OBJECT DETECTION FOR VISUALLY IMPAIRED DEEP LEARNING

[1]M.THARUN,[2]L.SNEHA,[3]M.MAHESH,[4]Mr.S. SRAVAN KUMAR
[123] Students, [4] Assistant Professor
Department Of Information Technology
Teegala Krishna Reddy Engineering College, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

Vision is one of the most important senses that help people interact with the real world. There are nearly 200 million blind people all over the world, and being visually impaired hinders a lot of day-to-day activities. Thus, it is very necessary for blind people to understand their surroundings and to know what objects they interact with. This project proposes an Android application to help blind people see through a handheld device like a mobile phone. It integrates various techniques to build a rich Android application that will not only recognize objects around visually impaired people in real time but also provide audio output to assist them as quickly as possible. For object recognition and detection, the application uses the Single Shot Detector (SSD) algorithm as well as You Only Look Once (YOLO), both of which are well-known deep learning-based object detection models. While YOLO is recognized for its high-speed, single-pass detection capability, SSD is preferred for mobile applications due to its balance between speed and accuracy. Both algorithms ensure efficient real-time performance.The application further utilizes Android TensorFlow APIs for deep learning processing and the Android TextToSpeech API to generate audio output. By combining these technologies, the application aims to provide a seamless and efficient way for visually impaired individuals to perceive and interact with their surroundings.

I. INTRODUCTION

It is easier for vision enabled people to carry out their everyday activities since they can clearly see all the objects in their surroundings, any obstacles they come across, other people and hence is easy to interact with these objects. Whereas, visually impaired people have to struggle a lot to deal with real world due to their everyday chores and jobs. There are more than a million blind people all over the world, and being visually impaired hinders a lot of day to day activities. Thus it is very essential for blind people to know their surroundings, and what objects they interact with to prevent accidents and make their life simpler. Visually impaired people always need someone to guide them throughout their day such as to cross roads, catch a bus, and many such activities. The main motive of developing this application is to assist blind people. This application aims to help the visually impaired people to know their surrounding objects that could be just basic everyday objects or can create obstacle in their activity. The application is build to recognize or detect some household objects like chair, table, bed, refrigerator, laptops etc and some on outdoor objects like cars, motorbikes, potted plant, people etc. The application will use mobile phone camera to scan the surrounding in real time and take the frames from the ongoing video. The frames will be sent to the next module where the SSD algorithm will create bounding boxes around the objects in the frame and classify them into given categories. At last the application will produce an audio output of the object detected which has the maximum confidence score among all other present in the frame. The frames are selected at a particular time interval to avoid the hindrance in the audio output.

Purpose of the Project

The primary purpose of this project is to develop an innovative Android application that empowers visually impaired individuals to gain a richer understanding of their immediate

surroundings. By leveraging real-time object recognition and audio feedback, this application aims to mitigate the challenges posed by visual impairment in everyday activities. The application integrates state-of-the-art deep learning models, specifically Single Shot Detector (SSD) and You Only Look Once (YOLO), to efficiently and accurately identify objects in the user's environment using a standard mobile phone camera. Furthermore, through the utilization of Android TensorFlow APIs for on-device processing and the Android TextToSpeech API for instant audio narration, this project seeks to provide a seamless, intuitive, and timely way for visually impaired users to perceive and interact with the objects around them, ultimately fostering greater independence and confidence in their daily lives.

Problem Statement

Object Detection for the Visually Impaired using Deep LearningBackground: Visually impaired individuals face significant challenges in navigating their environment safely. Traditional methods of assistance often rely on human intervention, static guide systems, or auditory cues, which may not be sufficient in complex or dynamic environments. With advancements in artificial intelligence (AI) and deep learning, there is an opportunity to develop systems that can provide real-time assistance by detecting and identifying objects in the environment..

