ABUAD Journal of Engineering Research and Development (AJERD)
ISSN (online): 2645-2685; ISSN (print): 2756-6811

Volume 7, Issue 1, 1-13

Robotic Assistant for Object Recognition Using Convolutional Neural Network

Sunday OLUYELE, Ibrahim ADEYANJU, Adedayo SOBOWALE

Department of Computer Engineering, Federal University Oye Ekiti, Oye, Ekiti State, Nigeria
sunday.oluyele.2826@fuoye.edu.ng/sunday.oluyele.2826@fuoye.edu.ng/ibrahim.adeyanju@fuoye.edu.ng/adedayo.sobowale@fuoye.edu.ng

Corresponding Author: sunday.oluyele.2826@fuoye.edu.ng, +2348130739580
Date Submitted: 14/11/2023
Date Accepted: 2/02/2024
Date Published: 12/02/2024

Abstract: *Visually impaired persons encounter certain challenges, which include access to information, environmental navigation, and obstacle detection. Navigating daily life becomes a big task with challenges relating to the search for misplaced personal items and being aware of objects in their environment to avoid collision. This necessitates the need for automated solutions to facilitate object recognition. While traditional methods like guide dogs, white canes, and Braille have offered valuable solutions, recent technological solutions, including smartphone-based recognition systems and portable cameras, have encountered limitations such as constraints relating to cultural-specific, device-specific, and lack of system autonomy. This study addressed and provided solutions to the limitations offered by recent solutions by introducing a Convolutional Neural Network (CNN) object recognition system integrated into a mobile robot designed to function as a robotic assistant for visually impaired persons. The robotic assistant is capable of moving around in a confined environment. It incorporates a Raspberry Pi with a camera programmed to recognize three objects: mobile phones, mice, and chairs. A Convolutional Neural Network model was trained for object recognition, with 30% of the images used for testing. The training was conducted using the Yolov3 model in Google Colab. Qualitative evaluation of the recognition system yielded a precision of 79%, recall of 96%, and accuracy of 80% for the Robotic Assistant. It also includes a Graphical User Interface where users can easily control the movement and speed of the robotic assistant. The developed robotic assistant significantly enhances autonomy and object recognition, promising substantial benefits in the daily navigation of visually impaired individuals.*

Keywords: *Convolutional Neural Network, Robotics, Object Recognition, Computer Vision, YOLO.*

1. INTRODUCTION

Robotic assistants are machines designed to work for human beings and with human beings [1]. Personal robotic assistant can help with day-to-day household tasks, making life more comfortable. Personal robotic assistants can help with day-to-day household tasks, making life more comfortable. These technological companions serve several purposes: from education to making the home run more efficiently, with less work for humans. They also make human homes smarter and technologically inclined to do repeated tasks easily and more efficiently [2]. Developing related systems is emerging from special-purpose robotic assistants to multi-functional robotic assistants, which can integrate diverse abilities such as person detection and tracking, object recognition and detection, human-robot interaction, etc. [3]. These robotic assistants are expected to be optimally functional and not limited to a particular location, time, and environment. They are also expected to guarantee the safety of their surrounding elements since their users are humans [4].

One area where these robotic assistants are useful is for visually impaired persons. Becoming visually impaired can be a life-changing experience and is likely to have far-reaching consequences for the person affected. Persons acquiring a visual impairment express a variety of emotional, cognitive, behavioural, and social responses to this significant loss [5]. Visually impaired people face many inconveniences when interacting with their surrounding environments and the most common challenge is to find dropped or misplaced personal items, which include keys, wallets, mobile phones, books, umbrellas, shoes, headsets, wristwatches, water bottles, pen, credit cards and sunglasses [6]. The person faces restrictiveness in closing most of the elemental tasks of their life, like recognizing items in a shopping mall, getting a chair in a gathering, and picking stuff up at random places [7]. Hence, there is a need for automated machines to assist in object recognition.

Initially, there were traditional methods used to assist visually impaired persons, which included Guide dogs [8], white canes [9] and Braille [10]. Recently, assistive technologies developed were smartphone-based only recognition systems, providing real-time identification and location through audio feedback [11]. However, these solutions were constrained to the limitations of smartphones, such as reduced accuracy in complex environments [11]. Another approach explored using a portable camera for object recognition and navigation assistance, offering hands-free operation and tactile feedback [12]. Yet, the potential impact of camera angles and lighting on accuracy were notable drawbacks [12]. A recent study aimed at Arabic-speaking users also implemented a deep learning approach for object recognition and audio feedback [13]. Despite

its cultural relevance, challenges such as dialect-specific training data requirements and privacy concerns were identified [13].

In contrast, this study takes a significant step forward by integrating a Convolutional Neural Network (CNN) object recognition system into a mobile robot designed as a robotic assistant. This shift to a mobile robotic assistant not only addresses the limitations of existing solutions but also emphasizes mobility, which solves camera angles and lighting issues. Doing this enhances the autonomy of visually impaired individuals in dynamic environments. The significance lies in providing a mobile robotic assistant capable of independently moving in a confined environment, recognizing objects, and delivering audio feedback, filling a gap in the current landscape of assistive technologies for the visually impaired.

This work aims to develop a Convolutional Neural Network-based robotic system to help persons with visual impairment recognize common objects in their environment. The specific objectives are to design a Convolutional Neural Network object recognition system, implement the designed system as a mobile robotic assistant to identify computer mice, mobile phones, and chairs, and evaluate the developed model's effectiveness and efficiency.

The paper comprised of the following key areas: Related Works, where we review prior research and provide context for this study; Methodology, in which we detail the technical aspects of the robotic assistant; Results, showcasing the performance of the system, which was backed by quantitative and qualitative data; and finally, a Conclusion that summarizes the findings, which emphasizes the significance of this work.

2. RELATED WORKS

In the quest to address issues relating to visually impaired people, several solutions have emerged, each striving to alleviate the challenges posed by visual impairment and enhances the daily lives of affected individuals. This section comprehensively explores related solutions and works, traversing subjects such as computer vision, artificial intelligence, assistive technologies for the visually impaired, artificial intelligence, object detection and recognition, robotics, and deep learning.

