A Survey of Object Classification and Detection Techniques in Assistance Systems for the Visually Impaired

Mrunmayi Deshpande[1], Sreenath M V[2], Shreya Nadella[3], Nirav Antony[4] and Nivedita Venkat[5]
[1]Final Year Student, Department of CSE, PES University, Bangalore, INDIA
[2]Assistant Professor, Department of CSE, PES University, Bangalore, INDIA
[3]Final Year Student, Department of CSE, PES University, Bangalore, INDIA
[4]Final Year Student, Department of CSE, PES University, Bangalore, INDIA
[5]Final Year Student, Department of CSE, PES University, Bangalore, INDIA

[1]Corresponding Author: mrunmayi.deshpande@gmail.com

Received: 09-01-2024 **Revised:** 24-01-2024 **Accepted:** 10-02-2024

ABSTRACT

The number of visually impaired individuals in the world is estimated to be 1 billion, as per WHO reports. Through a thorough examination of existing assistive technology and research, this paper provides a survey of object classification and detection techniques that are used in assistive technology by visually impaired individuals. We discuss the methodology's drawbacks, s and functionalities of these techniques, and observe how sufficient they are in meeting the needs and requirements of the targeted users, and how they can be improved. As a result of this study, we identify areas with room for improvement in object detection in the assistive technology domain.

Keywords-- Object Detection, Object Identification, Real-Time Systems, Visually Impaired

I. INTRODUCTION

Sight and visual perception of our environment plays a fundamental role in our daily lives. This not only enables us to navigate and comprehend what's around us but also allows us to identify objects in our surroundings. However, visually impaired individuals face significant challenges in their dayto-day life due to their inability of accessing and interpreting visual information.

Object detection and classification are fundamental components of visual assistance systems, allowing the visually impaired users to identify and locate objects in their environment.

Object detection, a task that sighted individuals effortlessly excel at but is proven to be extremely tedious for computers while humans locate and identify objects within their peripheral vision with high accuracy, poses a significant challenge for computer systems [1] Despite notable advancements in the field of computer vision, accurately identifying objects remains a non-trivial problem that demands extensive research and development [2].

Traditionally, computer vision algorithms relied on handcrafted features which often fell short in achieving results that are accurate and reliable however with the recent introduction of Convolutional Neural Networks (CNNs) a wave of research and exploration into leveraging deep learning for various computer vision applications was sparked. [3].This breakthrough showcased the power and potential of CNNs in achieving significant improvements in image classification accuracy, [4].

Despite these advancements, object detection remains an ongoing research challenge. Researchers are actively exploring new architectures, loss functions, and training strategies to tackle these challenges and improve the accuracy and efficiency of object detection systems.

In this survey paper, we have researched and compiled information about various object detection and classification techniques used in assistance systems for the visually impaired. We have collected and gone through a large number of papers that have been published in order to gain a deeper understanding of the problem at hand where there is room for improvement in these implemented techniques.

II. KEY CHALLENGES IN OBJECT DETECTION

Through the process of extensively examining published papers and prior research, we successfully pinpointed the key challenges faced in object detection within assistance systems designed for the visually impaired.

A. Dataset Size and Range of Objects

In most observed scenarios, the dataset that was used was limited in size and did not encapsulate all possible object. This in turn does not portray all of the possible scenarios and objects that the visually impaired

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)
https://ijemr.vandanapublications.com
https://doi.org/10.5281/zenodo.10648650

encounter in their daily lives. A limited dataset significantly increases the chances of biases and can potentially ruin the accuracy of the system's prediction. The lack of diverse examples may lead to skewed outcomes and inaccurate interpretations, which may potentially reinforce stereotypes and misconceptions. This limitation emphasizes the importance of incorporating a more comprehensive and diverse range of examples to enhance the effectiveness and fairness of the object detection system [5]

