



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Intelligence Eye for Blinds and Visually Impaired by Using Region-Based Convolutional Neural Network (R-CNN)

Lee Ruoy Yee^a, Hazalila Kamaludin^{a,*}, Noor Zuraidin Mohd Safar^a, Norfaradilla Wahid^a, Noryusliza Abdullah^a, Dwiny Meidelfi^b

^a Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia

^b Department of Information Technology, Politeknik Negeri Padang, West Sumatera, Indonesia

Corresponding author: *hazalila@uthm.edu.my

Abstract— Intelligence Eye is an Android based mobile application developed to help blind and visually impaired users to detect light and objects. Intelligence Eye used Region-based Convolutional Neural Networks (R-CNN) to recognize objects in the object recognition module and a vibration feedback is provided according to the light value in the light detection module. A voice guidance is provided in the application to guide the users and announce the result of the object recognition. TensorFlow Lite is used to train the neural network model for object recognition in conjunction with extensible markup language (XML) and Java in Android Studio for the programming language. For future works, improvements can be made to enhance the functionality of the Intelligence Eye application by increasing the object detection capacity in the object recognition module, add menu settings for vibration intensity in light detection module and support multiple languages for the voice guidance.

Keywords— Mobile application; Android; light detection; object recognition.

Manuscript received 10 Feb. 2021; revised 12 Apr. 2021; accepted 19 Oct. 2021. Date of publication 31 Dec. 2021. International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

The World Health Organization (WHO) estimates that there are 285 million visually impaired people worldwide [1]. Most mobile applications are designed for people with normal vision [2]. It is difficult for blind people to use mobile devices by using the accessibility features. The most versatile applications are not designed for the visually impaired, so they create a lot of confusion [3]. Blind and visually impaired people face several problems in their life. One of the most difficult tasks is identification of things that are useful in their daily life [4] for example accessing and using e-book applications as they must be able to navigate and use e-books in a manner equal to sighted people [5]. They also cannot detect light by themselves as these cannot be done just by touching, feeling, or smelling. They need to get the information on screen by hearing.

This Intelligence Eye application combined light detection and object recognition function with 80 object model classes for blind and visually impaired people. Blind and visually impaired people are the main users of this application. All the functions in this application are specially designed for their

ease. There are three modules included in Intelligence Eye: main menu, light detection module and object recognition module. Main menu module allows the users to select the functions by using swipe gesture and pressing the buttons on screen. There are two buttons in the main menu module which are light detection button and object recognition button. Users can also choose the function by swiping up and down on screen. A voice guidance is provided to guide users to select the functions.

Light detection module allows the users to detect the environment light by using the ambient light sensor of the smartphone and the system vibrates in different patterns according to the light value. When the light value is low, the vibration is slow. When the light value is high, the vibration is fast. When the light value is 0, the vibration stops. Object recognition module allows the users to identify and recognize an object by capturing the objects with live preview in real-time from the back camera of the smartphone. COCO (Common Object in Context) dataset that contains 80 object model classes is used as the object model class [6]. The recognition result is announced to users by using the Text-to-Speech (TTS) engine.

By using Intelligence Eye, blind and visually impaired people can identify and recognize the objects, detect the environment light without help from others. The application guides them by using voice. Users can use swipe gestures to control the application. This reduces their time and effort for learning to use the application. Following are terms and works related to Intelligence Eye:

1) *Blind*: According to Cambridge Learner's Dictionary, blind is defined as unable to see [7]. In Oxford Student's Dictionary, blind refers to lacking the sense of sight. Blindness is defined as the state or condition of being unable to see because of injury, disease, or a congenital condition [8].

2) *Visual Impairment*: Visual impairment, otherwise called vision impairment or vision loss, is a diminished capacity to see to a degree that causes issues not fixable by usual means, for example, glasses [9]. Visual impairment is usually defined as the best-corrected visual acuity worse than 20/40 or 20/60 [10].

