e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

OBJECT DETECTION SYSTEM FOR BLIND PEOPLE USING DEEP LEARNING (BLIND ASSISTANCE SYSTEM)

Prof. Shila Pawar[*1], Omkar Walke[*2], Kunal Rupnar[*3], Onkar Sable[*4], Gaurav Sonawane[*5]

[*1]Asst. Prof., Shri Chhatrapati Shivaji Maharaj College Of Engineering, India.

[*2,3,4,5]Student, Shri Chhatrapati Shivaji Maharaj College Of Engineering, India.

ABSTRACT

Object recognition, a key advancement in computer vision and machine learning, plays a crucial role in identifying and pinpointing objects within an image or video. Through object recognition, we detect individual items and accurately obtain details about them, such as their dimensions, form, and position. This study introduced an affordable assistive system designed for obstacle recognition and environmental representation to support blind individuals using deep learning methods. The TensorFlow object recognition API and SSDLite MobileNetV2 were utilized to develop the proposed recognition model. The pre-trained SSDLite MobileNetV2 model was trained on the COCO dataset, containing nearly 328,000 images across 90 distinct object categories. The gradient particle swarm optimization (PSO) algorithm was applied in this work to fine-tune the final layers and their respective hyperparameters of the MobileNetV2 model. Subsequently, the Google text-to-speech API, PyAudio, playsound, and speech recognition were incorporated to produce audio feedback of the detected items. A Raspberry Pi camera captures real-time video, where frame-by-frame object detection is performed with a Raspberry Pi 4B microcontroller. The suggested device is embedded into a head-mounted unit, intended to assist visually impaired individuals by detecting obstacles in their path, offering a more effective solution compared to a conventional white cane. In addition to the object detection model, a secondary computer vision model, named “ambiance mode”, was trained. In this mode, the final three convolutional layers of SSDLite MobileNetV2 were retrained via transfer learning on a weather-related dataset. This dataset includes approximately 500 images across four weather categories: cloudy, rainy, foggy, and sunrise. In ambiance mode, the system provides a detailed narration of the surrounding environment, resembling how a human might describe a landscape or a stunning sunset to someone with visual impairment. The performance of both the object detection and ambiance description functions was evaluated on both desktop and Raspberry Pi systems. The model's accuracy was assessed using metrics like mean average precision, frame rate, confusion matrix, and ROC curve. This affordable system is expected to be a valuable tool for enhancing the daily lives of visually impaired individuals.

Keywords: Deep Learning, Assistive Technologies, Mean Average Precision, Object Detection, Random Forest Classifier, SSD Mobilenet, Text-To-Speech.

I. INTRODUCTION

Visually impaired or blind individuals face extreme challenges in their daily lives due to significantly diminished vision. Some of these individuals rely on various assistive technologies, such as braille displays, augmented reality-based smart glasses, and screen reader software. According to estimates from the World Health Organization (WHO), more than 2.25 billion people globally and approximately 0.8 million people in Bangladesh are affected by this condition. This number is expected to double by 2022 and triple within the next 30 years. In densely populated countries in Asia or Africa, it becomes particularly difficult for blind individuals to navigate from one place to another using only a traditional white cane. While the white cane offers basic navigational aid by detecting obstacles, it fails to provide blind users with a more comprehensive understanding of their surroundings with greater accuracy. This challenging situation motivated the development of an electronic device with object detection and auditory feedback, aimed at increasing the independence of visually impaired individuals. This project seeks to improve navigation assistance and enhance environmental awareness for blind individuals.

