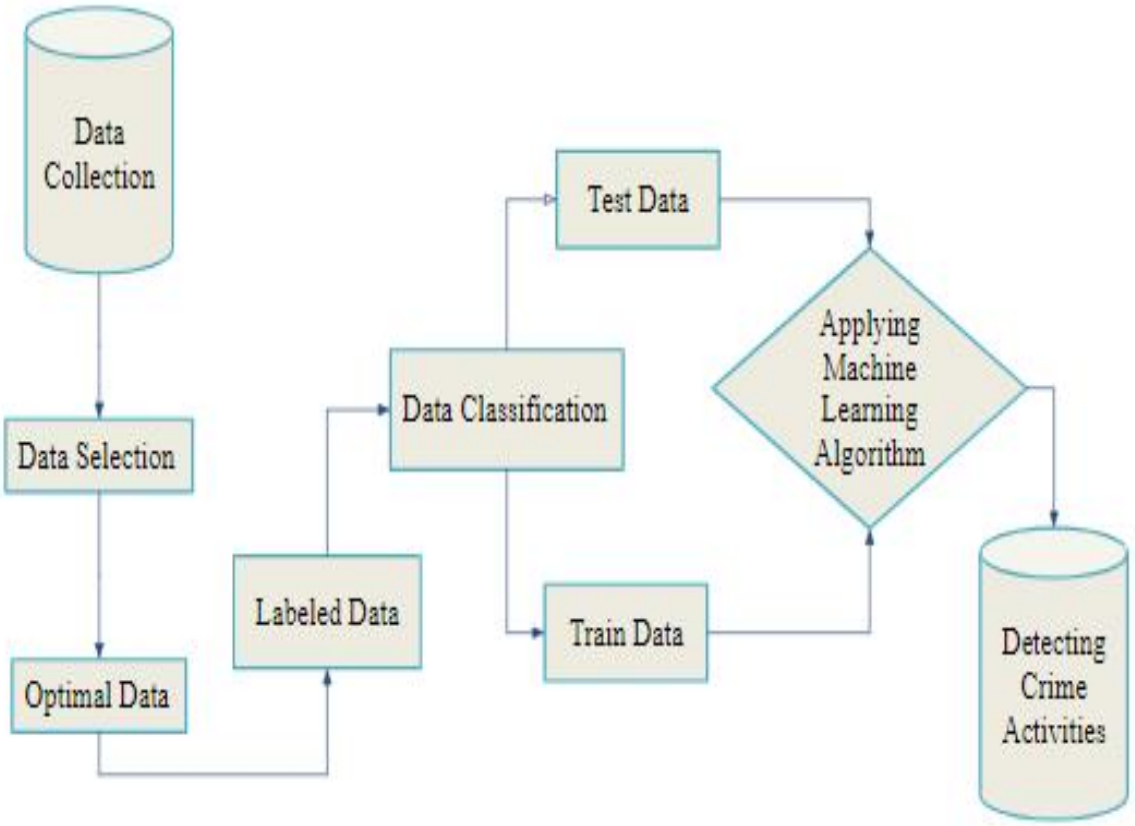

Figure 3.2.2: Methodology diagram of our project.

Figure 3.2.3 shows sample images in selected classes: velitation (300+ images), robbery (300+ images), pick-pocketing (300+ images), hijack (300+ images), stick fighting (300+ images).

