International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 12, ISSN No. 2455-2143, Pages 354-361
Published Online April 2024 in IJEAST (http://www.ijeast.com)

TALKIFY: REALTIME OBJECT DETECTION WITH VOICE ALERT

Chaitanya Singh, Sarthak Gupta, Zaheer Khan
B.Tech Student
Department of Computer Science and Engineering
Inderprastha Engineering College, Sahibabad, Ghaziabad, India

Chandra Pushpanjali
Professor
Department of Computer Science and Engineering
Inderprastha Engineering College, Sahibabad, Ghaziabad, India

Abstract— **For millions living with visual impairments, navigating the world safely and independently can be a constant struggle. Simple tasks like identifying objects in their path or crossing a busy street become significant challenges. TALKIFY emerges as a potential solution, aiming to empower these individuals with greater autonomy and confidence through a real-time object detection system with voice alerts. This project integrates advancements in computer vision and deep learning to bridge the gap between sight and touch.**

A strategically placed camera captures live video footage of the user's surroundings. This footage is then processed using sophisticated deep learning algorithms, trained on extensive datasets of labeled images. These algorithms enable the system to recognize a predefined set of objects within the user's field of view. Upon object identification, the system triggers a voice notification module, translating the object's label into clear and concise voice alerts delivered through earphones or a speaker. Imagine the system saying "Staircase ahead, three steps" as the user approaches one.

TALKIFY offers a multitude of benefits, transforming daily life for visually impaired individuals. By providing real-time object recognition, users can navigate both familiar and unfamiliar environments with greater ease, fostering a sense of freedom and independence. TALKIFY acts as a vigilant guardian, alerting users to potential hazards in their surroundings, leading to increased safety. Furthermore, the system empowers users with a sense of self-reliance and confidence, enhancing their participation in various aspects of life, from navigating workplaces to engaging in social interactions.

Keywords—**Real-time Object Detection, Voice Alerts Visually Impaired, Computer Vision, Deep learning, Independent Navigation System.**

I. INTRODUCTION

A World without Sight - The Challenge of Navigation forthe Visually Impaired

For millions of people worldwide, the ability to navigate their surroundings safely and independently is a constant struggle. Visual impairments, ranging from partial vision loss to complete blindness, pose significant challenges in everyday activities. Simple tasks like identifying objects in their path, recognizing landmarks, or even crossing a busy street become daunting obstacles. This often leads to a reliance on assistance from others, restricting their independence, social inclusion, and overall well-being.

Imagine a world where everyday tasks require constant help. Identifying a familiar object on a shelf in the grocery store, locating a specific department at work, or even navigating a social gathering can become overwhelming experiences. This is the reality for many living with visual impairments.

TALKIFY emerges from a desire to bridge this gap and empower visually impaired individuals to navigate their world with greater confidence and autonomy. This project explores the development of a real-time object detection system with voice alerts, aiming to transform the way visually impaired individuals interact with their surroundings.

II. LITERATURE SURVEY

The advancement of object detection technology has significantly contributed to numerous fields, including autonomous driving, surveillance, healthcare, and assistive technologies. Integrating voice alerts into real-time object detection systems aims to enhance the accessibility and utility of such systems, particularly for visually impaired individuals. This literature survey explores the evolution, methodologies, and applications of real-time objectdetection systems with voice alert functionalities, highlighting key contributions, technological progress, and existing challenges.

1. **Evolution of Object Detection Systems:** Object detection has evolved from traditional computer vision techniques to advanced deep learning methodologies. Early approaches relied on manual feature extraction and simple classifiers like Support Vector Machines (SVM) and Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005). These methods, although groundbreaking, were limited by their inability to handle complex variations in object appearances and backgrounds.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized object detection. Pioneering architectures like AlexNet (Krizhevsky et al., 2012) and VGGNet (Simonyan & Zisserman, 2014) demonstrated superior performance in image classification tasks, laying the groundwork for more complex detection frameworks. Region-based CNNs (R-CNN) (Girshick etal., 2014) and its successors, Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2015), integrated region proposal networks to enhance detection speed and accuracy.

