Object Detection and Localization for Visually Impaired People using CNN

Sony Tirkey[1], Anmol Ratan Tirkey[2], Cazal Tirkey[3]

[1]Student, CHRIST (Deemed-To-Be University), Bengaluru, Karnataka, India
[2] Student, JAIN (Deemed-To-Be University), Bengaluru, Karnataka, India
[3]Researcher, Bengaluru, Karnataka, India

Abstract - *Visually impaired people constitute a significant portion of the global population, with both permanent and temporary disabilities. WHO estimates around 390 lakh individuals are completely blind, and 2850 lakh individuals are purblind or visually impaired. To help them aid in daily navigation, numerous supporting systems are being developed which had numerous disadvantages. Our main objective is to create an auto-assistance system for the visually impaired. By using CNNs (Convolution Neural Network), a widely-used approach in deep learning models, our system achieves over 95% accuracy in object detection based on camera images. Identified objects are conveyed through voice messages, making it a valuable prototype for assisting the visually impaired.*

Key Words: Visually Impaired, Object Detection, CNN, Deep Learning, assistance.

1. INTRODUCTION

Visually impaired people constitute a significant portion of the population component, with tens of millions predicted to exist globally. Their integration into society is an essential and ongoing aim. A lot of work has gone into ensuring a health-care system, to help visually impaired people live a normal life, many guiding system approaches have been created. These systems are frequently created just for certain activities. However, these solutions can significantly improve such people's ability to move and their security.

The advancement of cutting-edge guiding systems to assist visually impaired persons is tightly linked to advanced technologies in image processing and computer vision, as well as the speed of the devices and unit processors. Regardless of the technology used, the application must work in real time with quick actions and decisions, as speed is crucial for taking action.

Choosing the best possible outcome is essentially a trade-off between the performance of the software component and the hardware capabilities. It is necessary to adjust the parameters to optimum. One of the primary goals of the aided system during a visually impaired person's indoor movement is to automatically identify and recognize objects or obstacles, followed by an auditory alert.

The image processing vision module described in this system is an integrated aspect of the platform dedicated to assist visually impaired people. Furthermore, the provided module can be used independently of the integrated platform. The proposed vision-based guidance system is created, built, and tested throughout experiments and iteratively optimized. The module follows the principle of producing a high-performance device that is also cost-effective for practical use. The module employs disruptive technology and permits updates and the addition of new functionality.

WORK DONE

Downloaded the project's model file.

2. EXISTING SYSTEMS

Convolutional Neural Networks (CNN), speech recognition, smartphone camera, and object personalization were all used in existing systems. The purpose is to help visually impaired people navigate indoor surroundings, recognize items, and avoid obstacles.

By using facial recognition for authentication, the Facial Identification and Authentication System provides a secure and personalized user experience while also ensuring that only authorized users may access the system and its features. Nonetheless, it is dependent on the accuracy of facial recognition technology, which can be influenced by lighting, changes in look, and other factors.

The Object Detection System (General Object Detection - Model 1) employs a pre-trained CNN model for general object detection, allowing the system to identify a wide range of items and providing real-time object recognition, which improves the user's comprehension of their surroundings. However, it is restricted to the items and categories in the pre-trained model. Objects that are not part of the model's training data may not be detected accurately. The Customized Object Detection System (Model 2) allows users to personalize the system by adding their own detection objects, increasing the system's versatility and usability for visually impaired users with special needs. However, users must take and label photographs, which can be time-consuming. Accuracy may also vary depending on the quality of photographs captured by users. Distance measurements may contain mistakes due to reliance on good camera and the assumption of a fixed focus length.

The detection of obstacles and navigation direction improves the user's safety when travelling through indoor spaces. It delivers real-time alerts and guidance to help you avoid obstacles and get to your destination. However, obstacle detection and navigation guidance may not be totally foolproof, and users must still use caution. Overreliance on the system may result in unanticipated outcomes.