Figure 3.2.3: Sample image data-set

In the entire data collection process, we tried to vary the movements, resolutions and sizes from one image to another image. After the data collection process, we need to annotation them. We manually label all the datasets by using labelImg tool. LabelImg is a visual photo annotation tool that allows us to create visual boxes amount a object in given image. After the labelImg process, it saves a xml file labeled image. Each text file contains four values like figure 3.2.4

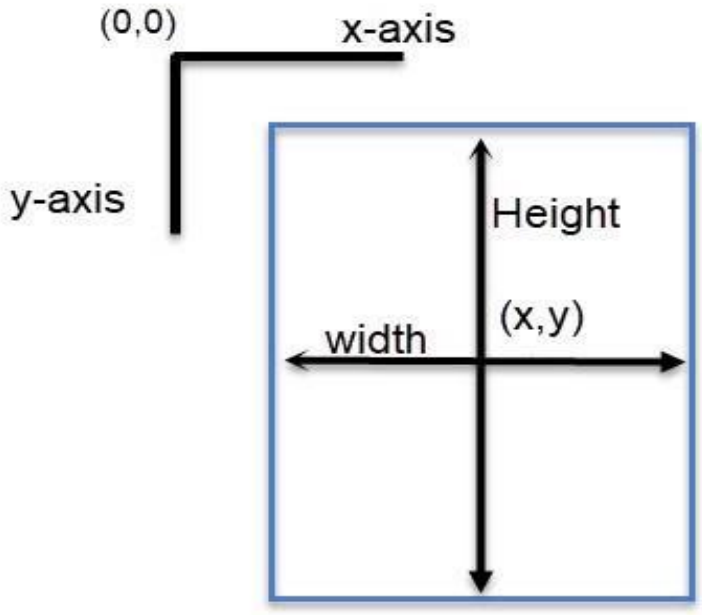

Figure 3.2.4: Determining the dimension and location of an object in a bounding box

In the above figure 3.2.4 shows that It annotated an image in the following form

Class name, x, y, width, height

Here,

class_name = class name of object

x = the coordinate of the X-axis

y = the coordinate of the Y-axis

width = width of bounding box

height = height of bounding box

By labeling all the images of our datasets, now we are ready to train our model. After the annotation process our model will clearly understand the position and movements of an object from an image. We split our data-set into two pants 80% data is splitted for training and 20% for testing. The model is developed on the training data-set and

evaluated using testing data set. Now the model is ready for training to detect criminal activities based on human movements.

3.3 Applied Mechanism and Training Approach: We develop a convolutional neural network (CNN) to detect an object by classification an image from the analysation of a crime scene. We train CNN by changing the different Convolutional layers for fastly detecting objects from the sequence of images. Our model was trained on google colab with free GPU and darknet. For the training process we have to set the YOLO V3 configuration file. We set up the configuration file based on how many classes we are training in our detector on.

Batch: It represents the number of training data in one iteration. When we load the data set to the memory we have two option:

i)We can load whole data set at once into the memory

ii) We can split the data-set into small parts and send them to memory.

This makes the training process faster, easier, and more accurate. Because if we load too large data into memory, it will slow down the training process. Because we are using lots of memory on the GPU. In our training process, we set batch=64, which means there are 64 images in a batch. Subdivision: Subdivisions are the split parts of the batch. A batch is further divided into many mini-batches. These mini-batches are called Subdivision. Since there is not much space on the GPU, the batch splits into smaller parts and sends them to the GPU for training. In our training process, we set the Subdivision= 16. So the batch is divided by 16, 64/16=4 images are sent for processing at one subdivision.

Max_batch= #num of classes*2000.We have selected 5 classes. So max_batch=10000

Steps= 80% of max_batch , 90% of max_batch. So steps= 8000,9000

Filter= (# of classes+5) *3. So filter= 30.

Width: 416. Network size of width.

Height: 416. Network size of height. All the images are resized during training.

At first we set batch = 16 and subdivision = 16 for our ultimate result. The process of configure our variables:

The following table 3.3.1 shows the different stages of CNN layer and filters during the training process of our model.

