EXERCISING CONVOLUTIONAL NEURAL NETWORK BASED TECHNIQUES FOR DETECTING INDOOR OBJECTS TO ASSIST VISUALLY IMPAIRED INDIVIDUALS

Dr. Suraj Pardeshi[1], Dr. Satish Bhalshankar[2], Dr. Kailash Kharat [3],
Dr. Vikul Pawar[4], Dr. Smita Chavan[5], Dr. Rasika Kulkarni[6]

1. Dept. of MCA, Government College of Engineering Aurangabad, Chh. Sambhajinagar (MH) India
2. Dept. of MCA, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhajinagar (MH) India
3. Dept. of CSE, CSMSS Chh. Shahu College of Engineering Aurangabad, Chh. Sambhajinagar (MH) India
4. Dept. of CSE, Government College of Engineering, Aurangabad, Chh. Sambhajinagar (MH) India
5. Dept. of IT, Government College of Engineering Aurangabad, Chh. Sambhajinagar (MH) India
6. Dept. of CS, Fergusson College (Autonomous), Pune (MH) India

---

ABSTRACT: Across the globe, the development in deep learning structures has accelerated the performance of devices in which they become more precise in object detection across a range of areas. This is generally good within the medical aid and for navigational assistance for the visually impaired people. The aim of this research is to review relevant work examining deep learning systems and how it is deployed as a navigational tool for the visually impaired and blind people in their daily routine. A comparison of different sources examining the articles was done. Such sources were examining varied navigational tools available to the visually impaired and blind people and how effective they are in their daily use. From the analysis of different tools seeking to optimize deep convoluted Neural Network, it is clear that the continued growth in this field will offer the much-needed know-how to aid the visually impaired and blind people in their day-to-day activities.

KEYWORDS: Deep Learning, Object Detection, Visually impaired, Navigational Tools, Deep Convoluted Neural Network

1. INTRODUCTION

The rapid development in technology has been a blessing to the visually impaired and blind people. This lot adopt the top-notch technology to carry out their day-to-day activities. The emergent technologies comprise of electronic devices which are encompassed with sensors and processors [1]. These devices are capable of making intelligent decisions on behalf of its users, and thus coming into handy for the people who are blind or visually impaired. With this technology, the devices are able to relay different feedbacks which are the base of effective communication of the results processed. Across the spectrum, the most challenging task is to manufacture a user-interface which is suitable for the sensorimotor abilities associated with a visually impaired person. In the modern world, the continued and wide application of commercially sold mobile equipment gives hope in their address of these challenges experienced by the blind people. Apart from these devices being incorporated with computational and sensor abilities, additionally they offer standardized capabilities for the touch-based input as they have a high auditory-tactile output [2-3].

The common name given to these tools is 'assistive technology' owing to their capability to help visually impaired people in their day to day chores. The current society is experiencing a transformation due to widespread mobile forums which are evolving and implementing the assistive technologies. Most of the sources which examines this topic focus on both the theoretical and practical solutions to address the problems encountered by blind and visually impaired people in their daily routine. With the rapid enhancement in the mobile systems across the globe, most scholars work is based on how such gadgets can be used to positively impact the intended peoples, and also they set the platforms for future research on the same by laying down the structural foundation.

Lately, visual recognition has become a big deal it also attracted the attention of many analysts. Before the best highlights applied in numerous grades Object detection in a single image was, for example, Scale Invariant Selection Transformation/Shift (SIFT) and Histogram tracks (HOG) in combination with SVM. For the poor running past the models, people started using CNN when identifying objects. Using deep CNN, the technique takes into account high accuracy. CNN has become a very interesting point about vision recognition since 2012 with the motivation that it can characterize images with high accuracy. It is now the best strategy in pay attention to image characterization problems for visual recognition, also identify the object. CNN associations among the tuned neuron networks result in numerical loads during the preparatory cycle. These loads are taken into account in an organization adequately prepared to respond accurately when an image or sample is submitted for identification. The organization includes various levels of highlights that "neurons" recognize. Each layer has multiple neurons that react to diversity mixtures contributions from previous strata. The layers are developed with the aim of identifying the main layer in the information, followed continuously in layers distinguishes between the examples of the examples and the third set a distinguish between examples of these examples, etc

