Object Detection with Speech Recognition Using AI Techniques

SANJAY P 1, Dr. Pamela Vinitha Eric 2

1(M. Tech Artificial Intelligence, Department of CSE, School of Engineering, Presidency University, Bangalore, sanjay.p8918@gmail.com)

2(Professor CSE, School of Engineering, Presidency University, Bangalore, pamelavinitha.eric@presidencyuniversity.in)

ABSTRACT

Object detection, which is an important part of video analysis and visual comprehension, has become a popular area of research recently. Traditional methods have limitations, so more powerful tools based on deep learning have been developed to learn semantic, high-level, and deeper features. These deep learning models have different network architectures, training methodologies, and optimization functions. Initially, computer vision systems focused on active and accurate object detection, but deep learning has greatly improved detection accuracy. Object detection is used to find meaningful objects in images and videos, such as people, buildings, or animals. It has applications in fields like image recognition and video surveillance. To locate objects, voice recognition is used to identify items, and then object recognition is applied. Voice guidance is also used to assist with object identification.

Keywords : Object detection, Deep learning, Convolutional neural networks (CNNs), YOLO (You Only Look Once), Data augmentation, Image recognition, Video surveillance, Voice recognition.

I. INTRODUCTION

The field of computer vision aims to enable machines to recognize and locate objects in images, which is known as object detection. This technology has the potential to free up human time by allowing machines to take on dangerous, laborious, or boring tasks. However, real-time object detection is a challenging task that has yet to be fully solved, as current approaches are ineffective and require lengthy training sessions. Additionally, recognizing and identifying all objects in a scene, especially when they are similar in appearance, is particularly challenging for machines.

In the mobile technology space, the Android platform has the largest market share and is increasingly used for creating mobile applications. Object recognition technology is still in its early stages on this platform, but its potential is significant, particularly in image tagging applications. Tagging images with relevant text data can assist users in finding the information they need quickly and efficiently, and computer vision algorithms can be used to identify and tag objects in images.

The visually impaired face a number of challenges in daily life, particularly with respect to identifying objects and navigating indoor spaces. However, recent advancements in computer vision and deep learning techniques have enabled the

development of systems that can accurately recognize objects and provide location information to visually impaired users through voice guidance.

II. RESEARCH SURVEY

[1] **An Intelligent Object Detection System with Voice Recognition Capability** is a research paper that proposes a system that integrates object detection and voice recognition capabilities. The system aims to allow users to interact with objects in their environment through voice commands, which can be useful for individuals with disabilities or in scenarios where hands-free interaction is necessary. The paper describes the architecture of the system, which includes a camera for object detection, a microphone for voice recognition, and a central processing unit for executing the commands. The system uses a deep learning-based object detection algorithm to detect objects in real-time and a voice recognition algorithm to interpret the user's spoken commands. The authors evaluate the system's performance and report high accuracy for both object detection and voice recognition tasks. The results suggest that the proposed system can be useful in various applications such as home automation, security systems, and healthcare.

[2] **A System for Object Detection and Voice Recognition** is a research paper that proposes a system that integrates object detection and voice recognition capabilities. The system is designed to enable users to interact with objects in their environment using natural language voice commands, which can be useful in various scenarios such as home automation, healthcare, and security. The paper describes the architecture of the system, which includes a camera for object detection, a microphone for voice input, and a central processing unit for executing the commands. The authors use deep learning-based algorithms for both object detection and voice recognition tasks, and they evaluate the system's performance using various metrics. The experimental results demonstrate that the proposed system achieves high accuracy and efficiency for both object detection and voice recognition tasks. The authors suggest that their system can be extended to support more complex scenarios and applications, such as real-time object tracking and multi-modal interaction.

[3] **Object Detection with Voice Recognition using Deep Learning** is a research paper that proposes a system that combines object detection with voice recognition using deep learning techniques. The system is designed to enable users to interact with objects in their environment through voice commands, which can be useful in various applications such as home automation, healthcare, and surveillance. The paper describes the architecture of the system, which includes a camera for object detection, a microphone for voice input, and a deep learning-based processing unit for executing the commands. The authors use convolutional neural networks (CNNs) for object detection and a recurrent neural network (RNN) for voice recognition tasks. The proposed system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks. The authors suggest that their system can be extended to support more complex scenarios and applications, such as multi-modal interaction and real-time object tracking.

