e-ISSN 2320 –7876 www.ijfans.org
Vol.11, S Iss 1, 2022

 UGC CARE Listed (Group -I) Journal

A Wearable Medicines Recognition System using Deep Learning for People with Visual Impairment

Nookala Venu[1*], Surendra Wani[2], Nilamadhab Dash[3], M. Sudha[4], Hima Bindu Katikala[5],

J. Sundararajan[6], Firos A[7]

[1]Department of Electronics and Communication Engineering, Balaji Institute of Technology and Science, Warangal, Telangana, India

[2]Department of Musculoskeletal Physiotherapy, MGM Institute of Physiotherapy, Aurangabad, Maharashtra, India

[3]Department of Computer science and Engineering, Centurion University of Technology and Management, Odisha, India

[4]Department of Electronics and Communication Engineering, Paavai Engineering College (Autonomous)

Namakkal, Tamil Nadu, India

[5]Department of Biomedical Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Andhra Pradesh, India

[6]Department of Electronics and Communications Engineering, NPR College of Engineering and Technology, Natham, Dindugul, Tamil Nadu, India

[7]Department of computer Science and Engineering, Rajiv Gandhi University (A Central University), Doimukh, Arunachal Pradesh, India

***Corresponding mail** id: venunookala@gmail.com

ABSTRACT

Visual impairment has had a significant impact on people and the world. Due to their more sensitive audible range and touch, Sensory Replacement Devices (SRDs), which convert visual information to audio or touch, are an accessible choice for those who are blind or visually impaired to improve their quality of life, employment opportunities, and education. Using scene-perception-based deep learning, we presented a spectacle with provision of vision-to-audio transfer system in this study to help visually impaired persons recognise and find familiar substances in their location. The scheme comprises of a Bluetooth voice feedback unit with a microphone, a wireless camera unit, and a mobile application running customised software. The camera element collects imageries from the environment then transfer them to an mobile software application. People with blindness who use the programme may get spoken instructions and audio aid thanks to the Bluetooth voice feedback

unit. The audio voice recognition and object identification replicas are loaded by the Android-based application. It has been discovered that using this technique may help people with limited eyesight locate and identify objects. With this system one can also take the medicine in which they are using daily.

Keywords: Sensor, deep learning, visual impairment, android based application.

1. Introduction

Impairment of vision is one of the world's most serious well-being issues, causing significant disruption in people's lives. Worse, it increases the strain on society and the family. Blindness affects 36 million of the world's 253 million visually handicapped people. As a result, developing visual assistive technology, such as Sensory Substitution Devices, is critical (SRDs) [1].

Researchers have worked hard to develop SRDs that can convert information from camera-captured ambient pictures into other sensory data for blind individuals. Others use electrodes or vibrators to convert optical input to electric touch or mechanical vibration [2], [3]. Some SRDs are designed to convert visual data into verbal or auditory commands. SRDs can assist the visually impaired with a number of tasks such as gesture, facial, and phrase recognition [4], [5]. Deep learning methods are being used to SRDs as artificial intelligence advances to assist the blind with increasingly difficult vision tasks like as public facility identification, interior scene description, and navigation [6].

The goal of this research is to create a spectacle, efficient vision-to-audio device for blind persons [7]. Figure 1 depicts the general architectural arrangement of the system. The device contains of a couple of glasses outfitted with a wireless photographic camera unit and Bluetooth headphones, as well as proprietary Android software and an app based on neural network models. Voice communication, voice recognition, and object identification are all possible with this sensory substitution technology. The consequences of the testing show that our technology can assist people with blindness in recognising and learning about common things in their surroundings [8]–[10].

