

**INDOOR NAVIGATION FOR THE VISUALLY IMPAIRED BY READING SHOP
TRADEMARKS IN SHOPPING MALL**
BY
BRANDON LING YI YUN

A REPORT
SUBMITTED TO
Universiti Tunku Abdul Rahman
in partial fulfilment of the requirements
for the degree of
BACHELOR OF COMPUTER SCIENCE (HONOURS)
Faculty of Information and Communication Technology
(Kampar Campus)

MAY 2022

REPORT STATUS DECLARATION FORM

Title: INDOOR NAVIGATION FOR THE VISUALLY IMPAIRED BY READING
SHOP TRADEMARKS IN SHOPPING MALL

Academic Session: MAY 2022

I BRANDON LING YI YUN
(CAPITAL LETTER)

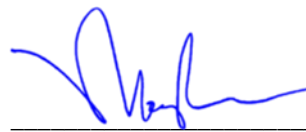
declare that I allow this Final Year Project Report to be kept in
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Bdn

(Author's signature)

Verified by,



(Supervisor's signature)

Address:

2130, Jalan Seksyen 2/4, Taman

Bandar Barat, 31900, Kampar,

Perak

Leung Kar Hang

Supervisor's name

Date: 09/09/2022

Date: 8 Sep 2022

| Universiti Tunku Abdul Rahman | | | |
|--|-------------------|-------------------------------------|-------------------------|
| Form Title : Sample of Submission Sheet for FYP/Dissertation/Thesis | | | |
| Form Number: FM-IAD-004 | Rev No.: 0 | Effective Date: 21 JUNE 2011 | Page No.: 1 of 1 |

FACULTY OF INFORMATION TECHNOLOGY AND COMMUNICATION

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 9/9/2022

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that **Brandon Ling Yi Yun** (ID No: **19ACB06748**) has completed this final year project/ dissertation/ thesis* entitled “Indoor Navigation for the Visually Impaired to Read Shop Trademarks” under the supervision of **Leung Kar Hang** from the Department of Computer Science, Faculty/Institute* of Information Technology and Communication , and _____ (Co-Supervisor)* from the Department of _____, Faculty/Institute* of _____.

I understand that University will upload softcopy of my final year project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,



(Brandon Ling Yi Yun)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled “**Indoor Navigation for the Visually Impaired By Reading Shop Trademarks in Shopping Mall**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.



Signature : _____

Name : Brandon Ling Yi Yun

Date : 09/09/2022

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof Leung Kar Hang, who had done his utmost to provide support to me throughout the project. I appreciate the time taken by my supervisor to provide valuable feedback to improve my project report writing.

I would also like to thank my family and friends, who have given me unwavering support from the beginning until the end of this project.

ABSTRACT

Indoor navigation for the visually impaired has been receiving much attention in recent years, and the increment of research or development related to the topic. The research and development of this system have played a vital role for the visually impaired to improvise a better future and improve their quality of life.

Signs and labels are actively used for indoor navigation systems; however, these are not beneficial for the visually impaired. Due to their disabilities, it is difficult for the visually impaired to identify shop trademarks, obstacles, and shops in a shopping complex.

As a result, this project aims to create a system that helps the visually impaired recognises trademarks, obstacles, and shops located in a shopping complex. Several components such as a camera, software application, and earbuds are required to execute the system. The camera will read images, which will then be processed by the software application recognising the character or object in the image. The information will then be converted to speech and delivered through earbuds or earphones.

The project will have several limitations, which may result in the output not being very accurate such as the quality image and processing power of the device. Both issues contribute to the performance of the system and, therefore, the accuracy of navigation to be affected.

TABLE OF CONTENTS

| | |
|--|------------|
| REPORT STATUS DECLARATION FORM | II |
| DECLARATION OF ORIGINALITY | IV |
| ACKNOWLEDGEMENTS | V |
| ABSTRACT | VI |
| TABLE OF CONTENTS | VII |
| LIST OF FIGURES | XI |
| LIST OF TABLES | XIV |
| LIST OF ABBREVIATIONS | XV |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1 PROBLEM STATEMENT AND MOTIVATION | 1 |
| 1.2 PROJECT SCOPES | 2 |
| 1.3 PROJECT OBJECTIVES | 2 |
| 1.4 IMPACT, SIGNIFICANCE AND CONTRIBUTION | 3 |
| 1.5 BACKGROUND INFORMATION | 3 |
| 1.6 REPORT ORGANIZATION | 4 |
| CHAPTER 2: LITERATURE REVIEW | 5 |
| 2.1 PREVIOUS WORK OF INDOOR NAVIGATION FOR THE VISUALLY IMPAIRED | 5 |
| 2.1.1 Indoor Navigation using Radio Frequency Identification (RFID) | 5 |
| 2.1.2 RGB-D camera based wearable navigation system for the visually impaired | 6 |
| 2.1.3 Development of an Indoor Navigation using Near Field Communication (NFC) | 7 |
| 2.1.4 Indoor Navigation and Product Recognition for Blind People Assisted Shopping | 10 |
| 2.1.5 Blind User Wearable Audio Assistance for Indoor Navigation Based on Visual Markers and Ultrasonic Obstacle Detection | 11 |
| 2.1.6 Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment | 12 |
| 2.1.7 Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System | 13 |
| 2.2 LIMITATION OF PREVIOUS STUDIES | 16 |

| | |
|--|-----------|
| 2.2.1 Indoor Navigation using Radio Frequency Identity | 17 |
| 2.2.2 RGB-D Camera Based Wearable Navigation System for Visually Impaired | 17 |
| 2.2.3 Development of an Indoor Navigation Near Field Communication (NFC) | 17 |
| 2.2.4 Indoor Navigation and Product Navigation for Blind People Assisted Shopping | 17 |
| 2.2.5 Blind User Wearable-Audio Assistance for Indoor Navigation Based on Visual Markers and Ultrasonic Obstacle Detection | 18 |
| 2.2.6 Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment | 18 |
| 2.2.7 Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System | 18 |
| 2.3 PROPOSED SOLUTION | 18 |
| 2.4 TECHNOLOGY REVIEW | 19 |
| 2.4.1 Programming Language- Python | 19 |
| 2.4.2 Firmware/ Operating System-Windows | 19 |
| 2.4.3 Yolo | 19 |
| 2.4.4 Convolutional Neural Network (CNN) | 21 |
| CHAPTER 3: SYSTEM METHODOLOGY/ APPROACH | 22 |
| 3.1 SYSTEM EVALUATION METHODOLOGY | 22 |
| 3.1.1 Confusion Matrix | 22 |
| 3.1.2 Accuracy Score | 22 |
| 3.1.3 Precision, Recall and F1-scores | 23 |
| 3.1.4 Classification Report | 23 |
| 3.1.5 System Architecture | 24 |
| 3.1.6 Datasets | 25 |
| 3.1.7 Timeline | 31 |
| CHAPTER 4: SYSTEM DESIGN | 33 |
| 4.1 PROJECT FLOW DIAGRAM | 33 |
| 4.2 SYSTEM FLOW DIAGRAM | 34 |
| 4.2.1 Shop Trademark Recognition | 34 |
| 4.2.2 Real Time Indoor Navigation System Methodology | 35 |
| 4.3 SYSTEM FUNCTION FLOW | 37 |
| 4.3.1 CNN Training Flow | 37 |

| | |
|---|-----------|
| 4.3.2 CNN Testing Implementation | 38 |
| 4.3.3 CNN Testing Modal with Camera Implementation | 39 |
| 4.3.4 Image to Speech Function | 40 |
| 4.3.5 Voice Activation Function | 41 |
| 4.3.6 Yolo Function | 42 |
| 4.3.7 Graph Implementation | 43 |
| 4.4 SYSTEM COMPONENT INTERACTION OPERATION | 45 |
| CHAPTER 5: SYSTEM IMPLEMENTATION | 46 |
| 5.1 SOFTWARE SETUP | 46 |
| 5.2 HARDWARE SETUP | 48 |
| 5.3 SYSTEM OPERATION | 49 |
| 5.3.1 Camera Display | 49 |
| 5.3.2 Speech Input and Voice Output | 51 |
| 5.4 IMPLEMENTATION CHALLENGES AND ISSUES | 52 |
| CHAPTER 6: SYSTEM EVALUATION AND DISCUSSION | 53 |
| 6.1 SYSTEM TESTING AND PERFORMANCE METRICS | 53 |
| 6.1.1 System Training result | 53 |
| 6.1.2 Test Set Predicted Result | 54 |
| 6.1.3 Confusion Matrix | 55 |
| 6.1.4 Accuracy, Recall, Precision, and F1 score Result | 56 |
| 6.1.5 Testing Report | 58 |
| 6.2 TESTING SETUP AND RESULT | 59 |
| 6.2.1 Starbuck Result | 59 |
| 6.2.2 Burger King Result | 62 |
| 6.2.3 Vivo Result | 64 |
| 6.2.4 Subway Result | 67 |
| 6.2.5 McDonald's Result | 70 |
| 6.2.6 Not Shop Trademark Result | 72 |
| 6.2.7 Voice Command Result | 73 |
| 6.3 COMPARISON OF TESTING RESULT BETWEEN REAL TIME AND TESTING DATASETS | 74 |
| 6.4 PROJECT CHALLENGES | 75 |
| 6.4.1 Unresolved Challenges | 75 |

| | |
|---|-------------|
| 6.4.2 Resolved Challenges | 75 |
| 6.5 OBJECTIVE EVALUATION | 76 |
| CHAPTER 7: CONCLUSION AND RECOMMENDATION | 78 |
| 7.1 Conclusion | 78 |
| 7.2 Recommendation | 79 |
| REFERENCES | 80 |
| APPENDIX A: WEEKLY REPORT | A-1 |
| APPENDIX B: POSTER | A-5 |
| APPENDIX C: PLAGARISM CHECK RESULT | A-6 |
| APPENDIX D: CHECKLIST | A-12 |

LIST OF FIGURES

| Figure Number | Title | Page |
|----------------------|---|-------------|
| Figure 2.1.1.1 | visualization of the system | 5 |
| Figure 2.1.2.1 | overview of the RGB-D camera based wearable navigation system | 6 |
| Figure 2.1.3 1 | overview of the RGB-D camera based wearable navigation | 8 |
| Figure 2.1.3.2 | starting point to destination point | 9 |
| Figure 2.1.3.3 | NFC system block flow diagram | 9 |
| Figure 2.1.4.2 | Blind Shopping distributed architecture | 11 |
| Figure 2.1.5.1 | the propose software architecture | 11 |
| Figure 2.1.6.1 | overview of the system | 13 |
| Figure 2.1.7.1 | system overview | 14 |
| Figure 2.1.7.2 | system algorithm | 14 |
| Figure 3.1.5.1 | system overview for real-time | 24 |
| Figure 3.1.6.1 | Burger King dataset | 25 |
| Figure 3.1.6.2 | McDonald's dataset | 25 |
| Figure 3.1.6.3 | Starbuck datasets | 26 |
| Figure 3.1.6.3 | Subway dataset | 26 |
| Figure 3.1.6.4 | Vivo datasets | 27 |
| Figure 3.1.6.5 | Not Shop Trademark dataset | 27 |
| Figure 3.1.6.6 | Burger King testing dataset | 28 |
| Figure 3.1.6.7 | McDonald's testing dataset | 29 |
| Figure 3.1.6.8 | Not Shop Trademark testing dataset | 29 |
| Figure 3.1.6.10 | Subway testing dataset | 30 |
| Figure 3.1.6.11 | Vivo testing dataset | 30 |
| Figure 3.1.7.1 | Project 2 Timeline | 31 |
| Figure 4.1.1 | project block flow diagram | 33 |

| | | |
|----------------|--|----|
| Figure 4.2.1.1 | block diagram of Shop Trademark recognition pipeline | 34 |
| Figure 4.2.2.1 | block diagram of real time indoor navigation system | 35 |
| Figure 4.3.1.1 | CNN Training Process Block Diagram | 37 |
| Figure 4.3.2.1 | CNN Testing Process Block Diagram | 38 |
| Figure 4.3.3.1 | CNN Testing Modal with Camera Block Diagram | 39 |
| Figure 4.3.4.1 | image to speech process | 40 |
| Figure 4.3.5.1 | voice activation implementation | 41 |
| Figure 4.3.6.1 | Yolo implementation | 42 |
| Figure 4.3.7.1 | graph creation implementation | 43 |
| Figure 4.4.1 | System Components | 44 |
| Figure 5.3.1.1 | camera display results | 49 |
| Figure 5.3.1.2 | graph map | 50 |
| Figure 5.3.2.1 | microphone icon pop out on computer menu | 51 |
| Figure 5.3.2.2 | unknown value input | 51 |
| Figure 5.3.2.3 | voice input “computer” | 51 |
| Figure 6.1.1 | step per epoch vs training accuracy | 53 |
| Figure 6.1.2.1 | testing results | 54 |
| Figure 6.1.3.1 | confusion matrix | 55 |
| Figure 6.1.4.1 | testing evaluation | 56 |
| Figure 6.1.4.2 | accuracy result using sklearn library | 57 |
| Figure 6.1.5 | testing report result | 58 |
| Figure 6.2.1.1 | Starbuck test result | 59 |
| Figure 6.2.1.2 | another Starbuck test result | 60 |
| Figure 6.2.1.3 | the image is shift to the right | 61 |
| Figure 6.2.2.1 | Burger King test result | 62 |
| Figure 6.2.2.2 | false prediction result of Burger King | 63 |
| Figure 6.2.3.1 | Vivo Result output | 64 |
| Figure 6.2.3.2 | placing the image closer to the camera | 65 |
| Figure 6.2.3.3 | shear angle of Vivo image | 66 |
| Figure 6.2.4.1 | Subway result | 67 |
| Figure 6.2.4.2 | the image is shift to right | 68 |
| Figure 6.2.4.3 | Another Subway testing result | 69 |
| Figure 6.2.5.1 | McDonald’s result | 70 |

| | | |
|----------------|---|----|
| Figure 6.2.5.2 | McDonald's result in a nearer view | 71 |
| Figure 6.2.6.1 | testing on the lab environment in Utar | 72 |
| Figure 6.2.6.2 | system detected shop trademark in the lab environment | 73 |
| Figure 6.2.7.1 | voice command and output result | 73 |

LIST OF TABLES

| Table Number | Title | Page |
|---------------------|-----------------------------|-------------|
| Table 2.2.1.1 | previous studies limitation | 16 |
| Table 2.4.3.1 | Yolo techniques description | 19 |
| Table 5.2.1 | hardware specification | 48 |

LIST OF ABBREVIATIONS

| | |
|------|------------------------------|
| FPS | Frame per second |
| CPU | Central Processing Unit |
| RAM | Random Access Memory |
| CNN | Convolutional Neural Network |
| YOLO | You Only Look Once |

CHAPTER 1: Introduction

1.1 Problem Statement and Motivation

According to [22], the visually impaired have trouble dealing with the environment that they are not familiar with. Thus, they require guidance all the time. The visually impaired are a community that is sidelined by society because they cannot do much work alone and must rely on others to help them most of the time. Nowadays, many products can be found in shopping malls. However, not one visually impaired has ever stepped into a shopping mall alone without the aid of others. This is because the visually impaired may injure the crowds if coming with a cane and shopping malls are crowded and noisy; it is impossible for the visually impaired to hear the sound of their aide cane. Besides that, many shopping malls do not provide any braille signs for the visually impaired. This makes visually impaired challenging to know if they are at the right shop or their current location. Lastly, visually impaired people are more likely to get injured due to knocking on objects such as a bench, flowerpots, banners, guard rails, etc. These objects are hard to avoid as the visually impaired may not know if the cane is beating on an object or a human. The motivation of this project is that the visually impaired community requires a guide to accompany them when they are going out. Sometimes, guides are not available to accompany them, which causes the visually impaired to reschedule or use a stick to go out. In a shopping mall, there are a crowd of people that which visually impaired would want to avoid because the visually impaired may hurt the nearest person with their stick. Not only that, the visually impaired uses their hearing sense to identify if obstacles are in front. However, it is noisy in a shopping mall, so it is hard for the visually impaired to hear the sound of their stick. Furthermore, the shopping complex does not provide any braille signs for the visually impaired to read; therefore, they must seek help from staff to tell them the shop's name or guide them towards it.

As a result, this project is proposed to help solve the problems faced by the visually impaired. This project will help the visually impaired be more independent and less dependent on their guide. The guide will also have more time to prioritise other matters that the system cannot handle.

1.2 Project Scopes

At the end of this project, the project will be developed with a camera that reads shop trademarks in the shopping mall. This project will be used only by the visually impaired, which will help navigate them inside a shopping complex and voice instruction to instruct the visually impaired on the direction to follow. The project will use a camera, built-in microphone earbud or earphone and Jupyter notebook software. The camera will capture the shop trademark image as an input dataset, and the software uses the OpenCV library to pre-process the image and CNN to classify the image. Next, the project uses the YOLO algorithm to detect objects in front. The project also uses OCR from Pytesseract to convert the word image to text. This will allow the system to read any signs display in the shopping mall. Lastly, using Pyttsx3 to convert text to speech.

1.3 Project Objectives

The project objectives are to help our system to achieves specific objectives, which are stated below:

- To provide voice instruction and voice command for the visually impaired to give instruction and receive instruction from the system and to read out shop trademarks.
- To implement shop trademark recognition for identifying and reading shop trademark
- To implement a system that can detect objects and humans and warn the visually impaired of the objects or person in front.

The purpose of the first objective is to help the visually impaired receive instruction from the system that the targeted shop trademark has been reached and to help the visually impaired to command the system to read out when reaching the targeted shop so that the system will not read out all shop trademarks it captures but read out only the commanded shop trademark. The next objective is to build a system that recognises shop trademarks as there are many types of designs, colours or patterns of shop trademarks; it is crucial the system recognizes. Otherwise, the visually impaired will have a hard time finding it. Finally, the final objective is to build a system that detects objects or humans so that fewer collisions will happen with the visually impaired.

1.4 Impact, Significance and Contribution

This project aims to help the visually impaired community to navigate themselves without much trouble. By achieving this, the visually impaired dares to live more independently. As a result, it will reduce the dependency on a guide when they are out alone.

Furthermore, society will benefit from this project because it helps the visually impaired community to have a higher chance of obtaining employment. Therefore, organisations could take the workforce from this community.

Lastly, the visually impaired community will participate more in various activities, and society will have a positive perspective on this community. Henceforth, the visually impaired community will have equal rights with the rest of the society.

1.5 Background Information

Shopping malls are vast and have many premises inside. These premises sell many varieties of products. The shopping mall also has many floor levels, and it is easy for some people to get lost. Since there are many shop premises in the shopping mall, people come to shop to purchase many goods. However, shopping malls have many signages to indicate direction or location to the shoppers. These signages are only applicable to people who have sight and not to the visually impaired.

Visually impaired are people born without sight. The visually impaired rely on a guide dog or cane when navigating themselves or walking outside alone. There are some technologies available to the visually impaired to use for navigation. Still, these technologies are expensive, such as smart cane, smartphones with indoor GPS, etc. smartphones have indoor GPS, but it cannot read shop trademarks and this is the same for the smart cane. There are some smart vision navigation systems for the visually impaired, but it does not have the features to read shop trademark.

An indoor navigation system that read shop trademark will help the visually impaired. The system will have a camera that views the surroundings, such as the objects or people in front, to alert the visually impaired. The system will read the shop trademark and inform the visually impaired. The visually impaired gives the system a command indicating the shop it is looking for so the system can inform the visually impaired if it captured it.

With this navigation system, the visually impaired will easily find the target shop trademark and avoid many objects in the shopping mall. The system will also help the visually impaired know if a person is in front, which will help the visually impaired when facing troubles. The

system will reduce dependency on a guide or cane and help provide the visually impaired with the same privilege average everyday person.

1.6 Report Organization

In the first chapter, the overview of the project such as the problem statement and motivation, project scopes, objectives, impact significance and contribution, and background information. In the second chapter, some research work such as reviewing existing work system and solutions that could be used in the project. This chapter also reviews the previous technology used by the authors and to identify the weakness of previous existing work together with the limitations it faces. This chapter also includes finding the most suitable solution. Chapter 3 covers mainly on the approach and methodology used in the project such as the software, components, and system specification. System specifications such as the methodology, tools, and system requirements are being discussed in this chapter and also identifying the implementation issue are also discussed. In chapter 4, it describes the system designs such as block diagrams and the flow of the system. In chapter 5, it will explain the system implementation such as software setup, configurations, and system operation. Chapter 6 covers the system performance and results. In this chapter, it will explain the result obtained, challenges being faced and objective remarks. Lastly, Chapter 7 will conclude the project and summarise the findings and the solution to solve the problem statement. This chapter will provide recommendations for improvement so that the project could perform better in the future.

CHAPTER 2: Literature Review

There are many ways of creating an indoor navigation system for the visually impaired. Some methods use visible light communication, computer vision guidance system, radio frequency identification (RFID), NFC, etc. These methods are explained below.

2.1 Previous work of Indoor navigation for the visually impaired

In this subsection will be explained each previous work done by others. These works will help in chapter 4 as the project will require a solution to solve the problem statement shown in 1.2.

2.1.1 Indoor Navigation using Radio Frequency Identification (RFID)

The author [3] uses an Arduino microcontroller, RFID Bluetooth chip, and a power regulator. The indoor navigation system that the author introduced is known as the PERCEPT. These components are then assembled and installed on a glove. The glove was to communicate through passive RFID tags and use Bluetooth technology with the android-based smartphone. The article also stated that the RFID tags are installed on each building door, and the microcontroller will keep track of the user action and the environment. The Bluetooth act as a transmission to exchange data between the microcontroller and the smartphone. The smartphone connects to the WI-FI, which establishes a connection between the smartphone and the PERCEPT server. The PERCEPT server will return navigation instructions via WI-FI and to the smartphone. This idea allows the user to know the surroundings in the building and is easy to implement as it does not require much material.

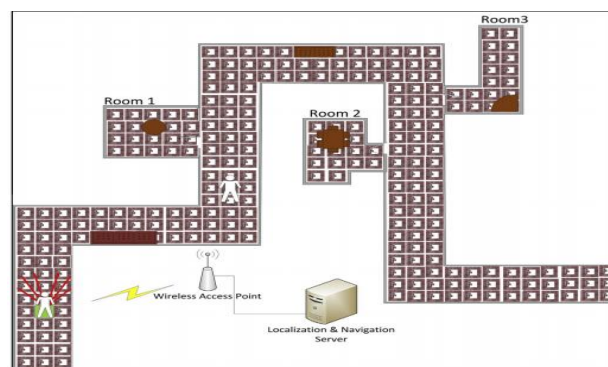


Figure 2.1.1.1: visualization of the system [3]

Figure 2.1.1.1 shows the system from [3], where a centralised server directs the user to their desired destination. The user walks through all corridors in the building while being equipped with wearable components, following instructions given by the central system via the

embedded headset. The system detects obstacles at any given time with the help of continuous localisation surveillance, which allows users to move safely and avoid obstacles.

2.1.2 RGB-D camera based wearable navigation system for the visually impaired

The author uses a head-mounted RGB-D sensor camera, smartphone user interface, and navigation software to build the mentioned project [19]. The article stated that the navigation software consists of modules such as a hybrid pose estimation algorithm, mapping algorithm, and dynamic path planning algorithm. Based on the article, the system will interact with the visually impaired for a start-up navigation task. The user uses voice commands with minimal information such as building name, room name or no, starting location address, etc. The starting location will require a localisation algorithm to correct minor deviants, approximating the current location. The author also mentioned they adapt a real-time navigation algorithm which acts as an example of the blind aided by a sighted person. The system stores information about the blind man travelling the path from the starting to the destination for reuse.

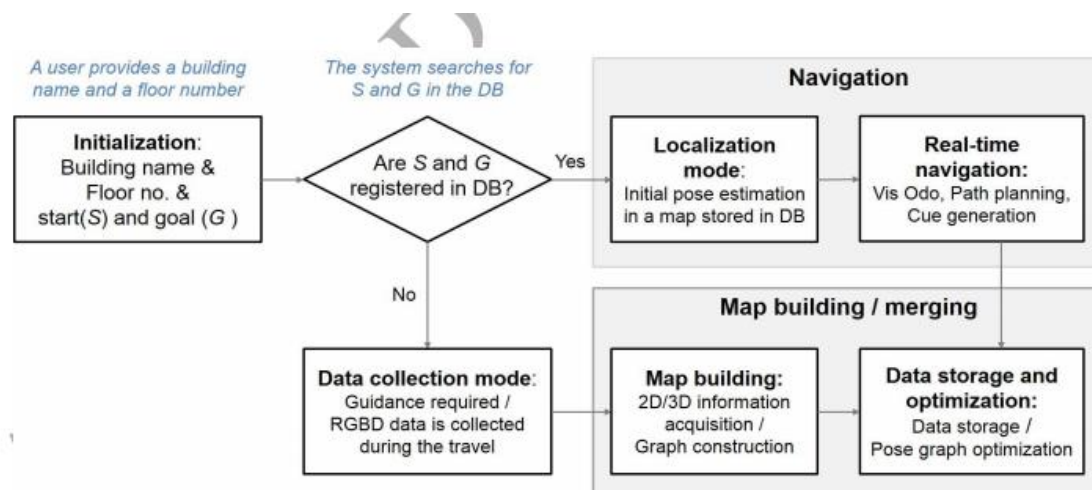


Figure 2.1.2.1 overview of the RGB-D camera based wearable navigation system [19]

From figure 2.1.2.1 shows the overview of the system. The system is initialised when the user provides information about the building, including its name and floor number. Next, the system determines start (S) and goal (G) registered in the database. If the system finds that the (S) and (G) are in the database, it will move towards localisation mode to determine the user's initial location. After that, the system will start path planning to record the user path and store the path in the system, which will be reused the next time. The path will store in a data storage. On the other hand, if the system cannot identify the (S) and (G) from the database, the system will collect the data by capturing the surrounding environment from the RGBD camera during

travel. The system will draw out the building map with 2D or 3D information acquisition and, lastly, store the information in the database for future use.

The article also mentioned that the author used an IMU sensor to read orientation prior and normal of point clouds using real-time normal estimation algorithm. They first get the major point cloud with a normal parallel to the gravity vector. They used the RANSAC-based least squared method to find the plane coefficient and find the major plane $\Pi_G = (\mathbf{n}, D)$, where $\mathbf{n} = (A, B, C)^T$ is a normal vector of a plane. The article also mentioned that the blind user pose is represented by T_n where n represents the frame number shown below which was obtained from the article:

$$T_0 = \begin{bmatrix} \mathbf{R}_0 & \mathbf{t}_0 \\ \mathbf{0}^T & 1 \end{bmatrix}, \text{ where } \mathbf{t}_0 = (0, 0, 0)^T$$

The RGB image will be converted to grayscale and using fixed-size gaussian kernel to smooth the image. Then they construct a Gaussian pyramid to detect any robust features at each gaussian pyramid level. The image of each level is divided by 80 x 80 sub-images. Fast corners are extracted, and any FAST corner associated with invalid depth will be discarded. The article also stated that a 9 x 9 square patch around each FAST corner was used as a feature for matching. Sum-of-absolute-difference (SAD) was used as a match score. Matches with reprojection error higher than the threshold were discarded from the inlier set and refined the motion again to obtain the final motion estimation. The path planning was performed using 2D space to reduce computation complexity in certain stages, while traversability was analysed in 3D.