Text-to-Speech Interaction and Speech Recognition offer natural and convenient interaction between the user and the smartphone. It recognizes human commands and delivers navigation instructions, making the system more user-friendly. However, the accuracy of speech recognition can vary based on the user's speech patterns, accents, and background noise. It is possible that commands will be misinterpreted.

Supports personalized object detection, allowing users to add objects tailored to their individual requirements. Cloud training improves the precision of personalized models, ensuring accurate object recognition. However, it may create latency and reliance on network connectivity. Concerns about privacy should be considered while uploading personal data to the cloud.

The experimental results and evaluations show promising accuracy in object detection and distance measurement, confirming the proposed approach. However, experimental results can differ depending on the testing environment and the quality of the data acquired. The accuracy percentages stated may not be achievable in all real-world conditions.

While the suggested systems provide essential features and benefits for visually impaired individuals, they also have limitations and possible issues that must be addressed. The careful examination of these benefits and drawbacks is critical for the successful deployment and enhancement of the overall system.

3. PROPOSED SYSTEMS

This research addresses at the difficulties that visually impaired people have when traversing indoor surroundings. The system attempts to enable real-time object recognition and localization by exploiting the capabilities of CNNs, while also providing users with complete and intuitive audio or haptic feedback.

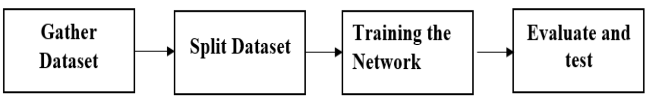

Fig -1: Deep Learning Steps

The following are the aims of the presented research paper:

Object Detection and Recognition: Use of CNN for precise object recognition, improving awareness of the indoor environment.

Object Localization: The use of a system to calculate or point out the location of an object in the frame of any image or video input.

Speech Interaction and Communication: Allows users to interact with the system by using voice commands and hearing instructions.

Personalized Object Detection: Entails developing customized object detection models and providing support for user-specific items.

Affordability and Accessibility: Develop a cheap solution by utilizing widely available cellphones with integrated sensors and functionalities.

Improved freedom and Safety: Improve the freedom and safety of visually impaired people by making indoor navigation and object recognition easier.

4. IMPLEMENTATION

A camera is used to capture the footage, which is then separated into frames. CNN classifiers are used for object detection, and pyttsx3 is used for text-to-speech conversion.

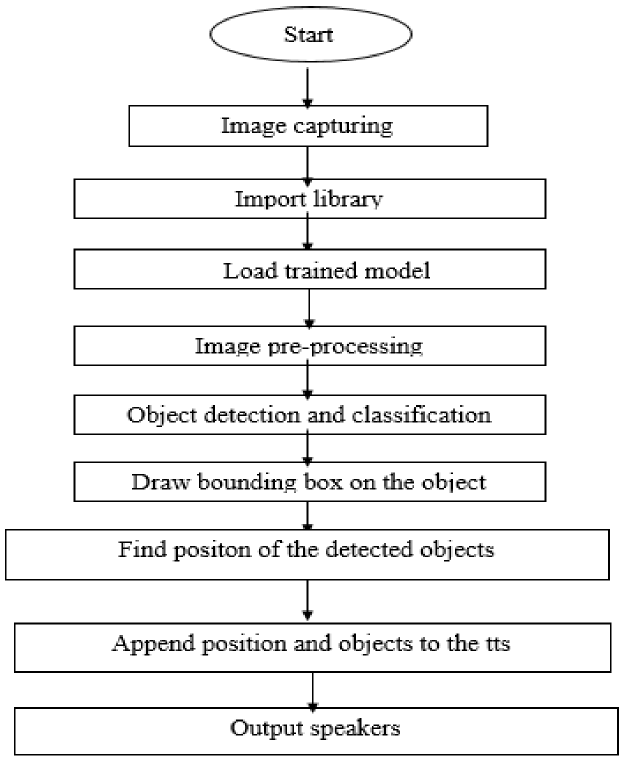

Fig -2: Workflow of the object detection algorithm.