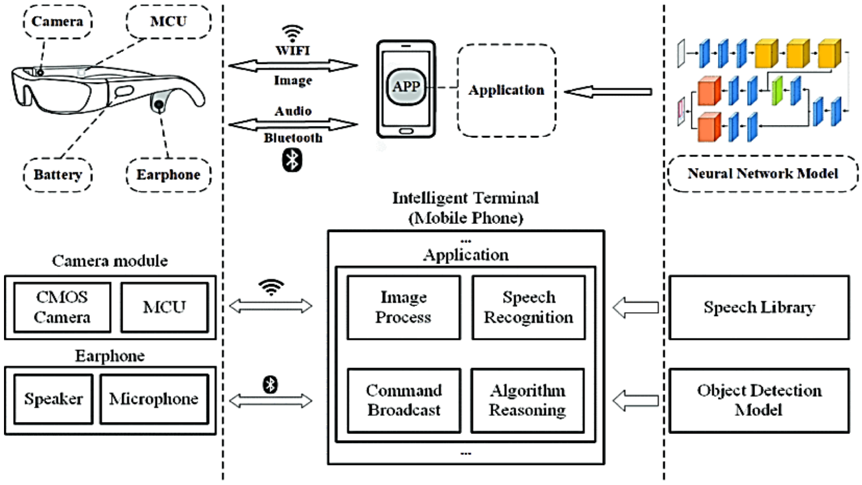


Fig. 1 Design architecture

2. Design of the proposed system

As illustrated in Fig. 1, our scheme comprises of a camera unit, Bluetooth earphones, an mobile interface, and a neural network model. The camera unit is made up of the OV2640 CMOS camera, the ESP32 control unit, and the power source. The OV2640 CMOS camera is used to continually record photos of the surroundings and communicate them in real time to the ESP32. Since the ESP32, an MCU with Wi-Fi capabilities, is available, the camera unit may interface wirelessly with a smartphone and relay data. The power source unit is a poly lithium battery with a 3.7V voltage and rechargeability.

Figure 2 depicts a flowchart for the operation of the camera unit. (a). The camera unit initially checks the Wi-Fi and camera capacity to ensure that the system is operational. While the camera unit is waiting, the Android phone is connected. After connecting, if the camera unit recognises a valid application command, it will begin gathering images and delivering them to the programme. The usage of Bluetooth earphones allows blind people to interact with the system more effortlessly. We employ Bluetooth bone-conduction earbuds with noise-cancelling features to guarantee the indicator from the proposed system is not interfered with through the signal from the surroundings.

The system's application is designed to collect signals, obtain relevant data, and communicate with customers. The open-source Android operating system, which is used to construct the software, has many benefits. The four sections of the software are algorithm reasoning, command broadcasting, voice recognition, and image processing.

The image processing unit handles the Wi-Fi connection, image receipt, and image scaling. Following their first interaction, the camera unit gets directives from the unit in order to capture photographs. Before being passed to the algorithm reasoning unit for additional processing, the photos are scaled.

The voice recognition unit recognises voice and converts it to text. The voice recognition unit must stay disconnected since it is helpful and accessible when the mobile phone's signal is low and Wi-Fi capacity is required for picture transfer. For these reasons, the unit employs Alpha Cephei's VOSK voice recognition technology. While the framework is being installed in the application, the unit takes the user's voice and converts it to text using the voice library.

The command broadcast unit converts text to voice and broadcasts it to users through the speaker. The text-to-voice engine for the command broadcast unit is also incorporated into the mobile or may be transferred from the Internet to assist users with a number of visual tasks. When the programme provides user advice, the unit identifies the language of the instructions and converts them to voice. The unit then establishes contact with a Bluetooth earpiece in order to relay the orders to clients.

The program's foundation is the algorithm reasoning unit, which is required for getting visual features. The unit might be used to incorporate deep learning object identification algorithms into mobile devices. For neural network reasoning, the programme employs the efficient and portable NCNN framework. Using the algorithm reasoning unit, the neural network model and framework are included into the software. The algorithm unit might then use the Java Native Interface to invoke the model (JNI). If the object recognition model is incorporated in the software, the instruction unit may receive a picture and provide the position and kind of substances in the image.

Figure 2 depicts the application's flowchart (b). The computer starts its typical starting procedure. Users speak instructions into the system, which uses the voice recognition unit to obtain them. If the instructions are correct, the system performs an object recognition algorithm to hunt for common things in the picture after receiving photos from the camera unit. If the required components are located, the system will notify the user of the object's position.

