DOKUZ EYLUL UNIVERSITY
ENGINEERING FACULTY
DEPARTMENT OF COMPUTER ENGINEERING

DETECTION OF PRODUCT INFORMATION FOR PEOPLE WITH VISUAL IMPAIRMENTS

by

İrfan Erdem ŞENYENER

Advisor

Asst. Prof. Dr.Yunus DOĞAN

November, 2022

Izmir

DETECTION OF PRODUCT INFORMATION FOR PEOPLE WITH VISUAL IMPAIRMENTS

A Thesis Submitted to the
Dokuz Eylül University, Department of Computer Engineering
In Partial Fulfillment of the Requirements for the Degree of B.Sc.

by

2019510102 İrfan Erdem ŞENYENER

Advisor
Asst. Prof.Dr.Yunus DOĞAN

November, 2022
Izmir

ACKNOWLEDGEMENT

We would like to take the opportunity to express our profound gratitude to our supervisor , Asst. Prof. Dr. Yunus DOĞAN,whose expertise understanding,useful information , practical advice and patience, added considerability to our gratitude experience and helped a lot in researching and writing this thesis.

We would like to express our special thanks to Dokuz Eylul University Computer Engineering Department for providing us with good education in the undergraduate program.

Thanks for all the encouragement!

İrfan Erdem ŞENYENER

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Based on the study of the World Health Organization (WHO), there were about 2.2 billion visually impaired people around the world in 2022, about 30% of the total population. Robust and efficient indoor object detection can help people with severe vision impairment to independently engage in any activity without help from other people.

Object detection is a computer technology related to computer vision and image processing that deals with the detection of instances of a priori objects of a certain class (such as faces, signs etc.) in digital images and videos captured by cameras. Detecting objects with a camera is challenging due to the following factors: 1) the wide variation in appearance and design (shape, color, texture, etc.) of objects in different places; and 2) large differences in camera view and image resolution of objects due to changes in position and distance between the blind user with wearable cameras and the target.

Object detection and recognition is a fundamental component for scene understanding. The human visual system is selective, robust, and fast . It is not only very selective, which allows us to distinguish among very similar objects, such as the faces of identical twins, but also robust enough to classify same category objects with large variances (e.g. changes of position,rotation, illumination, color and many other properties). Research shows that the human visual system can discriminate among at least tens of thousands of different object categories. However, it is extremely difficult to build robust and selective computer vision algorithms for object recognition which can handle very similar objects or objects with large variations. For instance, state-of-the-art object detection methods require hundreds or thousands of training examples and very long durations to learn visual models of one object category.

1.2 Problem Definition

With the growing food sector, a different product finds its place on the market shelves every day in the world. However, while the number of brands entering the sector is increasing day by day, marketing strategies have started to become more creative every day. As a result of this, products that are sometimes not understood by looking at the packaging began to be encountered on the shelves. While this is the case, it becomes even more difficult for visually impaired citizens to recognize the products.

With the advancement of technology, time has become more valuable for humanity, which can reach everything quickly. While the labor-force is decreasing in daily life and the use of technological tools has started to increase, it will be ensured that visually impaired citizens can solve their daily shopping without the help of sales consultants. This project is for the customer, who takes the product from the market shelves, by scanning the product with the help of the mobile application, and the product's name, brand, type, price, etc. enables them to learn their knowledge.

1.3 Motivation/Related Works

Visually impaired people do lead a normal life with their own style of doing things. However, they definitely face troubles due to inaccessible infrastructure and social challenges. One of the biggest challenges for a visually impaired person, especially the one with the complete loss of vision, is to navigate around places.Obviously, visually impaired people roam easily around their house without any help because they know the position of everything in the house. This is not the case for them when they are out of the house. Therefore, in this project, a project has been designed to make life easier in the markets where people spend most of their time during their life.

Many disability and assistive technologies have been developed to assist people who are visually impaired. The voice vision technology for the totally visual impaired pers offers sophisticated image-to-sound renderings by using a live camera.The Smith-Kettlewell Eye Research Institute developed a series of camera phone-based technological tools and methods for the understanding, assessment, and rehabilitation of blindness and visual impairment, such as text detection, crosswatch, and wayfinding.To help the visually impaired, Zandifar *et al.* used one head-mounted camera together with existing OCR techniques to detect and recognize text in the environment and then convert the text to speech.The Media Lab at the City College of New York has been developed a number of computer vision based technologies to help visual impaired people including banknote recognition, clothing pattern matching and recognition, text extract, and navigation and wayfinding.

Legally, equal accessibility to public places and services is now mandatory in many countries. Considering this situation in the Spring magazine published in 2008, technology was used to enable the visually impaired to benefit more from all the opportunities of the society. This volume describes the engineering and design principles and techniques used in assistive technology for the blind and visually impaired. Its features include; physiology of the human visual system and education of visual ability measurement methods; explaining many devices designed for everyday use in terms of general electrical engineering principles; There are hands-on project and research sections that will give the reader ideas for student study and self-education.

Rehabilitating Blind and Visually Impaired People: A psychological approach by Allan Dodds examines the influences of psychological factors on rehabilitation processes for visually impaired and blind people. Drawing on examples from a range of sensory and physical disabilities, this book emphasizes the importance of treating people individually, based on consideration of their psychological strengths and weaknesses as well as physical functioning. Written for workers with visually impaired people, this book is equally accessible to students and qualified workers, including rehabilitation workers, O & M specialists, occupational therapists, social workers and psychologists.

Despite many efforts, the question of how to apply this vision technology to help the visual impaired people understand their surroundings remains open.

1.4 Goal/Contribution

Today, visually impaired people, who can access products that make their lives easier in many areas with technology, have reached 39 million in the world population. Considering this number, besides helping them when they ask for help, they need to have an even more free life in which they can manage their lives without help. This project is to create a product that will make their lives easier in the markets they visit every day for their food and beverage needs. While they can get help from the sales consultants in the markets, many markets still do not write product information on the shelves with the visually impaired alphabet. While this is the case, this project that will guide them in every situation and everywhere will make a big difference in their lives.

1.5 Project Scope

Within the scope of this project, an application will be developed to facilitate the grocery shopping of the visually impaired people. In addition to the basic functions of the application, such as specifying the type of product, there are functions to determine the brand, expiration date and similar features.

This feature will be created using image processing algorithms. The step in which the feature defines the external information of the product will be scanned. As a result of scanning the product image, the internal information will be determined by means of the barcode. Afterwards, these features will be converted to audio files and returned to the customer.

1.6 Methodology/Tool/Libraries

In this project, classification algorithms will be used to create a model which will be suited best for image processing. Python and Learn libraries will be used to determine the fastest and the most efficient way to build a model.

Kotlin is expected to be used in an android application that is created for this project, according to functionality, Java can also be tested to determine the best language to use in application. Python will be used for the backend side.

Due to research it was found that most of the important information about a product is encoded in its barcode. This information will be decoded and turned into a more understandable state and presented to the user.

To convert text to speech, Python library gTTS or any similar library will be used in this project.

CHAPTER TWO

LITERATURE REVIEW

2. Related works

As part of research, many similar projects were examined. The following projects are the most significant for this project.

2.1 Lookout - Assisted Vision

"Lookout - Assisted Vision" is an application that uses computer vision to assist people with visual impairments. The application uses the phone camera to identify items, read a text for you, help with sorting mail or identify banknotes. It uses voice recognition and text to speech to communicate with the user. The functions identified previously are then selected and delivered by this method.

2.2 Aipoly Vision

"Aipoly Vision" is a product that uses image processing and a dataset, to identify objects in real time using the camera of the device. It can also read a text when detected, detect colors, plant or animal species and dishes around the world. Said dataset consists of user provided information of objects. Users can give an image to the application and enter keywords, mainly the name, and that image will contribute to the dataset. Currently that dataset consists of 5.000.000+ images and the application can identify up to 1000+ common objects, 900 dishes and 200 colors and 2000 plant or animal species.

2.3 Digit-Eyes

"Digit-Eyes" is an iOS application that helps visually impaired people to identify and recognise products when shopping. "Digit-Eyes reads manufacturer's UPC and EAN code and tells you the product name -- and often the full description, usage instructions and ingredients -- in 10 languages." Personal text and audio labels can be created to be read aloud by the iOS device.

As a part of research for building our model,a literature review was made. We shall briefly be discussing the review here with appropriate required details.

The first paper is titled "Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3 and was published in the year 2021. This paper uses algorithms and techniques like the OpenCV library, Yolo, Yolov3. The performance recorded in this paper indicates that it works better for smaller objects with future works mentioned as the expansion of the research on self explored dataset .

The next paper is titled "Reader and Object Detector for Blind" which was published in the year 2020. This paper uses algorithms and techniques such as Raspberry pi, OCR, tesseract, and tensorflow for carrying out the project. Text reading and object detection was successful but not for smaller than 16 font size is what was recorded in the performance of the paper. As the objective for future works, making it available for multiple languages is recorded as of now .

The paper that was studied next is titled "Obstacle Detection for Visually Impaired Patients" which was published in the year 2014. The techniques and algorithms used in this paper are stereoscopic sonar system, sound buzzers, voice IC- APR 9600. Wearable optical detection system is provided that provides full body vibration effect on obstacle detection. However, the device has a very limited range when compared to its own size and is also found difficult for users to comprehend the guidance signals in time .

The paper that was studied next is titled "Voice Based Smart Assistive Device for Visually Challenged" and was published in the year 2020. The Raspberry Pi, Deep Learning, conversational AI, speech recognition, Assistive Technology, and algorithms and methodologies were described in the article. After being trained on only 50 photos of each object, the model has an accuracy of 83 percent and detects campus objects that are commonly available. However, because it was trained on 8000 photos from the Flickr 8k dataset, the accuracy drops as the image complexity grows.

The next paper is titled "A Wearable Assistive Technology for the Visually Impaired with Door Knob Detection and Real-Time Feedback for Hand-to- Handle Manipulation" and was published in the year 2017. Algorithms and techniques such as YOLOv 2, Deep

Learning, Neural Network were used. The performance of the device is increased to folds if the hand detection is stable. The biggest difficulty, however, is the consistency of the hand detection performance. More images will be added to the database in the future, and the door knob identification feature will be extended to more general door handles.

