*2023 International Research Conference of Sri Lanka Technology Campus*
*Colombo, Sri Lanka. 14th-15th December 2023*

# Computer Vision and Deep Learning Based Assistance System for The Visually Impaired

Keirishan Balachandran
*Faculty of Engineering*
*SLTC Research University*
Padukka, Sri Lanka
balachandrank@sltc.ac.lk

Mithurshan Krishnanathan
*Faculty of Engineering*
*SLTC Research University*
Padukka, Sri Lanka
krishnanathanm@sltc.ac.lk

Swapna Premasiri
*Faculty of Computing & IT*
*SLTC Research University*
Padukka, Sri Lanka
swapnap@sltc.ac.lk

Mahima Weerasinghe
*Faculty of Computing & IT*
*SLTC Research University*
Padukka, Sri Lanka
mahimaw@sltc.ac.lk

***Abstract*—The creation of algorithms and methods for computer vision enables machines to comprehend visual input at a deep level, including photos and videos. By giving computers the ability to decipher and analyse visual data, identify things, and comprehend their surroundings, it seeks to mimic human vision. This project aims to develop an assistance system for the visually impaired. The proposed system will use a camera mounted on a wearable device to capture the surrounding environment. This system will function according to the computer vision techniques to detect and recognize objects, pathways, and pedestrian crossing and convert to text, in real-time. First, we have collected datasets of objects, pathways, and pedestrian crossing then we will annotate our data set with Roboflow and train with YOLO V8. By giving voice navigation it will be able to assist the user. If the objects detect then we will convert those objects into text format. After that, we will convert the text into the voice. The system will employ a deep learning (DL) model to classify the detected objects and a separate pre-trained DL model to provide audio feedback to the user through earphones. The obstacle detection module of this system will also use depth sensing techniques to detect and alert the user about potential obstacles in their path. The user will have the ability to interact with the system using voice commands, which are processed using natural language processing algorithms. The system's accuracy and effectiveness will be evaluated using user studies and benchmark datasets.**

***Keywords*—Computer vision, object detection, image segmentation, text recognition, audio feedback, and natural language processing***

## I. INTRODUCTION

As of 2017, the Sri Lanka Federation of the Visually Handicapped reported around 70,000 registered blind persons in Sri Lanka. To aid them, we have planned to implement a system. That is a computer vision and Deep learning-based assistant system for the visually impaired.

Computer vision techniques combined with deep learning can play a significant role in developing an assistant system for visually impaired individuals. Such a system can help them perceive and understand their surroundings, navigate safely, and access visual information. Object detection algorithms can be employed to detect and locate various objects in the environment, including obstacles, pedestrians, vehicles, and important landmarks. This system can offer real-time audio or haptic feedback to guide visually impaired individuals during navigation.

In this modern world, every activity is interrelated with technology. These technologies assist us in our day-to-day activities. Therefore, we have planned to implement an assistance system for the visually impaired with the help of technology. Visually impaired people are facing many difficulties in theirlives. They have to depend on others when they go out alone.Our aim is to make them function without others' assistance.Our system is to be designed in such a way that the visually impaired can be able to identify all the things that are on the road. There are plenty of assistance systems for the visually impaired in the world market, but our system we try to make that system more user-friendly. While the visually impaired person walking along the road, pathway, pedestrian crossing, and objects will be detected and give voice navigation about them. In addition to that, this system also helps to prevent accidents by giving voice alerts.