2. **Modern Object Detection Architectures**: State-of-the-art object detection systems leverage various CNN-based architectures. The Single Shot MultiBox Detector (SSD) (Liu et al., 2016) and You Only Look Once (YOLO) (Redmon et al., 2016) models introduced real-time detection capabilities, making them suitable for applications requiring immediate response. YOLO's unifiedarchitecture for both detection and classification significantly improved processing speeds, while SSD's use of multiple feature maps allowed for more precise detections at different scales.

Recent advancements include models like YOLOv3 (Redmon & Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), and Efficient Det (Tan et al., 2020), which incorporate complex backbone networks and novel optimization techniques to further enhance detection accuracy and efficiency. Transformer-based models, such as DETR (Carion et al., 2020), have introduced a new paradigm by utilizing attention mechanisms for direct object prediction without the need for region proposals.

3. **Integration of Voice Alerts:** The incorporation of voice alerts into object detection systems aims to create an intuitive interface for users, particularly benefiting thosewith visual impairments. Voice alerts provide immediateand comprehensible feedback about the surrounding environment, enhancing situational awareness and safety. Research on integrating voice alerts with object detection has explored various approaches. Early systems utilized pre-recorded messages triggered by detected objects, while more advanced systems employ text-to-speech (TTS) engines for dynamic and context-sensitive alerts. Forinstance, the work by Amrutha and Kumar (2019) demonstrated a real-time detection system with TTS- generated alerts, enhancing the mobility and independence of visually impaired users.
4. **Applications and Impact**: Real-time object detection with voice alerts has diverse applications across multipledomains. In assistive technology, these systems facilitate independent navigation and interaction for visually impaired individuals. Systems like Aira and Be My Eyes leverage object detection and remote assistance to guide users through complex tasks and environments.

In autonomous driving, voice alerts enhance driver assistance systems by providing real-time information about pedestrians, obstacles, and traffic conditions, thereby improving safety and decision-making. Similarly, surveillance systems benefit from immediate auditory notifications about potential security threats, enabling quicker responses.

Healthcare applications include aiding elderly and disabled individuals in home environments, where voice alerts can notify users of important events like doorbell rings or approaching hazards. Retail and customer service sectors also leverage these systems to enhance accessibility and user experience.

5. **Challenges and Future Directions**: Despite significant progress, several challenges remain in developing effective real-time object detection systems with voice alerts. Ensuring accuracy and reliability in diverse and dynamic environments is critical. Models must efficiently handle variations in lighting, weather, and occlusions to maintain high detection performance.

Latency is another crucial factor, as real-time systems must process and respond to inputs within milliseconds. Optimization techniques and hardware advancements, such as edge computing and specialized accelerators (e.g., GPUs, TPUs), are essential to achieve the required speeds.

The development of context-aware voice alerts is an ongoing research area. Alerts should be not only accurate but also contextually relevant and non-intrusive, avoiding unnecessary distractions. Natural language processing (NLP) advancements can enhance the personalization and contextual understanding of voice alerts.

III. PROPOSED METHODOLOGY

The proposed methodology for a project focused on innovation in speech recognition may vary based on the specific objective or novelty being pursued. Here is a generalized methodology that could be adapted to different areas of innovation within speech recognition.