2.1.3 Development of an Indoor Navigation using Near Field Communication (NFC)

NFC is a bidirectional range, wireless communication technology, and **figure 2.1.3.1** shows how NFC works [2]. It uses the slide rule interface to overcome the accessibility barrier of the touch screen by introducing the talking touch-sensitive interface, which uses a speech-based interface without any visual representation. It stated that the user navigation brushes their fingers up and down on their device to scan on-screen objects and used gestures to respond to the on-screen object they encountered. They also mentioned that users are guided by speech and non-speech devices whenever they move around. The article also stated they believe NFC technology to be a suitable solution for a navigation system as it allows mobile phone with GPS system to function not only outdoor but indoor as well. The article stated when a user

touches a URL tag which consist of the indoor map information on the map server, it will allow the mobile phone to obtain the address. The tags are placed in front of the building. The phone will connect to the URL via ORL which requests the map information from the map server. After that, the navigation application on the smart card will convert the map into 2D network with the topological relationship. The application will prothe usersuser of their destination by just specifying the person's, name and the application will use Dijkstra algorithm to find the shortest path.

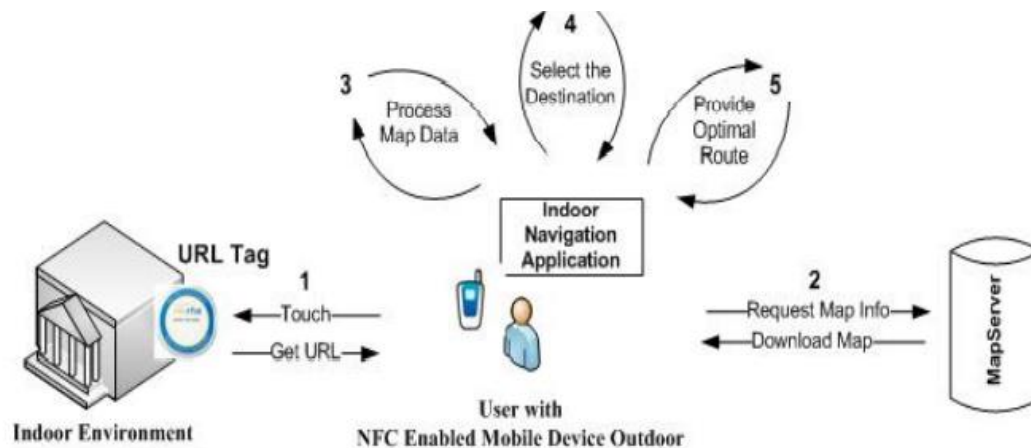
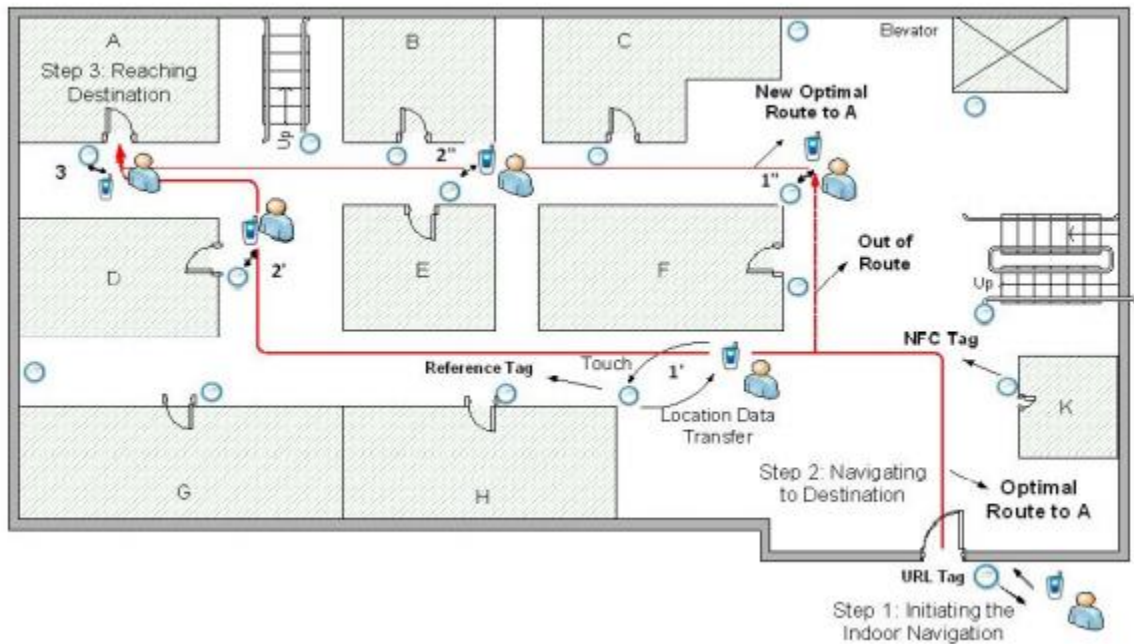


Figure 2.1.3.1: overview of the RGB-D camera based wearable navigation system [2]

The indoor environment of a building or complex has large number of reference tag according to the article [2]. The reference tag includes information of the location which contain of building identifier data, vector spatial data and floor identifier data. The author used vector spatial data because it allows efficient encoding topology, and network linkage can be employed efficiently [5].



2.1.3.2: starting point to destination point [2]

Based in **figure 2.1.3.2** shows the user travel in a building from the starting point and arrive to the destination point. Based on the article, the user touches the first reference tag she finds then the location data on the tag is transferred to the user mobile phone. The application will determine the path for user and provide user voice instruction such as 'turn left', 'turn right', etc. the user can touch the reference tag on her way to allow the system to know if the user is on the right path. By following the instruction given by the system, the user will arrive to the destination.

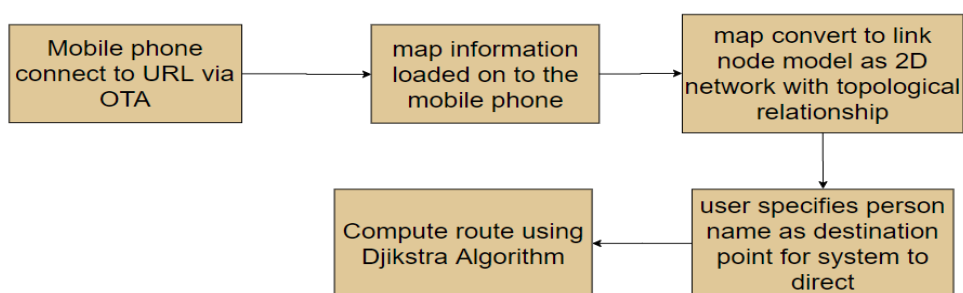


Figure 2.1.3.3 NFC system block flow diagram

Figure 2.1.3.3 shows the block diagram of the NFC system. It started when the mobile phone connects to a URL via OTA. A map information will be loaded on to the phone. Next the map converts to link node model as 2D network with topological relationship which will provide all the layouts. After that, user will only specified the person name as it is a destination point

for user to arrive and the system to direct. The system will use djikstra algorithm as a method to find the shortest path to allow user to arrive their destination point with the shortest point.

2.1.4 Indoor Navigation and Product Recognition for Blind People Assisted Shopping

The system used a platform known as “BlindShopping” which uses a map to identify ID from RFID that consist of cells to navigation, and product location information [6]. The platform also offers infrastructural support for whole purchasing process within the supermarket. The system consists of a navigation component that drives the user via audio messages to the aisle of the product that was located by their smartphone. The system has product recognition that can scan QR or UPC. According to the article, it implements the system into a smart phone where a Java ME application was developed to read RFID tags and deliver the code to an android application mobile phone through Bluetooth. The author also uses a Baracoda Tagrunner which is a portable RFID reader. This method will allow the blind person to opt an action such as using gesture or voice command via their smart phone. The blind person would either draw a “L” or issue command “location” to launch the supermarket navigation system which shown in **figure 2.1.4.1**. the navigation system will prompt user to either touch RFID floor marking or QR code that is attached to a shelf to determine user location. The system will prompt user of the product category they are searching. The smartphone will always maintain Bluetooth connection with the RFID reader to keep track of the blind person location. By issuing the command “product” or drawing a “P”, the system will ask the user to hold their smartphone to point on the shelf so that the camera can scan the the QR code and inform user of the product detail. The RFID tag on the floor will notify the system about the user location.



Figure 2.1.4.1 Baracoda Tagrunner [6]



Figure 2.1.4.2: BlindShopping distributed architecture [6]

Figure 2.1.4.2 shows the BlindShopping distributed architecture that was done by the author [6]. The mobile phone act as a mediator which take command from the blind man, scan QR or UPC to determine current location, produce voice instruction to the blind person. A RFID tag reader connected on the cane will read the tag on the floor and transmit the data via Bluetooth to the phone. A web service query is connected to the phone via WIFI will retrieve data from the server to the phone with the details when the blind person scans the QR or UPC.

2.1.5 Blind User Wearable Audio Assistance for Indoor Navigation Based on Visual Markers and Ultrasonic Obstacle Detection

The device consists of hardware module and software module. The user receives voice instruction through their headset shown in **figure 2.1.5.1** below together with the proposed software architecture [18].

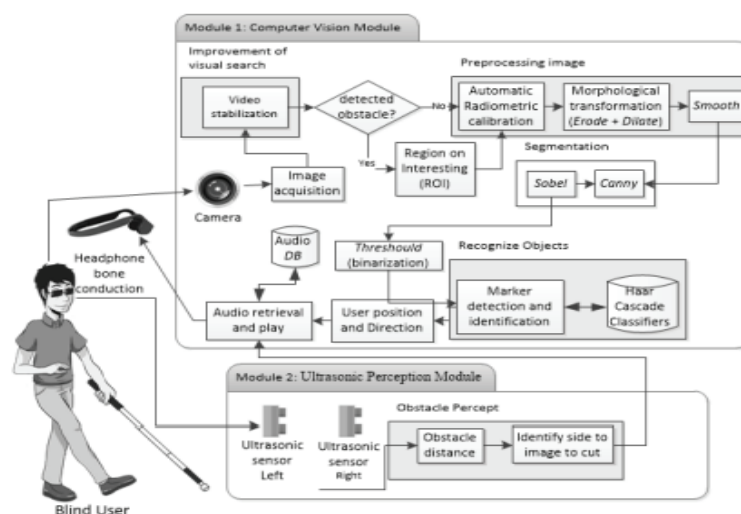


Figure 2.1.5.2: the propose software architecture [18],

The author uses a glasses model with RGB camera, two ultrasonic sensors and low-cost mini-PC to run the algorithm whereas the software module consists of Computer Vision module and ultrasonic perception module that operates in parallel mode. According to the article, the computer vision is used for recognizing the visual marker and obstacles in the environment. It will perform video stabilization to correct user motion while user is walking or moving. The author also uses image processing to adapt system brightness variation and smoothen noise effect. After image pre-processing, image segmentation is carried out. the author uses Sobel and canny filters to highlight or rebuild the thick and thin edges of the images. the author uses OpenCV as it consists of three of three algorithms to build Haar-like cascade which are Objectmarker, CreateSample and TrainCascade. The map of indoor environment is performed using printed visual marker arranged in known point. The marker attributes such as ID, audio information and other markers relationship are recorded in the database. The article mentioned that the marker identification process using proximity method and visual pattern analysis. The proximity method uses relative location and symbolic marker.

2.1.6 Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment

The system used a smartphone based indoor navigation system known as NavCog3 for visually impaired [8]. According to the article, it stated the system provides a turn-by-turn instructions to help visually impaired to correct their turns without visual aids. The system also provides nearby landmarks so that user can walk comfortably and confidently. According to the author, he uses a localization technique with a particle filter which allows the combination of BLE, beacon fingerprint, and pedestrian dead reckoning (PDR). The article mentioned the PDR component uses a filter to recover delay caused by the smoothing of BLE beacon RSSI. As a result, the component can be used to track user movement and makes localization stable. The beacon is placed on a suitable height where signals can receive by user. Next, the author developed a fingerprint machine that uses L1-DAR sensor to scan the environment and obtain the coordinates with error of centimetres-level. This will reduce radio-map creation time by 1/20 compared to point-to-point manual fingerprint according to the article [8].

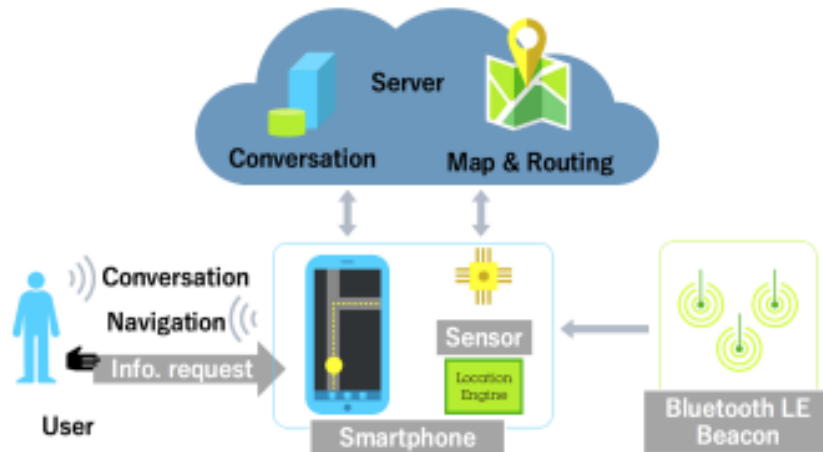


Figure 2.1.6.1: overview of the system [8]

The author also mentioned that the NavCog3 generates the navigation in-structure because of these route and semantic features. A conversation server is also implemented by the author to extract the search condition to filter shops that meet the condition where a user speech input is transcribed and submitted to the server. The server will return recommendation result else it will return conversation script.

Figure 2.1.6.1 shows the system overview of the author work. The user sends a request via voice toward the mobile phone. The mobile phone receives the request from the user and send a request to the server, the server will return the recommended result to the user if the value of the result is high otherwise, it will return conversation script according [8]. Beacons are used for user to connect to the server to obtain the map of the shopping mall. The beacon requires user to scan using LIDAR sensor which transmit the map routing to the server which the server will transmit the map routing to the user phone. User phone will use speech to inform user of the current location and instruct user of the direction.

2.1.7 Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System

The system was built with the reference of Sitixel World with journal article written by [9], [7], [17], [12]. Stixel World displays the scene using a few upright objects on the ground plane. The system was built with a miniature device to predict the surroundings to divide cloud data into free space and obstacles for blind man to walk through complex environment. The system consists of movable camera to provide onboard object detection in real time. It also has unobstructive haptic feedback that is given to user through a belt with vibration motor. According to the article, **figure 2.1.7.1** shows the system architecture and key capabilities. The

environment sensing is achieved by using a light structure camera that measure the depth of the field. It works best on indoor as it can detect unstructured wall, non-reflective surface, and non-absorptive surface. The algorithm being used are C++, OpenCV, and Point Cloud Library. The author mentioned that the system has the perception capabilities that consists of two component such as (1) free space searching, detecting obstacle and corresponding distances, and (2) object type recognition. They obtain the free parsing using a method mentioned by [12], the parsing uses algorithm 1 shown in **figure 2.1.7.2** where it traces ground height changes when user approaches a stair or an object.

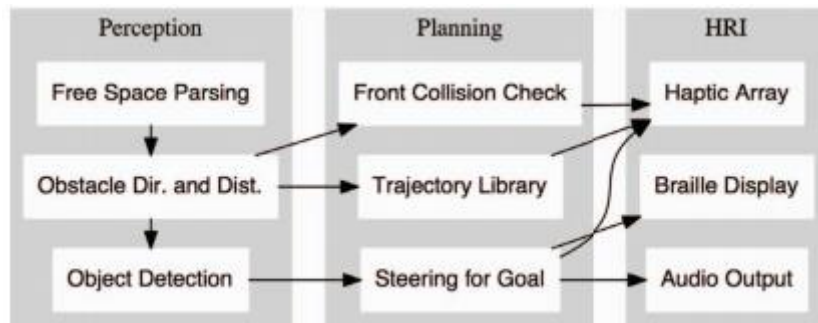


Figure 2.1.7.1: system overview [10]

Algorithm 1 Free Space Parsing

procedure FREESPACEPARSE(C)

Input: Point cloud C

Estimate surface normal N

Find ground plane G from the Stixel World, and estimate normal vector n_g of the ground plane

Rotate C based on n_g , obtain ground height h_g to translate the point cloud C

Compute ground-to-image frame transformation T_g

Find the occupancy grid, and extract the free space and distances d_o to objects/obstacles

Output: Free, walkable space

Figure 2.1.7.2: system algorithm [10]

They implement using resolution of 320 x 240 at 10 frame per second at point cloud C . The obstacles direction and distances can use full-scale window search to achieve computational efficiency for object recognition. For the object recognition, they used another algorithm which is a depth-based method. They used a feature vector of 2(vertical or horizontal) x 8 (height) and

a linear classifier for object classes according to [8]. Besides that, they implement Bluetooth communication for communicating between computer unit and receiver unit. The signal receiver will parse the information and actuate the haptic motor or braille display. The audio output will be synthesized with text-to-speech engine.

2.2 Limitation of Previous Studies

The limitation of previous studies is shown below in **table 2.2.1.1**.

Table 2.2.1.1:previous studies limitation

| Previous Studies | Limitation |
|--|--|
| Indoor Navigation using Radio Frequency Identification (RFID) | <ul style="list-style-type: none"> • Bluetooth drain phone battery quickly. • Every layout of the building must have at least one RFID tag which can be costly if the layout is big. |
| RGB-D camera based wearable navigation system for the visually impaired | <ul style="list-style-type: none"> • System could not merge both 2D and 3D images. • Voice command is not applicable in the system due to external noise interference. |
| Development of an Indoor Navigation using Near Field Communication (NFC) | <ul style="list-style-type: none"> • System does not detect obstacles. • All building layouts must have at least an URL tag which is costly if the layout is big. |
| Indoor Navigation and Product Recognition for Blind People Assisted Shopping | <ul style="list-style-type: none"> • Problem for visually impaired pinpointing their phone towards the QR code or Baracoda Code. |
| Blind User Wearable Audio Assistance for Indoor Navigation Based on Visual Markers and Ultrasonic Obstacle Detection | <ul style="list-style-type: none"> • Light intensity would affect the camera to not capture the marker. |
| Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment | <ul style="list-style-type: none"> • the system has lower accuracy in detecting small target such as elevator button, doorknob, etc. |

| | |
|--|--|
| | <ul style="list-style-type: none"> ● beacon uses radio signals that can be interfere easily |
| Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System | <ul style="list-style-type: none"> ● System could not operate well with audio feedback due noise interference from outside environment. |

2.2.1 Indoor Navigation using Radio Frequency Identity

The limitation of these study is due to the use of hardware component such as Bluetooth, and RFID tags. Bluetooth consume a lot of power which will drain the phone battery fast. Visually impaired will required turn of their Bluetooth for a long period which will result battery to finish quickly. The number of RFID tags required depending on the size of interior layout of the building, it will be costly to place every RFID tag in every section of the building. Not only that, but RFID also cannot help visually impaired to avoid any obstacles in front.

2.2.2 RGB-D Camera Based Wearable Navigation System for Visually Impaired

The limitation of this research is that the system could not merge both 2D and 3D images which means it can only perform one type of image. This system is very sensitive to external sounds thus may not respond to user voice command although user voice is louder.

2.2.3 Development of an Indoor Navigation Near Field Communication (NFC)

This system does not detect objects in front of the user. As a result, the user may still knock over any obstacles in front and hurt themselves. The system requires the whole building to have an URL tag paste at each room or corridor. This is costly if the building layout is big then every floor or room require to have at least one URL tag for the visually impaired to use.

2.2.4 Indoor Navigation and Product Navigation for Blind People Assisted Shopping

The limitation of this work is that the system requires visually impaired to scan QR code or RFID to determine user location. This in fact is difficult for visually impaired to pinpoint their phone towards the tag or QR code or Baracoda code. It is also difficult for user to hold their phone properly which may not scan the code successfully and thus, user may have to seek help from staffs or shoppers to help them scan.

2.2.5 Blind User Wearable-Audio Assistance for Indoor Navigation Based on Visual Markers and Ultrasonic Obstacle Detection

Based on the article, it stated the issue of the system limitation which is the camera cannot capture image if light intensity is high. The system will process longer or produce error if capture noisy image. The user will be confused at the moment when the system produce error or still processing.

2.2.6 Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment

The limitation of this article is that the system has lower chance in detecting small objects such as switch, elevator button, doorknobs, etc. according to the article, the author did not control participants exposure to their previous version of their work that causes an effect on the navigational error or subjective response. The author work did not test out on larger participant which the result could vary from the current result they obtained.

2.2.7 Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System

The limitation of this literature review is that the camera will capture blur image because of user movements which is difficult to capture object properly. The author uses sliding window classification which is too expensive for computation in real-time implementation. Other than that, the beacon uses radio wave which can be interfere if the number of people passing by the beacon. The beacon radio signal is short distance which user need to be closed to the beacon to receive connection.

2.3 Proposed Solution

This project aims to propose a solution that overcome the problem faced by visually impaired which is to help them read shop trademarks. The solution is to use computer vision, YOLO, CNN, and Pyttsx3 2.90 to produce a system that can read shop trademarks and produce voice instruction to the visually impaired. The system will use a camera to read shop trademark, surroundings, and obstacles so that user does not need to scan any tags or code to know their location. The system will run on a laptop as it consists of both Bluetooth and WIFI which the visually impaired can used a Bluetooth ear bud with built-in microphone or an earphone or a headset with microphone to connect to the laptop and listen to its instruction or giving commands.

2.4 Technology Review

2.4.1 Programming Language- Python

Python language is used in this project because of the wide number of libraries and the simplicity of coding the program. Python is a higher-level language which means its syntax is close to human language. Python can be used in web development, machine learning model, Artificial Intelligence, connecting database systems, etc. Python can run in any operating system such as Windows, Linux, Mac OS and raspberry pi. The syntax of python is lesser hence developers could code a program with fewer lines as a result reduce time in coding. The project runs python for it has many deep learning libraries such as Tensorflow, Pytorch, or Opencv which will help in creating models and running it. With lesser lines to code, it could simplify the work as lesser syntax errors would occur.

2.4.2 Firmware/ Operating System-Windows

The operating system that will be used in the project is windows. Windows is an operating system that is compatible to run many software's as compared to other operating systems. The window version that is used is windows 10 and this operating system comes in many updates which enhance the system security. The project will use windows as the platform to run the system.

2.4.3 Yolo

Yolo is an algorithm or model that uses a neural network to classify objects in real time. The algorithm is popular nowadays due to fast and accurate performance. Yolo uses CNN model to detect objects and only requires a single forward propagation through a neural network to detect objects[28]. The Yolo uses three techniques such as residue blocks, bounding box regression and intersection over union (IOU). Each technique are explained in the table below:

Table 2.4.3.1 Yolo techniques description [29]

| Techniques | Description |
|---------------|---|
| Residue block | <ul style="list-style-type: none"> The image is divided by different grids and each grid has a dimension of (w x h) [29]. An object will appear in every grid cells and be detected by the grid cells if the object is inside |

| | |
|-------------------------------|---|
| | the grid cell |
| Bounding box regression | <ul style="list-style-type: none"> The bounding consists of several attributes such as the width, height, class, and bounding box center which help for prediction of an object by predicting the width, height, and class of an object using probability to obtain the result for detection. |
| Intersection over union (IOU) | <ul style="list-style-type: none"> IOU provides an output box to surround an object perfectly [29]. The IOU equals to 1 when the actual box and bounding box are the same. IOU uses confidence score to determine the object class which obtained through the prediction of the bounding box done by each grid cells |

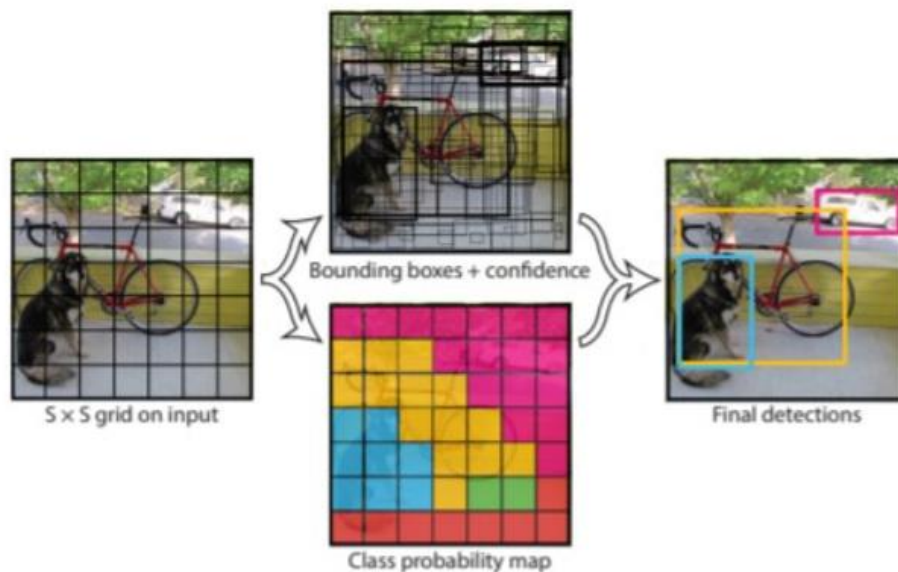


Figure 2.4.3.1 dog, cat, and bicycle [28]

Figure 2.4.3.1 shows an image of a dog, cat and a bicycle. The Yolo algorithm was applied using three of the techniques mentioned above. The image was divided into grid cells Each grid cell produces N bounding boxes which helps to obtain the confidence scores. The cells predict the class probabilities to form the class of each object [28]. Intersection over union (IOU) helps to ensure that the actual box and predicted box are equal, and this will help remove

any unimportant bounding boxes that do not match the characteristics of the object. The final detection is done when the unique bounding boxes fits perfectly on the object [28].

2.4.4 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is an artificial neural network (ANN), CNN composed of neurons that are capable of optimizing itself via learning [29]. The neurons will each receive an input for processes such as scalar followed by a nonlinear function. CNN contains multiple layers where the layer purpose is to detect the features in the input image. The last layer of the CNN contains the losses of function corresponding to the classes. CNN is widely used in pattern recognition within image, and it helps to encode the image-specific features into the architecture, making the network more suited for image-focused tasks hence reducing the parameters needed to build the model [29]. CNN contains 3 layers which are convolutional layer, pooling layer, and fully connected layer. The convolutional layer determines the output of the neuron through calculating the scalar product between the weight and the connected region of the input volume of the image. The pooling layer helps reduce the size of the image and preserve the important characteristics of the image. The pooling layer could reduce the number of parameters in the activation. The fully connected layer will produce class scores which are obtained from the activation to be used during classification.