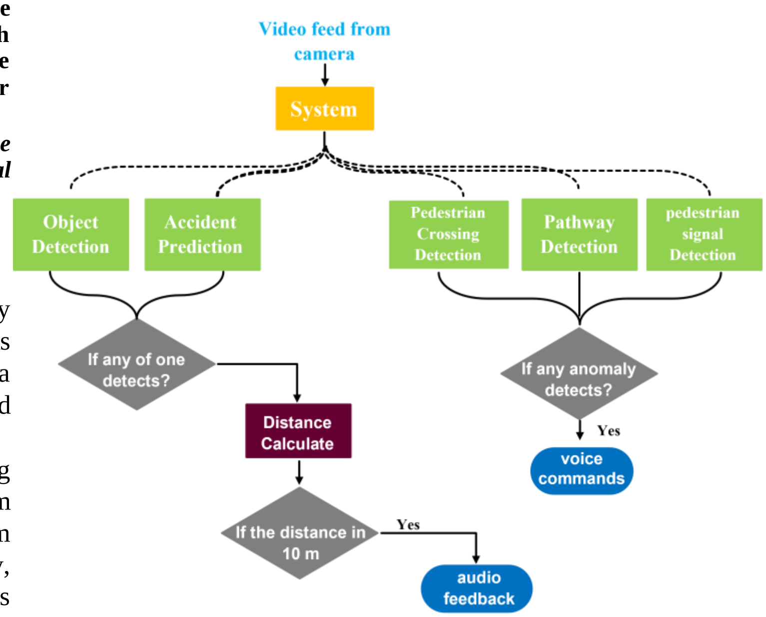


Fig. 1. Overview of project

## II. System Model

This block diagram clearly shows how our models will train. The training datasets will go through image pre-processing and feature extraction first. after finishing those, we need to train the model using the YOLO V8 algorithm. After the training process, we will get a trained model. Using the trained model, we need to recognize our system. so, for the recognition testing datasets will go through image pre-processing and feature extraction first. later the trained model will analyse the output from testing datasets and show us in the display. Now people understand what is happening in front of them.

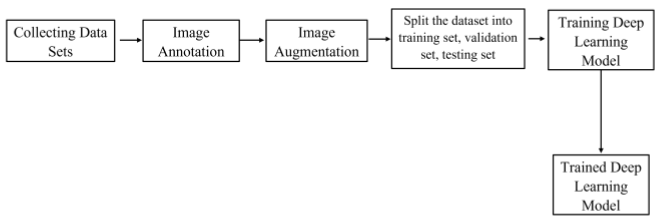


Fig. 2. Block diagram of models

## III. Related Work on Literature

To aid visually impaired people in social interactions, an electronic device with obstacle detection and face recognition has been developed. The proposed device was in the shape of smart eyewear that has an ultrasonic sensor and Pi camera with a Raspberry Pi installed. The Raspberry Pi, which is coupled to the ultrasonic sensors, receives data signals from the sensors for additional data processing and can detect impediments up to 1000 cm away from the user. A pi camera is used to recognize the person in front of the visually challenged according to the database list. This gadget makes it simple to recognize people. It is an inexpensive, power-efficient device [1].

Produced low-cost, dependable, transportable, user-friendly, sturdy Smart Glasses for those with visual impairments for easy navigation. The device is affordable and wearable like glass. The device features a built-in sensor that sends out ultrasonic waves in the direction the user is moving and has a maximum scanning range of 5–6 meters. The system's earliest iterations relied on simple image processing and computer vision techniques, and additional improvements were developed to identify a secure path for user movement. When an obstacle is there, the sensor picks it up and sends the information to the gadget, which then automatically produces a speech that the user may hear through an earpiece [2].

Presented a system for blind persons to use to detect and identify outdoor barriers. Their system was given a new obstacle detection dataset called OD that contains 15 typical objects and was used as a new baseline for outdoor obstacle detection. Three object identification techniques, including YOLO, SSD, and Faster RCNN, were utilized in their research to determine the optimum model for object detection. The evaluation results showed that the YOLO algorithm outperformed the other two algorithms [3].

Suggested a Blind Sight-Object Detection system that is based on cutting-edge object detection technologies and computer vision. The technology seeks to automate things that the human visual system can do. The features of the image are recognized and assigned to the correct class using image classification algorithms. The 123,287 hand-labeled images that make up the COCO dataset, which was employed in this experiment, are divided into 80 categories. The spatial relation- ships between things and their locations in the environment are described using this vast collection of data. This system uses the Object Detection Framework You Only Look Once to find objects. To distinguish the monetary denominations, a module for recognizing Indian currency was also created [4].

Presented a navigation aid for visually impaired persons to help them with autonomous direction. A camera, single-board DSP processor, wet floor sensor, battery, and a variety of sensors and processing components make up the entire system for those with vision impairments. To help the user become accustomed to the environment, the machine learning model was employed for object recognition. The proposed gadget was utilized to identify obstructions, stairs, potholes, speed bumps, wet flooring, and tight spaces in order to provide the best navigational direction. The output was delivered in the form of the vibration, to inform the user about the impediment [5].

## IV. Computer Vision Approach

Computer vision consists of techniques for obtaining, processing, analyzing, and understanding images. In general, it can produce high-dimensional information from the real world to numerical or symbolic information. In a scientific area, computer vision is involved with the theory behind artificial systems and it can extract information from images. The data from the image can be taken in many ways, such as video sequences, from multiple cameras, or multi-dimensional data from a medical scanner. As a technological improvement computer vision systems seek to apply computer vision theories and models. Sub-domains of computer vision include video tracking, object recognition, event detection, learning, indexing, motion estimation, and more [6].

