EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

A Mobile Lens: Voice-Assisted Smartphone Solutions for the Sightless to Assist Indoor Object Identification

Talal Saleem[1,*] and Dr.V. Sivakumar[2]

[1]1st Faculty of Computing and Engineering Technology, Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur, Malaysia
[2]2nd Faculty of Computing and Engineering Technology, Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur, Malaysia

Abstract

Every aspect of life is organized around sight. For visually impaired individuals, accidents often occur while walking due to collisions with people or walls. To navigate and perform daily tasks, visually impaired people typically rely on white cane sticks, assistive trained guide dogs, or volunteer individuals. However, guide dogs are expensive, making them unaffordable for many, especially since 90% of fully blind individuals live in low-income countries. Vision is crucial for participating in school, reading, walking, and working. Without it, people struggle with independent mobility and quality of life. While numerous applications are developed for the general public, there is a significant gap in mobile on-device intelligent assistance for visually challenged people. Our custom mobile deep learning model shows object classification accuracy of 99.63%. This study explores voice-assisted smartphone solutions as a cost-effective and efficient approach to enhance the independent mobility, navigation, and overall quality of life for visually impaired or blind individuals.

Keywords: Artificial intelligence, Deep learning, Indoor object identification, Mobile applications for the blind, Visual impairment.

Received on 11 02 2024, accepted on 30 05 2024, published on 28 06 2024

doi: 10.4108/eetiot.6450

1. Introduction

According to a report by the World Health Organization (WHO), approximately 2.2 billion people worldwide have visual impairment or blindness [1]. Over the past 30 years, there has been a 91% increase in severe visual impairment globally [2]. As the population of visually impaired or fully blind individuals continues to rise, the need for assistive aids becomes more critical. [3]. Visually impaired or blind individuals often use sticks to produce audio cues for obstacle detection while walking and rely on touch to recognize objects of interest. [4]. The object of interest is recognized through the sense of touch, by feeling the shape of the surface of the object [5]. Object detection, which involves predicting objects in a given image or video and determining their locations using boundary boxes, is a fundamental problem in computer vision [6]. However, deploying trained deep neural network models on resource-constrained mobile devices is challenging due to their large size and the limited computing power of mobile devices. [7]. Designing a deep convolutional neural network for mobile is particularly difficult due to constraints on battery, memory, and computational power (refer to Fig 1). It is not easy to design deep neural network architecture lightweight for mobile without significantly reducing object detection accuracy gets reduced drastically, and so finding solutions is challenging and complex [8].

*Corresponding author. Email: TP053459@mail.apu.edu.my

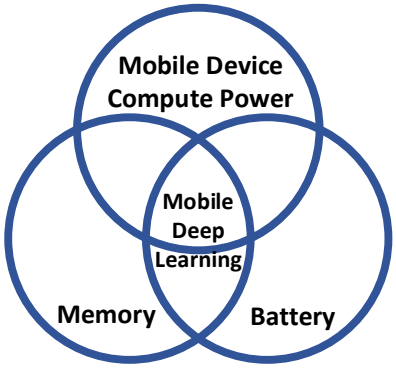

Figure 1. Mobile deep learning challenges to run locally on a mobile device, compute power, memory, and battery

Running deep learning models locally on mobile devices for real-time object detection using a built-in camera presents significant challenges in terms of latency and accuracy (Fig. 2). Addressing the issue of performing real-time object detection with efficient accuracy on mobile devices remains a key concern. Despite these challenges, smartphones are now ubiquitous due to their portability, lightweight design, and ease of use.

Figure 2. Challenges of Mobile deep learning-based object detection in real-time: accuracy and latency

There exists a research gap that necessitates further investigation into developing assistance for visually impaired or blind individuals. This includes enhancing camera image capture processing capabilities without reliance on cloud inference. Additionally, there is a need for systems that operate locally on mobile devices, independent of internet connectivity, to enable voice commands and provide real-time spoken descriptions of recognized objects. These advancements aim to significantly enhance indoor navigation assistance for the visually impaired.

