International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Assistive Object Recognition and Tracking System for the Visually Impaired Using CNN

[1] **Prof. Shwetha L**, Assistant professor, [2] **Gowthami T S**, Student, [3] **Madhushree B**, Student, [4]**Reshmashree M C**, Student
Department of Artificial Intelligence and Machine Learning Engineering
G Madegowda Institute of Technology, Bharathinagara, Maddur TQ,
Mandya dist-571422, India

Abstract - The visually impaired community faces significant challenges in perceiving surrounding objects, reading textual information, and safely navigating through unfamiliar environments. Although assistive tools such as white canes and guide dogs provide partial support, they lack the capability to deliver real-time environmental awareness and contextual understanding. Recent advancements in computer vision and deep learning have opened new possibilities for developing intelligent assistive systems that can enhance independent mobility and situational awareness for visually impaired individuals.

This paper proposes an Assistive Object Recognition and Tracking System for the Visually Impaired using Convolutional Neural Networks (CNN). The proposed system utilizes deep learning–based object detection, optical character recognition (OCR), and road or lane detection models to analyze the surrounding environment through a mobile device camera. The system detects and identifies objects, reads textual information such as signboards and labels, and recognizes road or lane patterns to support safe navigation. Spatial analysis techniques are applied to estimate object direction and distance, enabling meaningful feedback.

The system operates in both offline and online modes. Offline processing is achieved using optimized TensorFlow Lite models embedded within the mobile application, ensuring reliable operation without internet connectivity. When online connectivity is available, the system retrieves additional contextual information related to detected objects to enhance user understanding. Voice-based interaction and text-to-speech output are integrated to provide real-time auditory guidance, making the system accessible and user-friendly.
By combining CNN-based vision models, spatial awareness, and voice interaction, the proposed assistive system improves environmental perception, enhances navigation safety, and promotes independent living for visually impaired users.

***Key Words*:** Assistive Technology, Object Recognition, Convolutional Neural Networks, Computer Vision, Visual Impairment, Deep Learning, Voice Assistance

I. INTRODUCTION

Visual impairment significantly affects an individual's ability to perceive and interact with the surrounding environment, limiting independent mobility and access to everyday information. Visually impaired individuals often rely on traditional assistive tools such as white canes or guide dogs to navigate their surroundings. While these tools provide basic obstacle awareness, they do not offer detailed information about object identity, spatial positioning, or textual content present in the environment.

Advancements in mobile computing and artificial intelligence have enabled the development of intelligent assistive applications that can process visual information in real time. Modern smartphones equipped with high-resolution cameras and powerful processors serve as effective platforms for deploying computer vision–based assistive solutions. However, many existing applications focus on a single task such as object detection or text recognition and fail to provide a unified system capable of comprehensive environmental understanding.
Object recognition using Convolutional Neural Networks (CNN) has demonstrated high accuracy in identifying everyday objects. CNN-based models learn hierarchical visual features, allowing robust detection under varying lighting and environmental conditions. When combined with object tracking mechanisms, these models can

continuously monitor objects in motion, providing consistent spatial feedback to users.

Textual information such as signboards, labels, and printed instructions also plays a critical role in daily navigation. Optical Character Recognition (OCR) techniques enable systems to extract and interpret text from images, allowing visually impaired users to access written information through audio feedback. Similarly, road and lane detection models assist in identifying walking paths, road boundaries, and crossings, thereby improving navigation safety.

Voice-based interaction further enhances accessibility by allowing hands-free control and real-time auditory guidance. By integrating voice input and text-to-speech output, assistive systems can communicate detected information effectively without requiring visual interaction.

This paper proposes an Assistive Object Recognition and Tracking System for the Visually Impaired using CNN that integrates object detection, OCR, and road or lane detection into a single mobile application. The system supports offline functionality through embedded deep learning models and online enhancement through external information retrieval. The objective of this work is to improve environmental awareness, navigation safety, and independent living for visually impaired individuals.

2. Body of Paper

The proposed Assistive Object Recognition and Tracking System is designed to provide real-time environmental awareness by analyzing visual input captured through a mobile device camera. The system integrates multiple deep learning models and assistive technologies to detect objects, recognize text, identify road or lane structures, and deliver meaningful auditory feedback.