The paper is titled as “VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision”, published in 2020. It is a wearable device based on Raspberry pi, gTTs and YOLO. The text will be read out in English and at a slow speed that is saved as an mp3 file and future work is mentioned as location navigation that works in low-light conditions while remaining cost-effective.

“YOLO-compact:An Efficient YOLO Network for Single Category Real-time Object Detection”, published in 2020, is an efficient way for a location navigation that works in low-light conditions while remaining cost-effective.. The model would be more precise if the depth, width and precision is improved.

A 2017 paper titled “A Novel YOLO-based Real-time People Counting Approach” is based on YOLO- PC and CNN. IT provides us an automatic way to count people in a huge crowd and the accuracy states that it works perfectly fine for a huge crowd and for a single person as well. Future Work stated in the paper is to add abnormal behavior detection during counting and children counting.

As reviewed in the next paper, it is“Edge detection based boundary box construction algorithm for improving the precision of object detection in YOLOv3” published in 2017. YOLOv3, Edge Detection, YOLO, YOLO9000, Boundary Box Prediction, and Object Detection are all used in this research. The intersection over union for the proposed algorithm and YOLO v3 is determined, and the proposed approach outperforms YOLO v3 in terms of boundary box accuracy. When there are sharp objects in the image or there is too much noise, the model becomes constrained.

The next paper that was reviewed is titled as “Design of Deep Learning Algorithm for IoT Application by Image based Recognition” by Jeena Jacob and P Ebby Darney. The primary objective of the paper is to gather comprehensive survey data on the types of IoT deployments, along with security and privacy challenges with good recognition rates. The other objective of the research of this paper is to achieve a high recognition rate for IoT-based image recognition.

The last paper that was studied is titled "Multi Distance Face Recognition of Eye Localization with Modified GaussianDerivative Filter" by Subarna Shakya. This paper talks about how from a distance, distinguishing faces with high accuracy is a task. Eye localization works for a human's face and eye placement in the face of a person of various sizes and distances. This method is used to recognize the face rapidly and with a high degree of accuracy.As a result, the suggested system has a face recognition accuracy of 93.24% on average.

Florian Schroff, Dmitry Kalenichenko and James Philbin in their paper titled "FaceNet: 'A Unified Embedding for Face Recognition and Clustering'" present a unified system for face verification, recognition and clustering. It is based on learning a Euclidean embedding per image using a deep convolutional network. The network is trained to ensure that the square of the distances in the embedding space correspond to face similarity. This means that the faces of the same person have small distances and faces of dissimilar people have large distances. After embedding, face verification involves thresholding the distance between the two embeddings; recognition becomes a classification problem; clustering can be achieved using techniques such as K-means clustering.

G. Iannizzotto, C. Costanzo, P. Lanzafame and F. La Rosa in their paper "Badge3D for Visually Impaired" has presented a visual support system that provides acoustic information about the objects in the surrounding environment using preset barcode tags and reading their information about what they represent or where they are located. Barcode tags deployed in the environment can act as reliable stimuli that trigger local navigation behaviors to achieve global navigation objectives. To achieve this they used a proof-of-concept multimodal, sensor based application and obtained experimental results.

The application introduced by Chen *et al.* is helpful in the detection of obstacles and contains glasses, a long stick cane, and a mobile application. The object can be detected through glasses and the long stick cane. If the user falls, the information is uploaded on the online platform and is displayed on the mobile device.

Pogi and Mattoccia have developed a wearable system for VIPs using deep learning and 3D vision based on the smartphone. The system uses an RGB camera to capture frames and CNN to detect obstacles. The system's performance is exceptional, i.e., near to 98% through the LeNet architecture. The system is lightweight, small in size, and performs detection in real time, but it cannot categorize most of the daily life objects. The user is guided through the tactile and audio feedback. The system is tested on more than 20 individuals having a visual impairment of different degrees. The result is very promising with good detection performance.

Chae *et al.* developed a system for detecting the collision through image segmentation. It detects obstacles that the existing solutions cannot be detected i.e., walls, doors etc. The input frames are segmented into superpixels through the lattice algorithm. For segmentation results, a Graph–based merging algorithm is used to detect both non textured and textured objects suc- cessfully. The system is tested on sighted/blindfolded people, dynamic/static objects. The experiment's result shows that the proposed system efficiently detects the object in real-time.

Salavati and Mahvash have introduced a system for detecting obstacles through a camera based on the Deep Neural Network (DNN). The system comprises unsupervised deep neural networks for extracting global features of the image, whereas, for extracting local attributes, a supervised DNN is used.

Nguyen *et al.* have proposed an electro tactile device connected with the tongue that helps the user navigate. The device is supported by an antenna that enhances wireless communication.

Xiao *et al.* have introduced an outdoor assistive navigation system controlled by the Rasp- berry Pi and GPS sensor, guiding navigation. The device has a list of the recordings. The user selects any sound from the list to receive the navigation. The device also has a speech recognition system to enhance its capabilities. The system is tested in New York City College to evaluate its performance. The result shows that every direction is followed properly using the developed device.

Bharambe *et al.* have presented a device that uses an ultrasonic sensor for navigation and direction. This device is based on GSM and GPRS. GPS is used to receive the specific location of a particular person. The device is tested on blindfolded persons to evaluate their per- formance.

Martinez-Sala *et al.* have developed a sugar navigation device for VIPs. The designed device is tested on blindfolded persons. The system uses ultra-wide techniques that provide accurate information for the location. Its accuracy is about 95%.

Aladren *et al.* have introduced a system based on the RGB-D sensor that guides navigation. The RGB-D is used for obtaining the range and the color information. The individual wears the device to evaluate the performance. The results show that the algorithm is robust and gives approximately 95% accuracy.

Yamashita *et al.* have introduced a wearable device that helps VIPs navigate using RFID, Quasi-Zenith system, and Hololens of Microsoft. The Hololens contain multiple cameras and sen- sors which help in positioning. Four experiments have been conducted against each individual to assess the system's performance. So, in total, 16 versions of different scenarios were attained. Out of these scenarios, users could move on the right path without any collision in thirteen situations.

Mancini *et al.* have offered a vision-based device that guides VIP in jogging, walking, and running. The system consists of gloves, a camera, and a board equipped with motors. The system recognizes the accurate path with speeds around 10 km/h using gloves.

Zhao *et al.* introduced a hybrid system utilizing CNN and features of an image to get first-person vision (FPV) navigation. The system takes a sequence of videos as input and gives output in a map showing the user's position. The process of this system is divided into two sub-processes, i.e., estimation of a location and calculation of a move- ment. An AlphaMex-based CNN is developed to recog- nize scenes in real-time, and SIFT tracking algorithms are proposed to analyze the movement. The IMU sensor easily influences the noise, and the complete navigation system depends on the single camera instead of using GPS or any other sensors.

Kunz *et al.* have introduced the idea of Virtual Navigation to aid VIPs in facing problems in mysterious locations through a real walk while staying in a precise environment. For achieving this, a user needs a camera on the head and brings the laptop to the back.

Zhang *et al.* developed a navigation system for VIPs using ARCore and provided safety and efficiency. The dual-channel mechanism that interacts with humans is introduced to deliver continuous and accurate direction that ensures path safety from risky areas and obstacle detection. The information is conveyed to the user through an audio and haptic interface.

Tianmei Guo et.el proposed CNN which has less computation cost and examined different methods and optimization algorithms for learning rate set and image classification parameters. Shallow network for good recognition effect has been verified .

CHAPTER THREE

REQUIREMENTS

- **3.1 Functional Requirements**
  - ***3.1.1 Barcode and QR scanner***
    - **Detecting and identifying a wide range of objects**
      The system should be able to detect and identify a wide range of objects, including people, animals, vehicles, buildings, and other common objects.
    - **Providing object descriptions**
      The system should be able to provide descriptions of detected objects, using text-to-speech or other output methods, so that the visually impaired person can understand the nature and characteristics of the objects in their environment.
    - **Scanning and decoding barcodes and QR codes**
      The application should be able to scan and decode barcodes and QR codes, and should be able to do so accurately and reliably.
    - **Providing audio feedback**
      The application should be able to provide audio feedback to the user, so that they can hear the information contained in the scanned code. This could include text-to-speech output of the code's content, or other types of audio cues.
    - **Allowing the user to adjust the volume of the audio feedback**
      The application should allow the user to adjust the volume of the audio feedback, so that they can hear the information clearly, even in noisy environments.

- **Providing tactile feedback**

  The application should be able to provide tactile feedback to the user, such as vibrations or haptic cues, to confirm successful scans and to provide additional information about the scanned code.

- **Allowing the user to customize the tactile feedback**

  The application should allow the user to customize the tactile feedback, so that they can tailor it to their preferences and needs.

- ***3.1.2 Image processing***

  - **Accurate and reliable object detection**

    It is important that the system be able to accurately and reliably detect objects in the environment,so that visually impaired people can navigate safely and effectively.

  - **Real-time operation**

    The system should be able to operate in real-time,so that visually impaired people can receive information about products they want to take.

- **3.2 Non-Functional Requirements**

  - **Accessibility**

    The system should be accessible to people with visual impairments, and should be easy to use for people who are not familiar with technology.

