International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I4P130

Volume 72 Issue 4, 297-305, April 2024

Original Article

Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision

Myo Min Aung[1], Dechrit Maneetham[2], Padma Nyoman Crisnapati[3], Yamin Thwe[4]

[1,2,3,4] *Department of Mechatronics Engineering, Rajamangala University of Technology Thanyaburi, Thailand.*

[2] *Corresponding Author : dechrit_m@rmutt.ac.th*

Received: 09 October 2023 Revised: 08 February 2024 Accepted: 19 March 2024 Published: 24 April 2024

Abstract *- In the realm of computer vision and autonomous systems, object recognition and obstacle recognition are pivotal tasks, each contributing uniquely to the intelligent capabilities and safety of people and mobile robots. While object recognition focuses on identifying and classifying objects within digital images or video frames, obstacle recognition is dedicated to detecting and localizing obstacles or hazards in an environment. Object recognition, which utilizes machine learning, computer vision, YOLOv4 architecture, and the COCO dataset, is presented with a particular emphasis on visually impaired individuals. This study integrates YOLOv4 and the COCO dataset, aiming to advance object recognition while harnessing the benefits of obstacle recognition. The research encompasses hardware implementation, including a Raspberry Pi with an added 7-inch LCD and software implementation involving machine learning models. Test results reveal the system's robustness and real-time functionality. Furthermore, the user experience testing at the exhibition of Phramongkutklao Hospital garnered positive feedback, which is valuable input to build a user-centric approach in developing object recognition technology tailored to their needs. This research promises valuable contributions to intelligent systems' object recognition in complex environments.*

Keywords *- Computer Vision, Object recognition, Raspberry-pi, Vision impairment, YOLOv4.*

1. Introduction

Vision is our most important sense and is critical to daily activities of life. According to the statistics from WHO [1], at least 2.2 billion people have vision impairment or blindness. Assistive technologies have been developed for people who are blind or visually impaired, as they need mobility and independence, but many of these technologies are expensive and inaccessible. Therefore, there is a need for a cost-effective, lightweight, adaptable, and efficient solution that can operate in real-time with high accuracy [2]. Advancements in technology enable computers to see and analyze surroundings as humans do, utilizing computer vision. Computer vision can provide visual information for visually impaired individuals to understand the world although it cannot fully replace human vision that uses eyes and brain to see. Object recognition and obstacle recognition are critical tasks in computer vision which enable gaining intelligent capabilities and safety for people, autonomous vehicles, and robots/machines in the workplace [3]. Object recognition focuses on identifying and classifying various objects within digital images or video frames. Obstacle recognition serves a distinct purpose: the detection and localization of obstacles or hazards in an environment [4]. Object recognition, an essential component in computer vision, addresses the identification of objects within images, enabling applications like content-based image retrieval, augmented reality, and more [5].

However, its primary goal is to recognize and classify objects for tasks such as scene understanding, and often, it does not inherently consider the spatial relationships or dynamics between these objects [6]. To understand the scene, information of object recognition must be transferred as a voice. Receiving object information by the sense of hearing is fast and safe compared to the sense of touch. For blind individuals, existing object recognition systems are primarily available in English, rendering them less effective when faced with objects or text described in other languages [7]. Therefore, object recognition with the required add-on language is essential, especially in countries where English is rarely spoken [8]. Object recognition considers the context and motion of objects, prioritizing the detection of elements that may impede the intended path of blind people's devices [9]. This crucial distinction between object and obstacle recognition stems from the fact that not all recognized objects are obstacles, and not all obstacles are recognized as distinct objects [10].

An obstacle, in the context of object recognition for blind people, refers to any element that poses a potential danger, irrespective of its object category. For instance, a pedestrian crossing the road may be an object of interest (for a car) in object recognition but becomes an obstacle of utmost importance for blind people to avoid a collision [11].

The advantages of obstacle recognition are multifold. Firstly, it inherently addresses safety concerns by identifying hazards and allowing blind people to plan collision-free trajectories [12]. Secondly, it promotes proactive decision-making in real-time, enabling prompt actions to avoid or mitigate potential collisions.

Traditional human helpers, guide dogs, or commonly used white canes are hardly sufficient as standby companions for blind people. To gain scene understanding and knowledge acquisition, assisted tools and technology play an important role. In this research, object recognition is primarily studied, acknowledging its unique role in the broader landscape of computer vision and autonomous systems. The research objective is to develop a simple, low-cost, and wearable assistive device that enables object recognition for blind or visually impaired people, utilizing a fusion of machine learning, computer vision, YOLOv4 (You Only Look Once version 4) architecture, and the COCO (Common Objects in Context) dataset [14].

According to the review done in [15], YOLOv4 provides a faster and more accurate advanced detector compared to all available alternatives. In favour of object detection performance, YOLOv4-Lite was used in contrast to other works that used YOLOv4 [16, 17]. The YOLOv4 model and COCO dataset were chosen due to their stability, good accuracy and speed in real-time object detection tasks. This fusion brings distinctive advantages over the usual perception of the sense of touch for object recognition. The novelty of this work is conducting three-stage testing which was done to verify the performance and accuracy of the system. The user satisfaction survey was also conducted at a hospital exhibition to receive feedback and extend the system beyond the enhancement of object recognition.