B. Absence of User feedback

A lot of systems lack comprehensive information regarding user feedback on the practical effectiveness of the system .User feedback is essential to ensure that the system is performing the right task as well as identify potential usability issues and areas for improvement in the system's design and functionality by doing so the system can enhance the overall user experience.[6]

C. Image Related Issues

A commonly seen technique employed by researchers involves requesting users to capture an image of the target product, which is then processed by a model for object identification, classification, and generation of any other required output. However, this approach introduces several challenges, particularly for the visually impaired. These users will struggle to capture stable and clear images, as they may encounter obstacles while taking the image of the object in question[7][8]. These unclear images often lead to inaccurate identification of the object.

D. Others

Other challenges posed by existing object detection and classification techniques include angle and lighting dependency as well as high-cost implications. Existing models have been shown to exhibit unpredictable outcomes when the angles of the images change or the lighting under with the image is taken is different [9]. Another point to note is that higher accuracy in these systems comes at a higher cost, making them less accessible and affordable for both developers and end users [10].

The limitations impede the effectiveness and adoption of these object detection technologies for assisting visually impaired individuals.

III. DATASETS AND EVALUATION METRICS

In this section, we present an overview of the datasets and evaluation metrics frequently employed in object detection tasks in assistance systems aimed at aiding visually impaired individuals. We discuss the details of the datasets as well as an overview of the most commonly used evaluation metrics and their importance in object detection systems.

A. Datasets

PASCAL VOC 2007: This dataset is a very commonly used dataset that consists of annotated images covering a diverse set of object categories. The dataset consists of 9,963 photos in all, with 24,640 labeled samples across 20 different classes. It is seen being used to train a deep learning-based object detection model to assist visually impaired users in object recognition tasks. [10]

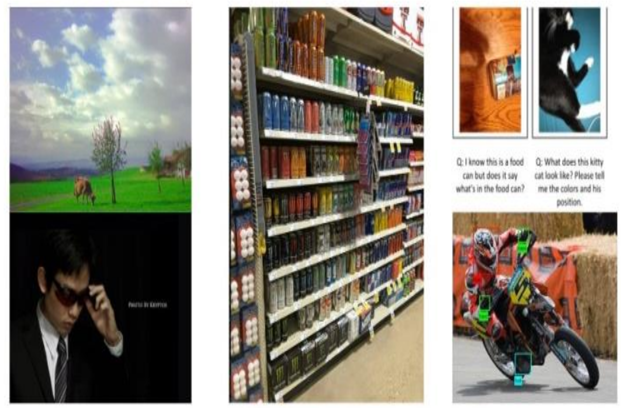

Figure 1: Sample Images from Datasets

- MS Coco: This is a widely used dataset that has more than 330,000 images and contains more than 80 object categories such as cars, boats, trucks, etc. The COCO dataset was developed to introduce more categories, than the other pre-existing datasets at the time, so as to increase the scope of computer vision tasks datasets. This dataset has been utilized in various research works for developing object detection systems for visually impaired individuals.
- SKU100K: SKU-110K is a large-scale visual dataset commonly used for object recognition and image classification tasks. It consists of around 111,000 images with annotations for over 11,000 different object classes.SKU110K was designed to address some limitations of existing datasets, such as the lack of object variety and limited coverage of fine-grained categories. It covers a broad range of object classes, Due to its large size and comprehensive coverage, SKU-110K has become a valuable resource in the field of computer vision research. It has been widely used for tasks such as object detection, and image classification
- ImageNet: ImageNet is a dataset that has played a crucial role in advancing the field of computer vision, particularly in the development and evaluation of deep learning models. The dataset is made up of roughly 14 million labeled images across 21,841 categories. ImageNet is most commonly used in general-purpose computer vision tasks.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)
https://ijemr.vandanapublications.com
https://doi.org/10.5281/zenodo.10648650

- VizWiz: This dataset was specifically created with the intention of being used in computer vision tasks to assist the visually impaired. The dataset consists of images taken by visually impaired individuals. VizWiz consists of approximately 31,000 images collected in real-world scenarios, covering a wide range of objects and environments. These images are annotated with text descriptions of the image. These annotations provide additional context and information about the visual content, which enables researchers to explore various computer vision tasks with the visually impaired in mind.