2. Related Work

Smartphone-based assistants to help visually challenged people, several efforts have been proposed, Shimakawa et al. [9] proposed mobile-based object detection using a camera to help visually impaired people identify objects. The author's work does not provide object recognition audio guidance.

Yu et al. [10] proposed a mobile-based traffic light recognition assistant for the visually impaired, employing the MobileNetV3 model as the backbone re-trained with a specific traffic lights dataset. The author's work reported a lower accuracy rate, emphasizing the potential risk of inaccurate predictions leading to accidents.

Jakhete et al. [11] propose an Android application that assists visually impaired people in recognizing objects using TensorFlow's Object detection API, Multibox, and YOLO frameworks via a mobile camera. The study lacked comprehensive evaluation, however, and did not provide evidence of detection accuracy.

Vaidya et al. [12] proposed a prototype based on the YOLO algorithm to detect objects and give feedback to visually impaired individuals both with and without internet connectivity on mobile devices.

Ramalingam et al. [13] proposed a mobile solution for object detection to assist the visually impaired leveraging TensorFlow's built-in object detection API. The study highlighted the need for research work to incorporate mobile use, highlighting the need for future improvements such as incorporating object distance, location, and interactive feedback. These identified research gaps form the basis of the current proposed research study.

3. Methodology

We have adopted this methodology to improve overall safely and independent mobility to assists visually impaired individuals to identify an object of interest and provide safe navigation with voice feedback, providing guidance on object distance and obstacles.

3.1. Proposed Work

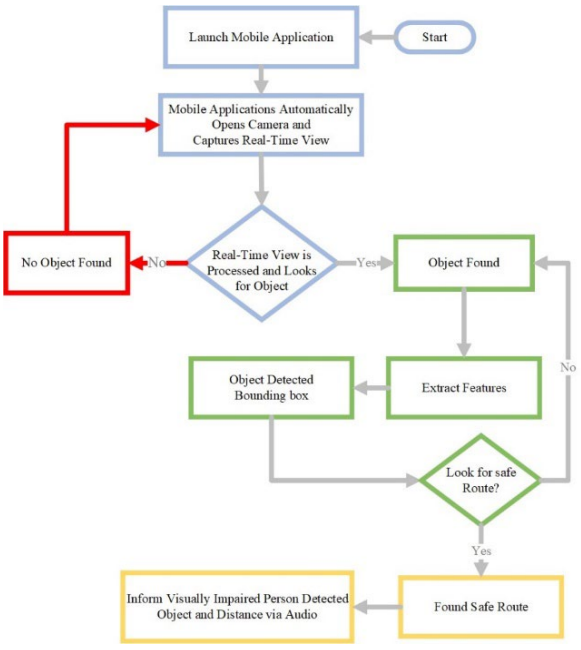

Figure 3. Proposed framework flow chart

The proposed framework uses a state-of-the-art smartphone-based on-device deep learning architecture to assist visually impaired or blind individuals with in-door navigation. Users can wear the smartphone using a lanyard or hold it while positioning the camera forward during in-door navigation, enabling real-time recognition of surrounding situations and providing object notification with efficient performance and accuracy. Our framework uses a mobile camera to detect objects in real-time, utilizing deep learning techniques to facilitate safe navigation and recognition of objects of interest. The speech module enables voice commands to classify and detect objects and provide real-time auditory commentary. Figure 3 illustrates our proposed framework, designed to aid the visually impaired by identifying everyday indoor objects or objects that generate obstacles in their indoor path.

4. Implementation

In this section, we outline the practical steps taken to implement the framework outlined in the methodology section.

4.1. Dataset

The proposed mobile deep learning framework for visually impaired people consists of two datasets.