Computer vision, a branch of AI, utilizes algorithms and optical sensors to mimic human visual perception and extract valuable information from objects. Combined with lighting systems, computer vision facilitates image acquisition and analysis [14]. AI's capability to analyze large volumes of data has been harnessed in e-commerce for tasks such as identifying browsing patterns, personalizing shopping experiences, fraud prevention, and decision-making based on credit checks and account information [15]. In conjunction with GPS technology, AI enhances safety by providing accurate and detailed information, such as lane detection and road type identification, even when obstructions are present [16]. AI has numerous medical applications across clinical, diagnostic, rehabilitative, surgical, and predictive practices. AI technologies can process and analyze vast amounts of data from different sources to detect diseases and guide clinical decision-making [17]. AI is also proficient in agriculture, where it can identify soil defects, nutrient deficiencies, and weed growth using computer vision, robotics, and machine learning.

Additionally, AI-powered bots can harvest crops at a higher volume and faster pace than human labourers' [18]. Natural Language Processing (NLP), another application of AI, enables humans to interact with machines. NLP aids in speech recognition, document summarization, machine translation, spam detection, named entity recognition, question answering, and predictive typing [19].

Rui Pedro and Figueiredo Garcia [20] studied object detection algorithms in digital images and their application in mobile robots. They developed a computer vision system integrated into the CAMBADA@Home robot from the University of Aveiro, aiming to recognize everyday objects. Two approaches were studied: one based on color histograms and the other based on image descriptors (SIFT and SURF) [21]. However, the study could be improved by adding the ability to learn new objects online and storing their visual descriptors in the database.

Ester Martinez-Martin and Angel P. del-Pobil [1] proposed a robot capable of detecting and recognizing objects in realistic, unconstrained scenarios in real-time. The system combined object-specific color, motion, and shape cues probabilistically to achieve real-time performance. The implementation required a Graphical Processing Unit (GPU) for efficient processing. The architecture consisted of three main modules: feature extraction, memory, and recognition [22]. They addressed the challenges assistive robots face in real-time object detection and recognition due to dynamic, cluttered environments and variations in lighting and object appearance.

Object recognition algorithms such as R-CNN, fast R-CNN, Faster R-CNN, and Single Shot Multi-Box Detector (SSD) have been developed in computer vision. R-CNN, introduced by Ross Girshick [23], combines region proposals with convolutional neural network (CNN) features for object localization and segmentation [24]. Fast R-CNN improved upon R-CNN by incorporating classification and bounding box regression, leading to more efficient processing [23] . Even further, Faster R-CNN introduced a region proposal network that eliminated the need for selective search, resulting in faster and more accurate object detection [24].

Single Shot Multi-Box Detector (SSD) discretizes the output space of bounding boxes into a set of default boxes, considering different aspect ratios and scales per feature map location [20]. This approach enables real-time object recognition with slightly lower accuracy than Faster R-CNN.

YOLO v3 (You Only Look Once) is another popular architecture that combines object recognition and detection. It operates at an impressive speed of 45 frames per second, achieving accuracy comparable to SSD but with three times faster processing time [24]. YOLO v3 uses TensorFlow and has been proven effective for recognizing multiple objects.

These algorithms have evolved to address issues related to speed and accuracy. While R-CNN was a significant step forward, subsequent methods like Fast R-CNN, Faster R-CNN, SSD, and YOLO v3 have achieved even faster and more accurate object recognition capabilities [24].

Technological advancements have been made to assist the visually impaired and promote independence [25]. While trained dogs and white canes are commonly used, they have limitations in providing a comprehensive understanding of the surroundings and long-range navigation [26]. Blind sticks are widely used, with approximately 48% of the blind population in the USA relying on them. The number of visually impaired individuals, especially those aged 50 and below 15, is expected to increase [27]. Nguyen et al. [28] proposed a cost-effective system for visually impaired individuals, utilizing a web-based object detection application based on the SSD-MobileNetV2 algorithm. Their system demonstrated high accuracy in single-object detection but encountered challenges with multiple objects of the same category.

Meanwhile, Potdar et al. [29] presented a live object recognition system employing a Convolutional Neural Network. The system uses a camera for real-time object detection and provides feedback through either audio or Braille text. Bhandari et al. [30] conducted a comprehensive review of deep learning systems applied to navigational tools for the visually impaired, emphasizing the significant impact of convolutional neural networks (CNN) and fully convolutional neural networks (FCN) in developing multifunctional technology. In a separate work, Parikh et al. [31] proposed an Android smartphone-based visual object recognition model for guiding visually impaired individuals in outdoor environments. The system employs a boosted optimized CNN, hosted on a cloud-connected server, to achieve measurable high accuracy in recognizing 11 outdoor objects. The visually impaired user captures video frames, sends them to the server for object recognition, and receives auditory feedback through a connected hearing device, offering an effective alternative to traditional guiding canes.

Shirley et al. [32] proposed a sensor-based multi-robot system with voice recognition to aid visually impaired individuals in navigating their surroundings. Equipped with Lidar, proximity sensors, and a Bluetooth transmitter and receiver, the system detects obstacles and provides real-time information to users through a Bluetooth headset. The machine learning algorithm, developed in Python, processes sensor data to deliver instructions through a speaker, offering enhanced independence and mobility for visually impaired individuals. Meanwhile, Najm et al. [33] focused on object detection with vocal feedback, employing the YOLO model for real-time object detection using a web camera. The system, utilizing OpenCV libraries and Google text-to-speech, delivers audible information about object locations to visually impaired users. Evaluation based on mean Average Precision (mAP) indicated good performance compared to previous approaches. Also, Adeyanju et al. [34] focused on developing a Convolutional Neural Network (CNN)-based object recognition system for uncovered gutters and bollards, often overlooked in urban and educational environments. The system, implemented with Python, OpenCV libraries, and Yolov4, demonstrated potential for aiding outdoor navigation for the visually impaired.

On the other hand, Gautam et al. [35] proposed a vision system with audio feedback to assist visually impaired individuals in grasping objects, addressing usability and complexity challenges associated with existing technologies. The system utilizes a Weighted Matrix Algorithm for object recognition, providing audio guidance to the user for locating and reaching desired objects, exemplified through the recognition of a water bottle. These advancements enhance the autonomy and accessibility of visually impaired individuals in diverse environmental settings.