B. Evaluation Metrics

Evaluation metrics are quantitative measures that are used to examine and assess the performance of different models.[11].These metrics are essential in object detection as

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$AP_k = \sum_{k=0}^{k=n-1} [Recall(k) - Recall(k+1)] * Precision(k)$$

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$n = number\ of\ classes$$

Figure 2: mAP Formula

they provide a measure through which different models can be compared, which aids the development of accurate and efficient systems. Although there are many kinds of metrics such as logarithmic loss, AUC (Area Under Curve) and F1 Score, mAP(Mean Average Precision) is one of the most commonly used. mAP is calculated by computing the AP(Average Precision) for each class and then finding the average of these values. AP can be found by calculating the AUC for the precision-recall curve of each class

IV. OVERVIEW OF OBJECT DETECTION ALGORITHMS

Various object detection algorithms have been used in assistance systems designed for the visually impaired to achieve their intended goals. This section provides a comprehensive examination of these algorithms and their overall functioning. We have separated the algorithms into 2 types: one-stage detection and two-stage detection.

A. One-Stage Detection

One-stage detection also known as Single-shot detection kind of object detection uses a single pass of the input image to generate a prediction on the location and presence of an object in an image. Although One-stage detection is more computationally efficient than two-stage detection it tends to be less accurate.

1) YOLO: YOLO also known as You Only Look Once is one of the most popular object detection algorithms. It is very widely used in real-time computer vision tasks due to its high speed and accuracy. It was first introduced by Joseph Redmon et al. [12] and has since undergone several iterations. YOLO makes use of an end-to-end neural network to predict bounding boxes and class probability in a single iteration. The architecture of YOLO which makes use of a simple CNN is shown in Figure 3. In the YOLO architecture the first 20 layers have been pre-trained using the ImageNet model. This is facilitated by adding a temporary average pooling layer and a fully connected layer. These pre-trained layers are then used to perform detection tasks. The final fully connected layer predicts the class probability and bounding box coordinates.[12].The input images in YOLO are divided into an S x S grid. Each

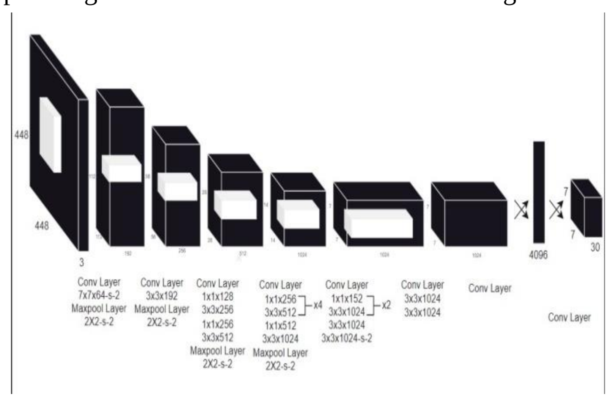

Figure 3: Yolo Architecture

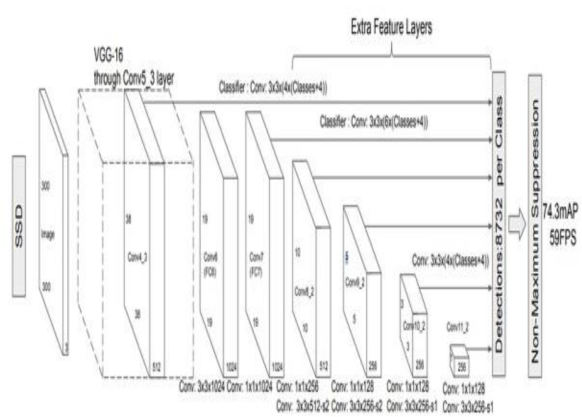

Figure 4: SSD Architecture

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10648650

cell in this grid is responsible for predicting the bounding box and confidence scores for it's respective cell. Each cell can have multiple bounding boxes. A predictor bounding box is assigned based on the highest IOU(Intersection over Union). YOLO also makes use of a post-processing step known as non-maximum suppression(NMS) to improve the accuracy and efficiency of the overall model.[13]