### A. Dataset Collection

Machine learning is divided into many subcategories. Supervised learning is one of the subcategories of machine learning and artificial intelligence. For supervised learning, we have to teach or train the machine using data that are well-labeled. Supervised learning is highly accurate rather to unsupervised learning. We need datasets for supervised learning. So, we collected datasets for objects, pathways, and pedestrians crossing while walking into the path.

### B. Training

For the training, we used the YOLO v8 algorithm. We cloned the YOLO GitHub repository into google Colab and started training the model. For the training, we used 50 epochs. Epochs mean the number of times the learning algorithm will go through the training datasets. For the training, it took a lot of time. It took 4 hours. Like that, each of the training takes a lot.

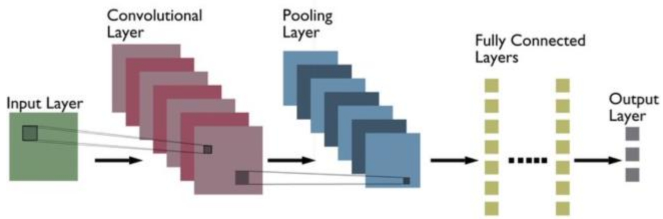


Fig. 3. Architecture of CNN

The algorithm we are planning to use here is Convolutional Neural Network (CNN) [7]. The convolutional neural network may be a profound learning neural organization outlined for preparing organized arrays of information such as pictures. Convolutional neural networks are broadly utilized in computer vision and have gotten to be the state of the art for numerous visual applications such as image classification, and have moreover found success in normal language processing for text classification. They are specifically designed to process pixel data and are used in image recognition and processing. So, we think our system will be most effective if we use Convolutional Neural Network (CNN).

Figure 3 shows how CNN algorithms work on this project. The first convolution layer simplified this complex image. The filtering process happens in this layer. The next layer is the pooling layer. This layer makes the process much faster and it creates a pooled feature map. After that, the upcoming layers extract small features, So the output will be highly accurate.

## V. Results

We have planned to show the results of our trained models on object detection, pathway detection, and pedestrian crossing detection. The objective of our research is to develop effective and accurate algorithms for detecting objects, pathways, and pedestrian crossings in real-world scenarios. We evaluated our models on a different dataset comprising different urban and suburban environments.

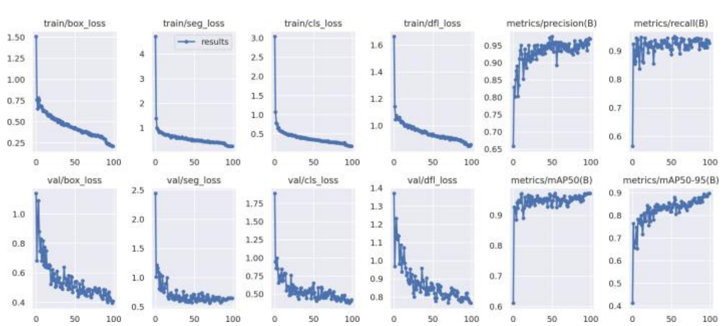


Fig. 4. Training, validation losses graph

In the above figure, 4 clearly shows the training losses and validation losses. Our object detection system was prepared to employ a state-of-the-art deep learning architecture, particularly leveraging the YOLO (You Only Look Once) algorithm. The model was initialized with a pre-trained YOLO modelon an expansive annotated dataset containing a wide range of objects, including vehicles, people, posts, and traffic lights, as well as pathways, pedestrian crossings, and pedestrian signals. This pre-training ensures that the model benefits from the knowledge gained on a large-scale dataset, enhancing its ability to generalize to diverse object categories. During the evaluation stage, we measured the execution of our object detection framework using well-established metrics such as accuracy, precision, recall, and the F1 score.