[4] **Voice Controlled Object Detection using Raspberry Pi** is a research paper that proposes a system that combines voice recognition with object detection using a Raspberry Pi microcontroller. The system is designed to enable users to interact with objects in their environment through voice commands, which can be useful in various applications such as home automation, surveillance, and healthcare. The paper describes the architecture of the system, which includes a camera for object detection, a microphone for voice input, and a Raspberry Pi for executing the commands. The authors use OpenCV and Haar classifiers for object detection and the Google Speech Recognition API for voice recognition tasks. The proposed system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks. The authors suggest that their system can be extended to support more complex scenarios and applications, such as real-time object tracking and multi-modal interaction. Overall, the paper demonstrates the feasibility and potential usefulness of using a Raspberry Pi micro controller for implementing a voice-controlled object detection system.

[5] **Voice-Based Object Detection using Convolutional Neural Network** is a research paper that proposes a system that combines voice recognition with object detection using convolutional neural networks (CNNs). The system is designed to enable users to interact with objects in their environment through voice commands, which can be useful in various applications such as home automation, healthcare, and surveillance. The paper describes the architecture of the system, which includes a camera for object detection, a microphone for voice input, and a CNN-based processing unit for executing the commands. The authors use transfer learning to fine-tune a pre-trained CNN model for object detection and a simple neural network for voice recognition tasks. The proposed system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks. The authors suggest that their system can be extended to support more complex scenarios and applications, such as real-time object tracking and multi-modal interaction. Overall, the paper demonstrates the potential usefulness of using CNNs for implementing a voice-controlled object detection system.

[6] **Object Recognition and Voice Recognition System based on FPGA** is a research paper that proposes a system that integrates object recognition and voice recognition capabilities using field-programmable gate arrays (FPGAs). The system is designed to enable users to interact with objects in their environment through voice commands, which can be useful in various applications such as home automation, healthcare, and surveillance. The paper describes the architecture of the system, which includes a camera for object recognition, a microphone for voice input, and an FPGA-based processing unit for executing the commands. The authors use support vector machines (SVMs) for object recognition and a simple neural network for voice recognition tasks. The proposed system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object recognition and voice recognition tasks. The authors suggest that their system can be extended to support more complex scenarios and applications, such as real-time object tracking and multi-modal interaction. Overall, the paper demonstrates the potential usefulness of using FPGAs for implementing a voice-controlled object recognition system.

[7] **Voice-Controlled Object Detection and Tracking for Smart Homes** by H. Kim, J. Kim, and J. Cho: This paper proposes a system for object detection and tracking in smart homes that can be controlled by voice commands. The system uses a camera for object detection, a microphone for voice input, and a Raspberry Pi for processing the commands. The authors use OpenCV and Haar classifiers for object detection, and the Google Speech Recognition API for voice recognition. The system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks.

[8] This paper proposes a system for object detection and voice recognition for robot navigation. The system uses a camera for object detection, a microphone for voice input, and a Raspberry Pi for processing the commands. The authors use a YOLOv3 model for object detection, and the Baidu Speech Recognition API for voice recognition. The system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks.

[9] This paper proposes a system for real-time object detection and voice recognition for assistive technology. The system uses a camera for object detection, a microphone for voice input, and a Raspberry Pi for processing the commands. The authors use a MobileNet-SSD model for object detection, and the Google Speech Recognition API for voice recognition. The system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks.

[10] **Voice-Controlled Object Detection and Recognition for Smart Surveillance System** by Y. Xie, Z. Sun, and L. Chen: This paper proposes a system for object detection and recognition in smart surveillance systems that can be controlled by voice commands. The system uses a camera for object detection, a microphone for voice input, and a Raspberry Pi for processing the commands. The authors use a Faster R-CNN model for object detection and recognition, and the Google Speech Recognition API for voice recognition. The system is evaluated using various performance metrics, and the experimental results show that it achieves high accuracy and efficiency for both object detection and voice recognition tasks.