CHAPTER 3: System Methodology/ Approach

The processes of the project were categorized into different phases in the development, which were project pre-development, data pre-processing, model training architecture building and data training, and prediction on test dataset.

3.1 System Evaluation Methodology

3.1.1 Confusion Matrix

In the confusion matrix, the sample will be categorised as 4 types: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). True Positives refers to the modal result that predicts the image correctly in the positive class. True Negatives refers to the modal result that predicts the image in the negative class correctly, ie. The image of a dog is predicted correctly by the modal as dog and there is not any image that is not dog being predicted as dog. False Positive refers to the modal predicted wrongly on the image in the positive class. For example, the image is a dog, but the system predicts the image as a dog. False Negative is the opposite of false positive which means the modal falsely predicts a negative class. Example, the image is not a dog, but the system predicts the image is a dog. The confusion matrix will provide a clear view of the number of images being predicted correctly by the system. As a result, the confusion matrix will help determine the accuracy, precision, recall and F1-score of the system.

3.1.2 Accuracy Score

Accuracy score is the measurement on the system performance of a classifier model. However, accuracy may not provide the correct measurement if the datasets are skewed. For example, the modal could produce 95% accuracy, but many predictions are incorrect on the selected data if the particular data only occupies 5% of the entire sample. Confusion matrix is required to determine the actual number of correct predictions made by the model. The formula to measure accuracy is the total number of correct predictions divided by the total number of samples.

$$\text{Accuracy} = \text{TP} / (\text{TP} + \text{TF} + \text{FP} + \text{FN})$$

3.1.3 Precision, Recall and F1-scores

Precision is the measurement of the number of predicted positive classes that were actually positive. The formula to calculate precision is stated below:

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

Recall is the measurement of the number of all positive classes that were actually predicted correctly by the model. The formula for recall is shown below:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

F1-scores is the measurement of both precision and recall. F1-scores use harmonic mean instead of arithmetic mean to remove extreme value [23]. The formula for F1-score is shown below:

$$\begin{aligned} \text{F1-scores} &= (2 * \text{Precision} * \text{Recall}) / (\text{Recall} * \text{Precision}) \\ &= \text{TP} / (\text{TP} + (\text{FN} + \text{FP}) / 2) \end{aligned}$$

3.1.4 Classification Report

The classification report provides the summary of the precision, recall and F1-score of each image class. The report helps to view the number of actual predictions that are true or false. The report also helps to determine which average is the most suitable to choose if the datasets are unbalanced. The report will include macro-average, weighted average and sample average. Micro-average will also be displayed when there is multi-label or multi-class with a subset of classes, as it corresponds to the accuracy and is the same for all the metrics. Weighted average takes each score per class multiply each class class support. Macro and Micro average are calculated as shown below:

$$\text{Macro-average} = \text{total of unweighted mean of each label} / \text{number of labels}$$

$$\text{Micro-average} = \text{total of } (\text{TP} + \text{FN} + \text{FP}) / \text{number of labels}$$

3.1.5 System Architecture

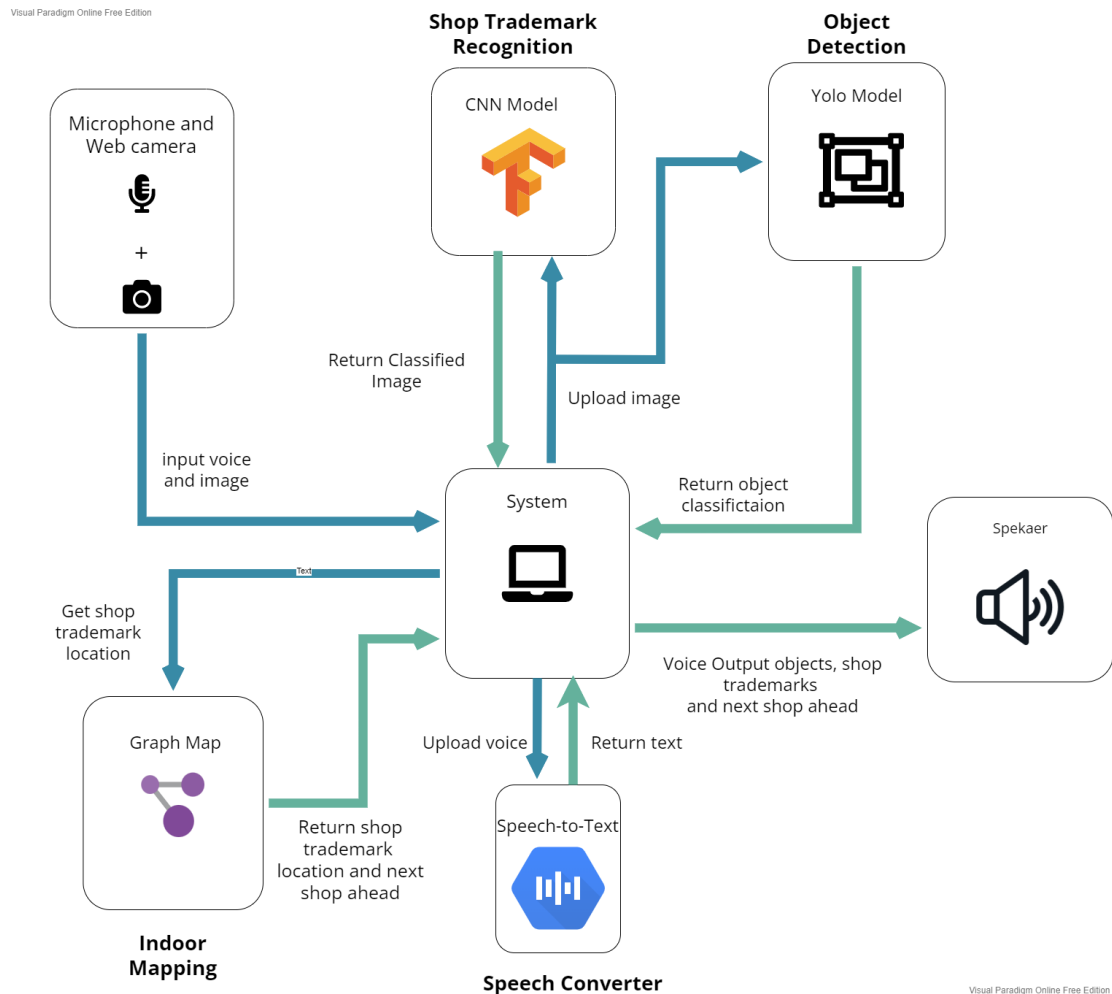


Figure 3.1.5.1 system overview for real-time

Figure 3.1.5.1 shows the overview of the system. The Microphone will first receive the input of the voice command from the visually impaired and the camera at the same time snapshot the surrounding environment. These inputs will pass into the system for processing. The system will process the speech to text through the speech converter and receive the text result. The system will upload the image into 2 models which are Yolo and CNN. The CNN model will process the image and classify the image with a shop trademark or not a shop trademark as the result. The CNN model will return the classified image back to the system. The system at the same time uploads the image into the Yolo model where the model will detect any objects in the image and return to the system. With the return classified image from the CNN model, the system will pass the value to the graph mapping function. In this function, it will verify the current shop trademark location and search the next shop ahead. The function will return both the

shop trademark location and next shop ahead to the system. The system will finally output all results such as the object detected, current shop trademark location and the next shop ahead.

3.1.6 Datasets

The CNN model requires training to enable the modal to recognise different shop trademarks hence, datasets are required to train and test the modal. The training, validation and testing sets will contain 6 classes of shop trademarks. The training datasets contain 113 images. The validation datasets contain 93 images and the testing dataset contain 103 images. The 6 classes that are trained, validated and tested are Burger King, McDonald's, Starbuck, Subway, Vivo and Not Shop Trademark. The Not Shop Trademark datasets are included to allow the CNN model to predict not shop trademark if the camera capture random objects without any shop signage shown.

Training Dataset



figure 3.1.6.1 Burger King dataset

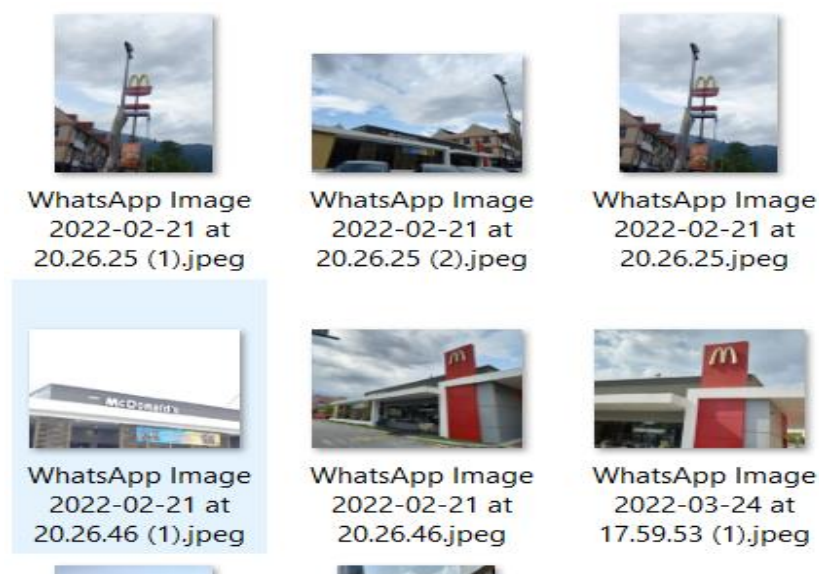


Figure 3.1.6.2 McDonald's dataset

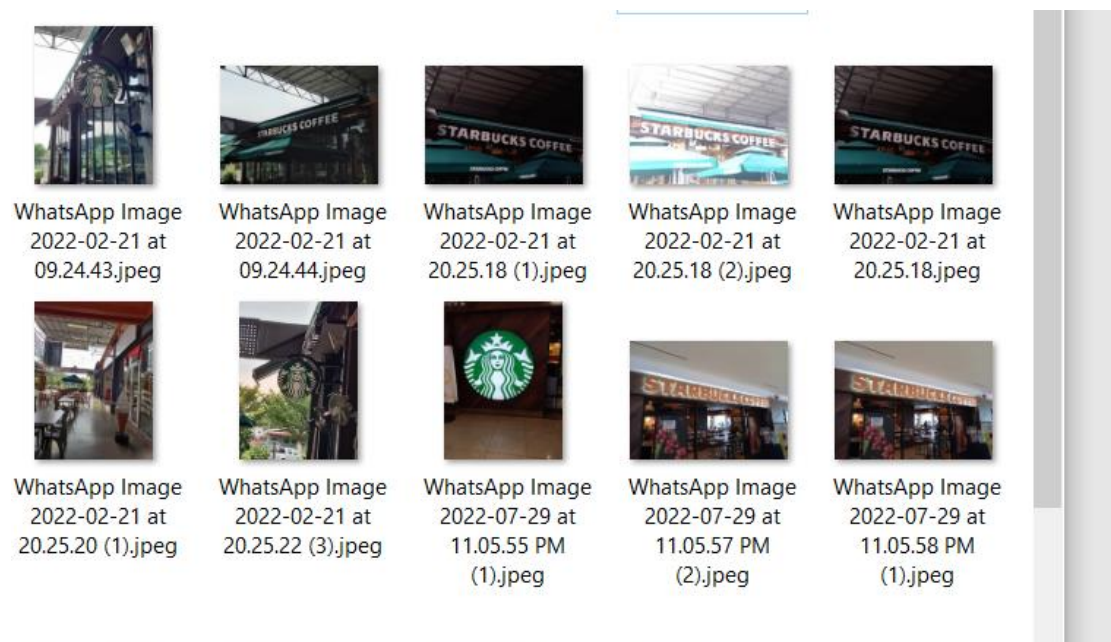


Figure 3.1.6.3 Starbucks datasets

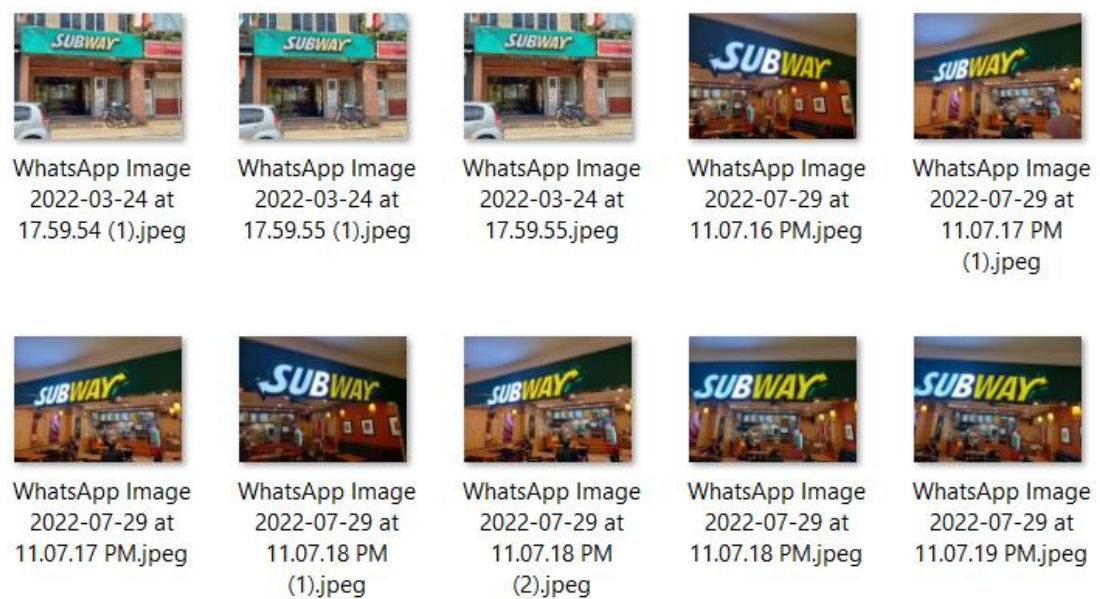


Figure 3.1.6.3 Subway dataset

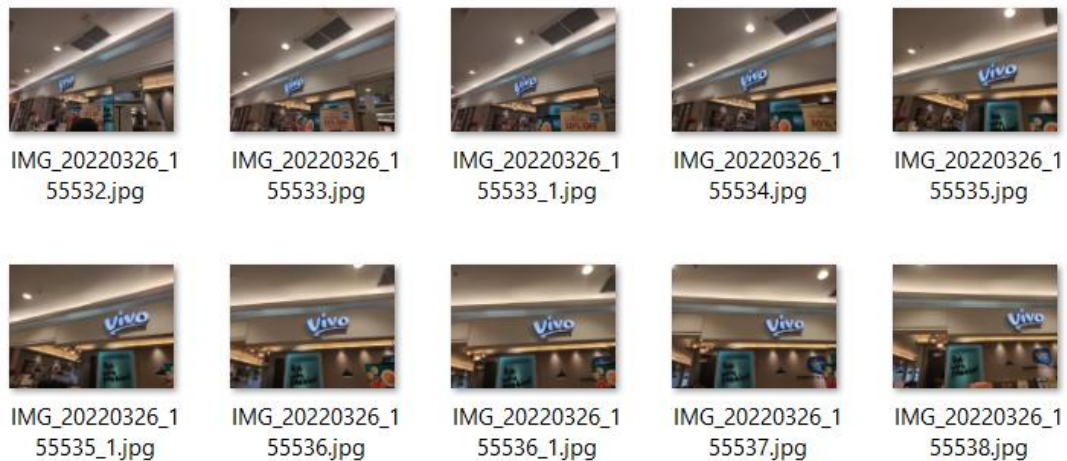


Figure 3.1.6.4 Vivo datasets

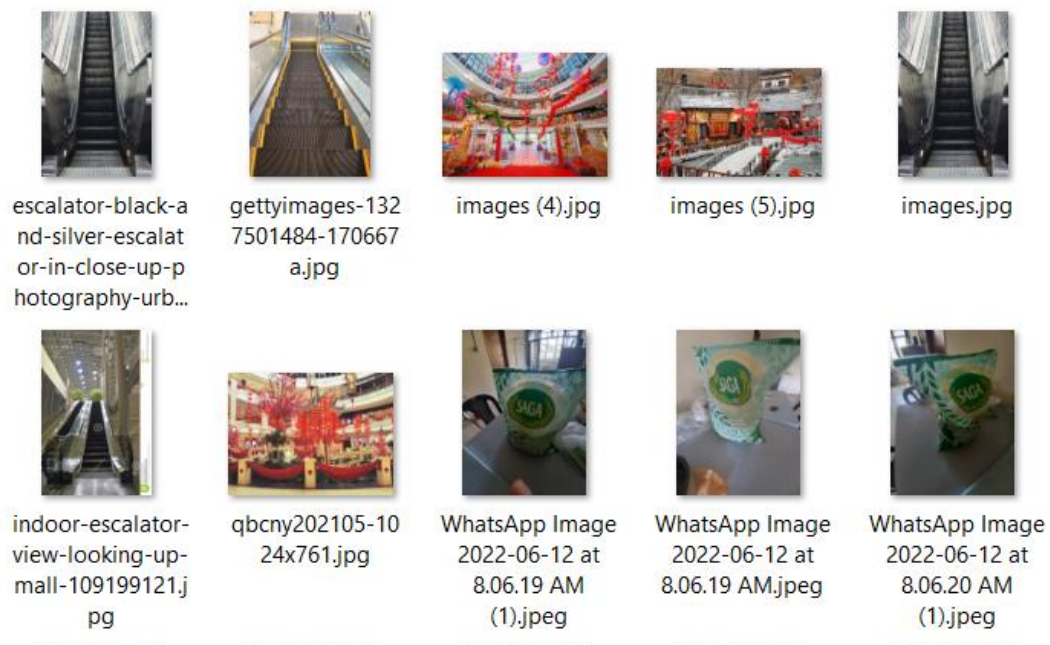


Figure 3.1.6.5 Not Shop Trademark dataset

Figure 3.1.6.1 - figure 3.1.6.5 show the datasets that are used for training. There are 14 pics of Burger King, 12 pics of McDonald's, 38 pics which do not have any shop trademark hence being classified as Not Shop Trademark. Since there are no modals in the system that detect shop trademarks, using these datasets to allow the CNN modal to recognize the environment without any shop trademark provided. There are 24 pics of Starbucks, 15 pics of Subway and 10 pics of Vivo. All of these datasets will be trained in the CNN modal so that it could predict shop trademarks when performing real-time.

Validation Dataset

The validation dataset plays a role in the CNN training as it helps the system to learn. The dataset consists of a total of 93 images of all 6 classes. The dataset contains 16 pics of Burger King, 10 pics of McDonald's, 25 pics of Not Shop Trademark, 17 pics of Starbuck, 19 pics of Subway and 6 pics of Vivo. All these images may have some slight differences with the training dataset and testing dataset such as the angle captured, and light intensity. When the modal is training, it will predict the images in the validation set to help the modal to learn and improve its accuracy and validation accuracy.

Testing Dataset

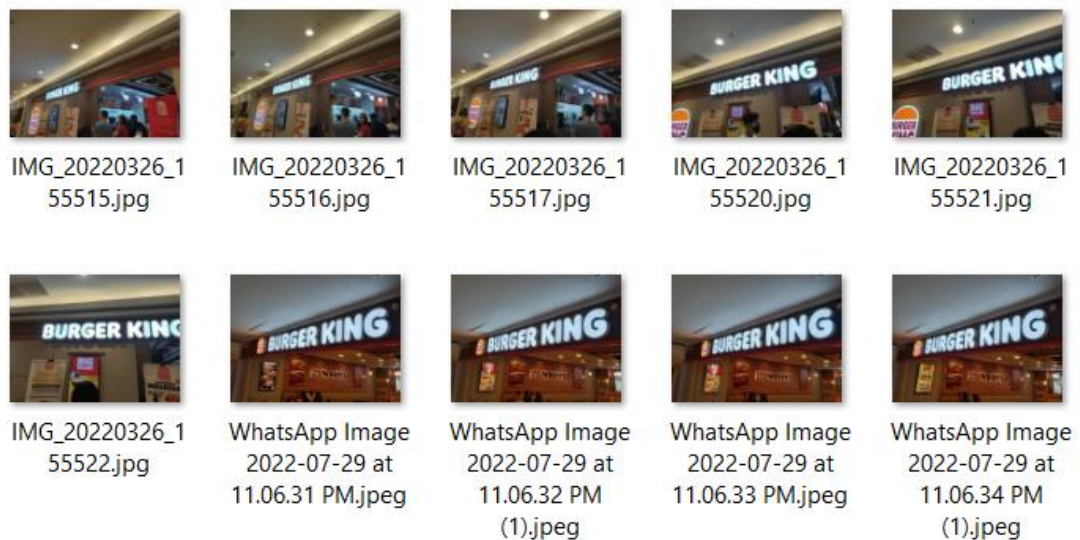


Figure 3.1.6.6 Burger King testing dataset

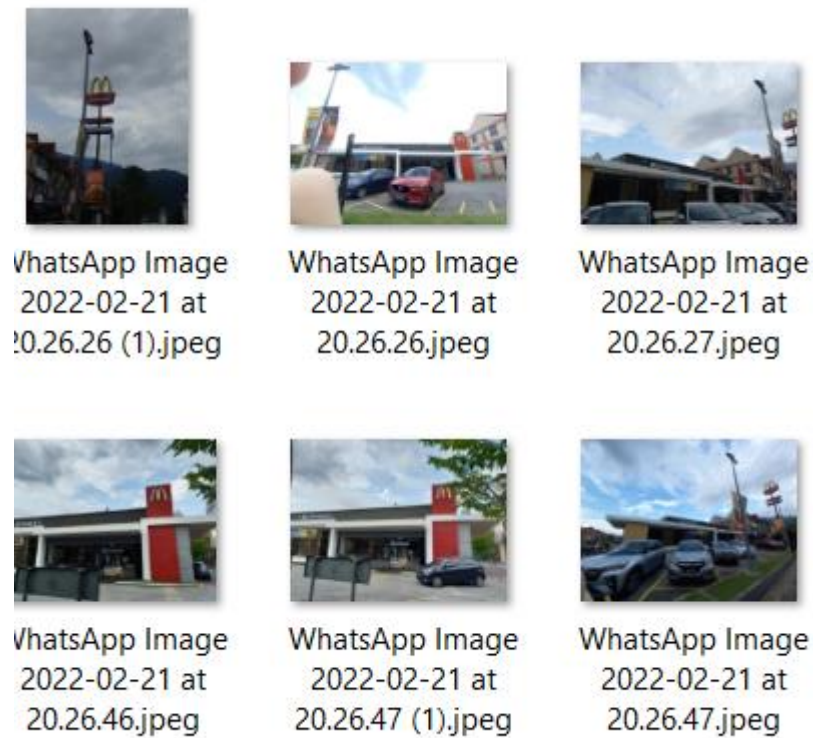


Figure 3.1.6.7 McDonald's testing dataset



Figure 3.1.6.8 Not Shop Trademark testing dataset

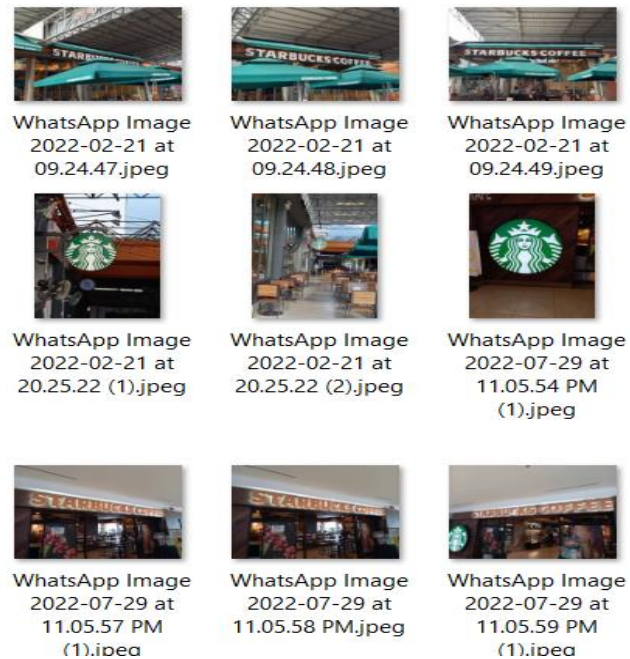


Figure 3.1.6.9: Starbuck testing dataset

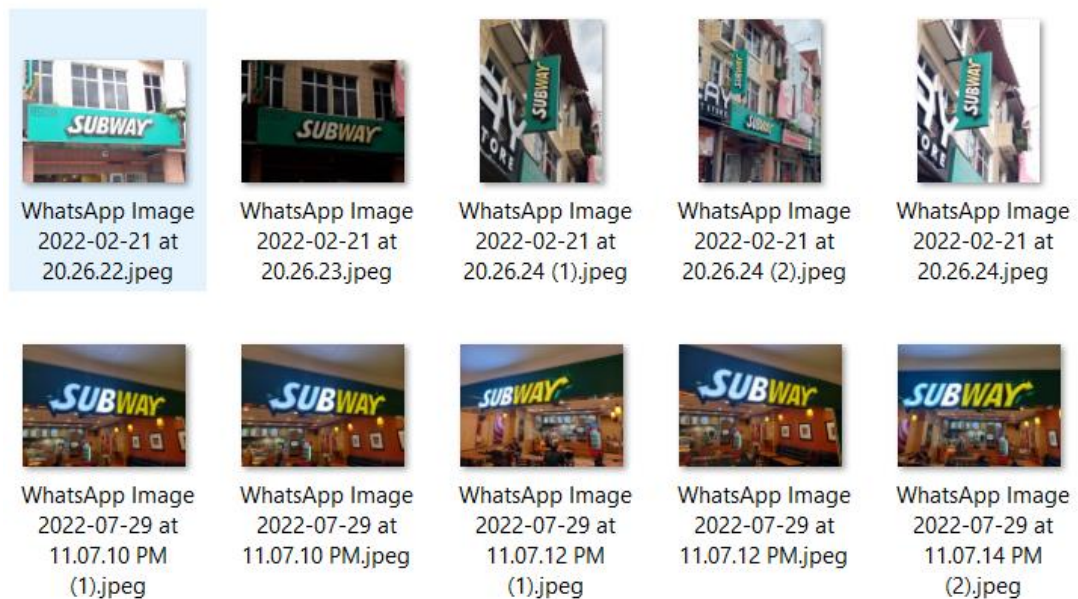


Figure 3.1.6.10 Subway testing dataset

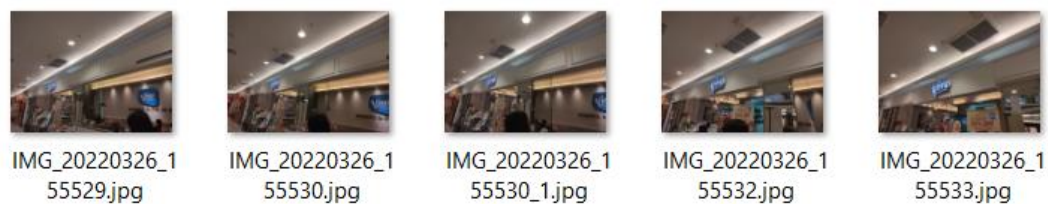


Figure 3.1.6.11: Vivo testing dataset

In the testing dataset, the modal will be tested on the testing datasets provided that some images will have different light intensity and angle of capture. **Figure 3.1.6.6** shows 18 pics of Burger King whereas **figure 3.1.6.7** shows 11 pics of McDonald's and **figure 3.1.6.8** shows 38 pics of Not Shop Trademark. Furthermore, **figure 3.1.6.9** shows 17 images of Starbuck meanwhile **figure 3.1.6.10** shows 15 pics of Subway and lastly, **figure 3.1.6.11** shows 9 pics of Vivo. These datasets are set into batches where the system will randomly select them to be tested in CNN testing. When the testing is done, the real-time testing is conducted where the camera will be turned on to identify these images displayed on the mobile phone where the results will display on the screen of the computer.