- **Performance**

  The system should be able to detect objects in real-time, and should not have significant delays in processing.
- **Reliability**

  The system should be reliable, and should not produce false positives or false negatives.
- **Security**

  The system should be secure, and should protect the privacy of users.
- **Scalability**

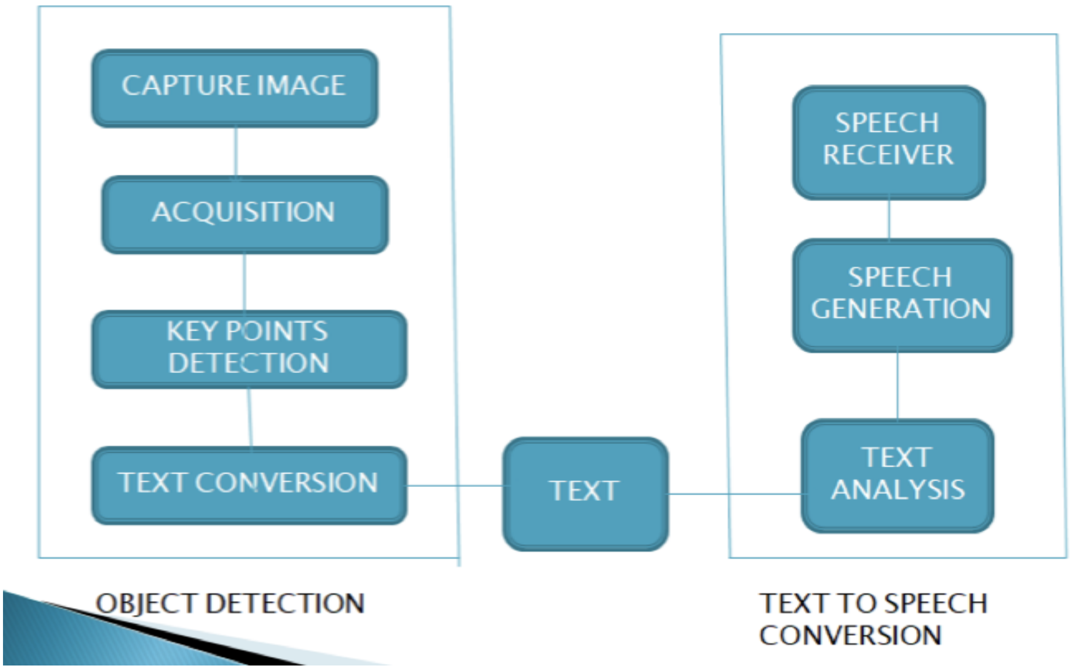
  The system should be able to handle a large number of users without experiencing performance issues.
- **Maintainability**

  The system should be easy to maintain and update over time.
- **Portability**

  The system should be portable, and should be able to run on a variety of devices.

CHAPTER FOUR

ARCHITECTURAL VIEW



4.1 Activity Diagram

In Figure 4.3 Activity Diagram is a part of the application for scanning product barcode or qr.

- Start: The process begins when the user launches the application.
- Wait for voice command: The application waits for the user to give a voice command to scan a barcode or qr code.
- Scan barcode: The user points the device's camera at the barcode and says the trigger word (e.g. "scan"). The application scans the barcode and retrieves the information it contains.
- Decode barcode: The application decodes the barcode and retrieves the information it contains.
- Look up product: The application searches a database or online resource to retrieve information about the product associated with the barcode.
- Give product information with tts: The application gives the product information to the user with text-to-speech. This may include the product name, price, manufacturer, and other details.
- Wait for voice response: The application returns to step 2 and waits for the user to give another voice command.
- End: The process ends when the user exits the application or performs another action.

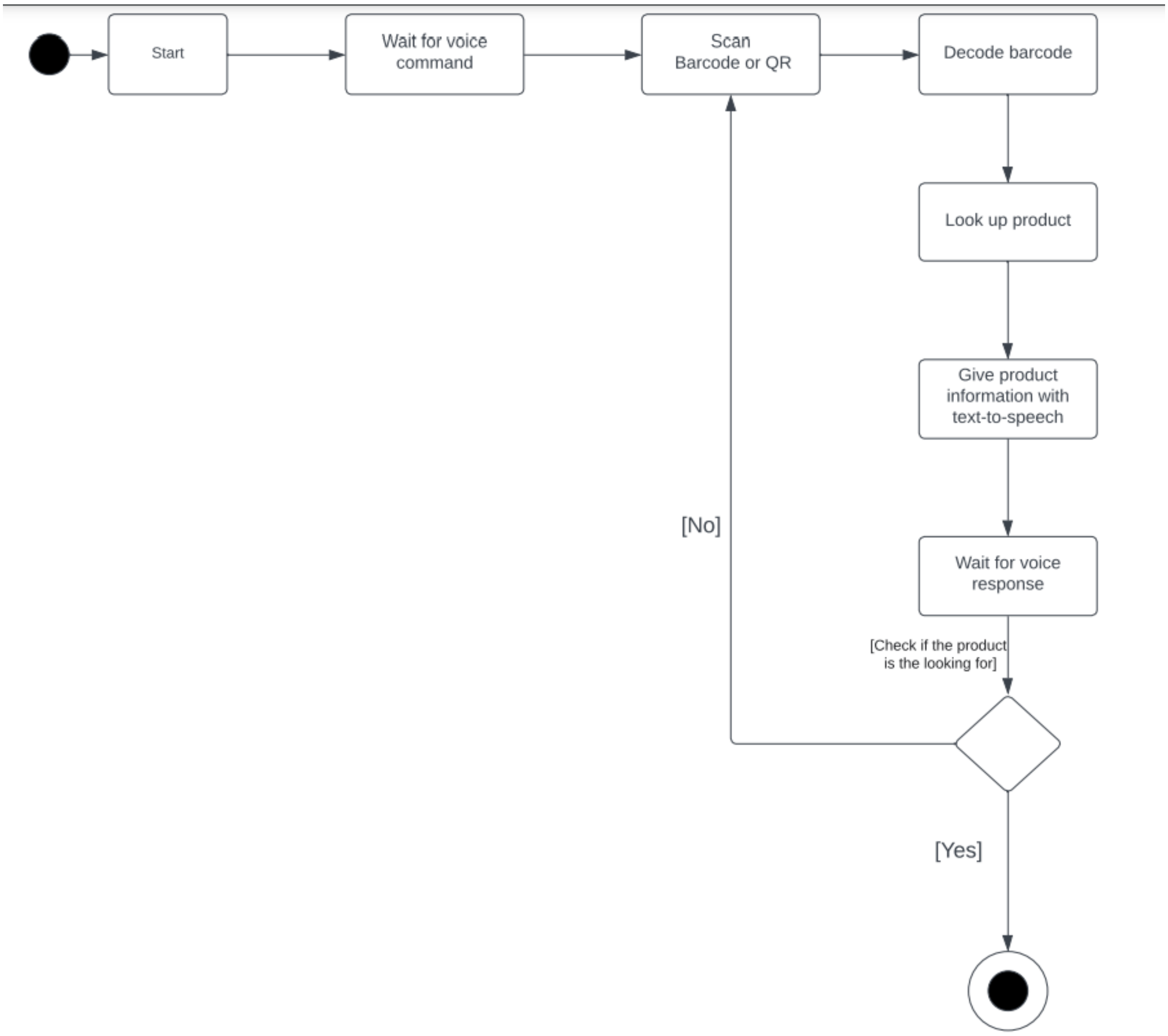


Figure 4.3 Activity Diagram

In Figure 4.4 Activity Diagram is a part of the application for scanning product barcode or qr.

This diagram shows the steps involved in using an image processing app for visually impaired people. The app is launched and the user selects an image. The image is pre-processed to prepare it for analysis, and then analyzed to extract information about the image. A description of the image is generated based on this analysis, and the description is output to the user. The process then ends.

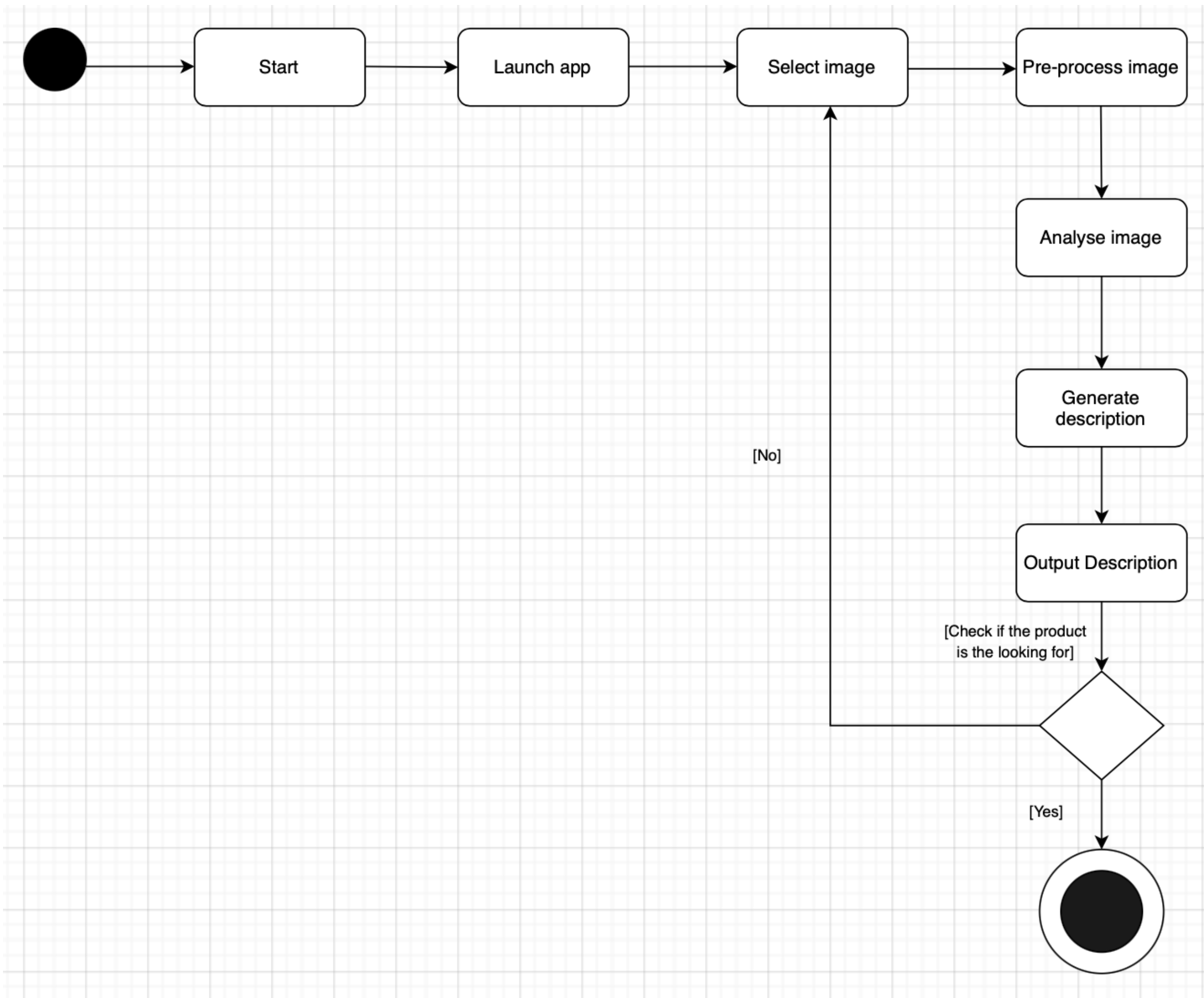


Figure 4.4 Activity Diagram

4.2 Class Diagram

The class diagram in Figure 4.6 showcases the class structure of the application.