	Layers	Filters	Size	Input	Output
0	conv	16	3 x 3/1	416 x 416 x 3	416 x 416 x 16
1	max		2 x 2/2	416 x 416 x 16	208 x 208 x 16
2	conv	32	3 x 3/1	208 x 208 x 16	208 x 208 x 32
3	max		2 x 2/2	208 x 208 x 32	104 x 104 x 32
4	conv	64	3 x 3/1	104 x 104 x 32	104 x 104 x 64
5	max		2 x 2/2	104 x 104 x 64	52 x 52 x 64
6	conv	128	3 x 3/1	52 x 52 x 64	52 x 52 x 128
7	max		2 x 2/2	52 x 52 x 128	26 x 26 x 128
8	conv	256	3 x 3/1	26 x 26 x 128	26 x 26 x 256
9	max		2 x 2/2	26 x 26 x 256	13 x 13 x 256
10	conv	512	3 x 3/1	13 x 13 x 256	13 x 13 x 512
11	max		2 x 2/1	13 x 13 x 512	13 x 13 x 512
12	conv	1024	3 x 3/1	13 x 13 x 512	13 x 13 x 1024
13	conv	1024	3 x 3/1	13 x 13 x 1024	13 x 13 x 1024
14	conv	50	1 x 1/1	13 x 13 x 1024	13 x 13 x 50

Table 3.3.1: Tiny-YOLO Architecture

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Results & Analysis: It is known that CNN played an impact role in the real time object detection model for crime scene analysis, but in this research we introduce a better optimized model YOLO V3 object detection algorithm. Our model has successfully implemented an intelligence surveillance system that can detect criminal activities in real time. The model achieved an average accuracy of 89% of training 10000+ iteration on the google colab with free GPU and darknet. The object detection model was evaluated with the help of a test data set. No object was detected when the confidence threshold was reduced 2 until all five objects could be detected at confidence threshold of 0.0013. 90% of correct data set on test set was achieved with an average time of 5.89 seconds.

