Blind Assistance System To Recognize Dynamic Object and Command Detection Using Image Processing Approach

Thesis submitted to the Department of Computer Science and Engineering in the partial fulfillment of the requirement for the Degree of M. Sc. In CSE (Evening) program.

Submitted by
Eshita Biswas
Student ID: 201-25-871

Supervised by
Dr. Md. Fokhray Hossain
Associate Dean & Professor
Department of Computer Science and Engineering
Daffodil International University

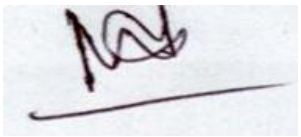
Daffodil International University (DIU)
102/1,Sukrabad Mirpur Rd, Dhaka, Bangladesh
3rd june, 2021

APPROVAL

This Project/internship titled **"Blind Assistance System To Recognize Dynamic Object and Command Detection Using Image Processing Approach**" ,submitted by **Eshita Biswas**, ID No: **201-25-871** 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 **03-June-2021.**

BOARD OF EXAMINERS

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



Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

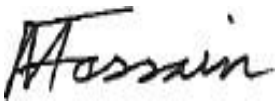
Dr. SheakRashedHaiderNoori
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

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

DECLARATION

I declare that, this thesis report is prepared by myself by own, **Eshita Biswas, ID: 201-25-871** to the department of Computer Science and Engineering, Daffodil International University, under the supervision of **Professor Dr. Md. Fokhray Hossain,** Assistant Dean & Professor, Department of CSE, Daffodil International University.
I furthermore declare that neither this thesis report nor any part of this thesis report has been submitted to any other university for the purpose of receiving a degree or diploma.

Supervised by



Professor Dr. Md. Fokhray Hossain
Assistant Dean & Professor
Department of CSE
Daffodil International University(DIU)

Submitted by:

Eshita Biswas
ID NO: 201-25-871
M. Sc. In CSE
Department of CSE
Daffodil International University(DIU)

Dedicated to my beloved parents..

Abstract

Human senses are the only medium to communicate with environment. Our brain makes a combination of different senses and convert it into a meaningful mixture. I, humans have five different senses as in two eyes to see, one nose to smell, two ears to hear, one tongue to taste and a skin to feel touch. All of these senses are connected deeply with each other. One sense always has the backup of another one. There are thousands of people in the whole world who are blind. They face many kinds of hindrances in their everyday life. They suffer badly in travelling here and there like shopping, work, educational institutions and so on due to their problem in sight. Hence, many papers have been published on making blind or visually impaired people's navigation easier. Different papers proposed different methods and different ways to navigate their way. My paper aims to represent a proposed system that helps a visually impaired, which detects various objects and finds right path for them to reach to their destination. My projected model is mainly based on a well-known object detect and identifier algorithm that needs to be trained by a dataset. Step by step, simulations using dynamic object detection algorithm results in a better accuracy to detect those objects in my proposed model. Hereby, my paper presents the idea of an object detection system based on object extractions; networks for segmentations match with the recognized dataset and locate those objects from the live video-image extractions, which gives an audio output.

Keywords: blind people; navigation system; objects detection; visually impaired; wearable device; path detection.

Acknowledgement

My heartiest thank goes to my thesis supervisor **Dr. Md. Fokhray Hossain** for his cordial support, valuable suggestions and guidance. He gave me all kinds of support through out my thesis period. He provided me all kind resources i needed for my thesis. I got all kind of inspiration and courage to do thesis from him. I extend my sincere thanks to my respected head of the depeartment **Professor Dr. Touhid Bhuiyan**, for allowing us to use the facilities available. I would like to thanks the other faculty members also , at this occasion. Last but not the least, I would like to thank our friends and family for the support and encouragement they have given me during the course of my work.

Finally, but most importantly, I want to convey my thanks to my parents for their love, support, compassion, and inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES:

LIST OF ACRONYMS :

RGB	Red, Green, Blue
BGR	Blue, Green, Red
YOLO	You Only Look Once
GPS	Global Positioning System
COCO	Common Object in Context
RCNN	Region-Convolutional Neural Network
GPU	Graphics Processing Unit
CUDA	Compute Unified Device Architecture
IDE	Integrated Drive Electronics

CHAPTER 1

INTRODUCTION:

1.1 Background of the research: The major sensory organ of a person is their eyes. One glimpse around us is enough to make us realize how visual is most important in our environment. Visually impaired and blind people come from all kinds of backgrounds. Many are elderly, some are young. They may be sportsmen and women, gardeners, farmers, chess players, teachers, typists, musicians, lawyers, housewives, computer programmers, physiotherapists, social workers, telephonist. Such people have many abilities and can achieve many things despite visual impairment or blindness, but there are times when they will appreciate and welcome practical assistance ."Eyes show the strength of one's soul." – is one of the expressions that i use to describe the deep connection with the mind and eyes. Unfortunately, there are millions of people in the whole world have problems with their eyesight. Some may have been blind since birth, and some other have lost their sight because of any accident or illness. Experts have found that about 750,000 people in Bangladesh are suffering from blindness while approximately 253 million people live with moderate to severe vision impairment and 36 million people are blind in a discussion. [1] At the stage of having no sight, eyes opt to other sense that will take over the responsibilities of eyes in every sphere. Blind people to cooperate with other senses like- they can use their hearing sense or smelling sense with proper training. However, only few of the sufferers take the help of existing technological aids to improve their situation. Hence the need for supportive devices for navigation are increasing day by day whereas some of the supportive tools are not available or do not have

easy access for all. Till now, the simplest and the most affordable tools for blind or visually impaired are trained dogs and the traditional white cane for their path detection [2]. However, these tools are very popular, but they cannot provide all information they need for safe movement in roads. According to WHO approximately 80% of vision impairment or blindness globally can be cured by Governments aids [3]. As there are different types of devices to prevent and treatment for blind or impaired people. One of the common device of assistive technology is Electronic Travels Aids (ETAs) which collect and transfer information about the surroundings to the user though sensor cameras, sonar scanners or laser scanners. This technology follows the rules of National Research Council [4-6]. Another technology is Electronic Orientation Aids (EOAs) that provide instruction to the user in unknown place [7, 8]. Position Locator Devices (PLD) uses GPS technology, which shows user's precise position. Wahab et al. researched the Smart Cane product development for object detection and developed accurate navigation instructions.

1.2 Aims and Objectives:

After observing all of these papers and so on, I have decided to propose a model for blind that will help them to choose right path to home, work or shop. It is not only handy but also it has easy access for each sphere of people. In my project, it presents a dynamic object and command detection system based on image processing by training faster RCNN that matches the needs of the virtual impaired or blind people in their everyday life. Here I use cameras in three angles- front, left and right to detect the obstacles or objects in their walkway or sideways. As visually impaired are more dependable on their hearing sense, my model will transfer an audio output to the user. That voice will guide him or her to their destination safely. My aim is to spread the thought that blind people are also capable of doing everything at ease.