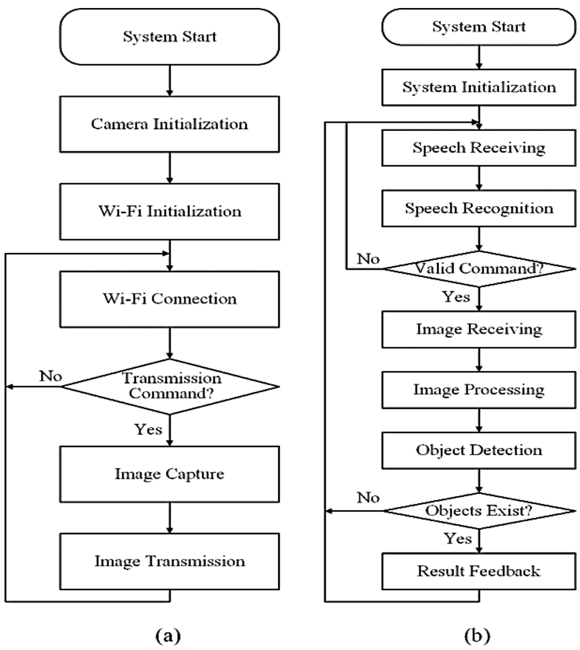


Fig. 2 a) unit of camera b) mobile application flow diagrams

The neural network model of the system is divided into two components. The first is a voice recognition voice library, and the second is an object recognition model that predicts the type and placement of things in the environment.

The voice library is imported as a model by the voice recognition unit. Because of the voice library, the voice recognition unit may securely give the recognition result after accepting user instructions. The VOSK website is used to download the Mandarin Chinese library into the programme.

To perform the object recognition function, we choose YoloV4-Tiny, a tiny object recognition model, and load it into the algorithm intellectual unit. YoloV4-Tiny employs segments of computation to extract maps, while YoloV3 Heads decodes the position and kind of bouncing boxes. It also employs data increase and pre-set anchors to increase the accuracy of the recognition findings. Figure 3 depicts the model's architecture. To directly identify the locations of the items, we split the picture into nine equally sized parts along its length. The item's position will be determined by the region in which the object's centre is put. For example, if the object's centre is in the intermediate, the direction will be intermediate.

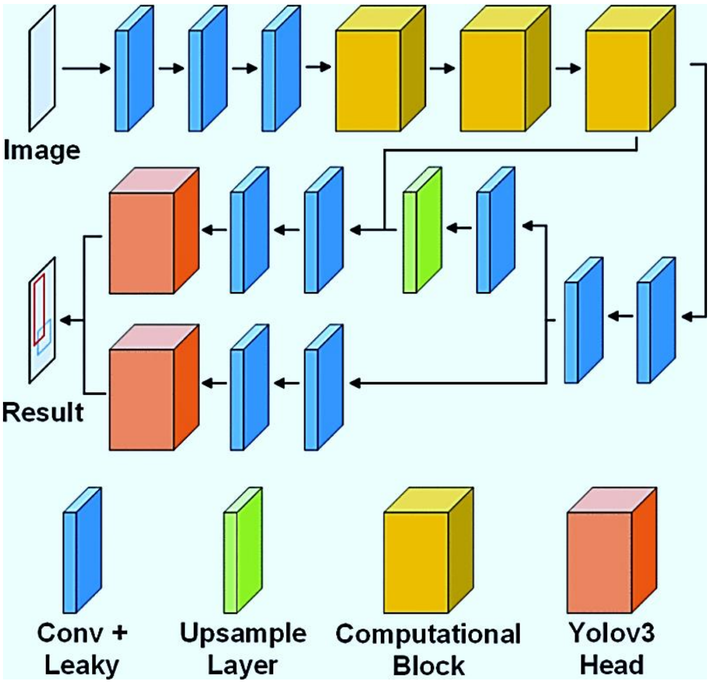


Fig. 3 Yolo Network architecture

3. Result and discussion

To ensure that the classification functions smoothly and proficiently, the cell phone's efficiency must meet a specific standard. We execute the Android application and evaluate system performance on a Xiaomi 8 smartphone with a 845 CPU and 4 GB of RAM. We monitor the battery life, Wi-Fi frequency band, viewing angle, and picture resolution of the camera unit. We keep track of how much time the software spends on verbal feedback, receiving single frames, and voice recognition. The prototypical size and execution period of the object recognition model utilised in the application are being monitored. Figure 4 depicts the system's performance.