[11] **An Automated Object Detection and Voice Recognition System for Smart Homes** by S. S. Asadullah, S. H. Bae, and S. K. Lim: This paper presents an automated object detection and voice recognition system for smart homes. The proposed system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize objects and commands spoken by users. The system can perform tasks such as turning on lights or appliances, adjusting the temperature, and opening doors based on voice commands.

[12] **Voice-Controlled Object Detection System for Intelligent Traffic Management** by H. Wu, Y. Chen, and J. Lu: This paper proposes a voice-controlled object detection system for intelligent traffic management. The system uses a deep neural network-based object detection algorithm to detect vehicles, pedestrians, and other objects, and a speech recognition system to recognize commands spoken by

users. The system can be used to manage traffic flow, monitor road conditions, and provide assistance to drivers and pedestrians.

[13] **An Embedded Object Detection and Voice Recognition System for Smart Shopping** by S. K. Jung and K. S. Kim: This paper presents an embedded object detection and voice recognition system for smart shopping. The proposed system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize products and commands spoken by users. The system can assist shoppers in finding products, providing information about products, and placing orders.

[14] **Real-Time Object Detection and Voice Recognition using Raspberry Pi** by V. B. Kamble and A. R. Kadam: This paper proposes a real-time object detection and voice recognition system using Raspberry Pi. The system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize objects and commands spoken by users. The system can be used in various applications such as surveillance, home automation, and robotics.

[15] This paper presents an object detection and voice recognition system for autonomous vehicles. The proposed system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize objects and commands spoken by passengers. The system can assist the autonomous vehicle in navigating through traffic and performing tasks based on voice commands.

[16]This paper proposes an object detection and voice recognition system for smart waste management. The system uses a deep neural network-based object detection algorithm to detect waste bins and a speech recognition system to recognize commands spoken by users. The system can assist in waste collection, optimize waste management processes, and reduce waste pollution.

[17] This paper presents a voice-controlled object detection and recognition system for smart factories. The proposed system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize objects and commands spoken by workers. The system can assist workers in monitoring and controlling production processes, improving efficiency and safety.

[18] This paper proposes a real-time object detection and voice recognition system for smart agriculture. The system uses a deep neural network-based object detection algorithm and a speech recognition system to detect and recognize crops and commands spoken by farmers. The system can assist farmers in monitoring and controlling crop growth, optimizing irrigation and fertilization, and improving yield.

[19] The paper presents an object detection and voice recognition system for smart cities. The system uses a deep learning-based object detection model and a speech recognition engine for voice recognition. The proposed system can detect and recognize various objects and vehicles on the streets, and respond to voice commands from pedestrians and drivers. The proposed system can help improve traffic safety and management in smart cities.

[20] The paper proposes a voice-controlled object detection and recognition system for smart retail. The system uses a deep learning-based object detection model and a speech recognition engine for voice recognition. The proposed system can detect and recognize various products on store shelves, and respond to voice commands from customers. The proposed system can help improve customer experience and sales in smart retail stores.

[20] The paper **An IoT-Based Voice-Controlled Object Detection and Recognition System for Smart Home Applications** proposes an IoT-based system for voice-controlled object detection and recognition in smart homes. The system consists of a Raspberry Pi as the central hub, connected to a variety of sensors and devices such as a camera module, a microphone, and a speaker.

The proposed system uses deep learning-based object detection algorithms, specifically the Single Shot Detector (SSD) algorithm, to detect and recognize objects in real-time. The voice recognition module is implemented using the Google Cloud Speech-to-Text API, which allows for accurate voice recognition even in noisy environments.

The system architecture includes three main components: the object detection module, the voice recognition module, and the user interface. The object detection module uses the SSD algorithm to detect and recognize objects in real-time. The voice recognition module uses the Google Cloud Speech-to-Text API to convert user voice commands into text commands, which are then used to control the object detection module.

The user interface is implemented using a web application, which allows users to view the real-time video feed from the camera module and control the object detection and recognition process using voice commands. The system also includes an alarm system that sends notifications to the user's smartphone in case of a security breach or an abnormal event.

The proposed system was tested in a real-world environment, and the results showed that it achieved high accuracy in object detection and voice recognition. The authors suggest that the system can be used in a variety of smart home applications, such as security monitoring, home automation, and energy management.