1.3 Problem Statement: I think the number one problem people who are blind, face low expectations. Others the biggest problem for like person, especially the one with the complete loss of vision is to navigate around the places. Arduous to travelling here and there like market, work,educational, institutions. Having little to no opportunity to support oneself, blind or low vision individuals are incapacitated from their independence. Navigating new environments is extremly difficult.

1.4 Research methodology: For dynamic object detection model is to detect all the objects was trained using faster RCNN algorithm which was implemented using python GluonCV library. At firstly process, it takes video as an input data, systematically process frames from the video input, calculate the approximate distance of the object from the blind people. Then it generates audio output according to the measured value. Here used 3 camera lens that covers fronts, right and left area.

1.5 Proposed System : The proposed system is to detect all the objects was trained using faster RCNN algorithm which was implemented using python GluonCV library. The model needs a process that takes video as an input data, systematically process frames from the video input, calculate the approximate distance of the object from the blind people. Then it generates audio output according to the measured value. Figure 1.1 & 1.2 provides a block diagram of the proposed system. Figure 1.1, include memory card where video image are stored, others ON/OFF button use for start or finish .

Fig 1.1: Diagram of the proposed system.

Fig 1.2 : Diagram of the proposed system.

In figure 1.2, I use three camera lens in my proposed system like front, right and left. So, it detects cover 3 sides. At the same time this 3 lens detects video image and stored in memory.

1.6 Conclusion: In this paper, a solution is proposed to help visually impaired person to travel safely. With the help of blind assist system, people can advance travel speed, diminish trivial collision, do not lose their way, and increase safety a compare to unaided equipment's. The lens for blind system developed through this research aids visually impaired peoples navigate smoother, both indoor and outdoor.

CHAPTER 2

RECOMMENDED ANALYSIS:

2.1 Introduction: Development in technology leads to the advancement in science and helps to revolution in several fields. For making a system, require different types of material and different types of product. For making a useable and reliable system, research the requirement, which makes the system more faxable for the system as well as for the user who are more likely use the system. For that, requirement analysis plays an important part of enhance the system and more useable for the user.

2.2 Exiting Work: Past or present, there have been different systems and devices related to the navigation assistance for visually impaired or totally blind people. Each of them has both advantages and drawbacks. Hence, in a paper, they proposed a system that is a sensor module to cover the head area for blind people. This object detector gives two different type output like buzzer mode and vibration mode and modes are controlled by the user. This device is using proximity IR sensor for detection which is placed on a stepper motor [12]. However, this system requires android mobile assistance. Another paper proposed an intelligent walking stick for the blind people. This system is using PCB unit which consists of microcontroller, Bluetooth HC05, MAX232, ADC 0808 and IR sensors [13]. Moreover, they are using RFID sensor and create an android application which gives the audio output. Whereas, this system cannot cover near to head area. There are other researches on visually impaired people that the walking stick system consists of a GPS that allows the VIP to know about the outdoor area. This system send massage to authorized people when emergency occurs by GSM. Though this system is not user friendly and only cover limited range of outdoor area [14]. Additionally, there is a paper in which they proposed a model where they use Gabor filtering, capture 2D image & convert them into RGB images into Gray scale to detect the staircase and escalator. They also use Microsoft Kinect and faster techniques to assist the blind people. Their system also works in

different environment such as foggy, blur, low light, and sunlight with different angle of view point. They only help impaired people to go through the staircase but it cannot deal with the dynamic objects [15]. In another paper, they present an assisting device which can support visually impaired to move safely and easily inside home or any complex corridors at home. They use AR, depth sensor, vision enhancement algorithms to prepare the smart guiding glass. It can only help blind people to move inside but cannot help them to move outside home like busy roads or other places because it cannot deal with the dynamic objects [16]. Furthermore, a project that uses YOLO model to get accurate directional knowledge and the relative distance. They use Microsoft Kinect to estimate the depth and plug-in for Unity-based game program to generate 3D audio output. Their project can detect real time objects and locate a blind person by giving 3D audio messages. However, it is not applicable for outside like roads because it only works with the static objects not the dynamic ones [17]. These different works and researches encouraged me a lot in my work. It also helps me with more and more informative ideas with my model.

2.3 Conclusion: A navigation system's purpose is to provide users with required andr helpful data to get to a destination point, monitoring their position in previous modeled maps. As i will see, researchers working in this field have yet to find effective, efficient, safe, and cost-effective technical solutions for both the outdoor and indoor guidance needs of blind and visually impaired people.

CHAPTER 3

METHODOLOGY:

3.1 Introduction: This chapter describes the methodology of the model for dynamic object and command detection for assisting blind people. It includes the explanation of dataset to be used in implementing and testing the model, as well as the description of how it detects any dynamic object and gives alarm. It also describes each part of the proposed method including faster RCNN algorithm which was implemented using python GluonCV library. There is different trained model based dataset named COCO, which is also used in my proposed model.

3.2 Model Design: The objective of the proposed dynamic object detection model is to detect all the objects was trained using faster RCNN algorithm which was implemented using python GluonCV library. The model needs a process that takes video as an input data, systematically process frames from the video input, calculate the approximate distance of the object from the blind people. Then it generates audio output according to the measured value. Figure 3.1 provides a block diagram of the model design.

Figure 3.1: Block diagram of the proposed model

Again, Figure 3.2 shows the workflow of the whole proposed model. The model works step by step according to the workflow.