3.1.7 Timeline

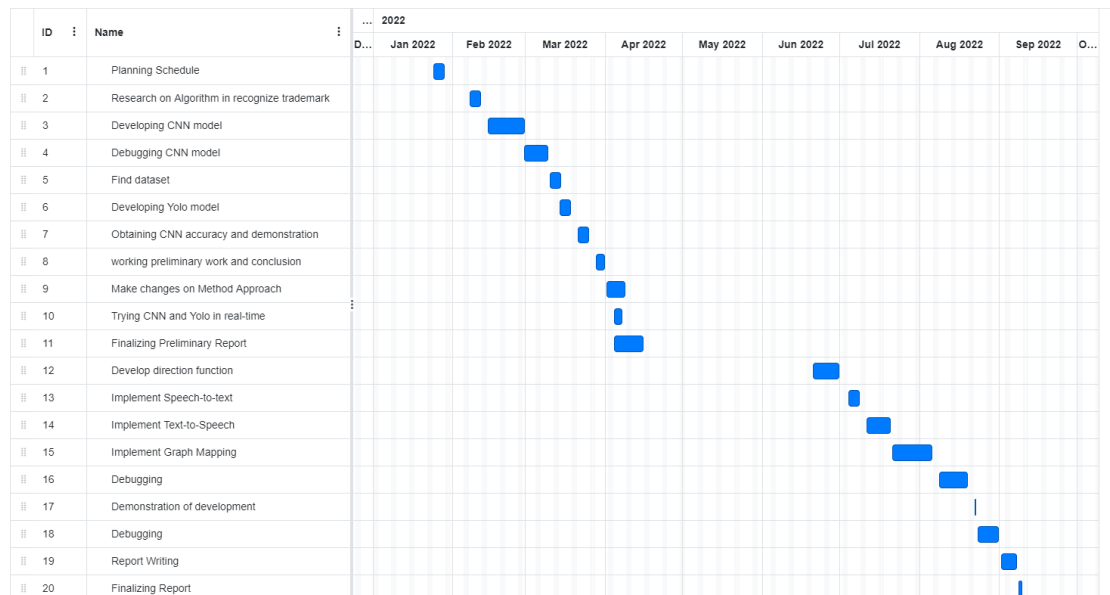


Figure 3.1.7.1 Project 2 Timeline

Figure 3.1.7.1 shows the project timeline to complete preliminary work. The project planning started on week 1, 24 January 2022- 28 January 2022 to schedule a timeslot to do research and meeting supervisor. On week 3, 07 February 2022- 12 February 2022, research on algorithm to recognize shop trademark as the pytesseract could not read some words of the shop trademark and also searching for an algorithm to detect objects. After searching a suitable algorithm which is CNN for shop trademark recognition and YOLO for object detection, on week 4, 14 February 2022- 26 February 2022, the developing on the CNN algorithm started. It took about 2 weeks to work however there was some errors occurring hence on week 5, 28 February 2022-9 March

2022, was debugging the error. It took almost 2 weeks to solve the error. On week 6, 10 March 2022- 13 March 2022, due to not finding a suitable dataset, the week was to snapshot some shop trademarks to test out the CNN algorithm. On week 7, 14 March 2022- 18 March 2022, the YOLO algorithm was developed for about 4 days to complete it. On week 8, 19 March 2022- 24 March 2022, the CNN algorithm was tested to obtain the accuracy and demonstrate to the supervisor for feedback. After that, on week 9, 26 March 2022-31 March 2022, was writing the preliminary work and conclusion for the project report which was submitted to the supervisor for preview. On week 10, 1 April 2022- 3 April 2022, changes on the Chapter 3 Method and Approach were made. On week 11, 3 April 2022-7 April 2022, continuing to develop the system using YOLO and CNN which runs real-time with camera capturing image of the surroundings. On week 12, 8 April 2022- 15 April 2022, finalizing the preliminary report before submitting on 15 April 2022.

On 20 June 2022, the project begins by developing the direction function for 10 days then start implementing the speech-to-text function in the project from 4 July until 8 July. After completing the speech-to-text function, the speech-to-text take place on 11 July and complete on 20 July. The graph implementation started on 21 July and requires to use 12 days as it requires to generate and set the sequence of the nodes. Next, the speech-to-text, text-to-speech and graph implementation are combined with the previous codes in project 1 and many bugs are found hence debugging take place on 8 August and debugged for 10 days. After debugging, a demonstration was done for the supervisor to review and provide feedback for improvement. Debugging continues on 23 August until 31 August and then proceed with report writing on the 1 Sept. The report writing took at least 5 days to write before finalizing on the 8 September and submitting on 9 September.

CHAPTER 4: System Design

4.1 Project Flow Diagram

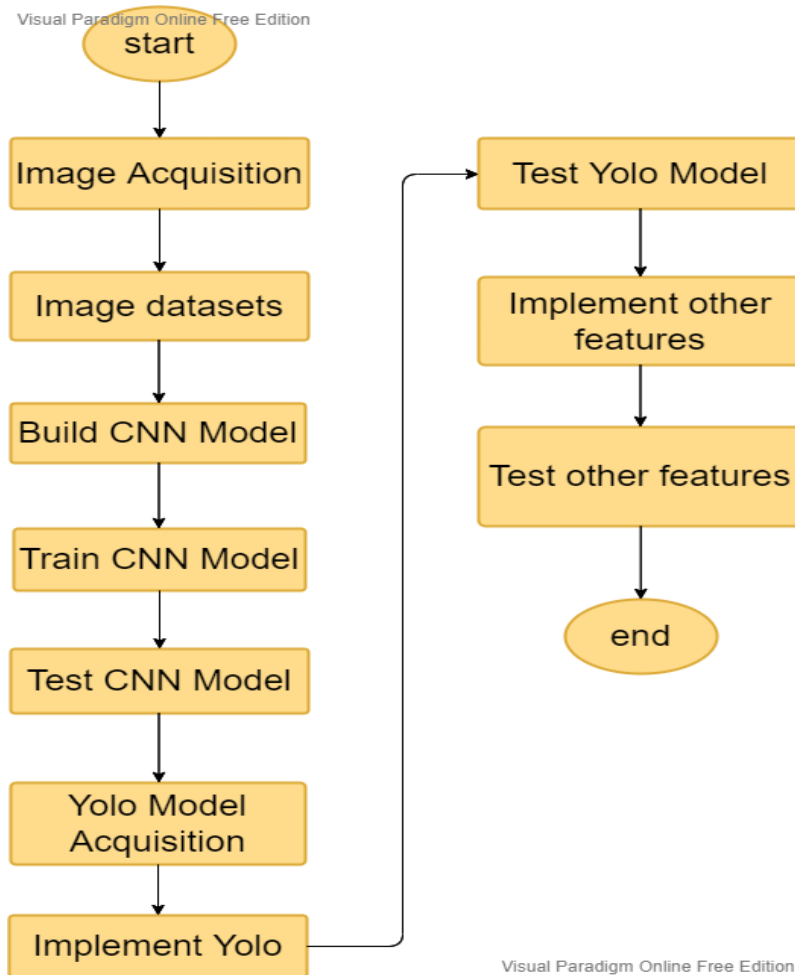


Figure 4.1.1 project block flow diagram

Figure 4.1.1 shows the project flow diagram where each block indicates the function implement into the system. The project starts by doing image acquisition. The images are obtained by snapping the shop trademarks located at the shop lots and shopping mall. After acquiring the images, the images are separate into 3 datasets such as training datasets, validation datasets and testing datasets. Next, develop the CNN model using tensorflow library where each layer is defined and set the number of epochs. Once the CNN model is done developed, the model is ready to be trained by loading the model with the training datasets Then, the CNN is ready to be tested using the testing datasets to determine the accuracy. The model is evaluated and compared with the training accuracy to determine the model performance. The Yolo model is to be acquired through online done by [28]. The Yolo model is pretrained where it can be tested

immediately. The Yolo model is implemented immediately and tested after that. Other than that, other features such as speech-to-text, voice to-text, direction, and next shop features are implemented into the system. Finally, these features are tested to ensure the features work before ending the project work.

4.2 System Block Diagram

4.2.1 Shop Trademark Recognition

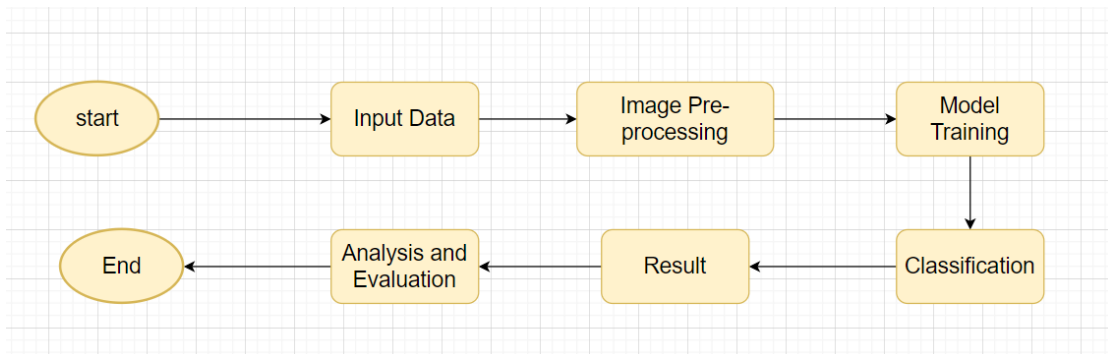


Figure 4.2.1.1 block diagram of Shop Trademark recognition pipeline

Figure 4.2.1.1 shows the flowchart of the shop trademark recognition system which will be used to classify shop trademarks. The method used in this system will be through deep learning, Convolution Neural Network using RMSProp as the optimizer. The model performance will be analyzed based on the accuracy of the training and testing set which will indicate the model would not be underfit or overfit. The datasets contain 3 type, training dataset, validation dataset and testing dataset. All 3 dataset would undergo image pre-processing where the images were resized, rescaled and labelled according to category so that it would be standardized (150x150) in 3 RGB channels for each image. For the CNN training set and validation set, each image will undergo an augmentation process to produce an extra jittered set in addition to the original train and validation set to increase the model robustness as well as reducing overfitting.

Next the training set and validation set will be loaded into the CNN model for prediction training. The train model will be validated using the prediction on the validation set and fine tuning is done to optimize the performance of the model where the parameters and features are assessed and adjusted. Finally, the trained model will be tested by predicting the test set. The test set was rescaled to allow the system to predict without processing too much time. The performance result will be obtained to

analyze and compare using various visualization techniques such as accuracy, precision, recall and f1-scores.

4.2.2 Real Time Indoor Navigation System Methodology

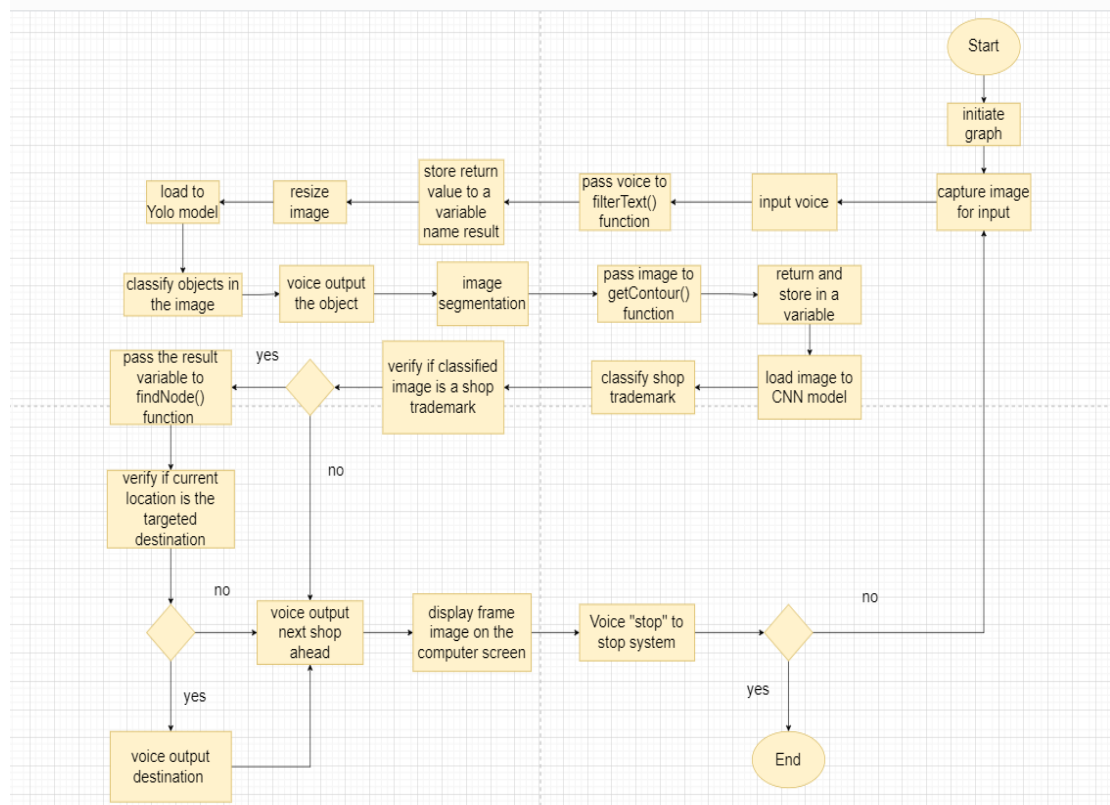


Figure 4.2.2.1 block diagram of real-time indoor navigation system

Figure 4.2.2.1 shows the block diagram of the indoor navigation system in real-time. The system shall start by initializing the graph mapping to create shop orders in the graph. Next, the system captured the environment using a webcam. The system will pause a moment to wait for the visually impaired to input a destination shop through voice, the voice input will pass to a function called filterText() function. The voice input will convert to text and the filterText() function will filter only the shop name and remove the unnecessary text. The shop name will be stored into a result variable which it will pass to another function later. The captured image will undergo resizing so that the models could process faster. The image is resized to (320x320) as the dimension and loaded into Yolo model where it will undergo preprocessing such as converting the colour from BGR to RGB, rescaling to turn each pixel's value to [0,1],

resizing it again to 320x320 dimension and mean subtraction to reduce the colour pixels but the mean is set 0 hence no colour value is reduced.

After preprocessing the image, the image is classified by the system to detect any objects found in the image where the object will be localized and labelled and ready to display on the screen when called. If any objects are detected, the system would voice output to warn the visually impaired to be cautious. The captured image would also undergo image segmentation as the captured image is copied to different variables and the image variable that undergoes image segmentation is the `img` variable. The `img` variable will undergo gaussian blurring, edge detection and dilatation where it will pass to the `getContour()` function. The function will find the contour based on the edges detected and draw the contour. The contour will be localized using a bounding rectangle where the bounding rectangle provides the coordination of the x-axis and y-axis. The return value will be the x-axis and will be stored to variable name `cor` where it will be validated. If the `cor` variable is more than 4000, the direction is right. If the `cor` variable is between 2400 and 4000 then the direction is front and if the `cor` is below 2400 then the direction is left. The direction variable will be displayed on the video frame when called.

Later, another copied image variable name `trademark` will be loaded to the CNN model. The CNN model will classify the image whether the image has a shop trademark. If the image has a shop trademark and returns to the next shop ahead. The result will be passed to the `findNode()` function with the result variable taken earlier to determine if the visually impaired has reached the targeted shop. If yes, then the system will output a voice message “You have arrived at your destination” and inform the next shop ahead. If no, the system will output a voice message of the next shop ahead. The system will display 5 video frames by calling the direction variable, `img` variable, the classification of the CNN and Yolo image and the original image captured from the beginning.

Finally, the system will wait for the visually impaired to command “stop”. If “stop” is commanded, then the system ends else it will loop back and capture new image and repeat the process.

4.3 System Function Flow

4.3.1 CNN Training Flow

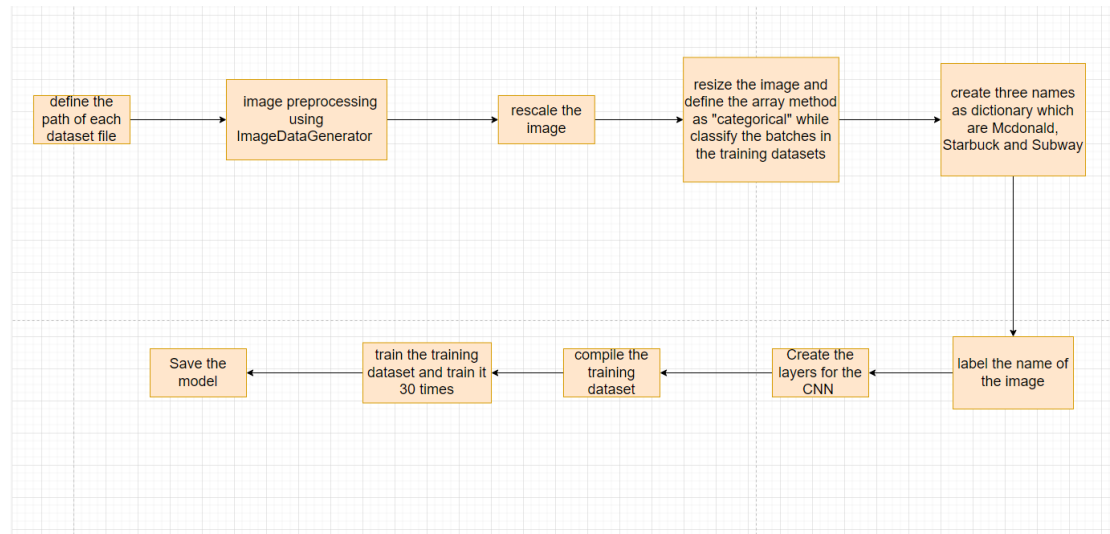


Figure 4.3.1.1 CNN Training Process Block Diagram

Based on **figure 4.3.1.1** shows the CNN Training Process Block Diagram. From the start, it defined the path of each dataset file so that the system knows where to search the datasets later. Next, the images are preprocessed such as rescaling, normalizing, fill mode, transformation and many more which will allow the image data to be cleaned and standardized. The rescaling is to allow the image to change the input value to another range as its input value may be big. Since the image consists of RGB value [0-255], the value can be rescaled to [0,1]. The normalizing process is to divide the input by standard deviation of the datasets and to standardize the input images. The fill mode is to fill up the spaces in the image with the nearest pixel value and stretch it. The transformation is to shift the image position such as rotation, horizontal flip, vertical flip and many other methods. After that, the next step is to resize the image to (150,150) as some image dimensions are big and some are small. The image is to return in 2D hot encoded array format which is selected as “categorical” for its class mode while allowing the system to classify the batches of images in the training dataset. Later, create six names and store it as a dictionary so that it can be used in the testing for labeling the predicted images. The six names set in the dictionary are Starbuck, McDonald’s, Burger King, Not Shop Trademark, Vivo, and Subway. This dictionary will be used for labeling the image and creating the layers of each CNN layer to allow each image to be classified in a few layers. After that, the model will compile the

training datasets. The datasets will be undergoing 30 epochs which means the datasets will loop 30 times to train the images and it will validate with the validate datasets together to determine the accuracy and validate accuracy. Once the loop is done, the model will be saved for future testing.

4.3.2 CNN Testing Implementation

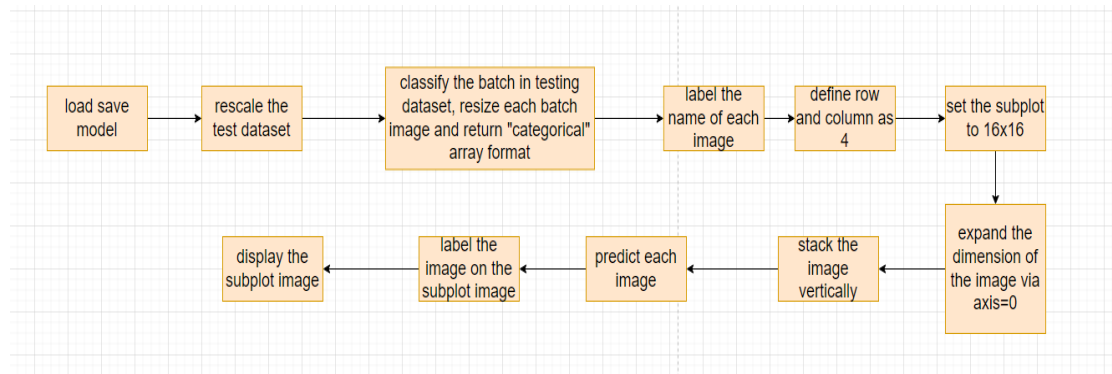


Figure 4.3.2.1 CNN Testing Process Block Diagram

Figure 4.3.2.1 shows the process of the CNN testing process. The process started by loading the save model. Next, the testing dataset was rescaled so that the image will produce 0 or 1 value. Later, the testing datasets will be classified by batch and the image size were resized to (150,150) and set the return value to 2D hot encoded array format. After that, label the name of each batch. The rows and columns were defined as 4 in the for-loop to produce 4 images in a row and 4 in a column. Set the subplot to 16x16 which means 16 coordinate on x-axis and 16 coordinate on the y-axis so that each image will place in 16cm width and 16cm height. Each image will be loop once and the image array dimension is expanded into 1D array as the axis is set to 0. The image array is then stack vertically which will then proceed to the next process that is the predict image. The image array will be predicted by the model and an output will be produced. The next process will be then using the label to label the image on the subplot when it is displayed. The label will follow the output result given by the model. Lastly, display the subplot image with the label place on the image.

4.3.3 CNN Testing Modal with Camera Implementation

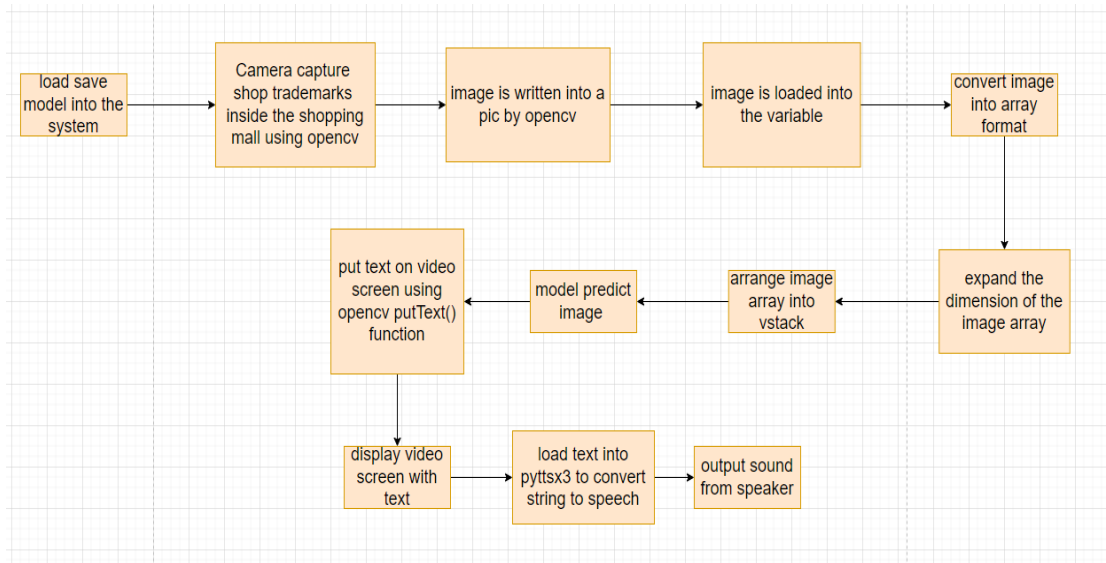


Figure 4.3.3.1 CNN Testing Modal with Camera Block Diagram

Figure 4.3.3.1 shows the process of the camera run for the CNN testing. The process starts by loading the saved model into the system then starts running the external camera where it will capture the shop trademark inside the shopping mall using opencv. Next, the image captured will be written into jpg format and store inside the system. The image will be loaded into a variable then the image variable will be converted into image array format. This image array dimension will be then expanded into 1D array dimension. Later, the image array is arranged to vertical stack or vstack. Each stack of image arrays will be loaded into the model to be predicted. The model will produce an output and will store into a variable. The dictionary that was defined earlier will label the name of the shop trademark depending on the output from the prediction of the model. The label will be put on the video screen layout using opencv. The text will be displayed on the screen to provide assurance that the label name correctly predicted the shop trademark. The text will load into pyttsx3 to convert the string text into speech which will be outputted via the speaker lastly.