4.1.1 Largescale Images COCO Dataset

Figure 4. Example images of COCO Dataset

The Common Objects in Context (COCO) dataset contains over 124,000 labelled images of large-scale common objects. It includes 80 different classes with 83,000 images used for training and 41,000 used for testing. Fig. 4 shows example COCO dataset images.

4.1.2 Custom Dataset

For our custom dataset preparation, we extracted frames from a short video of objects. Fig. 5 illustrates example images from our custom dataset, focusing on three indoor objects crucial for the visually impaired: walking stick, door, and wardrobe. The images were sourced from the internet using non-copyrighted material. It is also worth noting that most internet images are unsuitable for effective deep learning training.

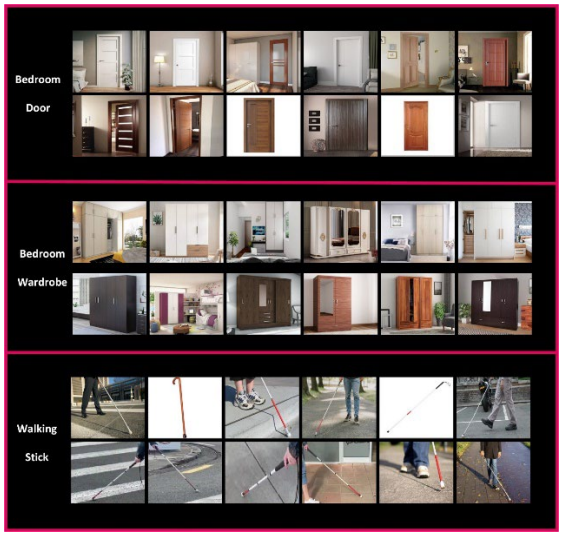

Figure 5. Custom dataset example images

To improve accuracy, a method was used of cropping the image and enlarging and focusing upon image object to increase system object detection robustness. Fig. 6 shows a custom dataset image trimming technique. Custom dataset images were labelled using the annotation tool LabelImg and saved as PASCAL VOC XML, the format required by SSD MobileNet V2.

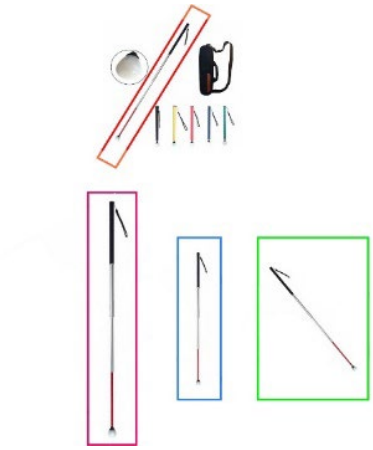

Figure 6. Custom dataset image trimming technique

4.1.2.1 Training and Testing Split

The custom dataset was split randomly, 80% for training the deep learning model and 20% for testing purposes. To facilitate training and testing of the custom deep learning object detection model, the custom dataset was converted into binary data storage in TensorFlow record format, namely Train.Record and Test.Record. Figure 7 depicts the TF-Record generation process.

Figure 7. PBTEXT and TF-Record Generation

4.2. Custom Model Training

To train our custom deep learning object detection model, we utilized the Colab cloud host. We retrained the SSD MobileNetV2 model usingthe TensorFlow object detection API. Fig 8 shows the custom model relearning process.

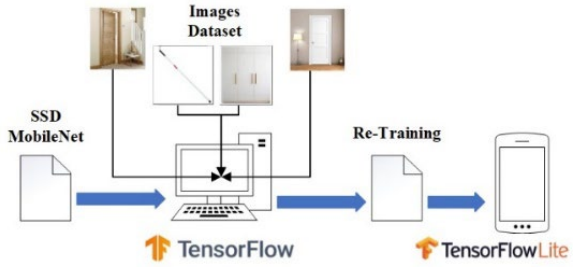

Figure 8. Custom object detection model re-learning using custom dataset.

