HUMAN ACTIVITY DETECTION USING YOLOV4

By

Rezwan Ahmed Siam
ID: 171-15-1422

Ahmed Nur-A-Jalal
ID: 171-15-1324

and

Tonima Aslam Barsha
ID: 171-15-1258

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

Supervised By

Mr. Ohidujjaman
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
December 2020

APPROVAL

This Project titled "**Human activity detection using YOLOv4**", submitted by Rezwan Ahmed Siam, ID 171-15-1422 , Ahmed Nur-A-Jalal, ID 171-15-1324 and Tonima Aslam Barsha, ID 171-15-1258 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.

BOARD OF EXAMINERS

____

(Name) **Chairman**
Designation
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

(Name) **Internal Examiner**
Designation
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

(Name) **External Examiner**
Designation
Department of -------
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Ohidujjaman, Lecturer (Senior Scale), 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:

Mr. Ohidujjaman
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Mushfiqur Rahman
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Rezwan Ahmed Siam
ID: 171-15-1422
Department of CSE
Daffodil International University

Ahmed Nur-A-Jalal
ID: 171-15-1324
Department of CSE
Daffodil International University

Tonima Aslam Barsha
ID: 171-15-1258
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Mr. Ohidujjaman**, **Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Fatema Tuj Johora, lecturer (Senior Scale)**,** Department of CSE, for her kind help to finish our project and also to other faculty member 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 discuss while completing the course work.

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

ABSTRACT

A simple activity recognition model can allow a single human person to monitor all our surrounding with the purpose to ensure safety and privacy while preserving maintenance cost and efficiency with the soaring level of precision. This monitoring system with real-time video surveillance could be deployed for patients and the elderly in a hospital or old age home along with various human activity on important area such as the airport. For speedy analyze of action and accurate result while working with complex human behaviour, we decided to use YOLOv4 (You Only Look Once) algorithm which is the latest and the fastest among them all. This technique uses bounding boxes to highlight the action. In this case, we have collected 4,674 number of different data from the hospital or different condition ourselves for fastest accuracy with the one of the largest data-set ever used in such kind of project. During our research, we had divided our action into three different class which are standing, sitting and walking. This model was able to detect and recognize multiple patients or other regular person activity and multiple human activities tracing support at once. After completing our project, this model manages an average accuracy of 94.6667% while recognizing image and about 63.00% while recognizing activity from video file. We also work on two other different projects with TensorFlow and OpenPose while the YOLOv4 perform better than those two. In future, more complex data can be added and prediction can be implemented which will improve and add utility to this project.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER - 1
INTRODUCTION

1.1 Introduction

Human activity is the continuous flow of single or distinct action essential in progression. Some specimen of human activity is a sequence of actions in which a subject enters a room, walk forward, sit down, stands up etc. Human activity recognition can widely apply to some real-world application like patient monitoring, surveillance of important location, activity-based search etc. and can be performed at the various abstract level. The human activity recognition is being studied by the student and the engineer and the student for a long time in every part of the world nowadays.

About 2.5 quintillion bytes of data get produced daily which is increasing day by day [**1**]. From this vast majority of data type, the video format is the most produced and monopolized format of them all. According to Google, the estimated YouTube server size will be 1 Trillion GB and about 400+ hours of video are uploaded on YouTube every minute [**2**]. Not only that, according to IDC they expect robust growth of surveillance camera market will be with CAGR of 12.9% for next five years with global revenue of nearly $49 billion by 2025 [**3**]

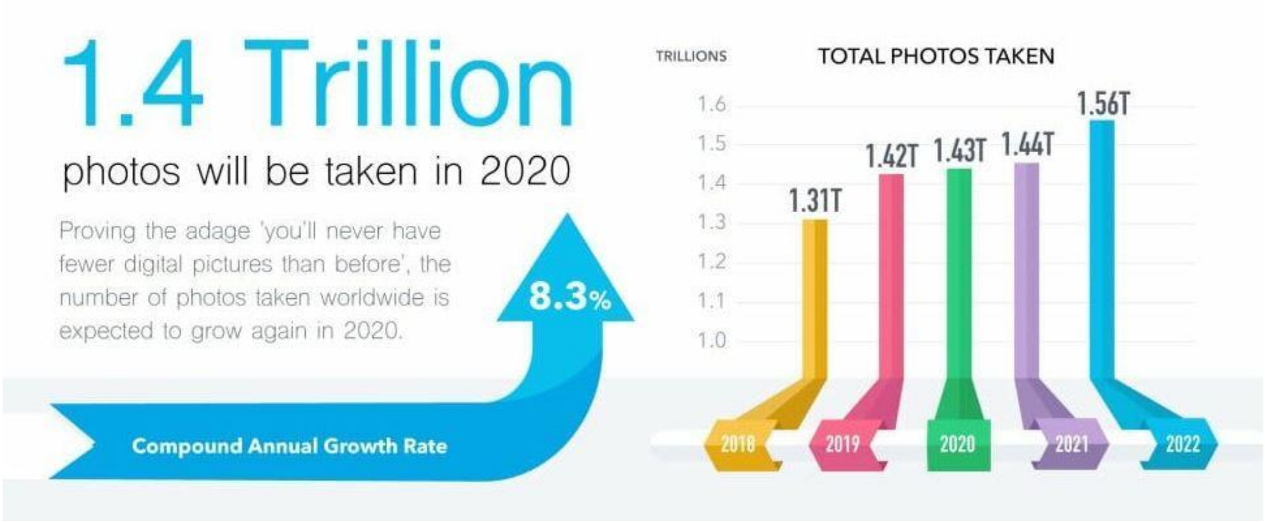


Fig 1.1: Estimated photo that will be taken near future.

While in a research blog it was published that humanity will be taking 1,436,300,000,000 photos in 2020 [**4**]. This huge amount of data is least processed which can be used after the process in a different form factor like surveillance, robotic vision, content-based video search and computer-human interaction.

In our project "Human Activity Recognition and Patient Monitoring System", we will mainly be focused on the various activities and detect these actions through video to monitor vital physiological sign. We have categorized patient's activity into three sections which are laying down, sitting, walking and standing. In this project, we will be using the YOLO (You Only Look Once) library to build a system that will detect human activity and monitor the patient. The YOLO library trains on image data and then adjusts action detection directly to be used in a project. In our project, we will be using the YOLOv4 as it is extremely quick and precise. The YOLOv4, mAP measured at .5 IOU, YOLOv4 is four-time faster and not only that we can change between faster speed and better accuracy by just changing the amount and data for the model, without any additional retraining of data required. [**5**]

1.2 Motivation

Human activity recognition which is a very critical monitoring system. The objective of human action detection is to inspect exercises from video successions or still pictures. During treatment, it is highly crucial for the doctor to continuously monitor the patient's activity and their vital physiological signs. That's why the patient monitoring system in the field of human activity detection has always been occupying a very principal position in the field of medical science. Not only patient monitoring, but this method can come in handy for security purposes in military quarter, Airport industrial factory etc.

The continuous improvement of artificial intelligence and deep learning algorithm not only helping us to transmit and get vital physiological signs to the medical personnel but also simplifies the quantification and as a result, rises the efficiency of the patient monitoring system. Human activity recognition can not only improve the patient in the medical sector but also can be used in a wide stage. The active or smart system can use HAR technology to monitor its residential area for better security purposes. [**6**] The aim of our research can also be to offer medical support, well-being services and health benefit to older adults and other security purposes for important infrastructure.

As we were studying Computer Science and Engineering in our final year, we already had some skills and knowledge to build a solution that can help us improve from a predicament. We decided to solve some of our social problems and we found out how to. We gather a group of students with the same peaceful purposes, "to monitor human movement" which can be revolutionary in the medical and other sector and the same project can also be used for security also after some modification.

After studying about the topic, a large amount of work has already been done in the video-based dataset, but a large amount of image that we have mentioned in the introduction has not been used rather used in the significantly lower number of datasets. This fact inspired us the most to work with a large number of images. But there was a catch as understanding motion and activity from some still image is quite a hard task.

This was very interesting for us because we were about to create an intelligent system that will detect human activity and monitor that activity intelligently. That's why we decided to took the challenge.

1.3 Rational study

Human understands better and faster when they monitor something with there own eyes. But there is a catch while solving this problem. Their camera view-point, the light condition, occlusions, complex background, long distance of human and low video quality makes this project more challenging. But we will try to compensate that problem with modern activity recognition library. In this paper, we will discuss the way to solve this issue.

Technology advancement has enabled machines or systems to understand or recognize the human actions from video or still image. In this modern era, it comes true that "Humans understand things better and faster by visualization". So, to compete with this modern world, there have already been a lot of works on realizing objects from the image. For erecting artificial systems computer vision is concerne with the notion and knowledgement so that it can gain information from images. In neoteric years, many deep-learning studies has already been done using powerful feature learning for activity recognition from vast-number of labelled datasets using many deep learning techniques. [**7**]

We have studied computer vision including fundamentals of image realization, feature detection and matching and also classification. We are focusing on this paper on action detection which is the recent topmost. If we take a glance at our daily life, what is happening every day we can see that many unethical, unsocial activities are occurring everywhere. Reducing those occurrences there developed various strict and innovative systems for halting them. After a survey in different hospitals and old age home also, we've come to know that there are lacking caring patients perfectly yet due to lack of enough staff, nurses, doctors moreover sometimes for over patients. So, if we can create an automation system in patients caring, it will be very progressive for a human being when there exists an artificial system where human activities are being recognized automatically from images. Added dataset can improve the projects detection perposes also. [8]

1.4 Research questions

While working on our project and writhing this research paper, we found some interconnected complication. Those question are:

- How will we collect all those data which will be needed for our research?
- With which approach we will process all those data?
- Which algorithm should be applied in our project for optimal result and accuracy?
- How can we improve accuracy and speed of our project?
- Can we update our project in the future according to our need?
- Will it be able to solve the current problem or patient monitoring?

1.5 Expected outcome:

From the study in this paper, our expectation to achieve efficient solution of:

- Detect particular and multiple action from a single project.
- Create dataset of common action of human being as a patient.
- Using object detection algorithm properly in action detection.
- Accuracy of our project over 90%
- Action detection can be used in video surveillance in a patient's cabin.
- We have completed all our expectation and the output of our accuracy and results are shown in the table.

1.6 Report layout:

This research paper consists of 5 chapters. The chapter 1 is for introduction, motivation and goals behind the research have been discussed in six sections.

- Section (1.1 introduction); Section (1.2 Motivation); Section (1.3 Rational study); Section (1.4 Research Question); Section (1.5 Expected Outcome); Section (1.6 Report layout)

Chapter 2, we covered the title background, prerequisites information relevant to the research into five sections.

- Section (2.1 Introduction); (2.2 Overview of HAR features and Related works); (2.3 Why YOLO v4); (2.4 Scope of the problem); (2.5 Challenges).

Chapter 3 illustrates the details of this research experiment which divided into eight sub-sections.

- Section (3.1 Introduction); (3.2 Research Subject and Instrumentation); (3.3.1 Our data collection procedure); (3.3.2 Number of Image and instances in the Training and Testing Dataset); (3.4 Data pre-processing); (3.5 Proposed Methodology); (3.6 YOLO detection architecture); (3.7.1 YOLO v4 Loss Function); (3.7.2 YOLO v4 Activation Function); (3.8 Implementation Requirement).

Chapter 4 announced the experimental outcome and discussion of that result into four sub section.

- Section (4.1 Introduction); (4.2 Experiments Result); (4.3 Compare with other platform); (4.4 Descriptive analysis of our result).