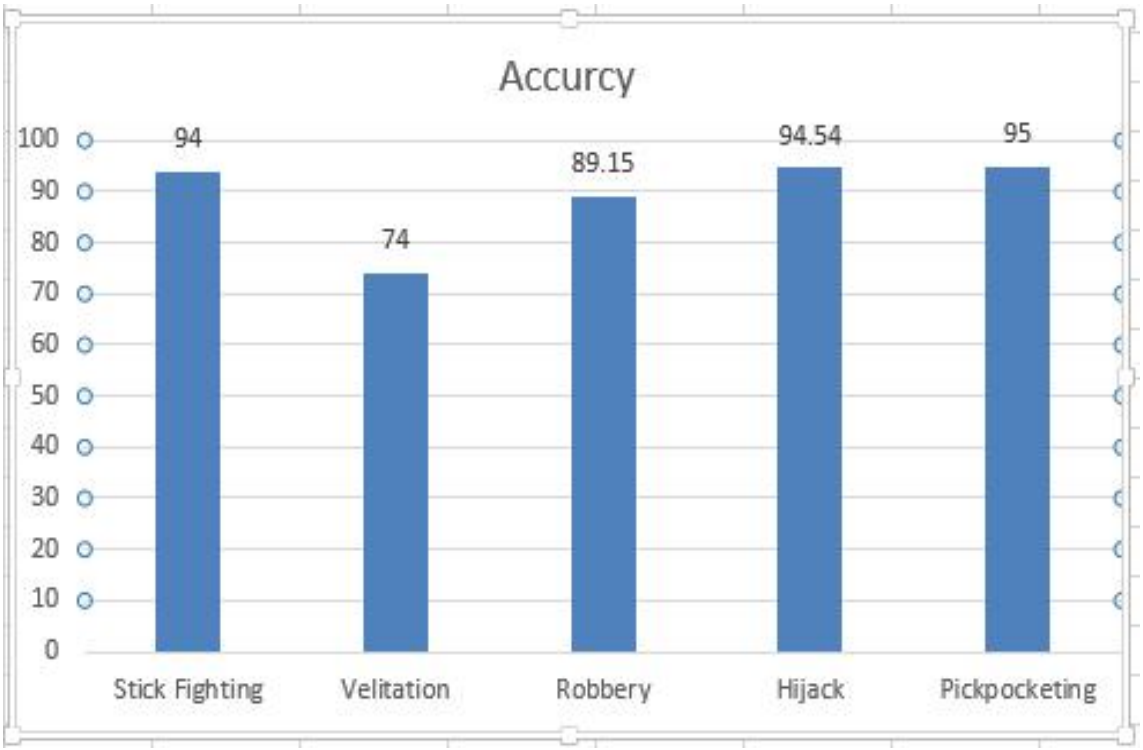

Figure 4.1.1: The percentage of detection accuracy for each class.

In the above figure, 4.1.1 shows the object detection accuracy for each class. We can see the accuracy of stick fighting class is 94%, velitation class is 74%, robbery is 89.15%, hijack is 94.54 and pick pocketing is 94%.

The following figure 4.1.2 shows the object detection percentage including the accuracy of the of outputs.

