VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2024, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 12, Number 2, July-December 2024 pp: 165-177

Towards Improved Assistive Technologies: Classification and Evaluation of Object Detection Techniques for Users with Visual Impairments

Shakeela Naz [1*] and **Fouzia Jabeen** [1]

[1]Department of Computer Science Shaheed Benazir Bhutto Women University Peshawar , Pakistan

Keywords:*Visually Impaired, Object Detection, Deep Learning, Datasets, Evaluation Metrics.*

Journal Info:
Submitted:
September 15, 2024
Accepted:
November 30, 2024
Published:
December 04, 2024

Abstract

Even though millions of people struggle to interact with the outside world due to visual impairments, vision is an essential part of our daily lives. Because of its ability to identify and navigate around objects in their surroundings, object detection a crucial component of computer vision has become a potentially helpful solution. This study offers a thorough analysis of object detection techniques utilizing a dual classification system that combines traditional and deep learning methods. In addition, we analyze the most popular evaluation metrics and datasets for these systems' training and evaluation. Unlike previous surveys, our work provides a unique perspective by carefully examining the latest advancements in both innovative deep learning models and traditional approaches. The survey's conclusion highlights current problems and recommends future research directions, highlighting the need for more effective models, diverse datasets, and multimodal data integration to improve assistive technologies for the visually impaired.

***Correspondence author email address:** nazshakeela38@gmail.com
DOI: 10.21015/vtcs.v12i2.1911

1 Introduction

Our eyes are our window to the outside world, but not for people who are visually impaired from birth, illness, accident, or old age. Because our daily activities heavily rely on our vision, they are severely disrupted in the absence of clear vision. As per the World Health Organization (WHO), visually impaired is defined as the inability to see objects between 0 and 6.1 meters ahead of us when our sight indicator is below the cutoff point, or 6/18. Some countries classify people as legally visually impaired if their field of vision is less than 20 degrees[1].

Therefore, creating new solutions is crucial to assisting Partially Visually Impaired (PVI) in carrying out these tasks quickly and effectively and inspiring them to engage with their social surroundings[2].

Due to recent technological advancements and its broad range of applications, object detection has been receiving more and more attention in recent years. In both academia and real-world applications, including robotic vision, autonomous driving, transportation surveillance, drone scene analysis, and security monitoring, this task is being thoroughly studied. Among the numerous elements and initiatives that have accelerated the advancement of object detection methods, the creation of deep convolution neural networks and GPU processing power deserve special recognition[3].

The object detection system functions exactly as a child would have learnt in school. In school, a child is trained by learning the names of different objects and shapes. He can identify all related objects that he has previously learnt once he has mastered every object. Similar to how a child learns, the machine must be completely trained using a variety of machine learning algorithms by adding all of the object names. All of the characteristics of related objects are kept in a database to facilitate the training process. The next step is to test an input image by comparing its features with the feature dataset that has been stored. An effective object recognition system will quickly and accurately produce the object's name [4].

Some methods for assisting visually impaired persons in finding objects in their environment have been proposed recently. Various research methodologies have been employed in this regard, such as artificial intelligence techniques, ultrasound signals, deep learning, and assisting visually impaired individuals in identifying their surroundings and avoiding obstacles, are also being studied [5].

All of the previous research was limited to giving a summary and evaluation of a small selection of object detection models for the visually impaired, despite the fact that there were other models available at the time. In most previous surveys, the models were classified as either one-stage detectors or two-stage detectors. In [6] number of topics have been covered in earlier works, including the analysis of conventional, two-stage, and one-stage object detection techniques, dataset preparation and readily available standard datasets, annotation tools, and performance evaluation metrics.

Similarly, in [7] the state-of-the-art object recognition, face recognition, and text-to-voice recognition methods proposed to assist the visually impaired are reviewed in detail. [8]. For this research, different pretrained models were reviewed for the detection of objects such as Region-based Convolutional Neural Network (R-CNN), Region-based Fully Convolutional Neural Network (R-FCN), single- shot multi box detector (SSD) and You Only Look Once (YOLO), with different extractors of characteristics such as VGG16, ResNet, Inception, Mobile Net. W. J. Chang [9] surveyed the current state of wearable and assistive technology and provides a critical evaluation of each system, pointing out its benefits and drawbacks. These models were used with a variety of extractors of characteristics, such as VGG16, ResNet, Inception, and MobileNet. R. Alex [10] provided a comprehensive survey of one stage and two-stage object detection techniques, as well as the most recent detection solutions and a number of noteworthy research trends, are thoroughly and methodically reviewed. We tried to cover both one-stage and two-stage object detection models in this paper, along with traditional methods for visually impaired individuals to detect objects. The number of models we have listed has not been thoroughly covered and analyzed in any previous work. Furthermore, we separated the detection models into two groups. The first focusses on traditional techniques, which are further divided into two groups: featured based and Tamplet Matching. The second focuses on deep learning models, which are further divided into CNN-based and transformer-based models. We also look at the most popular datasets and evaluation metrics for these systems' training and evaluation.