4.3.4 Image to Speech Function

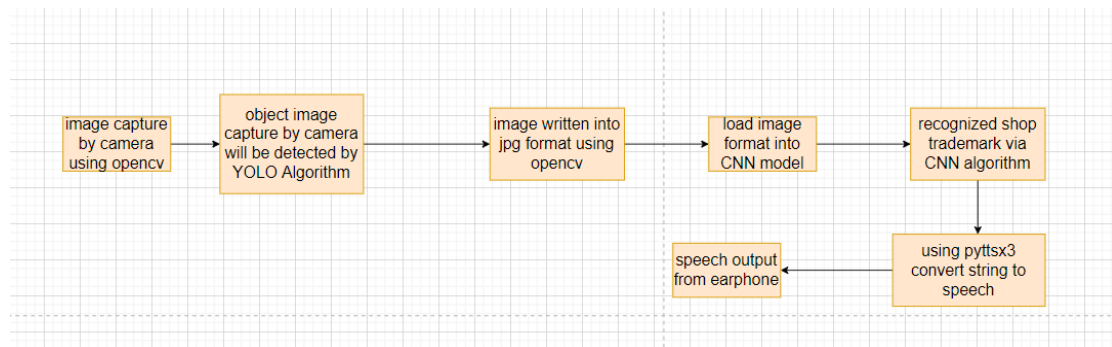


Figure 4.3.4.1 image to speech process

From the **figure 4.3.4.1**, the camera will capture the image of shop trademark and objects using opencv. The YOLO model will be called, and the video will pass into the function, the YOLO model will retrieve the config file and YOLOV3 file where the training datasets are all prepared. A text file name “cocoa.txt” consists of the names of the objects that the YOLO model can detect such as bench, bottle, mobile phone, etc. The video image will undergo preprocessing where the size, scale, colors, and mean value will be processed. Once the image is done preprocessing, the model will compute the layers of the image. Each layer will be classified so that the model can localized the object and predict the name of the object. The video will be written into jpg file to be store as image, this image will be loaded into the CNN model, the model will predict the shop trademark. Before the CNN model can perform prediction, training the model is required. A dataset of shop trademark is fitted into the model. The model will classify the image and compare it to the image in the validation dataset. The validation dataset may consist of the same images in the training dataset. The model will produce two accuracy results, one known as accuracy, and another validate accuracy. Accuracy is measured by the number of layers the image is classified in the model while validate accuracy is classified the image with validate dataset where the model will predict if both are the correct trademark. After that, the system will label the image to determine the shop trademark and load it into a text-to-speech function using pyttsx3 library. The library will convert text into speech hence any string value returned by the CNN function will immediately load into text-to-speech function. Lastly, the return value of text-to-speech will be outputted from the earphone.

4.3.5 Voice Activation Function

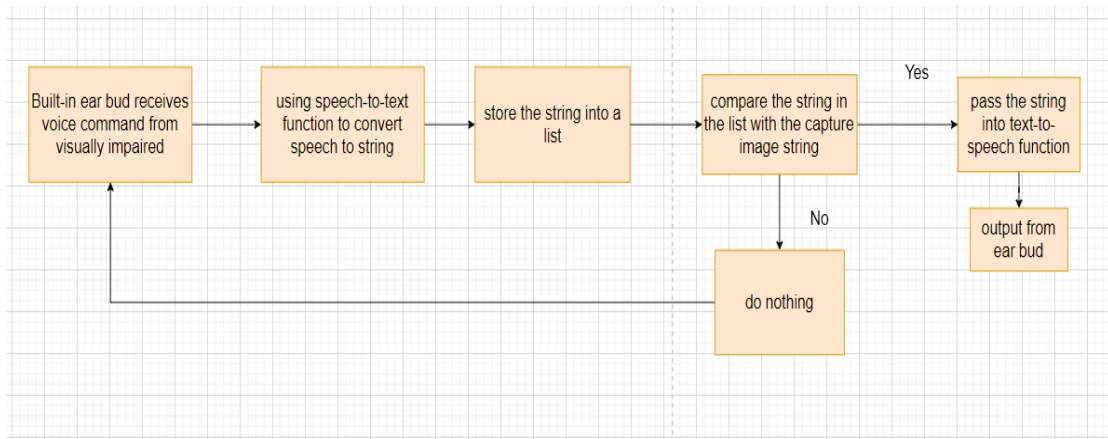


Figure 4.3.5.1 voice activation implementation

Based on **block diagram 4.3.5.1** shows the speech recognition process. The visually impaired will command using voice instruction which will be received by a built-in microphone earbud/earphone. Using a speech-to-text function to convert the speech back to text, the text will then store into a list. The text store in the list will be overwritten when a new command is inserted. The system will compare the text stored in the list with the labeled image captured from the camera. If the text and the image label are the same, then the text-to-speech function will be called and convert the text into speech which will output via earbud. If the string in the list is not the same as the shop trademark, then the system will do nothing and wait for the next command from the user.

4.3.6 Yolo Function

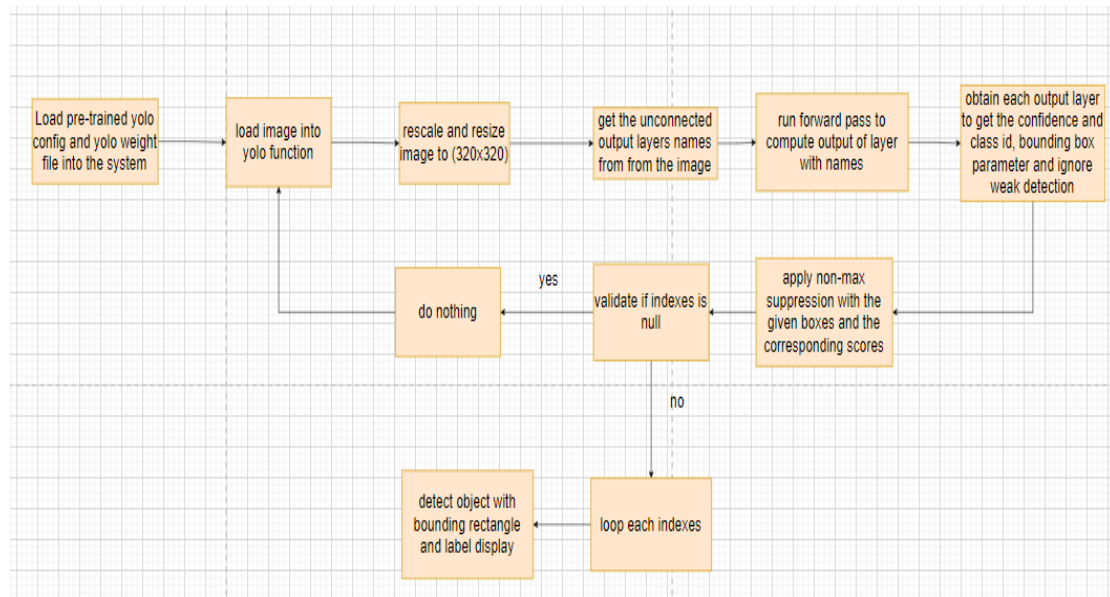


Figure 4.3.6.1 Yolo implementation

Figure 4.3.6.1 shows the Yolo implementation block diagram flow. The Yolo used in this project is a pre-trained file which means the system is not required to be trained as it is already trained and it is version 3. This Yolo could be tested immediately and could detect 85 objects. The Yolo function started by loading the yolo config file and yolo weight file into the system. Next, load the captured image into the Yolo function. The image will undergo rescaling, resizing and Red and Blue swapping process. The rescaling is to perform mean subtraction to scale the image by some factor and multiplying the input channel. The rescaling is set to 1/255 to change the pixels range to [0-1] and to normalise the input. The resizing is to reduce the dimension of the image so that the image could be processed faster. The Red and Blue swapping is to allow the colour BGR to change to RGB as opencv by default assumes the image is in BGR order. After augmented the image, the image is passed into a network where this network will first extract the unconnected output layers from the image which will then forward through the network to compute the output layer with names. After that, the function will obtain each output layer to get the confidence, class ids, and bounding box parameter. Any weak detection obtained from the output layer will be ignored if the confidence value is less than 0.5. The bounding box parameter helps to determine the coordination, width and height of the object. This will help in creating the bounding rectangle to focus on the detected objects. The class ids are the object labels, and the confidence is the accuracy of the object detection. Later, non-max suppression is

applied to prevent overlapping of the boxes focusing on one object. The non-max suppression boxes will remove any boxes with an IOU (intersection Over Union) more than 0.8. Then, the system will validate that the indexes variable is null. if the indexes variable is null, the system will do nothing and load the next image. If the indexes variable is not null, the system will loop each index in the variable and perform object detection with a bounding rectangle and label displayed. The indexes variable was defined to store the non-max suppression boxes result.

4.3.7 Graph Implementation

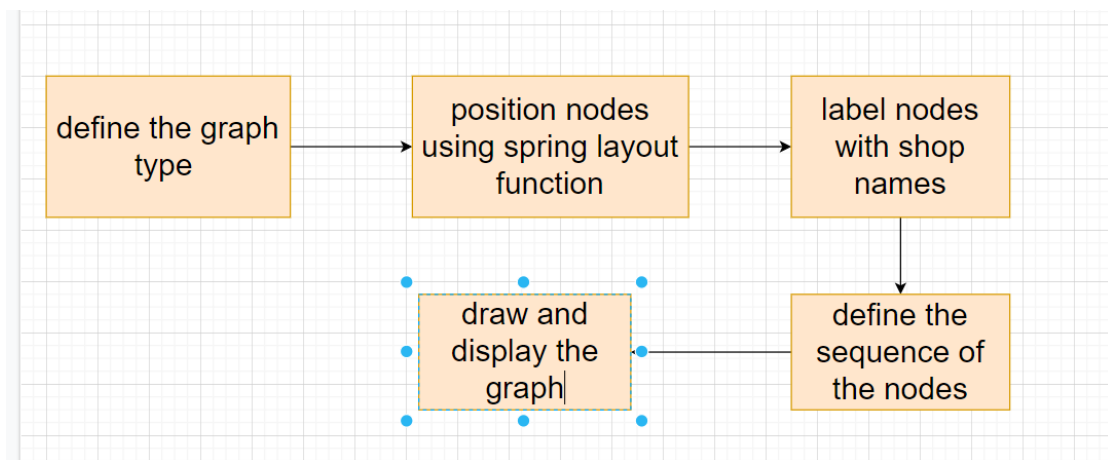


Figure 4.3.7.1 graph creation implementation

Figure 4.3.7.1 shows the graph creation implementation procedure which will be depicted as a navigational system. Although the graph will be a straight graph without any direction needed, the system will use the graph to inform the user of the next shop ahead of them. **Figure 4.3.7.1** describes the graph creation using networkx library. First, the graph type was defined as an undirected graph. Next, position the nodes using the spring layout function which uses the Fruchterman-Reingold force-directed algorithm. After that, create a shop name list and label each node. Each node follows the sequence of the list of shop trademarks. Then, once the nodes are labeled according to the sequence of the shop trademark, the nodes are arranged according to the defined sequence. Finally, display the graph which is shown in **figure 4.3.7.2**.

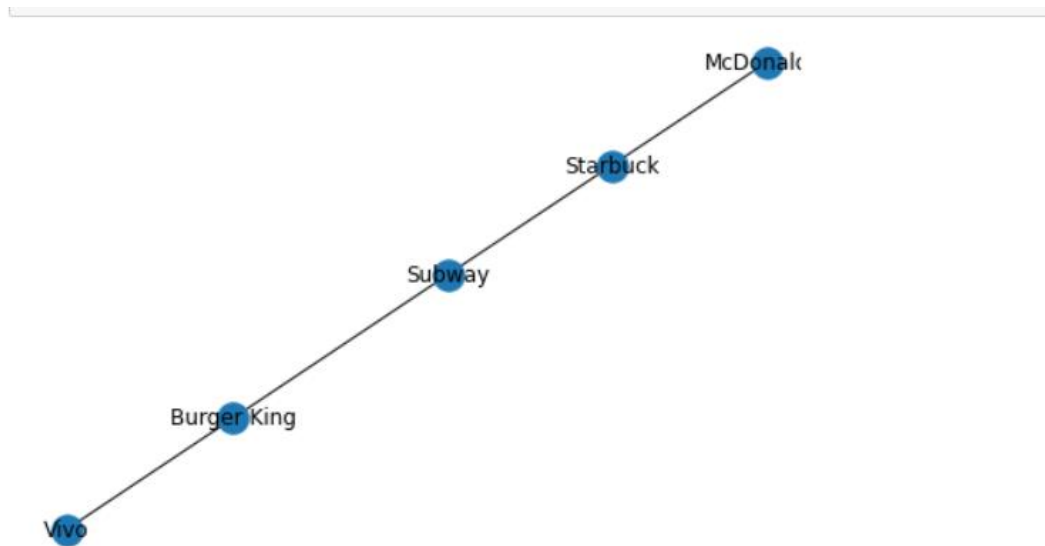


figure 4.3.7.2 graph display

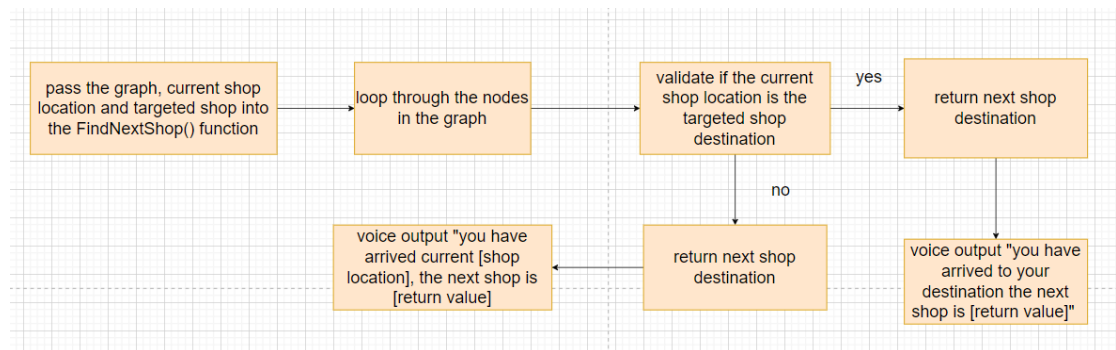


figure 4.3.7.3 findNextShop() function

Figure 4.3.7.3 shows the findNextShop() function flow. When the image was classified by the CNN modal, the label will be passed into the findNextShop() parameters together with the destination shop input by the visually impaired via voice. The function will loop the nodes in the graph and the function will return the next shop name regardless if the visually impaired has or has not arrived at the destination point. If the visually impaired arrive at the destination point, the function will return the next shop name and output the voice message “you have arrived at your destination, the next shop is [return value]”. If the visually impaired did not arrive at the destination point then the function returns the next shop name and outputs a voice message “you are now at [current shop location], the next shop is [return value]”.

4.4 System Component Interaction Operation

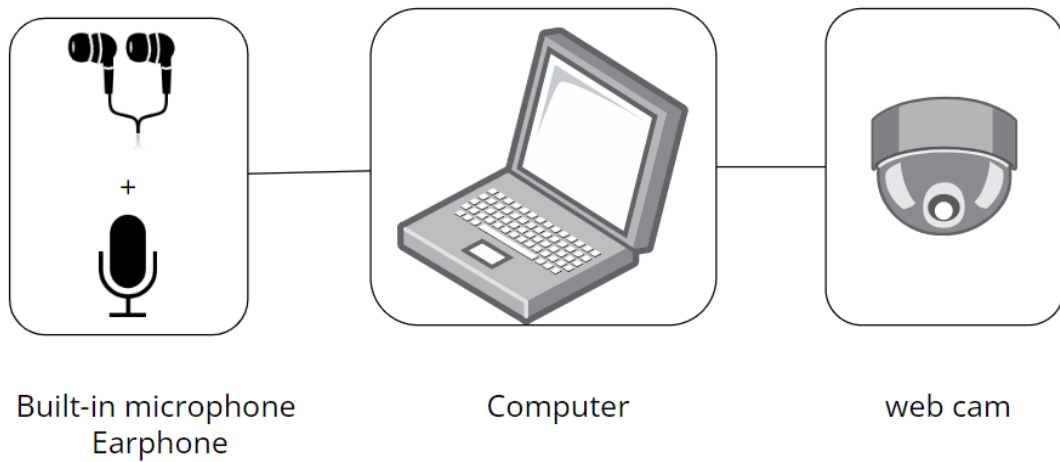


Figure 4.4.1 System Components

Figure 4.4.1 shows the system components connection where the built-in microphone earphone and web camera are connected to the computer. The built-in microphone earphone will receive the voice input from the visually impaired and will transmit the input to the computer. The computer running the software will convert the speech-to-text for processing. The webcam will capture the surroundings and transfer to the computer for processing. The software will process the image and use the models to classify the image.

The following are the steps of the operation process:

1. Built-in microphone earphone and webcam transmit input data to the computer software
2. The software processes the input data
3. The software will return the output and convert the output to sound
4. The output will sound through the earphones

CHAPTER 5: System Implementation

5.1 Software Setup

The software used in this project is a Jupyter notebook which the language is Python. Jupyter notebook is an open source software which allows editing and running the notebook documentation on a web browser.

The softwares that are used to build the system are described as follow:

Anaconda3 Prompt

Anaconda 3 is a console where the user will be required to type the command line to manage packages, environment and channels unlike Anaconda navigator users can perform those tasks without using any command line. The software will help install python libraries which will be used in the project.

Jupyter Notebook

Jupyter notebook is an open-source software that allows users to perform editing and running the notebook documentation on the web browser [25]. Jupyter notebook uses python language, and it runs the code by cells instead of one whole file. These cells help to detect errors in each line quickly and also to reduce interpreting time.

In this project, many libraries are required to help run the system. The libraries that are required are stated as follow:

Keras

Keras is an API that best follows the best practices for reducing cognitive load. Keras provides consistency and simplicity for an API. It helps reduce the number of user actions required and also provides clear and actionable error messages. It is widely used in deep learning frameworks [24].

Tensorflow

Tensorflow is a python library created and released by Google for fast numerical computing [26]. This library is used for building deep learning models as it simplifies the process of building the model. Tensorflow is used in this project to build the CNN model which allows the model to learn and predict images with great accuracy. This will help in building the shop trademark recognition.

Yolo version 3

Yolo means You Only Look Once, it is a real time object detector. YOLOv3 or Yolo version 3 is fast and accurate and has some advantages such as it makes predictions on a single neural network evaluation [29]. YOLOv3 classifies and localise to perform detection. The model is applied to an image where multiple locations and scales are read by the model and regions of the image with high scoring are considered as detection. The YOLOv3 used in this project is a pre-trained version done by [28] where it can detect 85 objects and does not need to be trained again.

Networkx

Networkx is a package in python to help create, manipulate, and understand the structure, dynamic and the function complexity of the network. This package is used in the project for creating the graph map for indoor mapping.

Pytsx3 2.90

Pytsx3 is a text-speech conversion library used in Python programming. It is compatible in both Python version 2 and 3 and it can operate offline. The library also does not require user to use an API (Application Programming Interface) to conduct text to speech conversion. Pytsx3 also run offline which is suitable in the project so that it can process text-to-speech when there is no internet.

Pytesseract

Python Tesseract is an Optical Character Recognizer (OCR) that read text from images. it can read all image types supported by the Pillow and Leptonica imaging libraries, including jpeg, png, gif, bmp, tiff, and others. It can print text on script instead of writing into a file.

Speech Recognition 3.8.1

Speech Recognition is a python library that allows user to use voice command to instruct the system to perform particular action. The library require user to train their speech to enable the system to understand user pronunciation. Once trained, the library will convert the speech to text and this text shall be input and the microcontroller will process the input and output the result.

The table below shows the command line to install the python libraries or packages using anaconda prompt.

Table 5.1.1 packages and command line

| Package/ libraries | Command line |
|--------------------|------------------------|
| Tensorflow | pip install tensorflow |
| Keras | pip install keras |
| Networkx | pip install networkx |

5.2 Hardware Setup

The hardware that will be involved in this project is a computer with a Windows 10 64-bit operating system, processor, and 8GB RAM and a Graphic card. A Web camera is required as the system will involve computer vision. Finally, an earphone or earbud with built-in microphone for the user to listen to voice instructions and perform voice activation.

Table 5.2.1 hardware specification

| Hardware | Specification |
|----------------|---|
| Laptop | <ul style="list-style-type: none"> • Windows 10 64-bits • 8GB RAM • Intel® Core™ i5-8600K Hexa-core@4.3GHz |
| Web Camera | <ul style="list-style-type: none"> • 30 fps • HD Resolution • Including Auto Focus • 2 MegaPixels • It does not require driver • It has USB 3.0 interface |
| Realme Buds Q2 | <ul style="list-style-type: none"> • Model: RM-630 • Color: Black / Blue • Frequency Response: 100Hz-10KHz • Output Sensitivity: 90 (+/- 3dB) |

| | |
|--|---|
| | <ul style="list-style-type: none"> • Microphone Sensitivity: -42dB • Impedance: 16Ω (+/- 15%) • Rated Power: 1mW • Drive Unit: 6.0mm Speaker • Pin: 3.5mm • Wire Length: 1.2 Meter • Product Weight: About 13g |
|--|---|

5.3 System Operation

The system operation explains the system display results when the system is executed. The displays are shown in the following:

5.3.1 Camera Display

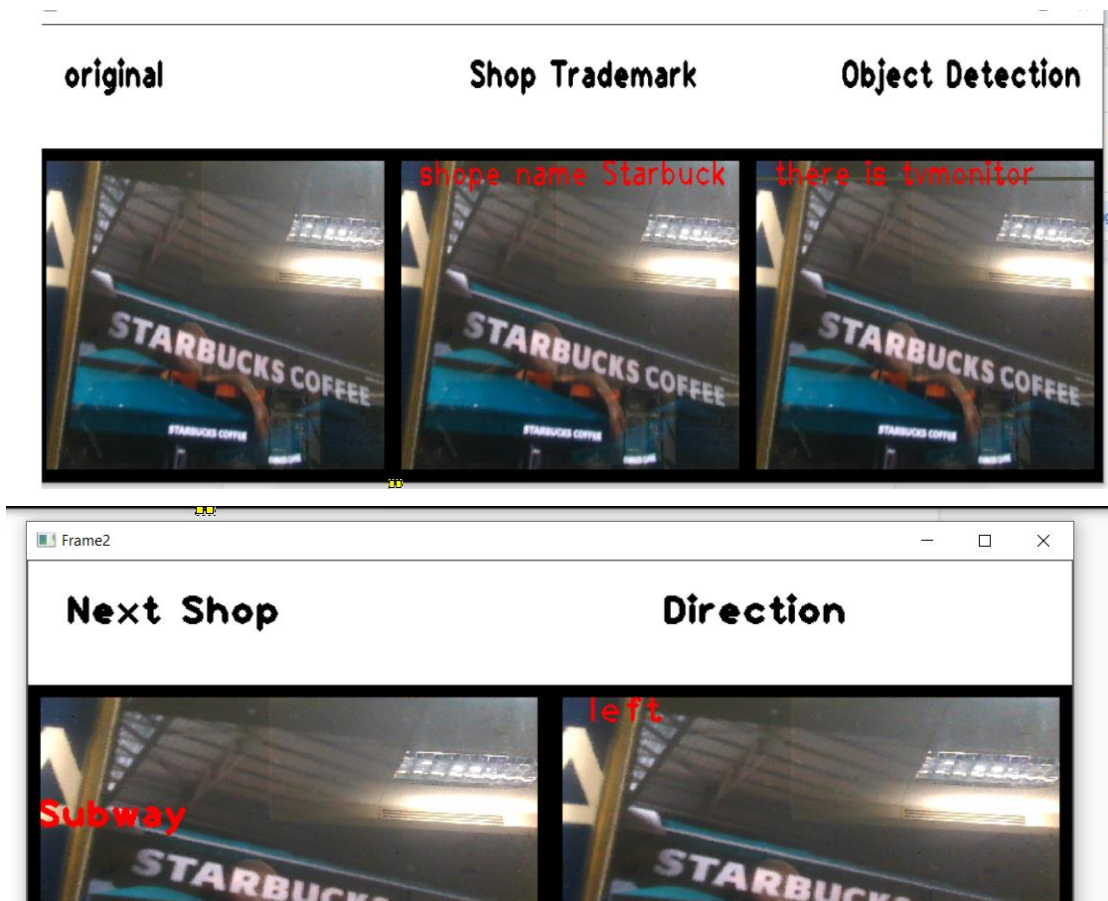


Figure 5.3.1.1 camera display results

Figure 5.3.1.1 shows the camera display results when the system is being executed. Two window frames will pop on the computer screen showing each frame and result. Each frame is labeled with a title to indicate the purpose of the display. From the top left shows the original pic until the bottom right shows the direction of the shop trademark. The title “original” depicts the original image captured by the camera. Next, the shop trademark depicts the result label in the image. In **figure 5.3.1.1**, the label in the shop trademark is “shop name Starbuck”. On the third column of the first row labeled object detection, this frame displays the detection of the object which shows in the image that the label is “there is tv monitor”. This frame only display the objects detected by the system and the system will voice out to the visually impaired for awareness. The bottom left labeled “Next Shop” displays the next shop after Starbuck which displays Subway as the next shop ahead. The next shop result is obtained from a graph map shown in **figure 5.3.1.2**.

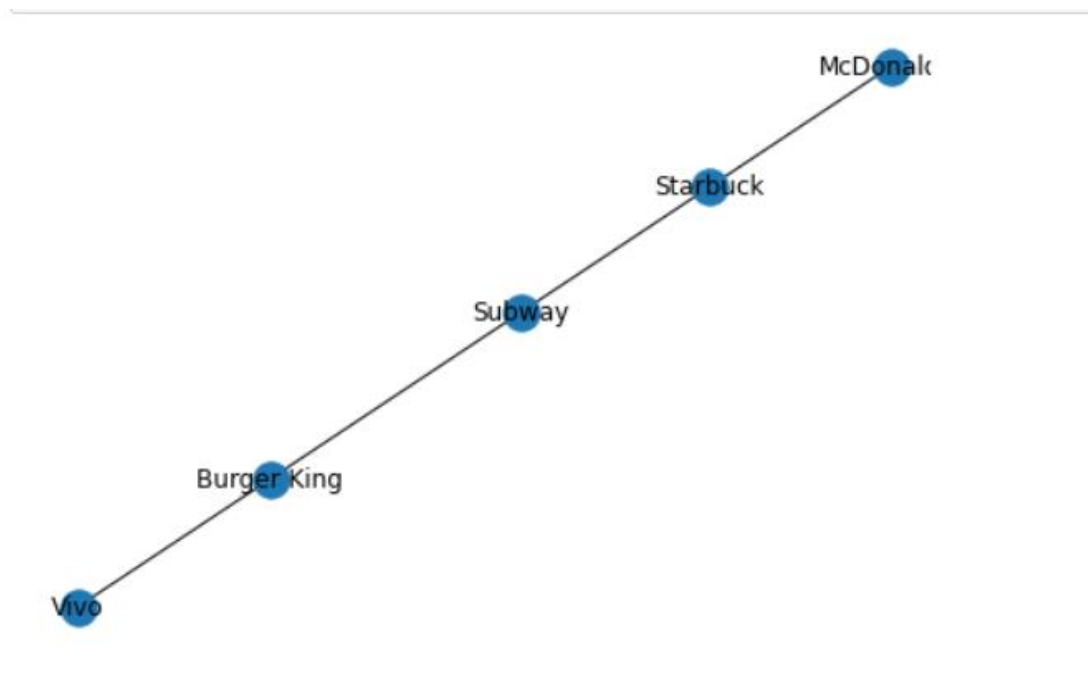


Figure 5.3.1.2 graph map

Figure 5.3.1.2 shows the map which initialized from the top right starting until the bottom left. The top right end is McDonald’s while the bottom left end is Vivo. The system indicates the next shop is traversing from the right to left by 1 edge. Hence, the next shop after Starbuck us Subway and the shop before Starbuck is McDonald’s. The result will return and label on the Next Shop frame shown in **figure 5.3.1.1**. The last

frame “direction” displays the direction of the shop trademark located. The label in the direction frame is left which means in the person view the shop is on the left.