2. Literature Review

Blind and visually impaired individuals need to rely on senses other than sight—such as hearing, touch and smell to gain awareness of their surroundings. The commonly used assistive method involves using their sense of touch for pattern perception. Haptics involves the use of active touch to perceive objects and forms. Mobility is crucial for them to conduct daily activities, and they use conventional mobility aids such as white canes, guide dogs, and assistance from trained guides or volunteers [18]. While a white cane has limited scene understanding and cannot easily identify certain types of objects, such as tables or chairs, without physical touch, the other aids require significant dependency on them. Smart canes have been developed to enhance the functionalities of a typical white cane. In [19], the white cane is modified by incorporating sensors to detect obstacles, steps, and pits in the user's path, and audio cues are added to notify the presence of obstacles. Another modified white cane, as mentioned in [20], utilizes ultrasonic sensors to detect obstacles at the waist (lower) and head (upper) levels.

However, these methods cannot fully replace vision and its ability to identify/recognize objects in the surroundings. Utilizing a camera with embedded computer vision proves to be a reliable solution for describing the objects in the scene with better accuracy in detection. Therefore, most works are developed with the help of a camera to capture an image or video first, followed by preprocessing and object description using machine or deep learning algorithms.

The detected images should be conveyed to visually impaired individuals through text-to-speech techniques [21]. In [22], Mohane and Gode proposed the SIFT (Scale Invariant Features Transform) algorithm for feature extraction. They compared the features extracted from images taken by a camera to those of known database objects. A visual substitution system for blind individuals based on video interpretation is presented in [23] and is based on object recognition and scene type prediction using patch extraction and matching.

Similar to white canes, wearables such as hand-held or head/chest-mounted devices are also useful mobility aids for blind or visually impaired people. In [24], an intelligent and smart helmet for visually impaired individuals is designed, making daily activities easier for them as it can identify text, recognize obstacles, and describe people in the scene. Nowadays, smartphones are widely used, and several mobile applications are designed to assist visually impaired individuals. The Intelligent Eye application presented in [25] provides assistance to visually impaired individuals by offering a set of useful features: light detection, color detection, object recognition, and banknote recognition. YOLOv5, COCO dataset, pyttsx3 and gTTS are the technologies used for image detection to speech generation. Machine learning approaches play a vital role in improving object detection and recognition. Reagan L. Galvez et al. [26] have demonstrated that classification and detection of objects are now accurately possible with the recent advancements in the field of deep neural networks in image processing. While Patrick Poirson et al. [27] focused on Single-Shot Detection (SSD), A. Kumar et al. [28] utilized a single-shot multi-box detector (SSD) algorithm in real time for detecting objects for visually impaired individuals.

IoT-enabled automated common object and currency recognition system [29] extends object recognition to the next level by identifying currency notes in the real-time scenario and sending all data about user mobility in both indoor and outdoor environments to the cloud. Single Shot Detector (SSD) model with MobileNet and Tensorflow-lite were used to perform object recognition and to notify users. According to [30,31], among mainstream object detection models, YOLO has a significant advantage in terms of detection speed. At the same time, SSD is better than YOLO in both accuracy and speed, but they would fail to maintain this accuracy with small objects.

3. Research Methodology

3.1. Obstacle Detection vs Object Recognition

Obstacle detection (Figure 1(a)) and recognition (Figure 1(b)) are integral components in the realm of computer vision and robotics, notably leveraging the advanced technologies mentioned in the introduction, such as machine learning, computer vision algorithms, YOLOv4 architecture, and the COCO dataset. Obstacle detection primarily focuses on identifying the presence and location of objects or hazards within the field of view and benefits from sensor modalities like LiDAR and cameras. In contrast, obstacle recognition, which delves deeper into semantic understanding and categorization, heavily relies on deep learning techniques, including Convolutional Neural Networks (CNNs) and Semantic Segmentation models. Both detection and recognition share the potential for sensor fusion, temporal modeling, and large-scale dataset utilization, presenting solutions to challenges such as occlusions, adverse weather conditions, and ethical considerations in applications like autonomous vehicles and robotics. The choice between detection and recognition is often guided by specific application requirements, incorporating a trade-off between computational complexity and semantic comprehension.