EXISTING SYSTEM

There are millions of people living in this world with difficulties in understanding the environment due to visual disabilities. Navigating around is one of the biggest challenges visually impaired people face. It is difficult for them to travel independently as they cannot analyze the position of the objects and the people surrounding them. In order to move around outdoors, visually impaired people need someone to guide them throughout. The white can is one of the most common aids for the visually impaired people. Though it is helpful in navigating around, it does not inform the user about the various obstacles until they are very close to them. Thus, due to the shortcomings of these conventional solutions, a lot of research is being done in order to develop better and advanced aids to assist the visually impaired people.

Disadvantages of existing system:

The existing methods for assisting visually impaired individuals in navigating their surroundings present several significant disadvantages. Primarily, the reliance on human guides severely limits independence and spontaneity, while the most common aid, the white cane, offers only tactile feedback for immediate obstacles at ground level. This lack of real-time, proactive information means users often become aware of hazards only when very close, hindering fluid movement and environmental awareness. Furthermore, the white cane provides no information about the identity of objects or people nearby, limiting social interaction and contextual understanding. Consequently, these limitations can lead to safety concerns in complex environments and potentially impact the user's confidence and psychological well-being, highlighting the critical need for more advanced and informative assistive technologies.

PROPOSED SYSTEM

Object detection is the principal objective of this system. It consists of object classification and object localization. Object detection is the process of categorizing the object into different classes that were defined before. In other words, object classification assigns a label to an entire image. That label is the name of the object present in that image. For instance, a computer is given an image of a cat and it will try to classify it and give the output as “Cat”. It is easy for us to identify the objects present in any image, but for a computer, object classification is a tedious task. In object localization, the computer tries to

isolate the object from the image by drawing a rectangular box around it which is also called Bounding Box. Thus, object detection is the combination of object classification and object localization in which we try to classify and isolate multiple objects present in the image. The output of this module will give us the name of the object.

Advantages of proposed system:

The proposed Android application offers significant advantages over existing methods for assisting visually impaired individuals. By integrating real-time object detection powered by sophisticated deep learning models like SSD and YOLO, the system provides users with timely and proactive information about their surroundings, going beyond the limited tactile feedback of a white cane. This enables greater independence and confidence in navigation by offering advance warnings about obstacles, identifying people, and recognizing objects of interest. The application's use of mobile phone technology makes it a potentially cost-effective and widely accessible solution. Furthermore, the audio output via the Android TextToSpeech API offers an intuitive and hands-free way for users to receive information, enhancing their ability to navigate safely and interact more effectively with the world around them. This combination of real-time awareness, object identification, and accessible technology promises a substantial improvement in the daily lives of visually impaired individuals.

Scope of the Project

The scope of this project centers on developing an Android application that empowers visually impaired users by providing real-time audio feedback about their surroundings. Utilizing the mobile device's camera, the application will employ deep learning-based object detection algorithms (SSD and/or YOLO) via the Android TensorFlow API to identify and locate common objects and people. Upon detection, the Android TextToSpeech API will instantly announce these findings to the user. The project focuses on achieving efficient, on-device real-time object recognition and clear audio output for a predefined set of relevant everyday objects, ensuring compatibility with standard Android smartphones. While advanced navigation, facial recognition, comprehensive scene understanding, multi-language support, external device integration, extensive customization, and offline model updates are beyond the current scope, this application aims to deliver a foundational yet impactful tool for enhancing the environmental awareness of visually impaired individuals.

II. LITERATURE SURVEY

The thesis from the Czech Technical University directly aligns with the core ambition of this project: leveraging the ubiquitous mobile phone camera as a tool to empower visually impaired individuals. Its focus on the Android system is particularly pertinent as your proposed application is also Android-based. The thesis's exploration extends beyond mere object detection, delving into specific assistive applications like banknote recognition and reading text labels.

These examples illustrate the diverse potential of computer vision to address the unique challenges faced by the visually impaired community. The development of a novel banknote recognition algorithm using BRISK descriptors and Gradient Boosted Trees Classifier underscores the value of tailored solutions for specific recognition tasks, suggesting that while general object detection is valuable, specialized modules could enhance the application's utility. Furthermore, the thesis's emphasis on creating an accessible mobile user interface for Google's real-time text recognition library is a critical consideration for your project.