3) *Artificial Intelligence*: Based on Cambridge Advanced Learner's Dictionary [7], intelligence refers to the ability to learn, understand and make judgments or have opinions that are based on reason. Artificial Intelligence (AI) is defined as the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages [11].

Machine learning systems can automatically develop information shaping systems, rendering them more appealing than manually creating the software [12]. Machine learning is an implementation of Artificial Intelligence which allows machines the ability to automatically learn and improve on knowledge without being directly programmed. Machine learning focuses on the creation of computer programs that can access data and know for themselves.

TensorFlow is a large-scale and heterogeneous running machine learning system [13]. TensorFlow Lite is used for mobile applications. It is a lightweight framework intended to run prototypes on mobile or embedded devices. It can be used in Android Studio to create an application with Artificial Intelligence. It consists of an API for users to plug into their application [14].

4) *Object Recognition*: Object recognition is a computer technology related to computer vision and image processing which identifies instances in digital images and videos of semantic objects of a certain type [15]. It is a technical discipline that looks for ways to simplify all the work a human visual system can do.

5) *Object Recognition Methods*: Object recognition methods are usually either machine-based approaches or deep learning approaches. For machine learning approaches, support vector machine (SVM) is used to do the classification. Deep learning approaches that are capable of detecting end-to-end artefacts without identifying specific features and are usually based on convolutional neural networks (CNNs) [16].

Neural network is a model of computation commonly used in machine learning to solve different tasks. Perceptron, the primitive neuron model, is a two-layer structure [17]. Feedforward networks such as Multi-layer Perceptron and

Radial Basis Function networks are the most widely used family of neural networks for pattern recognition [18]. Neural network performance increases the number of invisible layers by increasing them to some degree.

6) *Region-Based Convolutional Neural Networks (R-CNN)*: R-CNN combines rectangular region proposals with convolutional neural network features. It is a two-stage detection algorithm. In an image that could include an entity, the first stage defines a subset of regions. In each area, the second stage classifies the artefact [16]. R-CNN has three variants to optimize, speed up or enhance the results of these processes [16].

7) *Ambient Light Sensor*: Ambient light sensor is a component in the smartphone that detects the nearby amount of light. It is used to provide data on the brightness of the light in the region. This data is typically used to adjust the ambient screen brightness control to save the battery from the monitor while maximizing the visibility [19]. It can detect the light by the reading of light intensity.

8) *Light Intensity*: Light intensity relates to the frequency or quantity of light produced by a single source of light [19]. It is the estimate of the wavelength-weighted energy produced by a light source. There are different light levels for different environments. The light level is also known as illuminance. Lux is used to measuring the illuminance level.

9) *Review on Existing System*: Three existing systems which are TapTapSee [20], Seeing Assistant Home Lite [21] and Seeing AI [22] are selected. All the systems have their own strengths and weaknesses.

TapTapSee from Net Ideas, LLC is a free Android and iOS image recognition application for blind and visually impaired people with a very simple interface [20]. To take a photo, the back camera needs to be aimed at the object and double-tap the screen. The photo is automatically uploaded to the servers and when a match is found, it speaks the item name. Its user interface is clear and easy to use, and it is a good application for the blind and visually impaired people.

Seeing Assistant Home Lite is a community of assistants developed for everyday tasks on iOS and Android apps for mobile devices [21]. It is a mobile application with many features including light and colour detection, barcode scanner, and barcode reader [23]. It mainly uses the back camera of the devices to perform its assigned activities, but it frequently provides inconsistent data, particularly when in light detection mode.

Seeing AI [22] is a free mobile application designed for the low vision community to recognize people, text, currency, colour, and objects. It is a Microsoft research project that combines cloud and AI together to create an intuitive device designed to help the sight-impaired to navigate their day. Many useful functions are provided in this application. Users may be confused sometimes as too many features in the application increase the chance of mistakes.

10) *Comparison of the System*: Three existing systems are chosen to be compared with the proposed system, Intelligence Eye. The three systems are TapTapSee [20], Seeing Assistant Home Lite [21] and Seeing AI [22]. Table 1 shows the

comparison between the three existing systems and the Intelligence Eye on their characteristics and functions.