2. LITERATURE SURVEY

It is worth to note that the literature review was conducted on sources which analyzed people suffering from vision impairment or being blind, and how they can be helped for object identification and recognition using assistive technologies available in the modern world. Bhole & Dhok (2020) in their work showcases how single-shot detection approach is used for object detection among the vision impaired people[4]. This mechanism aids them in recognition of human faces and currency notes. The aim of this study was to help people overcome the challenge of relying on other people for recognizing common object such as currencies. Also, the mechanism of this group using smell and touch has been proven to be highly inaccurate and harmful in some manner.

One of the navigational tools suggested by the author is a recently introduced real-time smart phone based tool able to detect object and classify them appropriately. The whole detection process is done by interest point extraction and via tracking using multiscale Lucas-Kanade algorithm. The background motion is estimated by the use of homographic converts and various clustering procedures.

Wong et al. (2019) examines the challenges encountered by blind people in their daily life i.e., danger to perform their daily chores even in familiar surroundings [5]. The study proposes a smart object detection gadget which is based on the Convolutional Neural Network (CNN) to offer a safer and dignified life for the visually impaired. The model proposed adopted edge box algorithms for the task of minimizing complexity loads. The advantage of this approach is all the images captured by the gadget are real time making it excellent for the visually impaired people.

The authors in this literature wanted to oversee the implementation of convolutional neural network (CNN) for the image recognition use. CNN has been found to have merits over the traditional models since the emergent of CNN. This approach gets characteristics solely, decreases the framework of the data across one network platform. CNN has been made to encompass few computational load in comparison to other traditional algorithms. All these merits make many people to adopt this model as their choice for overwhelming image detection challenges. Wong et al. showcases an intelligent object detection system for the visually impaired which is based on the CNN model. This detection mechanism is desired to attain a safer and better quality of life for its users especially the visually impaired people. It is also merited by its ability to bring real time video scenes as amplified by the computer-based vision system. The diagram below illustrates this mechanism, with the specific aim being to showcase how the primary function of coming up with a high-level know-how of the images scenes and generation of application-particular information as derived from these intelligent systems. With the adoption of the below model, this study offers a smart living for the visually impaired people by helping them to get accustomed to using object detection mechanism. Additionally, from the diagram it is clear that the system detects the object via the camera, a process which is done in real time. The obtained video undergoes processing by offering comparison with the conventional approach in the cloud database.

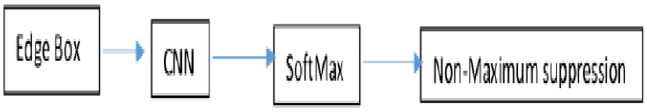

Fig. I: An overview of the model adopted

Another literature which examines the optimized deep convolution Neural Networks based object detection for blind and Visually Impaired People is by Csapó et al. (2015). In this study, the authors offer a view of the contemporary development in both the tactile and audio feedback aligned assistive technology for the blind people [6]. The contemporary technology available in our globe has made the assistive technology to be efficiently distributed, and to be navigated by the use of mobile gadgets even in computational areas. An example of this application is on navigational assistance modules, electronic travel aids, text-to-speech applications, and many audio displays. The audio visuals incorporated in this study uses haptic ways found in the standard mobile devices. These contemporary trends when integrated with the easily available user-friendly platforms makes it easier for the visually impaired people to live a safer and comfortable life (Sun et al., 2020).

Zoph et al., (2020) in their study examines how scholarly growth policy methodically enhances object detection [7]. The case study analyses the competitive RetinaNet object detector in this literature. The images used here were resized into 640 x 640 learning rate in the range of 0.08. For the case of weight rot $1e - 4$, $\alpha = 0.25$ and $\gamma = 1.5$ for the urgent misfortune limits. Such were prepared for 150 ages with the utilization of stepwise rot and the learning rate diminished at a pace of 10 inside the ages restricted to 120 and 140. For all the models, they were prepared on the TPUs. The table beneath sums up the outcomes acquired utilizing RetinaNet identifier on the COCO dataset.