Breve and Fischer [36] proposed a visually impaired aid system leveraging convolutional neural networks (CNN), transfer learning, and semi-supervised learning. The framework, designed for low computational costs and smartphone implementation, utilizes a smartphone camera to capture and classify images of the user's path. The proposed dataset for classifier training encompasses various indoor and outdoor scenarios, accommodating different lighting, floor types, and obstacles. The study achieves 92% and 80% classification accuracy in supervised and semi-supervised learning scenarios. Caballero et al. [37] contribute to inclusivity with a mobile application employing CNN for object recognition and text-to-speech technology to cater to the needs of the visually impaired community. The application analyzes images and provides auditory feedback, aligning with the United Nations' Sustainable Development Goals for disability inclusion. Bine, Costa, and Aylon [38] focus on automata recognition using CNNs for visually impaired individuals studying computer science. The method involves the late fusion of three CNNs, achieving 97% accuracy for automaton-type recognition and 91% for recognizing the number of states. Shaikh, Karale, and Tawde [39] introduce a cost-effective robotic system utilizing Raspberry Pi and the YOLO v3 algorithm for object recognition, attaining state-of-the-art results with 85% to 95% overall performance.

Jessica et al. [40] developed a Convolutional Neural Networks (CNN) based voice assistive system exclusively designed to aid blind individuals. This robotic system employs artificial vision, utilizing a high-resolution camera to capture images of the surroundings and open computer vision to process and detect objects. The CNN technique is then applied to process the captured image, compare it with the dataset, label it, and provide real-time audio feedback using a speaker and ultrasonic sensor, enabling blind users to navigate both indoor and outdoor environments efficiently. On a similar note, Kinra et al. [41] conducted a comprehensive review of deep learning-based object recognition techniques for the visually impaired. They highlighted existing technologies like the Path Force Feedback Belt, Eye Substitution, and Radio Frequency Identification Walking Stick, which utilize sensors for object detection but have limitations such as restricted detection range and lack of real-time response.

Additionally, a solution for indoor and outdoor navigation for the visually impaired was proposed by Alisetti et al. [42] . This personalized assistant integrates voice interaction with a reliable object detection module, ensuring efficient navigation by providing essential information through voice requests while minimizing unnecessary alerts and confusion for the user. Dahiya et al. [43] introduced a real-time assistive framework using Faster R-CNN with ResNet50 to identify public amenities for visually impaired individuals. Achieving 92.13% accuracy, the system recognizes symbols associated with facilities like restrooms, ATMs, metro stations, and pharmacies. Hsieh et al. [44] proposed a wearable guide device employing convolutional neural networks (CNN) to help blind or visually impaired individuals identify flat and safe walking routes. Using an RGB camera and deep learning, the device enhances independence and safety, providing valuable environmental information beyond traditional white canes.

In conclusion, the works reviewed have collectively aimed to assist visually impaired individuals in achieving autonomy and mobility. However, a notable limitation across these studies is the need for integration into a mobile robotic assistant for local environment navigation. This study bridges this gap by incorporating the object recognition model into a microprocessor and using a mobile robotic assistant that can adeptly navigate the local environment of visually impaired individuals. Furthermore, the developed model in this study addresses the specificity of objects encountered in real-life scenarios; for instance, Shaikh et al. [39] only relied on the COCO dataset while the acquired data for this study was from the local environment for training. This contribution enhances the practicality of assistive technologies for the visually impaired daily.

3. METHODOLOGY

The proposed object recognition system utilizes a single-board computer, specifically the Raspberry Pi, and a 5 Mega Pixel Raspberry Pi camera module. This lightweight camera can capture images and communicate with the Raspberry Pi within the Robotic Assistant for image processing. The system operates by continually streaming the environment using the Pi Camera and then processing it through the Raspberry Pi to recognize objects.

The Robotic Assistant is trained with data collected with images of chairs, mobile phones, and computer mice against which it matches the objects in the streams. When an object is accurately recognized, the Robotic Assistant sends a signal to the audio unit, which generates a sound saying the name of the recognized object.

3.1 Hardware Components

The system's design is divided into five main modules: a power supply, an input unit, a control unit, a robotic car, and a display unit (Liquid Crystal Display - LCD). The power supply module consists of a battery bank and a DC to DC step-up converter (module empty 3608) that converts the 5V DC input from the battery bank to a 12V DC output required to drive the DC motors within the robotic car. The battery bank supplies power to the Raspberry Pi and the DC motors.

A block diagram illustrating the hardware and software components of the system encompassing all the modules mentioned above is shown in Figure 1 and Figure 2.

This work utilized a battery bank consisting of four 15 V tiger-head batteries. The batteries are connected using series or parallel wiring, which enables more power storage than a single battery. A battery bank is a configuration that allows the storage of electricity generated by a solar PV system or other power sources, making it available anytime. The term "battery bank" is often used interchangeably with "power bank" or "battery pack." In various contexts, such as cordless tools, radio-controlled hobby toys, and battery-electric vehicles, the term "battery pack" is commonly used. The Pebble Pico Pocket-Sized 10,000 mAh Power Bank with a 2.1 A output was selected for this work. This power bank serves as the source of 5V DC power required by the system, and it draws power from the battery bank to provide the necessary voltage.

In computing, an input device is a peripheral (computer hardware equipment) that provides data and control signals to an information processing system, such as a computer or another information device. For this work, the input unit is classified as a visual input: the Raspberry Pi Camera and the HC-SR04 Ultrasonic sensor. The HC-SR04 is a 4-pin module, with the pins named Vcc, Trigger, Echo, and Ground, respectively. This sensor is widely used in various applications requiring distance measurement or object sensing. The module has two eye-like objects in the front, which form the Ultrasonic transmitter and receiver.