5.3.2 Speech Input and Voice Output

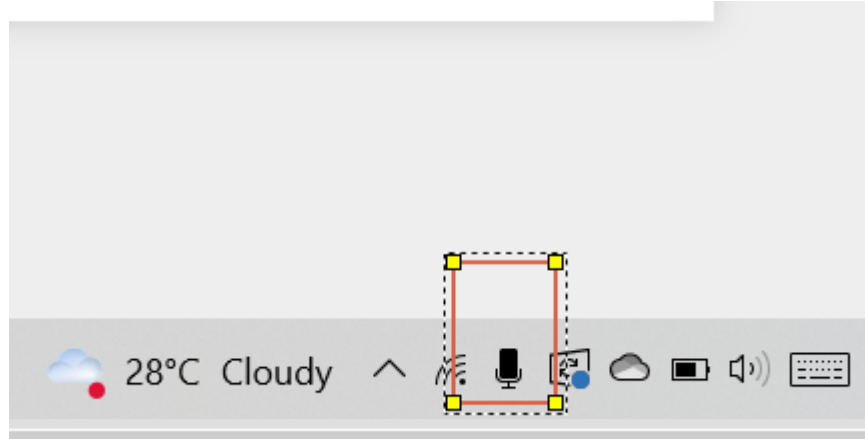


Figure 5.3.2.1 microphone icon pop out on computer menu

When the speech-to-text function is executed, a microphone icon will pop out on the bottom right window menu of the computer. This icon indicates the system is waiting for the input voice from the visually impaired. If the icon disappears from the window menu, it indicates time out or voice input received. The visually impaired are required to call “computer” so that the visually impaired acknowledge that the system is listening. If the visually does not call “computer”, the system will do nothing and proceed to inform the shop trademark name, object detected, next shop ahead and shop direction if arrive to the destination point.

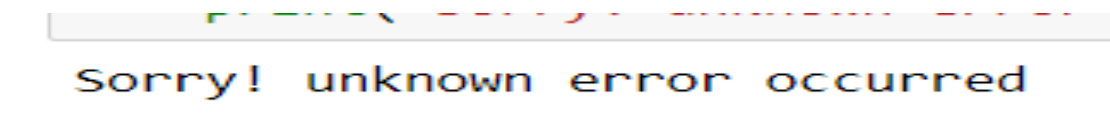


Figure 5.3.2.2 unknown value input

When the microphone timeout does not receive any input voice, the system will voice out the message “Sorry! unknown error occurred” to inform the visually impaired to speak again. This could occur if the visually impaired speak softly or too quickly which causes the system not to manage to detect the voice input.

Yes master! I'm here! which shop you want to go?

Figure 5.3.2.3 voice input “computer”

Figure 5.3.2.3 shows the input voice is “computer” and the system return a voice message “Yes master! I’m here! which shop you want to go?”

5.4 Implementation Challenges and Issues

In this project, the implementation challenges, and issues were the lagging issue due to the text-to-speech process and speech-to-text. When the text-to-speech or speech-to-text is executed, the whole system will pause a moment to allow the text-to-speech or speech-to-text to finish executing, resulting in the camera to capture lesser fps. Another challenge is the execution of the Yolo model. The Yolo model runs using CPU as a result the CPU consumption is high and slows down other processes running behind the system such as image augmentation, speech-to-text and CNN model. The Yolo model requires a high processing power hence will require all the processing unit from the CPU hence causing the system to run slower when executed and affect the classification of the CNN model because the image captured can be blurry or distorted resulting in many incorrect predictions. Next, The implementation of an indoor navigation system. The indoor navigation requires design using paid software tools to create the mapping of a shopping mall. Since the tools require monthly payment hence the indoor navigation system in this project uses graph mapping which is the simplest form to create however, it cannot track the user location and relies on the system to classify the image and inform the user of its current location. The CNN classification may not provide a correct prediction therefore may give the wrong location to the visually impaired.

Lastly, Training the CNN to classify images that are not shop trademarks is one of the challenges to face because the model will classify random views as a shop trademark. This will give false information to the visually impaired and also providing false location to the visually impaired due to the false result pass to the graph mapping.

CHAPTER 6: System Evaluation and Discussion

6.1 System Testing and Performance Metrics

6.1.1 System Training result

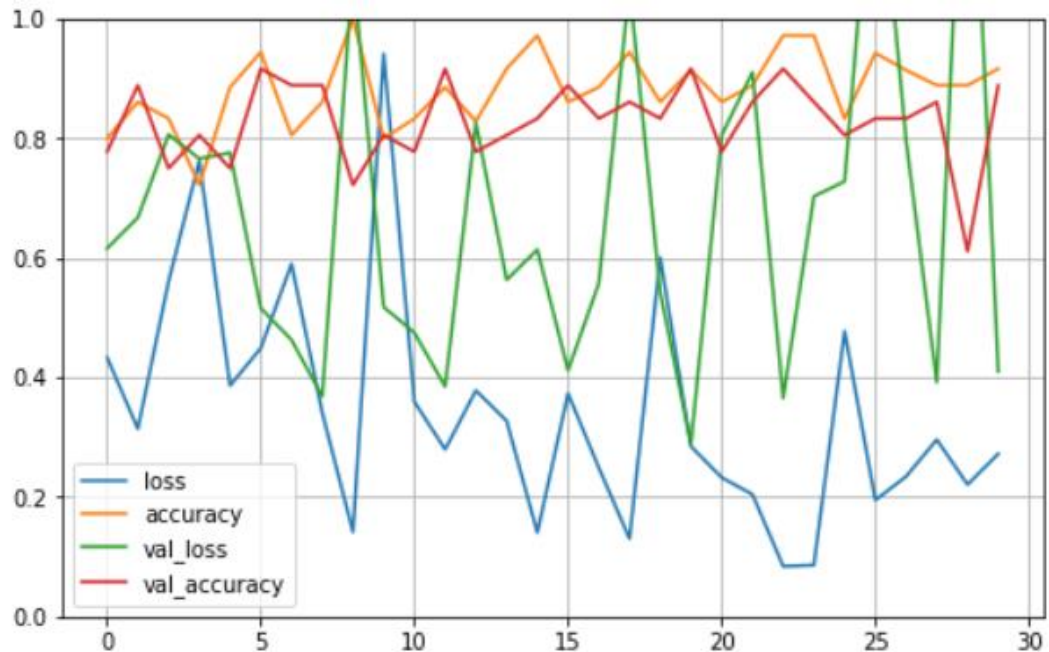


Figure 6.1.1 step per epoch vs training accuracy

Based on figure 6.1.1 shows the step per epoch vs the accuracy graph, the accuracy of the training data shows that it was between 1.0 and 0.8. For the accuracy validation, it was shown that it was between 1.0 and 0.8. Both accuracies show it is above 0.8 but lower than 1.0. This shows that the training modal was neither underfit or overfit because both accuracies were almost similar and closed. The system obtained the validation accuracy when it performed training prediction with the training dataset and validation dataset. The system performed 30 epochs to predict the image during the training process. The number of datasets for the training dataset was 113 images, validation dataset consists of 93 images, and testing dataset consist of 108 images. The val_loss and loss indicate the learning process of the model. When the losses were high, it was indicating the module was learning and predicting many errors. When the loss reduced over time, the model started to predict more accurately. The val_loss also shows the model makes many errors and slowly reduces over time when 30 epochs

were done. The result of the loss was between 0.4 and 0.2 while val_loss was fairly higher than 0.4.

6.1.2 Test Set Predicted Result

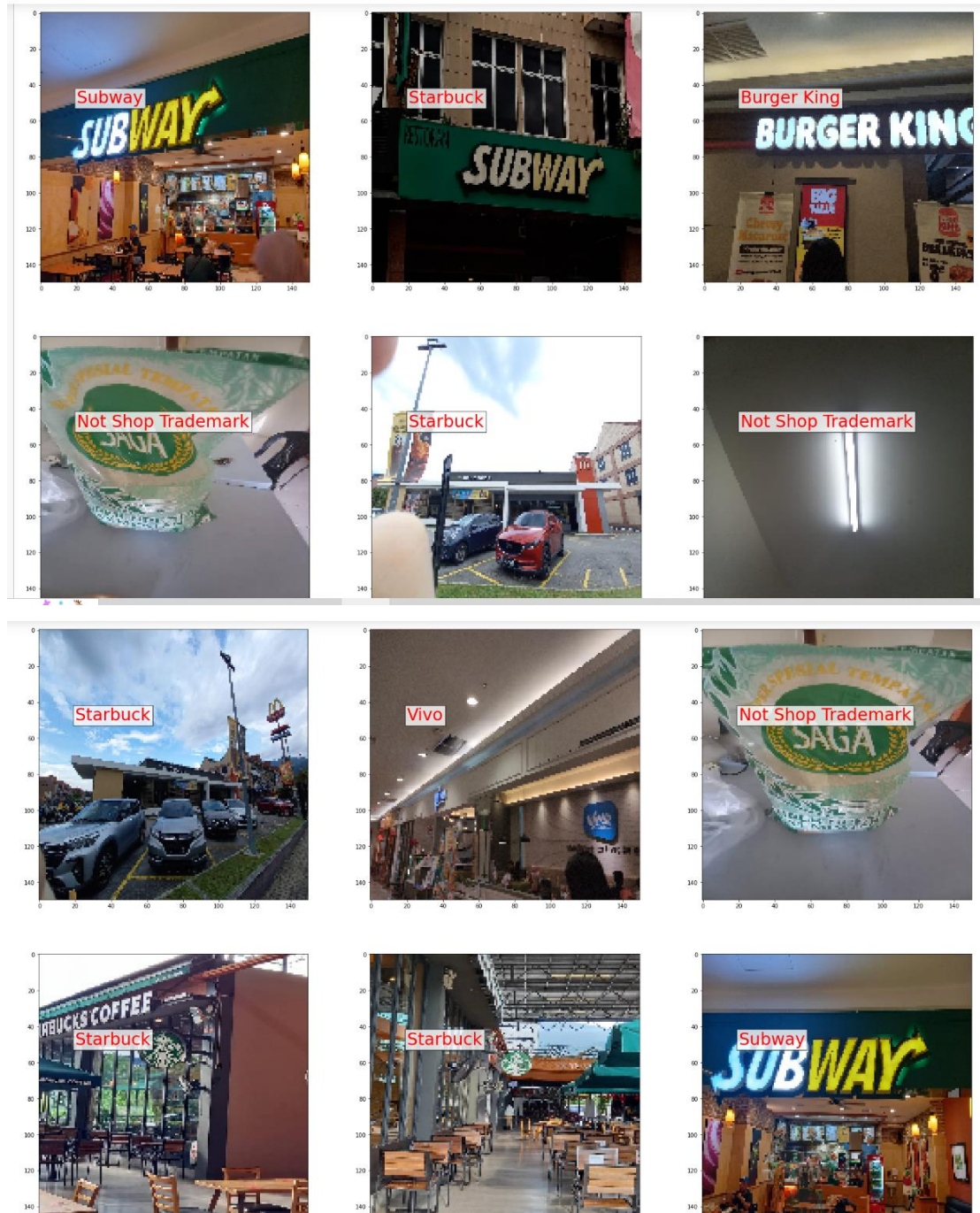


Figure 6.1.2.1 testing results

Figure 6.1.2.1 shows the output from the testing set where 12 images were displayed with the predicted labels. There were about 108 datasets in the testing sets, and about 14 images were detected wrongly from the overall. From **figure 6.1.2.1**, 3 images were

Bachelor of Computer Science (Honours)
Faculty of Information and Communication Technology (Kampar Campus), UTAR

shown to be wrongly predicted such as Subway was predicted as Starbuck, and McDonald's predicted as Starbuck. The system manages to classify images that are not shop trademarks at all as there are 3 images in **figure 6.1.2.1** are random object images. These random object images are used in the testing to allow the system to classify whether the image is a shop trademark or is not. This will help in preventing the real time system to falsely predict random objects as shop trademarks.

6.1.3 Confusion Matrix

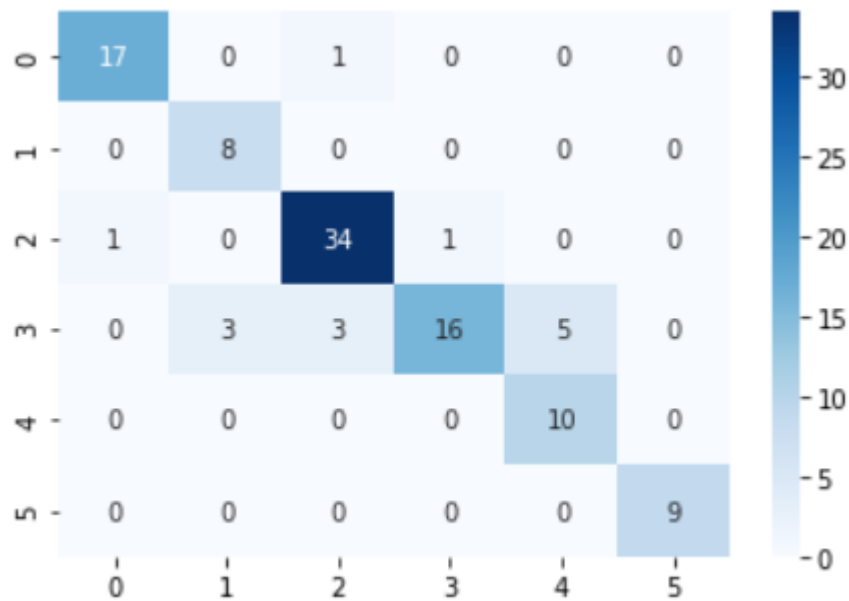


Figure 6.1.3.1 confusion matrix

Table 6.1.3.1 Class indices

| Class | Indices |
|--------------------|---------|
| Burger King | 0 |
| McDonald's | 1 |
| Not Shop Trademark | 2 |
| Starbuck | 3 |
| Subway | 4 |
| Vivo | 5 |

Figure 6.1.3.1 shows the confusion matrix of the predicted image. The matrix consists of 6 rows and columns. **Table 6.1.3.1** shows the class and indices which represent the label of the confusion matrix. Each row and column are labeled based on the arrangement of the shop trademarks. Therefore, the first row and column represents Burger King, while McDonald's represents second row and column, Not Shop Trademark third row and column and the next 4th, 5th, and 6th rows and columns are Starbuck, Subway and Vivo. from the confusion matrix shown in **figure 6.1.3.1** there were a total of 94 images were true positive which means 94 images were predicted correctly. 14 images were false negative which means 14 images were positive images but were predicted negative while another 14 images were false positive and these show that 14 negative images were predicted as positive images.

6.1.4 Accuracy, Recall, Precision, and F1 score Result

Accuracy Result

```
In [24]: model.evaluate(test_generator, steps=6)
6/6 [=====] - 2s 178ms/step - loss: 0.8360 - accuracy: 0.8333
Out[24]: [0.8360011577606201, 0.8333333134651184]
```

Figure 6.1.4.1 testing evaluation

Figure 6.1.4.1 shows the test set to be evaluated to determine the accuracy. The result shows that 0.8360 was the loss result while 0.8333 was the accuracy. The model was evaluated using the `model.evaluate()` function which was provided by the tensorflow library. The `model.evaluate()` function evaluated the test set by running 6 steps per epoch which means there are 6 steps to predict the batches of images before declaring the evaluation round finished. Since the `model.evaluate()` function relied on the number of steps to produce the final accuracy of the testing set, it is tedious to run several times to obtain the final accuracy hence, using sklearn library the final accuracy could be obtained immediately.

```

In [22]: report=classification_report(y_test_classes,y_pred_class,target_names=y_class_label)
print('Accuracy: {:.2f}\n'.format(accuracy_score(y_test_classes,y_pred_class)))
print('Precision: {:.2f}\n'.format(precision_score(y_test_classes,y_pred_class,average='macro')))
print('Recall: {:.2f}\n'.format(recall_score(y_test_classes,y_pred_class,average='macro')))
print('F1 Score: {:.2f}\n'.format(f1_score(y_test_classes,y_pred_class,average='macro')))

Accuracy: 0.87

Precision: 0.91

Recall: 0.86

F1 Score: 0.87

```

Figure 6.1.4.2 accuracy result using sklearn library

Figure 6.1.4.2 shows the accuracy result. using sklearn library to predict the test set, it showed the accuracy is 0.87 or 87%. Since the training accuracy was about 0.9167 or 92% while the testing accuracy was 87% therefore the modal was neither underfit or overfit as both training and testing accuracy value were near. The precision value of the modal was 0.91 or 91% . This indicated the modal had high accuracy in predicting correctly on the overall positive images as 91% of the positive images were predicted correctly while 9% of the positive images were wrong. Precision measures the overall positive images that how many are predicted correctly.

However, the recall value of the modal showed it was 0.86 or 86%. Recall measures the number of actual positive images were positive. For example, if the image is Burger king (True Positive) that went through the test and was predicted as another shop trademark (False negative) then the consequences will be bad because the visually impaired could not reach its destination. Thus, the recall rate describes the modal ability to predict the actual positive images. The modal can predict 86% actual positive images but 14% of the images were predicted wrongly hence the system could falsely inform the visually impaires that they are not at the right destination even though they are in fact arriving at their targeted shop. The recall value indicates the modal prediction to be trustable as the modal has less false negative number. The F1 score combines the precision and recall of a classifier into a single metric by a single harmonic mean. The F1-score for the overall system was 0.87 which was similar to the accuracy value.

6.1.5 Testing Report

In [23]: `print(report)`

| | precision | recall | f1-score | support |
|--------------------|-----------|--------|----------|---------|
| Burger King | 0.94 | 0.94 | 0.94 | 18 |
| McDonald | 1.00 | 0.73 | 0.84 | 11 |
| Not Shop Trademark | 0.94 | 0.89 | 0.92 | 38 |
| Starbuck | 0.59 | 0.94 | 0.73 | 17 |
| Subway | 1.00 | 0.67 | 0.80 | 15 |
| Vivo | 1.00 | 1.00 | 1.00 | 9 |
| accuracy | | | 0.87 | 108 |
| macro avg | 0.91 | 0.86 | 0.87 | 108 |
| weighted avg | 0.91 | 0.87 | 0.88 | 108 |

Figure 6.1.5 testing report result

Figure 6.1.5 shows the testing report result of each class of shop trademarks. The result shows the shop trademarks, accuracy, macro average, weighted average, support, precision, recall and F1-score. It shows that McDonald's, Subway, and Vivo have the highest precision value. This means that in these 3 shop trademark images, the system could classify accurately each image without any false positives. Burger King and Not Shop Trademark were 0.94 in the precision score while starbuck has the lowest precision score among all images. For the recall score, Vivo scored the highest with the score of 1.00 whereas Starbuck has the second highest score 0.94 while the lowest recall score was Subway with 0.67. For the F1-score, the highest was Vivo as both precision and recall were 1.00 thus both precision and recall score were accurate as compared to other shop trademarks. Burger King scored 0.94 for the F1-score as it has both precision and recall score to be 0.94 which was similar to Vivo as both precision and recall score were the same. The lowest F1-score is Starbuck as it obtained 0.73 due to its precision score being very low however it has higher recall which means the system will not falsely predict starbuck as another shop trademark. For McDonald's, Starbuck and Not Shop Trademark have higher precision scores but lower recall scores. The F1-score for each respectively were 0.84, 0.73 and 0.92. This showed that the system may falsely predict the trademark as another shop trademark. The accuracy of the system was overall 0.87 or 87%.

The macro average in the testing report describes the true positive, false negative and false positive of each score rate. From the figure it shows that precision obtained 0.91,

recall obtained 0.86 and F1-score obtained 0.87. For weighted average it describes the average the support weighted mean per label. Precision, recall and F1-score obtained 0.91, 0.87 and 0.88 for the weighted average.

6.2 Testing Setup and Result

The system is being tested using a webcam. The camera will capture the image from the phone and produce the classification result. The setup is described below.

6.2.1 Starbuck Result

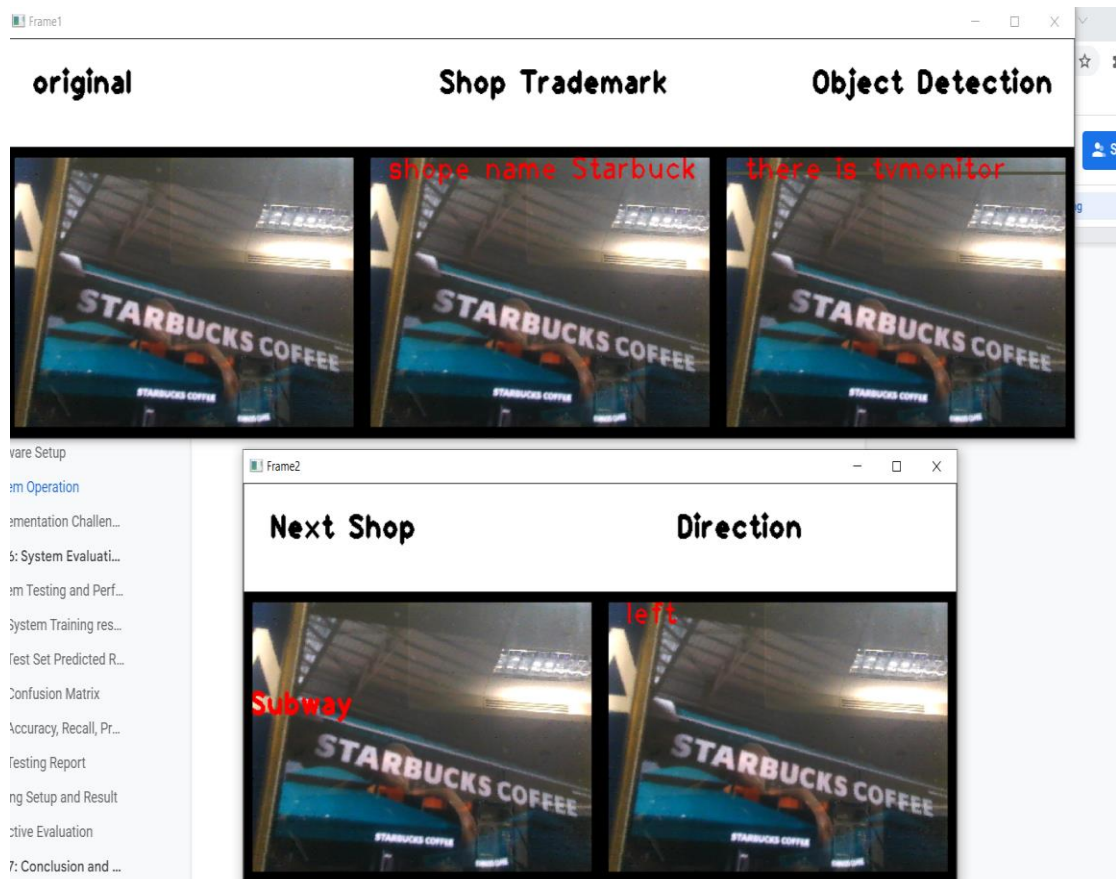


Figure 6.2.1.1 Starbuck test result

Figure 6.2.1.1 shows the test result when the Starbuck trademark is captured. The original frame shows the picture of Starbuck from a mobile phone and the next frame shows the classification from the system which is labeled “shop name Starbuck”. This means the system managed to classify the image correctly. On the object detection frame, the image was displayed on the mobile phone and there are reflections which the system may detect wrongly on the reflection and detect a tv monitor. The system labeled it on the object detection frame. On the next shop frame, the system displays

Subway as the next shop after Starbuck. The next shop will inform the user via voice output of the next shop ahead which follows the sequence of the graph map. The graph mapping takes the result of the classification and search it neighbour shops and return the next shop. The direction frame shows the shop is on the left since the image is a little bit to the left hence the system to obtain the coordination of the trademark is on the left. However, in **figure 6.2.1.2** shows another Starbuck image.



Figure 6.2.1.2 another Starbuck test result

Figure 6.2.1.2 shows another Starbuck image being classified by the system. The result shows that the system can classify correctly however the object detected by the system is wrong as the bounding rectangle is focusing on the trademark and the system predicted it as a clock. The next shop is shown to be Subway and the direction of the shop is on the left as the phone seems to be more to the left.

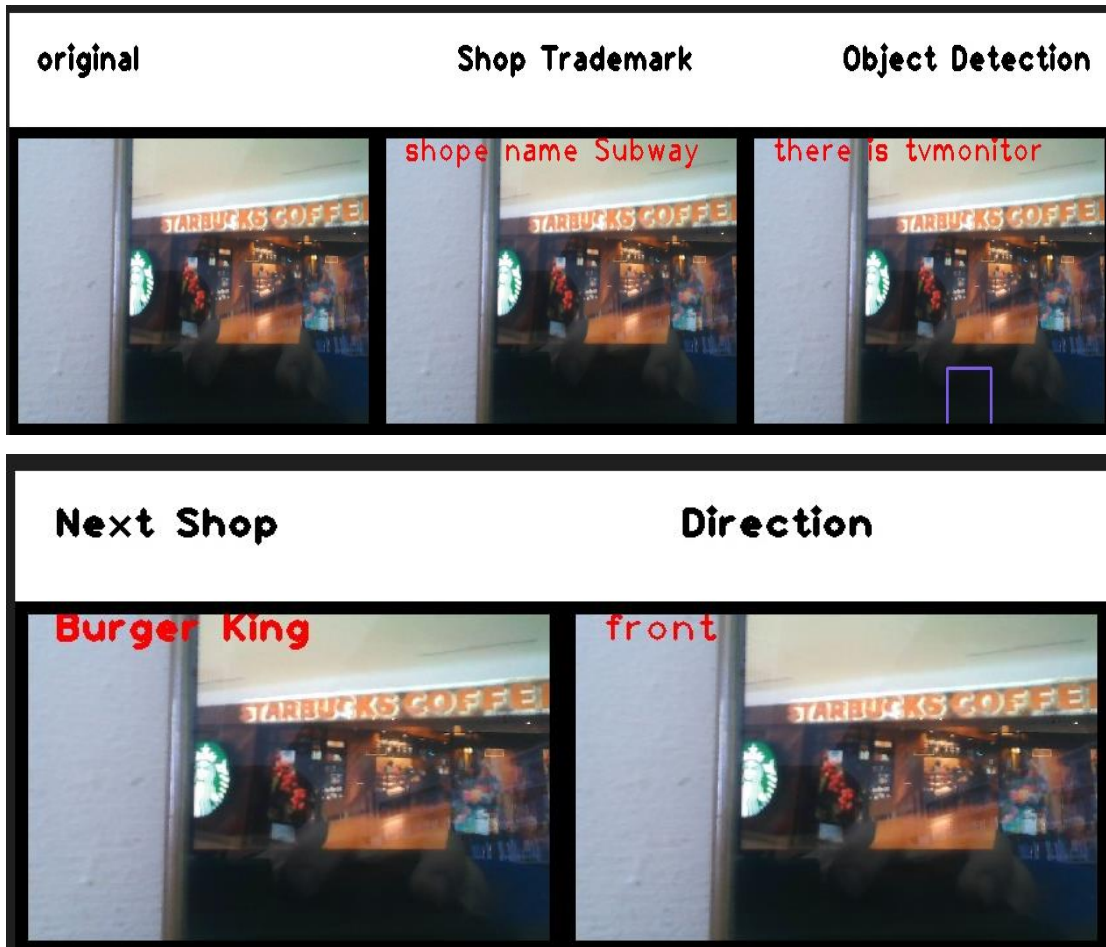


Figure 6.2.1.3 the image is shift to the right

Figure 6.2.1.3 shows the image is shifted to the right and the system classified the image as Subway although the actual result is Starbuck. The object detection detected a tv monitor but no tv monitor is shown in the image. The direction return by the system is front direction which means the shop is in front of the visually impaired. The expected result is right for the direction as the image is more towards the right however, the system may obtain the threshold coordinates to be in between 2400 and 4000 hence obtaining the result as front. The classification of the shop trademark and object detection may receive a value that is similar to Subway and a tv monitor. This shows that when the image is shifted to the left or right or middle, the system may provide different output sometimes.