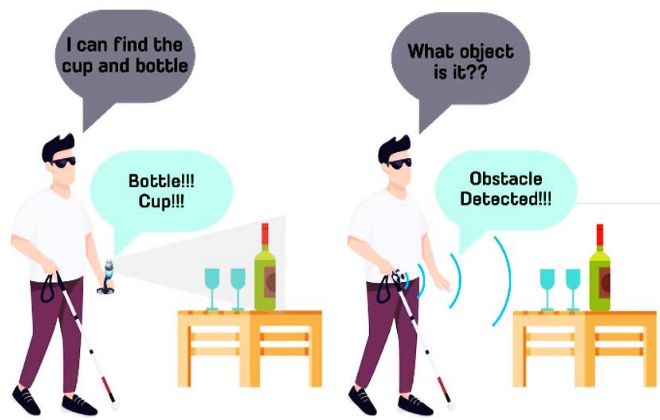
3.2. YOLOv4

In recent years, the landscape of object detection has evolved significantly with the emergence of state-of-the-art architectures like YOLOv4, underpinned by advanced technologies mentioned in the introduction. Using a single neural network, YOLO performs both prediction and categorization for identified objects by segmenting an image into smaller images and splitting them into a m×m square grid [32].

YOLOv4 epitomizes real-time object detection through its one-stage detection pipeline, incorporating components such as CSPDarkNet-53 as the feature extractor, SPP+PANet for context aggregation, and dense prediction for precise object localization and segmentation. CSPDarkNet-53, with its Cross-Stage Partial networks, enhances feature propagation and gradient flow. SPP+PANet efficiently captures multi-scale features and fuses contextual information, while dense prediction techniques empower pixel-wise object localization. The amalgamation of these technologies in YOLOv4 has propelled object detection capabilities to new heights, making it a leading choice for real-time applications across diverse domains, from autonomous vehicles to surveillance systems. The architecture of YOLOv4 consists of a backbone (CSP-DarkNet53), detection neck (SPP Block +PANet), and detection head (Dense Prediction) (Figure. 2).

3.3. System Overview

The object recognition system deployed on a Raspberry Pi 4 8GB, equipped with a USB camera, earphones, a power bank, and a 7-inch LCD display, represents a compact and versatile solution for real-time visual understanding. The Raspberry Pi 4's ample memory capacity and processing power enable efficient and responsive object recognition, while the USB camera (Logitech C310 HD) captures high-quality images or video feeds for analysis. With the integrated earphones and LCD display, the system provides both visual and auditory feedback, making it suitable for a range of applications, including assistive technologies for the visually impaired, interactive displays, or even surveillance systems. The use of a power bank (Remax 20K mAh-22.5W) ensures portability, allowing the system to operate autonomously without the need for a continuous power source, making it ideal for field applications where mobility and versatility are paramount (Figure. 3).



(a) (b)