At chapter 5, we discussed about Impact on society, Environment and Sustainability of our research about "Human Activity Detection"

- Section (5.1 Impact on Society); (5.2 Impact on Environment); (5.3 Ethical Aspects).

At last chapter 6 additionally discusses into four subsections about title summary, recommendation and implication study

- Section (6.1 Summary of the study); (6.2 Conclusion); (6.3 Recommendation); (6.4 Implication for further research).

CHAPTER - 2
BACKGROUND

2.1 Introduction

“Human Activity Detection” is such an important and challenging area in the same action in a plethora and even by the same individual. Moreover, action detection stills hold out a challenging problem due to the camera viewpoints, noise, occlusions, composite-dynamic surrounding, long distance etc.

Using different types of sensing technology such as computer vision and more recently using deep learning human activity recognition becomes very popular day by day. Action detection framework has been divided into two components in a typical action detection: action classification and action representation. [**9**] Normally, converting an action video into a series of action is a part of action representation and inferring action label from that vector is action classification [**10**]

The recent improvement in deep-learning techniques, CNN, viz has proved their capability in object detection (recognition and classification) in images.[**11,12**] Human detection plays a prime role in many domains and approach including video retrieval, gaming, intelligent video surveillance, home behavior analysis, entertainment, autonomous driving vehicle, human-robot interaction, health care and ambient assisted living. [**13,14**]

Unlike image processing, video processing considers temporal aspects along with the image frames. Therefore, video processing is a 4D processing and challenging task. Sometimes, activity recognition analyzed by human operators from CCTVs recorded video to detect the potential threats associated with it but in this way, manual intervention fails to provide real-time threat detection due to massive video data analysis requirement from multiple cameras. So, we need to develop automated video surveillance to get real-time threat detection.

In this paper, HAR includes walking, sitting and standing. Here, we proposed YOLOV4 based architectures for HAR. The YOLOv4 model will be used for detecting objects in images or video frames.

2.2 Why YOLOv4

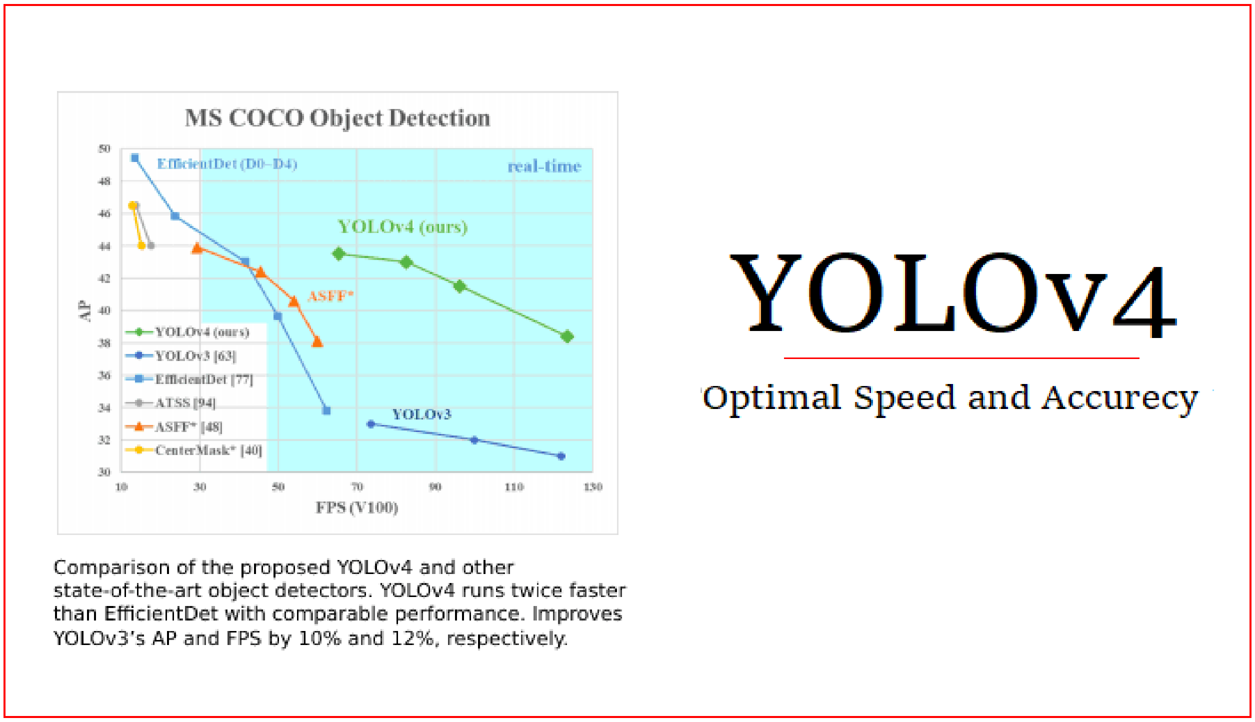
The full form of YOLO is You Only Look Once. There are a few versions out there but the YOLOv4 is the latest with the fastest and most accurate detection capability. The YOLOv4 detects objects in a significant way. YOLO didn't focus on the whole picture for an object or activity. It focuses on the main portion of a picture. The YOLOv4 applies a single CNN algorithm to the whole frame and divide the frame into matrix or grid. After this, the prediction for the action recognition boxes (bounding box) and the representation of the accuracy score comes into play.

The YOLO was developed fundamentally for detection and classification of object or an action. Surrounding box and class portend are made after one evaluation of the input image. YOLO is the best algorithm to detect object or action compared with R-CNN algorithm. R-CNN algorithm also detects an object. Both are the same category to detect objects but YOLOv4 uses a different approach.

The old method R-CNN algorithm focuses on several portions of an image and very difficult to detect an object. These algorithms focus on the full image, and then divide the image into some portions and portend surrounding boxes and probabilities for each portion. YOLOv4 is clever for doing object detection in real-time. YOLOv4 algorithm is best because it is working on real-time data. If we would know the YOLOv4 algorithm, first of all, we have to know what actually predicts it. To predict an object, it divides an object into 2 bounding boxes. YOLO algorithms divide an image into S*S grid. When the center of objects falls into the grid then the algorithm detects the object. To portend an object each surrounding box can be described using four descriptors:

- Center of a bounding box.
- Width.
- Height.
- Value is corresponding to a class of an object.

The main advantages of YOLO are it is fast and agile. In an appraisal, YOLOv4 acquire an ap value of 43.5 percent (65.7 percent ap50) on MS COCO dataset. YOLOv4 also pull off a real-time momentum of around 65 frames per second on a taste named Tesla V100, which beats the most agile and most accurate detector in terms of both speed and accuracy. [**15**]



Comparison of the proposed YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10% and 12%, respectively.

YOLOv4

Optimal Speed and Accurecy

Fig 2.1: YOLOv4 optimal speed and accuracy.

Its accuracy is better than R-CNN. This algorithm is the best algorithm to detect objects. YOLO can Detect 49 objects at a time. Its limitation to detect an object is 7*7 grids. YOLO algorithms increase the number of boxes and decrease the orbit for IoU for better results. It is faster to detect objects of another technique. If we use resize Image it gives us better results. YOLO algorithm Understand generalized object representation. YOLO cannot detect small objects because it focuses on-grid and portends one object.

If we differentiate YOLOv4 with the YOLOv3, The AP and the FPS have been escalated by 10 and 12 percent respectively. [**15**]

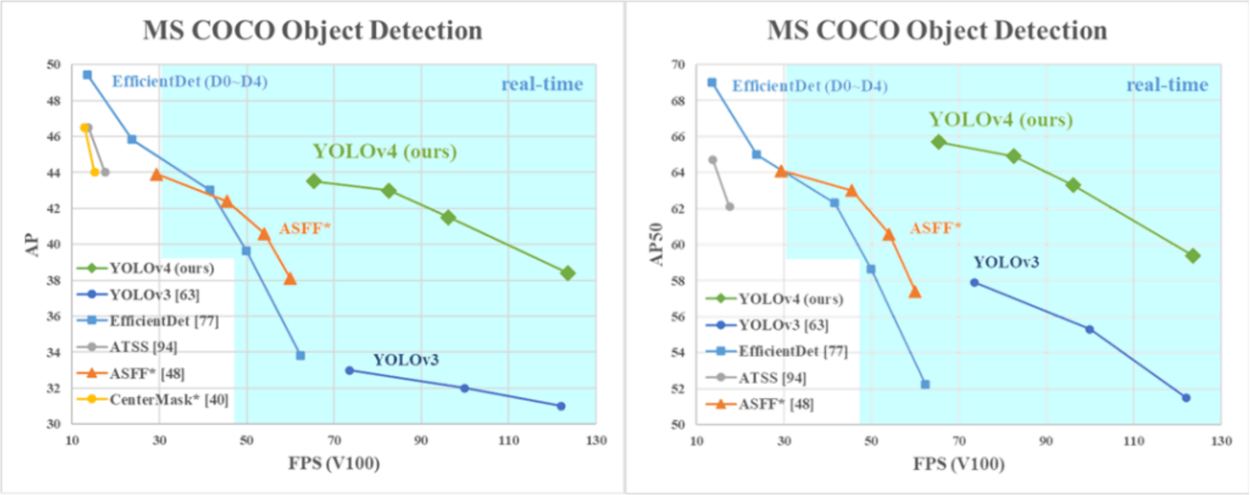


Fig 2.2 YOLOv4 vs YOLOv3

YOLOv4 have and excellent speed and accuracy and well-written documents which is a great contribution to the learners. The recent update from v3 to v4 also illustrate and encourage developing project with this open-source platform. With all those keeping in our mind, we chose YOLOv4 for our project.

2.3 Related work

Human action detection has already gained importance and been an interesting ongoing topic with regards to ongoing years because of its applications in various fields such as health, security and surveillance, amusement and intelligent environments. There also have been done a lot of work on human activity recognition and specialists have leverages various methodologies. For example, wearable, object-labelled and device-free to perceive human activities [18]. There has been such work like action-based, vision-based, motion-based, interaction-based, sensor-based etc. We can categorize approaches for HAR systems into two methods such as:

-Unimodal Methods represent human activities from data of a single modality, such as images. These include those approaches: space-time approach; stochastic approach; rule-based approach; shape-based approach methods [19].

-Multimodal methods combine features collected from different sources. Categories such as affective; behavioural; social networking methods [20]. Valentin Radu et al. [21]

proposed a multimodal RBM-based HAR with 81 % accuracy utilizing accelerometer and spinner signals.

Nowadays because of ease and advancement in sensor innovation, a large amount of the exploration field of HAR has moved towards a sensor-based methodology. In this methodology, various sorts of sensors are utilized to capture the conduct of people while they play out their day by day exercises [3]. Such as Fernando Moya Rueda et al. [22] had presented sensor-based HAR for walking, searching, picking so forth and accomplished 92.3 % accuracy by utilizing CNNs. There had been proposed by Zeng et al. [23] CNN-based HAR-based versatile sensor information and achieved 95% accuracy at that project. Moreover, in recent days, video processing systems in activity recognition have become high trust research areas that perceive various human activities like walking, jogging, running, opening doors, playing football etc. Michalis Raptis et al. [24] introduced a concealed hidden Markov-Model based Human Activity Recognition for six classes of activities, viz., jogging, kicking and so on and achieved the accuracy of 93% on full video. Also, Author Shinde et al. [25] achieved 87% accuracy on his proposed method using object detection engineering that has been utilized continuously in video processing as a lot of events were captured. Shubham Shindena et al. [26] presented an approach to detect, localize and recognize actions of interest in almost real-time from frames obtained by a continuous stream of video or from a video surveillance camera. In SVM and random forest-based machine learning techniques for HAR recognition, there done some work such as using acceleration generated by using cell phone [27,28] by Akram Bayat et al. [29] and achieved an accuracy of 91%. Keeping all these methods in mind for better accuracy, we proposed an object detection method using YOLO v4 on both still images and video surveillance.