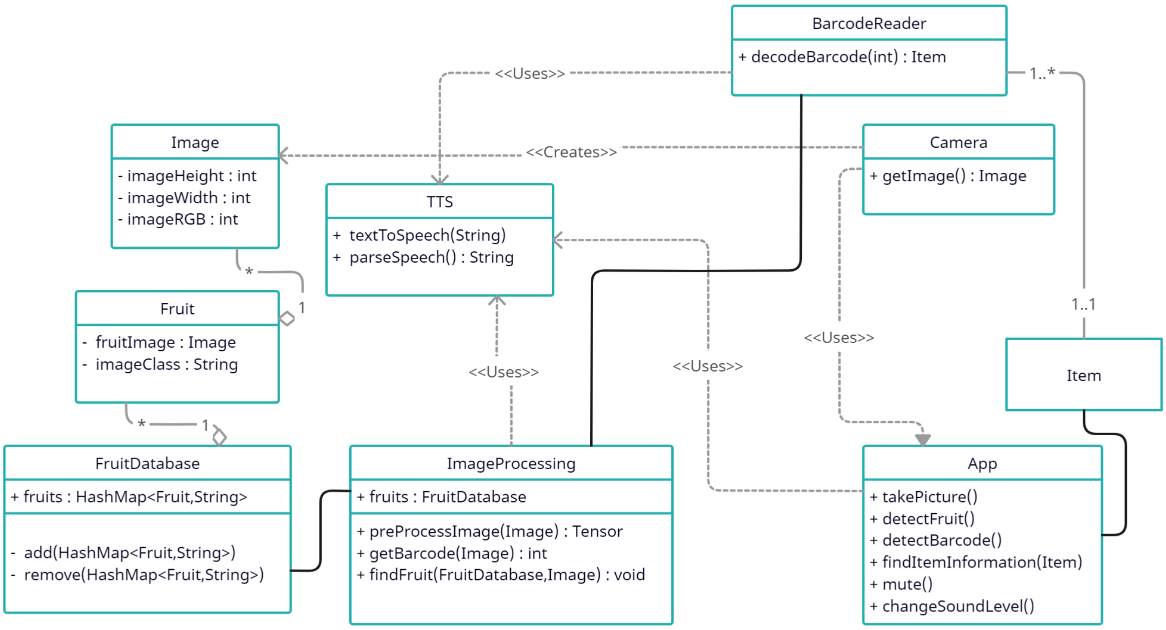


Figure 4.6 Class Diagram

4.3 Sequence Diagram

Sequence diagrams showcase the steps the application takes to execute a single command. In Figure 4.7, the user asks from the application to identify an item. The application then uses the corresponding classes to complete this task.

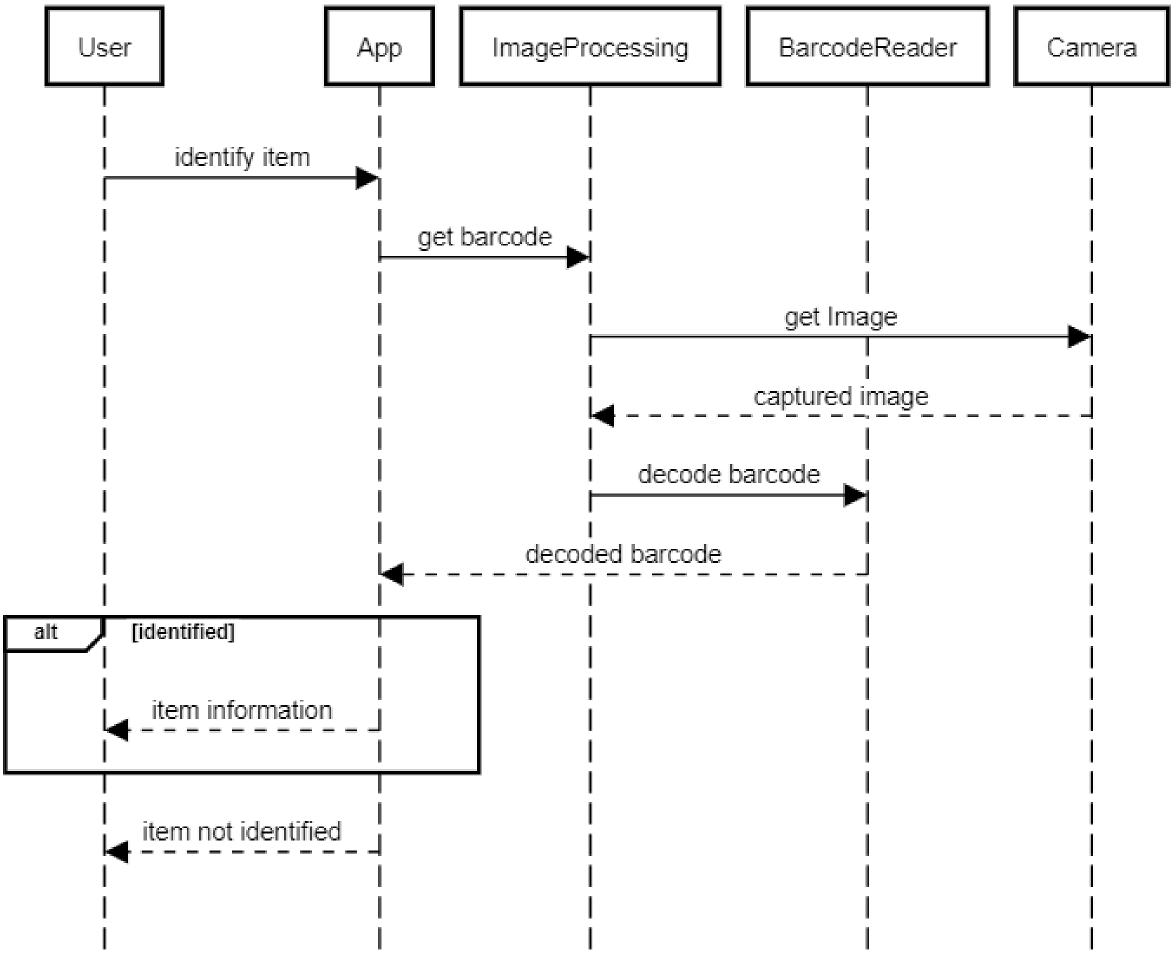


Figure 4.7 Sequence Diagram

In Figure 4.8, the user asks the application to identify a fruit. The application then uses the corresponding classes and a fruit database to identify the fruit.

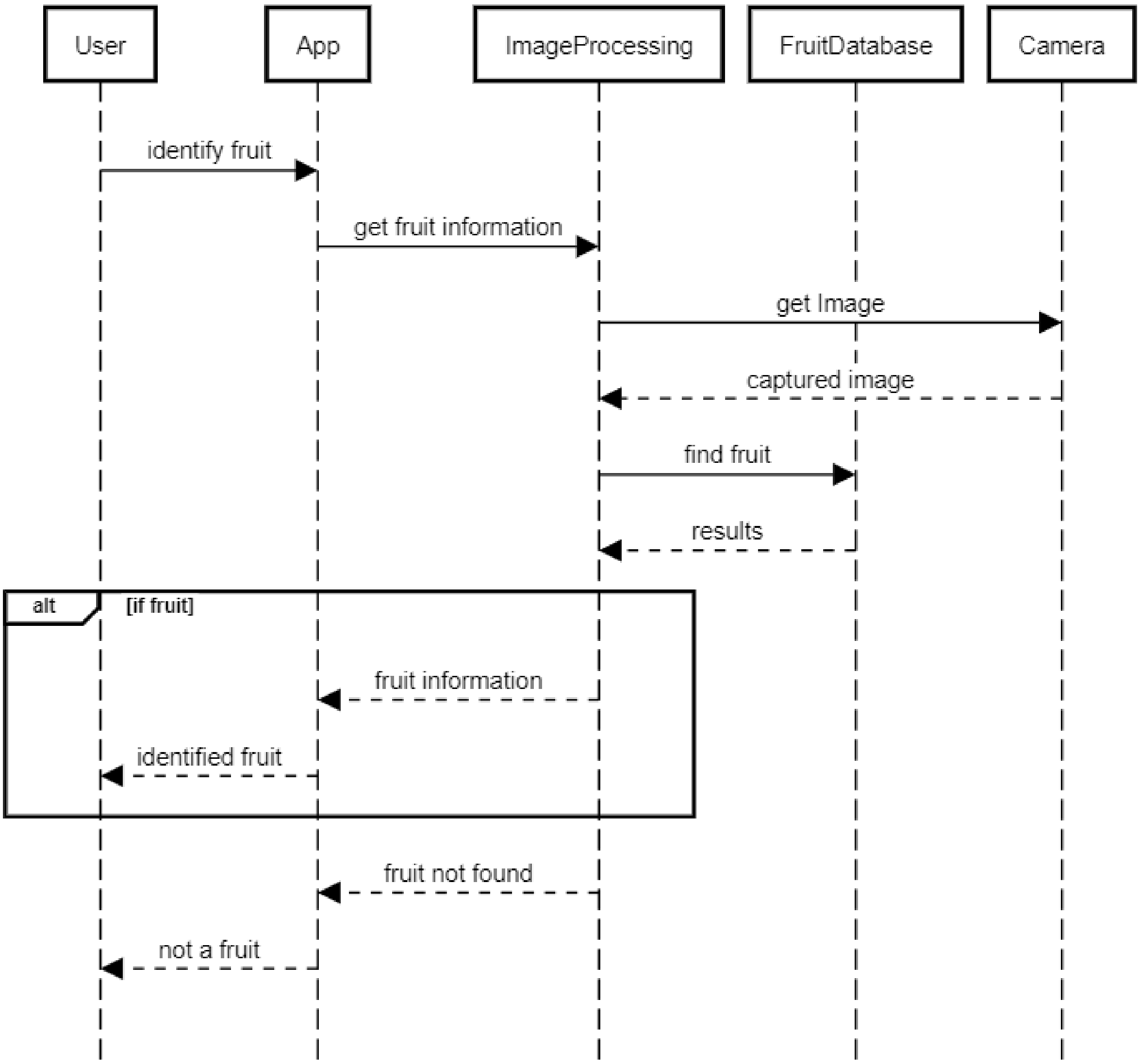


Figure 4.8 Sequence Diagram

CHAPTER FIVE

IMPLEMENTATION

5.1 Software Tools of Application

To establish the development environment for the Android application with real-time object detection, Android Studio needs to be downloaded and installed on the computer. Subsequently, an empty activity needs to be created in Android Studio after opening the application. The required dependencies , such as the TensorFlow Lite library for loading the pre-trained object detection model, must be added to the project.