This paper's remaining sections are organized as follows: Section 2. Object detection techniques 3. Dataset and available standard dataset 4. Evaluation matrices. Finally, Section 5 ends with a conclusion and future direction.

2 Techniques for Object Detection

The process of identifying a foreground object in a frame is called object detection. Any person, animal, or other object of interest could be the desired object [9]. Previous studies have explored a range of approaches to assist individuals with visual impairments in identifying objects and navigating obstacles. Several surveys have focused on deep learning methodologies, while others have examined artificial intelligence techniques more broadly. Some research has investigated the potential of ultrasound signals, whereas conventional methods have also been explored extensively. Together, these studies contribute a diverse set of insights into technological advancements aimed at enhancing spatial awareness and safety for those with visual impairments [5]. For example, A. Lavric [11] summarized the characteristics of some commercial assistance programs and looked at the border use of AI and visible light communications, highlighting how well they meet the needs of blind people. Similarly, A. Khan [12] examined research directions in smartphone-based assistive technologies for the blind and emphasized the necessity of technological developments, an accessibility-inclusive interface paradigm, and cooperation between computer professionals, medical specialists, usability experts, and domain users in order to fully realize the potential of information and communication-based interventions for the blind. In [13] the author offers a thorough analysis of the many assistive technologies and systems intended to help people who are blind or visually impaired. These technologies are categorised according to their functions and uses, such as carrying mode, object type, coverage, and analysis type. In [14] the state of research on object recognition for visually impaired individuals on mobile platforms was compiled and evaluated using the Systematic Literature Review approach. Wearable devices have also gained attention, as reviewed by K. Manjari [15], who reviewed all the pertinent wearable and handheld devices developed for visually impaired people, wearable devices have also gained attention. The main focus of the review was on the devices' prominent features, and an analysis was conducted based on a few factors such as power consumption, weight, economics, and user-friendliness. In order to address the changing needs of visually impaired people in their daily lives, more research focusses especially on improving real-time object detection[16]. Z. Q. Zhao [17] review of deep learning-based object detection frameworks and give a brief overview of a few particular tasks, such as pedestrian, face, and salient object detection, and concentrate on common generic object detection architectures. A. Balothiya [18] briefly addressed the various object detection algorithms for people with visual impairments, including the CNN family, SSD, and YOLO. These algorithms are compared based on a number of characteristics, including speed, cost, accuracy, and complexity. M. Deshpande [19] discussed several object classification and detection techniques (CNNs, Fast R-CNN, SSD, YOLO, ViT, and ResNet) that are particularly pertinent for helping visually impaired individuals. We divided object detection methods into two primary categories: Deep Learning Methods and Traditional Methods, which we will address individually.

2.1 Traditional Methods

Traditional object detection techniques are based on shallow trainable architectures and manually created features. Using complex ensembles that combine high-level context from object detectors and scene classifiers with numerous low-level image features, their performance quickly stalls [17].

2.1.1 Feature-Based Methods

To address different challenges faced by visually impaired people, feature descriptors like Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) [20][21][22] are often paired with traditional machine learning classifiers like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). Developed by David Lowe in 2004, SIFT provides robustness against changes in illumination and blur. It matches images under different scales and rotations. While SURF builds upon the concepts of SIFT, it uses a fast Hessian matrix computation to improve accuracy and speed in descriptor generation and location detection efficiency. These algorithms are widely used for tasks that do not require specific class knowledge, such as 3D mesh reconstruction and image stitching. If given enough training data, deep neural networks can classify items on an assembly line; however, less complex approaches such as color thresholding can accomplish comparable outcomes[23].

The Histogram of Directed Gradients (HOG) is a feature descriptor for visually impaired people that is used in computer-view and image analysis for object detection. The fundamental concept of a histogram is that a series of intensity gradients or boundaries represents the presence and structure of local objects within an image. The pictures are separated into cells, which are tiny, connected regions. Meanwhile, for every pixel in every cell, a gradient direction histogram is generated. The concatenation of these histograms is the descriptor. The HOG descriptor has a few key advantages over other descriptors. With the exception of object orientation, it is invariant to geometric and photometric transformations because it operates on local cells.[24].

2.1.2 Template Matching

For the purpose of object detection and recognition for visually impaired individuals, computer vision and image processing rely heavily on comparing similarities between various scenes. Template matching[25][26] is the widely used method compares different regions of a source image to a template image in order to find possible matches. High-level machine vision techniques such as edge detection, mobile robot navigation, object localization, and medical image registration depend on this technique. The process of template matching determines if an object is present in an image and where it is by calculating the difference measure between the template and potential matches. Numerous computer vision tasks, such as stereo correspondence, motion tracking, and 3D structure reconstruction, are based on this process[27].