The workflow of the system begins with image acquisition using the device camera. Captured frames are processed by CNN-based object detection models to identify common objects such as people, vehicles, obstacles, and daily-use items. The detected objects are tracked across consecutive frames to maintain continuity and reduce redundant announcements.

In parallel, the OCR module processes camera frames to extract readable text from the environment. This enables the system to recognize signboards, labels, and printed information. The road or lane detection module analyzes visual patterns to identify walking paths, road boundaries, and lane markings, assisting users during navigation in outdoor environments.

Spatial analysis is applied to detected objects to estimate their relative position and distance. Based on bounding box dimensions and object location, the system determines directional information such as left, right, or front and distance categories such as far, near, or very close. This spatial awareness allows the system to generate concise and context-aware voice feedback.

The system supports offline operation using optimized TensorFlow Lite models embedded within the application, ensuring functionality in environments without internet access. When network connectivity is available, the system retrieves additional contextual information related to detected objects, enhancing the quality of user guidance. The integration of multiple vision models and voice-based feedback makes the system suitable for real-world assistive applications.

II.BACKGROUND AND RELATED WORK

A. Literature Survey

Assistive technologies for visually impaired individuals have evolved significantly with advancements in computer vision and artificial intelligence. Early assistive systems primarily relied on ultrasonic sensors and basic obstacle detection mechanisms, which provided limited environmental information and lacked object identification capabilities.

With the introduction of machine learning and deep learning techniques, researchers began exploring vision-based assistive systems. CNN-based object detection models such as SSD, YOLO, and Faster R-CNN have been widely studied for real-time object recognition due to their high accuracy and efficiency. These models have been applied in assistive applications to identify obstacles and common objects in indoor and outdoor environments.
OCR technologies have also been extensively researched to support text recognition for visually impaired users. Modern OCR systems based on deep learning demonstrate improved accuracy in recognizing printed and handwritten text under varying conditions. Integrating OCR with mobile assistive applications enables users to access textual information through audio output.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Road and lane detection has been widely studied in the context of autonomous driving and intelligent transportation systems. Recent research has explored the adaptation of lane detection models for pedestrian navigation, enabling visually impaired users to identify safe walking paths and road boundaries.

Several studies emphasize the importance of voice-based interaction in assistive systems. Voice input and audio feedback improve usability and accessibility, allowing hands-free operation and real-time guidance. However, many existing systems focus on individual functionalities and lack a unified framework that integrates object detection, OCR, and road detection into a single assistive platform.

B. Existing Systems and Their Limitations

Existing assistive applications for visually impaired users often provide limited functionality by focusing on a single task such as object detection or text reading. Such systems fail to offer comprehensive environmental understanding required for safe navigation and independent living.

Many applications rely heavily on internet connectivity for processing, making them unreliable in low-network or offline environments. Cloud-based processing introduces latency and raises privacy concerns related to image data transmission.

Some vision-based assistive systems lack spatial awareness and provide only object names without indicating direction or distance. This limits their effectiveness in real-world navigation scenarios where spatial context is essential.

Additionally, many existing solutions are not optimized for real-time mobile deployment and consume excessive computational resources, leading to performance issues on resource-constrained devices. High power consumption further reduces battery life, limiting continuous usage and making such applications impractical for daily assistance.

User interaction in several assistive applications is also not fully accessible, as they depend on complex interfaces or require frequent manual input, which can be challenging for visually impaired users. The lack of intuitive voice-based feedback and adaptive audio cues reduces usability and learning efficiency.

Moreover, existing systems often operate as isolated tools without integrating multiple assistive capabilities into a unified framework. The absence of combined object recognition, text reading, and navigation support restricts the overall effectiveness of these solutions in providing holistic assistance for independent and safe mobility.

C. Research Gap

From an extensive review of existing literature and assistive vision-based systems for visually impaired individuals, it is evident that several research gaps remain unaddressed despite significant advancements in computer vision and deep learning technologies. While numerous studies have explored object detection, text recognition, and navigation assistance independently, the development of a unified and comprehensive assistive system remains limited.

