Volume: 13 Issue: VII Month of publication: July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73345

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

Vision Assist System based on Deep Learning Using Object Detection and Text Recognition with Voice Alert

P. Dinesh Kumar[1], D.Sathiya[2]
[1]Assistant Professor, [2]Student, Department of Computer Science and Engineering, Vivekanandha college of Technology for women Tiruchengode

***Abstract:** For blind people, recognizing a product in their daily routine is critical, so we proposed a method to identify products in their daily routine using a camera. A mobility method is employed to spot item of worry from the photo by advising person about recognized objects in order to separate an object from unnecessary background. The current project's goal is to model an object detector for visually impaired people and other commercial purposes by recognizing objects at a specific distance. Old techniques for object recognition required a large amount of training data, which takes more time and is quite complicated. Detection is employed in a variety of situations. Traditional methods of object detection rely on a large number of datasets, and training these datasets takes a long time. Training tiny or unnoticed objects is a more difficult task. People's minds and visual systems detect objects more accurately and quickly in real time, and they have conscious thoughts when detecting obstacles. Because of the abundance of data available, in addition to more advanced technologies and better working algorithms, classification and detection of multiple objects within the same frame has become simple and accurate. The project's main objective is to develop and implement a real-time object recognition system using a real-time camera. We can use Python as the front end to implement the system in real time. The experimental results show that the proposed system improves product identification accuracy.*

***INDEX TERMS**-object detection, text identification, OCR Technology, CNN Algorithm*

I. INTRODUCTION

The statistics of the World Health Organization (WHO) shows that the number of visually impaired is growing day by day. On average, the number of visually impaired is 285 million of whom 39 million are sightless and the rest 217 million are suffering from low vision. To keep doing their daily tasks, vision-impaired people usually seek help from others. And the most noteworthy part is when they explore a new place, they should understand the barriers' position and other objects in their course for their safe navigation. Secure and safe mobility is one of the most demanding events faced by vision-impaired people in the real-life environment. Being unable to track out and avoid blockage in their course, most often they become the victim of some unwanted troubles that might lead them to emotive misery or unasked incidents and their frequent mobility is being undercut need assistance from others or assistive devices to complete their day-to-day tasks including uninterrupted navigation and so on. However, ensuring secure and safe mobility for the visually impaired is a complex task that requires precision and effectiveness. One of the other serious issues, that are being faced by the visually impaired, is to recognize currency because of the likeness of paper surface and size among various classes. At the same time, visually impaired people are facing serious issues with newly released notes' sizes and colors. For example, the new 50 BDT and 200 BDT notes have identical colors, making it difficult for persons with low vision to identify and make appropriate transactions. This problem makes a great suffer in their daily activities when they deal with currencies. Identifying staircases is another matter of concern for the visually impaired since the failure to identify them can cause serious damages. Without seeing the stair, it is quite impossible to perfectly identify the edges of every plate of a stair. Among other problems, faced by people with visual disabilities, recognizing washrooms, chairs, tables, persons, etc. are mentionable.

The Blind Assistant System is a comprehensive and transformative solution aimed at enhancing the daily lives of visually impaired individuals, giving them greater autonomy in various aspects of life. By combining deep learning algorithms, computer vision, and speech synthesis technologies, this system offers a versatile, all-in-one assistive device that helps visually impaired people identify products, read text, navigate spaces, and interact with their surroundings more confidently.

The Blind Assistant System aims to significantly improve the independence and quality of life for visually impaired individuals by utilizing advanced technologies like deep learning, computer vision, and natural language processing. Its primary objective is to empower users to perform daily tasks without relying on external help. The system facilitates real-time product identification through the use of Convolutional Neural Networks (CNNs), allowing users to identify household items, groceries, and medications using a camera or wearable device. Additionally, it incorporates Optical Character Recognition (OCR) and CNN algorithm with neural networks to convert printed text into speech, enabling users to read labels, street signs, and books independently. Overall, the Blind Assistant System is designed to offer a comprehensive, intuitive, and secure solution that helps visually impaired individuals navigate the world with greater autonomy, safety, and confidence.