6.2.2 Burger King Result

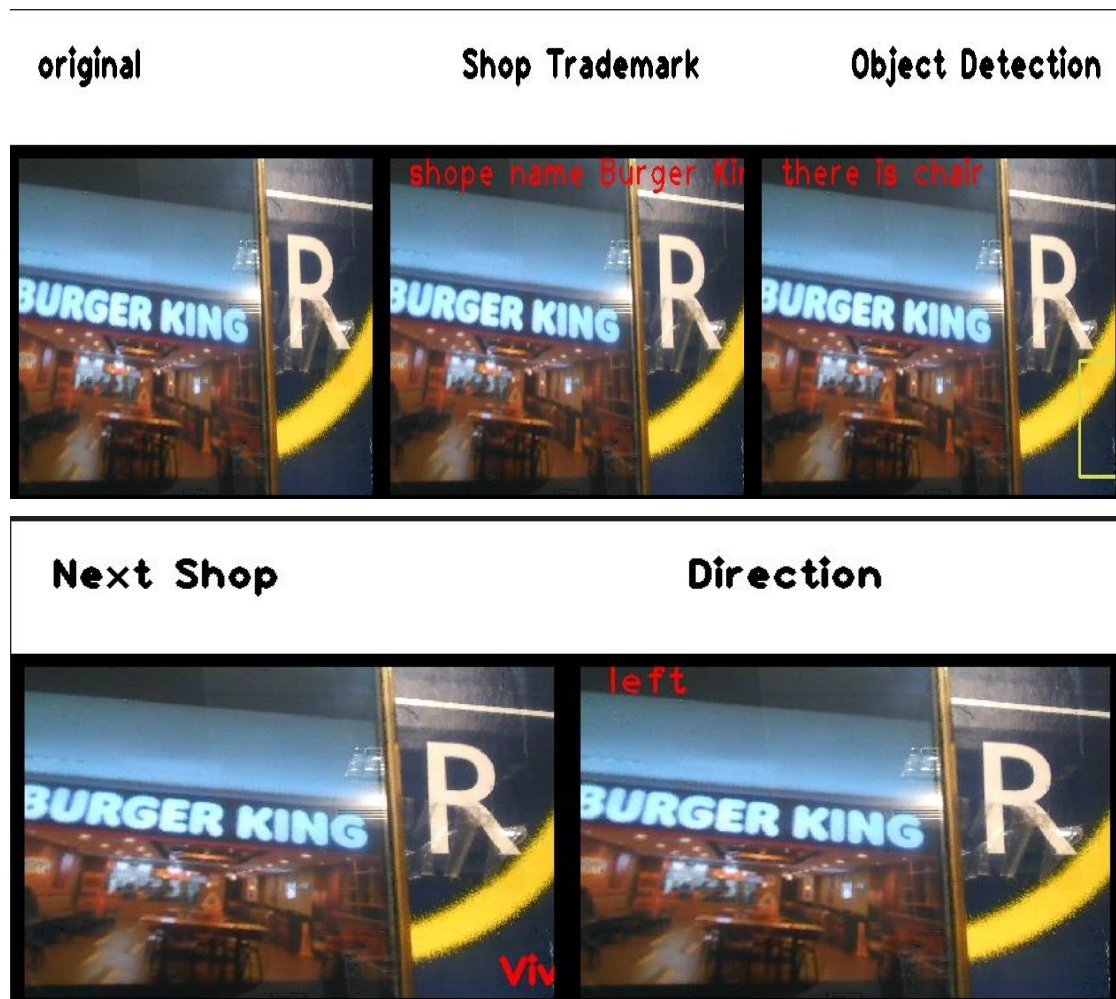


Figure 6.2.2.1 Burger King test result

Figure 6.2.2.1 shows the result of the classification of Burger King. The system managed to classify correctly on the shop trademark and also managed to detect a chair in the image. The objects in the image in **figure 6.2.2.1** are not shown clearly due to the lighting setting from the mobile phone being dark. Since the system classifies the image as Burger King the system will check the next shop from the graph and return the next shop which is Vivo. However, the Vivo label shown in **figure 6.2.2.1** shows that half of the Vivo letters are covered by the line as it was too much to the right. For the direction, the system managed to get the direction correctly as the trademark was more to the left hence the system could verify the image is on the left. However, **figure 6.2.2.2** shows the image of Burger King that is sheared to the left.

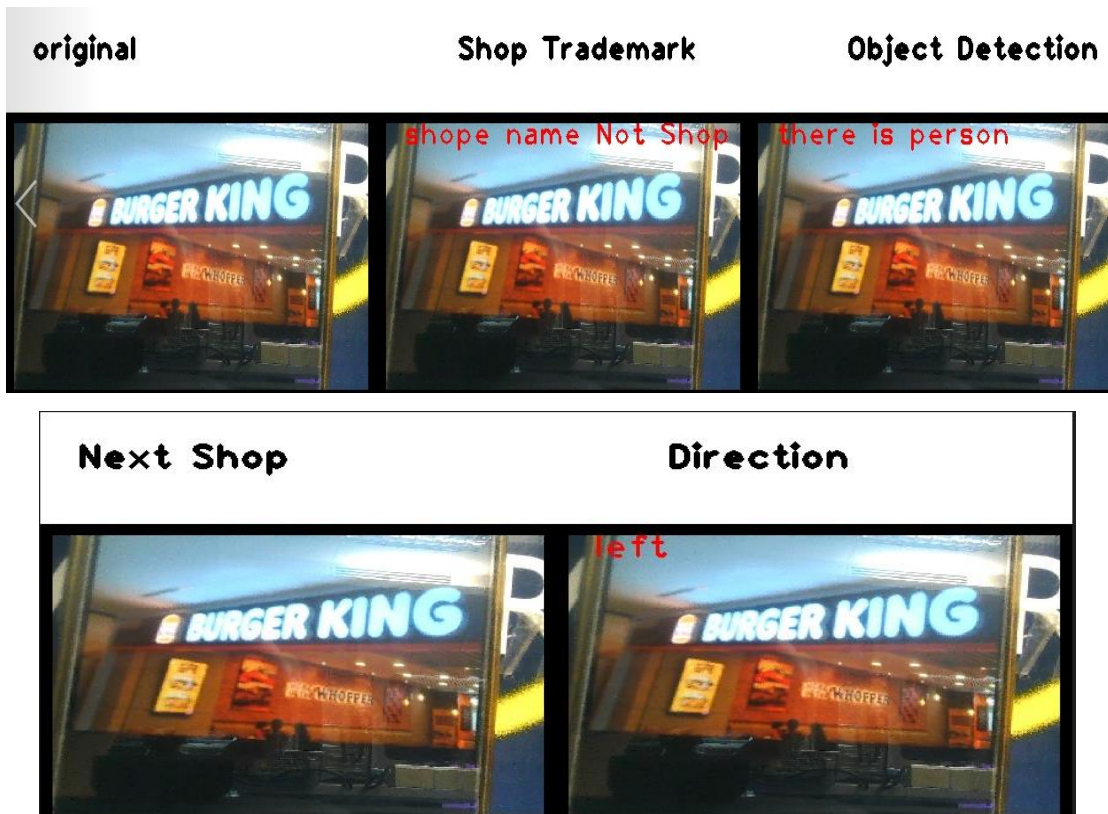


Figure 6.2.2.2 false prediction result of Burger King

The result shows in **figure 6.2.2.2** that the system classifies the image wrongly. The system classifies the image as a Not Shop Trademark which means the system classifies the image as a random object instead of a shop trademark. The object detection could detect a person in the image as it can be seen in **figure 6.2.2.2** but not very clearly. Due to the classification of the image as Not Shop Trademark, the next shop is empty as it does not exist in the graph mapping therefore there is not any label shown in the Next Shop frame. The direction shown in the direction frame is left although the image is just in front, the system detects another contour which causing the bounding rectangle to focus more on the left side of the image making the image to verify the as left due to the value returned from the getContour function.

6.2.3 Vivo Result



Figure 6.2.3.1 Vivo Result output

Figure 6.2.3.1 shows the shop trademark Vivo being captured by the camera. The result shown is Not Shop Trademark which is not the expected result. The system falsely predicted Vivo as Not Shop Trademark was due to the lack of Vivo images being obtained and trained causing the system to predict poorly on the image. The object detector frame shows that the system detected a refrigerator, but the image does not have any refrigerator inside. The image is shown more to the right however, the system verifies the image is on the left where the actual direction is on the right. This shows that the system can provide wrong direction sometimes due to the contour detection detecting the external things in the environment.

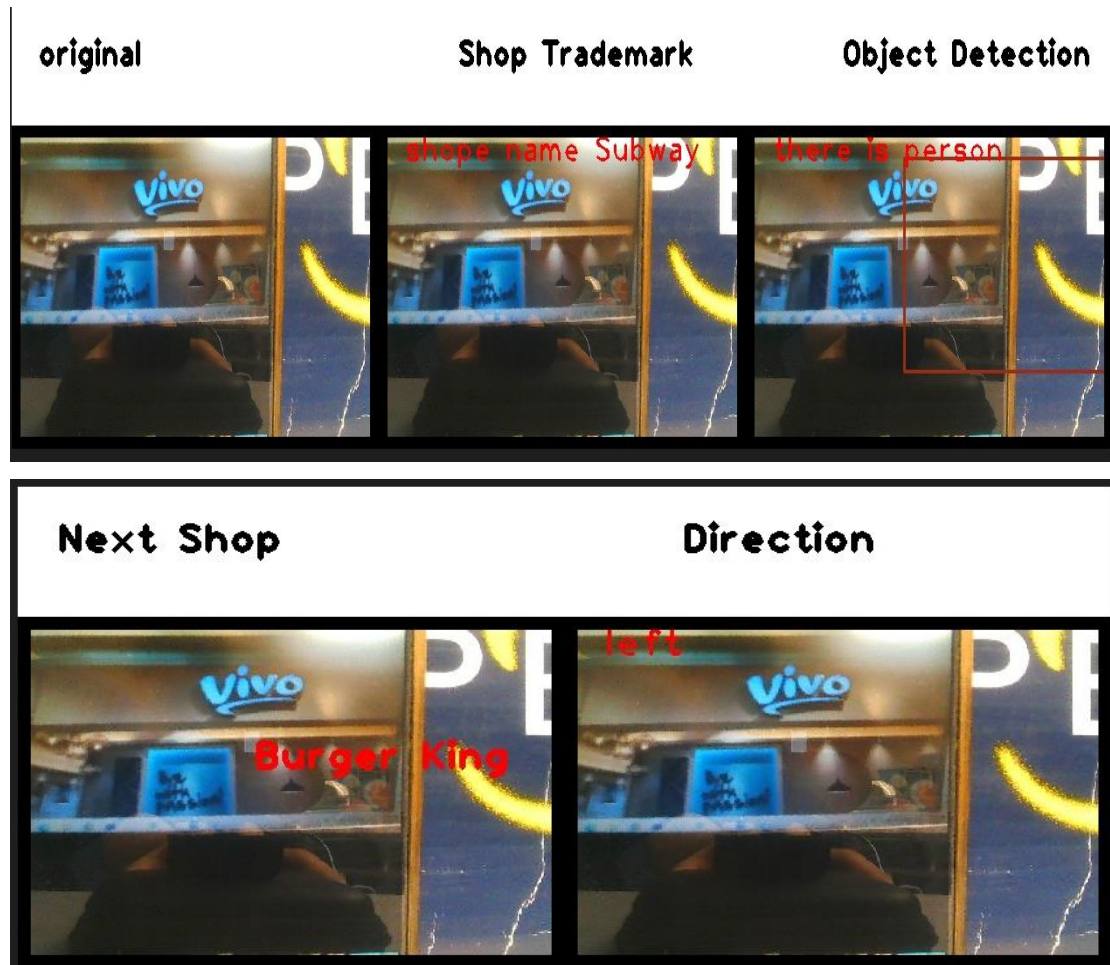


Figure 6.2.3.2 placing the image closer to the camera

The Vivo image was placed closer to the camera hence the system predicts the Vivo image as Subway. The system uses colours to learn the features of each shop trademark as a result the colour of Subway and the colour of Vivo may look similar in some aspects therefore the system may predict it as Subway. The object direction frame shows that the system detected a person in the image. This shows the pretrained yolo version may falsely predict objects in real time. The direction of the is on the left while the next shop is Burger King.



Figure 6.2.3.3 shear angle of Vivo image

Figure 6.2.3.3 shows that another image of Vivo is captured by the system however the image seems to be blurry. The system managed to classify the image as McDonald's while the system may detect the mobile phone as a tv monitor. Due to the colours of the image which has some similarity as McDonald's the system may falsely classified as McDonald's due to the colour value that is similar to McDonald's image. The direction of the shop trademark is on the left where the expected result is either front or left.

6.2.4 Subway Result



Figure 6.2.4.1 Subway result

Figure 6.2.4.1 shows the Subway being predicted by the system. The system could predict accurately on the image as Subway. The system could also detect people in the image although it is not clear in the image shown in **figure 6.2.4.1**. For the Next Shop frame, it is not showing the label because the system may place the label on the wrong position that is blocked by the border line. This is because the label follows the coordinates of the bounding rectangle of the object detection hence it will be hidden if the bounding rectangle coordinate is located at the borderline. For the direction, the system managed to determine the direction and the result shown is on the left.



Figure 6.2.4.2 the image is shift to right

Figure 6.2.4.2 shows the Subway image shift to the right. The image is classified as Subway by the system and the object detection detected a person in the image although the person in the image is not clear. The next shop followed the graph mapping which shows the next shop is Burger King. The direction of the shop trademark is determined to be in front by the system. The direction is still correct although the image is position to a little to the right.



Figure 6.2.4.3 Another Subway testing result

Figure 6.2.4.3 shows another Subway shop trademark. The system could predict this image correctly as shown in the shop trademark frame that the shop is Subway. On the object detection frame, it shows that the object detector detected a train in the image while for the Next Shop frame the system shows Burger King as the next shop based on the graph map. The Direction frame shows the shop trademark is on the left from the camera view.

6.2.5 McDonald's Result

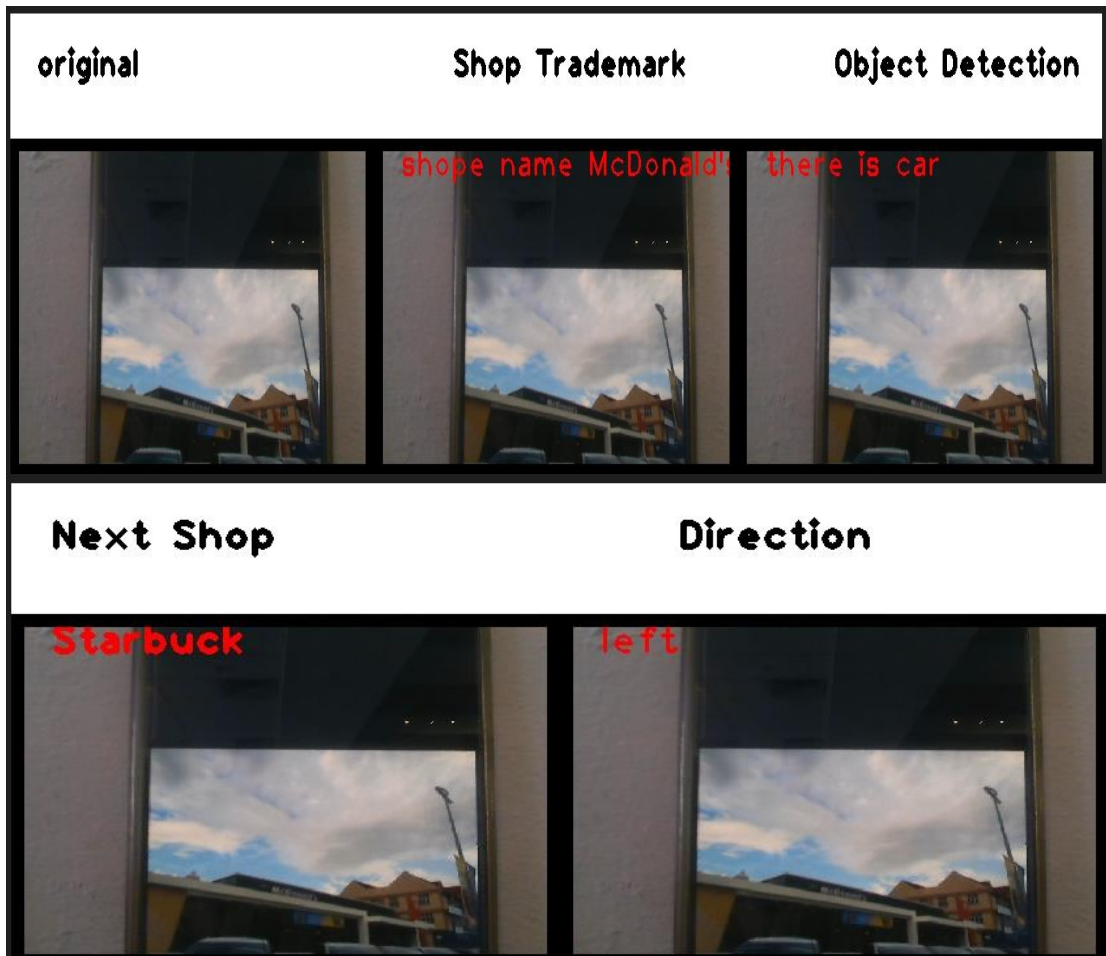


Figure 6.2.5.1 McDonald's result

Figure 6.2.5.1 shows the McDonald's result. The system classifies the image shop trademark as McDonald's. The system predicted correctly on the shop trademark and the object detected by the system is a car. **Figure 6.2.5.1** did not show clearly of the car but in the phone image there is cars hence the system will detect a car in the image. Since in the graph mapping shows the sequence of the shop location, the next shop after McDonald's is Starbuck which the system returns when verifying from the graph mapping. The direction of the shop trademark is on the left but the expected result should be "in front" instead of left as the image is place in front of the camera.



Figure 6.2.5.2 McDonald's result in a nearer view

Figure 6.2.5.2 shows the result when the camera captured McDonald's in real time. The system shows the CNN model classified the image as Starbuck although the actual image is McDonald's. This indicates the model may falsely predict some images in real time though compared to the image prediction through the testing set, the system can predict more accurately from the dataset than in real time. The Object Detection frame shows the Yolo model detecting a car in the image in real time. However, for the Direction frame shows the system verify the shop is on the left though the expected result for the direction should be front as the trademark is shown right in front of the camera.

6.2.6 Not Shop Trademark Result

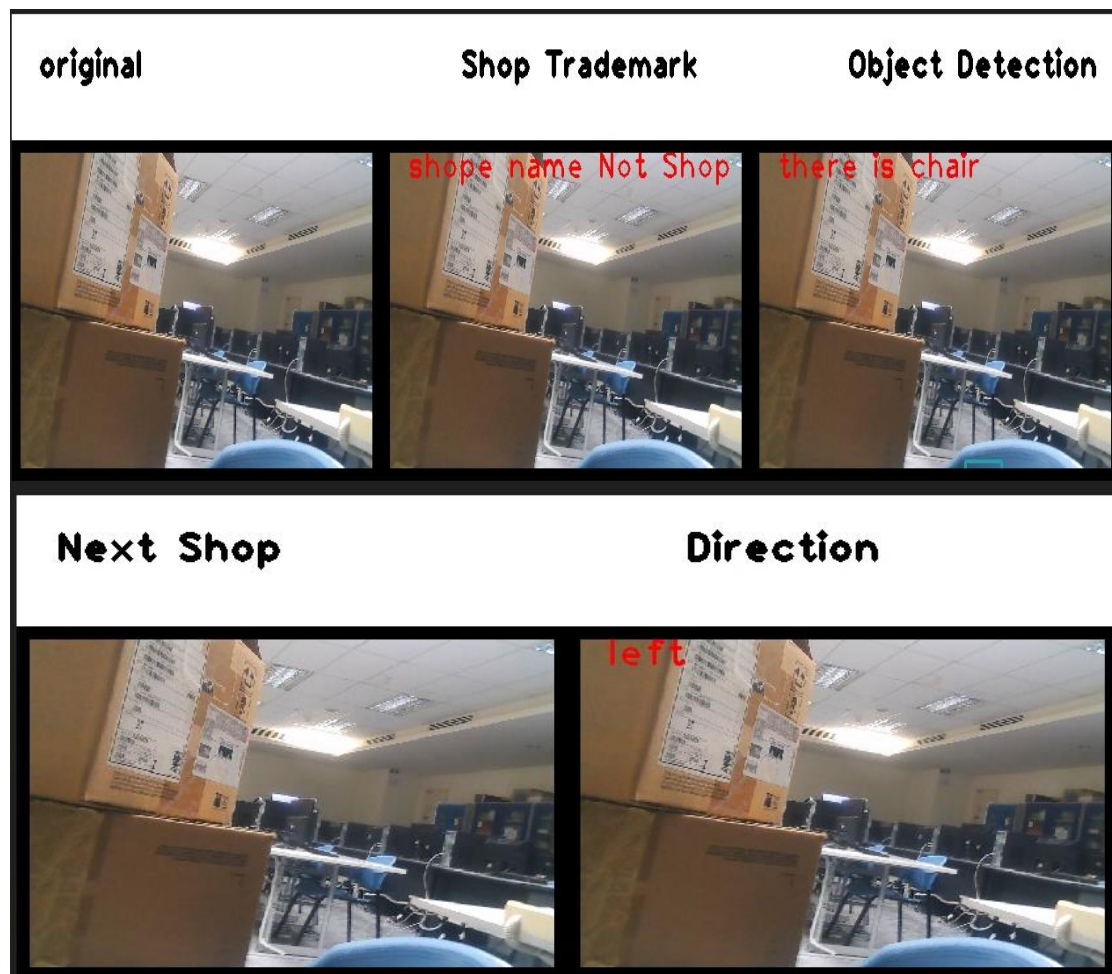


Figure 6.2.6.1 testing on the lab environment in Utar

Figure 6.2.6.1 shows that the camera captured the lab environment in Utar to test the system if the system could classify the image is not a shop trademark or is a shop trademark. The result in the shop trademark frame in **figure 6.2.6.1** is not shop trademark. Sometimes, the system may classify the environment as a shop trademark which is shown in **figure 6.2.6.2**. The Yolo model detected a chair in the image but for the next shop frame it shows nothing as in the graph mapping, all nodes represent a shop trademark and not shop trademark indicates there is no shop trademark detected by the system hence the graph will not have not shop trademark thus the result shows it is empty. For the Direction frame, the system detected contour and displays the result left.



Figure 6.2.6.2 system detected shop trademark in the lab environment

Figure 6.2.6.2 shows another result of the system capturing the same environment but showing another result such as the system classifying the image as McDonald's and the system detect tv monitor with a bounding box focusing on the air conditioner vent. This is because the image colour predicted by the system may have a similar result as the as colour of McDonald's and the object detector could only detect 85 objects, it does not have any training about air conditioner and could only detect based on the closest value or else display nothing on the object detection frame.

6.2.7 Voice Command Result

```
Yes master! I'm here! which shop you want to go?
The actual voice input is I'm looking for McDonald's
The text is split into ["I'm", 'looking', 'for', "McDonald's"]
The filter word is to search for ["McDonald's"]
```

Figure 6.2.7.1 voice command and output result

Figure 6.2.7.1 shows the voice command result. The function `voiceCommand()` will be called by the system and wait for the visually impaired to called out the name of the system "computer". When the system receives an input and validates if it is the word "computer" through speech-to-text conversion, the system will prompt "Yes master! I'm here! Which shop you want to go?" the system will wait for another voice input from the visually impaired to say the destination shop. The voice input "I'm looking for McDonald's" is received by the system and the system will convert speech-to-text where the text is pass to the `filterText()` function and split the text into a list of strings as such it is shown in the following and also in **figure 6.2.7.1**.



When the text is split into a list of strings, the system will loop and validates if any of the strings is a shop trademark. If true, the function will filter any string that is not a shop trademark and keep only the shop trademark string which is shown in **Figure 6.2.7.1**.

If the strings do not have any shop trademark, the function will return null back to the main function and the system will assign “Not Shop Trademark” as default value.

6.3 Comparison of Testing Result Between Real Time and Testing datasets

After running the system on the testing datasets and in real time, it shows that the system performs poorly in real-time due to many factors. In real-time, the environment factor such as the light intensity, vibration, shear range, and position may affect the classification of the image. The light intensity may affect the value of the colours as a result, making the image to be predicted as false positive images. Vibration causes the camera to capture distorted images or blurry images which makes the prediction to be poor. Shear range in images means the angle of capturing the image. Since the model relies on the datasets to train and learn, the model can only classify the few angles of the image and with the lack of datasets to train the model. The position of the image affects the classification as the image may processes a different value resulting in the model to classify the value closest to the shop trademark.

Another factor is Yolo model running in CPU. Yolo requires high processing power which will use lots of CPU power, and this will slow down other unit as the priority is given to Yolo to process. Therefore, the system will perform slower in real-time causing the camera to capture 1 frame-per-second and processed by the model. If the image is blurry or tilted, the model will predict wrongly due to the noisy effect. While the image in the testing datasets is preprocessed and the image taken is clear and not blurry which allows the system to predict easily with accurate classification.