4.3. Voice Commentary

Our framework integrated Android text-to-speech API with a deep neural network model to enable voice command to describe detected classified object recognition as real-time speech commentary.

4.4. Classify the objects and inform the visually impaired persons

For safe mobility, this module guides safe navigation using voice-guided feedback.

4.5. Mobile Application

Mobile application real-time frames are captured at the width of 640px and height of 480px. Mobile application real-time input generates neural network output. A mobile camera scans surroundings in real-time and sends video frames as input to a deep learning-based algorithm running locally on a mobile device. Fig. 9 shows the mobile framework process for the visually impaired using indoor object identification with voice.

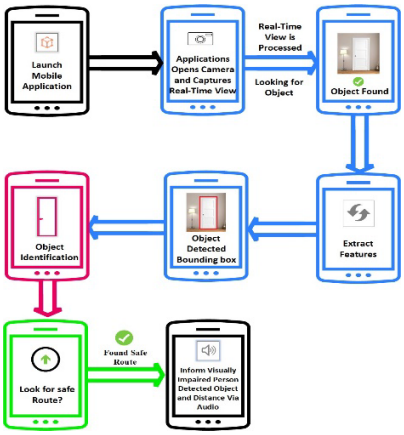

Figure 9. Mobile-based custom object detection framework for the visually impaired

5. Results & Discussion

We conducted experiments with a mobile deep learning-based object detection framework deployed locally on a Samsung Galaxy A12 mobile device running the Android 12 operating system, equipped with a 48-megapixel rear camera. The framework detects objects in real-time, classifies them, and provides voice feedback to assist visually challenged individuals.

5.1. Object Detection (Model 1)

The ssd_mobilenet_v2_fpnlite_320x320_coco17_tpu8 model, pre-trained on the COCO dataset, achieved the following accuracies in object detection and classification: 81% for laptop, 78% for bottle, 78% for cup, and 77% for remote (Fig. 10)

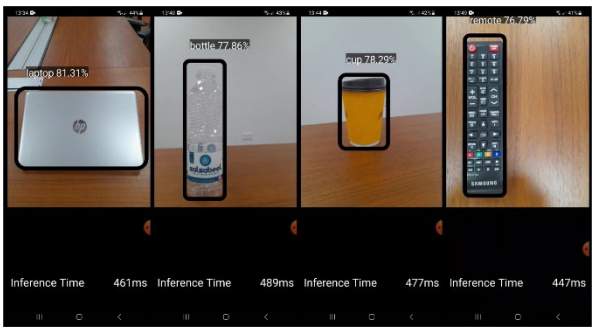

Figure 10. Mobile Small Object Detection

The accuracy evaluation results of the experiment for a mobile deep learning-based Android application designed for detecting small indoor objects for visually impaired people are shown in Table 1. The laptop class was trained with 3,707

images, the bottle class with 8,880 images, the cup class with 9,579 images, and the remote class with 3,221 training images. The mobile application uses a CNN with real-time video frame input size set to 300x300.

Table 1. Summarized Results Mobile Deep Learning-based Model-1 Android Application Indoor Small Object Detection.

	Mobile On-Device Local Model-1 Object Detection		
	Object	Inference Time	Accuracy (%)
1	Laptop	461ms	81%
2	Bottle	489ms	78%
3	Cup	477ms	78%
4	Remote	477ms	77%

The experiment results of the mobile deep learning-based Android application for the visually impaired, running the model locally on-device, achieved object detection and identification accuracies of 81% for couch, 82% for chair, 80% for bed, and 82% for refrigerator (Fig. 11).