In recent years, there has been significant research aimed at creating intelligent systems that ensure the safe and autonomous movement of visually impaired individuals. These assistive techniques typically incorporate artificial intelligence and advanced smartphone applications. For example, in [4], the authors presented a smart

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

eye Android app that can detect light, color, different objects, and banknotes for visually impaired users. The app uses the built-in light sensor of a smartphone for light detection, the OpenCV library for distinguishing colors, and a locally stored database for identifying objects. The application was developed using Android Studio with Java, and it incorporates CraftAR toolbox for image recognition and CNNdroid, an open-source library for running Convolutional Neural Networks (CNNs) on Android devices. A built-in voice engine reads the detection results aloud, enabling users to hear them. proposed a system to detect physical movement and kitchen activities using machine learning to assist blind individuals. This system applied scale-invariant feature transform (SIFT) to extract features from a dataset, and the random forest model, with optimized hyper parameters, achieved an accuracy of 0.808. Mukhiddinov and his team used an enhanced version of the YOLOv4 deep learning model to identify the freshness of fruits and vegetables. Their system classifies five categories of fruits and vegetables using a dataset collected under varying lighting conditions. The improved YOLOv4 model achieved 69.2% AP50 on test samples. They also developed a smartphone app that instantly detects rotten fruits and vegetables.

Most assistive devices consist of various hardware components with embedded circuits and sensors. For instance, an assistive device proposed in used Raspberry Pi and TensorFlow's object detection API to help blind people better understand their environment. The system featured eyeglasses that served as the base for the hardware components and provided real-time audio feedback via headphones. In this work, the Raspberry Pi 3B+ served as the primary processing unit, and a Raspberry Pi camera streamed live video. The TensorFlow object detection API was used to detect objects, and SSDLite MobileNetV2 was employed as the detection model. eSpeak was integrated into the feedback system to provide audible responses. While the system showed promising results, it still lacked certain features, such as reading the detected objects aloud more clearly. Nishajith et al. [8] developed a smart wearable cap that utilizes a Raspberry Pi 3, NoIR camera, earphones, and a power supply. The device takes a photo of the object using the NoIR camera and employs Google's Brain Team along with TensorFlow for object recognition. It can identify a range of objects from 90 categories, and earphones deliver voice output using a text-to-speech synthesizer like eSpeak. The SSD MobileNetV1 COCO model achieved a mean average precision (mAP) score of 37, but the speed of their system was relatively slow. In, the authors designed a lightweight, low-cost, portable, and hands-free navigation system for blind individuals. The system used Raspberry Pi Zero Was the central device, linking it to the Google Cloud Vision API for remote image processing and audio feedback through a speaker or headphones. The arrangement was integrated into spectacles, with a front-mounted camera that captures real-time images. The captured images were processed remotely through the Google Vision API using the user's Wi-Fi connection, and voice feedback was provided via Bluetooth. This device helps elderly and visually impaired individuals distinguish between locations, logos, text, facial expressions, and more. Although the system is effective, it lacks the ability to provide real-time feedback without needing the user to press a button. In, a smart glasses device was designed to recognize public signs in outdoor environments. The device utilizes an HD camera on the front of the glasses to capture video. The video is processed through the Intel Edison module and recognized using OpenCV techniques like SIFT and SURF. The battery and Intel module are positioned on the sides of the eyeglasses. A vision assistance system for blind individuals was also proposed in, using wearable sensors to capture images, which are then sent to the cloud for classification. The images are processed with the ResNet deep learning model to identify various objects. Kumar et al. Implemented a sensor-based device for guiding blind individuals by employing pre-trained deep learning models for object detection and person tracking

II. PROPOSED SYSTEM

In this study, a smart approach has been implemented to assist blind individuals in navigating their environment autonomously and gaining an understanding of the activities occurring around them.

This paper introduces two modes: object recognition and environmental description. Audio assistance is utilized to alert the user to the detected objects and the surrounding conditions.

2.1. Object Recognition

This paper employs the open-source MS COCO (Microsoft Common Objects in Context) dataset to build the object recognition model. The dataset comprises 165,000 annotated images across 90 categories.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Notably, the MS COCO dataset includes annotations for various tasks, including object recognition, captioning, key point identification, stuff image segmentation, panoptic segmentation, and dense pose estimation. For this study, only the object recognition aspect of the dataset is utilized.

Fig. 1 illustrates a few sample images from the MS COCO dataset. Subsequently, we adopted the deep learning-based Single Shot MultiBox Detector (SSD) for the object recognition model.

This particular model was selected due to its ability to detect objects in real-time and its superior frame processing speed. SSD is an object detection algorithm that generates bounding boxes around objects in an image or live video and predicts the corresponding object class.