Fig. 1 Illustration of (a) Obstacle Detection and Object Recognition

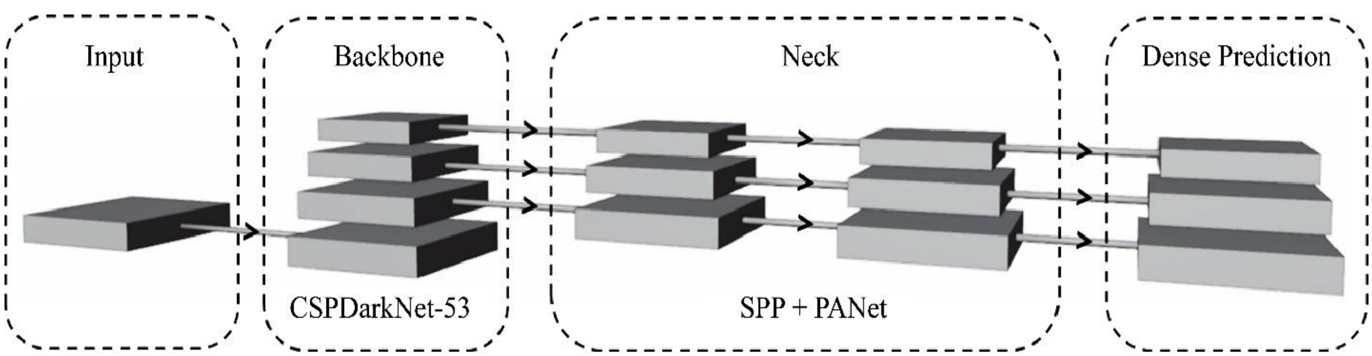


Fig. 2 YOLOv4 Architecture

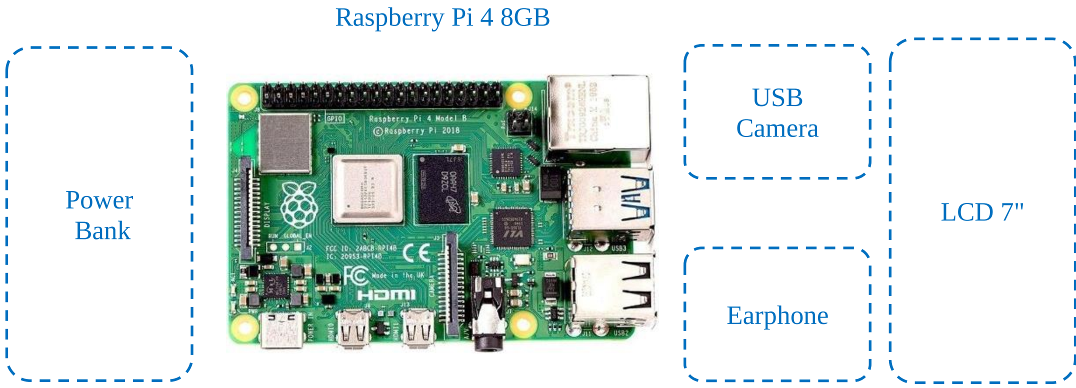


Fig. 3 System overview

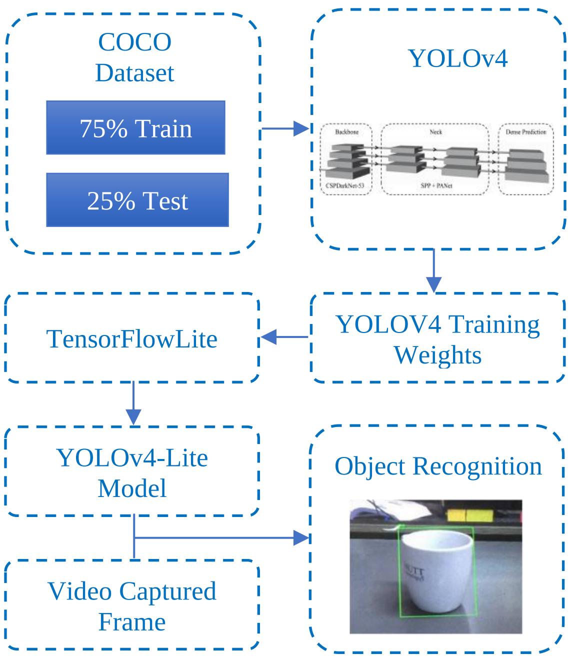


Fig. 4 Object Recognition Program Flow

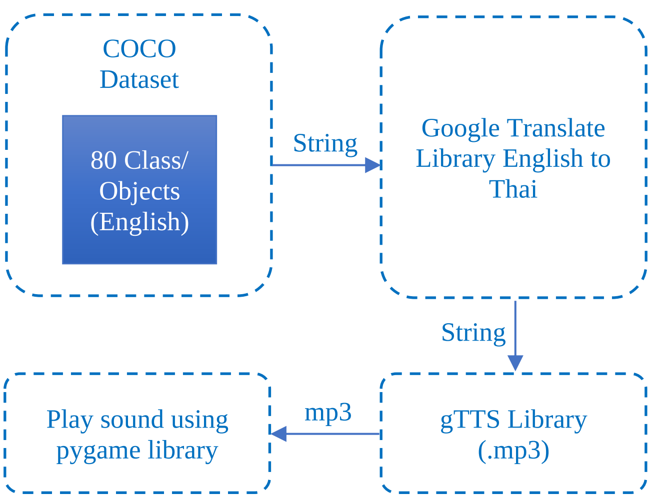


Fig. 5 Text-to-Speech Thai Language Flow

3.4. Machine Learning Model

One important process of machine learning is understanding a dataset thorough a process called training. COCO (Common Objects in Context) is a large-scale object detection, segmentation and captioning dataset widely used for research and practical applications [13]. The COCO dataset contains over 330K images and 80 different object categories with context recognition. In this research, the machine learning YOLOv4 model was trained on the COCO dataset with a 75% training and 25% testing split, intending to achieve a robust foundation for object recognition tasks.

This model has been fine-tuned to identify a wide range of objects with high accuracy and precision. After training, the model's weights are extracted, capturing the learned knowledge and enabling its deployment for real-time object recognition (Figure 4). To optimize real-time performance, the YOLOv4 model is converted into a more efficient YOLOv4-Lite model using TensorFlowLite. This lightweight variant maintains the core object recognition capabilities while reducing computational demands, making it well-suited for deployment on resource-constrained platforms like Raspberry Pi or mobile devices. The model is then integrated with a camera module, capturing video frames in real-time.

These frames are fed into the YOLOv4-Lite model for instant object recognition, enabling the system to identify objects within the camera's field of view and provide valuable insights or take appropriate actions. This application of YOLOv4-Lite as a real-time object recognition tool enhances its versatility, making it valuable for a wide array of applications, including surveillance, autonomous systems, and interactive displays. The text-to-speech process leveraging the COCO Dataset, which encompasses 80 object categories, involves a multistep transformation. Initially, it begins with converting a text string to audio. The string representing an object or class label from the COCO Dataset is translated from English to Thai using the Google Translate library's English-to-Thai function, facilitating cross-language accessibility.

Subsequently, the translated text is then converted into an audio file in the .mp3 format using the gTTS (Google Text-to-Speech) library, encapsulating the object's name in an audible form. Finally, the resulting .mp3 audio file is played using the pygame library, allowing users to audibly perceive the object's identification, which is particularly valuable for individuals with visual impairments or in scenarios where audio feedback enhances user experience and comprehension (Figure 5). This holistic approach bridges language gaps and aids in making object recognition more accessible and informative.