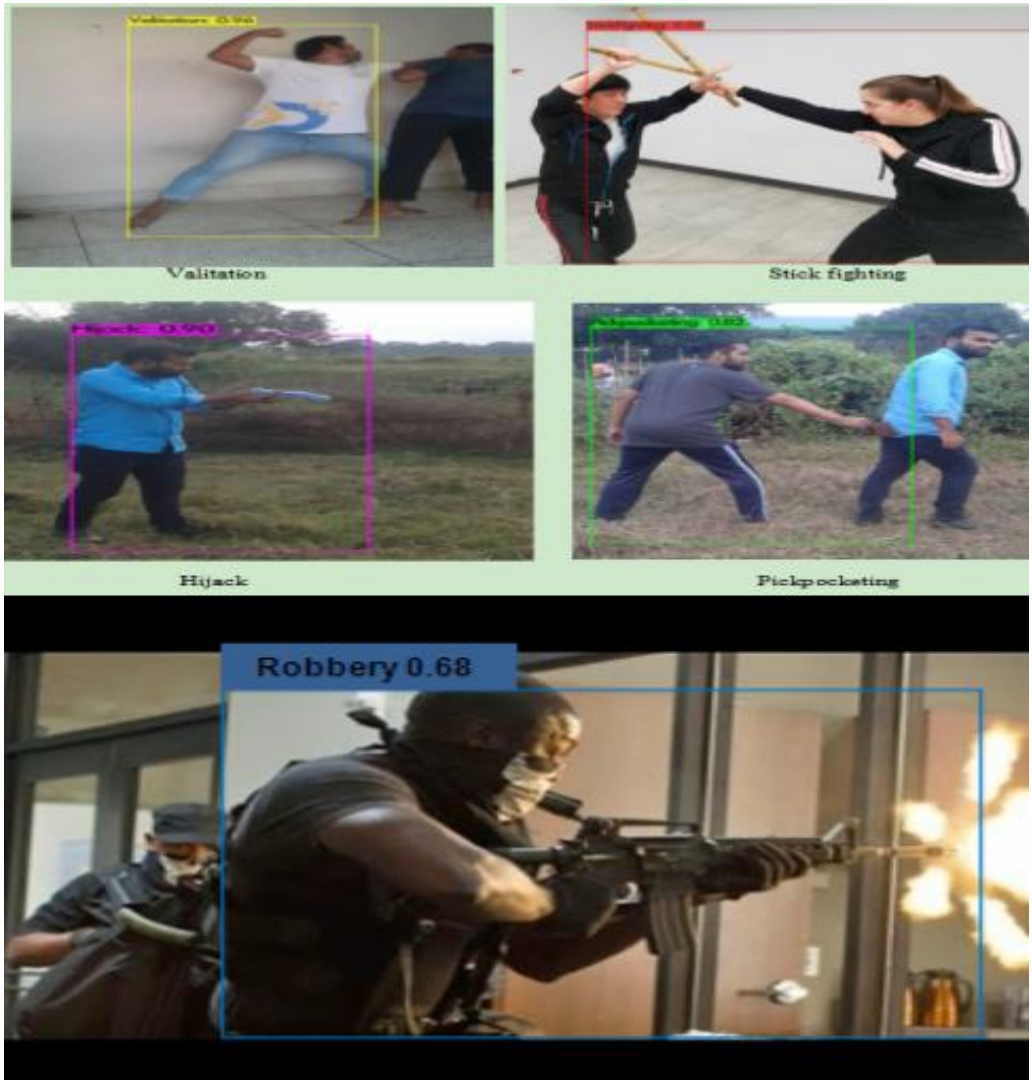

Figure 4.1.2: Object Detection Outputs on the Sample Images

We can see in the images the objects are successfully detected by our model. The 5 Example of object detection result on test data set square overlaid on images mark the bounding boxes and accuracy of the detected object.

4.2 Discussion: The results obtained in this study showed a promising future for crime evidence analysis. Crime scene reconstruction begins with identifying pieces of evidence at a crime scene. The objective is to find people or things involved in the crime scene that may be a clue to the arrest of the criminals. The implication of this study is to help the jury system in investigating crimes so as not to implicate innocent people. The system will greatly assist in decision making and focus on prosecuting the real criminals involved in crime. After obtaining a promising result, the study went further to evaluate the accuracy of a video file model and an image not in the crime evidence analysis data set used in training. This is to test its accuracy in a real-life crime scene. Unfortunately, there is some confusing data in our data set. That's why we have some false detection. This false detection can be easily solved if our data is more accurate. The level of intelligence displayed by the system is actually required whenever a crime scene analysis is required.

CHAPTER 5

IMPACT ON SOCIETY AND ENVIRONMENT, ETHICAL ASPECTS, SUSTAINABILITY PLAN

5.1 Impact on Society: Various types of crimes are committed in our society. In order to return to the streets, one has to accept various types of criminal activities such as theft, robbery, pick pocketing, fights, brawls. We think it is possible to reduce these activities by implementing our research based project. Because based on our project it is possible to detect real time crime occurrence and identify the right criminal. It is often seen that after committing the crime, the criminals leave the place quickly which makes it difficult to trace the real criminals later. Or because of not being able to identify the culprit at the right time, the wrong person is often identified as the culprit. Then due to the fact that the criminals do not recognize the crime in real time, the criminals take advantage of it again later to abuse it. Therefore, if the crime can be identified in real time and the criminal can be identified, it can be done quickly. As a result, there is a possibility that the crime rate will decrease in that place as we are able to take an instant action. We hope this research based project will be very impact for our society. And the incidence of crime will also decrease to a great extent. And if a crime has been committed, quick action can be taken. We face many problems in our daily life. Various types of criminal activities are happening in our society. As a result we have to suffer and sometimes lose valuable property. Most of the times we cannot do anything against this problem due to lack of proper evidence. Considering this problem we are trying to find a way out of it. Our model can contain the necessary visual data and proper evidence against these crimes. As a result, criminals can be easily identified by taking necessary action against them. It will help to provide strong security of our home or institutions.

5.2 Impact on environment: Our research based project will have a significant impact on the environment. Since it is possible to detect any crime activities through human behavior analysis in real time, it insures much more security. As a result of which

everyone will be able to roam freely on the road very easily. As a result, it also has an important impact on our environment. And also those who do such activities will also become very aware that our free movement plays a very important role for the environment. Everyone will be secured from such incident happening. We are living in a environment which is very important for us. So, we have to be careful about the environment. We must consider the safety of the environment before every action we take. If we can keep our environment safe then we ourselves will be safe. Our model is environment friendly and it can be used for environmental protection.