[21] The paper presents a voice-activated object detection system for smart cars. The proposed system can detect and recognize various objects on the road, such as cars, pedestrians, and traffic signs, through an onboard camera. The system uses deep learning techniques, specifically the YOLOv3 model, for object detection and recognition.

The authors then integrate a voice recognition module into the system, allowing drivers to control the system with their voice. The voice recognition module is based on the Baidu Speech API and can recognize multiple languages. The driver can use voice commands to start and stop the object detection system, change the detection mode, and get alerts for detected objects.

The proposed system was tested on a Raspberry Pi and achieved real-time object detection and recognition with an accuracy of over 90%. The system's performance was also evaluated in a simulated smart car environment, showing promising results for real-world applications.

Overall, the paper presents a voice-activated object detection system that can improve driving safety by providing real-time detection and recognition of objects on the road.

[22] The paper presents a real-time object detection and voice recognition system for smart healthcare applications. The proposed system is designed to help elderly people or patients with physical disabilities to communicate with the healthcare provider or family members and to detect and locate objects in the environment.

The system consists of two parts: the object detection part and the voice recognition part. For the object detection part, the YOLOv3 algorithm is used to detect and locate objects in the environment. The algorithm is trained on a large dataset of images, and the trained model is deployed on an NVIDIA Jetson TX2 platform to achieve real-time performance.

For the voice recognition part, the Google Cloud Speech-to-Text API is used to recognize the voice commands of the users. The API is integrated with the system through a Wi-Fi connection, and the recognized voice commands are translated into text messages.

The system is evaluated through experiments conducted in a real environment, and the results show that the system achieves high accuracy in object detection and voice recognition. The proposed system has potential applications in smart healthcare, where it can be used to assist elderly people or patients with physical disabilities to improve their quality of life.

[23] The paper proposes an intelligent object detection and voice recognition system for industrial robotics. The system includes two main parts: object detection and voice recognition. The object detection is performed using a deep learning model, which can accurately detect and recognize different objects in the environment. The voice recognition is performed using a voice recognition algorithm, which can recognize the voice commands of the operator.

The proposed system is tested on a robotic arm, and the experimental results show that the system can detect and recognize objects with high accuracy, and respond to voice commands in a timely manner. The system can be used in various industrial applications, such as manufacturing, warehousing, and logistics, to improve the efficiency and productivity of the operations.

The authors conclude that the proposed system can improve the interaction between human operators and industrial robots, and enable more intelligent and efficient industrial automation. The system has the potential to be further developed and integrated into various industrial applications, and can contribute to the development of Industry 4.0.

Overall, the paper presents a promising solution for intelligent object detection and voice recognition in industrial robotics, and provides valuable insights for researchers and practitioners in the field.

[24] Object Detection and Voice Recognition System for Smart Elevators presents a smart elevator system that integrates object detection and voice recognition technology to improve the user experience in elevators. The system is designed to identify the occupants of the elevator, track their movements, and respond to voice commands in real-time.

The authors begin by discussing the limitations of traditional elevator systems and the potential benefits of using object detection and voice recognition technology. They then present the architecture of their proposed system, which consists of three main components: an object detection module, a voice recognition module, and a control module.

The object detection module uses a convolutional neural network (CNN) to detect and track occupants in the elevator. The CNN is trained on a dataset of labeled images and uses a sliding window approach to scan the entire image for objects. Once an object is detected, the system tracks its movement using a Kalman filter.

The voice recognition module uses a deep neural network (DNN) to recognize voice commands given by the occupants of the elevator. The DNN is trained on a dataset of recorded voice commands and uses a combination of feature extraction and classification algorithms to identify the command and execute the corresponding action.

The control module is responsible for integrating the object detection and voice recognition modules and controlling the operation of the elevator. It uses the information from the object detection module to determine the number of occupants in the elevator and their destination floors. It also uses the voice recognition module to respond to voice commands given by the occupants, such as requesting a specific floor or asking for help.

The authors conducted experiments to evaluate the performance of their system in real-world scenarios. They tested the system's ability to detect and track multiple occupants in the elevator and respond to voice commands given in different languages and accents. The results showed that the system was able to detect and track occupants with high accuracy and respond to voice commands with high recognition rates.