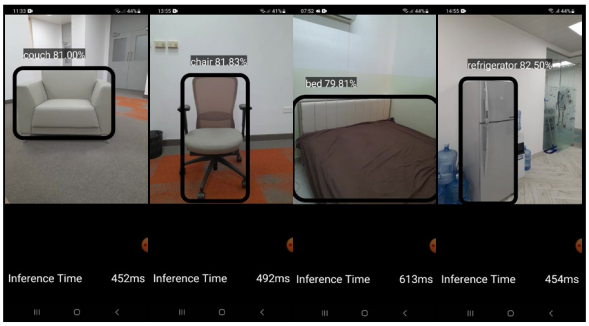

Figure 11. Mobile Large Object Detection

The experiment results of the mobile deep learning-based Android application for visually impaired people, focusing on large indoor object detection, are shown in Table 2. The couch class was trained with 4,618 images, chair class with 13,354 images, bed class with 3,831 images, and refrigerator class with 2,461 training images. The application processed real-time video frames with an input size of 300x300.

Table 2. Summarized Results Mobile Deep Learning-based Model-1 Android Application Large Small Object Detection.

	Mobile On-Device Local Model-1 Object Detection		
	Object	Inference Time	Accuracy (%)
1	Couch	452ms	81 %
2	Chair	492ms	82 %
3	Bed	613ms	80 %
4	Refrigerator	454ms	82 %

5.2. Custom Object Detection (Model 2)

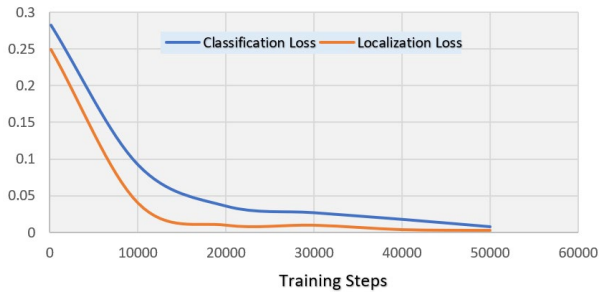

Figure 12. Custom model training Classification Loss and Localization Loss at 50,000 steps.

To optimize the model for better classification accuracy, the custom model underwent multiple retraining sessions. The model was trained for 50,000 steps until achieving a classification loss of 0.001. Figure 12 shows the training loss progression. Table 1 presents the total loss per training step for the custom model training.

Table 3. Custom model Training Loss.

	Custom model Training Loss					
	100 Steps	10,000 Steps	20,000 Steps	30,000 Steps	40000 Steps	50,000 Steps
Classification Loss	0.28	0.09	0.04	0.03	0.02	0.00
Localization Loss	0.25	0.04	0.01	0.01	0.00	0.00
Regularization Loss	0.15	0.11	0.09	0.07	0.06	0.06
Total Loss	0.69	0.24	0.13	0.11	0.08	0.07

The custom dataset-based mobile deep learning model Android application successfully detected and classified the following objects with high accuracy: walking stick with 99.50%, door with 99.76%, and wardrobe with 99.62% (Fig. 13).

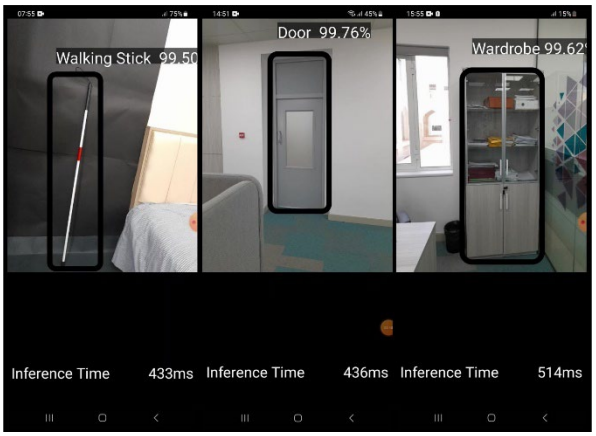

Figure 13. Custom Model Object Detection

The accuracy evaluation results of a mobile deep learning-based Android application for detecting small indoor objects among visually impaired individuals are presented. The custom dataset includes a walking stick class with 450 labeled images, comprising 360 for training and 90 for testing. Similarly, the door class was trained with 450 labeled images (360 for training, 90 for testing), and the wardrobe class with 450 labeled images (360 for training, 90 for testing). Real-time video input frames from the mobile camera were sized at 300x300 pixels. Table 4 displays the accuracy results of custom model-2 for object detection.