2) SSD: Single Shot MultiBox Detector known as SSD is an object detection algorithm first implemented and introduced by Liu et al. in 2016. [.]SSD was designed to detect objects of varying sizes on a single iteration. The main idea behind SSD is to combine feature maps of different resolutions to detect objects at multiple scales. It is a modification of the well-known VGG-16 architecture. By replacing the fully connected layers in VGG-16 with auxiliary convolutional layers the SSD model was able to extract features at multiple scales and progressively decrease the input size at each layer. SSD divides the input image into a grid of cells and each cell is responsible for detection objects within the region of the cell. Multiple objects of the same instance in a single image are detected with the help of anchor boxes in SSD. Anchor boxes are boxes that are assigned prior boxes which have a predefined shape and size. Similar to YOLO, SSD also utilised NMS to prune the bounding boxes. [14].

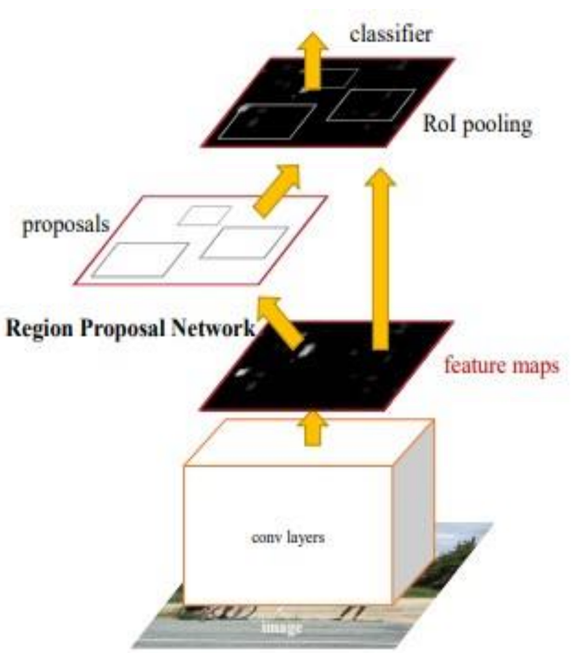

Figure 5: Faster R-CNN Architecture

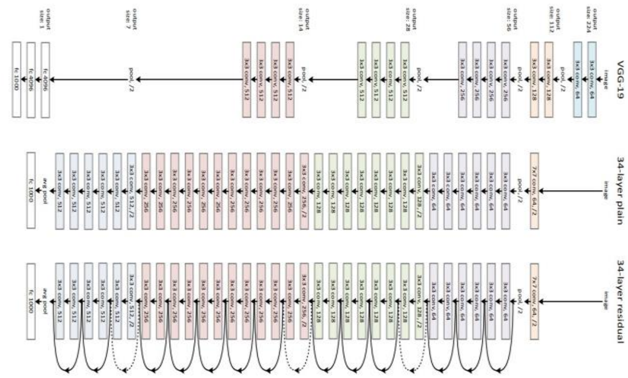

Figure 6: ResNet Architecture

B. Two-Stage Detection

Two-Stage detection also known as Two-shot detection is a method which involves the input image being passed twice to make predictions. The first pass involves generating a set of potential object locations and the second pass is used to refine these locations to make the final prediction. This approach is more accurate but is also more computationally expensive.