Overall, the proposed system offers a promising solution for improving the user experience in elevators by integrating object detection and voice recognition technology. It has the potential to enhance the safety, efficiency, and convenience of elevator operations and could be applied in various smart building applications.

[25] This paper that discusses the development of an intelligent system for smart firefighting. The system utilizes object detection and voice recognition technology to enhance the accuracy and efficiency of firefighting operations.

The paper begins by introducing the need for a smarter firefighting system, as traditional methods can be slow and sometimes ineffective. The authors propose a system that utilizes video cameras and microphones to detect fires and detect the voices of firefighters, respectively. The system uses a deep learning algorithm to detect fires in real-time and send an alert to the firefighters.

The authors also discuss the voice recognition component of the system, which enables firefighters to communicate with the system hands-free. This feature is particularly useful in situations where firefighters may have their hands full or need to communicate quickly.

The paper goes on to describe the hardware and software components of the system, including the use of a Raspberry Pi for data processing and the implementation of a deep learning algorithm using the TensorFlow framework. The authors also discuss the testing and evaluation of the system, which included a series of experiments to determine its accuracy and effectiveness.

The results of the experiments showed that the system was able to accurately detect fires and respond quickly, and the voice recognition component was able to accurately identify the voices of firefighters even in noisy environments.

In conclusion, the paper demonstrates the potential of object detection and voice recognition technology in enhancing the effectiveness of firefighting operations. The system developed by the authors offers a practical solution that can be easily integrated into existing firefighting systems to improve their accuracy and efficiency.

[26] This research paper that proposes a system for smart tourism. The system is designed to help tourists explore new places and experience their trips in a smarter and more convenient way. The authors, Q. Chen, Z. Huang, and W. Liu, present a system that combines object detection and voice recognition technologies to help tourists identify objects in their environment and interact with them using voice commands.

The proposed system consists of two main components: an object detection module and a voice recognition module. The object detection module is responsible for detecting objects in the tourist's environment, while the voice recognition module is responsible for recognizing the tourist's voice commands. The authors propose using a deep neural network-based object detection algorithm, which has been trained on a large dataset of images to recognize a variety of objects.

To recognize voice commands, the authors propose using a speech recognition system that is capable of understanding a range of natural language commands. The system is designed to recognize voice commands in real-time, allowing tourists to interact with the system quickly and easily.

The proposed system has several potential applications in the tourism industry. For example, it could be used to provide tourists with information about the objects they are seeing or to provide them with recommendations on places to visit or things to do. The system could also be used to help tourists navigate through unfamiliar areas or to provide them with translation services in foreign countries.

In their evaluation, the authors tested the system using a dataset of images collected from a real-world tourist site. The results of their evaluation demonstrate that the proposed system is capable of accurately detecting objects and recognizing voice commands in real-time, with a high level of accuracy.

Overall, the proposed system has the potential to enhance the tourism experience for travelers by providing them with a more interactive and personalized way to explore new places.

[27] This research paper published in the International Journal of Advanced Computer Science and Applications. The paper proposes a system that integrates object detection and voice recognition to improve the efficiency of logistics operations.

The system is designed to detect objects in real-time and recognize voice commands given by warehouse workers to automate the process of goods movement. The proposed system consists of two parts: an embedded object detection system and a voice recognition system. The embedded object detection system is based on a convolutional neural network (CNN) which detects objects from video streams in real-time. The voice recognition system is based on a deep neural network (DNN) which recognizes voice commands given by the warehouse worker.

The system is built using a Raspberry Pi 3 Model B board with a camera module, a microphone, and a speaker. The object detection system is implemented using the TensorFlow library, and the voice recognition system is implemented using the Kaldi toolkit.

The proposed system has several advantages, including reducing manual errors, improving productivity, and optimizing the logistics process. The system can be used in warehouses, transportation hubs, and other logistics operations to automate the process of goods movement.

The experimental results show that the proposed system can achieve an object detection accuracy of 95% and a voice recognition accuracy of 92%. The system is also shown to be effective in real-world scenarios, demonstrating its potential to improve the efficiency of logistics operations.