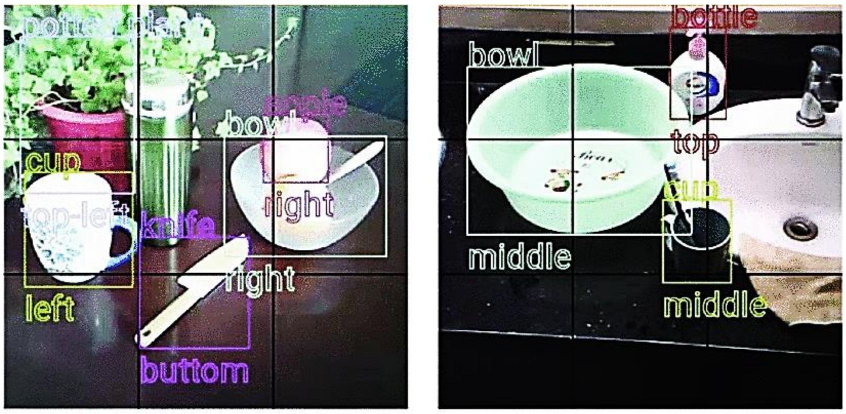
Parts	Parameters	Value
Camera module	Resolution	640*480
	Visual angle	68°
	Wi-Fi frequency band	2.4GHz
	Battery capacity	300mAh
Application	Single frame receive time	~0.03s
	Speech Recognition time	~2s
	Speech Feedback time	~1s
Object Detection model	Model size	23.0MB
	Model running time	~0.145s

Fig. 4 Performance of the system

The goal of experiments is to determine the overall efficacy of the system. The guiding principles of the Helsinki Declaration are followed throughout the investigation. The experiment is carried out at a specific location. A Xiaomi 8 smartphone with the trial-specific application must be installed, and the participant must wear glasses. Figure 4 depicts the system's human implementation. At all times, the test subject must provide vocal instructions to the machine. According to the system's instructions, the volunteer gets verbal input that is transformed into item recognitions.

Fig.5 Testing with human

The three indicators we employ to monitor task performance throughout the trial are the item's presence, kind, and position. Figure 5 depicts instances of command-answer designs used in the investigation. For instance, if the algorithm properly recognises and identify an apple on the accurate size of the picture, the record will include the terms "Found," "Apple," then "Right" in the relevant locations. Figure 5 depicts the work performance records. Some of the answers to the 12 problems are erroneous due to occlusion and the uneven lighting effect, while others are accurate.



Command: Where is the apple?
Answer: Found, apple, right.

Command: I want to find a bowl.
Answer: Found, bowl, middle.

Fig. 6 examples of the object recognition

No	Command	Answer
1	I want to find an apple.	Found☑, Apple☑, Right☑
2	I want to find a cup.	Found☑, Cup☑, Top-left☑
3	Where is the bottle?	Found☑, Bottle☑, Left☑
4	I want to find an orange.	Found☒, Orange☒, Middle☒
5	Is there a spoon?	Found☑, Spoon☑, Middle☑
6	Where is the scissors?	Found☒, Scissors☒, Right☒
7	Is there a refrigerator?	Found☑, Refrigerator☑, Middle☑
8	Is there an umbrella?	Found☑,Umbrella☑, Bottom-Right☑
9	I want to find a backpack.	Found☑, Backpack☑, Left☑
10	Where is the chair?	Found☑, Chair☑, Right☑
11	I want to find a toothbrush.	Found☑, Toothbrush☒ Middle☑
12	Is there a bowl?	Found☑, Bowl☑, Middle☑

Fig. 7 performance of the system

Conclusion

It is observed that the video camera satisfies the real-time criteria by distribution an image with a accuracy of 640*480 and a viewing angle of 68° to the phone in around 0.03 seconds. The object recognition algorithm forecasts the results in around 145 milliseconds on an Android phone with a Snapdragon 845 CPU operating at 2.8GHz. Although this period of time cannot satisfy real-time demands, it has no impact on the pictorial task since the user with blindness needs sufficient time to respond to information from the system. Finally, the system is functional and accessible if the mobile phone's CPU operates at a frequency greater than 2.8GHz. Figure 7 demonstrates how the scheme can accurately forecast the site and kind of large or medium-sized objects, such as a chair, umbrella, and cup. However, the system struggles to find little substances like a brush and cutters. Additionally, the algorithm struggles to detect concealed objects or scenes with uneven lighting effects. The results demonstrate that the system can identify unshaded, non-small items and recognise vocal instructions with accuracy in a room with balanced lighting.