5.3 Ethical Aspects: We need to be honest while researching anything. It can be seen that some researchers use fake datasets during research. Then they manipulate the data in such a way as to get better results without providing a well-research based data set. And these matters are completely unethical. Crime activities may not be accurately detected through certain poses from human movement if the data set is not properly arranged. So for our research, we conduct our own research and manage the data ourselves, basically like human movement during crime activities. We divide over 1600 custom data into five specific classes to build our model. We try to work with large amount of data to make our model work more accurately.

5.4 Sustainability Plan: When we research something, we must think about the sustainability of that work. In our research we have tried to create a model using machine learning approach that can detect crime activities from human poses. This type of research can have an important impact on human beings. And that is why we have thought about the sustainability plan for this research. So far we have worked with only 5 classes to detect some activities from human poses. But this task can be made more efficient by increasing the number of classes and collecting more datasets. Finally, one more method to add is through detection based on motion of human movements. And this method can accurately predict crime activities which can be much more impactful.

CHAPTER 6

SUMMARY, CONCLUSION, AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study: In this study, we want to show that a machine can understand and detect an environment only by analyzing human poses and movements. Our proposed model will play a key role like an active security guard. This research is done using the YOLOv3 object detection algorithm. At the beginning of the work, we selected 5 specific human poses that we commonly see at different crime scenes. Through these five selected gestures, we captured 1600+ images of ourselves. We classify all these images into five classes like stick fight … We split the whole data set into two folders named train data and test data. Train data is used for training our model and test data for testing the model. We prepared them for training by annotating them. Finally, we trained our model with the collected data set using Google colab with free [23] GPU. We then test our model using the test data set and get a 91.63% accurate result. We have some confusing data in our data-set that's why the model provide some false detection. Above all, our model can analyze human behavior from a video scene and understand whether a crime has taken place from it.

6.2 Conclusions: This study was able to meet its goal by successfully developing a machine learning forensic through transfer learning in YOLO using a custom data set collected and prepared in the study. The model was trained on google colab with free GPU achieving 89% accuracy. However, there was a limitation in the study regarding the integration of the model into mobile apps for effective end-user operations. Nevertheless, the tool app can be used to digitally keep physical crime case records and model identification of objects found in internal crime scenes or evidence analysis. Although

this system is not designed to prevent crime, knowledge of its existence will add to the reduction of crime in the nation.

6.3 Implication for Further Study: There is scope for further work on our research in the future. As the tendency to commit crimes is increasing day by day, at the same time, the demand to use the Intelligent video surveillance(IVS) system is increasing. Because SVM is the only way to detect a crime as well as take an effective step. Our proposed model can also be used as an IVS system. Because it is also able to identify a crime from a video sequence. By training our model more accurately, it is possible to create a complete project that will prevent any crime in society and reduce its damage.

We also need to work on data in the model further to be more accurate. Because the more accurate data we can train with, the more accurate the model will be.

We have used the YOLOv3 algorithm here. In the future, we may add something else algorithm that will be more effective. All in all, our model can be further updated in the future and marketed as a complete project for commercial purposes. That works like a security guard can detect a crime more effectively and identify the real culprits.

References

[1] Francisca Onaolapo Oladipo,Emeka Ogbuju,Femi Alayesanmi"An Object Detection and Classification Model for Crime Evidence Analysis Using YOLO" Preceding On Big Data Analytics & Innovation, Volume3, 2018,pp 18-26

[2] Koichino Goya,Xiaoxue Zhang, Koaki Kitayama"A Method for Automatic Detection of Crime for public Seccurity by Motion Analysis"Big Data Analytics & Innovation, Volume5, 2017,pp 17-23

[3] Held C, Krumm J, Markel P, et al.” Intelligent Video Surveillance”. Computer, 2012, 45(3):83-84. https://doi.org/10.1109/MC.2012.97

[4] Joseph Redmon, Ali Farhadi"YOLOv3: An Incremental Improvement",https://pjreddie.com/yolo/.

