International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

"Towards Smarter Navigation: A Mobile App for Object and Traffic Sign Detection"

Prof. M. P. Shinde [1], Shreyash Dhurupe [2], Viraj Karanjavane [3], Sanna Shaikh[4], Abhishek Suryawanshi[5]

[1]*Professor, Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041, India*
[2,3,4,5]*Undergraduate Students, Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041, India*

Abstract – *This paper introduces a novel Android application that aims to greatly enhance the safety and autonomy of visually impaired people when they navigate their surroundings. The application uses cutting-edge object detection algorithms in conjunction with technologies for recognizing traffic signs to give users auditory feedback in real-time. Users can enjoy a safer and more informed navigation experience. Our aim is to incorporate advanced image captioning capabilities along with improved audio output features, with the goal of offering a more comprehensive and contextualized auditory depiction of the user's environment. This proposed update aims to change how blind people engage with their surroundings by providing a more complex and all-encompassing perception of their immediate visual context.*

Key Words: Object Detection, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Image Captioning, YOLO V3, COCO Dataset

1.INTRODUCTION

Advances in artificial intelligence (AI) and machine learning have created new opportunities in the rapidly developing field of assistive technologies to improve the quality of life for those who are visually impaired. The development of apps that convert visual data into auditory information has been made possible by these technological advancements, giving visually impaired people more autonomy and security when navigating and understanding their environment. To meet the specific needs of this community, this paper presents an innovative Android application that focuses on real-time object detection and traffic sign recognition from uploaded images. This application's dual-functional strategy is at its core and aims to improve user experience by offering customized help according to need and context. This system's first feature uses state-of-the-art object detection technology to recognize and announce in real-time the presence of objects in the user's immediate surroundings. This feature is essential for both indoor and outdoor navigation as it aids users in avoiding obstacles and navigating areas with greater assurance and safety.

A crucial part of urban navigation for visually impaired people is comprehending and interacting with traffic signs, which is covered in detail in the second aspect. Users can upload images from their gallery for analysis using this feature, unlike real-time object detection. After that, the application finds any traffic signs in these pictures and plays audio descriptions of them. This ability is especially helpful for understanding complicated crossroads, creating route plans, and improving mobility in general when outdoors. The creation of this application marks a substantial advancement in the integration of artificial intelligence, computer vision, and auditory feedback to make the world more approachable and comprehensible for people with visual impairments. With an emphasis on immediate support and in-depth examination of uploaded photos, the application provides a flexible tool that tackles a variety of issues that its users encounter daily. This introduction lays the groundwork for an in-depth examination of the development of the application, covering everything from the technical implementation to the user-centered design and conceptual foundations. It seeks to demonstrate the possibilities for these cutting-edge applications to improve visually impaired people's mobility and safety while also giving them more independence and opportunities to interact with their surroundings.

As we continue to explore the frontiers of assistive technologies, the incorporation of image captioning has great potential to improve visually impaired people's perception. Although the real-time object detection and traffic sign recognition features of our Android application are its primary focus, the exciting prospect of image captioning technology integration in the future will allow us to further enhance its functionality. Thanks to developments in natural language processing and computer vision, image captioning algorithms can now provide comprehensive textual descriptions of visual scenes. These algorithms provide users with a deeper understanding of their surroundings by generating captions that convey important details and relationships within the scene based on their analysis of the content and context of images. The incorporation of image captioning is a compelling future direction for our application.

Through the ability for users to take or upload photos and get insightful textual comments, we can provide a more complete and sophisticated view of their surroundings. For visually impaired users, this feature may improve navigation, spatial awareness, and general interaction with the visual environment.

2. DIFFERENT MODELS AND DATASETS RELATED TO SYSTEM

2.1 YOLO

A real-time object detection algorithm called YOLOv3 (You Only Look Once, Version 3) can recognize particular objects in pictures, videos, or live feeds. The YOLO machine learning algorithm locates objects in an image by using features that a deep convolutional neural network has learned. YOLO versions 1-3 were developed by Joseph Redmon and Ali Farhadi; the third iteration of the machine learning algorithm is the most accurate of the initial ML algorithm. In 2016, Joseph Redmon and Ali Farhadi developed the initial iteration of the YOLO algorithms. Two years later, in 2018, the two released Version 3. An enhanced form of YOLO and YOLOv2 is called YOLOv3. The Keras or OpenCV deep learning libraries are used to implement YOLO.