II. RELATED WORK

Fahad ashiq, et.al,…[1] presented a smart system that performs real-time object localization and recognition. As soon as the system recognizes the object, it sends audio feedback to the user. For example, after identifying a known object (e.g., car), the user will hear the word "car." Moreover, the user's location and a snapshot of the most recent viewed scene are periodically stored at a server that can be accessed by the family members using an application to track the user. For object detection and recognition, mobileNet architecture is used because of its low computational complexity to run on low power end devices. Since wearable hardware resources are limited, and the system's feedback about the object's name needs to be as close as possible, complex state-of-the-art object recognition approaches might not be viable as the first-choice techniques. The main objective of this study is to develop a system for VIPs that contains the following features: It performs real-time object detection and recognition using a deep learning framework. It pronounces names of the objects that can be seen through the camera eye i.e., objects present in the current frame. It periodically sends the location of the user to a web server. It sends live feed/snapshots to the webserver. It provides a web-based interface that the family members of VIPs can use to track their movement while being at their homes. It is an on-demand feature that provides security and ensures the user's privacy. This paper presented a smart and intelligent system for VIPs to assist them in mobility and ensure their safety. The proposed system is based on the day-to-day requirements of VIPs. It assists them in visualizing the environment and providing a sense of the surroundings. They can recognize objects around them and sense the natural environment using CNN-based low-power Mobile-Net architecture. Moreover, a web-based application is developed to ensure the safety of VIPs. The user of this application can turn the on-demand function to share his/her location with the family. It is beneficial to their family as they can monitor the movement of VIPs and can track his/her location using the live feed from the camera. The experimental analysis shows that the proposed system provided satisfactory results and outperformed other devices in terms of supported features.

Wenguan wang, et.al,…[2] presented, to the best of our knowledge, the first comprehensive review of SOD with focus on deep learning techniques. We first carefully review and organize deep learning-based SOD models from several different perspectives, including network architecture, level of supervision, etc. We then summarize popular SOD datasets and evaluation criteria, and compile a thorough performance benchmarking of major SOD methods. Next, we investigate several previously under-explored issues with novel efforts on benchmarking and baselines. In particular, we perform attribute-based performance analysis by compiling and annotating a new dataset and testing several representative SOD algorithms. We also study the robustness of SOD methods w.r.t. various input perturbations. Moreover, for the first time in SOD, we investigate the robustness and transferability of deep SOD models w.r.t. adversarial attacks. In addition, we assess the generalization and hardness of existing SOD datasets through crossdataset generalization experiment. We finally look through several open issues and challenges of SOD in deep learning era, and provide insightful discussions on possible research directions in future. As an important problem in computer vision, salient object detection (SOD) from images has been attracting an increasing amount of research effort over the years. Recent advances in SOD, not surprisingly, are dominantly led by deep learning-based solutions (named deep SOD) and reflected by hundreds of published papers. To facilitate the in-depth understanding of deep SODs, in this paper we provide a comprehensive survey covering various aspects ranging from algorithm taxonomy to unsolved open issues. In particular, we first review deep SOD algorithms from different perspectives including network architecture, level of supervision, learning paradigm and object/instance level detection. Following that, we summarize existing SOD evaluation datasets and metrics. Then, we carefully compile a thorough benchmark results of SOD methods based on previous work, and provide detailed analysis of the comparison results. Moreover, we study the performance of SOD algorithms under different attributes, which have been barely explored previously, by constructing a novel SOD dataset with rich attribute annotations

Mehul mahrishi, et.al,...[3]established a basis for indexing digital video through the YOLO V4 Darknet Neural Network. Detection accuracy, Precision, Recall, and F1 Score are used to reflect overall performance and outcomes.