[5] Alexey Bochkovskiy,Chien-Yao Wang* , Hong-Yuan Mark Liao,"YOLOv4: Optimal Speed and Accuracy of Object Detection"arXiv:2004.10934v1

[6] Zhihao Chen,Redouane Khemmar,Benoit Decoux,Amphani Atahouet,Jean-Yves Ertaud, "Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility",978-1-7281-5546-3/19/$31.00 ©2019 IEEE.

[7] Ashok Veeraraghavan,Amit K. Roy-Chowdhury,Rama Chellappa, "Matching Shape Sequences in Video with Applications in Human Movement Analysis", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 12, DECEMBER 2005.

[8] QI-CHAO MAO1, HONG-MEI SUN 1,2, YAN-BO LIU1, AND RUI-SHENG JIA,” Mini-YOLOv3: Real-Time Object Detector for Embedded Applications” IEEE Received August 22, 2019, accepted September 11, 2019, date of publication September 16, 2019, date of current version September 27, 2019.

[9] Nixon M S, Aguado A S. “Object description - Feature Extraction and Image Processing “, 2002:247–290. https://doi.org/10.1016/B97 8-0-08-050625-8.50011-0

[10] Honghua Xu, Li Li!”, “Ming Fang.Movement Human Actions Recognition Based onMachine Learning” .https://doi.org/10.3991/ijoe.v14i04.8513

[11] Held C, Krumm J, Markel P, et al.” Intelligent Video Surveillance”. Computer, 2012,45(3):83-84. https://doi.org/10.1109/MC.2012.97

[12] Veeraraghavan A, Member S, Roychowdhury A K.” Matching shape sequences in video with applications in human movement analysis[C]” IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005:1896--1909. https://doi.org/10.1109/TPAMI.2005.246

[13] Chen C H. Handbook of Pattern Recognition and Computer Vision[M]. World Scientific Publishing Co. Inc. 2016. https://doi.org/10.1142/9503.

[14] Bobick A F, Davis J W. T”he Recognition of Human Movement Using Temporal Templates [J]”. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(3):257-267. https://doi.org/10.1109/34.910878

[15] Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P. and Wixson, L., 2000. A system for video surveillance and monitoring. VSAM final report, 2000(1-68), p.1.

[16] Sultana, F., Sufian, A. and Dutta, P., 2020. A review of object detection models based on convolutional neural network. Intelligent computing: image processing based applications, pp.1-16.

[17] Shinde, S., Kothari, A. and Gupta, V., 2018. YOLO based human action recognition and localization. Procedia computer science, 133, pp.831-838.

[18] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv: 1804.02767 (2018).

[19] O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv: 1511.08458 (2015).

[20] Bradski, G. and Kaehler, A., 2000. OpenCV. Dr. Dobb’s journal of software tools, 3(2).

[21] Adriaans, P., 1996. Data mining. Pearson Education India.

[22] Ayodele, Taiwo Oladipupo. "Types of machine learning algorithms." New advances in machine learning 3 (2010): 19-48.

[23] Oh, K.S. and Jung, K., 2004. GPU implementation of neural networks. Pattern Recognition, 37(6), pp.1311-1314.

DETECTING CRIME ACTIVITIES BASED ON HUMAN BEHAVIOR USING MACHINE LEARNING

ORIGINALITY REPORT

16%	12%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Daffodil International University Student Paper	2%
3	Qi-Chao Mao, Hong-Mei Sun, Yan-Bo Liu, Rui-Sheng Jia. "Mini-YOLOv3: Real-Time Object Detector for Embedded Applications", IEEE Access, 2019 Publication	1%
4	Zhihao Chen, Redouane Khemmar, Benoit Decoux, Amphani Atahouet, Jean-Yves Ertaud. "Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility", 2019 Eighth International Conference on Emerging Security Technologies (EST), 2019 Publication	1%
5	Submitted to Thadomal Shahani Engineering College Student Paper	1%