The rest of the paper is organized as follows: Section 2 presents the material and method. Section 3 presents the result and discussion of Intelligence Eye. Finally, Section 4 concludes the work, limitations and highlights the future direction for research of Intelligence Eye.

TABLE I
COMPARISON OF THE SYSTEM

| Characteristics Comparison | TapTapSee | Seeing Assistant | Seeing AI | Intelligence Eye |
|----------------------------|----------------------|--|--|--|
| Hardware Used | Camera of Smartphone | Camera of Smartphone, Light Sensor of Smartphone | Camera of Smartphone, Light Sensor of Smartphone | Camera of Smartphone, Light Sensor of Smartphone |
| Android Based | Yes | Yes | No | Yes |
| Audio Guidance | No | Yes | Yes | Yes |
| Swipe gestures | No | No | No | Yes |
| Light Detection | No | Yes | Yes | Yes |
| Object Recognition | Yes | Yes | Yes | Yes |

II. MATERIAL AND METHOD

A. Object-Oriented Software Development (OOSD)

Object-oriented software development (OOSD) approach is used in software development as a framework for assuring the software to meet defined needs [24]. There are six phases contained in the object-oriented life cycle which are object-oriented planning, object-oriented analysis, object-oriented design, object-oriented implementation, object-oriented testing and object-oriented maintenance.

1) *Object-Oriented Planning Phase*: In the object-oriented planning phase, the problem statement, objectives, project scope, expected results and project significance are identified. The work plan of developing the application is modified in this phase to ensure it can be completed within the timeframe. A Gantt chart is created for the Intelligence Eye application. Research studies on similar information and knowledge from online resources and journal articles are undertaken. Comparison between the existing applications and the proposed application is conducted to analyse functions provided in each application and the limitations of existing applications.

2) *Object-Oriented Analysis Phase*: In this phase, the established requirements are gathered and collected. Hardware and software requirements are analysed for the development of the application. Unified Modelling Language (UML) diagrams are used to visually represent the relationship and interaction between the classes [25]. The yEd Graph Editor is used to create the UML diagrams.

For hardware requirements, the computer model used is Lenovo Legion Y520 with Central Processing Unit (CPU) Intel Core i5-7300HQ, Graphic Processing Unit (GPU) NVIDIA GEFORCE GTX 1050 2GB, Random Access Memory (RAM) 8GB and runs in Window 10 operating system. There are six software that have been used to develop, design, operate and maintain Intelligence Eye which includes

Microsoft Word 2019, Microsoft Project 2019, Android Studio, TensorFlow Lite, Balsamiq Mockup 3 and yEd Graph Editor.

3) *Object-Oriented Design Phase*: Based on the requirement specifications stated in the previous phase, the design of Intelligence Eye is prepared in this phase. The prototype is created to define the information architecture and content of the application. A neural network model is designed by using TensorFlow Lite in this process to fulfil the functionality of object recognition.

4) *Object-Oriented Implementation Phase*: In this phase, units, which are the small parts of an application, are developed according to the interface design. A unit test is done for every unit in the initial testing stage. The neural network model is plugged into the Android Studio for object recognition. Accessibility functions such as voice guidance by using TTS is implemented to the application. Then, the units are combined to the system to develop the complete application.

5) *Object-Oriented Testing Phase*: In this phase, functional and non-functional requirements are tested to make sure that the completed application meets the established requirements during the analysis phase. To debug and improve the user experience, functional testing is carried out and correction is made once the bugs and issues are detected. The Intelligence Eye application is ready to be implemented at the end of the phase once the bug and issues are fixed.

6) *Object-Oriented Maintenance Phase*: In this phase, the developed application will be maintained and updated frequently. This will take three months to complete. However, this phase is not included in this project due to the time limitation.