Simply having powerful technology is insufficient; it must be presented and interacted with in a way that is intuitive and navigable for visually impaired users, likely relying heavily on

audio feedback as proposed in your application. The mention of a simple camera and image gallery application within the thesis suggests a holistic approach to enabling visually impaired users to interact with visual information through their mobile devices.

The introduction of the Single Shot Detector (SSD) methodology provides a strong technical justification for your choice of object detection algorithm. The key advantage of SSD lies in its ability to perform both object localization and classification within a single deep neural network (DNN) pass. This "single-shot" approach contrasts with earlier two-stage detectors that first propose regions of interest and then classify them, leading to greater computational efficiency and faster inference speeds. This efficiency is paramount for mobile applications where processing power and battery life are limited resources. The concept of discrete bounding box outputs and the use of multiple default boxes with varying aspect ratios and scales at each feature map are central to SSD's effectiveness. These default boxes act as anchors, allowing the network to simultaneously predict the presence of objects of different shapes and sizes. During prediction, the network outputs confidence scores for each object category within each default box, along with adjustments to refine the box's position and size to accurately enclose the detected object.

The performance metrics cited – 74% mean Average Precision (mAP) at 59 Frames Per Second (FPS) on an Nvidia Titan X with a 300x300 input, and 76% mAP with a 512x512 input – demonstrate a compelling balance between accuracy and speed, particularly at the lower resolution more suitable for mobile devices. The comparison highlighting SSD's superior accuracy even with smaller input sizes further strengthens its suitability for your Android application.

The findings from the literature survey have significant implications for the design and development of your Android application.

The Czech Technical University thesis emphasizes the importance of a user-centric approach. The successful implementation of your application will depend not only on accurate object recognition but also on how effectively the information is conveyed to the visually impaired user. The reliance on audio output, as proposed, aligns with the need for an accessible interface. Future development could explore more nuanced audio cues, spatialized audio, or haptic feedback to provide richer environmental information. The thesis's work on banknote and text recognition also suggests potential future extensions of your application beyond basic object identification. Imagine the added value of being able to identify currency or read aloud text from signs or documents using the same core technology.

The selection of SSD as a primary object detection model offers a strong foundation for achieving real-time performance on a mobile device. However, the specific implementation will require careful optimization using the Android TensorFlow Lite API. This involves converting the trained SSD model into a mobile-friendly format, potentially quantizing its weights to reduce its size and computational demands, and optimizing the inference pipeline for the specific hardware of target Android devices. Further research into the trade-offs between model size, accuracy, and inference speed will be crucial during the development process. Exploring different pre-trained SSD models or even fine-tuning a model on a dataset specifically curated for objects commonly encountered by visually impaired individuals in surrounding areas could further enhance the application's effectiveness. Consider the potential for incorporating user feedback to iteratively improve the model's accuracy and relevance.

Beyond SSD, other related object detection architectures and techniques could be explored for future iterations or specific use cases. For instance, the YOLO (You Only Look Once) architecture, also mentioned in your initial description, is known for its speed and could be investigated for scenarios where even faster processing is required, potentially at the cost of some accuracy. More recent advancements in lightweight convolutional neural networks (CNNs) designed for mobile devices could also offer improved efficiency.

Furthermore, incorporating techniques like contextual understanding or spatial reasoning could enhance the application's ability to provide more meaningful information about the user's surroundings. For example, knowing that a "chair" is detected near a "table" provides more context than just identifying two isolated objects.

Building upon this foundation, several avenues for future research and development emerge.

One key area is the creation of more sophisticated and context-aware audio feedback mechanisms. Instead of simply announcing the name of an object, the application could provide more descriptive information, such as the object's approximate distance, orientation, or its relationship to other detected objects. Integrating spatial audio could allow users to perceive the direction of detected objects more intuitively. Haptic feedback could complement audio cues, providing tactile information about nearby obstacles or objects of interest.