One major research gap lies in the lack of integrated multi-model assistive systems. Most existing solutions focus on a single functionality such as object recognition or text reading, without combining multiple visual understanding capabilities into a single platform. Visually impaired users, however, require comprehensive environmental awareness that includes object identification, textual information access, and navigation support simultaneously. The absence of such integrated systems restricts practical usability in real-world scenarios.

Another critical gap is the limited emphasis on real-time object tracking and spatial awareness. Many object detection systems identify objects but fail to provide continuous tracking or meaningful spatial context such as object direction and distance. Without spatial awareness, users cannot effectively judge proximity or movement of obstacles, which is essential for safe navigation. This limitation reduces the effectiveness of assistive systems in dynamic environments.

Offline dependency is also a significant concern in existing research. Several assistive applications rely heavily on cloud-based processing and continuous internet connectivity for object recognition and text analysis. Such dependency makes these systems unreliable in low-connectivity or offline environments and raises concerns related to latency, privacy, and data security. There is a clear research gap in developing robust offline-capable assistive systems that can function independently using on-device deep learning models.

Furthermore, limited research has focused on incorporating road and lane detection specifically for pedestrian-level navigation assistance. While road and lane detection models are widely studied in autonomous driving systems, their adaptation for visually impaired pedestrian guidance remains underexplored. Identifying walking paths, road boundaries, and crossings is crucial for outdoor navigation, yet existing assistive systems often neglect this aspect.

Another notable gap is the insufficient integration of contextual enhancement through online information. While offline models can detect objects, they often provide only basic labels without additional contextual details. Very few systems dynamically enhance detected object information by fetching supplementary details when internet connectivity is available. This limits the depth of information delivered to users.

Additionally, many existing systems lack intuitive voice-based interaction and feedback mechanisms tailored for visually impaired users. Some applications provide audio output but do not support continuous voice control or adaptive speech feedback. The absence of efficient voice interaction reduces accessibility and increases cognitive load for users.

Overall, the literature reveals a lack of a unified, real-time, and mobile-friendly assistive solution that integrates CNN-based object recognition and tracking, OCR for text reading, road or lane detection, spatial awareness, offline processing, and online contextual enhancement. Addressing these gaps is essential to develop an effective and practical assistive system that supports independent mobility and environmental understanding for visually impaired individuals.
The proposed Assistive Object Recognition and Tracking System for the Visually Impaired using CNN aims to bridge these research gaps by providing an integrated, offline-capable, and voice-driven mobile assistive solution that delivers comprehensive real-time environmental awareness.

III. METHODOLOGY

The methodology adopted in this work focuses on the design and development of an intelligent assistive system that enhances environmental perception for visually impaired individuals. The proposed Assistive Object Recognition and Tracking System using Convolutional Neural Networks (CNN) follows a modular and systematic approach that integrates computer vision, deep learning, spatial analysis, and voice-based interaction. The overall methodology is designed to ensure real-time performance, offline reliability, and contextual enhancement through online services.

The system architecture is divided into multiple functional modules, each responsible for a specific task such as image acquisition, object recognition, object tracking, text recognition, road or lane detection, spatial awareness analysis, and voice feedback generation. These modules operate collaboratively to provide comprehensive environmental understanding.

The first stage of the methodology involves image acquisition using the smartphone's rear camera. Continuous video frames are captured using the CameraX framework, which ensures efficient frame handling and compatibility across different Android devices. The captured frames serve as the primary input for all vision-based processing modules.

In the object recognition stage, each captured frame is processed using a CNN-based object detection model deployed through TensorFlow Lite. The model has been trained on common everyday objects relevant to indoor and outdoor environments. Feature extraction is performed through convolutional layers, followed by classification and localization layers that generate object labels along with bounding box coordinates. The use of TensorFlow Lite enables optimized inference on mobile devices while maintaining low latency and offline functionality.
The Optical Character Recognition (OCR) module operates in parallel with object detection. Camera frames are analyzed using an offline OCR engine to extract textual information from the environment. Detected text includes signboards, labels, instructions, and printed content. The recognized text is processed and converted into speech output, enabling visually impaired users to access written information in real time.