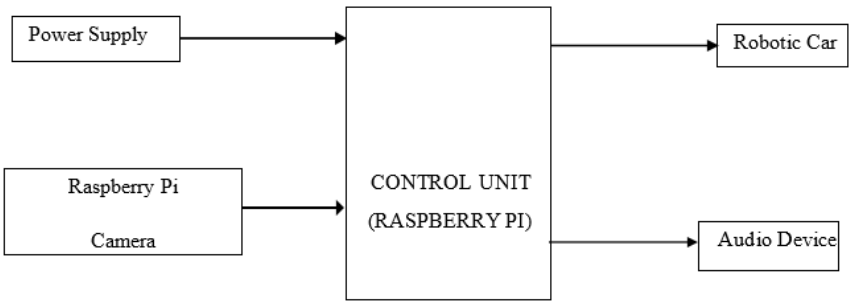

Figure 1: Block diagram of the proposed system's hardware component

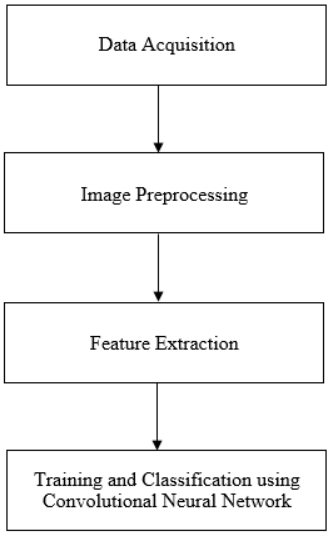

Figure 2: Block diagram of the proposed system's software component

The control unit consists of Raspberry Pi 3A+ (shown in Figure 3), a DC Motor – 100RPM 5V, and an L298N Motor Driver Module. This unit controls the Robotic Assistant, allowing it to move forward when the forward button is pressed/clicked, backward when the back button is pressed/clicked, left when the left direction button is pressed/clicked, and right when the right direction button is pressed/clicked. The system achieves this with the help of the DC motors, the L298N module, the Robotic wheels, and the Raspberry Pi 3A+.

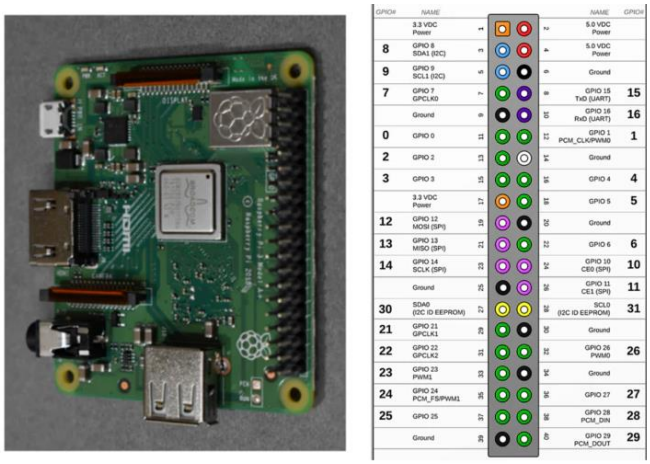

Figure 3: The Raspberry Pi 3a+ Pin configuration [16]

The output unit consists of a compact speaker with a 3.5mm jack connected to the Raspberry Pi audio port that allows the Robotic Assistant to output the pronunciation of the object that the Robotic Assistant has just recognized.

The robotic assistant was coupled using the materials listed in the previous paragraphs, such as robotic wheels, DC motors for movement, and a Pi Camera for image capture. The robotic assistant for object recognition was designed to navigate areas inside a house without sustaining any damage. The wheels are made with spur teeth gear motors, which enable the robotic assistant to move smoothly even when encountering small objects along its path. Figure 5 shows the coupled Robotic Assistant after implementation. Fritzing 0.9.3 was used for the schematic design of the circuit, which helped in the interconnection of the components used for the robot and also helped reduce damage to components due to bad connections. The schematic diagram in Figure 4 comprises a 1k ohm resistor, two (2) DC motors, an L293D module, a Raspberry Pi A+, an HC-SR04 Ultrasonic Sensor, and a 2k ohm resistor. The details of the diagram are shown in Figure 4.