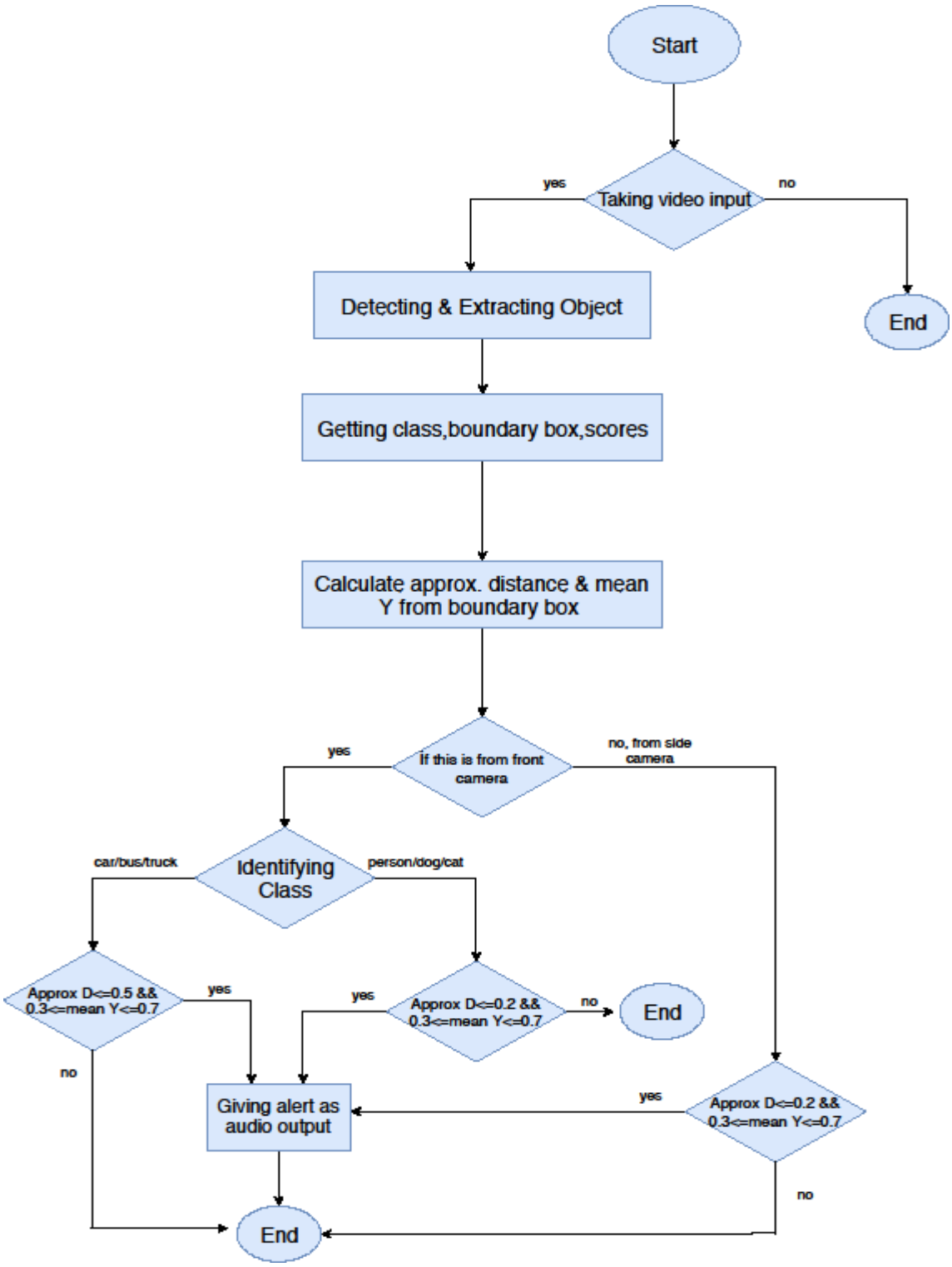


Figure 3.2: The flow chart of the proposed model

3.3 Description of the Flowchart: The proposed model has been followed some steps so that i can successfully detect my required object and give alert at the accurate time. This model at first takes a video as an input and starts working. When the video input is successfully taken, then by using the faster RCNN algorithm the objects are extracted from individual video frames. After that, i get 3 outputs in return such as class, bounding box and scores.

Here class defines what type of object it can detect. The object can be a car or truck or bus and it can also be a person or dog or cat. From the bounding box i can get Xmin, Xmax, Ymin, Ymax by which i can calculate approx. distance (D) and mean Y.

Scores defines the percentage of an object which is actually looks like that the object from the trained dataset. So that i can confirm about the actual object from the video frame by looking at the percentage.

Next i consider about the camera position, if the object is coming from the front camera and if is either a car or a bus or a truck then i measure the approx. D and mean Y. The model gives alert according to the compared calculation. If approx. D <= 0.5 and 0.3<= mean Y <= 0.7 then the model gives alert. On the other hand, if the detected object is a person or a dog or a cat and approx. D <= 0.2, 0.3 <= mean Y <= 0.7, then the system gives alert to the blind person.

However, if the object is detected from the side camera in that case approx. D <= 0.2 and 0.3 <= mean Y <= 0.7, then the alert is given to the blind person. By following the alerts, a visually impaired person can move easily outside home or crossing the road by himself.

3.4 Faster R-CNN Network: Region proposal network (RPN) for generating region proposals and a network using these proposals to detect objects are the two networks of Faster R-CNN. The time cost of generating region proposals is much smaller in RPN, when RPN shares the most computation with the object detection network.

An intuitive speedup solution is to integrate the region proposal algorithm into the CNN model. Faster R-CNN is constructing exactly a single, unified model composed of RPN (region proposal network) and fast R-CNN with shared convolutional feature layers.

This is a pre- train CNN network on image classification tasks. So i have to train my model with COCO dataset. Region proposal network is initialized by the pre-train image classifier. It gives two types of samples positive samples have intersection over union is greater 0.7 on the other hand negative samples have less than 0.3.

In MXnet faster RCNN returns raw outputs of class id, bounding box, scores. These three are mostly needed for my model. Calculation is depending on the bounding box. I can calculate approx. distance and mean Y from the bounding box. I get the approx. value of Xmax, Xmin, Ymax, Ymin from bounding box. A lot of intermediate values are generated but i use the approximate values for calculation. By using faster RCNN algorithm which it is implemented using python GluonCV library my proposed model is trained for detecting objects.

3.5 COCO Dataset: When it comes to object detection dataset, i selected COCO dataset. COCO stands for common object in context [4]. It is a large library which has loads of labeled object. The final version COCO have approximately 164k images in which 118k are trained. It presents 172 classes where 80 thing classes, 91 stuff classes and 1 classs is 'unlebeled'. This dataset has other highlights like- heavy pixel level annotations, complex context among stuffs and things. Those annotations of COCO are mostly useful for semantic segmentation or object detection. Here is the display of level hierarchy of COCO stuffing that includes all classes:

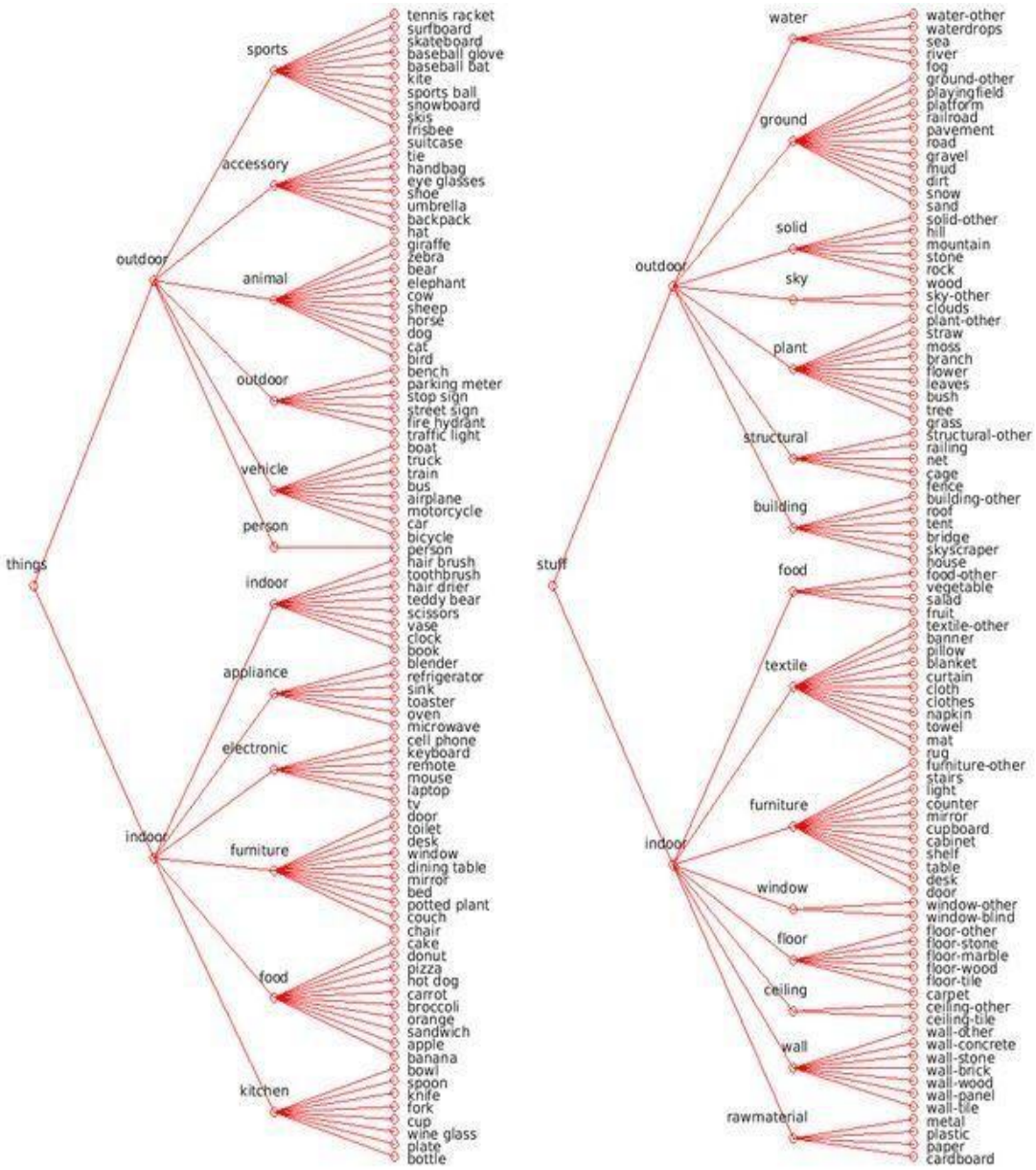


Figure3.3: Level hierarchy of COCO dataset

There's a connection between COCO dataset and GLUON to prepare the dataset. To train data, its necessary to pass through dataset transformation and then load it with GLUON.

Figure3.4: COCO dataset

3.6 GluonCV: GluonCV is a Deep Learning Toolkit for Computer Vision and natural language processing which provides state-of-the-art pre-trained models and modular APIs with flexible building blocks. Modular APIs are used for enabling efficient customization where user can train, design or interface by using efficient components across different models. This library is based on Apache MXNet. Since Python is an agnostic platform, GluonCV is implemented in Python and available for systems running Linux, OS, and Windows. The advantages of using GluonCV is that have 100 contributors worldwide on GitHub. I am using GluonCV in my project because it can work without GPU in some cases. Moreover, it can be simulated by different programming languages so that developers can use according to their preferences and it is also a lighted library.

3.7 Conclusion: By RCNN algorithm the objects was trained. It is constructing exactly a single unified model composed of RPN(region proposal network) and faster R-CNN with shared convolutional feature layers. Here, I get 3 outputs (fig:3.2) in return such as class, bounding box and scores. COCO are mostly useful for semantic segmentation or object detection. In methodology I was using Gluon CV in my work because it can work without GPU in some cases.

CHAPTER 4

IMPLEMENTATION:

4.1 Introduction: This chapter describes the implementation of the proposed system of dynamic object detection for assisting blind people. The system was implemented in 3 steps. Firstly, a model was trained for detecting objects. Then this model was used for object classification and localization. Finally, this system gave alert in given conditions. The model for detecting objects were trained using faster RCNN algorithm and it was implemented using python GluonCV library [1]. GluonCV is a deep learning toolkit which gives the opportunity to implement deep learning algorithms easily. It is integrated to Apache MXnet which is an open source framework [2]. MXnet is an effective tool for training and deploying neural networks and it supports almost all popular deep learning algorithms. My model was trained by the train_faster_RCNN.py file provided by gluonCV documentation [3]. This chapter also provides the results of the implemented system. Anaconda Spyder is used to run the whole process.

4.2 Implemented system files: The proposed system for dynamic object detection for assisting blind people consists of three files.

Table 1: Description of system files used in the model:

File Name	Description
mscoco.py	This script was provided in the official GluonCV website to download cocodataset and unzip them.

train_faster_RCNN.py	The file was provided in the official GluonCV website to train a model using faster RCNN algorithon.
Detection.py	This file contains the code for detecting objects and giving alerts in given conditions.
fRCNN_0000.params	This file contains the trained weights after training process.

The reason for choosing GluonCV library because it is easier to apply than other libraries like tensorflow . It is an open source toolkit. Dataset preprocessing and indexing is less time consuming here and it contains a consistent interface for retrain models.

4.3 Used Libraries:

Apache MXNet: The detailed form of MXNet is mixing and maximing network. It is an opensource deep learning framework. It supports almost all popular operating systems. MXNet framework can be used in C++, Python, R, Julia, Perl etc. Therefore, there is no need of changing platforms. Moreover, MXNet models are portable and they require less amount of memory.

GLuonCV: GluonCV is a deep learning toolkit. The main reason behind creating this toolkit was replicating experimental results from published papers as it is the most difficult part. GluonCV is a powerful tool specially for debugging and reproducing experiment results. It supports multiple algorithms and it can be applied on almost all operating system. In my system, i have used GluonCV api, for train my model and get weights.

OpenCV: OpenCV is a python library for real-time computer vision. In a word, it is used for image processing. It is popular for its simplicity and code readability. OpenCV supports a

multitude of algorithms and it can do high speed CUDA operation. OpenCV-python is basically the python api for python which combined OpenCV C++ api and python language together. In my system, i have used OpenCV for taking video inputs and showing outputs.

Numpy: Numpy stands for numerical python. It is a popular library of python which is used almost all deep learning projects. It supports large and multidimensional arrays. By using Numpy, it is easy to manipulate data and store them as matrix.

Matplotlib: Matplotlib is a python library which is used for plotting figures and charts. It is basically a 2D plotting visualization library. In my project, i have used Matplotlib to visualize data for multiple times.