2.4 Scope of the problem

Using this automation Human Activity Detection system, doctors can easily take observation of multiple patients at a time from their chamber or comfort zone also. Doctors also can keep an eye on the duty nurse or staff in the patient's cabin. What are they doing and are they doing their duty perfectly? This same system can also be in used for a lot of different purposes mentioned above.

In this paper, we implement the Human Activity Detection system in both still images and video with mentioned three human actions. Further processing we will be adding more action types. Adding more action type will add more variety.

2.5 Challenges

In the beginning, we needed lots of images to build the system. So, first, we met the challenge to collect image data. We collect about 4,674 pictures of the mentioned three actions for the image dataset.

The second challenge is to find a perfect object detection algorithm to face this issue. We use YOLOv4 basic algorithm for action detection. The big challenge is to find out which techniques work better on this dataset.

Chapter 3
RESEARCH METHODOLOGY

3.1 Introduction

Human activity detection plays an important role in this modern technology era. It’s a large field for research. Nowadays it has become a rising topic in the human interaction area. From the past decades, many researchers are working on this topic. Computers don’t have their own brain to detect anything. They can't read the humans mind. They only give us the output for what we trained for them. If the computer can understand the activity of humans, it can bring a lot of positive changes in the field of IoT. Nowadays HA is creating big chaos in the technology field.

In this chapter, we will discuss a more theoretical approach to this research. This chapter may come in handy to understand the main notion of this research work. First, we will discuss the aspect of HAD & then we will discuss the deep learning approach with data that we gathered from a different situation. We will end the chapter by giving a proper understanding of our implementation and requirements.

3.2 Research subject and instrumentation

Our thesis and research topic are “Human Activity Detection”. It can be interacted and implemented with the various algorithm & field of deep learning, machine learning, Image processing and neural network.

We will python programming language to implement our algorithm. We use YOLOv4 approach for better and faster detection. We use Google Colab for free and fast GPU acceleration which will speed up the data training.

3.3.1 Data collection procedure

We had to collect a lot of data in different condition, complex background and surrounding along with the hospital and other environment. So that, this project can give us the best accuracy at any location. That's why we divide into different group. We try to capture as much data as possible by divided into three group.

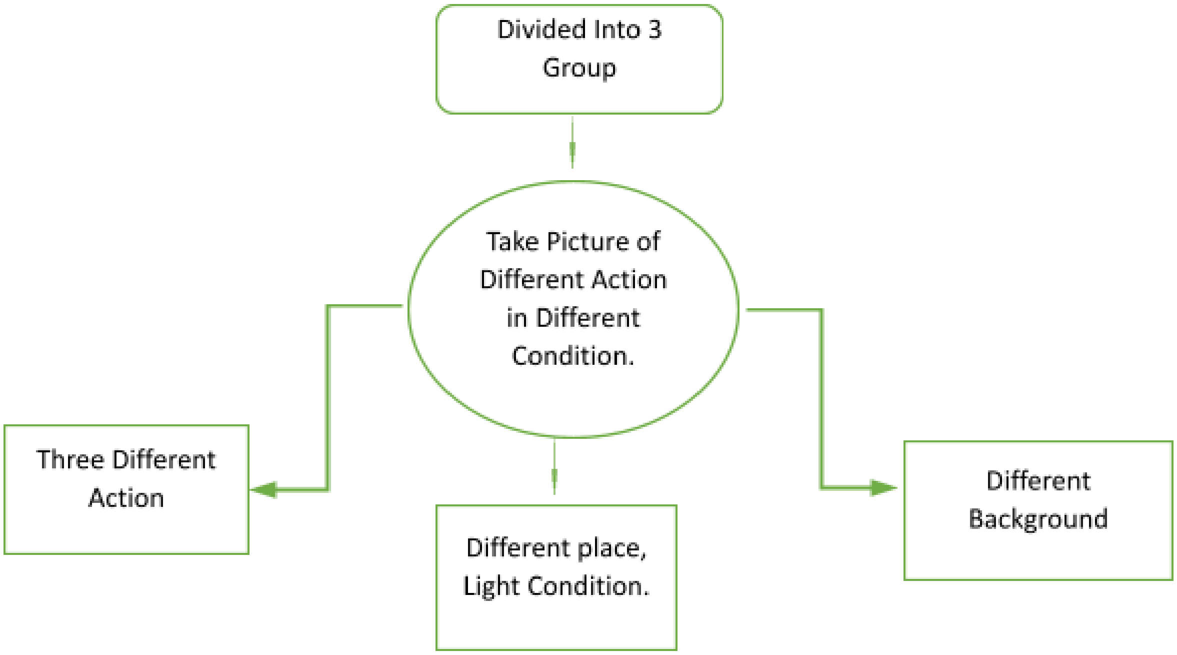

Divided Into 3 Group
Take Picture of Different Action in Different Condition.
Three Different Action
Different place, Light Condition.
Different Background

Fig 3.1: Data collection process

There are a number of factors works while collecting data. We gather a lot of action by acquiring video and image data and then by separating frame from those videos. We gather data from different light condition, noisy background and with a lot of different angle of view.

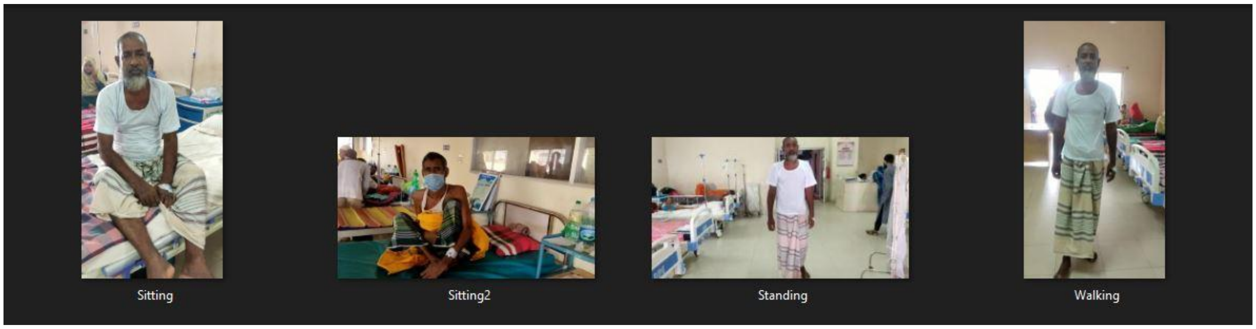


Fig 3.2: Data collected from different environments

To include more accuracy, we had to add more data from other critical and complex background. That's why we also chose to gather data from other condition.

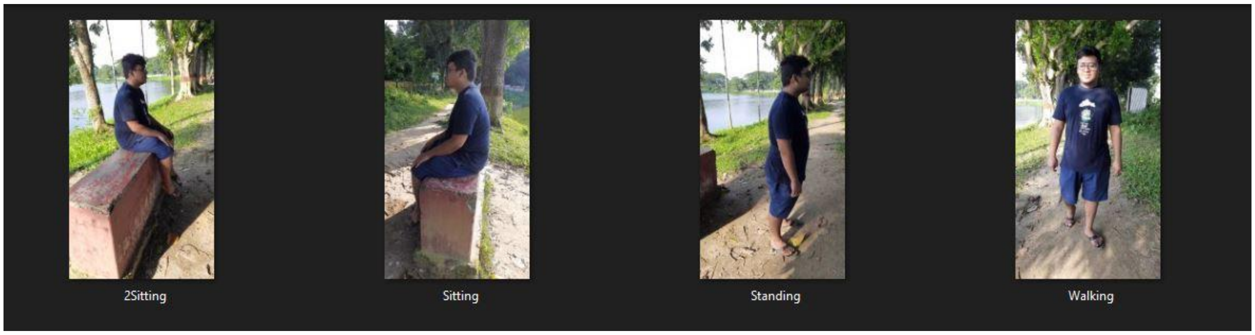


Fig 3.3: Data collected from different environments part 2

Following this way, we have gathered around 4,674 images to develop our project.

3.3.2 Number of image and instances in the training and testing dataset.

In this chapter we will provide the instances or the amount of our model data-set. For our model, we have gathered an amount of 4,674 data. For instances of class "Standing" we had trained about 1838 Images while for "Sitting" & "Walking" instances, we had trained 1705 and 1131 amount of images/data respectively.

In terms of percentage Standing, Sitting and walking class have a data set of 39.3239%, 36.4783% and 24.1976% respectively.

Dataset	Training Dataset	Percentage
Total Number of Image	4674	100%
Instances of class "Standing"	1838	39.3239%
Instances of class "Sitting"	1705	36.4783%
Instances of class "Walking"	1131	24.1976%

Table: 3.1: Amount of our dataset

3.4 Data pre-processing

As in this paper, we are only focusing on action that occur in the hospital or our surrounding, we need to preprocess all those 4,674 data/images that we've gathered. To prepare or train data, we need to classify each data in separate action.

For this reason, we used labelim. That create a data file which include the information about the image size and all the action value.

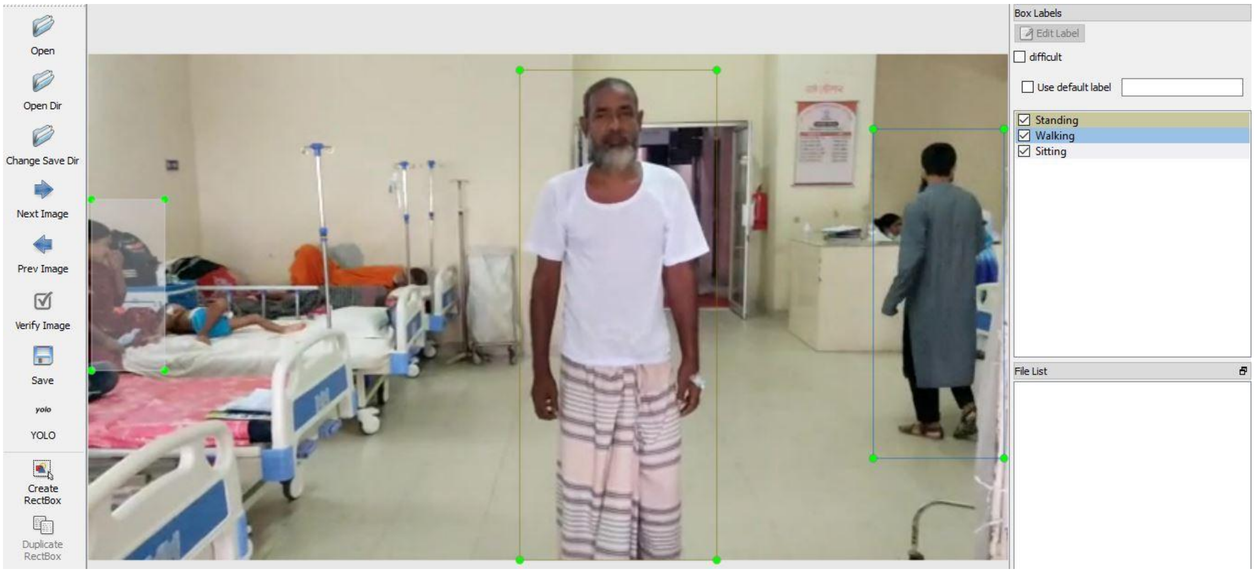


Fig 3.4: Preprocessing data

This way, we save the image information into a single .txt file extinction. YOLOv4 use that .txt file for recognizing each action in that image. Those data pre-processing process are given bellow.