Table 4. Custom model object detection accuracy.

	Mobile On-Device (Local) Model-1 Object Detection		
	Object	Inference Time	Accuracy (%)
1	Walking Stick	433ms	99.50%
2	Door	436ms	99.76%
3	Wardrobe	514ms	99.62%

Table 5 presents the results of object detection at distances of 30cm, 50cm, 80cm, 100cm, 120cm, 150cm, and 200cm. The results indicate that Model-1 successfully detected 14 out of 28 objects (laptop, bottle, cup, and remote) at different distances. For Model-1 objects (couch, chair, bed, and refrigerator), 24 out of 28 objects were detected at various distances. Additionally, our proposed custom mobile deep learning model detected 20 out of 21 objects across different distance ranges.

Table 5. Comparison of Model-1 and our custom Model-2 object detection at different distances.

	Model-1 Objects Detection (Laptop, Bottle, Cup, Remote)		Model-1 Objects Detection (Couch, Chair, Bed, Refrigerator)		Our Custom Model-2 Objects Detection (Walking Stick, Door, Wardrobe)	
	Actual Objects	Detected Objects	Actual Objects	Detected Objects	Actual Objects	Detected Objects
30cm	4	3	4	1	3	2
50cm	4	4	4	3	3	3
80cm	4	3	4	4	3	3
100cm	4	2	4	4	3	3
120cm	4	2	4	4	3	3
150cm	4	0	4	4	3	3
200cm	4	0	4	4	3	3
TOTAL	28	14	28	24	21	20

Table 6 provides a comparison of Model-1 and Model-2 with our custom mobile on-device object classification accuracy. The evaluation results demonstrate that Model-1 achieved 78.5% accuracy for small objects, with specific accuracies of 81% for laptop, 78% for bottle, 78% for cup, and 77% for remote. For large objects, Model-1 achieved an accuracy of 81.25%, with accuracies of 81% for couch, 82% for chair, 80% for bed, and 82% for refrigerator.

In contrast, Model-2, our proposed custom model, achieved significantly higher accuracy for custom objects, with specific accuracies of 99.50% for walking stick, 99.76% for door, and 99.62% for wardrobe, resulting in a total accuracy of 99.63%. These results highlight that our custom dataset-trained custom mobile deep learning Model-2 excels in real-time mobile object identification.

Table 6. Comparison of (Model-1) and (Model-2) Our Custom Model Mobile On-Device Object Classification Accuracy.

Comparison of (Model-1) and (Model-2) Our Custom Model					
Model-1		Model-1		Model-2 (Our Custom Model)	
Object	Accuracy	Object	Accuracy	Object	Accuracy
Laptop	81%	Couch	81%	Walking Stick	99.50%
Bottle	78%	Chair	82%	Door	99.76%
Cup	78%	Bed	80%	Wardrobe	99.62%
Remote	77%	Refrigerator	82%	--	--
Total	78.5%	Total	81.25%	Total	99.63%

Table 7 shows the comparison of the existing mobile deep learning-based object detection navigations systems for visually impaired or blind people. Our proposed mobile deep learning model achieved an accuracy of 99.63%.