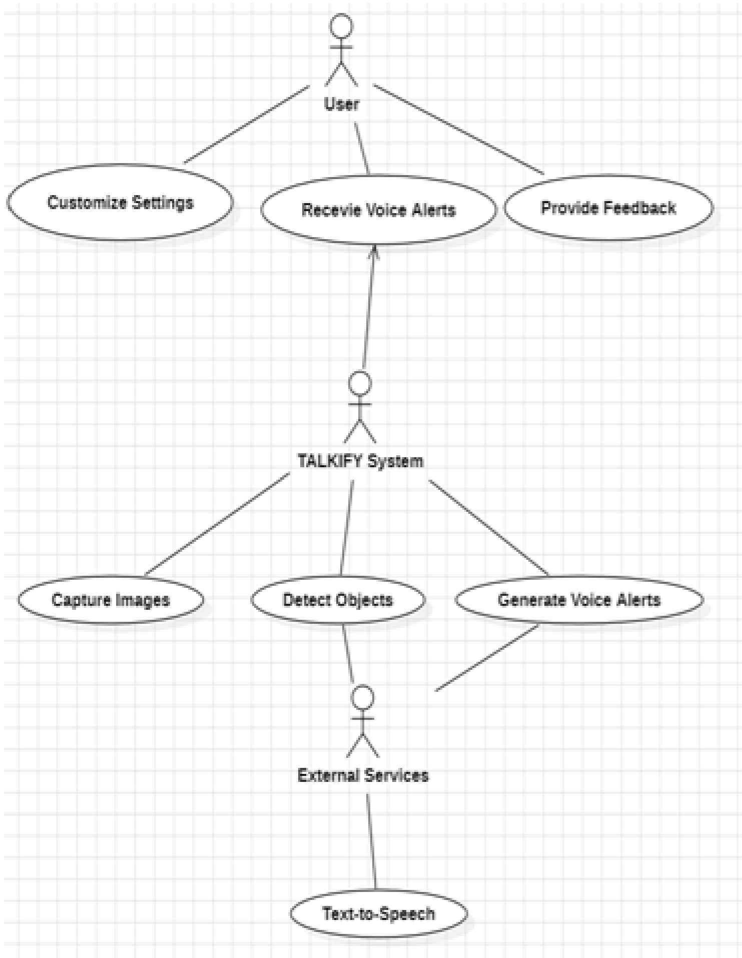
1. **Literature Review:** Begin by conducting an in-depth review of existing literature, research papers, and patents related to speech recognition. Understand current methodologies, algorithms, and challenges in the field.
2. **Problem Definition and Objective Setting:** Define the specific problem or area of innovation within speech recognition that your project aims to address. Set clear and achievable objectives based on the identified gaps or opportunities.
3. **Data Collection and Preprocessing:** Gather a diverse and representative dataset for training and evaluation. This dataset should cover different languages, accents, speech contexts, and potential challenges relevant to your project's focus. Preprocess the data to ensure quality and uniformity.
4. **Algorithm Development or Enhancement:** Develop new algorithms or enhance existing ones to address the identified challenges or to achieve the project's objectives. This could involve machine learning models, neural networks, signal processing techniques, or other innovative approaches.
5. **Model Training and Optimization:** Train and fine-tune the developed algorithms/models using the collected dataset. Optimize the models for accuracy, efficiency, and performance in the target environments or scenarios.
6. **Evaluation and Testing:** Evaluate the performance of the developed system rigorously using appropriate metrics. Test the system in various real-world scenarios to validate its effectiveness, accuracy, robustness, and scalability.
7. **Iterative Improvement:** Analyze the results and iterativelyrefine the models or algorithms based on the evaluation feedback. Incorporate improvements to address any shortcomings or limitations observed during testing. EthicalConsiderations and User Privacy: Ensure ethical considerations in data usage and model development, especially concerning user privacy and data security. Implement measures to protect sensitive information and adhere to ethical guidelines.

IV. USE CASE & SEQUENCEDIAGRAM

A) <u>USE CASE DIAGRAM:</u>

1. **Customize Settings:** Actor: User Description: Users can customize system settings, including the types of objects to detect the frequency of alerts, and the verbosity of thenotifications. Preconditions: The system must be operational, and the user must have access to the settings interface. Post conditions: User preferences are saved and applied to the system.
2. **Receive Voice Alerts:** Actor: User Description: Users receive voice alerts generated by the system based on detected objects. These alerts provide contextual information to help the user navigate their environment. Preconditions: The object detection module must detect an object, and the voice alert module must generate a corresponding alert. Post conditions: The user is informed about the detected objects through audio feedback.
3. **Provide Feedback**: Actor: User Description: Users can provide feedback on the system's performance, such as the accuracy of detections and the usefulness of voice alerts. Preconditions: The feedback interface must be accessible. Post conditions: Feedback is recorded and used to improve system performance and user experience.
4. **Capture Images**: Actor: TALKIFY System Description: The system continuously captures images from the device'scamera to process for object detection. Preconditions: The camera must be functioning correctly. Post conditions: Captured images are available for the object detection module to analyze.
5. **Detect Objects**: Actor: TALKIFY System Description: The system processes captured images to identify and classify objects in real-time using the YOLOv4 model. Preconditions: Images must be captured by the camera, and the object detection model must be active. Post conditions: Detected objects are identified, classified, and ready foralert generation.
6. **Generate Voice Alerts**: Actor: TALKIFY System Description: Based on detected objects, the system generates voice alerts that describe the objects and their locations. Preconditions: Objects must be detected and classified by the object detection module. Post conditions: Voice alerts are generated and ready to be delivered to the user.
7. **Text-to-Speech**: Actor: External Services Description: The Text-to-Speech (TTS) engine converts text descriptions of detected objects into spoken words. Preconditions: Text descriptions of objects must be generated by the system. Post conditions: Spoken alerts are created and played to the user through the device's speakers.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 12, ISSN No. 2455-2143, Pages 354-361
Published Online April 2024 in IJEAST (http://www.ijeast.com)