4. Results and Discussion

4.1. Software Implementation

The following pseudocode outlines a real-time object recognition system. It begins by defining a function to play audio files. It then loads a YOLOv4 model for object detection and class labels associated with recognized objects. The code opens a camera to capture video frames and enters a main loop where it periodically checks for elapsed time. When a second has passed, the system performs object detection on the captured frame, drawing bounding boxes and labels for objects with confidence scores above 0.5. Simultaneously, it plays an audio file associated with the recognized object. The frame with visual annotations is displayed, and the system listens for the user to press the "Escape" key to exit gracefully. Finally, it releases camera resources and closes windows when the program ends. This pseudocode provides a high-level representation of a real-time object recognition system and can be implemented in a programming language with appropriate libraries and frameworks.

```
function play_audio(filename):
# Open and play an audio file
audio = open_audio_file(filename)
audio.play()
# Load YOLOv4 model
model = load_yolov4_model()
# Load class labels
labels = load_class_labels()
# Open the camera
camera = open_camera()
current_object = ''
previous_object = ''
previous_time = get_current_time()
while True:
    frame = capture_frame(camera)
    if elapsed_time_since(previous_time) >= 1:
    # Perform object detection
    objects = detect_objects(model, frame)
    for obj in objects:
      if obj.confidence > 0.5:
        draw_bbox(frame, obj)
        draw_label(frame, obj.label)
        current_object = obj.label
        play_audio('/audio/'+ cur_obj+'.mp3')
    show_frame(frame)
    previous_time = get_current_time()
    previous_object = current_object
if user_presses_escape():
        break
release_resources(camera)
close_windows()
```

4.2. Hardware Implementation

In the initial phase of experimentation, as depicted in Figure 6, a compact and minimalistic setup was adopted for the Raspberry Pi 4B, following a headless concept. However, following valuable feedback from users, particularly doctors, we recognized the importance of enhancing the system's usability. Responding to their input, we introduced a significant modification in the second experiment, as illustrated in Figure 7. In this iteration, a 7-inch LCD monitor was added to the hardware setup. This addition served the vital purpose of enabling doctors and nurses to visually monitor and assess the accuracy of object detection in real time. The integration of this monitor not only addressed the users' needs but also substantially improved the system's functionality and extensibility, making it an indispensable tool for medical professionals engaged in object recognition tasks. A meticulous three-stage testing approach was conducted to assess the outcomes of this research comprehensively. In the first phase, the black box methodology was utilized to ensure the system's performance (Table 1). In the second phase, an evaluation was conducted using a random sample of 10 images from the COCO validation dataset, as presented in Table 2. Notably, the analysis reveals that the average detection time per image is approximately 1 second, regardless of the number of bounding objects in each image. Based on the data presented in the table, images with a blurred quality and those depicting only a portion of an object have an accuracy rate of less than 50% when compared to other images. Subsequently, in the third testing phase, an array of object detection scenarios was conducted, as documented in Table 3, to evaluate the system's robustness and functionality further.

Table 1. Black Box Testing

No.	Function	Status
1	Load YOLOv4 Model	Success
2	Load class label	Success
3	Capture image from the camera	Success
4	Object Recognition	Success
5	Play object audio (.mp3)	Success
6	Auto run when Raspberry Pi turned on	Success

Table 2. Testing Results on COCO Validation Dataset

Image	Detected Total Bounding Boxes	Detection Time (s)	Average Accuracy (%)
Image1	10	1.3	68.4
Image2	1	1.3	42
Image3	6	1.3	63.3
Image4	6	1.3	67.1
Image5	4	1.3	51.25
Image6	2	1.3	93
Image7	1	1.3	97
Image8	2	1.2	43
Image9	5	1.3	30.4
Image10	1	1.2	96