The threshold values are based on experimental observations of multiple videos. As this is a new approach, much work can be done in the future. For instance, an adaptive threshold value can be used instead, and the performance is still abysmal in the case of handwritten text in the videos. The framework used a basic binary search algorithm infused with Structural Similarity Index. The dataset used for training the neural network is made from 6000 video frames, and for testing, a suite of 50 videos was randomly selected. It is important to note that the manual annotation used for training has limitations. For example, during the random data collection to create a training dataset, one cannot always mention the annotators' context, accuracy, and qualification. However, to confine the scope of the study to a specific domain, the experiments were structured to include only participants with an educational context and to cover videos from educational portals. Furthermore, manual indexing is only used for evaluation and not for the generation of index points. As a result, our subjective assessment of the proposed algorithm using three different types of experiments is sufficient to determine its efficacy. The present study tackled some aspects; nevertheless, more in-depth optimization and benchmarking are necessary to study the effect on the efficiency of index point detection for standardization and relevant summaries. In the future, we shall explore the applicability of this technique and any potentially needed extensions. This research discovered that the frequency of words, n-grams, and the number of first-time words that appeared in a video provide valuable information for video segmentation by topic. It is therefore expected to do a long-term, real-world effect analysis to evaluate keywords in instructional videos. The research findings can also be improved by synchronizing speech with textual data. The spoken words can be extracted using speech transcripts and based on the context and semantics of the speech, and they can be associated with image text.

Usman masud, et.al,…[4] plan is to make an effective system which will assist visually impaired people through obstacle detection and scenes classification. The proposed methodology utilizes Raspberry-Pi 4B, Camera, Ultrasonic Sensor and Arduino, mounted on the stick of the individual. We take pictures of the scene and afterwards pre-process these pictures with the help of Viola Jones and TensorFlow Object Detection algorithm. The said techniques are used to detect objects. We also used an ultrasonic sensor mounted on a servomotor to measure the distance between the blind person and obstacles. The presented research utilizes simple calculations for its execution, and detects the obstructions with a notably high efficiency. When contrasted with different frameworks, this framework is a minimal effort, convenient, and simple to wear. Under the scope of various circumstances that have been encountered in this work, the overall system gives us an average efficiency of about 91% which is also a great enhancement for our project. Moreover, it has a rechargeable battery whose time is around 24 hours, so the user can recharge it during night. As the system is integrated with VNC Viewer, it can be connected to cell phone of the person. The KNFB reader may be used to convert text to voice. With a single tap, the KNFB Reader transforms text into high-quality voice, providing accurate, rapid, and efficient access to single and multi-page texts. We believe that this system can be a steppingstone for greater advancements in helping visually impaired individuals

Junhyung kang, et.al,…[5] introduced recently to overcome the challenging issues. Our survey paper explores the recent approaches in satellite and aerial imagery-based object detection research and aims to stimulate further research in this area by presenting comprehensive and comparative reviews. After researching a number of papers, we categorized the most important approaches into the six different categories. Further, we compare and analyze publicly available datasets to utilize and motivate research on object detection with the overhead imagery domain. Also, based on the difference in image sources, our paper surveyed the datasets with helpful information such as image resolution and size. We hope this paper will be helpful in developing more advanced deep learning-based approaches as well as understanding and discussing future research directions. Significant advancements and progress made in recent computer vision research enable more effective processing of various objects in high-resolution overhead imagery obtained by various sources from drones, airplanes, and satellites. In particular, overhead images combined with computer vision allow many real-world uses for economic, commercial, and humanitarian purposes, including assessing economic impact from access crop yields, financial supply chain prediction for company's revenue management, and rapid disaster surveillance system (wildfire alarms, rising sea levels, weather forecast).. This survey paper provides a comprehensive overview and comparative reviews on the most up-to-date deep learning-based object detection in overhead images. Especially, our work can shed light on capturing the most recent advancements of object detection methods in overhead images and the introduction of overhead datasets that have not been comprehensively surveyed before.