B. SEQUENCE DIAGRAM:

1. **Start System:** User: Initiates the system. TALKIFY System: Receives the start command and begins initialization.
2. **Initialize Components:** TALKIFY System: Initializes all necessary components, including the Camera Module.
3. **Capture Image:** Camera Module: Continuously captures images from the environment and sends the image data back to the TALKIFY System.
4. **Process Image:** TALKIFY System: Forwards the captured image to the Object Detection Module for processing. Object Detection Module: Analyzes the image to detect and classify objects. It sends the detected objects data back to the TALKIFY System.
5. **Generate Alert:** TALKIFY System: Sends the detected object data to the Voice Alert Module. Voice Alert Module: Generates a text description based on the detected objectsand sends this text to the TTS Engine for conversion.
6. **Text-to-Speech Conversion:** TTS Engine: Converts the text description into an audible voice alert and sends it back tothe Voice Alert Module.
7. **Receive Alert:** Voice Alert Module: Sends the generated voice alert to the user. User: Receives the voice alert and is informed about the detected objects.
8. **Provide Feedback:** User: Provides feedback on the system's performance or accuracy. TALKIFY System: Records and processes the feedback for future improvements.

Explanation of Interactions

User Interaction: The user starts the system, receives voice alerts, and provides feedback. This interaction ensures the system is user-friendly and meets the user's needs.

System Initialization: Upon starting, the TALKIFY system initializes all necessary modules, preparing them for real-time object detection and alert generation.

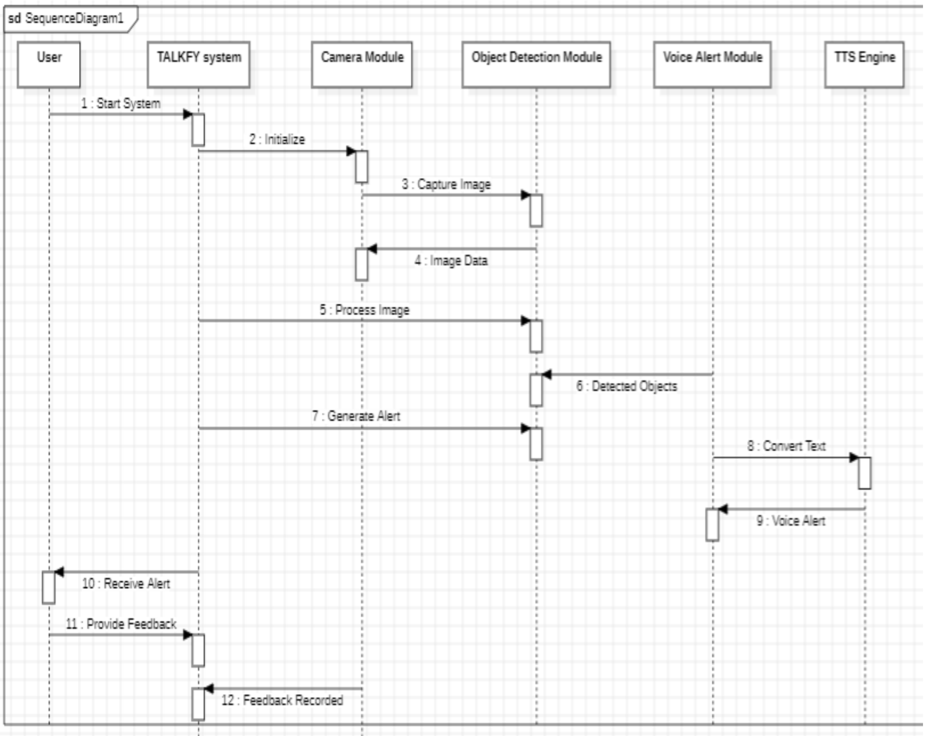
Continuous Image Capture: The Camera Module captures images continuously, feeding the system with real-time visual data.