Fig. 6 Hardware Implementation without LCD

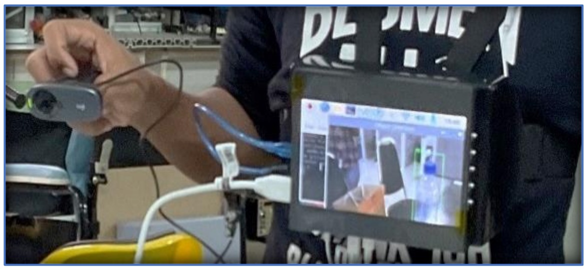
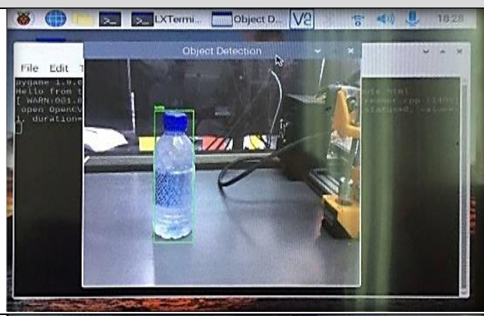
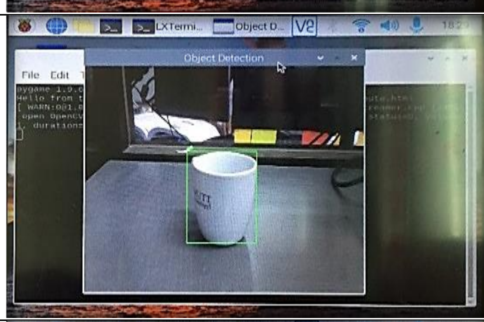
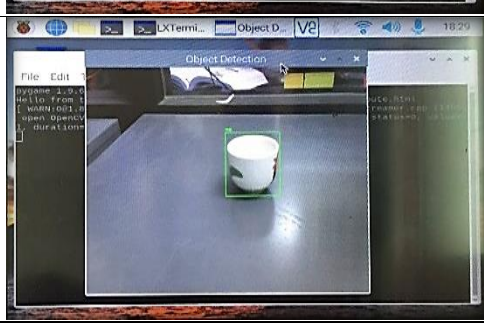
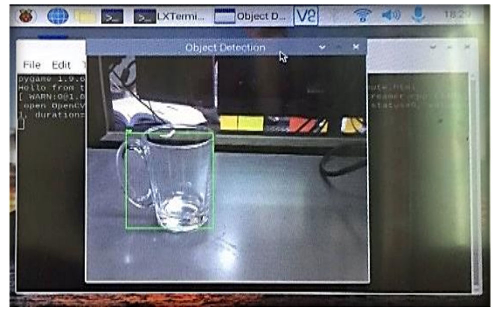
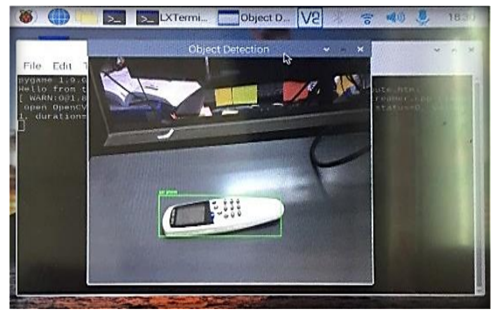
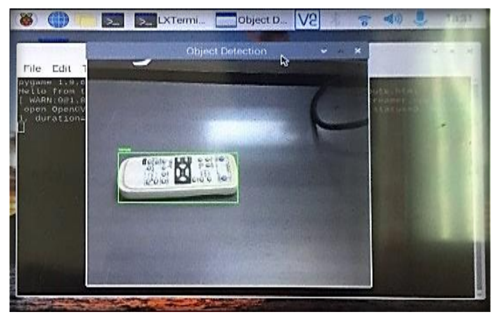
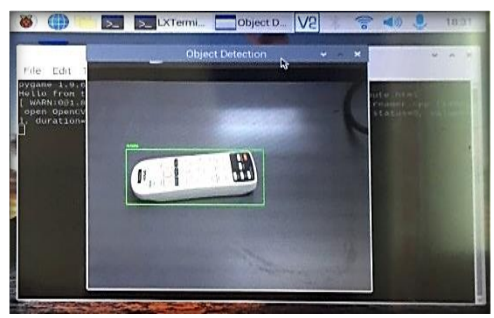
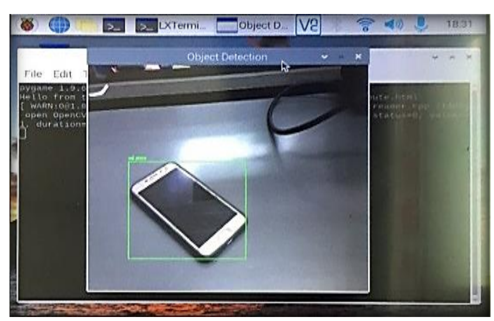


Fig. 7 Hardware Implementation with LCD

Table 3. Object Recognition Testing Results

Scenario	Status	Results
Bottle	Success	
Ceramic Cup	Success	
Small Cup	Success	
Transparent Cup	Success	
Remote AC1	Success	
Remote AC2	Success	
Remote AC3	Success	
Handphone1	Success	

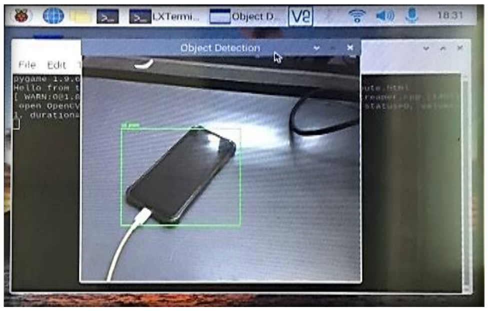
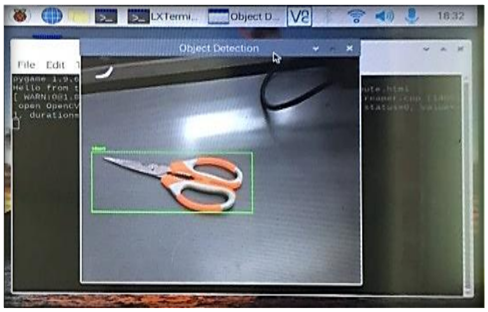
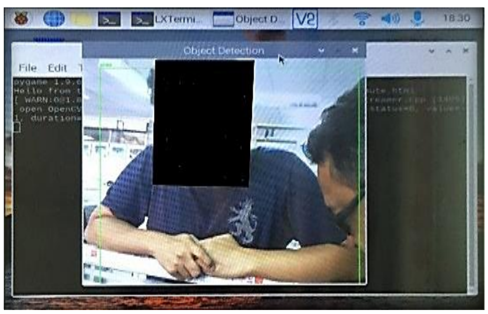
Handphone2	Success	
Scissors	Success	
People	Success	