Supriya v. Mahadevkar, et.al,…[6] applied in computer vision applications is reviewed in-depth in this article. The findings of a systematic literature review on machine learning styles are presented in this review. The authors intended to draw attention to the utilized learning types, adopted feature extraction techniques, methodologies, approaches, approved data sets, adopted application domains, and difficulties related to ML approaches in diverse sectors. This study planned, executed, and carried out different SLR phases on ML styles.

In the literature review for computer vision applications, other artificial intelligence methods— such as those based on deep learning and machine learning— have been used. Deep learning and machine learning-based techniques are popular thanks to easily accessible datasets and automated feature extraction methods. The authors investigated publicly accessible computer vision datasets. In this paper, we have discussed the different learning styles used in the field of Computer vision, Deep Learning, Neural networks, and machine learning. Some of the most recent applications of machine learning in computer vision include object identification, object classification, and extracting usable information from images, graphic documents, and videos. Some machine learning techniques frequently include zero-shot learning, active learning, contrastive learning, self-supervised learning, life-long learning, semi-supervised learning, ensemble learning, sequential learning, and multi-view learning used in computer vision until now. There is a lack of systematic reviews about all learning styles. This paper presents literature analysis of how different machine learning styles evolved in the field of Artificial Intelligence (AI) for computer vision. This research examines and evaluates machine learning applications in computer vision and future forecasting. This paper will be helpful for researchers working with learning styles as it gives a deep insight into future directions

Xufei wang, et.al,…[7] described in this paper, the penalty term in LCIoU of the bounding box regression loss function was improved. The new penalty term more comprehensively considers the aspect ratio relationship between the bounding box of ground truth and prediction, and includes the bounding box of prediction with the same aspect ratio as that of ground truth. The loss function, which is made up of the new penalty term, is called LICIoU . Experiments on the Udacity, PASCAL VOC, and MS COCO datasets have proved the effectiveness of LICIoU in improving localization accuracy of the model by using the one-stage target detector YOLOV4. The method improved the comprehensiveness of the positioning judgment of the bounding box of prediction, strengthened the effect of the penalty function, and improved localization accuracy of the model. According to the theoretical analysis and calculation results of LICIoU function, our algorithm is advanced to a certain extent. However, we believe that the main limitation of the research in this paper is that it is only verified in the YOLOv4 network at present. In order to better demonstrate the effectiveness of the proposed method, it needs to be verified in other neural networks in the future. At the same time, the loss function of better performance is also one of the contents of our future research. An object detector based on convolutional neural network (CNN) has been widely used in the field of computer vision because of its simplicity and efficiency. The average accuracy of CNN model detection results in the object detector is greatly affected by the loss function. The precision of the localization algorithm in the loss function is the main factor affecting the result. Based on the complete intersection over union (CIoU) loss function, an improved penalty function is proposed to improve the localization accuracy. Specifically, the algorithm more comprehensively considers matching bounding boxes between prediction with ground truth, using the proportional relationship of the aspect ratio from both bounding boxes. Under the same aspect ratio of the two bounding boxes, the influence factors of the prediction box on localization accuracy were considered

Sadia zafar, et.al,…[8] considerable portion of the world’s population that demands assistance to perform activities of daily life. Several devices have been developed with the help of emerging technologies to facilitate them in object/obstacle detection and recognition, navigation, and mobility, particularly in indoor and outdoor environments. This paper presented a comprehensive comparative analysis of assistive devices for VIPs. These devices are classified based on their functionality and working mechanism. The advantages of these devices along with the limitations are also discussed after performing a consolidated analysis of the devices. Moreover, a scorebased quantitative analysis of these devices has been performed based on their discriminant features. It is evident from the analysis that none of the systems/devices are providing up-to-the-mark performance. It is notable that each method maintains distinct feature(s) over the other and also has more landscapes than the other, but none of these sustained all the assessed features. It can be concluded that no device can be considered an ideal device. So, there is a need for developing an intelligent system that may cover all the essential features in order to support VIPs. This research work may assist the researchers/scientists who are passionate about developing the devices for VIPs. It would also help to select an appropriate device for a particular scenario. The paper presents a comprehensive comparative analysis of the state-of-the-art assistive devices for VIPs. These techniques are categorized based on their functionality and working principles. The main attributes, challenges, and limitations of these techniques have also been highlighted. Moreover, a score-based quantitative analysis of these devices is performed to highlight their feature enrichment capability for each category. It may help to select an appropriate device for a particular scenario.