Another important direction is the expansion of the application's object recognition capabilities. This could involve incorporating a wider range of everyday objects, as well as specialized objects relevant to specific environments or user needs local landmarks, transportation options, or commonly used tools. Continuous data collection and model retraining, potentially incorporating user feedback, will be essential for improving the accuracy and comprehensiveness of the object recognition. Exploring transfer learning techniques, where a model pre-trained on a large general dataset is fine-tuned on a smaller, more specific dataset, could be an efficient way to expand the application's vocabulary.

Furthermore, integrating the object recognition capabilities with other assistive technologies could significantly enhance the application's utility. For example, combining object detection with indoor navigation systems or GPS for outdoor navigation could provide a more holistic solution for environmental awareness. Exploring integration with smart glasses or other wearable devices could also offer a more seamless and hands-free user experience.

The broader context of assistive technology for the visually impaired is a rapidly evolving field. Advancements in artificial intelligence, sensor technology, and human-computer interaction are constantly creating new possibilities. Your project contributes to this important area by exploring the potential of mobile computer vision to empower visually impaired individuals. By focusing on real-time, accessible feedback, your application has the potential to significantly improve independence and quality of life. Continued research, user feedback, and iterative development will be crucial for realizing the full potential of this technology. Collaborations with organizations supporting the visually impaired community in Bhimavaram and beyond could provide valuable insights and ensure the application meets the specific needs of its target users.

III. SYSTEM DESIGN

SYSTEM ARCHITECTURE

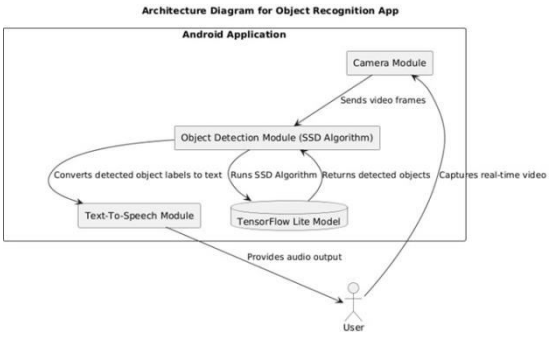


FIG: SYSTEMARCHITECTURE

IV. MODULE DESCRIPTION

Dataset:

Our Android application uses Neural Networks for object recognition. This requires an image dataset of the objects to train the classifier. In this project we have used COCO (Common Objects in Context) 2014 Database with 80 different object classes which have 83K training images, 41K Testing images. The dataset used is the labeled dataset which is useful to train the model. Some of the objects among 80 classes are as follows:

- **Neural Network Requirement:** It correctly identifies that neural networks, the core of modern object recognition systems, necessitate a substantial image dataset for effective training. The network learns to identify patterns and features associated with different object classes by analyzing a large number of labeled examples.
- **COCO 2014 Selection:** The choice of the COCO (Common Objects in Context) 2014 database is a strong one. COCO is a widely adopted and respected dataset in the computer vision community due to its:
- **Scale:** With 83,000 training images and 41,000 testing images, it provides a significant amount of data for robust model training, enabling the network to learn diverse representations of the 80 object classes.

Data Preparation

The fundamental first step in utilizing this dataset: downloading it directly from its official source, cocodataset.org. This action guarantees access to the standard, unaltered version of the COCO 2014 data, including both the images and their corresponding annotation files. This official source ensures the integrity and consistency of the data, which is paramount for reliable model training and evaluation. This downloaded data serves as the raw material that will undergo further processing to be compatible with our chosen deep learning models and training pipeline, ultimately enabling the object recognition capabilities of our Android application designed to assist visually impaired individuals in navigating their environment in areas.