B. System Analysis

1) *System Architecture Diagram*: In Intelligence Eye, through user interface, users can choose to use light detection function or object recognition function by using swipe gestures or pressing buttons. In the light detection function, it will request permission to use the ambient light sensor and vibration sensor of the device. TTS engine is used to instruct users to use the light detection function. The ambient light sensor responses to the system for the light intensity reading. Vibrator sensor responses to the vibration feedback according to the light intensity reading. In object recognition function, it requests training data and the training data responses to the system when the user wants to recognize an object. The TTS engine is required for the voice guidance and the announcement of result in object recognition function.

C. Requirement Analysis

1) *Functional Requirement*: Functional requirements are product features or functions to be implemented by developers to enable users to perform their tasks. For the light detection function, it allows the user to detect the brightness of environment light. A vibration feedback with different patterns according to the light intensity. For object recognition function, it allows the user to recognize the objects in real time by capturing images from the camera of

the device. The result is announced after the recognition process.

2) *Non-Functional Requirement:* A non-functional requirement is a measure that lays out parameters that can be used to determine the functionality of a process rather than specific behaviours. For usability requirements, users are able to learn and operate the application since it is user friendly to blind and visually impaired users. For performance, the response time is reasonable when operating the application. Intelligence Eye also must be available at any time. The maintenance features must be kept updated and repaired to improve the performance of the application.

D. Unified Modelling Language (UML)

1) *Use Case Diagram:* There are a total of nine use cases for this application which includes choose function, voice guidance, swipe gestures, get light value, voice guidance, vibration feedback, capture image, recognize object in scene and speak result. Users are allowed to choose a function by using swipe gestures and pressing the buttons on the main menu page of Intelligence Eye. Voice guidance is provided in the main menu to guide the user. In the light detection function, an ambient light sensor is used to get the light value. Vibrator sensor is used to vibrate in different patterns according to the light value to alert the users. Voice guidance is provided to guide users for light detection. In object recognition function, the user is able to capture an image and the neural network model recognizes the object in the scene. After the recognition process, the TTS engine is used to announce the result.

2) *Sequence Diagram:* In Intelligence Eye, users can choose a light detection function from the system interface. Once users choose to detect the light, ambient light sensor and vibration sensor are initiated and the value of light intensity is read. The light value is sent back to the system and the system will vibrate in different patterns according to the light value. Users can choose an object recognition function from the system interface. Users are able to capture images with the camera of the device. Once an image is captured, the neural network model recognizes the object in the scene by comparing it to the trained data. The result is then converted to text and announced by the TTS engine.

3) *Class Diagram:* There are six classes in Intelligence Eye which includes MainActivity, LightSensor, LightDetect, Object, ObjTrain and CameraActivity.

4) *Activity Diagram:* Intelligence Eye starts with a main menu. Users need to choose the function from the menu by pressing the buttons or using swipe gestures to continue using the application. There are two functions that can be chosen in the application which are light detection and object recognition. If users choose light detection, the application goes to light detection activity. The light detection protocol is started. If the "Start" button is pressed, then the application starts to detect the light. If the "Stop" button is not pressed, the application vibrates in different patterns according to the value of light intensity until the "Stop" button is pressed. If users choose object recognition, the application invokes the camera module. Image is captured and the neural network

model is invoked to recognize objects in the scene. The result is then announced by the TTS engine.

E. System Design

1) *Prototype:* Generally, a prototype is used to evaluate a new design to improve system analysts and users' precision. Figure 1 shows the splash screen of Intelligence Eye while Figure 2 shows the main menu of Intelligence Eye. Light detection interface as shown in Figure 3 while object recognition interface as shown in Figure 4.



Fig. 1 Splash screen



Fig. 3 Main menu



Fig. 2 Light detection interface



Fig. 4 Object recognition interface

III. RESULT AND DISCUSSION

A. Implementation of Intelligence Eye Application

Intelligence Eye is developed using Android Studio which is a software that is specially designed for Android development. Java and XML programming language is used to develop the application in Android Studio. XML is a programming language that is used to design the layout for the user interface. While Java programming language is used to compile and execute the module by using Java Development Kit (JDK).