Fig 1: Sample images from MS COCO Dataset.

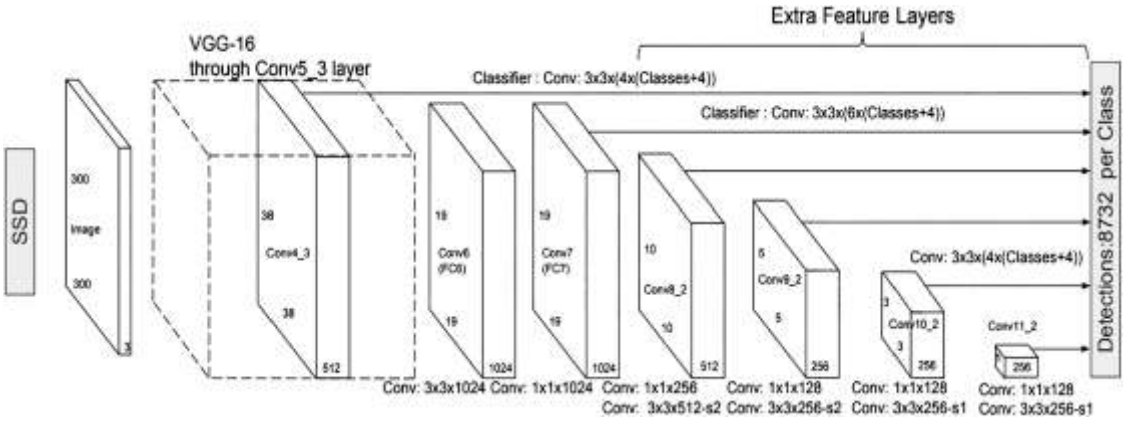

Fig 2: SSD network architecture.

Higher frame rate processing capabilities. SSD is an object detection algorithm that generates bounding boxes around objects in an image or live video and classifies the objects. The Single Shot MultiBox Detector (SSD) is straightforward to train and can be seamlessly integrated into a system to function as an object detector. It has been trained on several datasets, including MS COCO, and achieves a good mean average precision (mAP) score. The underlying principles of the Single Shot MultiBox Detector model are based on the following components:

Feature Map Extraction: Initially, the feature maps are derived from the VGG-16 network.

Single Shot: A single forward pass through the network is conducted to localize and classify the object.

MultiBox: The bounding boxes created during the detection process are primarily the MultiBox method introduced by Szegedy.

Detector: Since the network operates as an object detector, it focuses on identifying and categorizing the objects.

2.2. Ambiance mode

- A key contribution of this study is the implementation of the ambiance mode, a computer vision-based model trained on a dataset different from the one used in the object detection module of our project. As the name implies, ambiance mode is designed to describe the atmosphere or mood associated with a specific

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

location, person, or object. This mode serves as both a travel companion and a narrator for the user. The primary goal of the ambiance mode is to enhance the quality of life for blind individuals by providing a descriptive scenic narration of their surroundings.

- To activate the mode, the user of the proposed device must issue a voice command at the outset. Once enabled, the system monitors and recognizes the surrounding environment, tagging it with relevant labels. These tags enable the model to understand the context of the environment and generate an auditory description based on the identified surroundings. Finally, the corresponding audio narration is delivered to the user via voice output.