- Require total no of action class. In our case, we had three action class.
- Text file with the same path as all other image which we will train later.
- Text file with the proper naming of all action class.
- The main path that will contain the weight file.
- A configuration file with all layer described in YOLO architecture chapter.
- Pre-Trained YOLOv4 involutional weights.

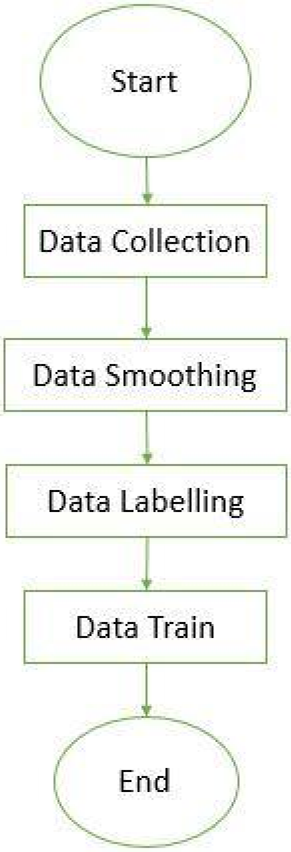


Fig 3.5: The flowchart for training and pre-processing in YOLOv4

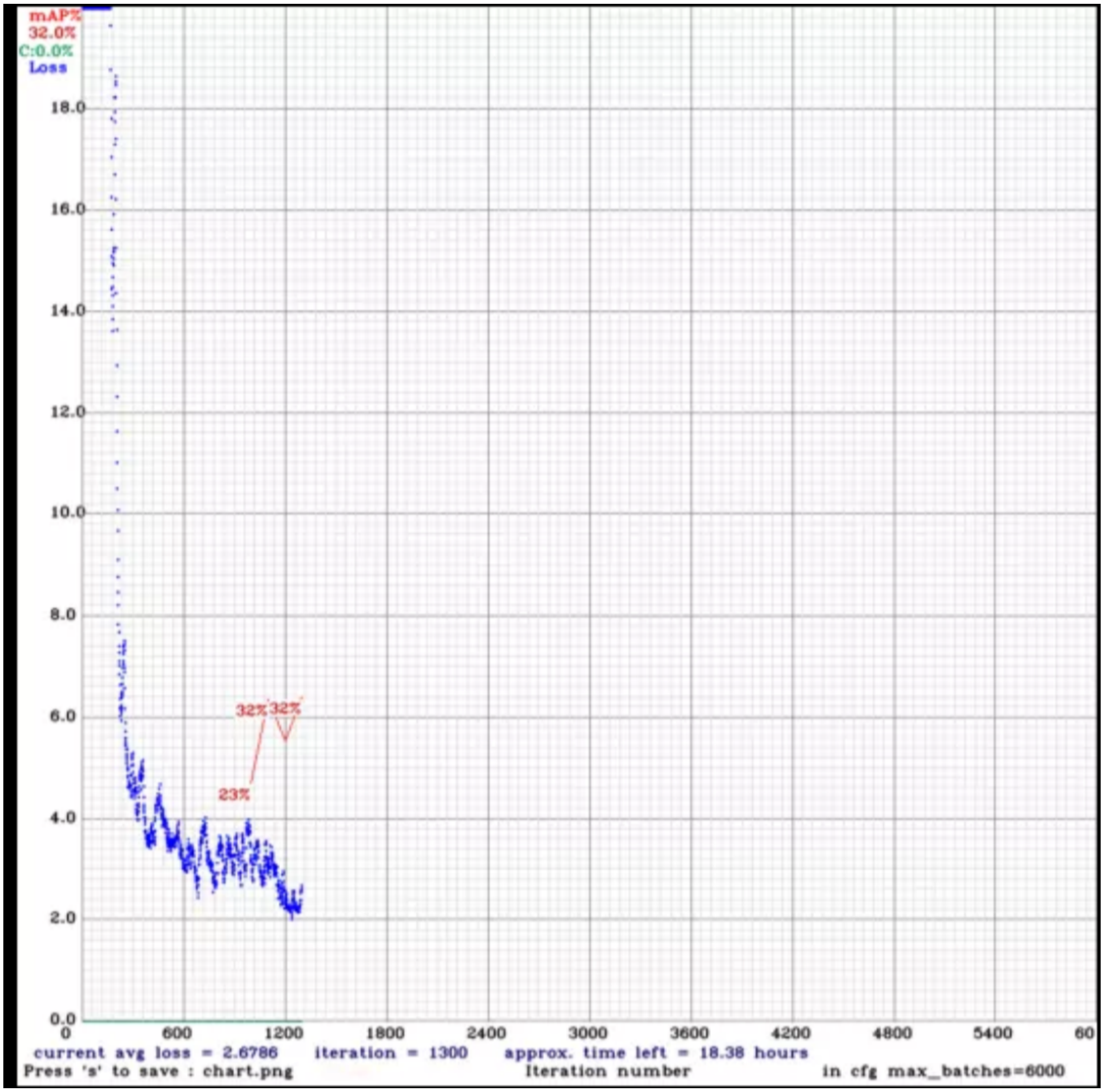


Fig: 3.6 YOLOv4 CFG weight

Fig 3.6 indicate the CFG diagram for training our model. This CFG model was captured when our system finished iteration of 1300 cycle. That moment, current average loss was 2.6786 which was significantly less than others.

3.5 Proposed methodology

The main work-flow that we implement in our project/paper is already explained in the fig 3.5. The method we used for labeling, training and testing the HAR model is You Only Look Once v4 (YOLOv4) which were well explained in details before. The dataset we used were collected from various situation which we named "Human Activity Dataset"

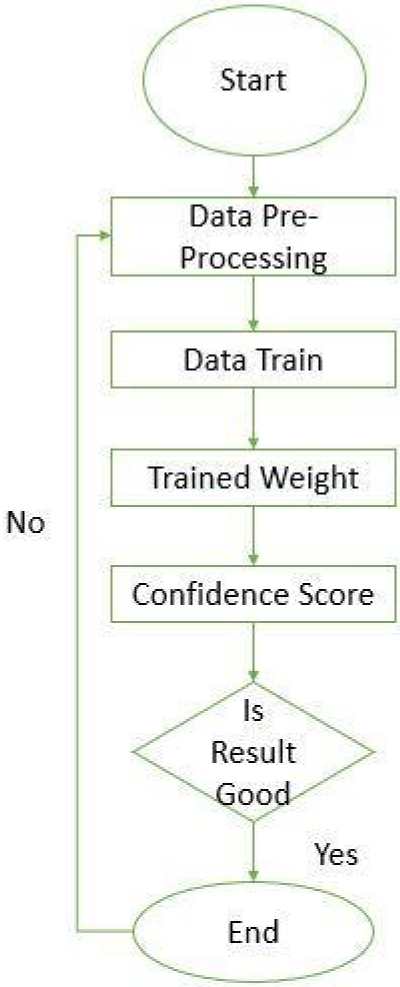


Fig 3.7: Working methodology flowchart

Given above is the main working methodology flowchart. That's the flowchart we followed to get a satisfactory result. We preprocess our collected data again and again for a satisfactory level of result.

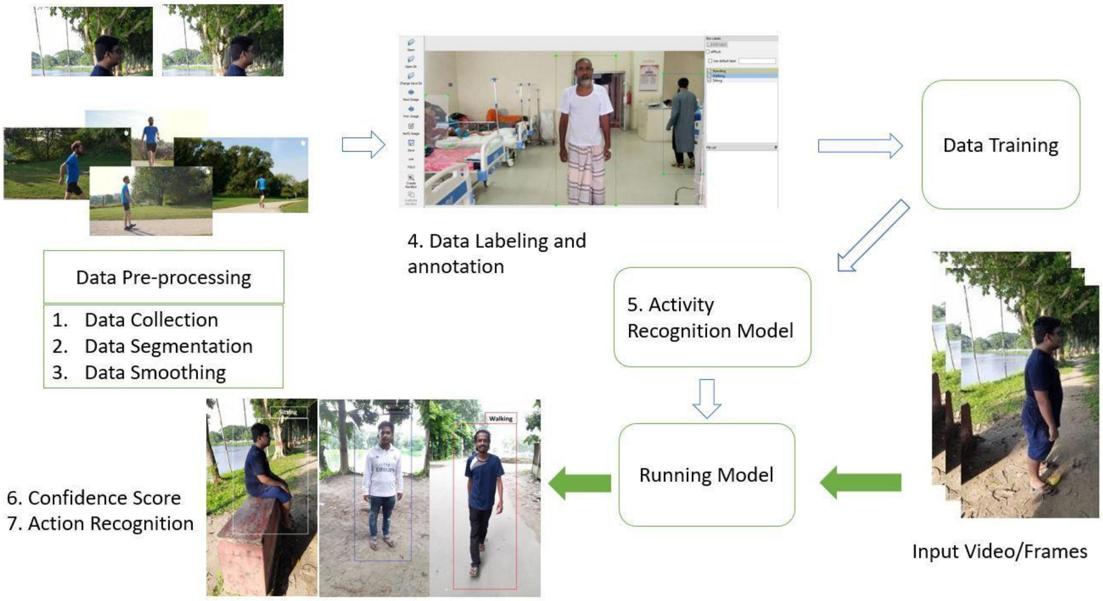


Fig 3.8: Working methodology

First, we trained our model with the congruous action. For testing the model, we input the video frame for the localization and then recognition of that action. The model will be able to recognize and then give the accuracy no.

3.6 YOLO detection architecture

In this topic, we will discuss farther about UOLOv4 along with the architecture. All of the YOLO data models are activity/object detection dataset. Those datasets are trained so that it can search for a subset of object class [30]

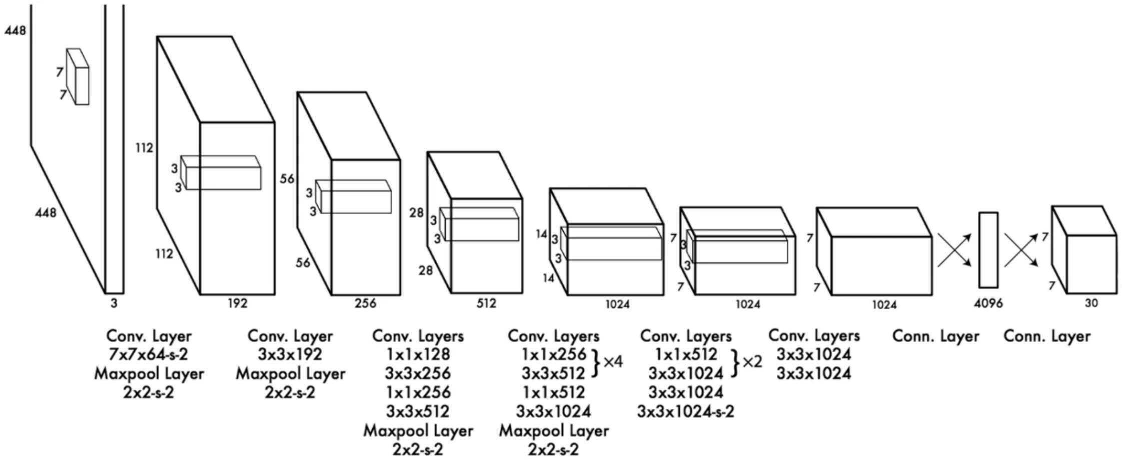


Fig 3.9: YOLO architecture

Most people in the research field still used to the YOLOv3 which was already giving us an excellent result. But the YOLOv4 had improved the fidelity and momentum of the two main attributes we generally use to qualify how the architecture and algorithms perform. [31]

The YOLOv4 is further improved approach of object detection. This applies a single CNN to an entire frame collected from video or just captured by a camera into the grid. After this, Prediction of those bounding box, then classify them into object or action and finally calculate the confidence score in a grid view. The main Architecture of YOLO has 24 conventional layers along with two associated layers. [32] YOLO takes an input image and then reside that frame into 448*448 pixels. Then the frame gets pre-processed through the conventional network. Then tensor gives accurate information about the coordination of the bounding box and the probability distribution of overall classes and attributes the system is trained for. [33]

3.7.1 YOLOv4 loss function

In our research paper, we've calculated the loss function of YOLOv4 which minimize the actual normalized distance between the dataset frame to target frame which was able to bring out convergence momentum which is more inch-perfect and fastened.