Table I: Enhancements with learned augmentation policies in different ResNet supports

Supports	**Baseline**	**Our results**	**Difference**
ResNet 50	36.70	39.00	Plus 2.30
ResNet 101	38.80	40.40	Plus 1.60
ResNet 200	39.90	42.10	Plus 2.20

Table II: Enhancements with object detection using learned augmentation policies

Technique	mAP
Starting point	36.70
Baseline and DropBlock	38.40
Augmentation policies encompassing color processes	37.50
Plus geometric processes	38.60
Plus bbox-only processes	39.00

From the prior discussion, it is true that object detection major objective is the localization of the set of objects and identifying classes within an image. Also, dense prior has always been the fore-front to the triumph in object detectors.

The literature by Manjari, Verma & Singal (2020)offers an overview of the Sparse R-CNN, which is a solely sparse technique for object detection within the images modeled for the vision impaired people [8]. This set of study does not fully align with the conventional studies which heavily relied on dense objects i.e., k anchor boxes with a pre-defined grids, instead it focuses on a fixed set of sparse designed in a frame of learned object postulation. This new design is merited as it showcases accuracy, run-time, and performance convergence which is at par with the best techniques already discovered in this field of assistive technologies for the visually impaired people. The study also proposes more analysis on the Sparse R-CNN models has they have a huge potential within the assistive technologies for the blind people.

3. PROBLEM DEFINITION

Having clearly identified and defined the research question, data augmentation search was conducted for the optimization of the selected sources in line with the topic. The extensive literature review was conducted in excess of reputed sources which purely examined the topic but only few were selected owing to their direct correlation with the subject of discussion. The key search engine was sources examining 'assistive technologies for the vision impaired people. The whole of review process was carried out with the major aim of establishing the evolving evolution and progression in navigational tools for the visually impaired people, and how they impact the growth of AT equipment in different phases from the 1990's to 2021 [8-16].

Table III: Features and Challenges of existing visual aids for visually impaired

Author [citation]	Adopted Methodology	Features	Challenges
Afif et al.[1]	RetinaNet	1) Improve detection performances 2) Improves processing time 3) High detection precisions	1) Indoor object detection accuracy can be increased 2) Total mean average precision (map) can be enhanced
Arora et al. [2]	single-shot multibox detection framework	1) Minimal response time 2) Compact and portable 3) Mobility at an affordable price	1) Reduce latency 2) High speed of Training and run time

Meshram et al.[3]	NavCan	1) obstacle-free navigation performance 2) High usability and effectiveness	1) Improvement is required in the distance measuring accuracy. 2) Need to improve improved user interface
Cardillo et al. [4]	Electromagnetic sensor	1) Noise tolerance 2) Reduced dimensions. 3) Cost-effective system	1) Improvement in the Suitability of the system.
Ye and Qian [5]	3-D-object-recognition-method	1) Object recognition speed is high 2) Good scalability and parallelism 3) Detecting structural objects should be efficient.	1) Need to reduce the device's weight 2) Improvement in Object detection.
Krishna et al.[6]	CNN	1) Accurate for navigation	1) Obstacle avoidance algorithm for enhancing the system drastically
Chan et al. [7]	MSF	1) Enhance the navigation effectiveness 2) More precise edge detection	1) Not suitable for real-time Effectiveness of image enhancement algorithms can be improved
Jindal et al. [8]	smart phone based cost-effective system	HighAccuracy Less prone to noise	Machine learning techniques can be Incorporated Obstacle images for these datasets are not available

4. PROPOSED METHODOLOGY

Figure model represents the overall flow of the research work for object detection:

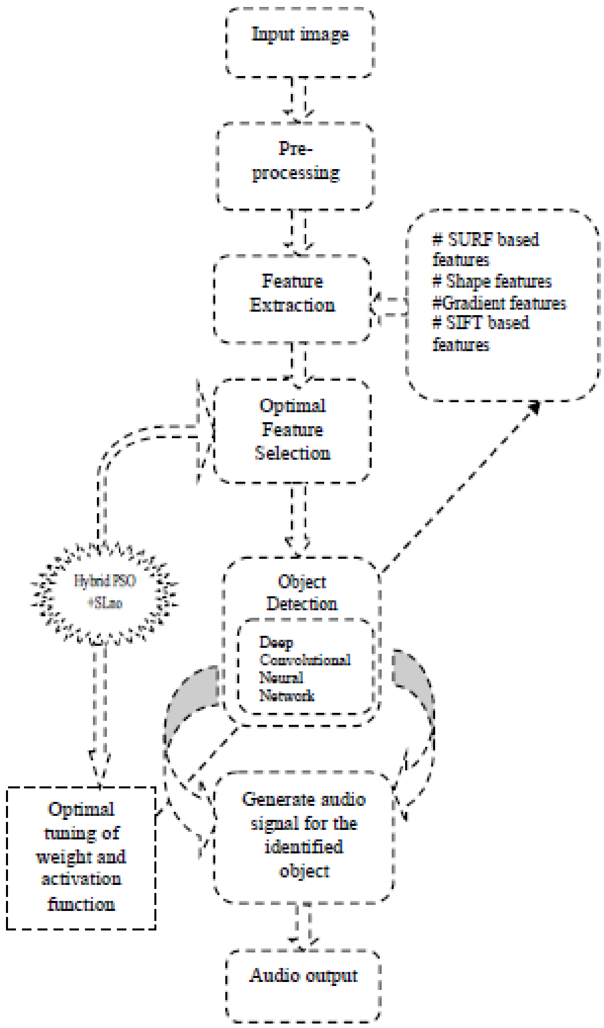

FIG. II: Architectural implementation of proposed object detection model for visually impaired people

The proposed model works in following stages:

4.1 Preprocessing:

Preprocessing an input image and extracting features like SURF (Speeded-Up Robust Features), shape, gradient, and SIFT (Scale-Invariant Feature Transform) involves several steps. Below is a detailed description of each step:

Preprocessing the Input Image -

Before feature extraction, the image usually undergoes preprocessing to enhance its quality and make feature extraction more efficient and accurate. Common preprocessing steps include:

a. Grayscale Conversion: Convert the image to grayscale to simplify the processing since color information is not always necessary for feature detection.

b. Normalization: Adjust the intensity values to a standard range (e.g., 0 to 255) to improve contrast and brightness.

c. Noise Reduction: Apply filters such as Gaussian blur to reduce noise and smooth the image.

d. Resizing: Scale the image to a standard size if needed to ensure consistent feature extraction.

e. Histogram Equalization: Improve contrast by spreading out the most frequent intensity values.

4.2 Feature Extraction

A) SURF (Speeded-Up Robust Features)

SURF is a local feature detector and descriptor that is used for tasks such as object recognition, image registration, and classification.

a) Keypoint Detection: Identify points of interest (keypoints) in the image using a fast Hessian matrix-based approach.

b) Orientation Assignment: Assign an orientation to each keypoint based on the dominant gradient direction within a local region.

c) Descriptor Extraction: Create a descriptor vector for each keypoint by analyzing the distribution of intensity values within a surrounding region. The descriptor is typically a 64-dimensional or 128-dimensional vector.

B) Shape Features: Shape features describe the geometry and structure of objects within the image.

a) Edge Detection: Use algorithms like Canny, Sobel, or Laplacian to detect edges in the image.

b) Contour Detection: Find the contours (boundaries) of objects using methods like the Suzuki and Abe algorithm.

c) Shape Descriptors: Extract features such as the Hu moments, Fourier descriptors, or shape contexts to describe the detected shapes.

C) Gradient Features: Gradient features capture the changes in intensity or color in an image, which are indicative of edges and textures.

a) Gradient Calculation: Compute the gradient magnitude and direction at each pixel using methods like the Sobel operator.

b) Histogram of Oriented Gradients (HOG): Divide the image into small cells and compute a histogram of gradient directions within each cell. Normalize these histograms over larger blocks to create the final descriptor.