1) *Text-to-Speech Engine:* Text-to-Speech engine is implemented in all activity classes for the voice guidance and to speak out the result of a recognized object.

2) *onSwipeTouchListener() Class:* onSwipeTouchListener() class is implemented in Menu activity and LightDetector activity for the swipe gestures function.

3) *Splash Screen Activity:* Splash screen activity is a module that acts as a start-up screen which appears when the application is launched. The logo of the application is animated, and a welcome message is presented by the TTS engine to welcome the user. It redirects to the main menu screen after 3 seconds.

4) *Menu Activity:* Menu activity is the first page shown to the user when the splash screen of Intelligence Eye is redirected. It consists of two buttons that redirect to light detection activity and object detection activity. Voice guidance is spoken to guide users to control the application.

Users can choose the function by pressing buttons or using the swipe gesture.

5) *LightDetector Activity*: LightDetector activity is the activity for light detection module of Intelligence Eye application which is to detect the environment light and give a vibration feedback to the user according to the light value. TTS engine is implemented in this activity to guide the user to swipe up for starting the detection and swipe down on screen to stop the detection. OnSwipeTouchListener() class is implemented in this activity to detect the swipe gestures of the user.

6) *Detector Activity*: Detector activity is the activity for object recognition module of Intelligence Eye application which is used to detect the objects around the blind users and give a verbal feedback for the detection results. TTS engine is implemented in this activity to speak out the name of the objects. COCO dataset which has 80 object classes is used as the model to detect objects in Intelligence Eye. This convolutional neural network is used in conjunction with an API model by TensorFlow Lite to transfer the images to it for inference.

B. Functional Testing

1) *Test Plan*: Test plans are developed to test whether an Intelligence Eye application meets its requirements. Test plans for each module of Intelligence Eye are tested and all the results are as expected.

2) *Functional Testing for Object Recognition Module*: A functional test is carried out to test on the number of objects detected by the object detection modules. A total of 80 objects are used to test for the object detection module. Figure 5 shows the results of animals that can be detected by the object recognition module. Figure 6 and Figure 7 show the results of common objects that can be detected. Figure 8 shows the food that can be detected while Figure 9 shows the kitchen utensils that can be detected by object recognition module. Figure 10 shows the results of sport equipment while Figure 11 shows the outdoor objects that can be detected. Figure 12 shows the common objects that can be detected by the object recognition module.



Fig. 5 Result of animal detected



Fig. 6 Result of common objects detected-1

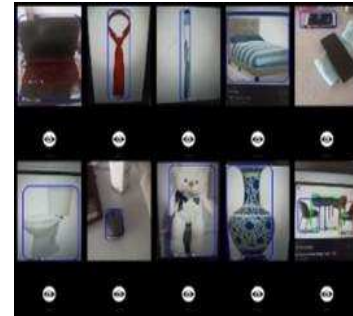


Fig. 7 Result of common objects detected-2



Fig. 8 Result of food detected



Fig. 9 Result of kitchen utensils detected



Fig. 10 Result of sport equipment detected



Fig. 11 Result of outdoor objects detected

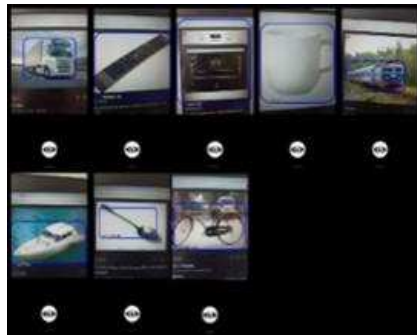


Fig. 12 Result of common objects detected-3

IV. CONCLUSION

Intelligence Eye application has been developed successfully and achieved the objectives which provide a user-friendly and ease to use application to visually impaired users. There are still some limitations on this application. In future work, more languages should be added and supported by the application to provide services to the users who do not understand English. A setting menu for vibration should be added in the light detection module to customize the vibration intensity. The detection capacity for the object recognition module should be increased to allow users to detect more objects. The performance of the application should be improved to guarantee a better user experience.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education (MOHE) Malaysia for supporting this research under Fundamental Research Grant Scheme Vot No. K216 Reference Code FRGS/1/2019/ICT04/UTHM/03/2 and partially sponsored by Universiti Tun Hussein Onn Malaysia.