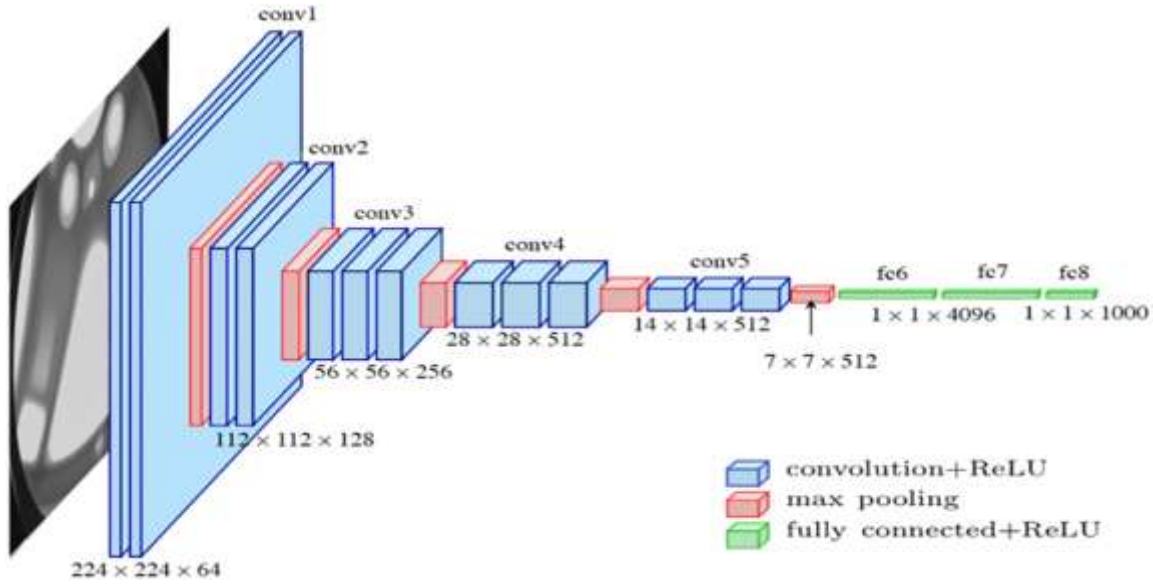

Fig 3: VGG-16 architecture

To implement the concept of the ambiance mode, this model was trained on a variety of scenes representing five different weather conditions. The categories include sunny, cloudy, foggy, sunrise, and rainy. A dataset of 500 images from each category was used, totaling 2,500 images. Fig. 3 presents a few sample images from the collected weather dataset from Kaggle. Of the dataset, 85 percent was allocated for training, and the remaining 15 percent was used for validation. It should be noted that the proposed ambiance mode system was initially tested on a desktop computer before being deployed on the embedded device. The specific implementation details are discussed in the following sections.

2.3. Hardware Implementation of the Proposed System

The primary objective of this research is to develop an innovative navigation assistance device for object recognition and environmental description. Below is a list of the hardware components used to implement this system:

- The Raspberry Pi 4 Model B+ is employed as the central processing unit due to its affordability, portability, and high performance in comparison to other embedded platforms.
- The Raspberry Pi Camera v2 serves as the video capture device. It is capable of recording video at 1080p30 resolution and is connected to the Raspberry Pi's CSI port through a 15 cm ribbon cable.
- A 10,000 mAh power bank is used as the main power source. It has an input of 5 V/1A and an output of 5 V/2.1A.
- The final prototype of the device is housed in a head cap. The Raspberry Pi is placed inside the cap, with the Pi camera mounted externally.

The basic hardware configuration using the Raspberry Pi 4 and Pi camera is shown in Fig. 3. The entire setup is lightweight, weighing approximately 140 grams (0.31 lbs), which includes the fabric head cap, Raspberry Pi, and Pi camera. These hardware components are integrated into the head cap to form the prototype of the proposed device. The internal and external views of the actual system respectively.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Once the device is powered on, the deep learning-based object detection model is loaded onto the Raspberry Pi. Real-time video is continuously captured using the attached Raspberry Pi camera. The proposed model then identifies various objects in the environment, and the corresponding audio output is delivered through earphones. The operation of object detection with the SSDLite MobileNetV2 model and audio feedback using Google text-to-speech is illustrated.

Fig 4: Elementary hardware arrangement.

Algorithm:

Algorithm for the proposed object detection with SSDLite MobileNetV2 and audio feedback using gTTS.