Object Detection Process: Captured images are processed by the Object Detection Module, which detectsand classifies objects, sending the information back to thesystem.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 12, ISSN No. 2455-2143, Pages 354-361
Published Online April 2024 in IJEAST (http://www.ijeast.com)

Alert Generation and Delivery: The Voice Alert Module generates textual descriptions of detected objects and uses the TTS Engine to convert these texts into audible alerts, which are then delivered to the user.

User Feedback: Users provide feedback on the system, which is recorded and used to improve system accuracyand user experience.



V. METHODOLOGY

1. **Inception: Dreaming the Vision:** Like a painter staring ata blank canvas, we envisioned a world where technology becomes the eyes for those with visual impairments. With empathy as our guiding star, we delved deep into understanding the challenges faced by the visually impaired, laying the groundwork for TALKIFY.
2. **Research and Analysis: Peering into the Depths**: Armed with compassion and curiosity, we scoured the landscape of existing solutions, dissecting their strengths and weaknesses. Through conversations with experts and potential users, we unearthed insights that served as the cornerstone for TALKIFY's design.
3. **Design Thinking: Sketching the Blueprint**: With a myriad of ideas swirling in our minds, we huddled together, donning our designer hats. Sketching, prototyping, and iterating became our daily rituals as we crafted a user- centric design that seamlessly integrates into the lives of ourtarget audience.
4. **Development: Breathing Life into the Code**: As the design took shape, our developers stepped into the limelight, wielding lines of code like seasoned craftsmen. Adopting an agile approach, we sprinted through iterations, transforming concepts into tangible features with each keystroke.
5. **Testing and Validation: Striving for Perfection**: With our creation nearing completion, it was time to put it through the crucible of testing. Rigorous testing sessions, both in simulated environments and real-world scenarios, unearthed bugs, refined functionalities, and validated the efficacy of TALKIFY.
6. **Deployment and Feedback: Releasing into the Wild**: With bated breath, we unleashed TALKIFY into the wild, eager towitness its impact. User feedback became our guiding star, illuminating areas for improvement and validating our efforts as we embarked on a journey of continuous refinement.
7. **Maintenance and Evolution: Nurturing Growth**: Like a gardener tending to a blossoming garden, we committed ourselves to the ongoing maintenance and evolution of TALKIFY. Regular updates, feature enhancements, and technological advancements ensured that TALKIFY remained at the forefront of accessibility innovation.
8. **Community Engagement: Fostering Connection**: Beyond mere technology, we sought to foster a vibrant community around TALKIFY. Through workshops,

outreach programs, and collaborative partnerships, we endeavored to empower and uplift individuals with visual impairments, creating a ripple of positive change.

VI. TALKIFY APPLICATIONS

1. **Assistance for Visually Impaired Individuals:** TALKIFY serves as a powerful tool for visually impaired individuals, providing them with real-time information about their surroundings. By detecting objects and relaying voice alerts, it assists in navigation and obstacle avoidance, empowering users to move confidently in various environments.
2. **Safety Enhancement in Public Spaces:** In public spaces such as streets, parks, and shopping malls, TALKIFY enhances safety by alerting users to potential hazards and obstacles. It can detect objects like curbs, obstacles on pathways, or moving vehicles, reducing the risk of accidents and improving overall safety.
3. **Indoor Navigation and Accessibility:** Within indoor environments like buildings, offices, and public facilities, TALKIFY helps users navigate complex layouts. It can identify key landmarks, entrances, exits, and obstacles, making indoor spaces more accessible and navigable foreveryone.
4. **Educational Support:** TALKIFY can be utilized in educational settings to assist students with visual impairments. It aids in identifying objects in classrooms, laboratories, and libraries, enabling independent learning and participation in educational activities.
5. **Assistance in Daily Activities:** In daily activities such as cooking, cleaning, and organizing, TALKIFY provides valuable assistance by identifying objects and providing relevant information. It helps users locate and differentiate between various items, promoting independence and efficiency in daily tasks.
6. **Workplace Accommodations:** TALKIFY can be integrated into workplace environments to support employees with visual impairments. It assists in navigating office spaces, locating equipment and supplies, and accessing digital interfaces, facilitating greater productivity and inclusion in the workplace.
7. **Enhanced User Experience for Wearable Devices:** When integrated into wearable devices such as smart glasses or headsets, TALKIFY offers a seamless and hands-free user experience. Users can receive real-time object detection and voice alerts directly through their wearable devices, providing discreet assistance and convenience.
8. **Accessibility Features in Smart Cities:** In the context of smart cities, TALKIFY contributes to accessibility initiatives by providing real-time assistance to individuals with disabilities. It can be integrated into urban infrastructure, public transportation systems, and smart devices to ensure inclusivity and equal access to city resources and services.
9. **Innovative Solutions for Assistive Technology:** TALKIFY represents an innovative solution in the field of assistive technology, demonstrating the potential of AI- driven systems to improve the lives of individuals with disabilities. Its seamless integration of object detection and voice alerts showcases the transformative power of technology in creating more accessible and inclusiveenvironments.