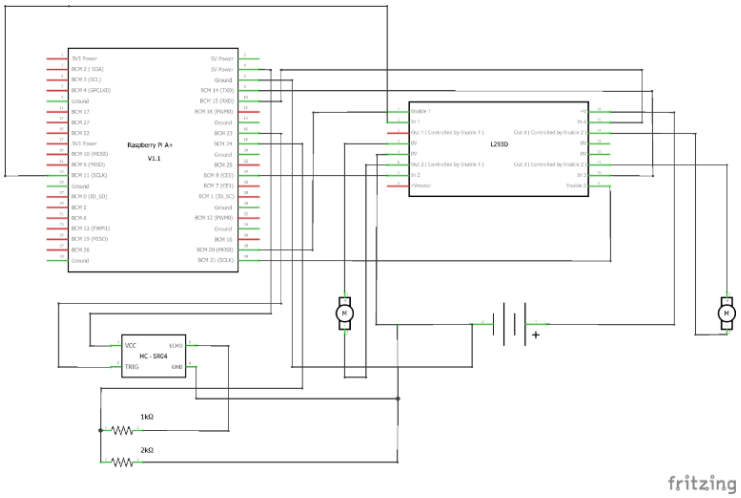

Figure 4: Schematic Circuit Design

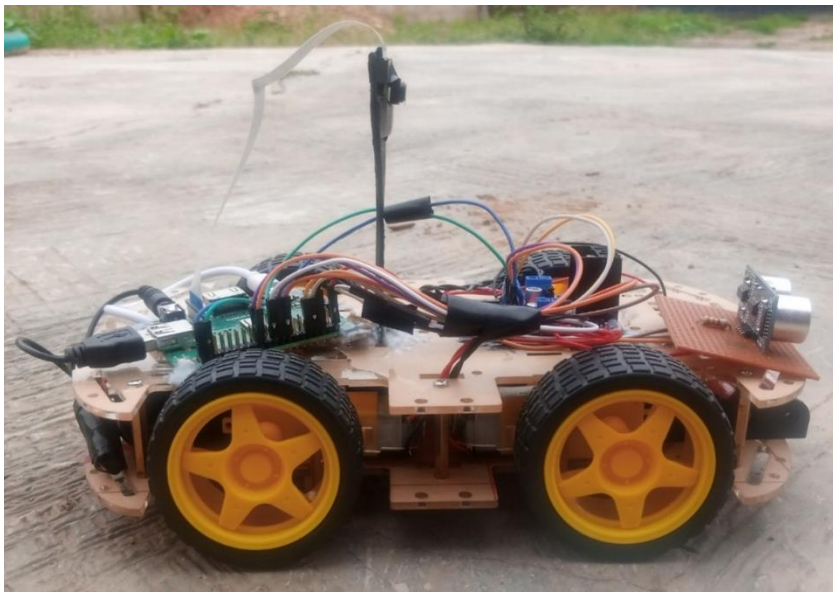

Figure 5: The complete Robotic Assistant for Object Recognition

3.2 Software Components

The software development of this work involves a series of steps, including data acquisition, image pre-processing, feature extraction, and training and classification using a convolutional neural network. Several tools were used to realize object recognition, including Microsoft Visual Studio and the Integrated Development Environment used for programming in Python. A few Python libraries were also utilized during development, such as PyTorch, OpenCV, TensorFlow Lite, Matplotlib, Keras, NumPy, etc.

3.2.1 Data acquisition

There are 2,895 images taken in broad daylight at Ikole-Ekiti State, Nigeria. Out of the total, 939 were captured for chairs, 932 for mice, and 1,024 for mobile phones, all with a resolution of 2,432 x 2,432 pixels. There are separate folders for each class. Table 1 summarizes the captured data's class, number, and size.

Table 1: Dataset acquisition summary

S/N	Class	No of Images	Resolution
1	Chair	939	2432 by 2432
2	Computer Mouse	932	2432 by 2432
3	Mobile phone	1024	2432 by 2432
4	Total	2,895	

3.2.2 Data pre-processing

Pre-processing is the step taken to format images before they are used for model training and inference; this includes resizing, orienting, and color corrections, among other things. Pre-processing aims to remove excessive white spaces, correct distortions at the edges, and detect and correct skew. In the pre-processing stage, the images were first cropped to focus on the objects for recognition. Then, padding was added to extend the area processed by the convolutional neural network (CNN). The kernel, the neural network filter, scans each pixel and converts the data into a smaller or larger format. Padding helps the kernel cover the image more effectively, allowing for more accurate analysis. Lastly, a Python script was used to resize the images to 224px by 224px.

Image annotation assigns labels to objects in an image, which can be done manually or automatically. The annotation of objects in pictures creates metadata for the objects in the dataset. The labels are predetermined by a machine learning engineer and selected to provide information to the computer vision model. As shown in Figure 6, labelImg was used for annotations for the robotic assistant.

Figure 6: Annotation of a mobile phone

3.2.3 The CNN architecture and feature extraction

This work uses a CNN model characterized by a structured architecture consisting of two convolutional layers, two pooling layers, and fully connected layers, as shown in Figure 7. The initial hidden layer adopts a convolutional structure with Rectified Linear Units (ReLU) as the activation function.

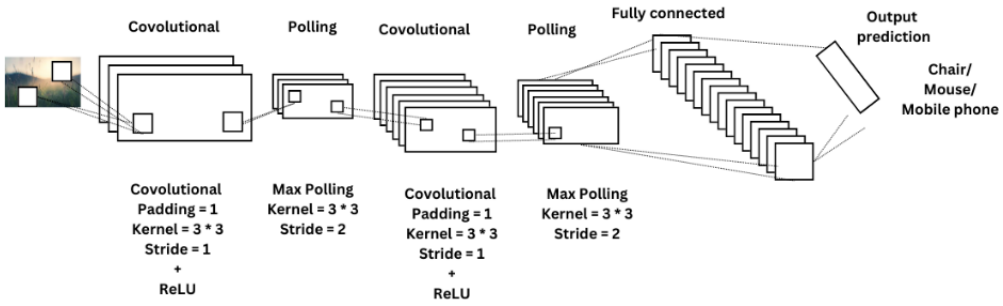

Figure 7: The architecture of the CNN model

For this work, YOLO was used, which is performed in the following three steps – firstly, the input image is divided into a G x G grid; then, a CNN that predicts y was run for each grid in the cell following the formulae in equation 1, where Pc represents the probabilities of detecting the objects and bx, by, bh and bw represents the properties of the bounding box, c1, c2, c3, …, cp is representation of the p-classes while k represents how many anchor boxes in the detection. Lastly, an algorithm to remove duplicate overlapping bounding boxes is run.

$$y = \left[p_c, b_x, b_y, b_h, b_w, c_1, c_2, \ldots, c_p, \ldots\right]^T \in \mathbb{R}^{G \times G \times k \times (5+p)} \quad (1)$$

Feature extraction starts with an initial set of measured data and constructs derived values (features) that are intended to be informative and non-redundant. This process facilitates subsequent learning and generalization steps and sometimes leads to better human interpretations. Feature extraction is closely related to dimensionality reduction. After completing the pre-processing, the pre-processed images are represented as a matrix containing pixels of large size. This representation is valuable for capturing the necessary information about the characters in the images. This process is referred to as feature extraction.

3.2.4 Training on Google Colab

This process was conducted in Google Colab using Python 3 and a Google Compute Engine backend with a GPU configuration of 2.50G and 2.16G of memory allocated for training. The training time was minimized using OpenCV version 3.2.0 and a GPU Tesla t$. The Yolov3 model was implemented by cloning the Darknet repository in Google Colab's virtual environment. Initially, the runtime was switched to GPU, and the Darknet GitHub repository was cloned into Google Colab's virtual environment. To manage the training images, a folder was created in Google Drive. Several key files, including a configuration file, were generated for the training process. The configuration file was configured to specify various training parameters. Google Drive was linked and mounted to the parent directory. OpenCV and GPU functionality were enabled by modifying the "makefile" in the Darknet folder. Subsequently, the run command was executed to compile the Darknet files, which were transferred from Google Drive to the Darknet directory within the Google Colab virtual machine. Scripts were executed to create the train and test data lists. In addition, another script generated a file containing essential annotations for the training data. After the completion of training, the detector training process was initiated. The weights file was downloaded after successful training, and the system's performance was evaluated.