The development of this application requires hardware and software tools and libraries. In the software part, an Android application was developed by working in the Android Studio environment using Kotlin. This method proved to be cost-effective and efficient in terms of time and efficient in terms of time and resources. To make the app easy to use and quick to operate,different software tools were added to enable the app to detect objects in real-time using the device's camera. The app can also draw rectangles around the detected objects to make them easily recognizable. In this way, the cost of the project was reduced, and time and money have been saved.

The objective of the application is to enhance time efficiency, hence to achieve this, various software tools and libraries have been utilized to develop a user-friendly interface.

To initiate the detection process, authorization for accessing the camera needs to be obtained, following which the frames are read from the camera. For displaying the frames, the desired user interface needs to be defined first. The Camera2 API has been implemented for extracting frames from the camera, while Tensorflow Lite has been utilized for detecting objects within these frames. Furthermore, to mark the detected objects on the frames, Canvas has been used for drawing rectangles around them.

The Camera2 API is a relatively new camera driver designed to operate on Android devices. It provides a range of features that can be leveraged for reading frames from the camera. The object detection model,integrated within the application, determines the presence of a known set of objects within an image or video stream, and provides relevant information regarding their position within the image.

After obtaining permission to access the camera, it is necessary to read the frames from the camera, therefore the first user interface defined in the desired user interface for displaying frames is presented. The Camera2 API was used to read frames from the camera, and Tensorflow Lite was used to detect objects within the frames. Canvas is used to draw a rectangle on the frame. The Camera2 API is a new camera driver that works on Android devices running version 5 or later. Given an image or video stream, an object detection model can determine which of a known set of objects may be present and provide information about their location within the image.

For instance, this screenshot of the sample application shows how the object is recognized and its location explained:

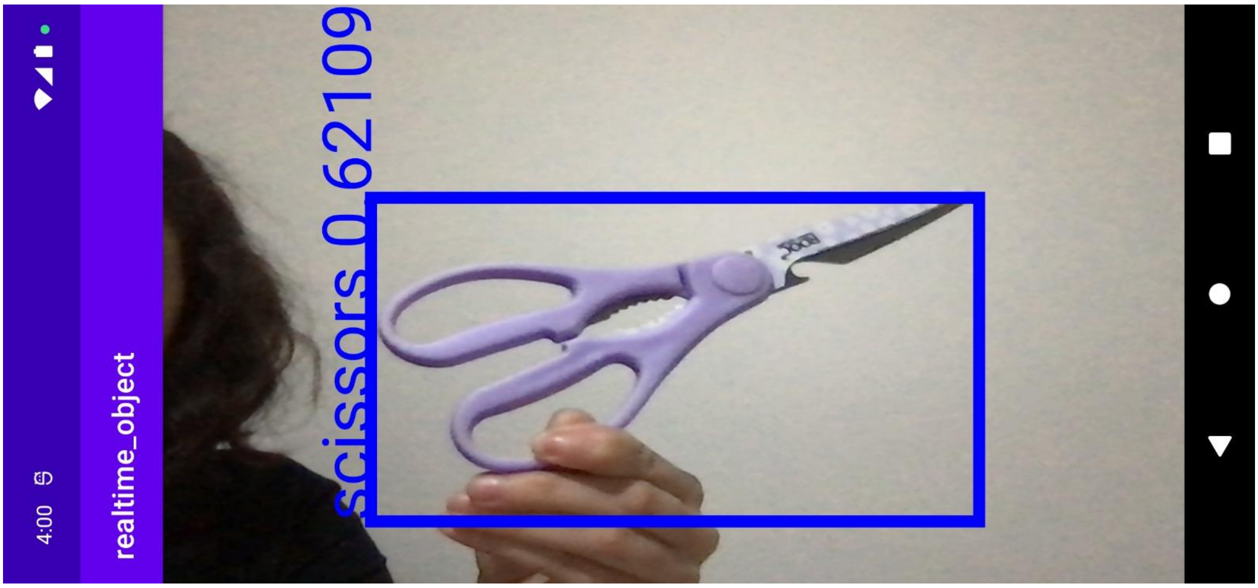


Figure 5.1 Scissors

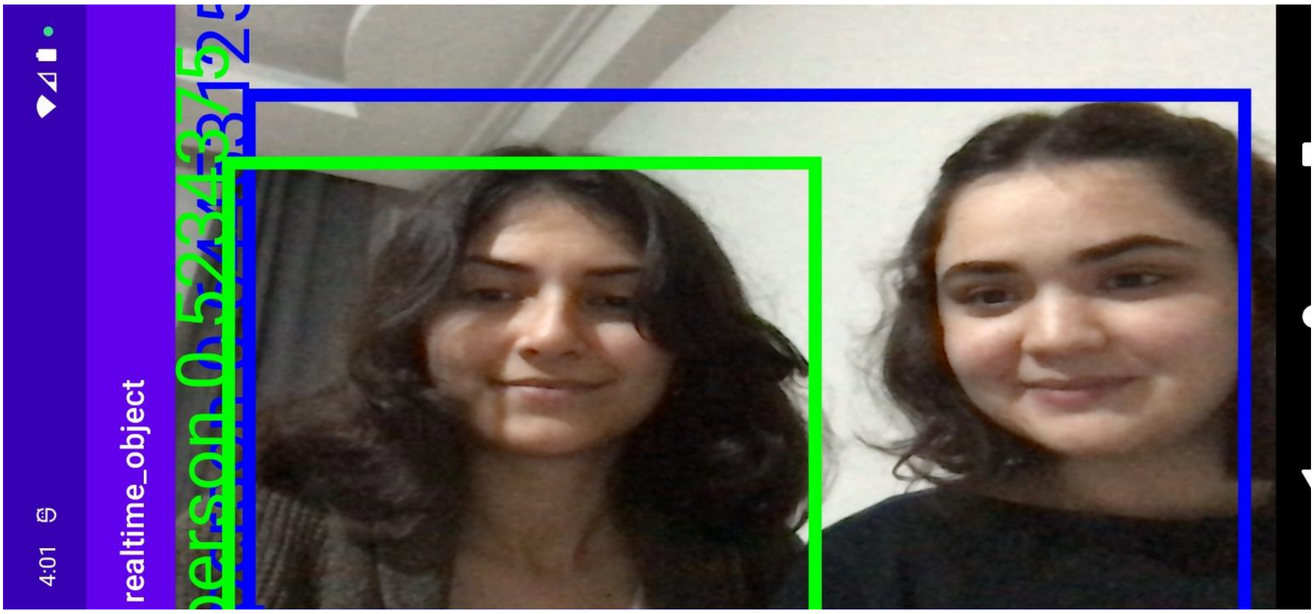


Figure 5.2 Persons

As can bee seen in Figure 5.1 and Figure 5.2, the application has utilized the TensorFlow Lite model, coupled with the Camera2 APIs, to read frames directly from the device’s camera. The TensorFlow mobile library has been employed to deploy these models on mobile devices, microcontrollers, and other edge devices. The library enables the selection of a new model or the retraining of an existing one. Additionally, this library is a symbolic math library based on data flow and differentiable programming, providing a flexi

Tensorflow light model plus camera two APIs are used to read frames directly in camera.TensorFlow Lite mobile library was used to deploy models on mobile, microcontrollers and other edge devices.It allows selecting a new model or retraining an existing model. Furthermore, this library is a symbolic math library based on data flow and differentiable programming.

The TextureView is the view object used to display the stream of video or frames in the application. TextureView is used to display a stream of content, such as from a camera preview, a video, or an OpenGL scene. OpenGL is an open-source graphics library used for rendering 2D vector graphics. OpenGL scene refers to a virtual environment created using OpenGL, different 2D objects and models can be rendered and displayed on a screen. It can be manipulated and interacted with in real-time, providing an immersive,interactive user experience.

The content flow can come from the application's process. The TextureView helps to display real-time video streams in the user interface of the application by using a fast and efficient solution called the hardware accelerated window. This is done by combining a view with a SurfaceTexture to enable the video stream to be shown in the application.

The texture view is the one which is used to show the stream of something the video or the frames. The texture view is used so that it will be the matching parent and the matching parent will give the texture view's id like this and it's back to the main activity. A TextureView is used to display a stream of content, such as from a camera preview, a video, or an OpenGL scene. The content flow can come from the application's process. The TextureView is used in the hardware accelerated window. The TextureView class is a view object that combines a view with a SurfaceTexture to enable the display in the app's user interface as shown in Figure 5.3.

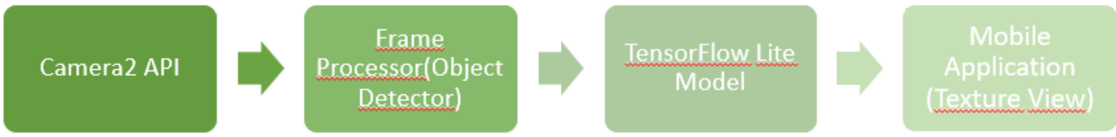


Figure 5.3 Api application

There also is a Keras model being developed at the same time. This model is being used to classify certain fruits and vegetables. Keras models are open-source machine learning models that are mostly used in image classification. This model uses CNN(Convolutional Neural Networks) which consists of several layers that can be summarized into two main layers: Feature Extraction and Classification.