VII. CHALLENGES AND LIMITATIONS DEPENDENCY ONINTERNET INFRASTRUCTURE:

1. **Real-time Processing Delay:** Challenge: TALKIFY relies on internet connectivity for accessing external services, such as cloud-based object detection models or text-to- speech engines. This dependency introduces latency in data transmission, which can result in delays in processing real- time images and generating voice alerts.

Limitation: Users may experience delays between object detection and receiving voice alerts, impacting the system's responsiveness and usability.

2. **Reliability and Stability:** Challenge: Variability in internet connectivity can lead to unreliable data transmission and service interruptions. Factors such as network congestion, bandwidth limitations, and signal strength fluctuations can affect the stability of TALKIFY's performance.

Limitation: Users may encounter instances where the system fails to detect objects or deliver voice alerts due to network disruptions, compromising the system's reliability in critical situations.

3. **Privacy and Security Concerns:** Challenge: Transmitting images and data over the internet raises privacy and securityconcerns, particularly regarding the storage and processing of sensitive visual information on external servers or cloud platforms.

Limitation: Users may hesitate to use TALKIFY due to concerns about potential data breaches, unauthorized access to personal information, or misuse of captured images, highlighting the importance of robust privacy measures and secure data transmission protocols.

4. **Accessibility in Remote Areas:** Challenge: In remote or underserved areas with limited internet infrastructure,access to high-speed internet connectivity may be unreliableor unavailable, hindering the adoption

and usability of TALKIFY in these regions.

Limitation: Visually impaired individuals residing in remote areas may face barriers in accessing real-time object detection and voice alert services, exacerbating disparities in accessibility and digital inclusion.

5. **Scalability and Cost Considerations:** Challenge: ScalingTALKIFY to accommodate a large user base requiressubstantial infrastructure resources, including server capacity, bandwidth allocation, and maintenance costs. Limitation: The scalability of TALKIFY may be constrainedby budgetary limitations or resource constraints, limiting its deployment and reach, particularly in resource-constrained environments or developing regions.

6. **Offline Functionality and Redundancy:** Challenge: Implementing offline functionality and redundancy mechanisms to mitigate internet dependency introduces additional complexity and overhead in system design and development.

Limitation: While offline capabilities can enhance the system's robustness and resilience, implementing effective offline modes may require significant engineering efforts and compromise real-time performance and functionality.