2.2 Deep Learning Based Object Detector

In order to recognize and categorize objects into different classes, supervised learning methods are frequently combined with neural networks (NN). The most popular application of NN in the field of computer vision is found in CNNs that use hierarchical feature learning, which makes use of the spatial information on the ever-increasing areas of images as the system processes them. More precisely, they drastically lower the parameters and, as a result, the computational power in comparison to conventional feed-forward fully connected networks due to their structure and characteristics (e.g., Local Receptive Fields, Shared Weights, Pooling)[28].

2.2.1 CNN-Based Methods

Object recognition is a popular artificial intelligence application for people with visual impairments and a very fruitful task for deep neural networks [29]. Deep learning techniques for regression, classification, and object detection are currently applied in two different ways to assist visually impaired individuals with object detection. The R-CNN, Fast R-CNN, and Faster R-CNN architectures all use the two-stage algorithm [30], Another algorithm is the one-stage algorithm [31], represented by the following symbols: which is symbolized by YOLO[32] [33] [34] [35] [36] [37], SDD[38] [39] [40], and so on[41]. R-CNNs, Fast R-CNNs, Faster R-CNNs [42] [43] [44][45] [46] [47] [48] and Mask RCNNs all use region proposal networks in their initial stages. After that, the bounding box regression and object classification processes propagate the region proposals down the pipeline[29].

By treating object detection as a straightforward regression problem, single-stage detectors YOLO and SSD were proposed to improve the detection efficiency to be suitable for real-time applications. They use an input image to learn the class probabilities and bounding box coordinates. One-stage detection algorithms, which are quicker but less accurate, only extract one feature at a time[2].

2.2.2 Transformer Based Method

Predicting bounding boxes and category labels for objects of interest is known as object detection. Convolutional neural networks, which have a head for localization and classification and a backbone for feature extraction, are commonly used in this process. A more recent alternative, called Detection Transformer (DETR) [49][50][51] frames object detection as a set-based matching problem. DETR eliminates hand-designed components like non-maximum suppression and anchor generation by modelling object relations and global image context using Transformers. DETR is innovative, but it has drawbacks like incompatibility with feature pyramids, limited input resolution because of the quadratic complexity of the self-attention module, and longer training times needed for convergence. As a result, there is a great need for DETR to be improved in an efficient manner [52].

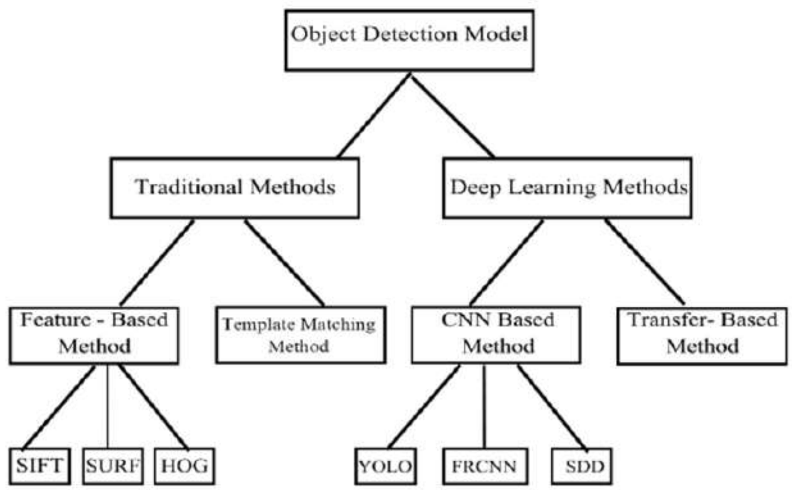

Figure 1. Techniques for Object Detection

3 Datasets

It is impossible to overstate the value of datasets for those who are visually impaired. Large-scale datasets are now central to the development and assessment of models for the visually impaired in the field of machine learning. Because these kinds of systems are used in real-world situations where real people's lives and livelihoods are at stake, it is critical that academics, advocacy organizations, and the general public comprehend the datasets contents and how they impact system performance [53].

Compilations of labelled photos make up datasets. Object position and class are included in the annotation. The number of classes in each dataset is predetermined. Groups of training and testing images are created from the class images. The PASCAL Visual Object Classes 2007 (VOC 2007), PASCAL VOC 2012, ImageNet, Microsoft Common Objects in Context (COCO), and OpenImages datasets are five of the most well-known datasets in

3.1 Analyses of Dataset

the generic object detection field [54]. Table 1 shows the most common images datasets.