The DloU misfortune is on the fundamental of IoU (Intersection over Union) which signify the middle distance of that particular bounding box which indicate the following formula (1) where Bgt is the target bounding box and B is the prediction box. Hare, the loss of function LoU is defined in formula (2); It shows the work function if bounding box overlap else there is no overlap if the gradient does not change.

$$\mathbf{IoU} = \frac{B \cup B\hat{}gt}{B \cap B\hat{}gt} \tag{1}$$

$$L_{LoU} = \mathbf{1} - \frac{B \cup B\hat{}gt}{B \cap B\hat{}gt} \tag{2}$$

That's why the GIoU function improves the loss of Lou in the case that the acclivity doesn't change while not overlapping another box. Which add some loss term for the function IoU. It is noted as the equation (3). If One of B or Bgt countermands the other bounding box, the penalty terms will not work, which can be defined as an IoU loss.

$$L_{GloU} = \mathbf{1 - IoU} + \frac{|\,C-B \cap B^{gt}\,|}{|\,C\,|} \tag{3}$$

To solve constraint, DIoU function was set on the motion which can be seen in the formula (4); where b and bst indicate the center point of that anchor image and targeted image one after another while the attribute p indicates the Euclidean distance between two centers while c represent the minimum rectangle distance that covers anchor and the targeted box. The loss function of YOLOv4 for DIoU can be noted as equation (5)

$$R_{DIoU} = \frac{p^2\,(b,b^{gt})}{c^2} \tag{4}$$

$$L_{GloU} = \mathbf{1 - IoU} + \frac{p^2\,(b,b^{gt})}{c^2} \tag{5}$$

After those, The CIoU is also introduced to our loss equation. There are a few upper hands in the CIoU loss function. (1) Those are this equation can increase the overlap area in both the ground truth box and the prediction box. (2) This equation can minimize the actual distance for the focal point.

The equation CIoU can be noted as formula (6) based on the formula (4) which can be calculated; where α represent an upper hand trade-off and “v” is called in formula (7) to measure the stability of the overall aspect ratio added.

$$R_{CIoU} = \frac{p^{2}(b,b^{gt})}{C^{2}} + \boldsymbol{\alpha\upsilon}, \tag{6}$$

$$V = \frac{4}{\pi^{2}} \textbf{(arctan } \frac{W^{gt}}{H^{gt}} - \textbf{arctan } \frac{w}{h}\textbf{)\^{}2} \tag{7}$$

The loss equation of the CIoU can be defined as for formula (8), and the main variable can be called at equation (9)

$$L_{GIoU} = \mathbf{1 - IoU} + \frac{p^{2}(b,b^{gt})}{C^{2}} + \boldsymbol{\alpha\upsilon}, \tag{8}$$

$$\alpha = \frac{\upsilon}{(1 - LoU) + \upsilon} \tag{9}$$

In this chapter, The DIou loss equation could be applied in NMS (Nonminimum Suppression) to delete unneeded bounding box. In this equation, both the distance of the action detection box and the center point of that bounding box is considered which can constructively above two-loss equation mistakes. [34]

3.7.2 YOLOv4 Activation Function

Activity Function or equation that runs on deep learning neural network which is in charge for mapping the input of the neuron to the output. Its main task is to expand the nonlinear change of the neural network dataset. In figure 3.7 the function is shown.

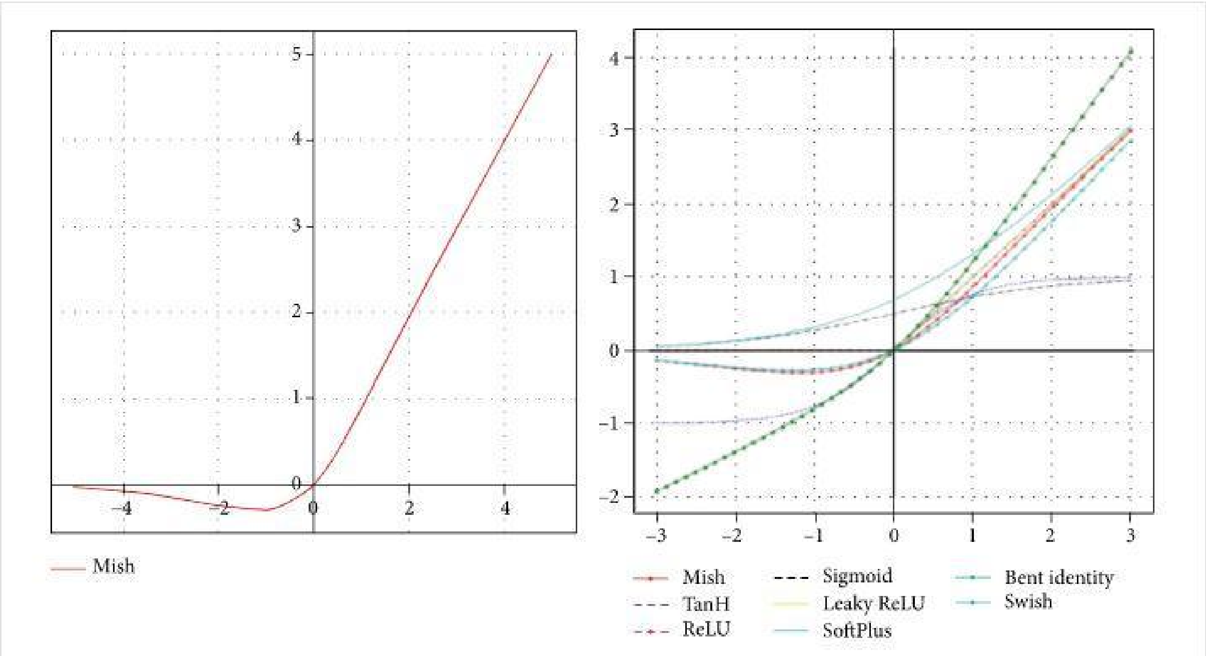


Fig: 3.10: YOLOv4 activation function

In our model, the Mish activation obligation is the activation function which we utilize which replace Leaky Relu that is a very tiny constant leak that has the much-updated function of Relu with Mish in YOLOv3 [34] Leaky Rule is the self-regular non-monotone deep learning neural activation and smooth activation function allowing the instruction into the deep learning neural network to obtain for preferable accuracy and generalization. It can be defined as equation (1) wherein that equation, it shows that c(x) $= \ln(1 + e^{x})$

	$F(x) = x. \tanh(c(x))$	**(2)**

This equation is specific then swish defined as equation (3) and Relu defined as formula (2) when performing on the experiments.

	$F(x) = \max(0, x),$	**(2)**
	$\mathbf{F(x) = x. sigmoid(x)}$	**(3)**

3.8 Implementation requirement

In this topic, we will discuss about system requirement to implement our project. This is the system we use to run our project titled “Human activation and patient monitoring system”. Some of the system requirement and essential developing environment is given bellow.

Implemented on system:

- Core i3 9th Gen (quad core 3.4ghz)
- 8 GB Ram
- NVDIA GeForce 610 GPU
- Windows/Linux/Mac OS
- Minimum 10 GB free space in SSD/Hard Disk

Recommended hardware and software requirement:

- Core i5 10th Gen or batter
- 16 GB Ram
- NVDIA GeForce 1060 or better
- Windows/Linux/Mac 64bit
- Fast NVMe SSD with 20 GB free space.

Developing Tools:

- Python Environment
- Anaconda Prompt
- Google Collab
- Lebellmg
- TensorFlow
- NumPy
- Darknet

Chapter 4
EXPERIMENTAL RESULT AND DISCUSSIONS

4.1 Introduction

In this chapter, we will talk about the experiments and the results on the groundwork of our test data to find out and compare our model's accuracy. For this purpose, we gathered and trained our system with a huge amount of data collected from different situation from which we can absorb the important information and features of actions detection from training dataset.

After that, we did all the improvement we could to get better speed and accuracy for our detection purpose. Better computing system will produce better result in this case. We get the average accuracy of about 94% after training those data 13th time. This accuracy can be even improved and mainly depends on interaction we perform with our taste subject.

To gather and compare the result of our experiment, we had to calculate the reliability and fidelity of The Human Activity Recognition and Patient Monitoring System. We have collected those data in numerical value from our numerous tastes and then we use the Average Accuracy and the Mean value for further comparison with another model. The Average Accuracy is used to calculate independent action's unassisted while the Mean Average Accuracy is used to calculate the model's fidelity combined.

4.2 Experiment results

For activity detection process, most of the time we need to work with complex background. For this reason, action detection is a difficult task for any detection model.

YOLOv4 is the most sophisticated algorithm that handle that part easily. We trained our model or system to detect three action class; Those are "Standing", "Sitting" and "Walking". After running and optimizing our model a few more time, our model is performing better than any other action detection works out there. Given bellow is the test data for detection of action from still image.

Action Class	Average Accuracy	Mean Average Accuracy
Sitting	95%	94.6667%
Walking	96%	
Standing	93%	

Table 4.1: Validation set accuracy for Still Image

Action Class	Average Accuracy	Mean Average Accuracy
Sitting	58%	63.00%
Walking	70%	
Standing	61%	

Table 4.2: Validation set accuracy for video file.

4.2.2 Comparing with others platform

While working on our research project, we also try two more approach. But mainly focused on the YOLOv4 for the best accuracy and performance.

TensorFlow is an Open Source Machine Learning Framework of Google for the programming od data flow across a variety of task. Tensors are just multidimensional arrays, and expansion of 2-dimensional tables to higher dimensional data. TensorFlow has many characteristics that make it appropriate for human activity detection.

Creating a reliable Machine Learning (ML) models which are capable of understanding and localizing multiple activity in a single image remained a key challenge in computer vision. But, with recent advances in Deep Learning, Activity Detection are simple to build then ever before. The object/activity detection API of TensorFlow is an open source platform developed on top TensorFlow that makes it simple to build, train and deploy models for activity detection. We build our first activity detection model with TensorFlow activity detection API.