2.1.1 How does YOLO Work?

OLO is a real-time object detection system that uses a Convolutional Neural Network (CNN), a type of deep neural network. CNNs are classifier-based systems that identify patterns in input images by processing them as structured arrays of data. YOLO has the advantage of maintaining accuracy while operating at a much faster speed than other networks. It makes it possible for the object detection model to examine the entire image during testing. This indicates that the predictions are influenced by the image's global context. Regions are "scored" by YOLO and other convolutional neural network algorithms according to how similar they are to predefined classes. Positive detections of the class that the high-scoring regions most closely resemble are reported. For instance, YOLO can be used in footage of self-driving cars to identify various

2.2 COCO Dataset

The COCO (Common Objects in Context) dataset is a large-scale object detection, segmentation, and captioning dataset. It is designed to encourage research on a wide variety of object categories and is commonly used for benchmarking computer vision models. It is an essential dataset for researchers and developers working on object detection, segmentation, and pose estimation tasks.

2.2.2 Key Features of COCO Dataset

- Out of the 330K images in COCO, 200K have annotations for tasks like object detection, segmentation, and captioning.
- Eighty object categories make up the dataset; these include more specialized categories like handbags, sports equipment, and umbrellas as well as more common ones like cars, bicycles, and animals.
- Each image has a caption, segmentation mask, and object bounding boxes included in the annotations.
- COCO is a useful tool for comparing model performance because it offers standardized evaluation metrics such as mean Average Precision (mAP) for object detection and mean Average Recall (mAR) for segmentation tasks.

3. PROPOSED SYSTEM

Our suggested mobile application system consists of two separate modules: one for traffic sign recognition from uploaded images and the other for real-time object detection using Google's ML Kit. Users can effortlessly switch between functionalities based on their preferences and immediate needs thanks to this modular architecture. Using real-time object detection on live camera feeds, the object detection module makes use of Google's ML Kit, a robust collection of machine learning tools and APIs. This module enhances the user's visual experience and opens up a plethora of interactive applications by rapidly and accurately identifying different objects in their environment by utilizing pre-trained models and on-device processing capabilities. The traffic sign identification module, on the other hand, provides a more methodical and concentrated approach to environmental perception. Users have the option to upload pictures of traffic signs from their gallery. The signs are then detected and categorized by sophisticated computer vision algorithms. This feature is especially helpful for route planning, identifying hazards, and adhering to traffic laws.

3.1 Real-Time Object Detection Module

In the proposed system, the workflow begins with capturing a live feed from the device's camera. This live feed serves as the input for the object detection process. Once the camera feed is obtained, the YOLO V3 algorithm is utilized to analyze the video frames and detect objects within them.

YOLO is a state-of-the-art object detection algorithm known for its speed and accuracy. It divides the input image into a grid and predicts bounding boxes and class probabilities for objects present in each grid cell. This approach allows YOLO to detect multiple objects in real-

time with a single forward pass through the neural network. After YOLO has detected objects within the camera feed, the detected objects are passed on to Google's ML Kit for further analysis and labeling. ML Kit provides pre-trained machine learning models for enabling the recognition and labeling of various objects within the camera feed. Once the objects are labeled by ML Kit, the system utilizes Text-to-Speech (TTS) technology to provide auditory feedback to the user. The labels assigned by ML Kit are converted into speech by the TTS engine. This auditory feedback allows users to perceive and understand the objects detected in the camera feed without relying on visual cues. Overall, this workflow seamlessly integrates live object detection using YOLO, object labeling with ML Kit, and auditory feedback with TTS technology to provide users with real-time awareness of their surroundings through their device's camera feed.