The comprehensive and diverse nature of the COCO (Common Objects in Context) dataset [55] [56] [49] [57] justifies its selection as the preferred dataset and benchmark for individuals with visual impairments. Because the COCO dataset can challenge models with a wide range of challenges, such as occlusions, diverse object sizes, and crowded scenes, which mirror the complexities encountered in real-world applications, researchers widely use it [58]. MS COCO is an open-source dataset comprising 328,000k images and 80 classes related to our daily activities. It is used for object detection, segmentation, key point detection, and captioning. The images are somewhat complex, but they are easily recognizable. The size of the dataset is 47 GB[59][60].

Many visually impaired people use transfer learning, a widely used computer vision technique made possible by large datasets such as ImageNet [61] [62] [63] PASCAL [64], and COCO [65]. ImageNet open-source nature and placement within the WordNet hierarchy both contribute to its increased usability [66]. Among these, ImageNet is especially useful because of its enormous size and variety—it comprises 14,197,122 images from 1,000 different object classes. Many tasks, including image classification [65] , object detection [67], and image captioning. . Its 47GB total size makes it a better option for transfer learning than other datasets because it offers a strong basis for creating precise and broadly applicable models [68][69][70]. For those with visual impairments, PASCAL VOC (Visual Object Classes) [58] [49] has been essential in the visual object recognition process since 2005. It provides standardized image data along with comprehensive annotations and evaluation protocols. It has made a major contribution to the advancement of computer vision technologies by making it easier to develop algorithms for tasks like image classification and object detection for the blind. From 2005 to 2012, the yearly VOC challenges played a pivotal role in promoting innovation within the industry. Notably, VOC12 represents a significant advancement over previous iterations with over 11,530 training images and 27,450 annotated objects spanning 20 object classes. The dataset is publicly available in XML format and is a favorite benchmark among computer vision researchers and practitioners due to its extensive coverage and open accessibility. PASCAL VOC is an important reference dataset in visual recognition research because of its rich history and broad adoption [59][71][72].

The publicly accessible Open Images dataset V4[73] comprises 9.2 million images with unified annotations for three tasks: image classification, object detection for visually impaired individuals, and visual relation detection. Each image was downloaded from Flickr without using predefined class names. This dataset includes 600 object classes with 15.4 M bounding boxes, 57 classes with 375 visual relationship annotations, and 30.1M image-level annotations with 19.8 k concepts[74]. Because all of the objects in the dataset were annotated by experts, it offers accuracy and consistency. This dataset includes multiple objects in all of its high-quality images [59].

Dataset for Object detection in Aerial Images (DOTA) [75] is a sizable dataset used for aerial image object detection. It is useful for creating and assessing aerial image object detectors. Various platforms and sensors are used to gather the images. Every image is between 800 × 800 and 20,000 × 20,000 pixels in size, and it includes objects with a broad range of sizes, shapes, and orientations. An arbitrary quadrilateral is used by specialists in aerial image interpretation to annotate the instances in DOTA images. In order to reflect changing real-world conditions, we will keep updating DOTA and expanding its size and scope.

4 Evaluation Matrices

Evaluation matrices offer important information about the advantages and disadvantages of an algorithm for a range of object sizes and complexity levels. Researchers and developers can use these metrics to improve their models and make them more suitable for use in the real world by utilizing the COCO benchmark, Pascal VOC, and Open image dataset for visually impaired individuals [58]. Based on the classification result, samples from both the positive and negative classes in the bi-class scenario can be divided into four groups. Based on the classification result, positive and negative class samples in the biclass scenario can be divided into four groups [75]. This can be denoted as the confusion matrix shown in Table 2 [76].

Several measures can be derived from this matrix:[41]

True positive rate:

$$\mathrm{TPR} = \frac{TP}{TP + FN}$$

Table 1. Comparison of Datasets

Research Articles	Dataset	Format	Images	Size	Source	Used for
[44][45][46], [47][48][38]	COCO 2017 Dataset	JSON and Image File	328,000	47GB	Open source	1) Object detection 2) Segmentation 3) Keypoint Detection 4) Captioning
[49][50][51], [52][53][54], [55]	ImageNet Dataset	WordNet Hierarchy	14,197,122	150GB	Open source	1) Image classification 2) Localization
[58][49][47][59]	Pascal VOC Dataset	XML Format	11,318	8.4GB	Open source	1) Classification 2) Object Detection 3) Semantic Segmentation
[47][61][62]	OpenImages Dataset	Image Level Label and Object Bounding Boxes	9,000,000	18TB	Open source	1) Object Segmentation 2) Visual Relationships
[63][64]	Dota Dataset	PNG Format	11,268	20GB	Not open source	1) Object Detection in Aerial Images

	Predicted Positive	Predicted Negative
Actually Positive	True Positive (TP)	False Negative (FN)
Actually Negative	False Positive (FP)	True Negative (TN)