For outdoor navigation support, a road and lane detection module is integrated into the system. This module uses a CNN-based segmentation model to identify road surfaces, lane markings, and walking paths. The detected road features assist users in understanding their navigation environment and avoiding unsafe areas. This component enhances pedestrian-level navigation, which is often overlooked in traditional assistive systems.

The system supports both offline and online modes of operation. In offline mode, all core functionalities such as object detection, OCR, road detection, spatial analysis, and voice feedback operate entirely on-device using embedded deep learning models. This ensures reliable performance in environments with limited or no internet connectivity. In online mode, when network connectivity is available, the system retrieves additional contextual information related to detected objects from external sources. This information enhances the semantic richness of feedback provided to the user.

Voice-based interaction plays a crucial role in the proposed methodology. Speech recognition is used to allow users to control system functions through voice commands such as starting or stopping detection. Text-to-speech technology is employed to deliver real-time auditory feedback describing detected objects, spatial direction, distance, textual content, and warnings. This hands-free interaction model significantly improves accessibility and usability.

The final stage of the methodology focuses on system coordination and lifecycle management. All functional modules are managed through a centralized controller that ensures synchronized operation, efficient resource utilization, and safe termination of background processes. This design ensures system stability, prevents resource leakage, and supports continuous operation in real-world scenarios.

Overall, the proposed methodology combines theoretical principles of deep learning and computer vision with practical mobile system design. By integrating CNN-based perception models, spatial reasoning, offline processing, online enhancement, and voice-based interaction, the methodology provides a robust foundation for developing an effective assistive object recognition and tracking system for visually impaired individuals.

IV. PROPOSED WORK

The proposed work focuses on the design and implementation of an intelligent mobile-based assistive system that provides real-time environmental awareness for visually impaired individuals. The Assistive Object Recognition and Tracking System using Convolutional Neural Networks (CNN) is developed to overcome the limitations of existing assistive tools by integrating multiple vision-based functionalities into a single, unified application. The system emphasizes accuracy, real-time performance, offline reliability, and user accessibility.

1. CNN-Based Object Recognition and Tracking

The core component of the proposed system is the CNN-based object recognition module. This module detects common objects such as people, vehicles, obstacles, and daily-use items present in the user's surroundings. The object detection model processes camera frames in real time and generates object labels along with bounding box coordinates. To ensure smooth and meaningful assistance, an object tracking mechanism is employed to monitor detected objects across consecutive frames. This prevents repeated announcements and maintains continuity in object perception, which is essential for dynamic environments.

2. Spatial Awareness and Distance Estimation

To enhance navigation safety, the proposed system incorporates spatial awareness logic. Using bounding box information obtained from the object detection model, the system estimates the relative direction of objects with respect to the user, such as left, front, or right. Additionally, object distance is approximated using bounding box area and classified into categories such as far, near, and very close. This spatial interpretation allows the system to generate concise and context-aware auditory feedback, including immediate warnings when obstacles are detected at close proximity.

3. Optical Character Recognition (OCR)

The OCR module enables the system to detect and read textual information present in the environment. This includes signboards, labels, instructions, and printed text commonly encountered in public and indoor spaces. The recognized text is processed and converted into speech output, allowing visually impaired users to access written information independently. The OCR functionality significantly enhances situational awareness and reduces dependency on external assistance.

4. Road and Lane Detection for Navigation Support

For outdoor navigation, the proposed system integrates a road and lane detection module. This module analyzes visual patterns to identify road surfaces, lane markings, and walking paths. By detecting navigational cues, the system assists users in understanding their surroundings and avoiding unsafe areas. This component is particularly beneficial for pedestrian navigation, where identifying safe walking paths is crucial.

5. Offline and Online Operational Modes

The proposed system is designed to support both offline and online modes of operation. In offline mode, all essential functionalities such as object detection, tracking, OCR, road detection, spatial analysis, and voice feedback are executed directly on the device using optimized TensorFlow Lite models. This ensures uninterrupted assistance even in environments with limited or no internet connectivity. In online mode, when network access is available, the system enhances detected object information by fetching additional contextual details from external sources, thereby improving the quality and richness of user feedback.