Kalyani kadam, et.al,…[9] presents Mask R-CNN with MobileNet V1 as a lightweight model for the detection of multiple image splicing forgery. It also provides a forged percentage score for multiple spliced images. The model specified in the literature evaluated only one image splicing dataset. However, the proposed model is evaluated on the MISD dataset for multiple image splicing and image splicing on CASIA 1.0, WildWeb, Columbia Gray, and Columbia Color datasets.

Also, comparative analysis of the proposed model is done with variants of ResNet such as ResNet 51,101, and 151. The proposed model achieves an average precision of 82% on the Multiple Image Splicing Dataset. The configuration of the proposed model is more efficient in terms of computing than variants of ResNet [39] used for the detection and identification of image splicing forgeries. The evaluation of the proposed model compared to variants of the ResNet [39] network shows that the proposed approach efficiently balanced efficiency and computational costs. The future work focuses on the use of various deep learning architectures such as GAN, MobileNet V2 with Mask-RCNN for detection and localization of multiple image splicing. Currently, the proposed model handles all the attacks/post-processing operations specified by table Image Splicing Dataset. However, in the future, we will try to evaluate our proposed model on a greater number of attacks and will compare evaluation results with and without attacks.

Wenfengzheng, et.al,…[10]including the characteristics of entity structure, hierarchy, and causal relationship between entities, semantic representation instead of image visual-feature as a visual reasoning model is proposed. Make the reasoning process more transparent and increase the comprehensibility of the model. It is convenient to decouple the system from the bottleneck of the analysis model. Improve U-Net so that the output of U-Net is not a mask of the scene object but a Distance Map for watershed segmentation. The output of U-Net can be directly used for watershed segmentation, equivalent to passing deep neural networks. The effect of watershed segmentation is optimized so that the method can obtain satisfactory image segmentation results under the condition of fewer samples. Use the model based on the attention mechanism to transform the natural language into a potential logical representation, which can be used to map natural language into a program tree-like machine translation. This paper demonstrates how the semantic representation can be used as an input and verifies that changing the representation of the image can further improve system performance. After replacing the visual feature, the accuracy of non-relational questions was significantly improved. Then the semantic vector was pre-processed by constructing a relation matrix. The semantic representation effect is competitive compared to visual representation, and the semantic representation is simple and easy to carry out other processes. After analysis, it was summarized that introducing semantic information was equivalent to a feature selection and extraction before input. The selected features were useful for answering questions. Compared with the feature extraction of CNN, the semantic information is more accurate and less redundant. So, it is easier to find the precise relationship when handling relational reasoning

III. EXISTING SYSTEM

Blind assistance is promoting a widely challenge in computer vision such as navigation and path finding. In existing system, two cameras placed on blind person's glasses, GPS free service, and ultra-sonic sensor are employed to provide. The necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. The two cameras are necessary to generate the depth by creating the disparity map of the scene, GPS service is used to create groups of objects based on their locations, and the sensor is used to detect any obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. And also implement vision substitute system designed to assist blind people for autonomous navigation. Its working concept is based on ‘image to sound’ conversion. The vision sensor captures the image in front of blind user. This image is then fed to MATLAB for processing. Process intuit processes the captured image and enhances the significant vision data. This processed image is then compared with the data base kept in microcontroller. The processed information is then presented as a structured form of acoustic signal and it is conveyed to the blind user using set of ear phones. Colour information from the interested objects evaluated to determine the color of the object. The color output is informed to the blind user through headphones.