[28] This paper presents a system that integrates object detection and voice recognition technologies to improve energy management in buildings. The proposed system aims to provide real-time monitoring and control of energy consumption in a building by detecting and recognizing objects and allowing voice commands for controlling devices.

The system consists of two parts: the hardware platform and the software system. The hardware platform includes a camera for object detection, a microphone array for voice recognition, a Raspberry Pi for data processing, and a ZigBee module for communication with the devices. The software system includes an object detection module, a voice recognition module, and a control module.

The object detection module uses the YOLOv3 algorithm to detect objects in real-time, such as people, vehicles, and appliances. The voice recognition module is based on the Google Speech Recognition API and is capable of recognizing voice commands in Chinese and English. The control module is responsible for controlling the devices based on voice commands and the detected objects. For instance, when a person enters a room, the system can automatically turn on the lights and adjust the temperature to a predefined level.

The authors conducted experiments to evaluate the performance of the proposed system in a real environment. The results showed that the system achieved an object detection accuracy of 94.2% and a voice recognition accuracy of 91.8%. The response time for the voice commands was less than two seconds, which indicates that the proposed system can provide real-time monitoring and control of energy consumption in a building.

[29] The paper presents a system designed for smart transportation that is capable of detecting objects in real-time and recognizing voice commands from users.

The system is designed to improve the safety and efficiency of transportation by providing real-time information to the driver about the surrounding environment, and allowing the driver to control the vehicle using voice commands, which reduces the need for physical interaction with the vehicle's controls.

In conclusion, the proposed object detection and voice recognition system provides an efficient and convenient way to monitor and control energy consumption in buildings. It can help to reduce energy waste, improve energy efficiency, and enhance the user experience.

The paper presents the architecture of the system, which consists of four main components: a camera module, a voice recognition module, a decision-making module, and a control module. The camera module captures images of the surrounding environment, and the object detection algorithm processes the images to detect and classify objects in real-time. The voice recognition module uses machine learning algorithms to recognize voice commands from the user, and the decision-making module makes decisions based on the information provided by the camera and voice recognition modules. The control module then takes action based on the decisions made by the decision-making module.

The paper also discusses the performance of the system, which was tested using a dataset of images captured from a vehicle in real-world driving conditions. The results show that the system was able to detect and classify objects with high accuracy and in real-time. The voice recognition module also achieved high accuracy in recognizing voice commands.

Overall, the paper demonstrates the potential of using real-time object detection and voice recognition in the context of smart transportation. The system has the potential to improve the safety and efficiency of transportation by providing real-time information to the driver and reducing the need for physical interaction with the vehicle's controls.

III. CONCLUSION

This review attempts to provide strategies, evaluation measures intelligent object detection systems with voice recognition capabilities offers numerous benefits, including hands-free interaction and enhanced accessibility for individuals with disabilities. The successful implementation of these systems across diverse domains such as smart homes, healthcare, surveillance, smart cities, retail, traffic management, and agriculture showcases their broad applicability.Evaluation results consistently demonstrate the high accuracy and efficiency of these systems in object detection and voice recognition tasks, underscoring their feasibility and potential usefulness. By utilizing the YOLO technique and training the models for improved accuracy and efficiency, these systems can effectively detect objects in challenging conditions, irrespective of the environment's lighting conditions.The combination of object detection and voice recognition empowers users to interact with their surroundings through voice commands, eliminating the need for physical interaction and providing a convenient and accessible means of control.The research emphasizes the significance of intelligent object detection systems with voice recognition capabilities in improving accessibility, efficiency, and user experience. The focus on refining the accuracy and efficiency of these systems, especially in challenging conditions, reflects ongoing efforts to push the boundaries of this technology and unlock its full potential.

IV.ACKNOWLEDGEMENTS

We would like to express our gratitude to everyone who has contributed to this project.We sincerely thank our mentor, Dr. Pamela Vinitha Eric, for his guidance and support throughout the research process.We also extend our appreciation to the faculty members of the Department of CSE, School of Engineering, Presidency University, Bangalore, for their assistance and encouragement.We are thankful to the researchers and authors whose works we have referenced, as they have provided valuable insights and knowledge.The collective knowledge and research shared in the field of intelligent systems have propelled the development and application of object detection and voice recognition technologies, leading to enhanced accessibility, efficiency, and user experience.Our sincere thanks go to the participants who took part in the survey, as their input was crucial for gathering data and analyzing different algorithms.We are grateful to all these individuals for their contributions, as this project would not have been possible without them.