FFpyplayer: FFpyplayer is a library for playing and writing media files. In my project, i have used it for giving output as audio form.

4.4 Dataset Preparing:

The model for detecting objects was trained using COCO dataset. Its large dataset helps me to detect as much object as possible. GluonCV provides a script to download this large dataset with annotation. Once downloaded, the files were unzipped. There were three separate folders - train2017, val2017 and annotations. The first two folders contain images for training and validation. The last folder contains json formatted files of annotation. But all of the files of annotation folder were not used. Only instances_train2017.json and instances_val2017.json were used.

After that, gluoncv.data.COCODetection method was used for loading images and labels. No transformation was done in this step. Here is the result of loading data and creating index. This screenshot was captured from Ipython console of Spyder IDE.

```
loading annotations into memory...
Done (t=20.55s)
creating index...
index created!
loading annotations into memory...
Done (t=1.00s)
creating index...
index created!
Num of training images: 117266
Num of validation images: 4952

In [2]:
```

Figure 4.1: A sample of indexed training and validation data

Here after removing unlabeled images the number of total training image is 117266 and the number of validation image is 4952. After indexing, it was checked if the indexing was created correctly. Bounded boxes and class ids were found from train label and it was found that indexing was successful. The 85th image from the train2017 folder was chosen for the test which contained a bus with a person inside it. The result of the test is given below. This figure shows the properly labeled bounded box.

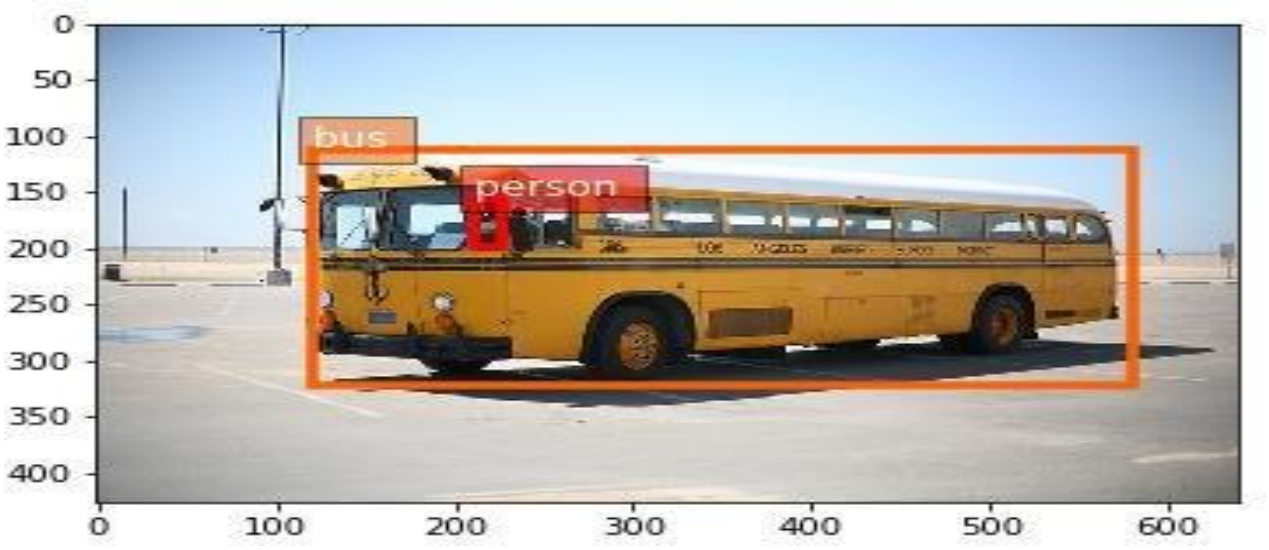


Figure 4.2: A sample of validation of indexing

4.5 Faster RCNN Algorithm Implementation: Faster RCNN algorithm was implemented in train_faster_RCNN.py file. This file was provided by GluonCV official website [3]. This file takes some arguments while running for example pochs, seeds, dataset etc. to train this model, 18 epochs were done and batch-size was 20. By declaring pre trained base false, this model was prevented from using pre trained weights and the weights were updated during the training process. In this step raw images were turned into tensors. A tensor is a mathematical entity which describes a linear mapping from one set of objects to another. GluonCV is integrated to MXnet which Follows BCHW(Batch x Channel x Height x Width) format of tensor. Faster RCNN is a powerful algorithm which can deal with raw images with various aspect ratios.

gluoncv.data.batchify.Append() is used for this process. Here images are not stacked or padded and unchanged with their own shape or aspects. For iterating through the whole dataset for multiple times during the training, Dataloader() method was used to load data in small batches.

In MXnet faster RCNN returns class id, scores and boxes. Score means how much perfect is the predicted object. Bounded boxes are co-ordinates of detected object in the frame. During the training process, faster RCNN behaves differently than other algorithms. A huge amount of intermediate values is generated. During training the model, rpn_score and rpn_box was generated which are the raw outputs from RPN's convolutional layer.

To train the model for object detection, four type of losses were involved throughout the training. This loss functions are builtin mxnet gluoncv classes when were built with different equations. RPN class loss was calculated by sigmoid binary cross entropy loss function,RPN box losses was calculated by huberloss function . RCNN class loss was calculated by softmax cross entropy loss and RCNN box loss was calculated by huberloss. To speed up the training process, MXnet let the CPU to pre calculate RPN training targets. RCNN training targets were created from the intermediate outputs with stored target generator.

After calculating the loss function and training targets, the training loop was iterated according to the batch size and epochs. When the iterations were completed, the trained weights were saved as fRCNN_0000.params.

4.6 Detection of Objects: The detection part and relative distance measurement part were done in detection.py file. In this part, fRCNN_0000.params file was loaded in the network. For capturing video and getting the frames, opencv library was used. Here some frame preprocessing was done. For testing purpose, i run the detection process using webcam.

Figure 4.3: Taking input using webcam.

4.7 Hardware Setup: Three webcam were used in the whole process. These cameras were attached to a spectacle. One camera was set at the frontal part of the spectacles for getting the front view. Other two were used for getting the view of left and right side. By using these spectacles with three cameras, now it is possible to cross the road safely during a traffic signal.

4.8 Detected Outputs: OpenCV captures frames in BGR color format but for detecting objects RGB format is needed. Therefore, each frame was converted into RGB format. Frames which are shorter than 512px, they are resized into 512px and max size was set to 700px as the longer side of the image remains smaller than 700px. After that, the frame was fed into the network which returns three values – class id, scores and bounding boxes. cv_plot_bbox()method from gluoncv was used to plot the bounding boxs,scores and classes on the output frame and cv_plot_image() was used to plot the output frame with these values.