Table 2. Confusion Matrix

False negative rate:

$$\mathrm{FNR} = \frac{FN}{TP + FN}$$

True negative rate:

$$\mathrm{TNR} = \frac{TN}{TN + FP}$$

False positive rate:

$$\mathrm{FPR} = \frac{FP}{TN + FP}$$

A models precision is its capacity to recognize only pertinent objects. It is the proportion of accurate positive forecasts. A model's recall is its capacity to locate all pertinent cases, or all ground-truth bounding boxes. It represents the proportion of accurate positive predictions made out of all available ground truths. The precision × recall curve represents a trade-off between precision and recall for various bounding box confidence values produced by a detector. For unbalanced dataset Precision and recall [77] are especially useful metrics. With the help of these metrics, which assess both the quantity and quality of predictions, model performance can be understood in more detail. For tasks where the significance of false positives (FPs) or false negatives (FNs) varies independently, using precision and recall may not be sufficient [78].

The definition of precision and recall are shown in equations (1) and (2), respectively [**?**]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

The F1 score can be determined by calculating the precision and recall rates. F1 is the weighted harmonic average of recall and precision, as defined by eq 3:[78].

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

One basic metric used to assess the accuracy of an object detection algorithm is Average Precision (AP) [11][79] It determines the dataset's average precision-recall curve for each class. Benchmarking performance is facilitated by AP, which offers an overall evaluation of the algorithm's accuracy in identifying objects across various classes [58]. Due to its sensitivity to confidence threshold selection, AP might not be the best choice for tasks requiring exacting precision or recall.[78]

The definition of Average Precision are shown in eq 1[80, 81].

$$\text{AP} = \int_0^1 p(r)\, dr \tag{1}$$

In conventional object detection, the metric mean Average Precision (mAP) [7],[81] is widely used to evaluate performance in terms of regression and classification accuracies [77]. The model's overall performance is assessed using the mAP value, which is numerically equivalent to the average value of the AP sum across all categories [80]. It is sensitive to class imbalance and has the ability to mask subpar performance in particular classes. Furthermore, the computation of mAP is computationally expensive due to the complex calculations involved [78].

The definition is shown in Equation 5:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{2}$$

5 Conclusion and Future Work

This survey has examined the development of object detection methods designed for people with visual impairments, emphasizing the benefits and drawbacks of both conventional and deep learning approaches. While deep learning techniques have shown a lot of promise, particularly in terms of robustness and accuracy, real-time application and computational efficiency still need to be improved. In the future, research should focus on evaluating object detection models in terms of hardware requirements, particularly for mobile and wearable devices, to ensure they are practical for everyday use. Additionally, there should be a focus on improving the identification of common, daily objects to make assistive technologies more reliable and useful for visually impaired user.

Credit author statement

Shakeela Naz: Conceptualization,Methodology , Data curation, Visualization , Supervision, Software Validation.
Fouzia Jabeen: Experimental Evaluation , Validation, Critical Analysis.

Compliance with Ethical Standards:

It is declare that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] J. Y. Mambu, E. Anderson, A. Wahyudi, G. Keyeh, and B. Dajoh, "Blind reader: An object identification mobile-based application for the blind using augmented reality detection," *2019 1st Int. Conf. Cybern. Intell. Syst. ICORIS 2019*, vol. 1, no. August, pp. 138–141, 2019.

[2] M. Elgendy, C. Sik-Lanyi, and A. Kelemen, "A novel marker detection system for people with visual impairment using the improved tiny-yolov3 model," *Comput. Methods Programs Biomed.*, vol. 205, p. 106112, 2021.

[3] L. Jiao *et al.*, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[4] M. Bansal, M. Kumar, M. Kumar, and K. Kumar, "An efficient technique for object recognition using shi-tomasi corner detection algorithm," *Soft Comput.*, vol. 25, no. 6, pp. 4423–4432, 2021.

[5] M. Kadhim and B. Oleiwi, "Blind assistive system based on real time object recognition using machine learning," *Eng. Technol. J.*, vol. 40, no. 1, pp. 159–165, 2022.

[6] J. Kaur and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," *Multimed. Tools Appl.*, vol. 81, no. 27, pp. 38297–38351, 2022.

[7] A. N. E. al., "Study of deep learning techniques for differently abled applications and detection techniques," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 10, pp. 5817–5829, 2021.

[8] S. A. Sanchez, H. J. Romero, and A. D. Morales, "A review: Comparison of performance metrics of pretrained models for object detection using the tensorflow framework," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 844, no. 1, 2020.

[9] W. J. Chang, L. B. Chen, C. H. Hsu, J. H. Chen, T. C. Yang, and C. P. Lin, "Medglasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients," *IEEE Access*, vol. 8, pp. 17013–17024, 2020.

[10] R. Alex, "Object detection algorithms: A review," *pp. 5440–5442*, 2020.