IV. PROPOSED SYSTEM

The human eye is the organ which gives us the sense of sight, allowing to observe and learn more about the surrounding world than we do with any of the other from sense. We use our eye in almost every activity we perform, whether reading, working, watching television, writing a letter, driving a car, and in countless other ways. Most people probably would agree that sight is the sense they value mare than all the rest. The ability to see clearly depends on how well these parts work together. Light rays bounce off all objects. If a person is looking at a particular object, such as a tree, light is reflected off the tree to the person’s eye and enters the eye through the cornea (clear, transparent portion of the coating that surrounds the eyeball). The proposed system develops an object detection method combining top-down recognition with bottom-up image segmentation. There are two main steps in this method: a hypothesis generation step and a verification step. In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust to object deformation and background clutter. The improved Shape Context is used to generate a set of hypotheses of object locations and figure ground masks, which have high recall and low precision rate.

In the verification step, we first compute a set of feasible segmentations that are consistent with top-down object hypotheses, then we propose convolutional neural network procedure to detect and recognize the objects. We exploit the fact that feature vectors for object detection and recognition. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. Then implement text recognition techniques to identify the text strokes from uploaded images using Optical character recognition algorithm. Then recognized text can be converted into voice. The experimental results reveal the performance of the proposed work in about real time system. Fig 1 shows the proposed work.

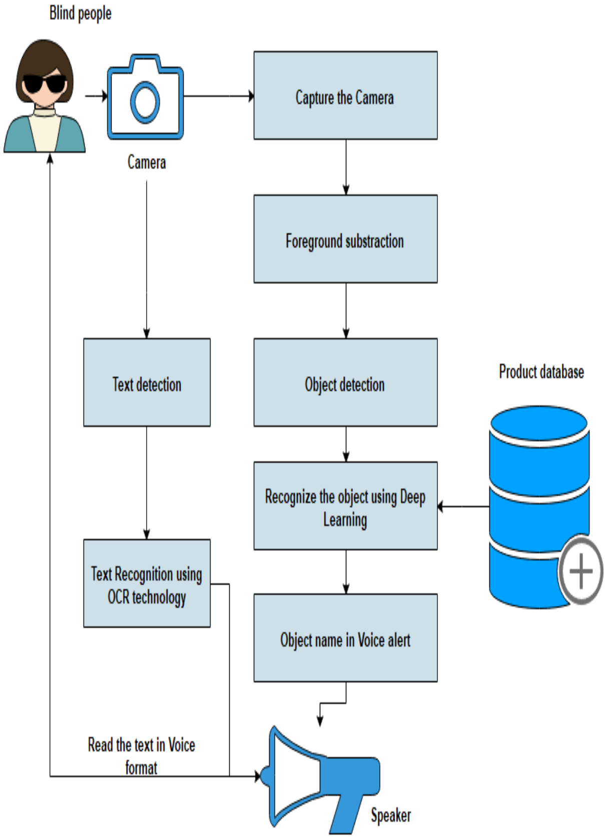

Blind people
Camera
Capture the Camera
Foreground substraction
Text detection
Object detection
Product database
Recognize the object using Deep Learning
Text Recognition using OCR technology
Object name in Voice alert
Read the text in Voice format
Speaker

Fig 1: Proposed work

A. IMAGE CAPTURE

Computer vision is a multidisciplinary field that studies how computers can be built to learn at a high level from electronic images or videos. From an engineering standpoint, it strives to automate that the human eye can perform. The automatic extraction, analysis, and comprehension of helpful data from a single image or a sequential of images is the focus of computer vision. It entails the creation of a conceptual and algorithmic foundation for automatic visual comprehension. The footage is recorded using a webcam, as well as the panels from the video are separated and pre-processed. First, continuously capture objects from the webcam and adapt to the processing. After extracting the desired object from the image taken and converting it to coding format.