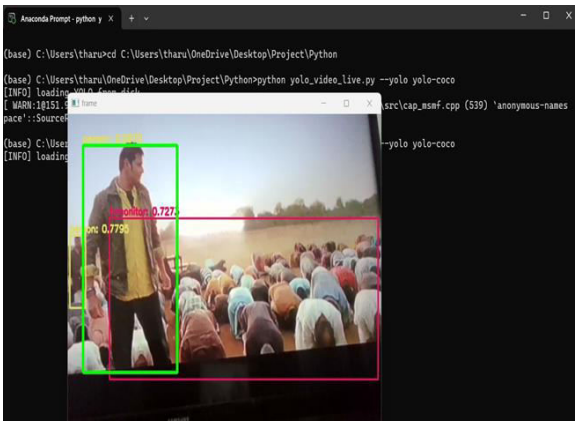
Data Labeling

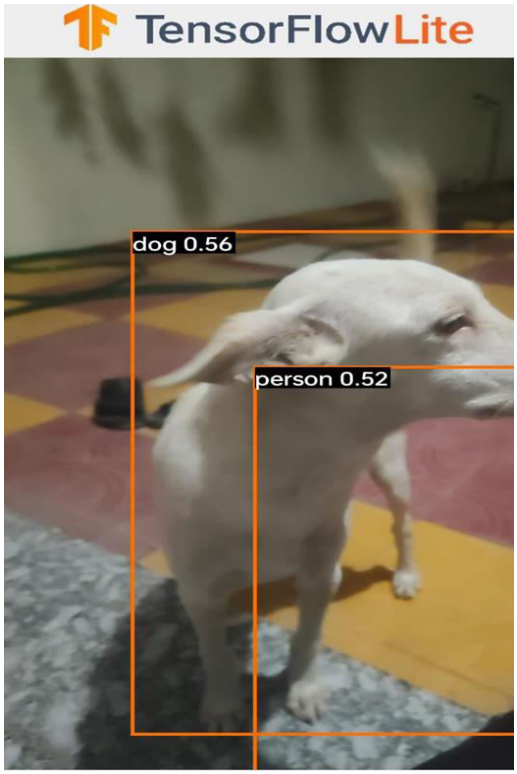
The module details the crucial process of ensuring that the images in the COCO dataset are correctly associated with information about the objects they contain. While a significant portion of the COCO dataset comes with pre-existing annotation files, the mention of using LabelImg software suggests a potential need for manual intervention in the labeling process. This could involve verifying the accuracy of the downloaded annotations, refining the bounding boxes around objects for greater precision, or even labeling additional images if the project's scope expands beyond the standard COCO annotations to include more specific objects relevant to the context.

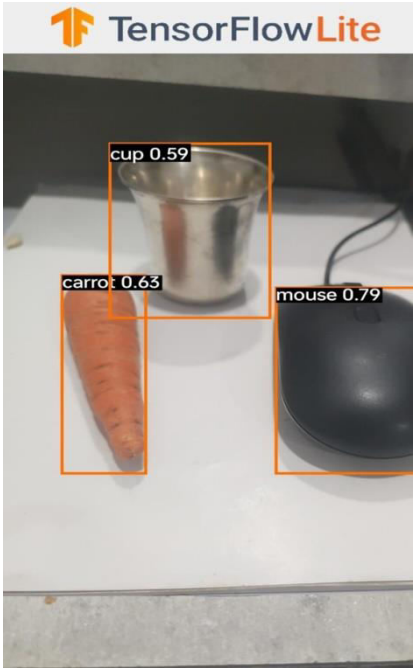
The annotation files themselves contain structured information that is essential for training object detection models. The object_class parameter specifies the category of the detected object, providing the semantic meaning for the visual information. The unique object_id allows for distinguishing between individual instances of the same object within a single image, which can be important for certain training methodologies or for future tracking features. The parameters x_coordinate for centre and y_coordinate for centre define the central location of the object within the image frame, while width and height specify the spatial extent of the object. These four parameters together define the precise bounding box that encloses each identified object. This detailed spatial information is critical for training the model to not only classify objects but also to accurately localize them within the image, which is the core task of object detection that your Android

application relies upon to identify surroundings for visually impaired users in an area. The accuracy and consistency of these labels directly impact the performance of the trained models; any errors or ambiguities in the labeling can lead to the model learning incorrect associations and ultimately affecting the reliability of the object recognition in the final application.