VIII. PROPOSED ENHANCEMENTSAND ADAPTATIONS

1. **Enhanced Object Detection Algorithms:** Investing in research and development to enhance object detection algorithms can significantly improve the accuracy and efficiency of the TALKIFY system. By incorporating cutting-edge deep learning techniques and leveraging larger and more diverse datasets, we can achieve better object recognition performance across various environmental conditions.
2. **Multi-Modal Sensory Integration:** Expanding the capabilities of TALKIFY to integrate multiple sensory modalities, such as incorporating depth sensing technology or thermal imaging, can provide more comprehensive environmental awareness for users. By combining visual data with depth or thermal information, the system can offer richer contextual understanding, leading to more precise object detection and enhanced user safety.
3. **Contextual Awareness and Semantic Understanding:** Integrating contextual awareness and semantic understanding into the system can elevate its intelligence and usability. By analyzing the spatial relationships between detected objects and inferring their semantic significance (e.g., identifying potential obstacles or hazards), TALKIFY can provide more informative and actionable voice alerts tailored to the user's specific needs and preferences.
4. **Adaptive User Interface:** Designing an adaptive user interface that caters to the diverse needs and preferences of users can enhance the overall user experience. Customizable alert settings, personalized voice profiles, and intuitive gesture-based controls are some features that can empower users to tailor the system to their individual preferences and usage scenarios, ensuring maximum accessibility and usability.
5. **Seamless Integration with Wearable:** Devices Enabling seamless integration with wearable devices, such as smart glasses or wrist-worn sensors, can enhance the portability and convenience of the TALKIFY system. By leveraging the capabilities of wearable technology, users can access real-time object detection and voice alerts in a hands-free manner, further enhancing their mobility and independence in navigating their surroundings.
6. **Continuous Feedback Loop and User Engagement**: Establishing a continuous feedback loop and fostering active user engagement are essential for the ongoing improvement and refinement of the TALKIFY system. Regular user surveys, usability testing sessions, and community feedback forums can provide valuable insights into user preferences, pain points, and emerging needs, guiding future development efforts and ensuring that the system remains relevant and impactful.

IX. CONCLUSION

TALKIFY, a proposed real-time object detection system with voice alerts, presents a glimpse into a future where visuallyimpaired individuals can navigate their surroundings with greater confidence and independence. By leveraging advancements in computer vision and deep learning, TALKIFY offers a solution that addresses the limitations of traditionalassistive technologies.

This project has explored the challenges faced by visually impaired individuals in navigating their environments. The literature survey highlighted the need for a comprehensive system that goes beyond basic object detection, emphasizing the importance of accuracy, reliability, and user-centered designprinciples.

The proposed system model outlines how TALKIFY will utilize deep learning models and real-time processing to identify objects and provide clear voice alerts. The use case diagram illustrates how users can interact with the system for navigation and personalization. The methodology section details the development process, focusing on data acquisition, model selection, and user interface design.

Envisioning the broader impact, the applications section explores how TALKIFY can empower users in various aspects of daily life, fostering independence, safety, and social inclusion. Finally, the proposed enhancements and adaptations section outlines possibilities for expanding TALKIFY's functionalities, including advanced object recognition, user interface customization, and integration with other technologies. TALKIFY holds the potential to bridge the gap in navigation for visually impaired individuals. Its success hinges on continuous development, user feedback integration, and open-source collaboration. As TALKIFY evolves, it can become a transformative tool, empowering users to navigate their world with greater confidence, freedom, and a sense of belonging in an inclusive society.

X. REFERENCES

[1]. Gibert, A., Garcia, M. A., & Tortajada, C. (2013). A reviewon electronic travel aids for blind people. Sensors, 13(2), 1945-1973.https://www.mdpi.com/2079-9292/8/11/1281

[2]. Guo, Y., Zhou, Y., Li, Z., & Li, H. (2019). A computer vision-based navigation system for the visually impaired using deep learning and semantic segmentation. Sensors, 19(14),3230.https://www.sciencedirect.com/science/article/ pii/S0957417422017432

[3]. Lee, J. M., Kim, J. H., & Kim, S. W. (2017). Development of a smartphone-based electronic travel aid system for the visually impaired. Sensors, 17(10), 2421.https://link.springer.com/article/10.1007/s10209-023-01086-8

[4]. Lindsay, S. A., Arentsen, J. J., & Howell, W. C. (2012). Guide dogs and the blind: A review of the evidence on costs and outcomes. Journal of Visual Impairment & Blindness, 106(3), 172-184. https://pubmed.ncbi.nlm.nih.gov/18432492/

[5]. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu,C., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2

[6]. Löwenstein, G., Caspi, O., & Samuels, M. A. (2018). Guide dogs for the blind: History, development, and the future. Journal of Veterinary Behavior: Clinical Applications and Research, 25, 54-62. https://www.sciencedirect.com/science/article/pii/030437628390010X

[7]. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[8]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 141-150).

[9]. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60. https://www.researchgate.net/publication/334279066_A_sur vey_on_Image_Data_Augmentation_for_Deep_Learning

[10]. Tzeng, E., Maire, M., & Belongie, S. (2015). Deep domain-adversarial image mapping. In Advances inneural information processing systems (pp. 2468-2476).