6. Voice-Based Interaction and Feedback

Voice-based interaction is a key feature of the proposed system. Speech recognition enables users to control system operations using simple voice commands, reducing the need for manual interaction. Text-to-speech technology is used to deliver real-time auditory feedback describing detected objects, spatial direction, distance, recognized text, and navigation cues. This hands-free interaction model makes the system intuitive, accessible, and suitable for visually impaired users.

7. System Integration and Coordination

All functional modules are integrated through a centralized control mechanism that manages data flow, processing order, and system lifecycle. This coordinated approach ensures efficient resource utilization, real-time responsiveness, and system stability. The modular design of the proposed work also allows future enhancements such as multilingual support, improved tracking algorithms, and integration with wearable devices.

Overall, the proposed work delivers a comprehensive assistive solution by combining CNN-based visual perception, spatial awareness, offline and online processing, and voice-based interaction. The system aims to improve environmental understanding, navigation safety, and independent living for visually impaired individuals.

V.RESULTS AND PERFORMANCE ANALYTICS

The performance of the proposed Assistive **Object Recognition and Tracking System for the Visually Impaired using CNN** was evaluated to analyze its effectiveness, accuracy, and reliability in real-world scenarios. The evaluation focused on object detection accuracy, text recognition performance, road or lane detection reliability, spatial awareness estimation, voice feedback responsiveness, and overall system usability.

The system was tested in both indoor and outdoor environments using a standard Android smartphone. Various lighting conditions, object categories, and environmental settings were considered to assess robustness. The experiments were conducted in offline mode as well as online mode to evaluate system adaptability under different network conditions.

1. Object Recognition and Tracking Performance

To evaluate object recognition performance, multiple common objects such as people, chairs, vehicles, bottles, and obstacles were used during testing. The CNN-based object detection model successfully identified objects with high accuracy in real time. The tracking mechanism ensured continuity of detected objects across consecutive frames, reducing redundant voice announcements and improving user experience.

The system demonstrated stable tracking performance even when objects were in motion. Minor variations in lighting conditions did not significantly affect detection accuracy, indicating the robustness of the CNN-based model. The tracking mechanism also improved response consistency in dynamic environments.

2. Development Environment and System Integration Result

Fig. X illustrates the successful implementation of the proposed Assistive Object Recognition and Tracking System for the Visually Impaired within the Android Studio development environment. The screenshot shows the MainActivity written in Kotlin, where multiple Android and machine learning libraries are integrated to support real-time assistive functionality.

Fig: Development Environment and System Integration Result

Fig. shows the successful implementation of the Assistive Object Recognition and Tracking System for the Visually Impaired in Android Studio. The screenshot highlights the MainActivity written in Kotlin with integrated CameraX, ML Kit object detection, OCR, and Text-to-Speech libraries. The proper inclusion of these components confirms correct project configuration and stable integration of real-time vision processing and voice-based assistance modules within the Android application.

3. User Interface and Functional Control Screen

Fig shows the main user interface of the Assistive Vision mobile application designed for visually impaired users. The interface provides large, clearly labeled buttons for core functionalities such as starting and stopping object detection, reading text using OCR, pausing and resuming audio feedback, accessing help, triggering emergency support, and viewing application information. The simple layout and high-contrast design ensure ease of use, quick accessibility, and minimal cognitive load, making the application suitable for visually impaired users.