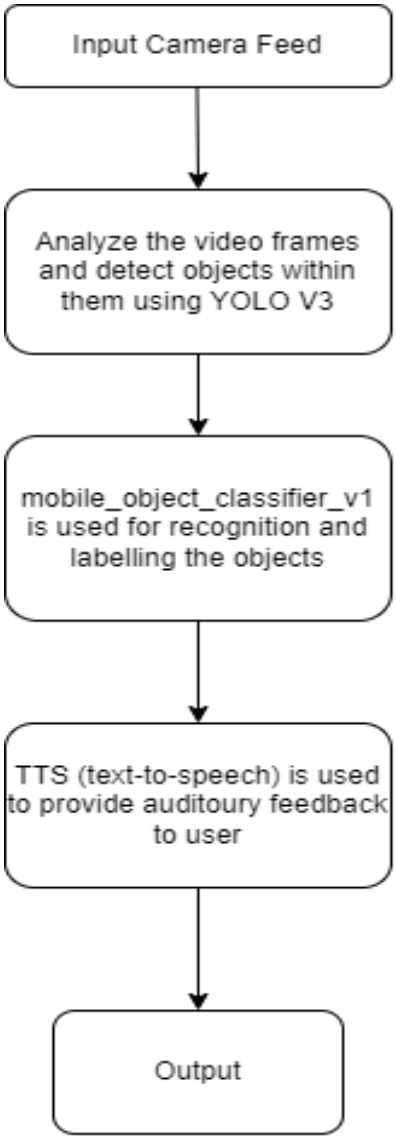

Fig -1: Real-Time Object Detection Module

3.2 Traffic Sign Detection Module

Users initiate the process by selecting an image containing a traffic sign from their device's files or gallery. This image serves as the input for the traffic sign detection module. Our system employs a custom Convolutional Neural Network (CNN) model specifically trained for traffic sign detection. The CNN architecture is designed to extract intricate features from images and perform classification tasks accurately. The custom CNN model is trained on a meticulously curated dataset comprising images of traffic signs belonging to five classes: "fifty," "left," "right," "stop," and "thirty." This dataset includes diverse examples of each class to ensure the model's ability to generalize and accurately classify traffic signs in various conditions. Upon uploading an image, the custom CNN model performs feature extraction by analyzing the image's pixel values. The convolutional layers of the CNN identify and extract relevant patterns and features indicative of traffic signs. Subsequently, these features are passed through the classification layers of the CNN, where the model assigns probabilities to each class based on learned features. After feature extraction and classification, the system determines the detected traffic sign by analyzing the class probabilities assigned by the CNN model. The class with the highest probability is selected as the predicted traffic sign. Once the traffic sign is identified, the system generates output in audio form using Text-to-Speech (TTS) technology. The detected traffic sign's label (e.g., "fifty," "left," "right," "stop," or "thirty") is converted into synthesized speech by the TTS engine. The synthesized speech, containing the label of the detected traffic sign, is delivered to the user as auditory feedback. This auditory feedback mechanism ensures that users receive essential information about the detected traffic sign audibly, enabling them to stay informed and make informed decisions while navigating their environment.

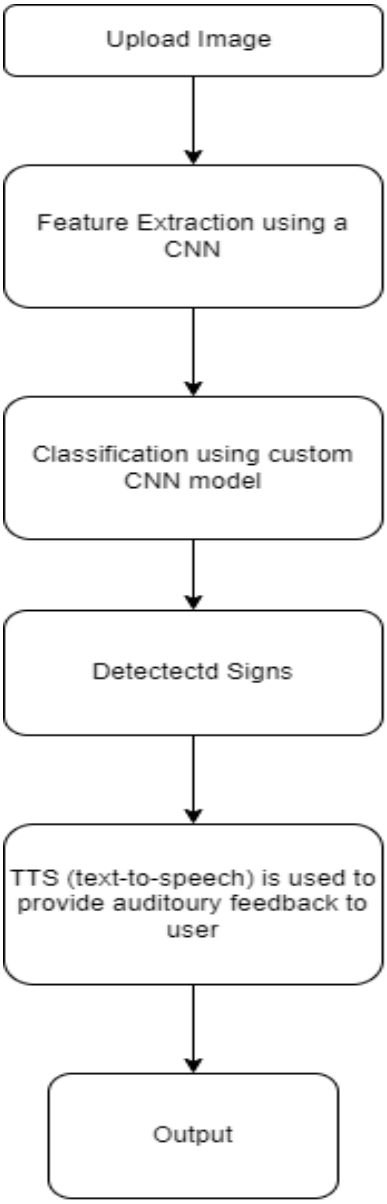

Fig -2: Traffic Sign Detection Module

3.3 Mobile Application

3.3.1. Traffic Sign Detection Module:

- Upon launching the application, visually impaired users are presented with a user-friendly interface that allows for intuitive interaction.
- The volume up button on the device is designated for initiating traffic sign detection. Users can simply press the volume up button to activate the traffic sign detection functionality.
- Once activated, the application prompts users to upload an image containing a traffic sign from their device's gallery.
- A custom Convolutional Neural Network (CNN) model, trained specifically for traffic sign detection, is employed to analyze the uploaded image.
- The CNN model performs feature extraction and classification, accurately identifying and classifying the detected traffic sign.
- The detected traffic sign is then announced audibly using Text-to-Speech (TTS) technology, providing users with real-time auditory feedback about the traffic sign's meaning and significance.