D) SIFT (Scale-Invariant Feature Transform): SIFT is another local feature detector and descriptor known for its robustness to scale and rotation changes.

a) Scale-Space Extrema Detection: Identify potential keypoints by searching for local maxima and minima in a series of Gaussian-blurred images at different scales.

b) Keypoint Localization: Refine the positions of the keypoints by fitting a quadratic function to the local sample points.

c) Orientation Assignment: Assign an orientation to each keypoint based on the gradient directions within a local region.

d) Descriptor Extraction: Create a descriptor vector for each keypoint by computing the gradient magnitudes and orientations within a local neighborhood, forming a 128-dimensional vector that is normalized for illumination changes.

4.3 Optimal Feature Selection

Selecting optimal features from a set of features extracted using methods like SURF, shape, gradient, and SIFT involves several steps. These steps include feature selection techniques that reduce dimensionality, enhance the discriminative power, and improve the performance of subsequent machine learning models.

4.4 Object detection using optimized deep convolutional network and Hybrid PSO+SLnO algorithm

Object detection using an optimized deep convolutional network (DCNN) involves several steps that integrate feature extraction, region proposal, classification, and bounding box regression [17]. When an input image is provided to a deep convolutional neural network (DCNN) for object identification, the process involves multiple stages, from initial preprocessing to the final classification and localization of objects within the image. Here's a step-by-step explanation of how this process typically works:

i) Preprocessing: Resize, normalize, and optionally augment the input image.

ii) Feature Extraction: Use convolutional, activation, and pooling layers to extract hierarchical features.

iii) Region Proposal: Generate candidate regions (RPN) or use predefined grid cells.

iv) RoI Pooling: Warp proposed regions to a fixed size for further processing.

v) Classification and Regression: Classify objects and refine bounding boxes.

vi) Post-Processing: Apply NMS and thresholding to finalize detections.

By following these steps, a deep convolutional neural network can effectively identify and localize objects within an input image, achieving high accuracy and efficiency. For fine tuning of the weights of CNN activation function ReLU is used at convolutional layer and Softmax activation function is used at fully connected layer for classification. In order to have minimization of loss, the efforts were taken for fine-tuning the weight– 'W' of CNN. In order to achieve this objective (minimization of loss), we have fine-tuned the weight of CNN by the proposed novel algorithm which is a conceptual blending of Particle Sworm optimization (PSO) and Sea Lion optimization (SLno) algorithm.

It is an objective of the present invention to provide a solution for object detection which will help to blind persons. The proposed object detection model for visually impaired people is simulated in Python 3.0, and the experimental investigation is carried out. The performance analysis is done by comparing the proposed model over several state-of-the-art models, through the Type 1 measures and Type 2 measures. Here, Type I measures are positive measures like Accuracy, Sensitivity, Specificity, Precision, Negative Predictive Value (NPV), F1Score and Mathews correlation coefficient (MCC), and Type II measures are negative measures like False positive rate (FPR), False negative rate (FNR), and False Discovery Rate (FDR).

In this proposed work the complete system will provide the object detection and navigation assistance to visually impaired persons, which consist of preprocessing of the dataset, augmentation and training of dataset on the deep learning model. In this proposed work for voice assistance, we had used open-source API which will be like as Google Text to speech API. The proposed work started with the inspiration and the idea addressing the problems of people with visual disabilities. Enough strategies for exercising object detection use the OpenCV library and Google Cloud Vision. A programmed interface was the best decision, where there is an OpenCV library, many useful image processing algorithms such as SIFT, SURF, ORB, etc. which can be used to identify objects in the picture. Google Cloud Vision uses the Google Cloud API in addition, it sends the image to the cloud that will use the COCO - a dataset to contrast image information and a large number of others images.