1) Faster R-CNN: Faster R-CNN (Region-based Convolutional Neural Network) is a popular object detection algorithm introduced to improve upon previous approaches by combining a region proposal network (RPN) with a convolutional neural network (CNN) for object detection tasks. [15].The architecture of this model consists of two main components: the RPN and the Fast R-CNN. The RPN generates region proposals, which are potential bounding boxes around objects, using a sliding window approach. It simultaneously predicts objectness scores and adjusts the coordinates of the proposals. The RPN is a fully convolutional network that shares the convolutional layers with the Fast R-CNN, making the overall computation more efficient. The Fast R-CNN takes these region proposals and feeds them into a CNN, which extracts features from each proposed region. These features are then used for object classification and bounding box regression. Faster R-CNN achieves high accuracy by training the RPN and Fast R-CNN together in a multi-task learning framework. The RPN generates region proposals and learns to classify them as either object or background, while the Fast R-CNN uses these proposals for object classification and bounding box refinement. The Faster R-CNN architecture has become a foundation for many state-of-the-art object detection systems. Its combination of region proposal generation and feature extraction using a shared CNN has led to significant improvements in both accuracy and speed.

2) ResNet: ResNet models can be trained to recognize and classify objects, enabling visually impaired individuals to identify and distinguish various objects in their environment. By leveraging ResNet's deep architecture

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)

https://doi.org/10.5281/zenodo.10648650

and skip connections, the model can learn complex visual representations and make

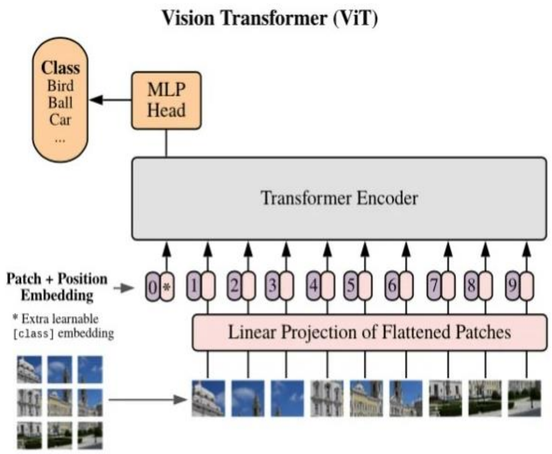

Figure 7: ViT Architecture

accurate predictions. ResNet models can analyze and interpret complex scenes, providing a high-level understanding of the environment. ResNet models can be utilized in conjunction with assistive technologies such as cameras or smartphones. By capturing images and running them through the ResNet model, real-time object recognition and scene understanding can be provided to visually impaired users, helping them navigate their surroundings more effectively. ResNet models can also be applied to text recognition tasks, enabling visually impaired individuals to access written information in their surroundings. By leveraging ResNet's deep layers and powerful feature extraction capabilities, the model can accurately recognize and extract text from images, providing text-tospeech capabilities or converting the text into accessible formats.By combining ResNet with natural language processing techniques, visually impaired individuals can benefit from image captioning systems. ResNet can extract visual features from images, which can then be fed into a language model to generate textual descriptions or captions for the images, enhancing the understanding of visual content.

3) Vision Transformers: ViTs can be trained to recognize and classify objects, which can aid visually impaired individuals in identifying and distinguishing various objects in their environment. By capturing images through cameras or smartphones, the ViT model can provide real-time object recognition and describe the objects to the user. This can assist visually impaired individuals in navigating unfamiliar places, detecting obstacles. The key idea behind ViTs is to treat an image as a sequence of patches rather than a grid of pixels. The image is divided into smaller patches, which are then flattened and treated as input tokens. These tokens are fed into a transformer model, allowing the ViT to capture global relationships and dependencies among the patches. ViTs can be trained to recognize and extract text from images, enabling visually impaired individuals to access written information in their surroundings. ViTs combined with natural language processing can generate textual descriptions or captions for images. This can assist visually impaired individuals by providing verbal descriptions of images, enhancing their understanding of visual content. By integrating ViTs with localization technologies and sensors, visually impaired individuals can receive real-time guidance and directions. ViTs can analyze the surroundings, recognize landmarks, and provide audio cues to assist with navigation. It is worth noting that while Vision Transformers have shown promising results in computer vision tasks, their application in assisting visually impaired individuals is an active area of research. Customizing and optimizing ViTs specifically for the needs and challenges of visually impaired users would require further investigation, data collection, and user-centered design considerations to ensure their practical effectiveness and usability.