4.9 Approximate Distance Measurement and Alerting: In my proposed system, it was not possible to get the real distance of detected object. Because focal length is different for different cameras and it is not a food approach to calculate focal length manually each time while using a different camera. To solve this issue, approximate distance was used instead of real distance which worked perfectly for this system

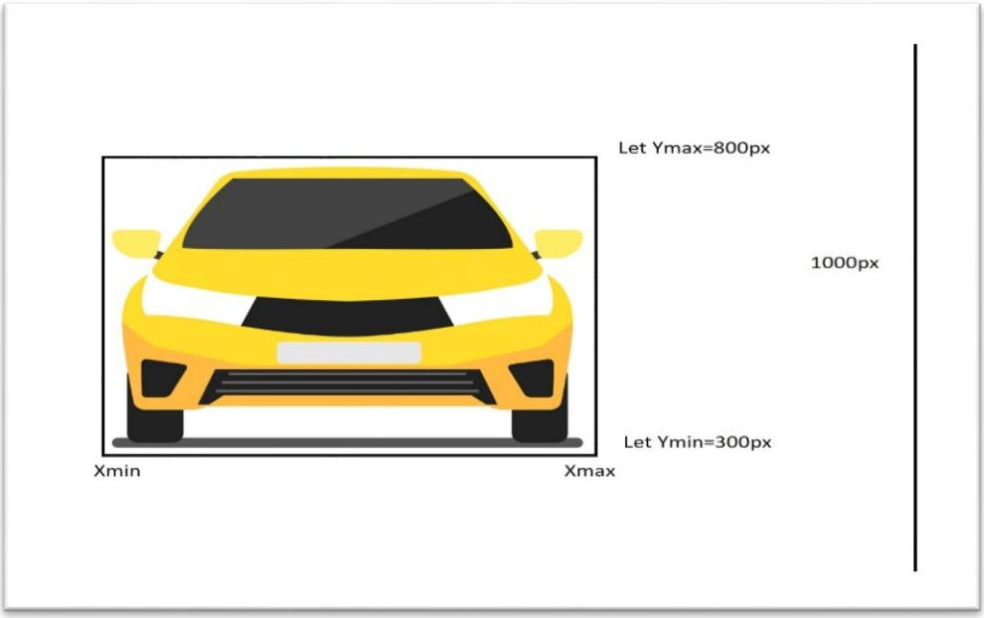


Figure 4.4: Measuring relative distance.

Figure 4.4 refers to a detected object with bounding boxes. Each bounding box contains four values(Ymin,Xmin,Ymax,Ymin) which are the co-ordinate of bounding box on the frame. Here, an equation was proposed for approximate distance –

Approximate distance = frame height – (Ymax –Ymin)

Ymean=(Ymax+Ymin)/2

For the simplicity of the measurement and less memory use, approximate distance and mean of Y were divided by frame height. Now how this process is working with approximate distance and Y mean, that can be cleared with little example.

Lets,

frame height= 1000px

Ymin= 300px; Ymax=800px

Therefore, approximate distance = (1000-(800-300)) px= 700px and MeanY = (800+300)/2= 550 px

After scaling approximate distance= (700/1000) px =0.7px and meanY= (550/1000) = 0.55px

The proposed system is giving alert for four classes (car,bus,truck,person,cat,dog) which means there are two categories here – vehicles and living instances.

If the detected object is car, bus or truck, the system will alert when approximate distance is less than 0.5 but if the detected objected is person, dog or cat the system will alert when approximate distance is less than 0.3. But there was also an issue arisen here. If some objects are close to the camera but there is not any of collision.

Figure 4.5: Detected objected with different localization in frame.

It can be observed from figure 4.5 that, there is not any possibility of collision with the left and right car but they are very close to the camera. Therefore, the system will give wrong alerts. Toovercome this issue, the system will only give alert if MeanY of detected object is in the range of 0.3 to 0.7.

If the detected object is person, cat or dog, the system will alert when approximate distance is less than 0.3 as they are living instances and can avoid collision easily. However, MeanY technique has also been followed to avoid unnecessary objects which are very close to the camera but there is not any change of collision.

Earlier it was mentioned that we are using three cameras – one for front side and the other two are for side views. For left and right camera, the alert is given when approximate distance is less than 0.2 and MeanY of detected object is in the range of 0.3 to 0.7.

Finally, the alert has been created using audio with the help of FFpyplayer library. The whole process continues while getting video frames as input. However, the perfection of the detection process depends on the resolution of the cameras to some extent.

4.10 Conclusion: The system was implanted in 3 steps. Firstly, a model was trained for detecting objects. Then this model was used for classification and localization. (Table1-) is the model of the system file. This files used in the model that is consists of three files. I choose Gluon CV library because it is easy to apply than other libraries like TensorFlow . (In fig:1) the RCNN params file was loaded in the network. For testing purpose I run the detection process using webcam. In(fig:2) refers to a detected object with bounding boxes. Another (fig:3) it mean that it can be observed there is no any possibility of collision with left and right car but they are very close to the camera. Threrefore, the system give wrong alerts.

CHAPTER 5

RESULTS ANALYSIS:

5.1 Introduction: After implementation I found a outcome from my system. Using python ffplayer library to each detected objects is given in the audio form. Comparison about train weight with pretrain weight show in this chapter. Also I had made a test for accuracy rate.

5.2 Outcome: The final output for each detected object is given in the audio form and python ffplayer library was used to play the audio. I had made a test for checking accuracy with our trained weight. Then i did the same thing with the weights of pretrained model faster_RCNN_resnet50_v1b_voc from model zoo.

Figure 5.1: Detected objects during the test

Table 2: Outcome through the test in the model

The output i found through this test is given below:

Item	Frames	Detected Objects	Wrong Prediction
Weights of trained model for the System	30	8	4
Weights of pretrained model	30	5	2

I have run this test for four times with different video frames and specified them as sessions. After that, i created a bar chart for the result. Number of frames in session01, session02, session03, session04 are respectively 30,60,15,24.