Fig. 8 Hospital exhibition

Table 4. User satisfaction survey result

Satisfaction factors	Mean Score of Satisfaction
Usability	
Simplicity	5
Portability	3.9
Audibility	4.7
Ergonomics	4.8
Efficiency	
Detection Speed	4.8
Recognition Accuracy	4.8
Overall Satisfaction	**4.7**

The results stemming from these comprehensive examinations were unequivocal and highly promising. Remarkably, the system demonstrated an impeccable success rate, executing all functional programs flawlessly at an impressive 100% efficiency. Additionally, the object detection component exhibited exceptional performance, affirming its capacity to identify objects consistently and accurately across various scenarios. These positive outcomes underscore the system's reliability and its potential for practical application in real-world contexts.

The demonstration of this assisted system took place at Phramongkutklao Hospital in Thailand on July 25, 2023 (Figure. 8). Visitors were invited to test the system and collected their invaluable feedback for the improvement of object recognition technology designed specifically for individuals with visual impairments. This event served as a pivotal platform to gather insights from the users (the doctors, the nurses and other participants) that helped to refine and enhance the system based on their unique needs and preferences. A user satisfaction survey was also conducted about the usability and efficiency of our assistive device by using the 5-point Likert scale analysis ranging from 1-very unsatisfied to 5-very satisfied. The result with a sample size of 15 participants shows that the overall satisfaction of the user is high except for a minor concern over using a 7″ LCD panel that impacts the portability of the device (Table 4). Encouragingly, the exhibition yielded many positive comments from the visitors, further reinforcing the importance and potential impact of research in providing valuable solutions for the visually impaired community. These affirmations from users underscore the significance of user-centric design and its pivotal role in shaping the evolution of assistive technologies like object recognition systems.

5. Conclusion

In conclusion, this research delved into the critical domains of object recognition within the realm of computer vision and autonomous systems. Object recognition is primarily oriented toward identifying objects in digital images and video frames. Leveraging the YOLOv4 architecture and the COCO dataset, this research aimed to advance object recognition for visually impaired people with the potential to improve safety and autonomy in a variety of contexts. This research demonstrates the successful deployment of YOLOv4, trained on the COCO dataset, on the Raspberry Pi 4 platform with a 7″ LCD, obtaining an impressive real-time detection rate of 1 second per frame.

Additionally, the identification of detected objects is audibly communicated in Thai, allowing for immediate recognition by individuals with visual impairments. This innovative method increases safety and facilitates enhanced object identification and localization, thereby assisting the visually impaired community in a significant way. Rigorous testing demonstrated the system's robustness and real-time functionality, achieving a 100% success rate. The limitation of this work is that participants in the satisfaction survey were non-visually impaired people. Therefore, research outcomes might be somewhat different from those in testing with targeted users, visually impaired people who need to interpret the world using other senses.

Fortunately, user-level testing was done by doctors and nurses who have had experience with actual targeted users; therefore, the user's feedback is rather useful for further development. Targeted users will do further testing, and more objects will be trained to strengthen object recognition and extend to obstacle recognition in support of mobility/navigation in an indoor environment. The research outcome could be further extended to use in the movement and navigation of mobile robots. To some extent, this research makes a significant contribution to the evolution of intelligent systems, object recognition and navigating in complex and dynamic environments, with a focus on safety and accessibility.

Acknowledgments

The authors extend heartfelt gratitude to all individuals and organizations who contributed to the success of this research. Especially, doctors and nurses who participated in the user experience testing at the exhibition hall of Phramongkutklao Hospital are really appreciated for providing essential feedback that contribute to the development of the object recognition system. Rajamangala University of Technology Thanyaburi and Valaya Alongkorn Rajabhat University provided the resources and facilities needed for this research. Authors really appreciate for their help and support.