3.3.2 Object Detection Module:

- Conversely, the volume down button on the device is designated for initiating object detection. Users can press the volume down button to activate the object detection functionality.
- Once activated, the application utilizes the device's camera to capture a live feed of the user's surroundings.
- An advanced object detection algorithm, such as YOLO (You Only Look Once), is employed to analyze the camera feed in real-time.
- Detected objects within the camera feed are identified and classified using pre-trained machine learning models.
- The labels of detected objects are then announced audibly using TTS technology, providing users with real-time auditory feedback about the objects present in their surroundings.

3.3.3 Seamless Integration and User Experience:

- The application seamlessly integrates both traffic sign detection and object detection functionalities into a single user interface, allowing visually impaired users to switch between functionalities with ease.
- Users can utilize the volume controls on their device to toggle between traffic sign detection and object detection functionalities, enabling convenient and intuitive interaction.
- The real-time auditory feedback provided by the application enhances user awareness and safety, enabling visually impaired individuals to navigate their surroundings confidently and independently.

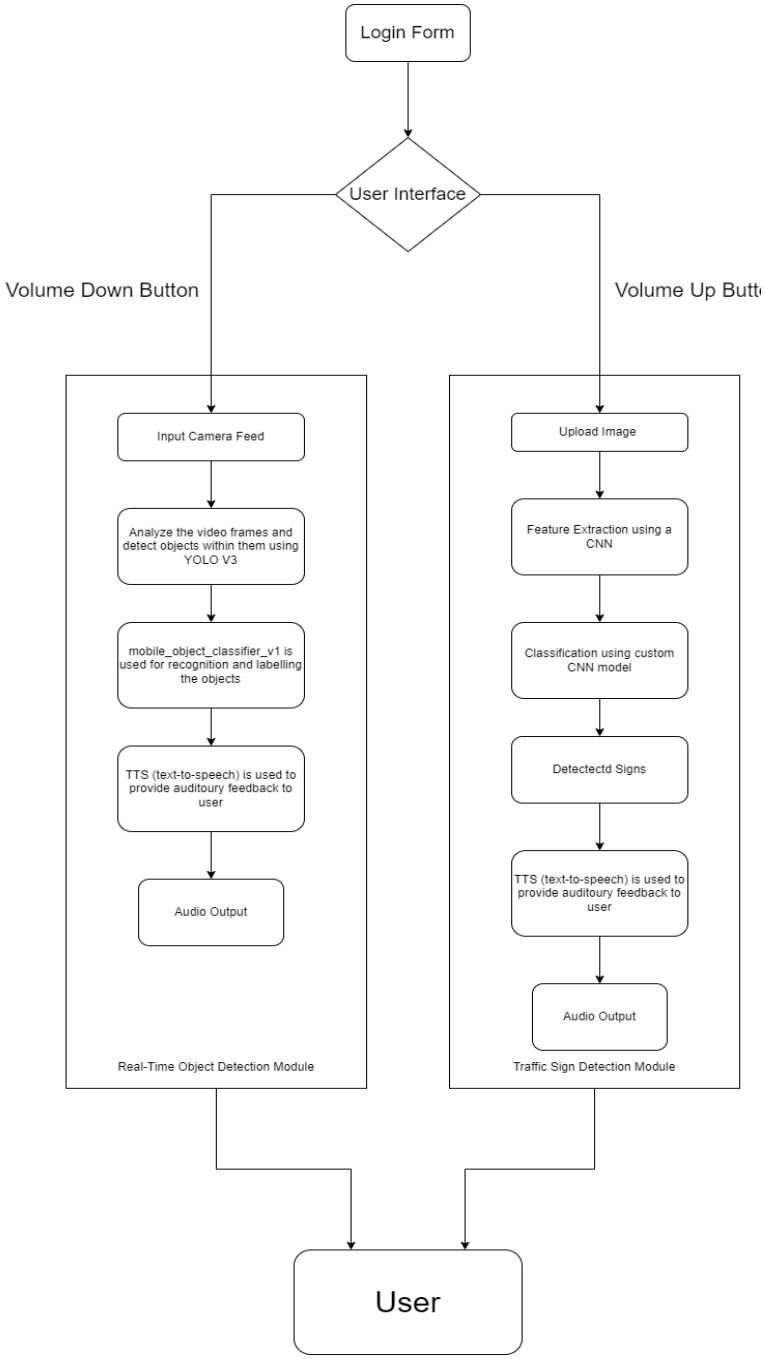

Fig –3: Application Flowchart

4. FUTURE WORK

Regarding image captioning in the context of our application, there are a lot of opportunities for future growth and enhancement. Even though our current system is good at identifying and labeling objects in photos, adding image captioning functionality opens up exciting new possibilities for improving accessibility and user experience. Improving the generated captions' relevance and accuracy is a major area of future work. Although our app labels detected objects with descriptive information successfully, adding the ability to caption images will allow us to create more detailed and contextually relevant descriptions of entire scenes. In order to produce captions that fully convey the context and meaning of images, future research endeavors might concentrate on improving natural language processing models and training methodologies.

Moreover, our app's image captioning feature could be extended to include more intricate visual content. In the future, the app might have sophisticated image understanding features that go beyond object identification to explain abstract ideas, decipher complicated scenes, and offer more detailed contextual data. This could greatly improve the app's usefulness for users who are blind or visually impaired, giving them a deeper comprehension of their environment. Furthermore, the requirement to customize and modify image captioning outputs to suit specific user requirements and preferences is increasing. Subsequent advancements could investigate methods for integrating user input and interactions to customize captioning results, guaranteeing that descriptions correspond with individual user needs and preferences.