REFERENCES

- [1] World Health Organization, "Visual Impairment and Blindness 2010," World Heal. Organ., 2012.
- [2] J. S. Sierra and J. S. R. De Togoires, "Designing mobile apps for visually impaired and blind users: Using touch screen based mobile devices: iPhone/iPad," in ACHI 2012 - 5th International Conference on Advances in Computer-Human Interactions, 2012.
- [3] E. Chen, Y. Lin, C. H. Chen, and I. F. Wang, "BlindNavi: A navigation app for the visually impaired smartphone user.," in Conference on Human Factors in Computing Systems Proceedings., 2015.
- [4] R. Tapu, B. Mocanu, A. Bursuc, and T. Zaharia, "A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People," in Proceedings of the IEEE International Conference on Computer Vision, 2013.
- [5] N. Hashim, M. Saleh Ba Matraf and A. Hussain, "Identifying the Requirements of Visually Impaired Users for Accessible Mobile Ebook Applications", JOIV : International Journal on Informatics Visualization, vol. 5, no. 2, pp. 99-104, 2021.
- [6] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, and J. Hays, "Microsoft COCO: Common Objects in Context," 2015.
- [7] C. McIntosh, Cambridge Advanced Learner's Dictionary, 5th ed. Cambridge, United Kingdom: Cambridge University Press, 2013.
- [8] Oxford, Oxford Student's Dictionary, 8th ed. Cambridge, United Kingdom: Oxford University Press, 2016.
- [9] World Health Organization, "Change the Definition of Blindness," World Heal. Organ., 2015.
- [10] H. Hollands and J. C., "The Prevalence of Low Vision and Blindness in Canada," Eye 20, pp. 341–346, 2016.
- [11] M. Deuter and J. Bradbery, Oxford Advanced Learner's Dictionary. Oxford, United Kingdom: Oxford University Press, 2014.
- [12] P. Domingos, "A Few Useful Things to Know about Machine Learning," 2016.
- [13] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning.," 2016.
- [14] "TensorFlow Lite." <https://www.tensorflow.org/lite> (accessed Oct. 05, 2019).
- [15] A. Uçar, Y. Demir, and C. Güzelış, "Object Recognition and Detection with Deep Learning for Autonomous Driving Applications," 2017.
- [16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014.
- [17] S. Asht and R. Dass, "Pattern Recognition Techniques: A Review," 2012.
- [18] H. Wu, Y. Zhou, Q. Luo, and M. A. Basset, "Training Feedforward Neural Networks Using Symbiotic Organisms Search Algorithm," Comput. Intell. Neurosci., 2016.
- [19] K. Dhondge, B. Choi, S. Song, and H. Park, "Optical Wireless Authentication for Smart Devices Using an Onboard Ambient Light Sensor.," in 23rd International Conference on Computer Communication and Networks (ICCCN), 2014.
- [20] "TapTapSee." <https://taptapseeapp.com/> (accessed Oct. 05, 2019).
- [21] "Seeing Assistant Home Lite." <http://seeingassistant.tt.com.pl/> (accessed Oct. 05, 2019).
- [22] "Microsoft." <https://www.microsoft.com/en-us/ai/seeing-ai> (accessed Oct. 05, 2019).
- [23] B. Leporini and M. Buzzi, "Home Automation for an Independent Living: Investigating the Needs of Visually Impaired People," 2018.
- [24] S. K. Dora and P. . Dubey, "Software Development Life Cycle (SDLC) Analytical Comparison and Survey on Traditional and Agile Methodology," Natl. Mon. Ref. J. Res. Sci. Technol., 2013.
- [25] F. Alhumaidan, "A Critical Analysis and Treatment of Important UML Diagrams Enhancing Modeling Power," 2012.