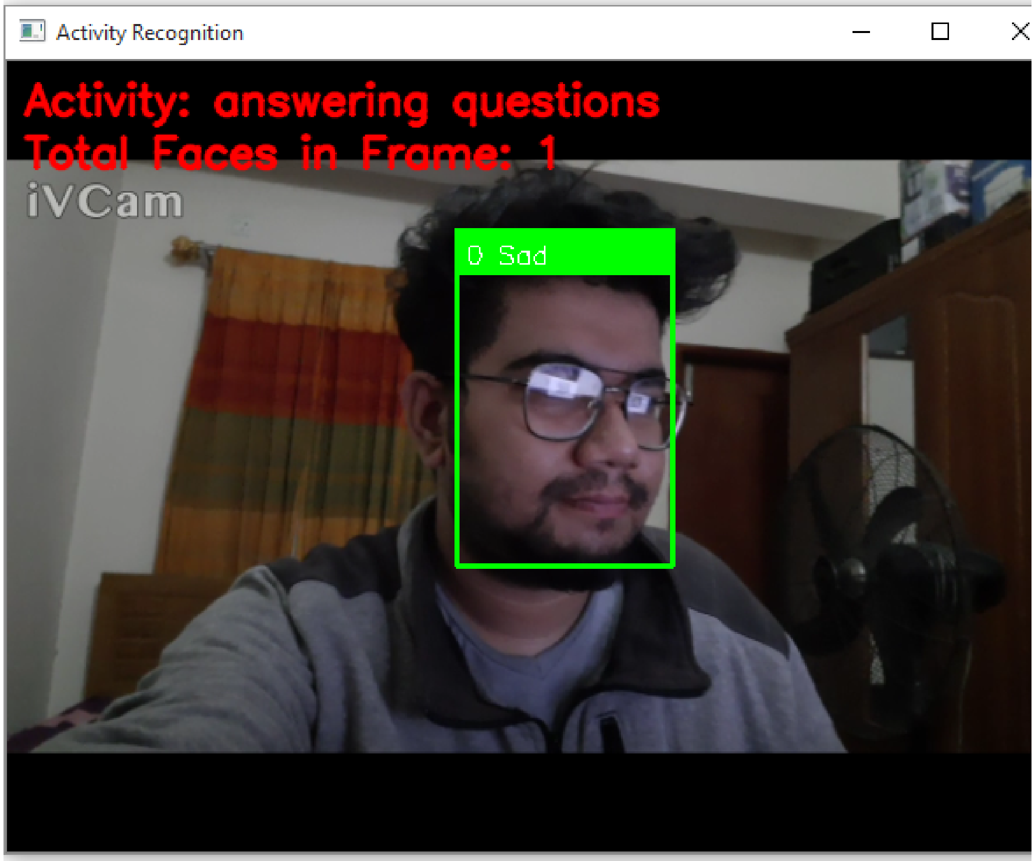


Fig 4.1: Human activity detection using TensorFlow

In the fig 3.11 we can see that, TensorFlow can be able to properly detect human's activity with great accuracy.

OpenPose is the first multi-person real time platform to collectively detect key points for the human body, hand, face and foot (135 key points in total) on a single image. Researcher at Carnegia Mellon University suggested this method as well. I the form of python, C++ implementation and unity plugin, the published this.

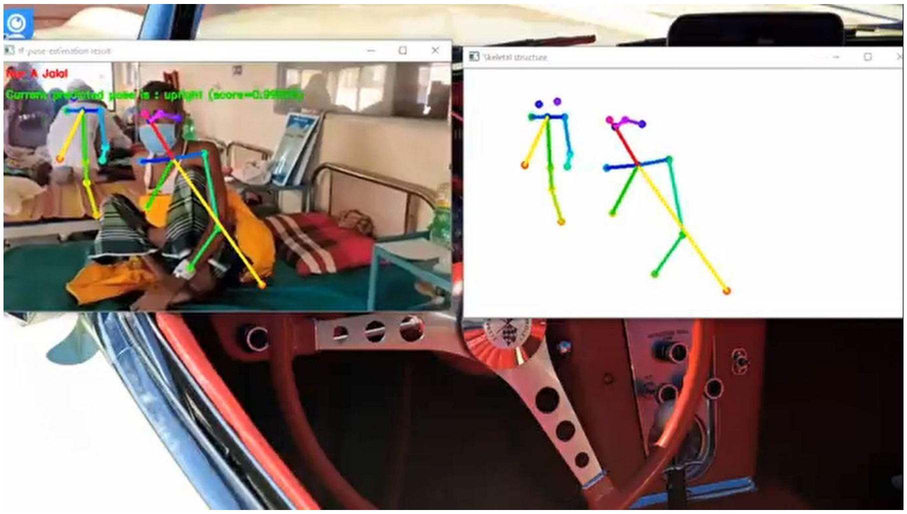

tf-pose-estimation result
Current predicted pose is : upright
Skeletal structure

Fig 4.2: Human activity detection using OpenPose

Using the OpenPose, pose estimate & detection has been minimally implemented. The MobileNet (a CNN originally trained on the ImageNet wide visual detection task dataset), or binary classification of poses, (sitting or upright), was retrained (final layer) on a data set. In the bellow, we will compare those above-mentioned models of ours.

We have run all those models and then gather data to find out the most efficient approach. Those measurements are given bellow.

Attribute	Project -1	Project-2	Project-3
Training Platform	TensorFlow	YOLOv4, Danknet, OpenCV, NumPy	OpenCV OPENPOSE, Motplot, NumPy
Data:	Google Pretrained Weight	Manually anoted- 6000 Data 2000 Data 1000 Data	Live Data using 2D Mapping
Detection Platform	Local + Live	Local Live (possible)	Local + Live
Accuracy	Overall: 80%	1000 Data: 85% 2000 Data: 95% 6000 Data:	Overall 90%
Status	good	Strong	Strong

Table 4.3: Comparing YOLOv4 project with others platform.

Given below are the advantage and disadvantages of our three models.

Attribute	Project 1	Project 2	Project 3
Advantages	1.Can detect emotion.	1.Faster training time using Colab	1.It can be implemented for prediction model.
Drawbacks	1.Run slow on GPU and CPU. 2. Low frame rate 3. No prediction on next activity. 4.Requre high performance GPU	1.Many Data Require 2.Too many activity trains overlap the detection. 3. Require very high performing GPU and CPU for local PC test.	1. Require High configure PC. 2. Low frame Rate.

Table 4.4: Comparing advantage and disadvantage of YOLOv4 project with others platform.

From the above result, we can say that all of those models have their own advantages. But the activity detection model using the YOLOv4 has the most meaningful purposes. It can be implemented with less configuration on Colab unlike others.

4.3 Descriptive analysis of our result

Before training the data, we had divided them into three part. For first 1000 data, we named it test data 01, secondly, for 2000 data; we renamed the data set to test data 02 and finally we trained all the data we had in our disposal. After training with largest dataset, the system was giving us the best accuracy.

In the Table 4.2.1 we can see that, the test accuracy for walking was about 96% which had the highest average precision in our test model. Given bellow is the output for action walking in still image detection.

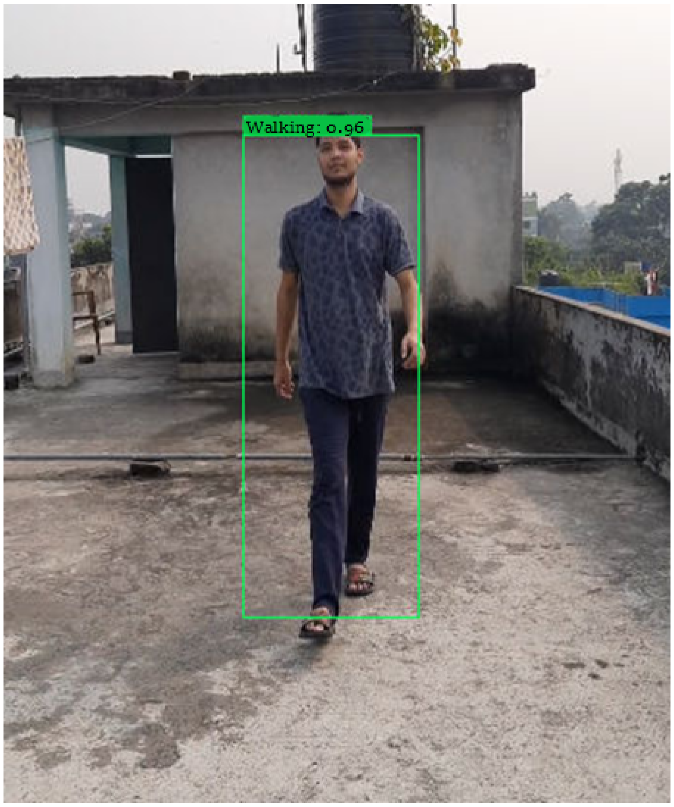


Fig 4.3: Action detection of walking

In the data table 4.2.1 we can see that, the action class “Sitting” have the second highest average accuracy of over 0.9500. This value indicates that, our action class “Sitting” have precision of over 95%. Let’s see our detection for action class sitting.

Finally, for the standing class which perform an average precession of 0.9300. Which means, this action has average accuracy of 93%. This action class had the lowest accuracy of all our action class. Let’s see the result image.

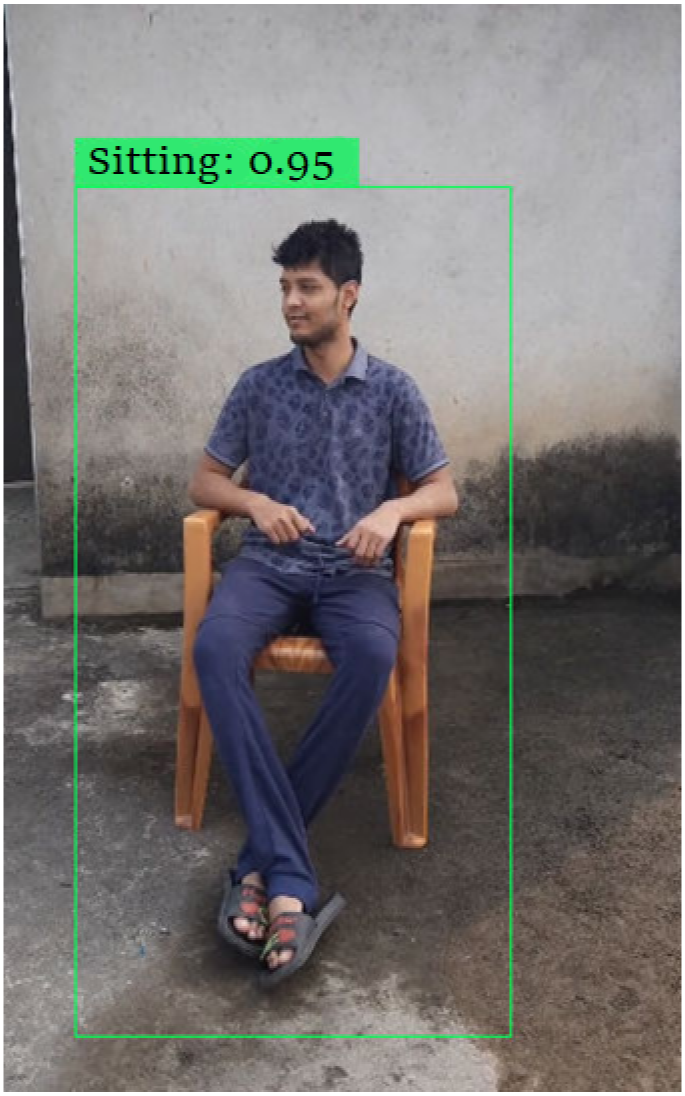
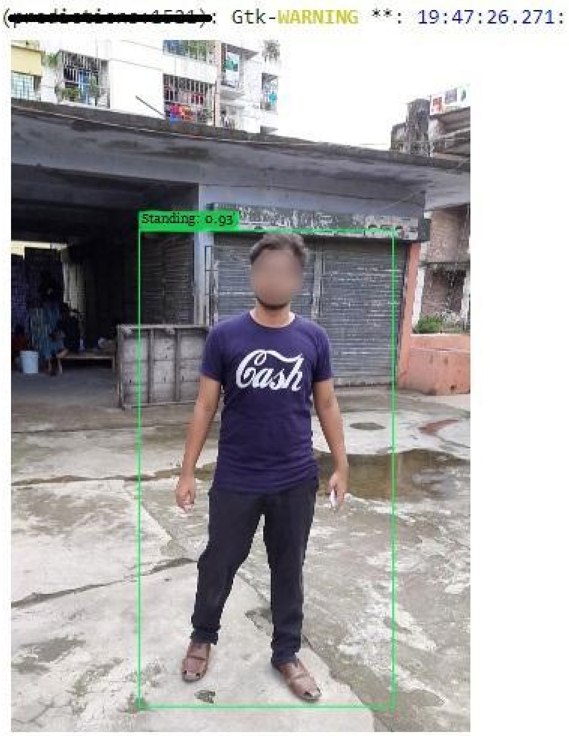


Fig 4.4: Action detection of sitting

After calculating, we get the average accuracy of 94.6667% which is on the top of the line if we compare with other research paper.



```
([redacted]): Gtk-WARNING **: 19:47:26.271:
```

Fig 4.5: Action detection of standing

As, we can see that we get the highest accuracy on every action recognition class in the still images. Our trained model can even detect all action even in the most complex and congested images where there are a lot of people. Let’s see the accuracy of our project in video feed.