1. **Step 1**: Initialize the Raspberry Pi 4 and load the object recognition model.
2. **Step 2:** Optimize the hyperparameters of the SSDLite MobileNetV2 model using Particle Swarm Optimization (PSO).
3. **Step 2.1:** Randomly initialize the swarm of potential solutions within the search space based on the values provided in Table 2.
4. **Step 2.2**: Evaluate the fitness of each particle according to a defined fitness function.
5. **Step 2.3**: Update the velocity and position of each particle using its best position and the best position found by the entire swarm.
6. **Step 3:** Initialize the Pi camera and configure it to stream video frames.
7. **Step 4:** Set up the audio output device for providing feedback.
8. **Step 5:** Start the video feed from the Pi camera.
9. **Step 6:** For each frame:
10. **Step 6.1:** Pass the frame through the object detection model.
11. **Step 6.2:** Retrieve the corresponding output, i.e., the class of the detected object.
12. **Step 6.3:** Generate an audio feedback message and play it through the headphones to announce the identified class.
13. **Step 7**: Repeat Step 6 for every subsequent frame in the video stream.
14. **Step 8:** If the video stream is stopped or the program is interrupted, release any resources that were utilized by the program.

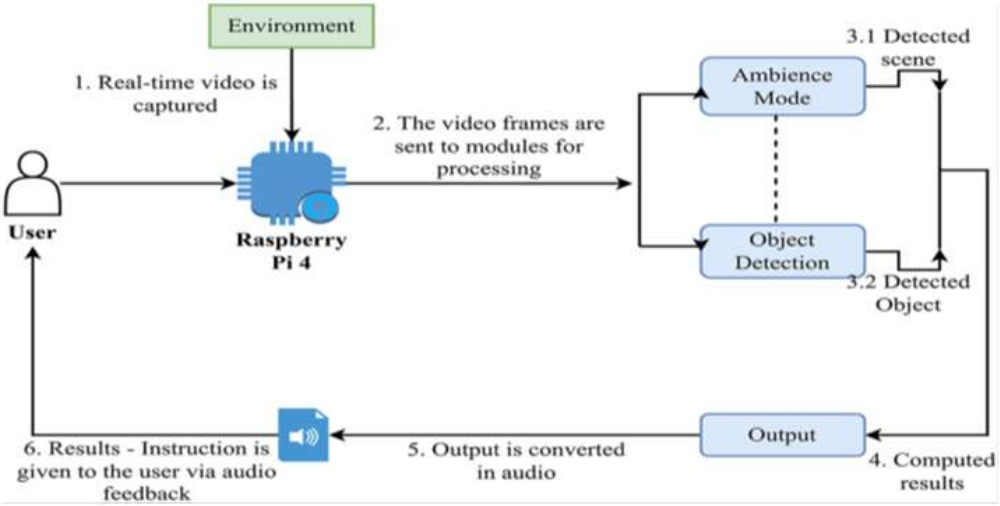

Fig 5: Working sequences of object detection with SSDLite MobileNetV2 and audio feedback using gTTS

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

III. RESULT AND DISCUSSION

This section outlines the software and hardware outcomes of the proposed automatic voice-guided object recognition and environmental description device designed for visually impaired individuals. Initially, the performance of the deep learning-based object detection and ambiance mode is presented on a desktop personal computer. Finally, the results of the proposed hardware design are discussed. Additional materials (videos) showcasing the performed hardware experiments with the Raspberry Pi device are available for reference.

IV. CONCLUSION

The primary challenge for visually impaired individuals is ensuring their safety, reliability, and accurate navigation guidance, which introduces significant difficulties in their movement and orientation. In this study, an automatic assistive framework has been developed for object detection and environmental description, designed to assist blind individuals in navigating independently without external assistance. The proposed system utilizes deep learning methods, implemented with the help of a Raspberry Pi 4 and Pi camera. The pre-trained SSDLite MobileNetV2 model has been trained on the COCO dataset for object detection and an open-source weather dataset for the ambiance description mode. The Particle Swarm Optimization (PSO) method was applied to optimize the final layers and associated hyperparameters of the MobileNetV2 model. The system is user-friendly, energy-efficient, and cost-effective, suitable for both indoor and outdoor navigation. The combined voice-assisted feature provides image-to-text conversion through speech, allowing visually impaired users to comprehend what they are encountering in their environment.