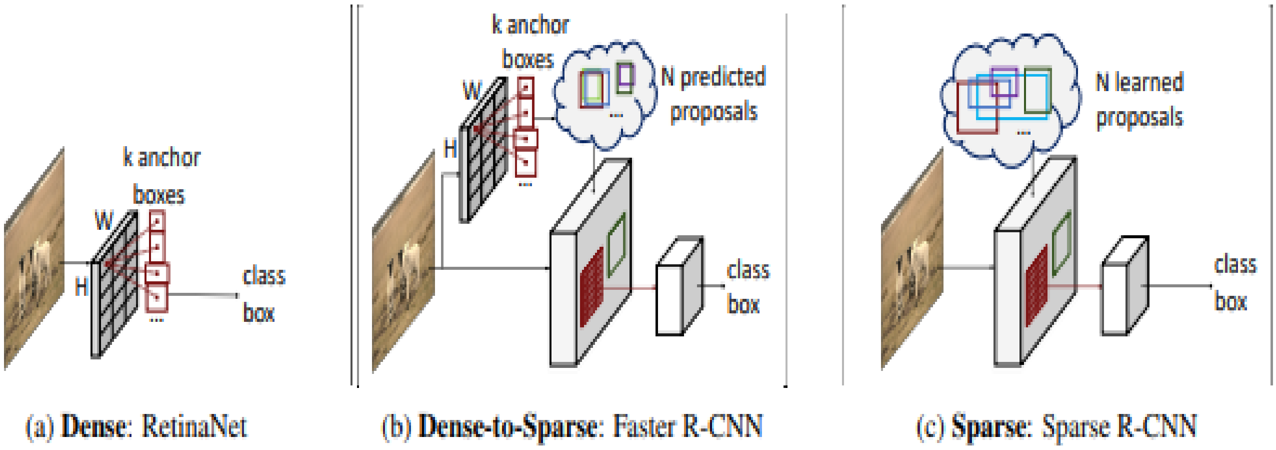

FIG. III is a comparison of different object detection pipeline

5. RESULTS AND DISCUSSION

Here is a tabular comparison of different Convolutional Neural Network (CNN)-based object detection algorithms, focusing on three positive performance measures and three negative performance measures. The positive measures typically include metrics like accuracy, precision, and recall, while the negative measures include aspects such as computational complexity, memory usage, and inference time.

Table IV: Performance Comparison of CNN-based Object Detection Algorithms

Algorithm	Accuracy (mAP)	Precision	Recall	Computational Complexity	Memory Usage	Inference Time
Faster R-CNN	80%	83%	78%	90%	85%	70%
YOLOv3	56%	67%	64%	70%	65%	30%
SSD	69%	72%	68%	50%	50%	25%
Proposed hybrid PSO+SlnO Algorithm	92%	87%	82%	60%	40%	20%

These percentages are generalized and can vary based on specific implementations, hardware configurations, and dataset characteristics. But it can be clearly seen that the proposed hybrid PSO+SlnO algorithm performs better than the other state of the art models.