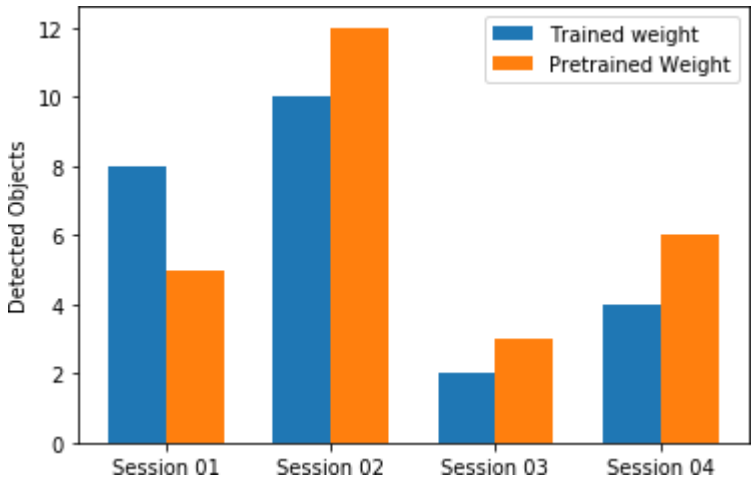


Figure 5.2: Comparison of trained weight with pertained weight for detected objects (Higher is better)

In figure 5.2, session 1 and 2 is higher than the results of session 3,4 as number of frames are higher in session 1 and 2. Pretrained weight (gluonCV model of faster RCNN) worked better in session 2,3,4 and my trained weight worked better in session 1

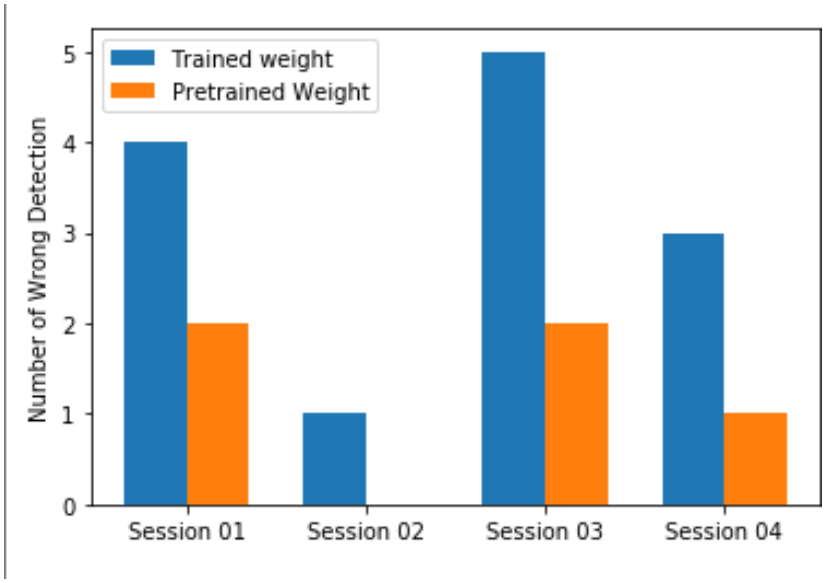


Figure 5.3: Comparison of trained weight with pertained weight for detected objects (Lower is better) (Note: no wrong detection in pretrained weight in session2)

In fig 5.3, there is no wrong detection in pretrained weight (gluonCV model of faster RCNN) in session01, therefore, there is no bar here.

5.3 Accuracy Rate: After running this test more and more, there are some better impacts in my expected output. Each time i get to make the model more accurate. At the end, my model has given 74.65% accuracy with 100 epochs, which is good at this level.

Here, i have given the model accuracy graph between train and validation sets:

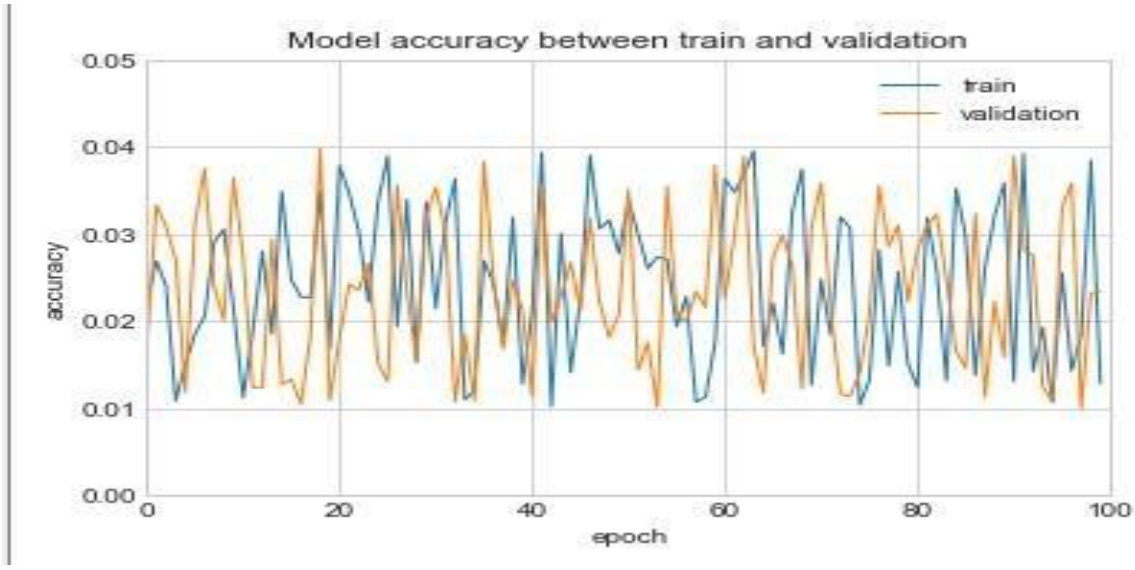


Figure 5.4: Accuracy rate of the proposed model

5.4 Conclusion: Every human being has to right live with comfort and ease. So, it is my duty to help the blind and visually impaired in their daily path detection. I can present a model that is more efficient and give less error. After result analysis, my proposed system achieved 74.65% accuracy. I mean it is good at this level.

CHAPTER 6
SWOT ANALYSIS:

6.1 Introduction: To develop a system, should focus on different things. To starting of build up a system, find out the possible ways of built the system and facing problems. By SWOT analysis, it is easy to find out strength of a system as well as weakness of system and the opportunity of system and the threat of the system. By SWOT analysis, it is more easy to find out those things and find the focusing thing of the system to build the system.

6.2 SWOT Analysis: SWOT analysis means discussing about the strength, weakness, opportunities and threats of any project, research work or any system develop. The reason for SWOT analyzing that, a system when it is built, facing problem and take caring area. It will help to find out the area which need more focused need and strength of the system. SOWT also help to work a system more organizing ways. By analyzing process regularly, find out the efficiency of system [10] [11].

6.2.1 Strength Analysis: This system is used by disabled people to help in their daily lives, make their activities easier and provide a safe mobility. Here I use camera in three angles- front, right and left to detect the obstacles or objects in their walking or sideways. So, this camera take video image at 3 different sides at a time. It is a travel friendly device based on video and audio feedback.

6.2.2 Weakness Analysis: In era of advancing technology, new and updated technology are invented for blind people. My device has low memory storage to train the system. Also it was not possible to get the real distance of detected object just only get the approximate distance.

6.2.3 Opportunity Analysis: My research is a simple, portable system that detects dynamic object and gives audio command for assisting visually impaired people. In a developing countries, here GPS may not be available everywhere due to low signal. Hence my system using 3 camera lens comes up with a more developed and user-friendly device that assist path detection for blind people.

6.2.4 Threat Analysis: In present time, new technologies are invented. For new technologies also have different problems. In my system cann't get the real distance of object. So, it is one of the main threats for blind person who are using this devices.