Table 7. Comparison of the existing Mobile Deep Learning-based object detection navigations systems for visually impaired or blind people

Researcher	Existing Mobile Deep Learning-based object detection navigations systems for visually impaired or blind people			
	Mobile Deep Learning Model	Object Detection	Voice Feedback	Accuracy
Nguyen et al. [14]	SSD MobileNet-V2	✓	✓	60%
Ramalingam et al. [13]	Custom Mobile DL Model	✓	✓	77%
Shimakawa et al. [9]	Custom Mobile DL Model	✓	✗	80%
Vaidya et al. [12]	YOLOv3-tiny	✓	✗	85.5%
Yu et al. [10]	Custom Mobile DL Model	✓	✗	96%
Our Proposed Framework	Our Custom Mobile DL Model	✓	✓	99.63%

6. Conclusion

Enhancing the recognition capabilities of visually impaired individuals and facilitating their interaction with household items such as doors or wardrobes can significantly improve their independence and ease their indoor navigation. In this research study, we propose a deep learning framework that runs locally on a mobile device to identify and recognize objects in real time using a built-in camera, providing voice commentary to assist visually impaired individuals in safe indoor navigation.

To evaluate the performance of our custom-trained model in object detection accuracy, we conducted experiments and achieved a remarkable accuracy of 99.63%. Our results demonstrate that our custom dataset-trained mobile deep learning model excelled in detecting objects such as walking sticks, wardrobes, and doors. This underscores the effectiveness of our approach in detecting indoor objects for the benefit of visually impaired individuals.

References

[1] G. Qiao, H. Song, B. Prideaux, and S. (Sam) Huang, "The 'unseen' tourism: Travel experience of people with visual impairment," *Ann. Tour. Res.*, vol. 99, p. 103542, 2023.

[2] R. Bourne *et al.*, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study.," *Lancet Glob. Heal. - Elsevier*, no. December, 2020.

[3] M. Biswas *et al.*, "Prototype Development of an Assistive Smart-Stick for the Visually Challenged Persons," *Proc. 2nd Int. Conf. Innov. Pract. Technol. Manag. ICIPTM 2022*, vol. 2, pp. 477–482, 2022.

[4] Y. Tange, T. Konishi, and H. Katayama, "Development of vertical obstacle detection system for visually impaired individuals," *ACM Int. Conf. Proceeding Ser.*, no. 1, 2019.

[5] L. He, R. Wang, and X. Xu, "PneuFetch: Supporting Blind and Visually Impaired People to Fetch Nearby Objects via Light Haptic Cues," *Ext. Abstr. 2020 CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–9, 2020.

[6] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[7] Z. Huang *et al.*, "Making accurate object detection at the edge: review and new approach," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2245–2274, 2022.

[8] B. Chen *et al.*, "MnasFPN: Learning latency-aware pyramid architecture for object detection on mobile devices," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 13604–13613, 2020.

[9] M. Shimakawa, K. Matsushita, I. Taguchi, C. Okuma, and K. Kiyota, "Smartphone apps of obstacle detection for visually impaired and its evaluation," *ACM Int. Conf. Proceeding Ser.*, pp. 143–148, 2019.

[10] S. Yu, H. Lee, and J. Kim, "Street crossing aid using light-weight CNNs for the visually impaired," *Proc. IEEE/CVF Int. Conf. Comput. Vis. Work.*, 2019.

[11] S. A. Jakhete, P. Bagmar, A. Dorle, A. Rajurkar, and P. Pimplikar, "Object Recognition App for Visually Impaired," *2019 IEEE Pune Sect. Int. Conf. PuneCon 2019*, pp. 18–21, 2019.

[12] S. Vaidya, N. Shah, N. Shah, and R. Shankarmani, "Real-Time Object Detection for Visually Challenged People," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 311–316, 2020.

[13] D. Ramalingam, S. Tiwari, and H. Seth, "Vision connect: A smartphone based object detection for visually impaired people," *Int. Conf. Comput. Vis. Bio Inspired Comput. Springer, Cham*, 2020.

[14] H. Nguyen, M. Nguyen, Q. Nguyen, S. Yang, and H. Le, "Web-based object detection and sound feedback system for visually impaired people," *2020 Int. Conf. Multimed. Anal. Pattern Recognition, MAPR 2020*, 2020.