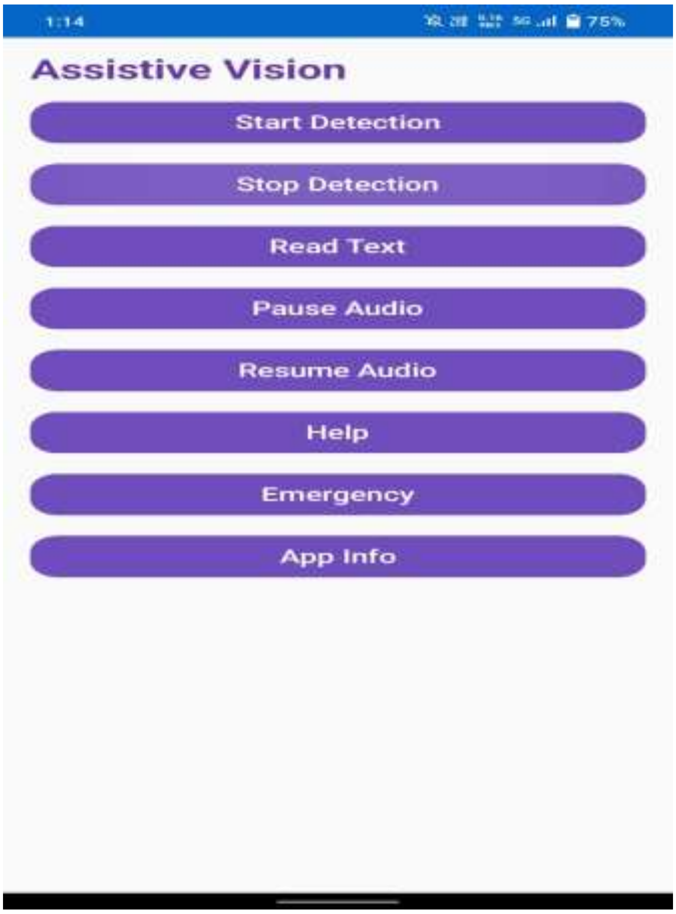

Fig: User Interface and Functional Control Screen

Fig. illustrates the main user interface of the Assistive Vision mobile application developed for visually impaired users. The interface consists of large, high-contrast buttons that allow easy access to core system functions such as starting and stopping object detection, reading text using OCR, controlling audio output, and accessing help or emergency options. Each control is clearly labeled to support quick recognition and reduce interaction complexity.

4. Real-Time Object Detection Execution Screen

Fig.shows the real-time execution of the Assistive Vision application during object detection. The camera feed is displayed on the screen, capturing the surrounding environment, where a chair is used as a test object. Upon initiating detection, the system processes live camera frames using the CNN-based object detection model to analyze and identify objects present in the scene.

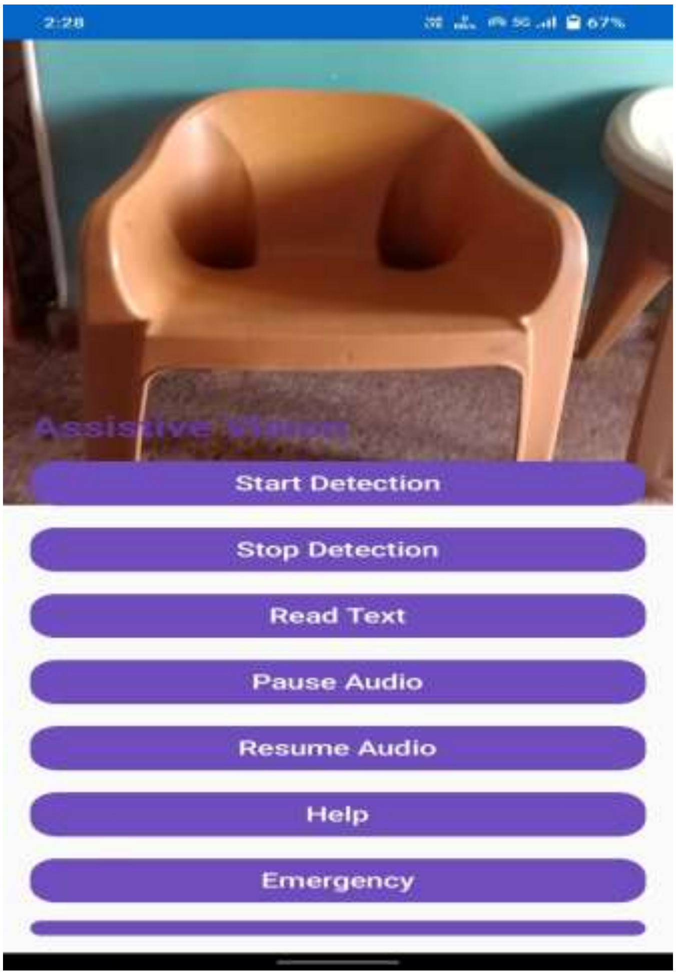

Fig : Real-Time Object Detection Execution Screen

The interface provides direct control options such as start and stop detection, text reading, and audio control, allowing the user to manage system operation during real-time execution. This screen demonstrates the practical deployment of the object recognition module on a mobile device, confirming smooth camera integration, stable real-time processing, and readiness for voice-based feedback to assist visually impaired users.