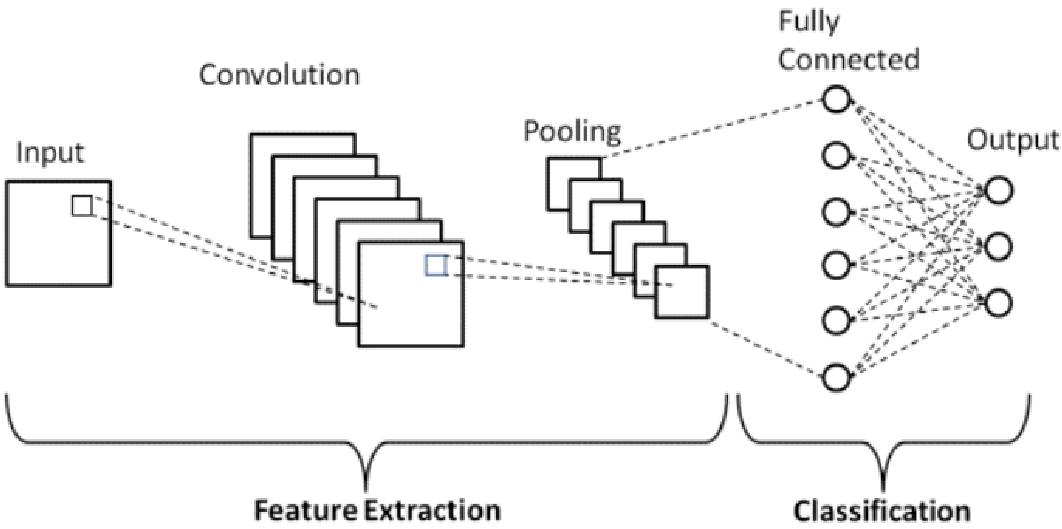


Figure 5.4 CNN Layers

In Figure 5.4, feature Extraction, the given input is preprocessed for the classification layer. This is done in several sub-layers inside the Feature Extraction layer in Figure 5.5.

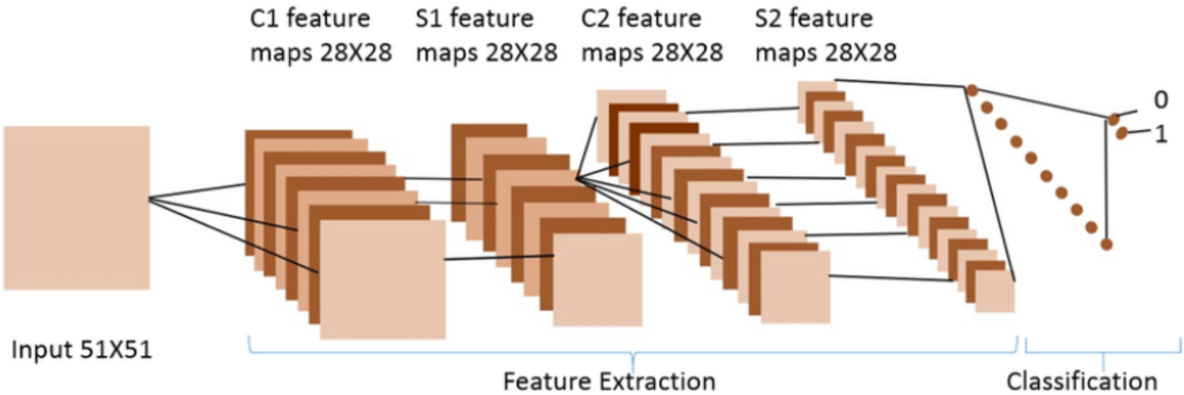


Figure 5.5 Feature Extraction Layers

To train the model; first, inputs are taken from a dataset, then reshaped and compressed into a smaller, significant set of data. Then the output is directed to the Classification layer.

The Classification layer makes predictions based on the output from the Feature Extraction layer. Every iteration of the model learns from the previous outcome to maximize the accuracy of the model.

This model is trained with a public dataset from Kaggle. Kaggle is an online website that is a repository to large quantities of datasets. The model is built using Python, with the Google Colab environment. Several libraries such as TensorFlow and aforementioned Keras, combined with Google Colab created an environment where this model can be made without Cloud knowledge.

When trained, the model is approximately %90 accurate, as shown in Figure 5.6 and Figure 5.7. This means that it is ready to be used. It can be used to find and classify fruits from a picture.

Figure 5.6 Predictions of the Model

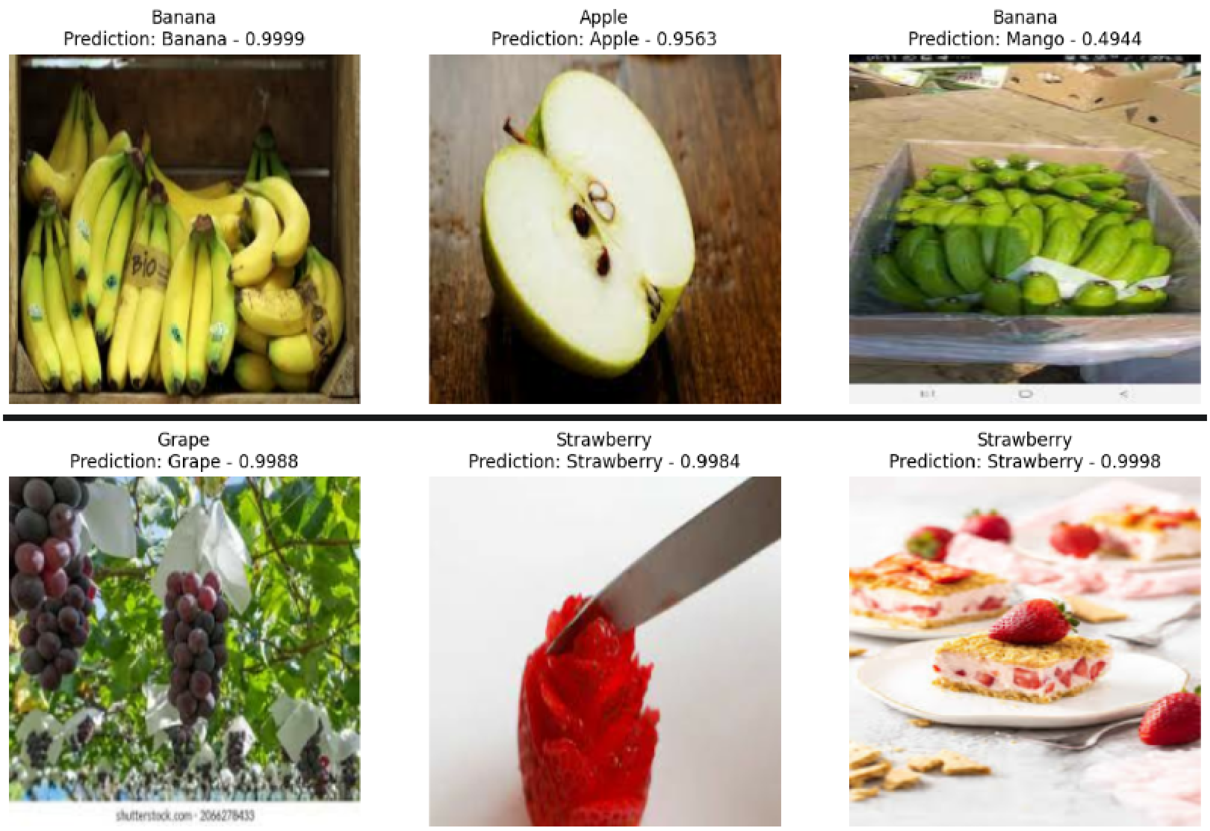


Figure 5.7 Predictions of the Model

The model is trained on a pre-trained TensorFlow model called MobileNet with Sequential layers that differ in density. Activation relu is used as activation for each layer. Image sizes are rescaled to 160x160 pixels and then RGB values are normalized into [0 1] range as shown in Figure 5.8 and Figure 5.9.

Figure 5.8 Example Picture

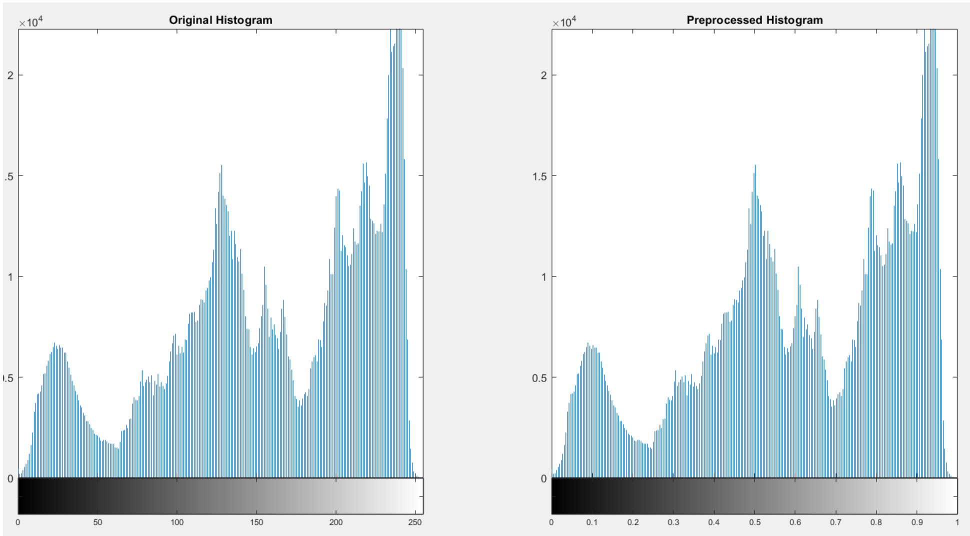


Figure 5.9 Original and After Normalization Histogram of Figure 5.8

The dataset consists of 5 different classes of fruits, which are: Appe, Banana, Mango, Grape and Strawberry. The dataset contains on average 1940 pictures for each fruit to train, and 20 pictures to test the model as shown in Figure 5.10.