V. METHODS OF IMPLEMENTATION

Through our research we were able to identify that assistance systems for the visually impaired fall into 1 of 2 categories primarily: Software and Hardware. With Software leaning more towards the mobile application side and hardware referring to wearable device that the user will have to carry with them.

A. Software Focused Implementation

In the last few years with the surge in the number of mobile applications we can see more and more assistive application for the visually impaired being developed. These applications can detect objects in real-time or can take image input for classification [16]. YOLO and SSD are most commonly seen being used in the real-time detection systems whereas Faster R-CNN sees more use in the image-based detection applications.

B. Hardware based Implementation

There are multiple physical devices as well that make use of different object detection and classification technology to provide assistance to the visually impaired. There are devices such as wearable eyeglasses and smart bands which incorporate object detection techniques and models in order to function as required. Viziyon, a wearable assistive technology for visually impaired individuals makes use of ResNet to enable object recognition, scene understanding, and navigation assistance through the wearable device.

In terms of accuracy, ResNet has demonstrated impressive performance in image classification tasks, achieving top-1

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10648650

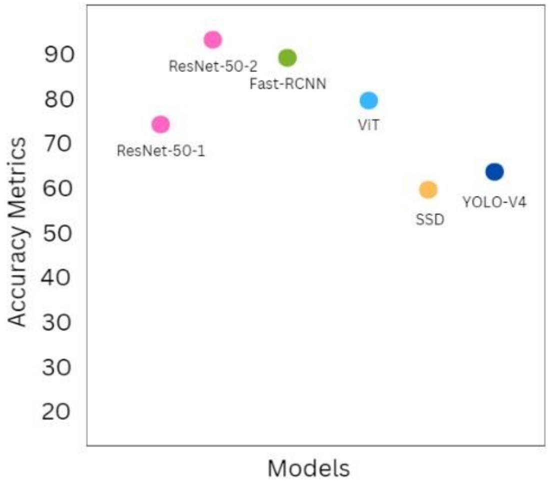

Figure 8: Model Comparison

and top-5 accuracy rates of 75.3% and 92.2% respectively on the ImageNet dataset.[19] Its deep architecture with skip connections allows for effective feature extraction, leading to high accuracy in recognizing and classifying objects. Vision Transformer (ViT) has emerged as a promising model for computer vision tasks, although specific accuracy figures in the given context are not mentioned. However, ViT has shown competitive performance in image classification tasks, often approaching or surpassing the accuracy of traditional convolutional neural network (CNN) models like ResNet.[20]Fast R-CNN, a widely used model for object detection, has demonstrated good accuracy in localizing and recognizing objects, with reported accuracy rates of 91% and recall rates of 94% in certain applications. Models like SSD and YOLO, designed for real-time object detection, have also achieved high accuracy in detecting objects with varying aspect ratios and scales[21]. For instance, the standalone SSD detector achieved a mean average precision (mAP) score of 67.8% for a specific class[22]. It is important to consider that reported accuracy can vary across different datasets, evaluation metrics, and application domains. The choice of model depends on specific task requirements and available resources. Additionally, the accuracy of these models can be further enhanced through fine-tuning, dataset augmentation, and architectural improvements tailored to the specific application context.