References

[1] World Health Organization, Blindness and Vision Impariment, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] Kanak Manjari, Madhushi Verma, and Gaurav Singal, “A Survey on Assistive Technology for Visually Impaired,” *Internet of Things*, vol. 11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Rabia Jafri et al., “Computer Vision-Based Object Recognition for the Visually Impaired in an Indoors Environment: A Survey,” *The Visual Computer*, vol. 30, pp. 1197-1222, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[4] Idris Jeelani et al., “Real-Time Vision-Based Worker Localization & Hazard Detection for Construction,” *Automation in Construction*, vol. 121, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Afshan Latif et al., “Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review,” *Mathematical Problems in Engineering*, vol. 2019, pp. 1-21, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Shervin Minaee et al., “Image Segmentation Using Deep Learning: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523-3542, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Vidula V. Meshram et al., “An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 5, pp. 449-460, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Wilson Luu et al., “A Holistic Model of Low Vision Care for Improving Vision-Related Quality of Life,” *Clinical and Experimental Optometry*, vol. 103, no. 6, pp. 733-741, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mohamed Dhiaeddine Messaoudi, Bob-Antoine J. Menelas, and Hamid Mcheick, “Review of Navigation Assistive Tools and Technologies for the Visually Impaired,” *Sensors*, vol. 22, no. 20, pp. 1-29, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Usman Masud et al., “Smart Assistive System for Visually Impaired People Obstruction Avoidance through Object Detection and Classification,” *IEEE Access*, vol. 10, pp. 13428-13441, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Abhishek Gupta, “Deep Learning for Object Detection and Scene Perception in Self-Driving Cars: Survey, Challenges, and Open Issues,” *Array*, vol. 10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Shuchang Xu et al., “Virtual Paving: Rendering a Smooth Path for People with Visual Impairment through Vibrotactile and Audio Feedback,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1-25, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Silvia Rostianingsih, Alexander Setiawan, and Christopher Imantaka Halim, “COCO (Creating Common Object in Context) Dataset for Chemistry Apparatus,” *Procedia Computer Science*, vol. 171, pp. 2445-2452, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yamin Cheng et al., “Rethinking Vision Transformer through Human–Object Interaction Detection,” *Engineering Applications of*

Artificial Intelligence, vol. 122, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Peiyuan Jiang et al., "A Review of Yolo Algorithm Developments," *Procedia Computer Science*, vol. 199, pp. 1066-1073, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Mang Ning et al., "YOLOv4-object: An Efficient Model and Method for Object Discovery," *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, Madrid, Spain, pp. 31-36, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Joo Woo et al., "A Study on Object Detection Performance of YOLOv4 for Autonomous Driving of Tram," *Sensors*, vol. 22, no. 22, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes, "Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review," *IETE Technical Review*, vol. 39, no. 1, pp. 3-18, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Sharada Murali et al., "Smart Walking Cane for the Visually Challenged," *2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Agra, India, pp. 1-4, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Pisak Chinchai et al., "A White Cane Modified with Ultrasonic Detectors for People with Visual Impairment," *Journal of Associated Medical Sciences*, vol. 55, no. 3, pp. 11–18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Koppala Guravaiah et al., "Third Eye: Object Recognition and Speech Generation for Visually Impaired," *Procedia Computer Science*, vol. 218, pp. 1144-1155, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Vikky Mohane, and Chetan Gode, "Object Recognition for Blind People Using Portable Camera," *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, Coimbatore, India, pp. 1-4, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[23] Hanen Jabnoun, Faouzi Benzarti, and Hamid Amiri, "Visual Scene Prediction for Blind People Based on Object Recognition," *2017 14th International Conference on Computer Graphics, Imaging and Visualization*, Marrakesh, Morocco, pp. 21-26, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[24] P. Jagadesh et al., "Design and Development of an Intelligent and Smart Helmet for Visually Impaired," *2023 9th International Conference on Advanced Computing and Communication Systems*, Coimbatore, India, pp. 249-253, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Milios Awad et al., "Intelligent Eye: A Mobile Application for Assisting Blind People," *2018 IEEE Middle East and North Africa Communications Conference*, Jounieh, Lebanon, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Reagan L. Galvez et al., "Object Detection Using Convolutional Neural Networks," *TENCON 2018 - 2018 IEEE Region 10 Conference*, Jeju, Korea (South), pp. 2023-2027, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Patrick Poirson et al., "Fast Single Shot Detection and Pose Estimation," *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, pp. 676-684, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ashwani Kumar, S.S. Sai Satyanarayana Reddy, and Vivek Kulkarni, "An Object Detection Technique for Blind People in Real-Time Using Deep Neural Network," *2019 Fifth International Conference on Image Information Processing*, Shimla, India, pp. 292-297, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Md. Atikur Rahman, and Muhammad Sheikh Sadi, "IoT Enabled Automated Object Recognition for the Visually Impaired," *Computer Methods and Programs in Biomedicine Update*, vol. 1, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Meeth Shetty, "A Review on Deep Learning Object Detection: YOLO vs SSD," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 5, no. 2, 2021. [Publisher Link]

[31] Lu Tan et al., "Comparison of RetinaNet, SSD, and YOLO v3 for Real-Time Pill Identification," *BMC Medical Informatics and Decision Making*, vol. 21, no. 324, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] N. Aishwarya et al., "Skin Cancer Diagnosis with Yolo Deep Neural Network," *Procedia Computer Science*, vol. 220, pp. 651-658, 2023. [CrossRef] [Google Scholar] [Publisher Link]