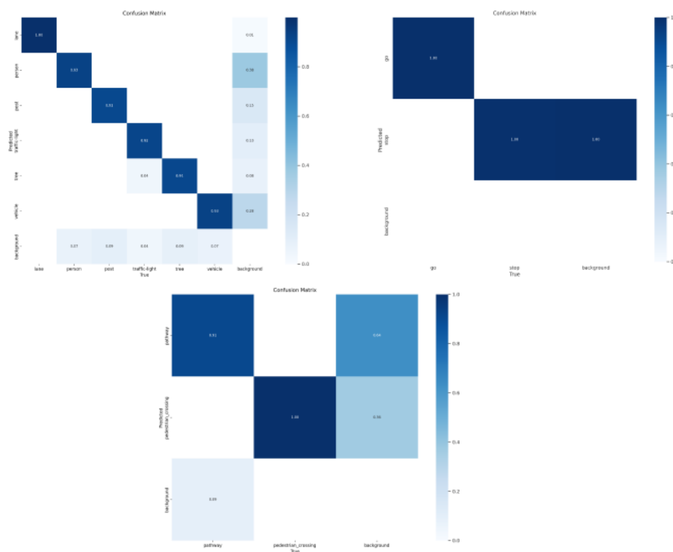

Fig. 5. Validation graphs

The confusion matrix depicted in Figure 5 provides a comprehensive evaluation of the performance of our trained model. The matrix illustrates the distribution of true positive (TP), truenegative (TN), false positive (FP), and false negative (FN) pre-dictions across different classes. This visualization is crucial for assessing the model's ability to accurately classify objects within the given dataset. High values along the diagonal indicate successful predictions, while off-diagonal elements highlight instances of misclassification. The precision, recall, and F1-score metrics derived from the confusion matrix furtherquantify the model's efficacy in terms of both precision and recall, offering insights into its overall classification performance. The detailed analysis presented in this section servesas a critical reference for understanding the strengths and potential areas of improvement in our model's object detectioncapabilities.

## VI. Future Works

As part of our future work, we are planning to enhance the functionality of our system by connecting the mobile phones of visually impaired individuals via Bluetooth. This connectivity will enable our system to provide valuable assistance, including location tracking, checking current weather conditions, and facilitating communication with housemates. To further enhance the object detection capabilities of our system, we are exploring the integration of sensor technologies. Specifically, we plan to incorporate ultrasonic sensors to augment object detection. These sensors offer real-time distance measurements, allowing for a more robust and accurate detection of obstacles and objects in the user's vicinity. By leveraging sensor datain conjunction with our deep learning-based object detection framework, we aim to provide a comprehensive and reliable assistance system for individuals with visual impairments.

## VII. Conclusion

The proposed assistance system for the visually impaired can be a powerful tool for accomplishing their daily tasks, providing them with greater independence, and improving their quality of life. The system can recognize objects, read printed materials, and provide navigation assistance by leveraging the latest advancements in computer vision and deep learning. The development of such a system requires a multi-disciplinary approach, bringing together experts in computer vision, deep learning, natural language processing, and user experience design. The system should also be designed with inclusively in mind, ensuring that it meets the needs of a diverse range of visually impaired individuals. Overall, the development of this system for visually impaired individuals is an exciting area of research and has the potential to significantly impact millions of people worldwide.

## References


[1] K. S. Srinivas, K. Sahithya, G. L. Tejaswi, K. H. Gopal, and B. P. Karthik, "A new method for recognition and obstacle detection for visually challenged using smart glasses powered with raspberry pi," *International Journal of Engineering Applied Sciences and Technology*, vol. 5, no. 1, pp. 408–412, 2020.

[2] P. Hegde, N. Devathraj, S. Sushma, and P. Aishwarya, "Smart glasses for visually disabled person," *International Journal of Research in Engineering and Science (IJRES)*, vol. 9, no. 7, pp. 62–68,

[3] W. Tang, D.-e. Liu, X. Zhao, Z. Chen, and C. Zhao, "A dataset for the recognition of obstacles on blind sidewalk," *Universal Access in the Information Society*, pp. 1–14, 2021.

[4] A. Annapoorani, N. S. Kumar, and V. Vidhya, "Blindsight: Object detection with voice feedback," *International Journal of Scientific Research & Engineering Trends*, vol. 7, no. 2, pp. 644–648, 2021.

[5] S. Yadav, R. C. Joshi, M. K. Dutta, M. Kiac, and P. Sikora, "Fusion of object recognition and obstacle detection approach for assisting visually challenged person," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2020, pp. 537–540.

[6] S. Minocha, A. Roychowdhury, P. Dittia, and P. Leekha, "Personal driver assistance system (pdas) based on raspberry pi,"

[7] C. Sruthi and A. Lijiya, "Signet: A deep learning based indian sign language recognition system," in *2019 International conference on communication and signal processing (ICCSP)*, IEEE, 2019, pp. 0596–0600.