[11] A. Lavric, C. Beguni, E. Zadobrischi, A. M. Căilean, and S. A. Avătămăniței, "A comprehensive survey on emerging assistive technologies for visually impaired persons: Lighting the path with visible light communications and artificial intelligence innovations," *Sensors*, vol. 24, no. 15, 2024.

[12] A. Khan and S. Khusro, "An insight into smartphone-based assistive solutions for visually impaired and blind people: Issues, challenges and opportunities," *Springer Berlin Heidelberg*, vol. 20, no. 2, 2021.

[13] S. E. E. Profile, "A survey on an intelligent system for persons with visual disabilities," *Aust. J. Eng. Innov. Technol.*, vol. November, pp. 97–118, 2021.

[14] Z. A. N. Hisham, M. A. Faudzi, A. A. Ghapar, and F. A. Rahim, "A systematic literature review of the mobile application for object recognition for visually impaired people," in *2020 8th Int. Conf. Inf. Technol. Multimedia, ICIMU 2020*, pp. 316–322, 2020.

[15] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, 2020.

[16] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, 2017.

[17] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019.

[18] A. Balothiya, A. Paliwal, and N. Rathi, "Object detection assistance for visually impaired people - a brief review," *vol. 10*, no. 2, pp. 15–20, 2022.

[19] M. Deshpande, M. V. Sreenath, S. Nadella, N. Antony, and N. Venkat, "A survey of object classification and detection techniques in assistance systems for the visually impaired," *Int. J. Eng. Manag. Res.*, vol. 1, no. 1, pp. 44–51, 2024.

[20] P. Umale, A. Patil, C. Sahani, A. Gedam, and K. Kawale, "Planer object detection using surf and sift method," *Int. J. Eng. Appl. Sci. Technol.*, vol. 6, no. 11, pp. 36–39, 2022.

[21] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, "An object tracking system based on sift and surf feature extraction methods," in *Proc. - 2015 18th Int. Conf. Network-Based Inf. Syst. NBiS 2015*, pp. 561–565, 2015.

[22] A. Agarwal, D. Samaiya, and K. K. Gupta, "A comparative study of sift and surf algorithms under different object and background conditions," in *Proc. - 2017 Int. Conf. Inf. Technol. ICIT 2017*, pp. 42–45, 2018.

[23] J. A. Gonzalez, A. L. Garcia, and C. S. Torres, "Computer vision for the visually impaired: A review," *Journal of Visual Impairment & Blindness*, vol. 115, no. 6, pp. 582–593, 2021.

[24] D. Rodriguez and D. J. K. Nunes, "Assistive technologies for blind and visually impaired people," *Springer International Publishing*, vol. 17, no. 2, 2019.

[25] F. Marchal, A. E. E. Bouna, and C. C. M. S. Barka, "Smartphone-based object recognition and navigation systems for the visually impaired: A systematic review," *IEEE Access*, vol. 7, pp. 89811–89824, 2019.

[26] S. Aravind, S. Balaji, and R. A. Vishal, "Development of object detection system for blind people using deep learning and neural networks," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 2423–2434, 2021.

[27] A. Hakim and A. K. El-Sayed, "Visual object recognition for the blind using color and depth information," in *Proc. 3rd Int. Conf. Technol. Adv. Comput. Electr. Eng.*, pp. 156–160, 2017.

[28] N. Tahiri, N. F. G. M. Daoud, and B. D. Bennani, "An adaptive object detection system for visually impaired people using smartphones," in *Proc. 2018 16th Int. Conf. Inf. Technol. New Generations*, pp. 367–373, 2018.

[29] Y. Chen, J. Wang, Y. Zheng, and Z. Yu, "A survey of object detection techniques for the visually impaired," *Journal of Optical Society of America A*, vol. 37, no. 7, pp. 1156–1166, 2020.

[30] X. Sun, Z. Zhang, and J. Li, "Computer vision-based object detection for the visually impaired: A review of applications and challenges," *Journal of Visual Communication and Image Representation*, vol. 72, 2023.

[31] S. Dariwal, V. Singh, and D. Verma, "Object detection for visually impaired using convolutional neural networks," in *2023 IEEE Int. Conf. Comput. Vision Comput. Imaging*, pp. 1594–1601, 2023.

[32] N. Desai and R. K. Gupta, "Development of an object recognition system for visually impaired individuals using ai and deep learning models," *Procedia Computer Science*, vol. 197, pp. 278–283, 2023.

[33] F. Barros, G. A. Silva, and E. S. Ferreira, "Integrating vision and touch: A technology for blind people to identify objects using smartphone cameras," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 6, pp. 472–481, 2020.

[34] A. Mohamed, T. Sharaf, and S. Al-Qatami, "Object recognition for the visually impaired using convolutional neural networks," *International Journal of Computing and Digital Systems*, vol. 10, no. 1, pp. 15–23, 2021.