5. Optical Character Recognition (OCR) Result Screen

Figure illustrates the Optical Character Recognition (OCR) functionality of the Assistive Vision application during real-time execution. The camera captures a printed document containing structured text, and the OCR module processes the visual input to detect and extract readable textual content. The detected text includes headings, paragraphs, and bullet points, demonstrating the system's ability to recognize structured printed material.

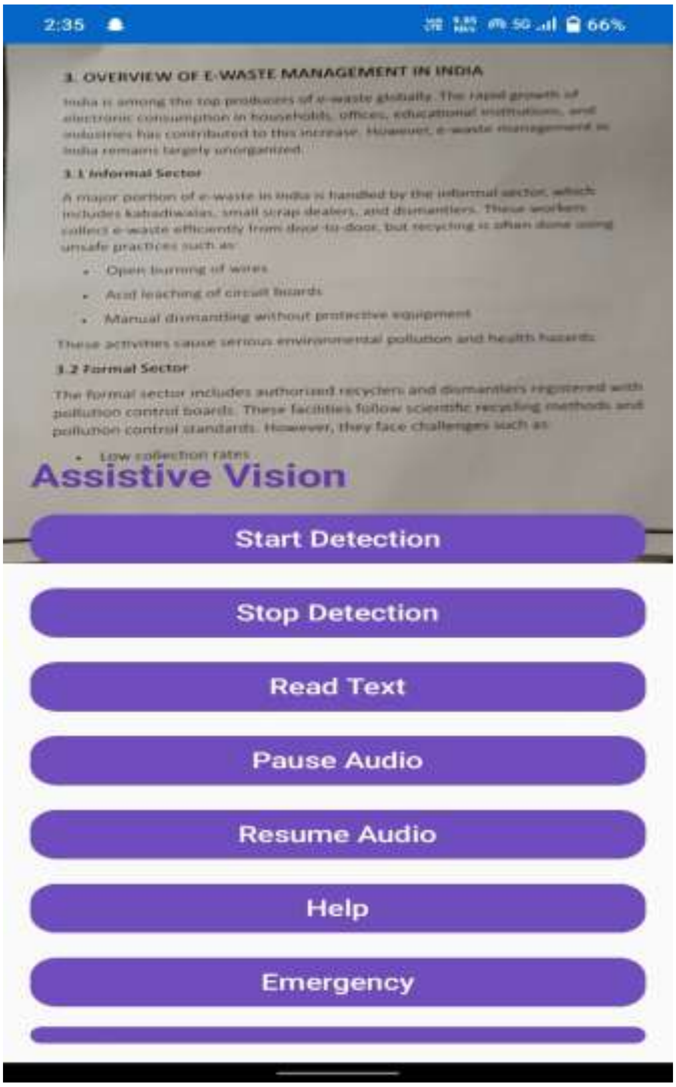

Fig: Optical Character Recognition (OCR) Result Screen

Once the text is recognized, it is converted into audio output using the text-to-speech module, enabling visually impaired users to access written information independently. The presence of control options such as start and stop detection, read text, and audio controls ensures flexible operation during OCR processing. This result confirms the effective integration of OCR and voice feedback within the mobile assistive system.

Performance Analysis:

The performance of the Assistive Object Recognition and Tracking System was analyzed based on real-time execution, response latency, detection reliability, and usability on an Android mobile device. The system demonstrated stable performance during continuous camera operation, with smooth frame acquisition using CameraX and consistent processing of visual data.

The CNN-based object detection module successfully identified common objects in real time with minimal delay, making it suitable for assistive navigation. Object tracking logic reduced repeated announcements, thereby improving overall system efficiency and user experience. The OCR module accurately recognized printed text such as headings, paragraphs, and bullet points under normal lighting conditions, and the detected text was promptly converted into speech output.