Finally, in-real time the shop trademark captured by the camera would capture the external objects which may affect the classification as the system due to the colour value interference. Unlike the testing datasets, the images do not have any external objects captured to be processed hence the colour value and edges from the datasets will be approximately similar to the one from the training dataset.

6.4 Project Challenges

6.4.1 Unresolved Challenges

There were some challenges encounters during the development phase and the challenges were stated following:

1. System performance

The system performance is too slow and lagging when being executed. Due to the speech-to-text and text-to-speech implementation, the system will take more than 3 second to process the image and produce the output. The Yolo model uses high consumption of the CPU unit to process to the object detections causing the camera to take only 1fps and together with the implementation of speech-to-text and text-to-speech, the system will slow down even more and allowing the system to delay in processing.

2. Real-time image capturing environment

The environment affects the system prediction on the shop trademarks. Due to lighting in the environment, the image captured by the camera be too bright or too dark causing the system to predict falsely on the image as the system relies on the colours to classify. Hence, the system will predict many false results during real-time.

6.4.2 Resolved Challenges

There were some challenges that were encountered however it was solve during the development phase.

1. Direction of the shop trademark

This function is to allow the system to inform the visually impaired when the targeted shop locates when the system captured it. The function requires the coordination of the shop trademark so that the system could identify it is on the left or right. The solution to solve this is to get the contour in the image, with the contour detected by the system, the bunding rectangle can be obtained hence using the edge of the bounding rectangle to set it as the coordinates. Although the contour may not provide an accurate result, it does work in some condition.

2. Datasets

Shop trademark datasets are very much not available online hence without any of the datasets it is impossible to train the model. To resolve this challenge, the datasets were obtained by taking pictures from the streets and shopping mall. Using these datasets, this will allow the model to train and test. However, the picture taken by the phone has low resolution which the model will take more period to train, and the number of datasets obtained is not many due to the limited storage capacity.

3. Training objects for object detection

This project requires to detect objects so that it can inform visually impaired of the objects in front however training a model to detect is tedious and may requires a lot of effort and time. The model to detect objects may need a lot of datasets to train and localize the object and the model will need GPU to run the system to allow the model to learn quickly. The solution to resolve this challenge is to use the pretrained Yolo model done by [27]. The pretrained model managed to detect 85 objects and the model is compatible to run in CPU. This pretrained Yolo model helps the system to detect many objects although some detection may be false.

6.5 Objective Evaluation

Based on the objectives stated in **1.3**, the system should recognize shop trademarks when being captured by the camera hence based on the testing result shown in **6.2**, the system could perform real time classification however the accuracy could vary with the testing on the datasets. In real-time, the CNN model will classify the image The system could detect objects in the surroundings however certain objects such as air conditioner, flowerpot, escalator are undetectable by the system due to the limitation of the Yolo model. The system manages to convert the text-to-speech which can be shown in **5.2.2** the system prints the output of the voice and the system receives voice input when the visually impaired say “computer”. If the system receives the input computer, the system will voice out the message prompting the visually impaired to tell the targeted shop trademark.

Following the performance of the system, it managed to achieve the objectives of this project however some improvement shall be made because the classification, object

detection may not produce very high accuracy in real-time. The direction of the shop trademark requires improvement as it uses contour and bounding rectangle to determine the coordination of the x-axis. The direction should be more accurate if the system implements shop trademark detection which will get a more accurate result of the direction of the shop trademark.

CHAPTER 7: Conclusion and Recommendation

7.1 Conclusion

The visually impaired community are too afraid to go to shopping malls alone as there required guide or guide dog to help them to move around. Due to many obstacles such as objects and many people in the shopping mall, it is harder for visually impaired to have less or always avoid collision inside the shopping mall. There are no braille words available for visually impaired to read or to inform their location which cause them to not know the location of the targeted premise, or they enter the right premise inside the shopping mall.

The motivation of this project is to help visually impaired to read shop trademark inside shopping mall and to avoid object collision inside the shopping mall by designing an image processing device to help identify shop trademarks. Hence, the shop trademark recognition and object detection are important for the project to fulfil the objectives. The shop trademark recognition is meant to help identify shop trademark and will read out the name to the user. The object detection will inform the visually impaired of the objects in front hence allowing the visually impaired to avoid it.

So, to determine how to build the system, YOLO and CNN algorithms are used in the project. By using CNN, it will train the system to identify the shop trademark and predict the name of the shop trademark. The image will be first pre-process to clean the images then it will load into the CNN model. The CNN model will loop 30 times to identify the images and validate the image with the validation datasets. Once complete, a testing dataset contains 23 datasets divided into 3 batches that were predicted by the model. The YOLO algorithm will help determine the objects and when the camera captures an object, it will return a result which will be inserted into the parameter of the pyttsx3 or speech engine which will tell the visually impaired of the object. The YOLO algorithm contains two files which are already trained and just require to be inserted into the system. A text file name as cocoa.txt consists of names of objects that the two-files recognized. When the camera captures an object, it will localise the object and predict the name of the object which will be displayed on the screen. At the end, the final delivery will help to resolve the problem with the right implementation and algorithm used.

7.2 Recommendation

The system has some flaws such as slow performance in running Yolo, false prediction or detection by both CNN and Yolo model. Wrong direction provided by the system hence some recommendations are needed to improve the system in the future. The recommendations are stated as follows:

1. Implement GPU type of Yolo model to allow the system process the image faster and smoother
2. Train the CNN model to perform shop trademark detection so that the system could provide more accurate direction of the shop trademark and inform the visually impaired when the system detected a shop trademark.
3. Run more datasets so that the CNN model could recognise more shop trademarks and perform more accurate predictions in real time.
4. Design an indoor mapping to allow the system to keep track of the visually impaired location inside the shopping mall.
5. Implement multithreading to allow text-to-speech and speech-to-text to run concurrently which reduce lagging in the system

References

1. B. Gozick, K. P. Subbu, R. Dantu, and T. Maeshiro, "Magnetic Maps for Indoor Navigation." *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 12, pp. 3883-3891, 2011, doi: 10.1109/tim.2011.2147690.
2. B. Ozdenizci, V. Coskun, and K. Ok, "NFC Internal: An Indoor Navigation System." *Sensors*, vol. 15, no. 4, pp. 7571-7595, 2015, doi: 10.3390/s150407571.
3. C. Tsirmpas, A. Rompas, O. Fokou, and D. Koutsouris, "An indoor navigation system for visually impaired and elderly people based on Radio Frequency Identification (RFID)." *Information Sciences*, vol. 320, pp. 288-305, 2015, doi: 10.1016/j.ins.2014.08.011
4. D. Gusenbauer, C. Isert, and J. Krösche. "Self-contained indoor positioning on off-the-shelf mobile devices .." *IEEE Xplore*. 2010
5. D. I. Heywood, S. Cornelius, and S. Carver, *An Introduction to Geographical Information Systems*. UK: Prentice Hall, 2006
6. D. López-de-Ipiña, T. Llorido, and U. López, "Indoor Navigation and Product Recognition for Blind People Assisted Shopping." *Ambient Assisted Living*, pp. 33-40, 2011, doi: 10.1007/978-3-642-21303-8_5.
7. D. Pfeiffer, F. Erbs, and U. Franke, "Pixels, Stixels, and Objects." *Computer Vision – ECCV 2012. Workshops and Demonstrations*, pp. 1-10, 2012, doi: 10.1007/978-3-642-33885-4_1.
8. D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, "NavCog3." *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2017, doi: 10.1145/3132525.3132535..
9. H. Badino, U. Franke, and D. Pfeiffer, "The Stixel World - A Compact Medium Level Representation of the 3D-World." *Lecture Notes in Computer Science*, pp. 51-60, 2009, doi: 10.1007/978-3-642-03798-6_6.
10. H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarre, and D. Rus, "Enabling independent navigation for visually impaired people through a wearable vision-based feedback system." *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, doi: 10.1109/icra.2017.7989772.
11. M. Weyn, *Opportunistic Seamless Localization*. Antwerp, Belgium, 2011.
12. M.-Y. Liu, S. Lin, S. Ramalingam, and O. Tuzel, "Layered Interpretation of Street View Images." *Robotics: Science and Systems XI*, 2015, doi: 10.15607/rss.2015.xi.025.

REFERENCES

13. R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology." *Insights into Imaging*, vol. 9, no. 4, pp. 611-629, 2018, doi: 10.1007/s13244-018-0639-9.
14. S. A. Cheraghi, V. Namboodiri, and L. Walker, "GuideBeacon: Beacon-based indoor wayfinding for the blind, visually impaired, and disoriented." *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1-10, 2017
15. T. Gallagher, E. Wise, B. Li, A. G. Dempster, C. Rizos, and E. Ramsey-Stewart, "Indoor positioning system based on sensor fusion for the Blind and Visually Impaired." *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-9, 2012, doi: 10.1109/ipin.2012.6418882.
16. T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen. "A Probabilistic Approach to WLAN User Location Estimation .." SpringerLink. Apr. 2002
17. T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth, "Stixmantics: A Medium-Level Model for Real-Time Semantic Scene Understanding." *Computer Vision – ECCV 2014*, pp. 533-548, 2014, doi: 10.1007/978-3-319-10602-1_35.
18. W. C. S. S. Simoes and V. F. De Lucena, "Blind user wearable audio assistance for indoor navigation based on visual markers and ultrasonic obstacle detection." *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, doi: 10.1109/icce.2016.7430522.
19. Y. H. Lee and G. Medioni, "RGB-D camera based wearable navigation system for the visually impaired." *Computer Vision and Image Understanding*, vol. 149, pp. 3-20, 2016, doi: 10.1016/j.cviu.2016.03.019.
20. Y. Li, A. Tsai, P. Mumford, W. Lin, and I.-chou Hong. "Continuous high precision navigation using MEMS inertial sensors aided RTK GPS for mobile mapping applications." Semantic Scholar. Jan. 01, 1970.
21. "Raspberry Pi 3 Model B+." Cytron Marketplace. <https://my.cytron.io/p-raspberry-pi-3-model-b-plus>. [Accessed: 03-March-2022]
22. M. Uribe-Fernández, N. SantaCruz-González, C. Aceves-González, and A. Rossa-Sierra, "Assessment of How Inclusive Are Shopping Centers for Blind People." *Advances in Intelligent Systems and Computing*, pp. 86-97, 2018, doi: 10.1007/978-3-319-94622-1_9.

REFERENCES

23. Sarang Narkhede, "Understanding Confusion Matrix," *Medium*, May 09, 2018. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>. [Accessed: 06-Sep-2022]
24. Keras, "Home - Keras Documentation," *Keras.io*, 2019. <https://keras.io/>. [Accessed: 04-Sep-2022]
25. What is the Jupyter Notebook? — Jupyter/IPython Notebook Quick Start Guide 0.1 documentation," *jupyter-notebook-beginner-guide.readthedocs.io*. https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html. [Accessed: 04-Sep-2022]
26. <https://www.facebook.com/jason.brownlee.39>, "Introduction to the Python Deep Learning Library TensorFlow," *Machine Learning Mastery*, May 04, 2016. <https://machinelearningmastery.com/introduction-python-deep-learning-library-tensorflow/>. [Accessed: 04-Sep-2022]
27. J. Redmon, "YOLOv3: An Incremental Improvement," *Yolo: Real-time object detection*. [Online]. Available: <https://pjreddie.com/darknet/yolo/>. [Accessed: 06-Sep-2022].
28. G. Karimi, "Introduction to YOLO Algorithm for Object Detection," *Engineering Education (EngEd) Program / Section*, Apr. 15, 2021. <https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/>. [Accessed: 04-Sep-2022]
29. K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *arXiv:1511.08458 [cs]*, Dec. 2015,

Appendix A: WEEKLY REPORT

FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

| | |
|--|-------------------|
| Trimester, Year: Y3S3 | Study week no.:11 |
| Student Name & ID: Brandon Ling Yi Yun | |
| Supervisor: Professor Leung Kar Hang | |
| Project Title: Indoor navigation for visually Impaired to read shop trademark in malls | |

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

1. Managed to display multiple video frame in one window
2. Managed to produce bounding rectangle on the objects

2. WORK TO BE DONE

1. Resolving Pytsx3 runandwait() function that causes system lagging
2. Debugging bounding rectangle when recognizing shop trademarks as it does not display when testing
3. Displaying legend on to the video window
4. Test the system performance and finding defects if any

3. PROBLEMS ENCOUNTERED

1. Unable to run tensorflow in GPU
2. Voice activation will slow down the system due to conversion of text to speech

4. SELF EVALUATION OF THE PROGRESS



Supervisor's signature

23 Aug 2022



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

| | |
|--|------------------|
| Trimester, Year: Y3S3 | Study week no.:9 |
| Student Name & ID: Brandon Ling Yi Yun | |
| Supervisor: Professor Leung Kar Hang | |
| Project Title: Indoor navigation for visually Impaired to read shop trademark in malls | |

| |
|--|
| 1. WORK DONE [Please write the details of the work done in the last fortnight.] <div style="margin-top: 10px;"> <ol style="list-style-type: none"> 1. Create graph for indoor mapping 2. System able to inform user the next shop trademark ahead </div> |
| 2. WORK TO BE DONE <div style="margin-top: 10px;"> <ol style="list-style-type: none"> 1. Improve system performance such as time complexity 2. Train more datasets so system can predict accurately of the shop trademark </div> |
| 3. PROBLEMS ENCOUNTERED <div style="margin-top: 10px;"> <ol style="list-style-type: none"> 1. CNN models can predict the shop trademark however will falsely predict some shop trademarks 2. System will contour other unnecessary objects captured by the camera and will provide wrong direction sometimes 3. System will provide wrong location due to falsely predicting the shop trademark resulting the system to tell user of their current location and the next shop to arrive 4. Voice activation causes system to slow down due to the process of converting text-to-speech 5. System uses cpu process hence lower cpu unit will result system performance </div> |
| 4. SELF EVALUATION OF THE PROGRESS <div style="margin-top: 10px; height: 40px;"></div> |



 9 Aug 2022

Bdn

FINAL YEAR PROJECT WEEKLY REPORT*(Project I / Project II)*

| | |
|--|------------------|
| Trimester, Year: Y3S3 | Study week no.:8 |
| Student Name & ID: Brandon Ling Yi Yun | |
| Supervisor: Professor Leung Kar Hang | |
| Project Title: Indoor navigation for visually Impaired to read shop trademark in malls | |

1. WORK DONE*[Please write the details of the work done in the last fortnight.]*

1. No work done

2. WORK TO BE DONE

1. Create graph for the indoor mapping
2. Initiate current location in the graph
3. Allow system to inform user the direction to the destination

3. PROBLEMS ENCOUNTERED

1. CNN models can predict the shop trademark however will falsely predict some shop trademarks
2. System will contour other unnecessary objects captured by the camera and will provide wrong direction sometimes
3. System can't identify the direction to the destination point

4. SELF EVALUATION OF THE PROGRESS


Supervisor's signature

2 Aug 2022



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

| | |
|--|-------------------|
| Trimester, Year: Y3S3 | Study week no.: 7 |
| Student Name & ID: Brandon Ling Yi Yun | |
| Supervisor: Professor Leung Kar Hang | |
| Project Title: Indoor navigation for visually Impaired to read shop trademark in malls | |



1. WORK DONE

[Please write the details of the work done in the last fortnight.]

1. Obtained contour for the characters

2. WORK TO BE DONE

1. Display direction of the shop trademarks
2. Create graph for the indoor mapping

3. PROBLEMS ENCOUNTERED

1. CNN models can predict the shop trademark however will falsely predict some shop trademarks
2. System will contour other unnecessary objects captured by the camera and will provide wrong direction sometimes


4. SELF EVALUATION OF THE PROGRESS

Supervisor's signature
27 Jul 2022

Bdn

Student's signature

APPENDIX B: Poster



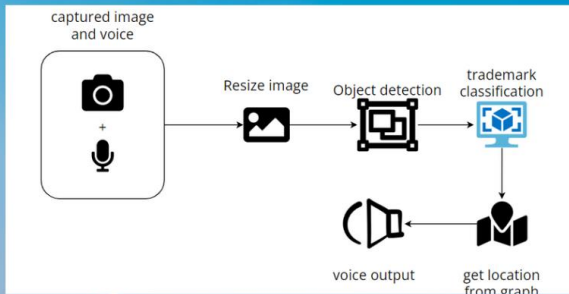
Universiti Tunku Abdul Rahman

Indoor Navigation for Visually Impaired By Reading Shop Trademark in Shopping Mall

INTRODUCTION

To design a system to help visually impaired to read shop trademarkmarks in shopping mall, detects object and listen to command from the visually impaired

METHOD




DISCUSSION


This project uses CNN to recognize Shop Trademark and YOLO to detect objects around the environment to warn the visually impaired of the objects ahead

RESULT

original Shop Trademark Object Detection



original Shop Trademark Object Detection



CONCLUSION

At the end of the project, the system can recognize shop trademark, detects objects in front and listen to command from the visually impaired and them to become more independent when going out alone

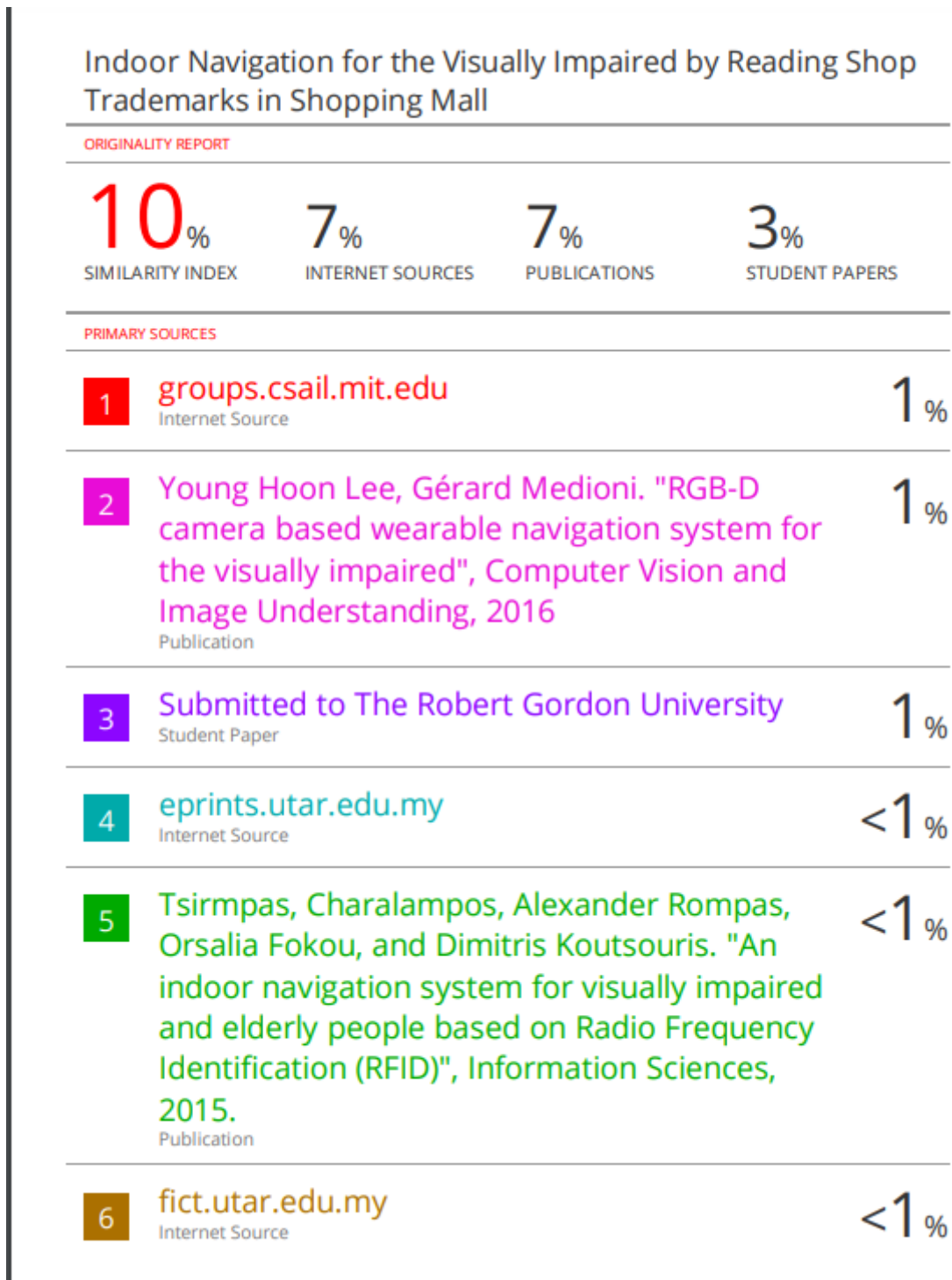
Supervisor

Prof Leung Kar Heng

Student

Brandon Ling Yi Yun

APPENDIX C: PLAGARISM CHECK RESULT



| | | |
|----|---|------|
| 17 | Ozdenizci, Busra, Kerem Ok, Vedat Coskun, and Mehmet N. Aydin. "Development of an Indoor Navigation System Using NFC Technology", 2011 Fourth International Conference on Information and Computing, 2011. Publication | <1 % |
| 18 | Lee, Young Hoon, and Gérard Medioni. "RGB-D camera based wearable navigation system for the visually impaired", Computer Vision and Image Understanding, 2016. Publication | <1 % |
| 19 | ece.anits.edu.in Internet Source | <1 % |
| 20 | Submitted to Universiti Tunku Abdul Rahman Student Paper | <1 % |
| 21 | "Technological Trends in Improved Mobility of the Visually Impaired", Springer Science and Business Media LLC, 2020 Publication | <1 % |
| 22 | library.isical.ac.in:8080 Internet Source | <1 % |
| 23 | www.grafiati.com Internet Source | <1 % |
| 24 | scikit-learn.org Internet Source | <1 % |

| | | |
|----|--|------|
| 7 | www.ri.cmu.edu Internet Source | <1 % |
| 8 | www.section.io Internet Source | <1 % |
| 9 | www.frontiersin.org Internet Source | <1 % |
| 10 | Diego López-de-Ipiña. "Indoor Navigation and Product Recognition for Blind People Assisted Shopping", Lecture Notes in Computer Science, 2011 Publication | <1 % |
| 11 | ijstr.org Internet Source | <1 % |
| 12 | acikerisim.isikun.edu.tr Internet Source | <1 % |
| 13 | res.mdpi.com Internet Source | <1 % |
| 14 | bhadreshsavani.medium.com Internet Source | <1 % |
| 15 | Lecture Notes in Computer Science, 2015. Publication | <1 % |
| 16 | export.arxiv.org Internet Source | <1 % |

| | | |
|----|---|------|
| 25 | Busra Ozdenizci, Vedat Coskun, Kerem Ok. "NFC Internal: An Indoor Navigation System", Sensors, 2015 <small>Publication</small> | <1 % |
| 26 | Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, Chieko Asakawa. "NavCog3", Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17, 2017 <small>Publication</small> | <1 % |
| 27 | www.soundtransit.org <small>Internet Source</small> | <1 % |
| 28 | Submitted to Universiti Teknologi Malaysia <small>Student Paper</small> | <1 % |
| 29 | dspace.mit.edu <small>Internet Source</small> | <1 % |
| 30 | link.springer.com <small>Internet Source</small> | <1 % |
| 31 | www.researchgate.net <small>Internet Source</small> | <1 % |
| 32 | Submitted to University of Hertfordshire <small>Student Paper</small> | <1 % |
| 33 | atrium.lib.uoguelph.ca <small>Internet Source</small> | <1 % |

| | | |
|----|---|------|
| 34 | Prathyusha Chalasani, S. Rajesh. "Lung CT Image Classification using Deep Neural Networks for Lung Cancer Detection", International Journal of Engineering and Advanced Technology, 2020 Publication | <1 % |
| 35 | arxiv.org Internet Source | <1 % |
| 36 | repository.tudelft.nl Internet Source | <1 % |
| 37 | staff.utar.edu.my Internet Source | <1 % |
| 38 | "Artificial Intelligence and Evolutionary Computations in Engineering Systems", Springer Science and Business Media LLC, 2018 Publication | <1 % |
| 39 | "Image Analysis and Processing - ICIAP 2017", Springer Science and Business Media LLC, 2017 Publication | <1 % |
| 40 | Submitted to Coventry University Student Paper | <1 % |
| 41 | Submitted to University of Bedfordshire Student Paper | <1 % |

| Universiti Tunku Abdul Rahman | | | |
|---|------------|----------------------------|------------------|
| Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes) | | | |
| Form Number: FM-IAD-005 | Rev No.: 0 | Effective Date: 09/09/2022 | Page No.: 1 of 1 |



FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

| | |
|------------------------------|--|
| Full Name(s) of Candidate(s) | Brandon Ling Yi Yun |
| ID Number(s) | 19ACB06748 |
| Programme / Course | Computer Science CS |
| Title of Final Year Project | Indoor Navigation for the Visually Impaired to Read Shop Trademark |

| Similarity | Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR) |
|---|--|
| Overall similarity index: <u>10</u> % Similarity by source Internet Sources: <u>7</u> % Publications: <u>7</u> % Student Papers: <u>3</u> % | |
| Number of individual sources listed of more than 3% similarity: <u>0</u> | |
| Parameters of originality required and limits approved by UTAR are as Follows: (i) Overall similarity index is 20% and below, and (ii) Matching of individual sources listed must be less than 3% each, and (iii) Matching texts in continuous block must not exceed 8 words <i>Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.</i> | |

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Leung Kar Hang

Date: 8 Sep 2022

Signature of Co-Supervisor

Name: _____

Date: _____

APPENDIX D: CHECKLIST**UNIVERSITI TUNKU ABDUL RAHMAN**
**FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY
(KAMPAR CAMPUS)**
CHECKLIST FOR FYP2 THESIS SUBMISSION

| | |
|-----------------|---------------------|
| Student Id | 19ACB06748 |
| Student Name | Brandon Ling Yi Yun |
| Supervisor Name | Prof Leung Kar Hang |

| TICK (✓) | DOCUMENT ITEMS |
|-----------------|---|
| | Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item. |
| | Front Plastic Cover (for hardcopy) |
| ✓ | Title Page |
| ✓ | Signed Report Status Declaration Form |
| ✓ | Signed FYP Thesis Submission Form |
| ✓ | Signed form of the Declaration of Originality |
| ✓ | Acknowledgement |
| ✓ | Abstract |
| ✓ | Table of Contents |
| ✓ | List of Figures (if applicable) |
| ✓ | List of Tables (if applicable) |
| ✓ | List of Symbols (if applicable) |
| ✓ | List of Abbreviations (if applicable) |
| ✓ | Chapters / Content |
| ✓ | Bibliography (or References) |
| ✓ | All references in bibliography are cited in the thesis, especially in the chapter of literature review |
| ✓ | Appendices (if applicable) |
| ✓ | Weekly Log |
| ✓ | Poster |
| ✓ | Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005) |
| ✓ | I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report. |

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

Bdn

(Signature of Student)

Date: 09/09/2022