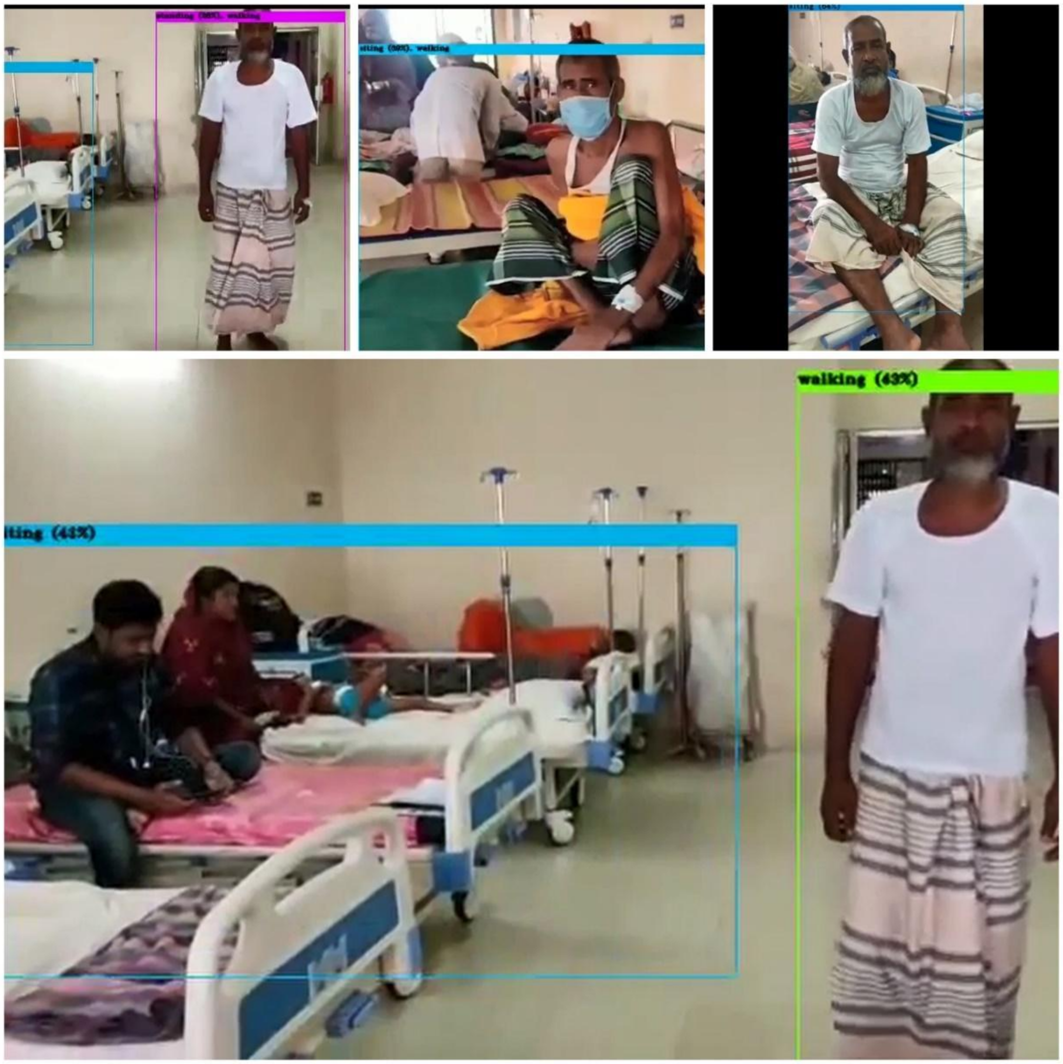


Fig 4.6: Action detection from video

From Table 4.2.2 we can see that our action class sitting have accuracy of 58% while walking and sitting have the accuracy of 70% and 61% respectively. The average detection accuracy was 63%

Chapter 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on society

The social consequences of video surveillance system are very extensive.

Human activity detection for many kinds of control applications is commonly used in healthcare and assisted living facilities. Protection and privacy are two promising considerations that match the efficiency and validation of video surveillance system with the caliber of applications for patient monitoring. Human activity detection is a system that typically uses a network of camera to control protection against theft, violence, terrorism or other similar problems in a particular area (private or public)

In society, this model can profoundly affect all the general population. It will help both the destitute and all kind of individual. So, through this model, we can solve some of the social problem mentioned above.

5.2 Impact on environment

Due to the rise in the trend of digitalization, our society is completely dependent on technology and its application. Human activity detection is also one of those technology. Now, it’s impossible to separate technology and human life. The regular and extensive use of technology has also affected people’s need and demands. Technology can be used in the life of an individual to perform distinct and routine task including traveling, walking, sitting, communication, learning new things, business or comfort.

Human activity detection is a set of approaches that can be used in a wide variety of application, including smart home and healthcare. Deep learning technique such as recurrent neural network and CNN (one dimensional coevolutionary neural networks) have recently been shown to provide state of the art outcome on difficult task of activity detection. But when in terms of environmental impact, this approach has little to no devastation.

Since we have already seen the damage the technology has created, we were well aware of this. Technology influence the way people act, exercise, interact, learn and think. It helps society and determines how individuals interact on a daily basis with each other.

During our research, we keep that in mind that we were also the one to combat the climate change as well as its effect.

5.3 Ethical aspect

Detecting human activity means paying another person near and continuous attention. In addition, the model of activity detection not only pays attention to anyone but also for a specific reason, to pay attention to some entity (a individual or group) which create an ethical situation.

Before working on research, we also need to look for what impact it will contribute toward our society. Human activity detection with the purpose of surveillance is the initiative that sometimes works with or without the authorization or dispensation of a person or an individual. The main purposes maybe for better management of the particular institute or the safekeeping of that person or other purposes.

But this creates an ethical in the world of computer science. The ethics for such action detection considers on the basic principle of how the surveillance system is implemented. That's why the surveillance system must be used whenever it is required to.

Despite posing a threat to privacy and the risks of its abuse by officials, it is difficult to ignore the utility of this system. What is most critical is that these systems must be built in ways that not only preserve privacy and liberty while protecting individuals from security risks, but must also be able or detect abusive uses through the use of technology such as logging, encryption and authorization control mechanisms.

In general, an activity detection system is considered a powerful tool to combat crime and protect individuals and property from harm. The integrated use of video monitoring with a rapid response police forces expected the police's scope and allow for a reduced cost of better policing. The legal implication of the activity detection are very large. It is crucial that the use of video surveillance in the community where it is implemented in both lawful and appropriate.

Human activity detection ethics takes into account the legal implication of how surveillance is employed. Other than that, our project can be impactful in society as it can be helpful in a huge amount of application. For the social realm, our activity detection model can presume a strong viral example without most of the ethical issue.

5.4 Sustainability plan

Sustainability plans are basically the roadmaps for long term goals, planning and strategies. It can be emphasized in three ways like the sustainable relationship between organizations, the values that project promotes and the sustainability of services.

The notion of sustainability is mainly depended by three fundamental point, which are economic, environmental and social which are also informally know as profit, the earth and humans. It allows the generation in the future to fulfil their required needs.

There are a few sustainable development goals that we can follow for our model. Sustainable growth is a social problem rather than an environmental one. Significant changes in human potential are needed through education and healthcare reforms. In human-to-human interaction and interpersonal relationships, human behavior identification plays an important role. It is difficult to extract because it contains information about a person's identity, their personality, and psychological condition. This research paper can be beneficial for next sustainable research in those above-mentioned research field in the upcoming future. So, our research can be sustainably developed for the betterment of everyone’s future.

Chapter 6
SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FURTHER RESEARCH

6.1 Summary of the study

We are attempting to speak to a profound learning strategy with YOLO v4 in Human Activity Detection. The entire exploration short summary is given below Stage:

Stage I:

- Data collection
- Data Annotation

Stage II:

- Data training and labeling
- Data validation and test set

Stage III:

- Detect action using YOLO v4

Stage IV:

- Calculation
- Result and discussion

6.2 Conclusion

In this research paper, we assessed a continuous methodology for human movement identification, picture arrangement dependent on YOLOv4 (You Only Look Once) from complex scenes. The techniques approved with our difficult dataset where are many jumble and uproarious information for checking more exactness. It can recognize more than one individual's various exercises utilizing additional jumping encloses a solitary picture. Different activity recognition methods and a few research points that are connected to activity investigation in still pictures have been talked about in our paper. In future we are planning to add more features in this project that would make this more usable and would revolutionize human activity monitoring system.

- In future we will implement more data set about patients' current condition which will detect patient's injury, tension etc.
- Prediction can be applied in this project to make it more usable as it can be used for security purposes.

Though it is hard working to detect action from still image and video file, we have obtained a satisfying result from our model.

6.3 Recommendation

It will be smarter to development the measure of preparing information. The dataset needs information from more various points and different wellspring of lights. Attempt other diverse profound learning calculations or deep learning algorithm to look at which is better for "Human Act". This model can be additionally utilized for different activity discovery issues like home, industry activity location issue, any sensitive area. A huge dataset request to work to perform different activity recognition task.

6.4 Implication for further research

Every single framework has been shaped with future progression openings point. In future, this framework will be quicker and more productive. Diminishing handling time is one of the significant issues.

We will redesign this for better execution from now. We need to proceed the research in this field and work with Human Activity Prediction for next step. We will attempt to recognize the more convoluted activities. We will work with recordings. We will attempt to get more exactness by applying different methods.

We will make a join programming by which we can make a report of social orders exercises. The modified upsetting structure will be made for unfortunate exercises.

REFERENCES

[1] Bernard Marr Contributor Enterprise Tech How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read May 21, 2018,12:42am EDT url:https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#75b5af2f60ba

[2] Available at: <https://www.quora.com/What-is-the-total-size-storage-capacity-of-YouTube-and-at-what-rate-is-it-increasing-How-is-Google-keeping-up-with-the-increasing-demands-of-Youtube%E2%80%99s-capacity-given-that-thousands-of-videos-are-uploaded-every-day#:~:text=is%20revealed%20yet.-,But%20estimate%20size%20of%20youtube%20may%20be%20in%20some%20Exabyte,what%20size%20it%20is%20increasing.> [Accessed 6 December 2020]

[3] Mike Jude Research Director Worldwide Video Surveillance Camera Forecast, 2020–2025 #May 2020 - Market Forecast url:https://www.idc.com/getdoc.jsp?containerId=US46230720

[4] Life In Focus. 2020. *How Many Photos Will Be Taken In 2020? - Life In Focus*. [online] Available at: <https://focus.mylio.com/tech-today/how-many-photos-will-be-taken-in-2020> [Accessed 6 December 2020].