3.2.5 Implementation

The implementation of this work was divided into two main aspects: hardware and software. The hardware component encompasses the control unit, power supply, audio jack unit, Raspberry Pi camera, and the robotic car. Meanwhile, the software aspect involved the writing of codes, compilation, and uploading to the Raspberry Pi, as well as the creation of a web-based application using the PHP programming language for controlling the robot with the following functions - move forward, left, right, backward, stop and speed, a screenshot of this Graphical User Interface can be found in Figure 7.

The operational sequence is illustrated in the flowchart in Figure 9, describing the systematic procedure for the developed system. Upon turning on, the system initializes key components, including the camera and Raspberry Pi. The user connects with the Raspberry Pi through WiFi, gaining control via a web portal (GUI in Figure 8). Subsequently, the system engages the camera to capture live feeds, performs real-time object recognition, and extracts essential information such as class ID, confidence level, and bounding box for detected objects. A text-to-speech conversion occurs, articulating the identified object through the speaker. The user can control the robot's speed and direction through the web portal concurrently with the ongoing object recognition process.

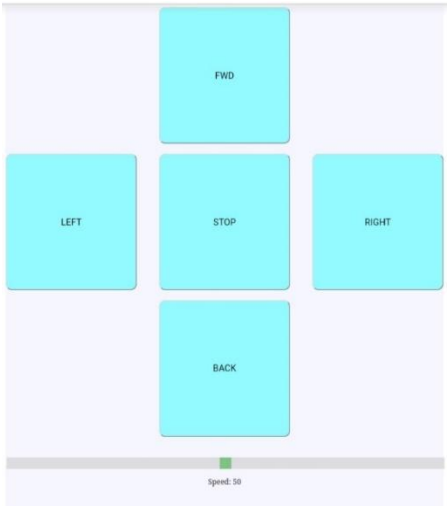

Figure 8: Graphical user interface for controlling robot via web

3.3 Performance Metrics

The system was evaluated using empirical and qualitative evaluations. The empirical evaluation, which relates to the real-time usage of the robotic assistant, comprises accuracy, which is the degree to which the result of a measurement, calculation, or specification conforms to the correct value or standard. The accuracy of this work will be determined by how closely the object pronounced by the robotic assistant matches the object streamed by the Pi camera.

The qualitative evaluations comprise four (4) metrics: precision, recall, accuracy, and f1-score; these metrics only concern the object detection section of the robotic assistant. Table 2 addresses the parameters for these metrics and their descriptions:

- Precision: The ratio of true positive predictions to the total predicted positives and is expressed by:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- Recall: The ratio of true positive predictions to the total actual positives and is expressed by:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- Accuracy: The ratio of correct predictions to the total predictions and is expressed by:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

- F1-Score: A balanced metric that combines precision and recall, computed as:

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5)$$

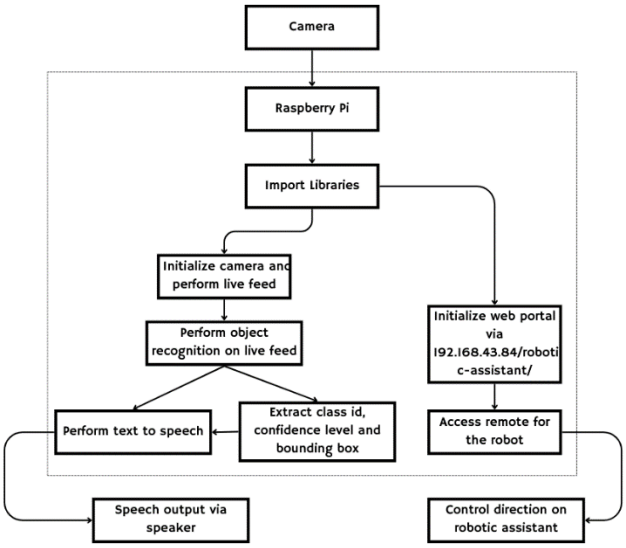

Figure 9: Flowchart of the system.

Table 2: Performance metrics parameters and their descriptions

Value	Description
True Positive (TP)	This is an outcome when the Robotic assistant correctly recognize the object
False Positive (FP)	This is an outcome when the Robotic Assistant incorrectly recognize an object as one of the classes.
True Negative (TN)	This is an outcome when the Robotic Assistant correctly recognize an object not in any of the classes.
False Negative (FN)	This is an outcome when the Robotic Assistant incorrectly recognize an object not in any of the classes.

4. RESULTS AND DISCUSSIONS

The trained model underwent testing to come up with some results and evaluate its performance. The results in Table 3 summarize the empirical accuracy metrics described in the methodology section. In real-time testing, the robot demonstrated a recognition rate of 5 out of 10 for chairs, 10 out of 10 for computer mice, and 8 out of 10 for mobile phones. Moving on to the qualitative evaluation results, Table 4 showcases the recognition outcomes using a confidence threshold 0.5. The model exhibited a precision of 0.79, recall of 0.96, accuracy of 0.8, and an f1-score of 0.87, providing insights into the overall effectiveness of the object recognition system. Further details regarding average precision, false positives, and true positives for each class are presented in Table 5. Average Precision quantifies the weighted mean of precisions for each class at a threshold. The results Average Precision values of 93.77% for mobile phone, 94.12% for computer mouse, and 93.33% for Chair, reflecting the model's effectiveness in accurate object recognition.

Table 3: Empirical results of the system

Class	Numbers tested	Numbers recognized
Chair	10	5
Computer Mouse	10	10
Mobile phone	10	8
Total	30	23

Table 4: Result of all classes' evaluation

Metric	Value
Confidence Threshold	50%
Precision	0.79
Recall	0.96
Accuracy	0.80
F1-Score	0.87
True Positive	99
False Positive	26
False Negative	4
True Negative	21
Average Confidence Level (IoU)	80.26%

Table 5: Average precision for each classes

Class	True Positive	False Positive	Average Precision
Mobile Phone	21	15	93.77%
Mouse	57	5	94.12%
Chair	21	6	93.33%

It is vital to note that the testing dataset covers 30% of the total images used in developing the model. The distribution details can be found in Table 6.