V. OUTPUT SCREENS







VI. CONCLUSION

Visually impaired individuals face numerous challenges in recognizing and interacting with their environment—whether it be identifying everyday household items or navigating safely in outdoor settings like roads and public spaces. Although Braille provides a valuable means for reading and communication, it does not address the need for real-time object awareness and situational understanding in dynamic environments.

To bridge this critical gap, our project presents a robust and intuitive Android application specifically designed to assist visually impaired users in perceiving their surroundings more effectively. Leveraging state-of-the-art object detection algorithms such as YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector), the app is capable of detecting and recognizing objects in real-time with high accuracy. Once an object is identified, the application instantly delivers audio feedback by announcing the object's label along with a confidence score, thus enabling users to respond appropriately and confidently.

This intelligent system transforms a basic camera feed into a powerful assistive tool,

allowing users to hear what they cannot see. Whether it's recognizing a chair in a room, identifying a passing vehicle, or locating a doorway, the application provides meaningful information that enhances autonomy, safety, and overall quality of life. By combining deep learning, real-time processing, and audio guidance through text- to-speech, the application stands as a practical, accessible, and impactful solution for the visually impaired community.

Moving forward, this project has the potential to evolve further with additional features such as scene understanding, obstacle distance estimation, multi-language support, and integration with wearable devices. Ultimately, it aims to empower visually impaired users by offering them a greater sense of independence, spatial awareness, and confidence in their everyday lives.

REFERENCES

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),779–788.https://arxiv.org/abs/1506.02640
This paper introduces the YOLO algorithm, which is widely used for real-time object detection. It is one of the core technologies integrated into our application for rapid and efficient recognition.

[2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *SSD: Single Shot MultiBox Detector*. European Conference on Computer Vision (ECCV), 21–37. https://arxiv.org/abs/1512.02325
SSD offers a good trade-off between speed and accuracy, making it suitable for mobile-based real-time applications. This algorithm forms the backbone of our object detection model optimized for Android.

[3] TensorFlow Lite. (2023). *TensorFlow Lite Documentation*. TensorFlow Developers. https://www.tensorflow.org/lite
TensorFlow Lite is the framework used to deploy machine learning models on Android devices. It enables efficient deep learning computation suitable for real-time mobile applications.

[4] Android Developers. (2023). *TextToSpeech | Android Developers*. Android API Reference. https://developer.android.com/reference/android/speech/tts/TextToSpeech
TextToSpeech API is used to convert detected objects into spoken audio. It plays a vital role in communicating object recognition results to visually impaired users.

[5] OpenCV. (2023). *Open Source Computer Vision Library*. OpenCV Organization. https://opencv.org/
OpenCV is used for image pre-processing and manipulation. It supports capturing camera frames and preparing input data for the detection models.

[6] Microsoft. (2023). *Seeing AI – Talking Camera App for the Blind Community*. Microsoft AI. https://www.microsoft.com/en-us/ai/seeing-ai
Microsoft's Seeing AI app is a benchmark in assistive technology for the blind. It inspired the concept of integrating real-time audio feedback with object detection.

[7] WHO. (2019). *World Report on Vision*. World Health Organization. https://www.who.int/publications/i/item/world-report-on-vision
This report highlights the global challenges faced by visually impaired people. It provides the motivation and context for the development of this assistive application.

[8] Pyttsx3 (n.d.). *pyttsx3 · Text-to-Speech Conversion Library in Python*. Read the Docs. https://pyttsx3.readthedocs.io/
Pyttsx3 is used in the prototype testing phase for offline voice feedback. It supports text-to-speech without internet, useful for standalone Python-

based testing.

[9] Tan, M., & Le, Q. V. (2020). *EfficientDet: Scalable and Efficient Object Detection*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10781–10790. https://arxiv.org/abs/1911.09070
EfficientDet introduces a scalable model for edge devices with high accuracy. Though not used directly, it serves as a reference for future improvements.

[10] Google Developers. (2023). *ML Kit for Android*. Google Machine Learning. https://developers.google.com/ml-kit
ML Kit provides machine learning APIs that can be integrated into Android apps. It is considered for future enhancement of detection and OCR capabilities.