For every person's movement in the indoor environment, the process image acquisition > image processing > acoustic notification is looped. The total processing time is calculated by adding the three processing periods, which determines the acquisition rate for the input image frames. The process must be quick enough so that possible roadblocks may be avoided on time.

The image processing method is used to detect a specific object, more specifically traffic sign recognition. We made use of the cv2 integrated OpenCV function.

Image acquisition, pre-processing, CNN model, object detection and localization, feedback generation, and user interface are all components of the implemented system architecture. Image acquisition is the process of capturing photographs of the indoor environment using a camera module (e.g., web camera, depth sensor). Image pre-processing techniques are used to improve image quality and eliminate noise. Design and training are required for the Convolutional Neural Network (CNN) model for object detection and localization. The input photos are then processed through the trained CNN to detect and localise objects, barriers, and landmarks.

The user receives auditory or haptic input regarding the detected objects and their positions. To convey information to the user, a user-friendly interface is created, which may be a mobile application or wearable device.

CNN Model Development:

The first stage is Dataset Collection, which requires amassing a broad indoor dataset that includes various settings, items, barriers, and landmarks. Precise annotations serve as the foundation for training and evaluation. In the Model Architecture phase, a Convolutional Neural Network (CNN) is painstakingly developed to solve real-time item detection, harnessing spatial features for accurate indoor object identification and positioning. The model's efficiency is improved through rigorous training using the acquired dataset and accuracy optimization. The third step, Integration, incorporates the CNN model into the system architecture, which improves real-time object recognition and localization. This improves overall functionality, allowing for more informed decisions in interior environments.

Hardware Integration:

A suitable camera module was chosen and integrated into the system to facilitate image acquisition. To effectively manage the tasks of image processing, Convolutional Neural Network (CNN) inference, and feedback generation, a microcontroller or processor was selected. To convey the outcomes of the analysis to the user, auditory and haptic output devices were incorporated. The integration of these components resulted in development of functional and robust systems.

User Testing and Validation:

To facilitate picture acquisition, a suitable camera module was selected and put into the system. A microcontroller or CPU was chosen to successfully manage the activities of image processing, Convolutional Neural Network (CNN) inference, and feedback production. Auditory and haptic output devices were used to communicate the results of the analysis to the user. The combination of these elements led in the creation of functional and robust systems.

5. RESULTS

The result indicates that the CNN (Convolutional Neural Network) program for object recognition was implemented effectively. By detecting and assisting them with the obstacle or object found, the aim is to help persons who are purblind improve their quality of life. This application can be used to differentiate between objects and assist those with disabilities, according to the proposed paradigm.

6. CONCLUSIONS

A system-based aiding network has been proposed to help purblind and fully blind people. The template that matches the procedures completed by experimenting with OpenCV has developed a successful multiscale and useful method for the applications employed within the environment. Finally, the identified items are output as an auditory message with the object's name. The clarity of the image obtained by the user will determine the accuracy. The real-time implementation of the technology is promising, providing real-world benefits for visually impaired persons traversing indoor settings. Our technology has the potential to become an indispensable tool, enabling greater freedom and inclusivity for visually impaired individuals, with user-centric input and incremental enhancements. This project shows the positive impact of cutting-edge technology and opens the way for future advances in assistive systems, making the world more accessible to all.

REFERENCES

[1] Samkit Shah, Jayraj Bandariya, Garima Jain, Mayur Ghevariya, Sarosh Dastoor, “CNN based Auto-Assistance System as a Boon for Directing Visually Impaired Person”, 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) 23-25 April 2019, DOI: 10.1109/ICOEI.2019.8862699.

[2] Bin Jiang, Jiachen Yang, Zhihan Lv, Houbing Song, “Wearable Vision Assistance System Based on Binocular Sensors for Visually Impaired Users”, IEEE Internet of Things Journal (Volume: 6, Issue: 2, 31 May 2018), DOI: 10.1109/JIOT.2018.2842229.