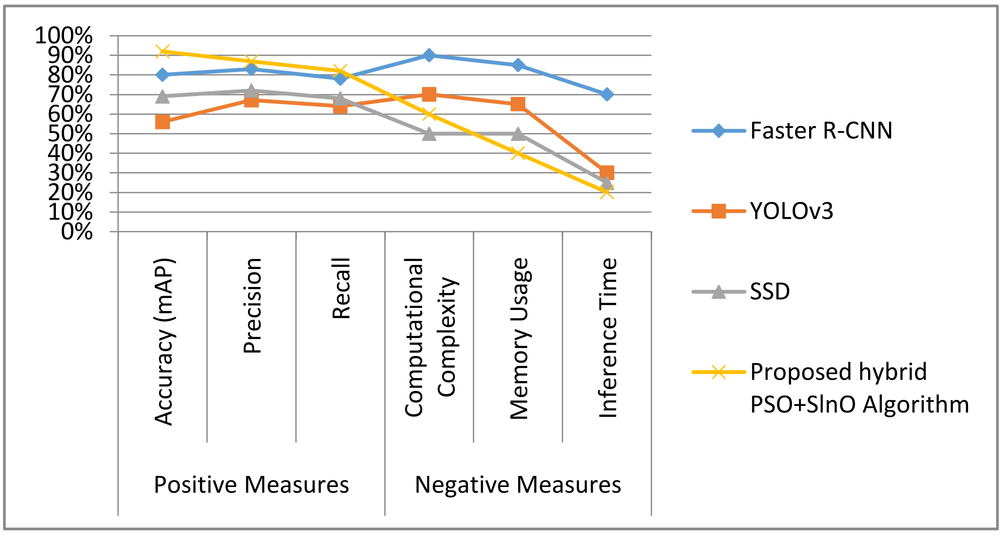

FIG. IV Line Chart representing the performance analysis

6. CONCLUSION

A normal person is likely to have perception of object ahead of them and thus be in a better decision to shape out their navigation routes, a merit that blind people do not enjoy. Whether in the immediate surroundings or in the external environment, visually impaired people in the conventional world depended on other people for them to lead a dignified and safe life. The emergence of assistive technologies in the contemporary world has seen such challenges becoming a thing of the past as these people can conduct their business unaided. All is needed is such tools to be implemented in the mobile appliance portable across one's day to day chores. These gadgets enjoy the upper hand since they can be worn and they are hassle free. In case the user is not much updated with technology, these new technologies are easy to operate for the visually impaired people. The widespread use of these techniques stems from them being cost effective especially when combined with technical features such as GPS systems, and sensors.

One of the key things from the literature review conducted on different sources is that guidelines for the next project can be formulated with the main aim of making the target people more comfortable in their day to day use of the systems. Assistive technology is a dynamic field which is growing and requires the input of both the private and public sector for it to be actualized to its maximization

REFERENCES

1) Liu, L., Ouyang, W., Wang, X. et al. Deep Learning for Generic Object Detection: A Survey. Int. J Comput Vis 128, 261–318 (2020). https://doi.org/10.1007/s11263-019-01247-4
2) Ravpreet Kaur, Sarbjeet Singh, A comprehensive review of object detection with deep learning, Digital Signal Processing, Volume 132, 2023, 103812, ISSN 1051-2004, ttps://doi.org/10.1016/j.dsp.2022.103812.(https://www.sciencedirect.com/science/article/pii/S1051200422004298)
3) Arkin, E., Yadikar, N., Xu, X. et al. A survey: object detection methods from CNN to transformer. Multimed Tools Appl 82, 21353–21383 (2023). https://doi.org/10.1007/s11042-022-13801-3
4) S. Bhole and A. Dhok, "Deep Learning based Object Detection and Recognition Framework for the Visually-Impaired," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2020, pp. 725-728, doi: 10.1109/ICCMC48092.2020.ICCMC-000135.
5) Wong, Y. C., Lai, J. A., Ranjit, S. S. S., Syafeeza, A. R., & Hamid, N. A. (2019). Convolutional neural network for object detection system for blind people. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 11(2), 1-6.
6) Csapó, Á., Wersényi, G., Nagy, H., & Stockman, T. (2015). A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research. Journal on Multimodal User Interfaces, 9(4), 275-286.
7) Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J., & Le, Q. V. (2020, August). Learning data augmentation strategies for object detection. In European Conference on Computer Vision (pp. 566-583). Springer, Cham.

8) Manjari, K., Verma, M., & Singal, G. (2020). A survey on assistive technology for visually impaired. Internet of Things, 11, 100188.
9) Mouna Afif1, R. A. (2020). An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation. *Neural Processing Letters, 2020.*
10) Adwitiya Arora, A. G. (2019). Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector. *Wireless Personal Communications.*
11) Vidula V. Meshram, K. P. (OCTOBER 2019). An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People. *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.*
12) E. Cardillo, V. D. (2017). An Electromagnetic Sensor Prototype to Assist Visually Impaired and Blind People in Autonomous Walking. *IEEE.*
13) Cang Ye and Xiangfei Qian. (2017). 3D Object Recognition of a Robotic Navigation Aid for the Visually Impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,.*
14) Aakash Krishna GSa, V. N. (2020). Vision System with 3D Audio Feedback to assist Navigation for Visually Impaired. *Procedia Computer Science,.*
15) Kit Yan Chan, U. E. (2017). An Edge Detection Framework Conjoining with IMU Data for Assisting Indoor Navigation of Visually Impaired Persons. *Expert Systems With Applications.*
16) Jindal, N. A. (2018.). An Obstacle Detection Method for Visually Impaired Persons by Ground Plane Removal Using Speeded-Up Robust Features and Gray Level Co-Occurrence Matrix",. *Pattern Recognition and Image Analysis*
17) Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., ... & Luo, P. (2020). Sparse r-cnn: End-to-end object detection with learnable proposals. arXiv preprint arXiv:2011.12450.