Table 6: Dataset acquisition summary

Class	No of Images for Training	Number of images for test
Mobile Phone	167	50
Mouse	167	50
Chair	167	50
Total	501	150

4.1 Comparing with other studies

In comparing the developed robotic assistant with existing studies (details in Table 7), the system achieved an accuracy of 80%. While this accuracy is slightly lower than the reported results of [13] (83.9%) and [39] (85%-95%), this study's innovation lies in the integration of a mobile robot, Raspberry Pi, Pi camera, and speaker into a more comprehensive solution.

Table 7: Comparing the Robotic Assistant with other studies

Author	Accuracy	Implementation method
[45]	69%	Used a wearable camera
The developed Model	80%	Mobile robot + Raspberry pi + Pi-camera + Speaker
[13]	83.9%	Raspberry pi in a box + Pi-camera + Speaker
[39]	85%-95%	Software

4.2 Marketability of the system

The analysis (presented in Table 8) of comparable assistive devices for visually impaired individuals reveals that the developed Robotic Assistant for Object Recognition, priced at $172.59, offers a more cost-effective solution than existing alternatives. Popular products such as the OrCam MyEye Pro and Yahboom ROS Robot Kit are priced at $4,250.00 and $579.99, respectively. This system provides a competitive advantage in terms of cost efficiency, making it a more accessible and budget-friendly option for individuals seeking assistance with object recognition and navigation.

Table 8: Cost analysis of related home automation device visually impaired persons

S/N	Product name	Manufacturer	Price
1	OrCam MyEye Pro - The Most Advanced Wearable Assistive Device for The Blind and Visually Impaired. Featuring Smart Reading, Face Recognition, Color & Product Identification, Orientation	OrCam	$4,250.00
2	Yahboom ROS Robot Kit Sunburst RDK X3 Based on Horizon Sunrise X3 Board SLAM Map Navigation Depth Camera Visual Recognition ROS2 Educational Robotic	Yahboom	$579.99
3	Ray Electronic Mobility Aid for the Blind	Caretec	$299.95
4	WeWALK Ultrasonic Smart Cane	WeWALK	$325.00
5	Robotic Assistant for Object Recognition using CNN	-	$172.59

5. CONCLUSION

This work represents a journey towards enhancing autonomy and independence for visually impaired individuals. We integrated a mobile robot into existing assistive technologies, making it stand out as a distinctive and innovative feature in this present study, offering significant benefits in the real world. While the accuracy of the developed model was measured to be 80%, this was counterbalanced by the unique functionalities introduced to make the system a mobile robotic assistant. This study work contributes to the world of existing assistive technologies and holds promise for real-world applications helping visually impaired individuals have a better life. Further advancements could be offered by incorporating a more extensive dataset, allowing the system to recognize a broader range of objects. Additionally, refining the model through continuous learning could enhance its accuracy.

REFERENCES

[1] Ester, M., & Angel, d. P. (2017). Object Detection and Recognition for Assistive Robots. IEEE Robotics & Automation Magazine, 1(1), 1-12.
[2] Anderson, J., & Rainie, L. (2018). Improvements ahead: How humans and AI might evolve together in the next decade. Pew Research Centers: Artificial Intelligence and the future of humans, 4.
[3] Alisha, P., Kanika, M., & Leah, F. (2018). Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 18, 1-13.
[4] Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Soler, G. J., & Brauchitsch, B. v. (2017). Artificial Intelligence and Robotics and Their Impact on the Workplace. IBA Global Employment Institute, 1, 1-120.
[5] Stevelink, S., Malcolm, E., & Fear, N. (2015). Visual impairment, coping strategies and impact on daily life: a qualitative study among working-age UK ex-service personnel. BMC Public Health, 1, 1-15.