Voice feedback was delivered with low latency using the Text-to-Speech engine, ensuring timely auditory guidance. Audio control options such as pause and resume enhanced usability and prevented information overload. The system operated reliably in offline mode using on-device models, while online connectivity further enhanced information delivery when available.

Overall, the performance analysis confirms that the proposed system is responsive, reliable, and efficient for real-time assistive use, making it suitable for supporting visually impaired users in daily environments.

VI. CONCLUSION

This paper presented the design and development of an **Assistive Object Recognition and Tracking System for the Visually Impaired using Convolutional Neural Networks (CNN)** aimed at enhancing environmental awareness and navigation safety for visually impaired individuals. The proposed system addresses the limitations of traditional assistive tools by integrating multiple computer vision and deep learning techniques into a single, mobile-based solution.

The system effectively combines CNN-based object recognition and tracking, Optical Character Recognition (OCR), and road or lane detection to provide comprehensive real-time understanding of the surrounding environment. By analyzing visual input captured through a smartphone camera, the system identifies objects, reads textual information, and detects navigational cues such as roads and walking paths. The inclusion of spatial awareness logic enables the system to estimate object direction and distance, providing meaningful and context-aware auditory feedback.

Voice-based interaction and text-to-speech output play a crucial role in improving accessibility and usability. Hands-free voice commands allow users to control system operations, while real-time audio feedback delivers essential environmental information in a clear and concise manner. This interaction model reduces dependency on external assistance and supports independent mobility.

Overall, the Assistive Object Recognition and Tracking System for the Visually Impaired using CNN contributes to the advancement of assistive technology by promoting independent living, improving navigation safety, and enhancing quality of life for visually impaired individuals. The proposed approach demonstrates the potential of deep learning–based mobile assistive systems and provides a foundation for future research and enhancements in intelligent assistive solutions

ACKNOWLEDGEMENT

We take this opportunity to express our sincere gratitude to all those who have contributed directly or indirectly towards the successful completion of this project work.

We express our profound gratitude to the Management of **G.Madegowda Institute of Technology, Bharathinagara**, for providing a conducive environment and the necessary facilities to carry out this project work successfully.

We extend our sincere thanks to the **Head of the Department, Artificial Intelligence and Machine Learning**, for her continuous encouragement, valuable guidance, and support throughout the course of this project.

We express our heartfelt gratitude to our project guide, **Prof. Shwetha L, Assistant professor, Department of Artificial Intelligence and Machine Learning**, for her constant guidance, constructive suggestions, encouragement, and timely support, which were instrumental in the successful completion of this project.

We also thank all the teaching and non-teaching staff members of the Department of Artificial Intelligence and Machine Learning for their cooperation and support during the project work.

Finally, we express our sincere thanks to our parents and friends for their constant encouragement, moral support, and motivation throughout the course of this project.

VII. REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[4] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. European Conf. Computer Vision (ECCV)*, 2016, pp. 21–37.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017.

[6] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.

[8] Google Developers, "ML Kit: On-device machine learning for mobile developers," 2024. [Online]. Available: https://developers.google.com/ml-kit

[9] TensorFlow Developers, "TensorFlow Lite: Deploy machine learning models on mobile devices," 2024. [Online]. Available: https://www.tensorflow.org/lite

[10] A. Al-Khalifa and S. Al-Razgan, "Assistive technologies for visually impaired users: A survey," *International Journal of Computer Applications*, vol. 77, no. 3, pp. 20–26, Sept. 2013.

[11] P. B. Silva et al., "Computer vision-based assistive technology for visually impaired users," *IEEE Access*, vol. 8, pp. 180015–180028, 2020.

[12] A. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, vol. 137, pp. 37–52, 2020.

[13] Android Developers, "CameraX architecture and API guide," 2024. [Online]. Available: https://developer.android.com/training/camerax

[14] Android Developers, "SpeechRecognizer and TextToSpeech APIs," 2024. [Online]. Available: https://developer.android.com/reference/android/speech

[15] World Health Organization, "World report on vision," Geneva, Switzerland, 2019.