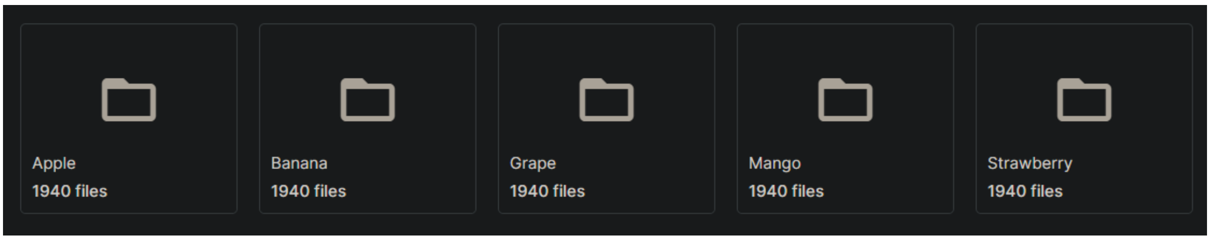


Figure 5.10 Structure of the Dataset

This model is then converted into a model.tflite file to be used in the application. Android Studio has built in dependencies for TensorFlow Lite that allows a Keras model to be used in an application.

There also is a QR and Barcode reader. To read Barcode and QR ,There are various libraries available that provide barcode detection functionality.And most optimal option for project was choosing **barcode detection library ZXing (**pronounced "zebra crossing")which is a widely-used barcode scanning library that supports multiple barcode formats such as QR codes, UPC codes, EAN codes, and more.

8 697447 970366

Figure 5.11 Example Barcode

According to our research, in Turkiye, the most commonly used barcode format is the EAN-13 (European Article Numbering) barcode. EAN-13 (Figure 5.11) is widely used for retail products and consists of a 13-digit number. The first few digits of the EAN-13 barcode represent the country or region code, followed by the manufacturer code, the numbers on the 6th right-most to the 2nd right-most represent the product code given by the manufacturer that is standardized in Turkiye .This barcode format is used for various purposes, including product identification, inventory management, and sales transactions .With these numbers one can identify and get information about the product using government issued manuals and in this case api's.

Next and main modules in this project are text-to-speech and image-to-speech conversion.

In the project, after detecting a certain object, it is written what that object is on the outside of the frame. Therefore output given as text.

Conversion takes place in several steps:

1. Get a text as input, which it first must analyze and then transform into a phonetic description.

2. Then it generates the prosody. From the information now available, it can produce a speech signal.

Natural Language Processing (NLP) module:

Produces a phonetic transcription of the text read, together with prosody.

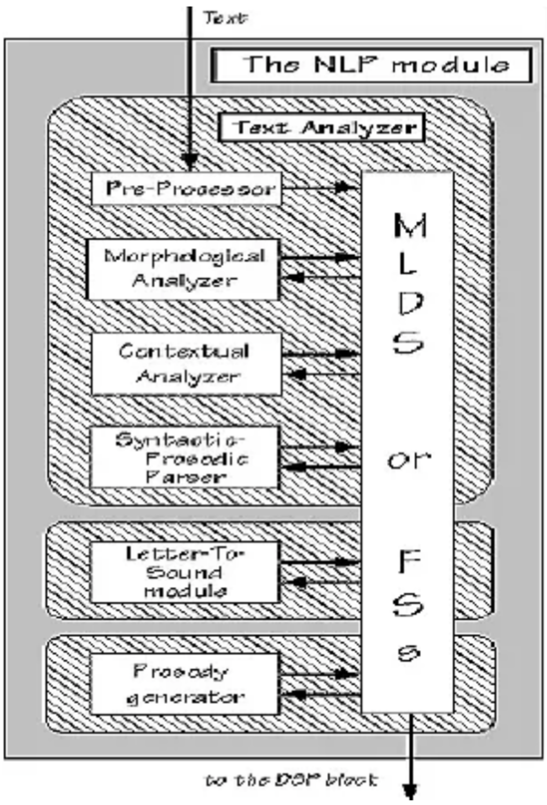


Figure 5.12: Operations of the natural Language processing module of a TTS synthesizer.

Figure 5.12 illustrates the operations performed by the Natural Language Processing (NLP) module within a Text-to-Speech (TTS) synthesizer. The NLP module is responsible for processing and analyzing input text to generate appropriate linguistic structures and representations for the synthesis process.

Digital Signal Processing (DSP) module:

It transforms the symbolic information it receives from NLP into audible and intelligible speech.

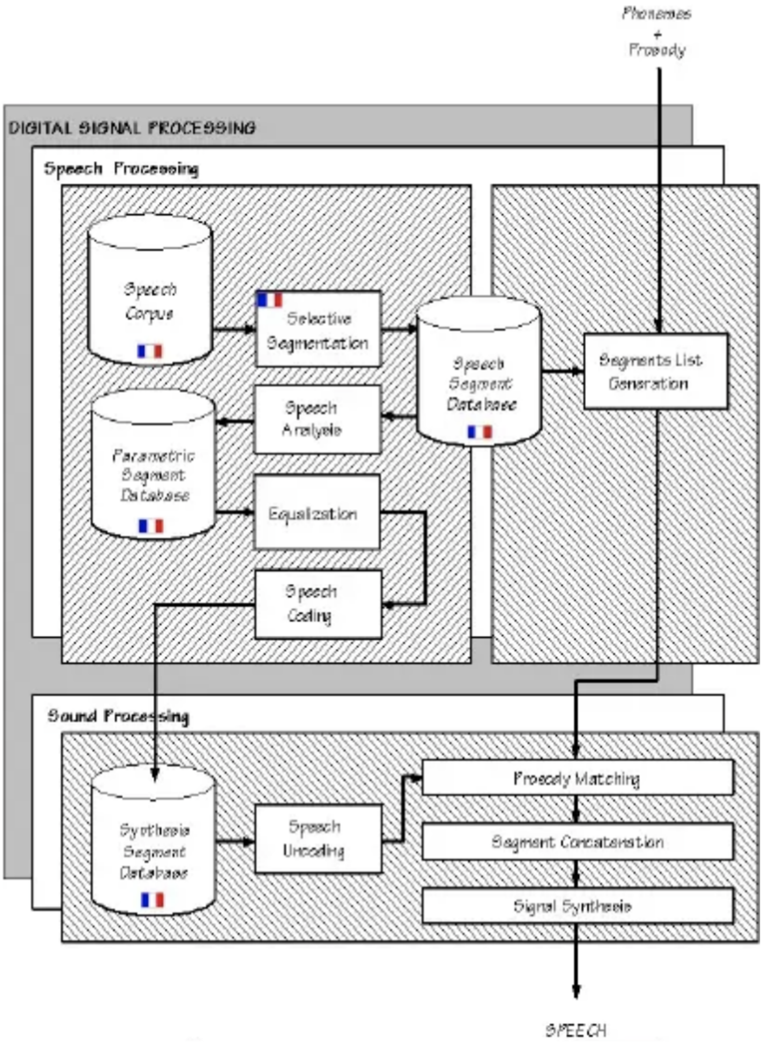


Figure 5.13: The DSP component of a general concatenation-based synthesizer.

Figure 5.13 shows the DSP component of a concatenation-based synthesizer. The DSP module is responsible for processing recorded speech units, such as phonemes or diphones, to produce natural-sounding synthesized speech. It performs tasks like signal processing, spectral analysis, time-domain manipulation, and waveform generation to ensure accurate speech synthesis.

Code of the conversion :

```kotlin
import androidx.appcompat.app.AppCompatActivity
import android.os.Bundle
import android.speech.tts.TextToSpeech
import android.util.Log
import android.widget.Button
import android.widget.EditText
import java.util.*

// Extending MainActivity TextToSpeech.OnInitListener class
class MainActivity : AppCompatActivity(),TextToSpeech.OnInitListener {
    private var tts: TextToSpeech? = null
    private var btnSpeak: Button? = null
    private var etSpeak: EditText? = null

    override fun onCreate(savedInstanceState: Bundle?) {

        super.onCreate(savedInstanceState)
        setContentView(R.layout.activity_main)

        // view binding button and edit text
        btnSpeak = findViewById(R.id.btn_speak)
        etSpeak = findViewById(R.id.et_input)

        btnSpeak!!.isEnabled = false

        // TextToSpeech(Context: this, OnInitListener: this)
        tts = TextToSpeech(this, this)

        btnSpeak!!.setOnClickListener { speakOut() }

        }
```

Figure 5.14:Code snippet for implementing Text-to-Speech functionality in an Android application

Figure 5.14 presents a code snippet demonstrating the implementation of Text-to-Speech (TTS) functionality in an Android application. The code utilizes the AndroidX library and extends the AppCompatActivity class. The TextToSpeech.OnInitListener interface is also implemented to handle the initialization of the TTS engine.

```
override fun onInit(status: Int) {
    if (status == TextToSpeech.SUCCESS) {
        val result = tts!!.setLanguage(Locale.US)

        if (result == TextToSpeech.LANG_MISSING_DATA || result == TextToSpeech.LANG_NOT_SUPPORTED) {
            Log.e("TTS","The Language not supported!")
        } else {
            btnSpeak!!.isEnabled = true
        }
    }
}
private fun speakOut() {
    val text = etSpeak!!.text.toString()
    tts!!.speak(text, TextToSpeech.QUEUE_FLUSH, null,"")
}

public override fun onDestroy() {
    // Shutdown TTS when
    // activity is destroyed
    if (tts != null) {
        tts!!.stop()
        tts!!.shutdown()
    }
    super.onDestroy()
}
```

Figure 5.15:Code snippet for Text-to-Speech functionality implementation and usage

Figure 5.15 presents a code snippet that demonstrates the implementation and usage of Text-to-Speech (TTS) functionality in an Android application. The code snippet includes functions for initializing the TTS engine, enabling speech synthesis, and shutting down the TTS engine when the activity is destroyed.

CHAPTER SIX

TEST AND EXPERIMENTS

6.1. Test Conditions

6.1.1. Image Processing Accuracy

Verifying that the application correctly identifies and recognizes the food items appearing in the camera view.

Test recognition accuracy with food types commonly encountered by visually impaired users.

6.1.2. User Interface Accessibility:

Checking that the user interface (UI) elements are designed with accessibility in mind.

Verifying that the application follows accessibility guidelines, such as providing descriptive labels for images and using appropriate color contrasts and font sizes.