VI. CONCLUSION

In conclusion, various models have been utilized for objection detection and identification particularly helping visually impaired individuals models like ResNet showed great accuracy metrics when it came to image classification, and Vision Transformer (ViT) has also emerged as a promising model, showcasing competitive accuracy in image classification tasks.

Fast R-CNN, SSD, and YOLO have shown good accuracy in object detection and localization. It is important to note that accuracy can vary vastly just by using the same model with different datasets eg.ResNet-50.The choice of the model depends on the task at hand and the resources that are available Further research and improvements are necessary to customize these models specifically for the needs and challenges of visually impaired users. Overall, these models provide valuable tools for assisting visually impaired individuals in tasks such as object recognition, text extraction, and navigation, paving the way for more inclusive and accessible technologies.

ACKNOWLEDGMENT

First and foremost, we would like to thank our mentor, Prof. Sreenath, for his guidance, expertise, and continuous encouragement. Their insightful feedback and constructive criticism played a vital role in shaping this research work. Furthermore, we are grateful to our panel consisting of Prof. Bharthi and Prof. Alpha for their valuable input and discussions. Their feedback and suggestions significantly contributed to the improvement of our work.

REFERENCES

[1] Russakovsky, O. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211-252,

[2] Ren, S. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*, pp. 91-99.

[3] Ross Girshick, Jeff Donahue, Trevor Darrell & Jitendra Malik. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Reognition (CVPR)*, pp. 580-587.

[4] Krizhevsky, A. (2012). Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097-1105.

[5] Koenecke et al. (2018). *Buolamwini Gebru*.

[6] Cruz, E., Olivares, A & Favela, J. (2018). *International Journal of Human-Computer Studies, 109*, 58-71.

[7] Smith, J. (2019). Exploring the usability of smartphone apps for supporting independent

travel by people with visual impairments. In: *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 155-166.

[8] Johnson, K & Robertson, R. (2020). Investigating the accessibility of mobile technologies for people with visual impairments. *Universal Access in the Information Society, 19*(2), 369-379.

[9] Rahman & M. S., et al. (2020). Real-time object detection for visually impaired and blind people using deep learning. In: *3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 928932.

[10] Mills, D. (2019). Towards affordable, accessible, and available assistive technology. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. LBW2102.

[11] S. Bhole & A. Dhok. (2020). Deep learning based object detection and recognition framework for the visually-impaired. *Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India*, pp. 725-728.

[12] Joseph Redmon, Santosh Divvala, Ross Girshick & Ali Farhadi. (2016). You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 779-788.

[13] R., Kundu. (2023). *YOLO: Algorithm for Object Detection Explained [+Examples]. V7.*

[14] Everingham, M., Eslami, S. M. A., Gool, L. V., Williams, C. K. I., Winn, J & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision, 88*(2), 303-338.

[15] Girshick, R. (215). Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448.

[16] Jawaid Nasreen, Warsi Arif, Asad Ali Shaikh, Yahya Muhammad & Monaisha Abdullah. (2019). Object detection and narrator for visually impaired people. *IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS).*

[17] C.K. Amarasinghe, R.D.S.P. Pinto & K.N. Sudusinghe. (2022). Guardian - smart assistant tool for visually impaired people. *4th International Conference on Advancements in Computing (ICAC).*

[18] Sreeraj Ma, Jestin Joyb., Alphonsa Kuriakosec., Bhameesh. M. Bd, Anoop. K. Babud & Merin Kunjumond. (2020). VIZIYON: Assistive handheld device for visually challenged. In: *Third International Conference on Computing and Network Communications (CoCoNet'19), Procedia Computer Science, 171.*

[19] K. He, X. Zhang, S. Ren & J. Sun. (2016). Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[20] A. Dosovitskiy et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In: *Proceedings of the International Conference on Learning Representations (ICLR).*

[21] J. Redmon et al. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.