[5] Redmon, J., 2020. *YOLO: Real-Time Object Detection*. [online] Pjreddie.com. Available at: <https://pjreddie.com/darknet/yolo/> [Accessed 6 December 2020].

[6] SAGE Journals. 2020. *A Review On Applications Of Activity Recognition Systems With Regard To Performance And Evaluation - Suneth Ranasinghe, Fadi Al Machot, Heinrich C Mayr, 2016*. [online] Available at: <https://journals.sagepub.com/doi/full/10.1177/1550147716665520> [Accessed 6 December 2020].

[7] Brownlee, J., 2020. *A Gentle Introduction To Object Recognition With Deep Learning*. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/object-recognition-with-deep-learning/> [Accessed 6 December 2020].

[8]2020. [online] Available at: <https://www.researchgate.net/publication/224190337_Supporting_patient_monitoring_using_activity_recognition_with_a_smartphone> [Accessed 6 December 2020].

[9]Shi Q, Cheng L, Wang L, Smola A. Human action segmentation and recognition using discriminative semi-markov models. IJCV. 2011;93:22–32.

[10]Feichtenhofer C, Pinz A, Wildes RP. Spatiotemporal multiplier networks for video action recognition. In: IEEE conference on computer vision and pattern recognition (CVPR); 2017. p. 7445–54

[11] Krizhevsky A, Sutskever I, Hinton G. E. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012. pp. 1097-1105

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.

[13] Pantic M, Pentland A, Nijholt A, Huang T. Human computing and machine understanding of human behavior: a survey. Artificial intelligence for human computing. Berlin: Springer; 2007. p. 47–71.

[14] Kidd C, Orr R, Abowd G, Atkeson C, Essa I, MacIntyre B, Mynatt E, Starner T, Newstetter W. The aware home: a living laboratory for ubiquitous computing research. Cooperative buildings: integrating information, organizations, and architecture. Berlin: Springer; 1999. p. 191–8.

[15] Synced. 2020. *YOLO Is Back! Version 4 Boasts Improved Speed And Accuracy*. [online] Available at: <https://syncedreview.com/2020/04/27/yolo-is-back-version-4-boasts-improved-speed-and-accuracy/#:~:text=In%20experiments%2C%20YOLOv4%20obtained%20an,as%20EfficientDet%20with%20comparable%20performance.> [Accessed 6 December 2020].

[16] Robertas Damaševičius, Mindaugas Vasiljevas, Justas Šalkevičius, Marcin Woźniak, "Human Activity Recognition in AAL Environments Using Random Projections", Computational and Mathematical Methods in Medicine, vol. 2016, Article ID 4073584, 17 pages, 2016.

[17] 2. N. A. Capela, E. D. Lemaire, and N. Baddour, “Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients,” PLoS ONE, vol. 10, no. 4, Article ID e0124414, 2015.

[18] Vrigkas M, Nikou C and Kakadiaris IA (2015) A Review of Human Activity Recognition Methods. Front. Robot. AI 2:28. doi: 10.3389/front.2015.00028

[19] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., et al. (2011). “Real-time human pose recognition in parts from single depth images,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Colorado Springs, CO), 1297–1304.

[20] Wu, Q., Wang, Z., Deng, F., Chi, Z., and Feng, D. D. (2013). Realistic human action recognition with multimodal feature selection and fusion. IEEE Trans. Syst. Man Cybern. Syst. 43, 875–885. doi:10.1109/TSMCA.2012.2226575.

[21] Radu V, Lane N. D, Bhattacharya S, Mascolo C, Marina M. K, Kawsar F. Towards multimodal deep learning for activity recognition on mobile devices. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct. ACM; 2016, September. pp. 185-188.

[22] Moya Rueda F, Grzeszick R, Fink G, Feldhorst S, ten Hompel M. Convolutional neural networks for human activity recognition using body-worn sensors. In: Informatics, Vol. 5, No. 2. Multidisciplinary Digital Publishing Institute. 2018. p. 26.

[23] Zeng M, Nguyen L. T, Yu B, Mengshoel O. J, Zhu J, Wu P, Zhang J. Convolutional neural networks for human activity recognition using mobile sensors. In: 6th International conference on mobile computing, applications and services. IEEE; 2014, November. pp. 197-205.

[24] Raptis M, Sigal L. Poselet key-framing: a model for human activity recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013. pp. 2650-2657.

[25] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 7263-7271.

[26] YOLO_based_Human_Action_Recognition_and_Localizati.pdf

[27] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance. Vis Comput. 2013;29(10):983–1009.

[28] Kastrinaki V, Zervakis M, Kalaitzakis K. A survey of video processing techniques for trafc applications. Image Vis Comput. 2003;21(4):359–81.

[29] 10. Bayat A, Pomplun M, Tran DA. A study on human activity recognition using accelerometer data from smartphones. Procedia Comput Sci. 2014;34:450–7.

[30] Roboflow Blog. 2020. *Breaking Down Yolov4*. [online] Available at: <https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/> [Accessed 6 December 2020].

[31] Medium. 2020. *Introduction To Yolov4: Research Review*. [online] Available at: <https://heartbeat.fritz.ai/introduction-to-yolov4-research-review-5b6b4bd5f255> [Accessed 6 December 2020].

[32] Medium. 2020. *YOLO — You Only Look Once, Real Time Object Detection Explained*. [online] Available at:

<https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006> [Accessed 6 December 2020].

[33] International Conference on Robotics and Smart Manufacturing (RoSMa2018) YOLO based Human Action Recognition and Localization

[34] Flower End-to-End Detection Based on YOLOv4 Using a Mobile Device

[35] Redmon, J. and Farhadi, A., 2020. *Yolov3: An Incremental Improvement*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1804.02767> [Accessed 6 December 2020].

APPENDIX

Appendix A: research reflection

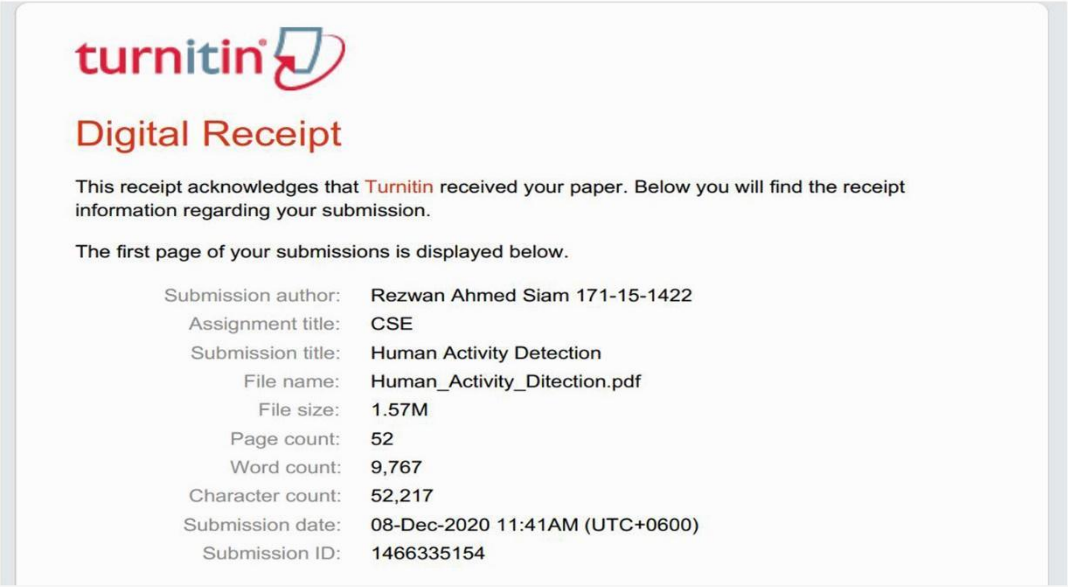
In this chapter, we will introduce you to the research reflection. To work on this research titled "Human Activity Detection using YOLOv4" we had to work on a group which was challenging and pleasant. While collecting data, we meet a lot of new peoples. Due to covid-19, we had a hard time collecting all those data. But in the end, we managed to pull this through.

We gather a lot of experience in the management part. We work so herd to maintain our time and workflow. We had to go to difference places including hospital for complex images or data. There we spent our time with new people, and took their help. Due to covid-19 maintaining the current workflow was a hard task. Hope, this research paper where we put all our effort would help all who need the best out of it.

Appendix B: related issue

While working on our project, we faced a lot of complexity. We learn python programming along with machine learning, deep learning and neural networking. We had to learn every aspect of the new YOLOv4 as this is a brand-new platform.

We had to label the image carefully for the best accuracy which was a very tiring task. We had to label about 4,674 amounts of data. Wrong labeling wound have created distortion in our project. After that, due to our system limitation we had to train all those data in Google Colab which takes a lot of time. The model also takes a lot of time due to our system limitation. But by following our recommended system, you can run this model in your own system without any problem. After all this, we had done it with lots of new experience. We ran our model in our system successfully. Hope, our research paper will come in handy for an enormous amount of people.



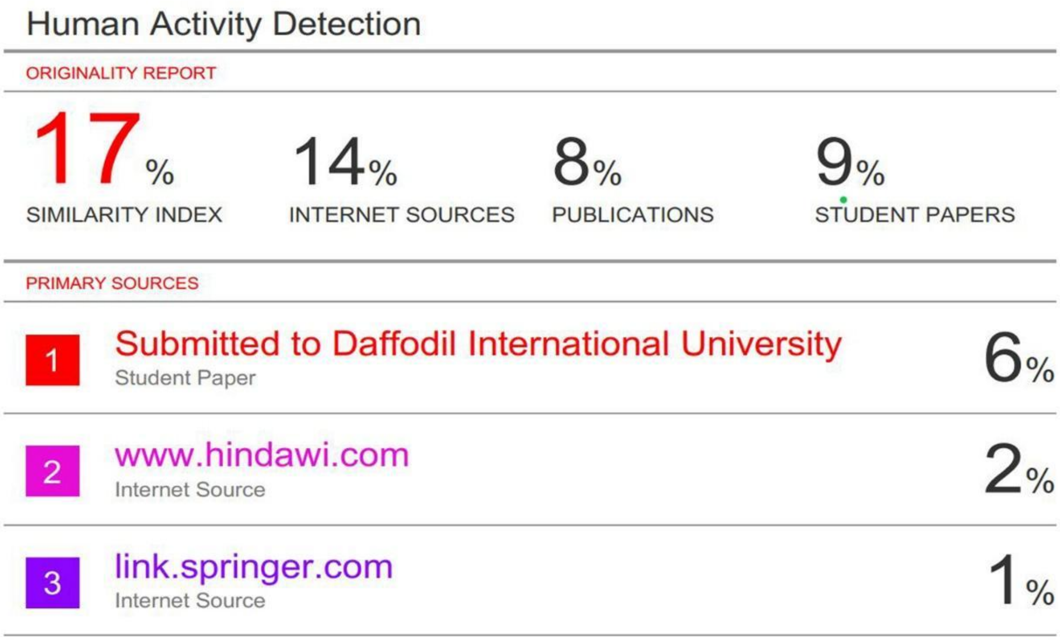
turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Rezwan Ahmed Siam 171-15-1422
Assignment title:	CSE
Submission title:	Human Activity Detection
File name:	Human_Activity_Dition.pdf
File size:	1.57M
Page count:	52
Word count:	9,767
Character count:	52,217
Submission date:	08-Dec-2020 11:41AM (UTC+0600)
Submission ID:	1466335154



Human Activity Detection

ORIGINALITY REPORT

17%	14%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	www.hindawi.com Internet Source	2%
3	link.springer.com Internet Source	1%