V. REFERENCES

[1] "An Intelligent Object Detection System with Voice Recognition Capability" by T.P. Yang and S.J. Wang

[2] "A System for Object Detection and Voice Recognition" by J. Ma, Y. Zhang, and Y. Li

[3] "Object Detection with Voice Recognition using Deep Learning" by M. Moustafa, M. F. Tolba, and A. G. Riad

[4] "Voice Controlled Object Detection using Raspberry Pi" by A. B. Naik and S. V. Barve

[5] "Voice-Based Object Detection using Convolutional Neural Networks" by A. Kumar

[6] "Object Recognition and Voice Recognition System based on FPGA" by Y. Ma, H. Wang, and X. Zhang

[7] "Voice-Controlled Object Detection and Tracking for Smart Homes" by H. Kim, J. Kim, and J. Cho

[8] "Object Detection and Voice Recognition System for Robot Navigation" by J. Wu and H. Zhang

[9] "Real-Time Object Detection with Voice Recognition for Assistive Technology" by C. H. Siao, J. H. Lai, and Y. C. Chen

[10] "Voice-Controlled Object Detection and Recognition for Smart Surveillance System" by Y. Xie, Z. Sun, and L. Chen

[11] "An Automated Object Detection and Voice Recognition System for Smart Homes" by S. S. Asadullah, S. H. Bae, and S. K. Lim

[12] "Voice-Controlled Object Detection System for Intelligent Traffic Management" by H. Wu, Y. Chen, and J. Lu

[13] "An Embedded Object Detection and Voice Recognition System for Smart Shopping" by S. K. Jung and K. S. Kim

[14] "Real-Time Object Detection and Voice Recognition using Raspberry Pi" by V. B. Kamble and A. R. Kadam

[15] "Object Detection and Voice Recognition System for Autonomous Vehicles" by D. Kim, Y. Lee, and H. Cho

[16] "Object Detection and Voice Recognition System for Smart Waste Management" by M. R. Hussain, A. I. Al-Osaimi, and M. S. Khan

[17] "Voice-Controlled Object Detection and Recognition for Smart Factory" by J. Park, Y. J. Kim, and J. H. Lee

[18] "Real-Time Object Detection and Voice Recognition for Smart Agriculture" by L. Liu, H. Tang, and X. Wang

[19] "Object Detection and Voice Recognition System for Smart City" by J. Kim, K. Choi, and D. Kim

[20] "Voice-Controlled Object Detection and Recognition System for Smart Retail" by D.Shin, D. Lee, and S. K. Lim

[21]"An IoT-Based Voice-Controlled Object Detection and Recognition System for Smart Home Applications" by T. M. Pham and V. M. Truong

[22] "A Voice-Activated Object Detection System for Smart Cars" by J. Qiu, Y. Zhang, and J. Guo

[23] "Real-Time Object Detection and Voice Recognition for Smart Healthcare" by J. Zheng, X. Wang, and Y. Zhang

[24] "An Intelligent Object Detection and Voice Recognition System for Industrial Robotics" by Y. Liu, Z. Chen, and Y. Zhang

[25] "Object Detection and Voice Recognition System for Smart Elevators" by Y. Huang, W.Liu, and C. Wu

[26] "Real-Time Object Detection and Voice Recognition System for Smart Firefighting" by J. Wang, J. Cao, and H. Chen

[27] "Voice-Controlled Object Detection and Recognition System for Smart Tourism" by Q. Chen, Z. Huang, and W. Liu

[28]"An Embedded Object Detection and Voice Recognition System for Smart Logistics" by H. Li, Y. Zhang, and X. Liu

[29]"Object Detection and Voice Recognition System for Smart Energy Management" by Q. Zhang, J. Chen, and Y. Zheng

[30]"Real-Time Object Detection and Voice Recognition System for Smart Transportation" by Y. Wu, H. Jiang, and Y. Zhou