Future developments could focus on enhancing the contextual understanding, atmosphere, and adaptability of the assistive device's design. The weather dataset used for training the ambiance mode, which was relatively limited, could be expanded to improve the accuracy of the ambiance model. Currently, the audio feedback for ambiance mode relies on predefined text, which is then converted to speech using Google's text-to-speech engine. However, the integration of advanced natural language processing (NLP) techniques could make interactions sound more natural and dynamic. While the Raspberry Pi's storage, RAM, and processing speed are sufficient for this small-scale project, future work could involve deploying the assistive system on more powerful hardware, such as the Jetson Nano or a cloud server, to leverage additional computational resources and enhance the model's ability to describe more complex scenes. Moreover, the model could be trained on various situational and architectural datasets to improve the device's ability to describe real-world environments, making it more like a human companion for blind individuals. The implementation details of this work will be made open-source, allowing other researchers to contribute and adapt the project using diverse datasets to increase its adaptability to real-life scenarios. Incorporating additional sensors could further improve the system's ability to gather detailed information about its surroundings, thus enabling it to identify specific situations more accurately.

V. REFERENCES

[1] A. Bhowmick, S. Hazarika, An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends, J. Multimodal User Interfaces 11 (2017) 1–24.

[2] P. Ackland, S. Resnikoff, R. Bourne, World blindness and visual impairment: despite many successes, the problem is growing, Community Eye Health 30 (2017) 71–73.

[3] I. Khan, S. Khusro, I. Ullah, Technology-assisted white cane: evaluation and future directions, PeerJ 6 (2018) 1–27.

[4] M. Awad, J.E. Haddad, E. Khneisser, T. Mahmoud, E. Yaacoub, M. Malli, Intelligent eye: a mobile application for assisting blind people, in: IEEE Middle East and North Africa Communications Conference, 2018, pp. 1–6.

[5] S. Bhatlawande, S. Shilaskar, A. Abhyankar, M. Ahire, A. Chadgal, J. Madake, Detection of exercise and cooking scene for assistance of visually impaired people, in: G. Ranganathan, R. Bestak, X. Fernando (Eds.), Pervasive Computing and Social Networking, Springer Nature Singapore, Singapore, 2023, pp. 493–508.

[6] M. Mukhiddinov, A. Muminov, J. Cho, Improved classification approach for fruits and vegetables freshness based on deep learning, Sensors 22 (2022).

[7] M.A. Khan, P. Paul, M. Rashid, M. Hossain, M.A.R. Ahad, An AI-based visual aid with integrated reading assistant for the completely blind, IEEE Trans. Human- Mach. Syst. 50 (2020) 507–517.

[8] A. Nishajith, J. Nivedha, S.S. Nair, J. Mohammed Shaffi, Smart cap - wearable visual guidance system for blind, in: International Conference on Inventive Research in Computing Applications, 2018, pp. 275–278.

[9] M. Cabanillas-Carbonell, A.A. Chávez, J.B. Barrientos, Glasses connected to Google vision that inform blind people about what is in front of them, in: International Conference on e-Health and Bioengineering, 2020, pp. 1–5.

[10] F. Lan, G. Zhai, W. Lin, Lightweight smart glass system with audio aid for visually impaired people, in: IEEE Region 10 Conference, 2015, pp. 1–4.

[11] B. Jiang, J. Yang, Z. Lv, H. Song, Wearable vision assistance system based on binocular sensors for visually impaired users, IEEE Int. Things J. 6 (2019) 1375–1383.

[12] N. Kumar, S. Sharma, I.M. Abraham, S. Sathya Priya, Blind assistance system using machine learning, in: J.I.-Z. Chen, J.M.R.S. Tavares, F. Shi (Eds.), International Conference on Image Processing and Capsule Networks, Springer International Publishing, Cham, 2022, pp. 419–432.

[13] Y. Bouteraa, Smart real time wearable navigation support system for BVIP, Alex. Eng. J. 62 (2023) 223–235.

[14] Z. Xie, Z. Li, Y. Zhang, J. Zhang, F. Liu, W. Chen, A multi-sensory guidance system for the visually impaired using YOLO and ORB-SLAM, Information 13 (2022).

[15] M. Mukhiddinov, J. Cho, Smart glass system using deep learning for the blind and visually impaired, Electronics 10 (2021).