[35] J. Jiang, Z. Li, and W. Zhuang, "Deep learning-based object detection for the visually impaired: A survey," in *Proceedings of the 2022 IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2324–2332, 2022.

[36] H. Huang and X. Yan, "Ai-assisted object detection for visually impaired people: Current trends and challenges," *Sensors*, vol. 23, no. 4, pp. 937–946, 2023.

[37] M. Zhu and L. Lu, "Assistive object detection for the visually impaired using convolutional neural networks," in *Proceedings of the 2021 IEEE International Conference on Image Processing*, pp. 1867–1871, 2021.

[38] A. Chowdhury, M. R. R. Basha, and A. Shafqat, "Deep learning-based object detection for visually impaired individuals: A comparative study," *Neural Computing and Applications*, vol. 33, no. 12, pp. 6693–6705, 2021.

[39] S. Balaji and P. Rajasekaran, "Object detection for visually impaired using deep learning models," in *Proceedings of the 2020 International Conference on Computational Intelligence*, pp. 275–279, 2020.

[40] P. Smith and B. Johnson, "Smartphone-based object recognition system for visually impaired people: A review of technologies and applications," *Journal of Visual Impairment & Blindness*, vol. 115, no. 4, pp. 290–301, 2021.

[41] V. Raghav, M. Kumar, and A. Sharma, "Detection and identification of objects for visually impaired using transfer learning techniques," in *Proceedings of the 2023 International Conference on Artificial Intelligence*, pp. 324–330, 2023.

[42] R. Bastos, M. Silva, and J. Santos, "Computer vision for the blind: Object detection using smartphone cameras," *Computer Vision and Image Understanding*, vol. 214, p. 103242, 2022.

[43] Y. He, X. Li, and Y. Zhang, "Object detection for visually impaired people using smartphone cameras," *Journal of Assistive Technologies*, vol. 15, no. 4, pp. 235–244, 2021.

[44] L. Li and Y. Wang, "Efficient object detection algorithms for visually impaired people," in *Proceedings of the 2020 International Conference on Artificial Intelligence and Robotics*, pp. 142–148, 2020.

[45] L. Liu, J. Yang, and K. Zhang, "Mobile-based object detection using deep learning for visually impaired people," *Applied Sciences*, vol. 13, no. 11, pp. 2761–2774, 2023.

[46] H. Wang and Y. Zhao, "Vision-based object detection for the visually impaired using convolutional neural networks," in *Proceedings of the 2022 IEEE International Conference on Robotics and Automation*, pp. 1345–1351, 2022.

[47] S. Park and H. Lim, "Visual object detection for visually impaired: A deep learning approach," *Neurocomputing*, vol. 426, pp. 70–79, 2021.

[48] X. Zhang and F. Wang, "Object detection and recognition for visually impaired people using smartphone cameras," *Journal of Visual Impairment & Blindness*, vol. 114, no. 3, pp. 236–245, 2020.

[49] J. Liu and T. Zhang, "Ai-based object detection for visually impaired people using rgb-d cameras," in *Proceedings of the 2023 IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2127–2134, 2023.

[50] S. Rodrigues and F. Pereira, "Object detection for visually impaired individuals using convolutional neural networks," *Machine Learning Applications*, vol. 8, no. 2, pp. 120–130, 2023.

[51] T. Chen and Z. Li, "Improving object recognition for visually impaired people with deep learning models," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1991–2002, 2021.

[52] Y. Tian and F. Zeng, "Object detection for visually impaired using transfer learning-based cnn models," in *Proceedings of the 2022 International Conference on Robotics and Automation*, pp. 1076–1082, 2022.

[53] J. A. Gonzalez, A. L. Garcia, and C. S. Torres, "Computer vision for the visually impaired: A review," *Journal of Visual Impairment & Blindness*, vol. 115, no. 6, pp. 582–593, 2021.

[54] D. Rodriguez and D. J. K. Nunes, "Assistive technologies for blind and visually impaired people," *Springer International Publishing*, vol. 17, no. 2, 2019.

[55] F. Marchal, A. E. E. Bouna, and C. C. M. S. Barka, "Smartphone-based object recognition and navigation systems for the visually impaired: A systematic review," *IEEE Access*, vol. 7, pp. 89811–89824, 2019.

[56] S. Aravind, S. Balaji, and R. A. Vishal, "Development of object detection system for blind people using deep learning and neural networks," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 2423–2434, 2021.

[57] A. Hakim and A. K. El-Sayed, "Visual object recognition for the blind using color and depth information," in *Proc. 3rd Int. Conf. Technol. Adv. Comput. Electr. Eng.*, pp. 156–160, 2017.