[22] W. Liu et al. (2016). SSD: Single shot multibox detector. In: *Proceedings of the European Conference on Computer Vision (ECCV).*

[23] Pierre Dognin, Igor Melnyk, Youssef Mroueh, Inkit Padhi, Mattia Rigotti, Jarret Ross, Yair Schiff, Richard A & Young Brian Belgodere. (2020). Image captioning as an assistive technology: Lessons learned from vizwiz 2020 challenge. *arXiv:2012.11696v1 [cs.CV]*.

[24] Joseph Redmon & Ali Farhadi. (2018). YOLOv3: An Incremental Improvement. *arvix.*

[25] Savera Sarwar,Danish Channa,Muhammad Turaba,Aisha Chandio & M. Uzair Sohu.(2022). Advanced Audio Aid for Blind People. *arXiv:2212.00004v1 [cs.HC]*.

[26] Yuxuan Cai, Hongjia Li, Geng Yuan, Wei Niu, Yanyu Li1,Xulong Tang, Bin Ren & Yanzhi Wang. (2009). YOLObile: Real-time object detection on mobile devices via compression-compilation co-design. *arXiv:2009.05697v1 [cs.CV]*.

[27] Meera Santhana krishnan, Sharmikha Sree Rajarathinam, Valarmathi K. (2019). Effective shopping method for visually impaired people using optical character recognition. *International Journal of Advanced Technology and Engineering Exploration.*

[28] Prabu Selvam, Joseph Abraham & Sundar Koilra. (2022). A deep learning framework for grocery product detection and recognition. *Research Square.*

[29] A. Labeeshan, S.A.R.L.P. Satharasinghe, A.M. Dhananjana, T.D. Rupasinghe, U.U. Samantha Rajapaksha, Prasanna S & Haddela (2021). iRetina: An intelligent mobile application for the visually impaired in an indoor environment. *IEEE 16th International Conference on Industrial and Information Systems (ICIIS).*

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-14, Issue-1 (February 2024)

https://doi.org/10.5281/zenodo.10648650

[30] Ankit Sinha, Soham Banerjee & Pratik Chattopadhyay. (2022). An improved deep learning approach for product recognition on racks in retail stores. *arXiv:2202.13081v1 [cs.CV]*.

[31] Chih-Hsien Hsia,Tsung-Hsien William Chang,Chun-Yen Chiang & Hung-Tse Chan. Mask R-CNN with new data augmentation features for smart detection of retail products. *MDPI*, *12*, 2902.

[32] Sunit Vaidya, Naisha Shah, Niti Shah & Radha Shankarmani. (2020). Real-Time object detection for visually challenged people. *4th International Conference on Intelligent Computing and Control Systems (ICICCS).*

[33] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y & Berg, A. C. (2016). SSD: Single shot multibox detector. In: *European Conference on Computer Vision*, pp. 21-37.

[34] S. Shukueian. (2022). How single shot multibox detector (SSD) real-time object detection technique works?. *Medium.*

[35] Anish Aralikatti, Jayanth Appalla, Kushal S, Naveen G S, Lokesh S & Jayasri B S. (2020). Real-time object detection and face recognition system to assist the visually impaired. *Journal of Physics: Conference Series, 1706, First International Conference on Advances in Physical Sciences and Materials.*

[36] Therese Yamuna Mahesh, Parvathy S S, Shibin Thomas, Shilpa Rachel Thomas & Thomas Sebastian. (2020). CICERONE- A real time object detection for visually impaired people. *IOP Conference Series: Materials Science and Engineering, 1085*.

[37] https://www.v7labs.com/blog/yolo-object-detection.

[38] Swapnil Bhole & Aniket Dhok (2020). Deep learning based object detection and recognition framework for the visually-impaired. Conference: *Fourth International Conference on Computing Methodologies and Communication (ICCMC).*

[39] B. Li, J. P. Munoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi & M. Yousuf. (2019). Vision-based mobile indoor assistive navigation aid for blind people. *IEEE Transactions on Mobile Computing, 18*(3), 702–714.