[3] Md. Milon Islam, Muhammad Sheikh Sadi, and Thomas Bräunl, "Automated Walking Guide to Enhance the Mobility of Visually Impaired People", IEEE Transactions on Medical Robotics and Bionics (Volume: 2, Issue: 3, August 2020), DOI: 10.1109/TMRB.2020.3011501.

[4] Devashish Pradeep Khairnar, Rushikesh Balasaheb Karad, Apurva Kapse, Geetanjali Kale, Prathamesh Jadhav, "PARTHA: A Visually Impaired Assistance System", 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), DOI: 10.1109/CSCITA47329.2020.9137791.

[5] Ajinkya Badave, Rathin Jagtap, Rizina Kaovasia, Shivani Rahatwad, Saroja Kulkarni, "Android Based Object Detection System for Visually Impaired", 2020 International Conference on Industry 4.0 Technology (I4Tech), DOI: 10.1109/I4Tech48345.2020.9102694.

[6] Ruiqi Cheng, Kaiwei Wang, Longqing Lin and Kailun Yang, "Visual Localization of Key Positions for Visually Impaired People", 2018 24th International Conference on Pattern Recognition (ICPR), DOI: 10.1109/ICPR.2018.8545141.

[7] Payal T. Mahida, Seyed Shahrestani, Hon Cheung, "Localization Techniques in Indoor Navigation System for Visually Impaired people", 2017 17th International Symposium on Communications and Information Technologies (ISCIT), DOI: 10.1109/ISCIT.2017.8261229.

[8] Vidula V. Meshram, Kailas Patil, Vishal A. Meshram and Felix Che Shu, "An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People", IEEE Transactions on Human-Machine Systems (Volume: 49, Issue: 5, October 2019), DOI: 10.1109/THMS.2019.2931745.

[9] Vidya N. Murali, James M. Coughlan, "Smartphone-based crosswalk detection and localization for visually impaired pedestrians", 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), DOI: 10.1109/ICMEW.2013.6618432.

[10] Jawaid Nasreen, Warsi Arif, Asad Ali Shaikh, Yahya Muhammad, Monaisha Abdullah, "Object Detection and Narrator for Visually Impaired People", 2019 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), DOI: 10.1109/ICETAS48360.2019.9117405.

[11] Cang Ye, Xiangfei Qian, "3D Object Recognition of a Robotic Navigation Aid for the Visually Impaired", IEEE Transactions on Neural Systems and Rehabilitation Engineering (Volume: 26, Issue: 2, February 2018), DOI: 10.1109/TNSRE.2017.2748419.

[12] Akshaya Kesarimangalam Srinivasan, Shwetha Sridharan, Rajeswari Sridhar, "Object Localization and Navigation Assistant for the Visually challenged", 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), DOI: 10.1109/ICCMC48092.2020.ICCMC-00061.

[13] Joel A. Hesch and Stergios I. Roumeliotis, "An Indoor Localization Aid for the Visually Impaired", Proceedings 2007 IEEE International Conference on Robotics and Automation, DOI: 10.1109/ROBOT.2007.364021.

[14] K. Matusiak, P. Skulimowski, P. Strurniłło, "Object recognition in a mobile phone application for visually impaired users", 2013 6th International Conference on Human System Interactions (HSI), DOI: 10.1109/HSI.2013.6577868.

[15] Nouran Khaled, Shehab Mohsen, Kareem Emad El-Din, Sherif Akram, Haytham Metawie, Ammar Mohamed , "In-Door Assistant Mobile Application Using CNN and TensorFlow", 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), DOI: 10.1109/ICECCE49384.2020.9179386.

[16] Hsueh-Cheng Wang, Robert K. Katzschmann, Santani Teng, Brandon Araki, Laura Giarré, Daniela Rus, "Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System", 2017 IEEE International Conference on Robotics and Automation (ICRA), DOI: 10.1109/ICRA.2017.7989772.