[58] N. Tahiri, N. F. G. M. Daoud, and B. D. Bennani, "An adaptive object detection system for visually impaired people using smartphones," in *Proc. 2018 16th Int. Conf. Inf. Technol. New Generations*, pp. 367–373, 2018.

[59] Y. Chen, J. Wang, Y. Zheng, and Z. Yu, "A survey of object detection techniques for the visually impaired," *Journal of Optical Society of America A*, vol. 37, no. 7, pp. 1156–1166, 2020.

[60] X. Sun, Z. Zhang, and J. Li, "Computer vision-based object detection for the visually impaired: A review of applications and challenges," *Journal of Visual Communication and Image Representation*, vol. 72, 2023.

[61] S. Dariwal, V. Singh, and D. Verma, "Object detection for visually impaired using convolutional neural networks," in *2023 IEEE Int. Conf. Comput. Vision Comput. Imaging*, pp. 1594–1601, 2023.

[62] N. Desai and R. K. Gupta, "Development of an object recognition system for visually impaired individuals using ai and deep learning models," *Procedia Computer Science*, vol. 197, pp. 278–283, 2023.

[63] M. Mohammad and N. Rahman, "Real-time object detection for visually impaired people in outdoor environments," in *Proceedings of the 2024 IEEE International Conference on Computer Vision*, pp. 211–218, 2024.

[64] S. Patel and D. Kumar, "Deep learning approaches for real-time object detection in assistive technologies," *Journal of Machine Learning and Artificial Intelligence*, vol. 31, no. 4, pp. 250–264, 2023.

[65] H. Cho and T. Lee, "Smart glasses for the visually impaired using object detection and augmented reality," in *Proceedings of the 2024 IEEE International Symposium on Augmented Reality*, pp. 55–62, 2024.

[66] P. Xu and G. Wu, "Adaptive object detection and recognition for visually impaired people with variable environmental conditions," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 6, pp. 1451–1463, 2024.

[67] R. Singh and S. Gupta, "Vision-based navigation for the blind using object detection and pathfinding algorithms," *International Journal of Human-Computer Interaction*, vol. 40, no. 7, pp. 477–491, 2024.

[68] Y. Liu and X. Zhang, "Integration of object detection and gps for assistive navigation systems for the blind," in *Proceedings of the 2024 International Conference on Assistive Technologies*, pp. 98–105, 2024.

[69] H. Wang and L. Zhou, "Deep learning for object detection and obstacle avoidance in visually impaired assistance systems," *Neural Networks*, vol. 140, pp. 52–63, 2024.

[70] J. Zhang and F. Li, "Real-time object recognition and navigation assistance for the visually impaired using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 110, p. 101234, 2024.

[71] A. Yadav and P. Jain, "Development of a wearable object detection system for visually impaired using deep neural networks," in *Proceedings of the 2024 IEEE Wearable Technology Conference*, pp. 210–215, 2024.

[72] J. Park and S. Choi, "Vision-based object detection and recognition for the visually impaired using deep convolutional neural networks," *Journal of Vision and Technology*, vol. 40, no. 2, pp. 112–123, 2024.

[73] H. Li, B. Singh, M. Najibi, Z. Wu, and L. S. Davis, "An analysis of pre-training on object detection," *arXiv*, 2019.

[74] A. Kuznetsova *et al.*, "The open images dataset v4 unified image classification, object detection, and visual relationship detection at scale," *Int. J. Comput. Vis.*, vol. 128, no. 7, pp. 1956–1981, 2020.

[75] N. Joshi, S. Maurya, and S. Jain, "Real-time object detection and identification for visually challenged people using mobile platform," *CEUR Workshop Proc.*, vol. 2823, pp. 47–59, 2021.

[76] J. Miao and W. Zhu, "Precision - recall curve (prc) classification trees," *Evol. Intell.*, vol. 15, no. 3, pp. 1545–1569, 2022.

[77] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics, and methods," *Applied Sciences*, 2020.

[78] Y. Kniazieva, "Object detection: Key metrics for computer vision model performance." https://labelyourdata.com/articles/object-detection-metrics. Accessed: Sep. 14, 2024.

[79] Q. Fan, "Few-shot object detection with attention-rpn and multi-relation detector," pp. 4013–4022, n.d.

[80] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of retinanet, ssd, and yolo v3 for real-time pill identification," *BMC Med. Inform. Decis. Mak.*, vol. 8, pp. 1–11, 2021.

[81] Z. Zhou, X. Lan, S. Li, C. Zhu, and H. Chang, "Feature pyramid ssd: Outdoor object detection algorithm for blind people," pp. 650–654, 2019.

[82] G. H. Fu, L. Z. Yi, and J. Pan, "Tuning model parameters in class-imbalanced learning with precision-recall curve," *Biometrical J.*, vol. 61, no. 3, pp. 652–664, 2019.