6.1.3. Usability testing:

Observe visually impaired users interacting with other visually impaired apps to assess their overall usability.

Identify any areas of confusion, difficulty, or frustration during the observation.

Make the necessary adjustments and improvements to increase the usability of the application.

6.1.4. Voice Command Integration:

Test the integration of voice commands within the application.

Ensure that voice commands are accurately recognized and trigger the intended actions.

Figure 6.1:Testing application image processing part

Figure 6.1 illustrates the process of fruit scanning using image processing techniques within a mobile application.

6.2. Testing Phase

6.2.1. Unit Testing

Individual units of code,such as functions, methods and classes are tested separately.
The focus is on testing the logic and behavior of application components.

6.2.2. Integration Testing

The interactions and compatibility between different modules or components of the application are verified.
It tests how different parts of the application, such as image recognition algorithms, speech output, work together.

6.2.3. Functional Testing

The features and functions of the application are tested from the end user's perspective.

Verify that all features are working correctly, including image recognition, voice commands, and accessibility.

6.2.4. Accessibility Testing

Compliance with accessibility guidelines for visually impaired users is verified.

Effective navigation and interaction is provided for visually impaired users.

6.2.5. Performance Testing

It is tested for any lags, delays or crashes that may affect the user experience.

6.2.6. Regression Testing

Verify that existing functionality is not affected after changes are made.

Rerun relevant tests to ensure that changes haven't introduced new bugs.

CHAPTER SEVEN

CONCLUSION

In this project , for visually impaired people has been successfully developed, aiming to provide a user-friendly and accessible experience. The project utilized Android Studio as the development environment.

Through rigorous testing phases, including unit testing, integration testing, functional testing, accessibility testing , performance testing , regression testing , important results have been achieved.

The application demonstrates accurate image recognition capabilities, allowing visually impaired users to identify and recognise various food items using the device's camera. The integration of speech output functionality ensures clear and concise delivery of information to users. Accessibility features have been thoroughly tested to ensure compatibility with assistive technologies such as screen readers, enabling easy navigation and interaction for visually impaired users.

Performance testing was conducted to optimize the application's use of resources and ensure smooth operation across different devices and network conditions. Usability testing with visually impaired users provided valuable feedback to improve the overall user experience and address areas of confusion and difficulty.

The project report highlights the successful implementation of the image processing market application to meet the specific needs of visually impaired users. The iterative development process, driven by user feedback and rigorous testing, has resulted in a robust and accessible application.

Nevertheless, it is important to note that further improvements and enhancements can still be made. Future development efforts could focus on expanding the range of recognised objects, improving voice command integration, and incorporating additional accessibility features to accommodate a wider range of disabilities.

Overall, this image processing market Android mobile application project contributes to the goal of empowering visually impaired people by providing them with a tool to access and navigate the world of visual information. The successful development and testing of this application lays the foundation for future advances in assistive technologies and contributes to a more inclusive digital environment for visually impaired people.

REFERENCES

Afif, M., Ayachi, R., Said, Y., Pissaloux, E., & Atri, M. (2020). An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation. *Neural Processing Letters*, *51*(3), 2265-2279.

Ajala, J. (2021). Object Detection and Recognition Using YOLO: Detect and Recognize URL (s) in an Image Scene.

Aladren, A., López-Nicolás, G., Puig, L., & Guerrero, J. J. (2014). Navigation assistance for the visually impaired using RGB-D sensor with range expansion. *IEEE Systems Journal*, *10*(3), 922-932.

Annapoorani, A., Kumar, N. S., & Vidhya, V. (2021). Blind-Sight: Object Detection with Voice Feedback. *International Journal of Scientific Research & Engineering Trends*, *7*(2), 644-648.

Aralikatti, A., Appalla, J., Kushal, S., Naveen, G. S., Lokesh, S., & Jayasri, B. S. (2020, December). Real-time object detection and face recognition system to assist the visually impaired. In *Journal of Physics: Conference Series* (Vol. 1706, No. 1, p. 012149). IOP Publishing.

Ashiq, F., Asif, M., Ahmad, M. B., Zafar, S., Masood, K., Mahmood, T., ... & Lee, I. H. (2022). CNN-based object recognition and tracking system to assist visually impaired people. *IEEE Access*, *10*, 14819-14834.

Bharambe, S., Thakker, R., Patil, H., & Bhurchandi, K. M. (2013, April). Substitute eyes for blind with navigator using android. In *2013 Texas Instruments India Educators' Conference* (pp. 38-43). IEEE.

Chae, S. H., Kang, M. C., Sun, J. Y., Kim, B. S., & Ko, S. J. (2017). Collision detection method using image segmentation for the visually impaired. *IEEE Transactions on Consumer Electronics*, *63*(4), 392-400.

Chen, L. B., Su, J. P., Chen, M. C., Chang, W. J., Yang, C. H., & Sie, C. Y. (2019, January). An implementation of an intelligent assistance system for visually impaired/blind people. In *2019 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1-2). IEEE.

Coughlan, J., & Manduchi, R. (2009). Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. *International Journal on Artificial Intelligence Tools*, *18*(03), 379-397.

Dewangan, R. K., & Chaubey, S. (2021). Object Detection System with Voice Output using Python.

Dodds, A. (2013). *Rehabilitating blind and visually impaired people: A psychological approach*. Springer.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Guo, T., Dong, J., Li, H., & Gao, Y. (2017, March). Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (pp. 721-724). IEEE.

Gupta, S., Chakraborti, S., Yogitha, R., & Mathivanan, G. (2022, May). Object Detection with Audio Comments Using YOLO V3. In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 903-909). IEEE

Hersh, M. A. (2008). *Assistive technology for visually impaired and blind people* (Vol. 1). M. A. Johnson (Ed.). London: Springer.

Iannizzotto, G., Costanzo, C., Lanzafame, P., & La Rosa, F. (2005, September). Badge3D for visually impaired. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops* (pp. 29-29). IEEE.

Isewon, I., Oyelade, O. J., & Oladipupo, O. O. (2012). Design and implementation of text to speech conversion for visually impaired people. *International Journal of Applied Information Systems*, *7*(2), 26-30.

Karmarkar, M. R. R., & Honmane, V. N. Object Detection System For The Blind With Voice Guidance. *International Journal of Engineering Applied Sciences and Technology*, *6*(2), 2455-2143.(2021)

Kaggle, “Fruit Classification”,
https://www.kaggle.com/datasets/utkarshsaxenadn/fruits-classification

Kunz, A., Miesenberger, K., Zeng, L., & Weber, G. (2018, July). Virtual navigation environment for blind and low vision people. In *International Conference on Computers Helping People with Special Needs* (pp. 114-122). Springer, Cham.

Lakde, C. K., & Prasad, P. S. (2015, April). Navigation system for visually impaired people. In 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC) (pp. 0093-0098). IEEE.

Mahendru, M., & Dubey, S. K. (2021, January). Real time object detection with audio feedback using Yolo vs. Yolo_v3. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 734-740). IEEE.

Mancini, A., Frontoni, E., & Zingaretti, P. (2018). Mechatronic system to help visually impaired users during walking and running. *IEEE transactions on intelligent transportation systems*, *19*(2), 649-660.

Martinez-Sala, A. S., Losilla, F., Sánchez-Aarnoutse, J. C., & García-Haro, J. (2015). Design, implementation and evaluation of an indoor navigation system for visually impaired people. *Sensors*, *15*(12), 32168-32187.

Murali, M., Sharma, S., & Nagansure, N. (2020, July). Reader and Object Detector for Blind. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0795-0798). IEEE.

Nguyen, T. H., Le, T. L., Tran, T. T. H., Vuillerme, N., & Vuong, T. P. (2013, April). Antenna Design for Tongue electrotactile assistive device for the blind and visually-impaired. In *2013 7th European Conference on Antennas and Propagation (EuCAP)* (pp. 1183-1186). IEEE.

Papageorgiou, C.P., Oren, M., & Poggio, T. (1998, January). A general framework for object detection.In Sixth International Conference on Computer Vision(IEEECat.No.98CH36271)(pp.555-562).IEEE.

Patel, P., & Bhavsar, B. (2021). Object Detection and Identification. *International Journal*, *10*(3).

Poggi, M., & Mattoccia, S. (2016, June). A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning. In *2016 IEEE symposium on computers and communication (ISCC)* (pp. 208-213). IEEE.

Potdar, K., Pai, C. D., & Akolkar, S. (2018). A convolutional neural network based live object recognition system as blind aid. arXiv preprint arXiv:1811.10399..

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

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 779-788).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Salavati, P., & Mohammadi, H. M. (2018, October). Obstacle detection using GoogleNet. In *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 326-332). IEEE.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

Shah, S., Bandariya, J., Jain, G., Ghevariya, M., & Dastoor, S. (2019, April). CNN based auto-assistance system as a boon for directing visually impaired person. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 235-240). IEEE.s

Shaikh, S., Karale, V., & Tawde, G. (2020). Assistive Object Recognition System for Visually Impaired. *International Journal of Engineering Research & Technology (IJERT)*, *9*(9), 736-740.

Srikanteswara, R., Reddy, M. C., Himateja, M., & Kumar, K. M. (2022). Object Detection and Voice Guidance for the Visually Impaired Using a Smart App. In *Recent Advances in Artificial Intelligence and Data Engineering* (pp. 133-144). Springer, Singapore.

Xiao, J., Ramdath, K., Losilevish, M., Sigh, D., & Tsakas, A. (2013, June). A low cost outdoor assistive navigation system for blind people. In *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)* (pp. 828-833). IEEE.

Yamashita, A., Sato, K., Sato, S., & Matsubayashi, K. (2017, December). Pedestrian navigation system for visually impaired people using HoloLens and RFID. In *2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI)* (pp. 130-135). IEEE..

Zhang, X., Yao, X., Zhu, Y., & Hu, F. (2019). An ARCore based user centric assistive navigation system for visually impaired people. *Applied Sciences*, *9*(5), 989.

Zhao, Q., Zhang, B., Lyu, S., Zhang, H., Sun, D., Li, G., & Feng, W. (2018). A CNN-SIFT hybrid pedestrian navigation method based on first-person vision. *Remote Sensing*, *10*(8), 1229.