In this research, we established a wearable, efficient, and instantaneous vision-to-audio sensory transfer system that can provision voice interaction via voice credit, acknowledgement, and voice input. The system's functionality can be expanded in the forthcoming by adding more machine learning methods, such as minor thing recognition, semantic division, and depth approximation. This will help people with visual impairments with object ability to handle, collision avoidance, route planning, pattern recognition, and taking the medicine.

References

[1] L. Pillette, F. Lotte, B. N'Kaoua, P. A. Joseph, C. Jeunet, and B. Glize, "Why we should systematically assess, control and report somatosensory impairments in BCI-based motor rehabilitation after stroke studies," *NeuroImage: Clinical*, vol. 28, no. August, p. 102417, 2020, doi: 10.1016/j.nicl.2020.102417.

[2] M. Caubet *et al.*, "Bayesian joint modeling for causal mediation analysis with a binary outcome and a binary mediator: Exploring the role of obesity in the association between cranial radiation therapy for childhood acute lymphoblastic leukemia treatment and the long-term risk of insulin resistance," *Computational Statistics and Data Analysis*, vol. 177, p. 107586, 2022, doi: 10.1016/j.csda.2022.107586.

[3] J. Habulezi and T. N. Phasha, "Provision of Learning Support to Learners with Visual Impairment in Botswana: A Case Study," *Procedia - Social and Behavioral Sciences*, vol. 69, no. Iceepsy, pp. 1555–1561, 2012, doi: 10.1016/j.sbspro.2012.12.098.

[4] M. Mardaniyan Ghahfarrokhi, E. Banitalebi, M. Faramarzi, and R. Motl, "Feasibility and efficacy of home-based neurofunctional exercise vs. resistance exercise programs for ambulatory disability of multiple sclerosis patients with cognitive impairment," *Multiple Sclerosis and Related Disorders*, vol. 58, no. July 2021, p. 103400, 2022, doi: 10.1016/j.msard.2021.103400.

[5] Z. Shi, D. D. Langleben, C. P. O'Brien, A. R. Childress, and C. E. Wiers, "Multivariate pattern analysis links drug use severity to distributed cortical hypoactivity during emotional inhibitory control in opioid use disorder," *NeuroImage: Clinical*, vol. 32, no. May, p. 102806, 2021, doi: 10.1016/j.nicl.2021.102806.

[6] T. Motoyoshi, N. Tetsumura, H. Masuta, K. Koyanagi, T. Oshima, and H. Kawakami,

"Tangible gimmick for programming education using RFID systems," *IFAC-PapersOnLine*, vol. 49, no. 19, pp. 514–518, 2016, doi: 10.1016/j.ifacol.2016.10.608.

[7] J. V. Shanmugam, B. Duraisamy, B. C. Simon, and P. Bhaskaran, "Alzheimer's disease classification using pre-trained deep networks," *Biomedical Signal Processing and Control*, vol. 71, no. PB, p. 103217, 2022, doi: 10.1016/j.bspc.2021.103217.

[8] B. Valéry, V. Peysakhovich, and M. Causse, "Hear me Flying! Does Visual Impairment Improve Auditory Display Usability during a Simulated Flight?," *Procedia Manufacturing*, vol. 3, no. Ahfe, pp. 5237–5244, 2015, doi: 10.1016/j.promfg.2015.07.596.

[9] T. Farooqui, "A potential link among biogenic amines-based pesticides, learning and memory, and colony collapse disorder: A unique hypothesis," *Neurochemistry International*, vol. 62, no. 1, pp. 122–136, 2013, doi: 10.1016/j.neuint.2012.09.020.

[10] D. Dulin and C. Serrière, "Special needs education in the blind population: effects of prior expertise in raised line drawings on blind people's cognition," *Procedia - Social and Behavioral Sciences*, vol. 1, no. 1, pp. 549–553, 2009, doi: 10.1016/j.sbspro.2009.01.099.