6.3 Conclusion: By SWOT analysis, it is easy to know about the system. Every human being has right to like with comfort and easy. So. I am trying to waken their daily path detection through my proposed model. I wish that the blind people will be helped by this system

CHAPTER 7
CONCLUSION:

7.1 Conclusion: I designed and implemented the proposed system that ensures the mobility and safe navigation for blind people. The system must require prior knowledge about the object size, shape or distance. My work results in a simple, portable system that detects dynamic object and then gives audio command for assisting visually impaired people. It is a travel friendly device based on video and audio feedback. Several experimental results in outdoor and indoor environments showed that this system is very helpful for this user group. Additionally, more and more disabled people are suffering everyday due to the lack of perfect path detector device. Especially there is not much development in this system in Bangladesh. Here GPS may not be available everywhere due to low signal. Hence my paper comes up with a more developed and user-friendly device that assist path detection for blind people. I have implemented the system with different techniques to design my predicted model so that i can present a model that is more efficient and give less error. The proposed system achieved 74.65% accuracy with 100 epochs. Every human being has the right to live with comfort and ease, so it is my duty to help the blind and visually impaired in their daily path detection. It is said that even eyes are useless when the mind is blind. Thus, i have tried to waken their mind through my proposed model with an audio output to waken their eyes to give them a better world.

7.2 Future Scope: My initial steps are hopefully promising and helpful for visually impaired people. My future improvements depend on flourishing training dataset. For now, I am notifying the blind person in what distance a car is coming. Actually, for now I am measuring the distance of a car from the person whereas; this system can also be trained into two ways. In order to do so, i will store 1000 pictures of the front side of a car will be stored in one folder and 1000 pictures of other side of a car will be stored in another folder. So that the system will detect not only the distance of a car but also it will tell in which side the car is coming or in which side of the car, the blind person is standing. The more pictures i will store to train the system the most accurately the system can give alert. Furthermore, i will use Raspberry Pi-style Jetson Nano along with the three web cameras for making the prototype of our proposed model. This type of Raspberry Pi is a powerful low-cost AI computer. This developer tool is a board tailored for running machine learning models and using them to carry out tasks such as computer vision. These are some challenges for making this proposed system more efficient in future. Due to lack of time, I am now proposing this model in short but when i will make this device more compatible for business purpose to improve the visually impaired people's life.

REFERENCES:

1. https://www.dhakatribune.com/bangladesh/2018/10/12/experts-750-000-people-suffer-from-blindness-in-bangladesh.
2. http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/8365/14366004_MSc%20in%20CSE.pdf?sequence=1&isAllowed=y
3. COCO Dataset. [Online]. http://cocodataset.org/#home
4. Blasch B.B., Wiener W.R., Welsh R.L. Foundations of Orientation and Mobility. 2nd ed. AFB Press; New York, NY, USA: 1997
5. MXNET apache. [Online]. https://mxnet.apache.org/
6. GLUONCV mxnet. [Online]. https://gluon-cv.mxnet.io/
7. Prof. Seema Udgirkar1, Shivaji Sarokar2, Sujit Gore3, Dinesh Kakuste4, Suraj Chaskar. Object Detection System for Blind People. International Journal of Innovative Research in Computerand Communication Engineering; Vol. 4, Issue 9, September2016.
8. Kher Chaitrali S., Dabhade Yogita A., Kadam Snehal K., Dhamdhere Swati D., Deshpande Aarti V. An Intelligent Walking Stick for the Blind. International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.
9. Train Faster RCNN. [Online]. https://gluon-cv.mxnet.io/build/examples_detection/train_faster_RCNN_voc.html
10. Bai, J., & Lian, S., & Liu, Z., & Wang, K., & Liu, D. (2016). Smart Guiding Glasses for Visually Impaired People in Indoor Environment [PDF file]. China, DC: Author. Retrieved from https://arxiv.org/ftp/arxiv/papers/1709/1709.09359.pdf

11. A.Sangami1, M.Kavithra2, K.Rubina3, S.Sivaprakasam. Obstacle Detection and Location Finding For Blind People. International Journal of Innovative Research in Computer and Communication Engineering; Vol. 3, Special Issue 2, March 2015.

12. Farcy R., Leroux R., Jucha A., Damaschini R., Grégoire C., Zogaghi A. Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view; Proceedings of the Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion; Los Alamitos, CA, USA. 9–11 July 2006.

13. Liu J., Liu J., Xu L., Jin W. Electronic travel aids for the blind based on sensory substitution; Proceedings of the 2010 5th International Conference on Computer Science and Education (ICCSE); Hefei, China. 24–27 August 2010.

14. Bharambe S., Thakker R., Patil H., Bhurchandi K.M. Substitute Eyes for Blind with Navigator Using Android; Proceedings of the India Educators Conference (TIIEC); Bangalore, India. 4–6 April 2013; pp. 38–43.

15. Jiang, R., & Lin, Q., & Qu, S. (2016). Let Blind People See: Real-Time Visual Recognition with Results Converted to 3D Audio [PDF file]. UK, DC: Author. Retrieved from http://cs231n.stanford.edu/reports/2016/pdfs/218_Report.pdf

16. White C.E., Bernstein D., Kornhauser A.L. Some map matching algorithms for personal navigation assistants. Trans. Res. C Emerg. Tech. 2000;8:91–108. doi: 10.1016/S0968-090X(00)00026-7.

17.Kammoun S., Marc J.-M., Oriola B., Christophe J. Toward a better guidance in wearable electronic orientation aids; Proceedings of the IFIP Conference on Human-Computer Interaction; Lisbon, Portugal. 5–9 September 2011.

18. Sánchez J., Elías M. Guidelines for designing mobility and orientation software for blind children; Proceedings of the IFIP Conference on Human-Computer Interaction; Janeiro, Brazil. 10–14 September 2007.

19. Wahab A., Helmy M., Talib A.A., Kadir H.A., Johari A., Noraziah A., Sidek R.M., Mutalib A.A. Smart Cane: Assistive Cane for Visually-impaired People. Int. J. Comput. Sci. Issues.2011;8:4.

20. Vision impairment and blindness. (n.d.). Retrieved from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

21. Hossain, E. (2010) 'Terrain detection and hardware implementation for walking support system', Masters thesis, International Islamic University, Malaysia.

22. Prasun Shrivastava, Akash Singh, Praveen Anand , V Sagar , "Medico Stick :An Ease to Blind & Deaf "IEEE Sponsored 2nd International Conference On Electronics And Communication System (ICECS 2015)".

Turnitin Originality Report

Processed on: 03-Jun-2021 12:50 +06

ID: 1599503998

Word Count: 6838

Submitted: 1

201-25-871 By Eshita Biswas

Similarity Index	Similarity by Source
10%	Internet Sources: 8% Publications: 4% Student Papers: 4%