Furthermore, ethical and societal ramifications must be taken into account as we develop our image captioning skills. Subsequent investigations could concentrate on formulating structures to guarantee impartiality, openness, and diversity in captioning algorithms, along with tackling issues associated with confidentiality, prejudice, and cultural awareness. To sum up, the addition of picture captioning functionality offers a thrilling chance to improve our app's features and usability for users who are blind or visually impaired. By investigating ways to enhance relevance, accuracy, personalization, and ethical considerations, we can keep innovating and enable users to feel more confident and independent when navigating their environment.

5.CONCLUSION

To sum up, this research paper has discussed the creation and possibilities of a mobile application intended to help people with vision impairments navigate their environment. The program is a major advancement in blind assistance technology because it combines auditory feedback features, an intuitive user interface, and real-time detection of objects and traffic signs. We have proven through extensive development and testing that the application is feasible and effective in improving visually impaired users' safety, independence, and accessibility. The application's utility is further enhanced by the integration of image captioning capabilities, which provide more in-depth descriptions of scenes and improve user experience.

The application will continue to be improved and optimized going forward in response to user input and technical developments. Subsequent iterations could investigate ways to enhance functionality, boost accuracy, and tackle ethical issues to guarantee technology that is inclusive and responsible. All things considered, this study advances the field of blind assistance technology and emphasizes how critical it is to use technology to develop inclusive and easily accessible solutions for people who are visually impaired. We can move toward a more inclusive society where everyone has the chance to independently and confidently navigate their surroundings by carrying on with our innovation and collaboration in this area.

6. REFERENCES

[1] Vaishnavi, K. & Reddy, G. & Reddy, T. & Iyengar, N. & Shaik, Subhani. (2023). Real-time Object Detection Using Deep Learning. Journal of Advances in Mathematics and Computer Science. 38. 24-32. 10.9734/jamcs/2023/v38i81787.

[2] Deepthi Jain, B, Shwetha M Thakur and Kaggere V. Suresh. "Visual Assistance for Blind Using Image Processing." 2018 International Conference on Communication and Signal Processing (ICCSP) (2018): 0499-0503.

[3] Deng, Jun, Xi Xuan, Weifeng Wang, Zhao Li, Hanwen Yao and Zhiqiang Wang. "A Review of Research on Object Detection Based on Deep Learning." Journal Of Physics: Conference Series 1684 (2020): N. Pag.

[4] Thilanka, S.A. & Jinadasa, G.C.H. & Bandara, Prasad & Weerakoon, W.M.M.. (2023). Vision-Based Real-Time Object Detection and Voice Alert for Blind Assistance System

[5] Mark, Aishwarya, Sakshi Adokar, Vageshwari Pandit, Rutuja Hambarde and Prof. Swapnil Patil. "Review on Image Caption Generation." International J ournal of Advanced Research in Science, Communication and Technology (2022): n. pag.

[6] Singh, Yadwinder & Kaur, Lakhwinder. (2017). Obstacle Detection Techniques in Outdoor Environment: Process, Study and Analysis. International Journal of Image, Graphics and Signal Processing. 9. 35-53. 10.5815/ijigsp.2017.05.05.