[6] Yi, C., Flore, R. W., R. C., & Tian, Y. (2013). Finding Objects for Assisting Blind People. Network Modeling Analysis in Health Informatics and Bioinformatics, 2, 71-79.
[7] Singh, G., Kandale, O., Takhtani, K., & Dadhwal, N. (2020). A Smart Personal AI Assistant for Visually Impaired People. International Research Journal of Engineering and Technology (IRJET), 7(6), 1450-1454.
[8] Rickly, J., Halpern, N., Hansen, M., & Welsman, J. (2021). Travelling with a Guide Dog: Experiences of People with Vision Impairment. Transport Inequalities, Transport Poverty and Sustainability, 13(5), 2840.
[9] Attia, I., & Asamoah, D. (2020). The White Cane. Its Effectiveness, Challenges and Suggestions for Effective Use: The Case of Akropong School for the Blind. Journal of Education, Society and Behavioural Science, 33(3), 47-55.
[10] Encalada, E. G., Jordán, C. d., Chicaiza, V. E., & Pazmiño, S. J. (2022). Enhancing Reading Competence through the Braille System for Visually Impaired People: A Preliminary Study. International Journal of Teaching and Learning, 1(1), 65-77.
[11] Shaikh, S. (2020). Assistive Object Recognition System for Visually Impaired. International Journal of Engineering Research & Technology (IJERT), 9(9), 736-740.
[12] Mohane, V., & Gode, C. (2016). Object recognition for blind people using portable camera. 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), 1-4.
[13] Alzahrani, N., & Al-Baity, H. (2023). Object Recognition System for the Visually Impaired: A Deep Learning Approach using Arabic Annotation. Applications of Neural Networks for Speech and Language Processing, 12(3), 541.
[14] Wiley, V., & Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. International Journal of Artificial Intelegence Research, 2(1), 28-36.
[15] Kakkar, S., & Monga, V. (2017). A STUDY ON ARTIFICIAL INTELLIGENCE IN E-COMMERCE. International Journal of Advances in Engineering & Scientific Research, 4(4), 62-68.
[16] Aswinvenu. (2019). Element14 Community. Retrieved 2023, from https://community.element14.com/products/roadtest/rv/roadtest_reviews/648/raspberry_pi_3_model_5
[17] Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. BMC Medical Informatics and Decision Making, 21(1), 1-23.
[18] Biswal, A. (2021). Top 10 Artificial Intelligence Applications. Simplilearn Journal on Artificial Intelligence, 1, 50-62.
[19] Khensous, G., Labed, K., & Labed, Z. (2023). Exploring the evolution and applications of natural language processing in education. Revista Română de Informatică și Automatică, 33(2), 61-74.
[20] Pedro, R., & Garcia, F. (2015). Object recognition for a service robot. Master's thesis, University of Aveiro,, 1(1), 1-80.
[21] Wei, L., Dragomir, A., Dumitru, E., Christian, S., Scott, R., Cheng-Yang, F., & Alexander, B. (2015). SSD: Single Shot MultiBox Detector. Journal for Computer Vision and Pattern Recognition, 5, 21-37.
[22] David, C. C., Fiorella, S., & Marina, I. (2022). A Framework for Safe and Intuitive Human-Robot Interaction for Assistant Robotics. In Proceedings of the 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation, 1-4.
[23] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(1), 142-158.
[24] Anand, J., & Divyakant., M. (2020). A Comparative Study of Various Object Detection Algorithms and Performance Analysis. International Journal of Computer Sciences and Engineering, 8, 158-163.
[25] Zientara, P. A., Lee, S., Smith, G. H., Brenner, R., Itti, L., Rosson, M. B. & Narayanan, V. (2017). Third Eye: A Shopping Assistant for the Visually Impaired. IEEE Computer Society Journal, 50(2), 16-24.
[26] Elmannai, W., & Elleithy, K. (2017). Sensor-Based Assistive Devices for Visually-Impaired People: Current Status. Sensors for Globalized Healthy Living and Wellbeing, 17(3), 565.
[27] Chaudhari, G., Deshpande, A., & Liu, K. (2017). Smart Robotic Assistant for Visually Impaired. National Science Foundation, 12(7), 345-352.
[28] Nguyen, H., Nguyen, M., Nguyen, Q., Yang, S., & Le, H. (2020). Web-based object detection and sound feedback system for visually impaired people. International Conference on Multimedia Analysis and Pattern Recognition, 1-6.
[29] Potdar, K., Pai, C. D., & Akolkar, S. (2018). A Convolutional Neural Network based Live Object Recognition System as Blind Aid. Computer Vision and Pattern Recognition, 1(1).
[30] Bhandari, A., Prasad, P., Alsadoon, A., & Maag, A. (2019). Object detection and recognition: using deep learning to assist the visually impaired. Disability and Rehabilitation: Assistive Technology, 16(2), 1-9.
[31] Parikh, N., Shah, I., & Vahora, S. (2018). Android Smartphone Based Visual Object Recognition for Visually Impaired Using Deep Learning. International Conference on Cryptography, Security and Privacy, 420-425.
[32] Shirley, Rane, K., Rao, K. H., B, B. B., Agrawal, P., & Rawat, N. (2023). Machine learning and Sensor-Based Multi-Robot System with Voice Recognition for Assisting the Visually Impaired. Journal of Machine and Computing, 3(3), 206-215.

[33] Najm, H., Elferjani, K., & Alariyibi, A. (2022). Assisting Blind People Using Object Detection with Vocal Feedback. 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), 48-52.
[34] Adeyanju, I. A., Azeez, M. A., Bello, O. O., & Badmus, T. A. (2022). Development of a Convolutional Neural Network-Based Object Recognition System for Uncovered Gutters and Bollards. ABUAD Journal of Engineering Research and Development (AJERD), 5(1), 147-154.
[35] Gautam, S., Sivaraman, K., Muralidharan, H., & Baskar, A. (2015). Vision System with Audio Feedback to Assist Visually Impaired to Grasp Objects. Procedia Computer Science, 58, 387-394.
[36] Breve, F. A., & Fischer, C. (2020). Visually Impaired Aid using Convolutional Neural Networks, Transfer Learning, and Particle Competition and Cooperation. IEEE International Joint Conference on Neural Network, 1-8.
[37] Caballero, A., Catli, K. E., & Babierra, A. G. (2020). Object Recognition and Hearing Assistive Technology Mobile Application using Convolutional Neural Network. 2020 International Conference on Wireless Communication and Sensor Networks, 41-48.
[38] Lailla, B., Yandre, C., & Linnyer, A. (2018). Automata Classification with Convolutional Neural Networks for Use in Assistive Technologies for the Visually Impaired. 11th PErvasive Technologies Related to Assistive Environments Conference, 157-168.
[39] Shaikh, S., Karale, V., & Tawde, G. (2020). Assistive Object Recognition System for Visually Impaired. The International Journal of Engineering Research & Technology (IJERT), 9(9), 736-740.
[40] Jessica, A., Veena, S. H., Srivarshini, S., Krishna, R. G., & Mounica, M. (2022). Convolutional Neural Networks based Voice Assistive System for Blind People. International Conference on Communication and Electronics Systems (ICCES), 7, 1608-1613.
[41] Kinra, A., Walia, W., & Sharanya, S. (2023). A Comprehensive and Systematic Review of Deep Learning Based Object Recognition Techniques for the Visually Impaired. 2023 2nd International Conference on Computational Systems and Communication (ICCSC), 2, 1-6.
[42] Nikhil, A. S., Swarnalatha, & Lav, M. (2019). Guiding and Navigation for the Blind using Deep Convolutional Neural Network Based Predictive Object Tracking. International Journal of Engineering and Advanced Technology, 9(3), 306-313.
[43] Dahiya, D., Gupta, H., & Dutta, M. K. (2020). A Deep Learning based Real Time Assistive Framework for Visually Impaired. 2020 International Conference on Contemporary Computing and Applications (IC3A), 106-109.
[44] Hsieh, Y.-Z., Lin, S.-S., & Xu, F.-X. (2020). Development of a wearable guide device based on convolutional neural network for blind or visually impaired persons. Multimedia Tools and Applications, 29473–29491.
[45] Yi, C., Flores, R. W., Chincha, R., & Tian, Y. (2013). Finding Objects for Assisting Blind People. Netw Model Anal Health Inform Bioinform, 2(2), 71-79.
[46] Neto, L. B., Grijalva, F., Maike, V. R., Martini, L. C., & Florencio, D. (2017). A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users. IEEE, 47(1), 52-64.