B. OBJECT DETECTION

Due to routine changes in object motion and variability in scene size, occluded, appearance variants, and self-image and illumination changes, object tracking and detection is among the crucial areas of research. In particular, feature selection is critical in object tracking. It is linked to many real-time applications such as vehicle perception and video surveillance. To overcome the detection problem, monitoring linked to attribute movement and appearance is used. The majority of the algorithm is devoted to the tracking algorithm, which is used to help smooth the video stream. Few methods, on the other hand, make use of previously information available about object shape, colour, texture, and so on. Tracking algorithm that combines the above-mentioned object parameters. In this module, feature extraction is used to extract features from an image and match them with a trained database. Finally, the object is recognized and labeled using a machine learning algorithm.

C. LABEL TO SPEECH CONVERSION

Message synthesis (TTS) is indeed the fully automated conversion of a message into speech that sounds as close to a native speaker of the language trying to read the text as possible. TTS (text-to-speech synthesizer) is a technology that allows a computer to speak to you. The TTS system receives text as input, then a computer program known as the TTS motor analyses this same text, pre-processes it, and synthesizes speech using mathematical models. As an output, the TTS engine generates audio signals in an audio format. The text-to-speech (TTS) synthesis procedure is divided into two stages.The first is text analysis, in which the regarding the mode is transcribed into a phonetic or other linguistic representation, as well as the second is speech waveform generation, in which the output is generated from this pronunciation and prosodic information. These two stages are commonly referred to as high-level and low-level synthesis. The input text could be information from the text editor, standard ASCII from e-mail, a mobile text message, or scanned text from a newspaper, for example. This same character string is then which was before and analyzed to produce a phonetic representation, which is typically a sequence of characters of morphemes with some additional info for correct enunciation, duration, and stress. The information from the high-level synthesizer is finally used to generate speech sound with the low-level synthesizer.For this purpose, Android includes a Text to Speech class. To use this class, you must first create an object of it and then specify the Speech synthesizer (). In this hearer, you must specify the Message to Speech object's properties, such as language and pitch. The developed scheme would thus enable the blind people to shop by themselves using technology. The camera captures show the interface of a proposed system, that would transform the digitized object of the good or service to audio, allowing the blind person to function independently.

D. TEXT DETECTION AND RECOGNITION

Analyse text strokes in this module using an algorithm for feature extraction such as the optical character detection algorithm. This same concept of feature detection in imaging and computer vision refers to methods that aim to compute abstractions of image data and make local decisions at each image point whether or not an image feature of a given type exists at that point. This same eventually results features will be image domain subsets, typically in the form of discrete points, continuous curves, or connected regions. A reduced image processing operation is featuring detection. That is, it is typically performed as the initial operation on an image and examines each pixel to see if a feature is prevalent at that pixel.When this is a part of a larger method, the algorithm will generally only investigate the image in the features region. After detecting a text region in an image, text is extracted from that text region using personality descriptive terms and structure configuration. These methods are used to convert text-filled images in to the editable formats and to process input images. To amalgamate the extracted features during the feature quantization process, a BOW with OCR model is used. All detectors' key points are subjected to OCR. Character structure is composed of various oriented strokes that serve as basic text character elements. A pulse of written material is described as an area surrounded by two parallel bounding segments at the pixel level. Their orientation is considered to be stroke orientation, as well as the spacing between the two is considered to be stroke width. Then, to confirm this same character classes, their correlating character classifiers are called. If the majority of enquired protagonists exist, the information extraction application will return a positive response; otherwise, it will return a negative response. For people, the recognised message could be converted into voice.

V. RESULTS AND DISCUSSION

The Vision Assist System was evaluated based on product recognition accuracy, text extraction accuracy, response time, and user satisfaction.

Metric	Model Used	Achieved Performance
Object Recognition Accuracy	YOLOv11, EfficientNet	94.30%
OCR Text Recognition Accuracy	EasyOCR, Tesseract OCR	92.50%
Response Time for Object Detection	YOLOv8	120 ms

Response Time for OCR Processing	EasyOCR	180 ms
TTS Processing Latency	Google TTS	80 ms

Table 1: Performance table

The proposed framework results are shown in following figures.

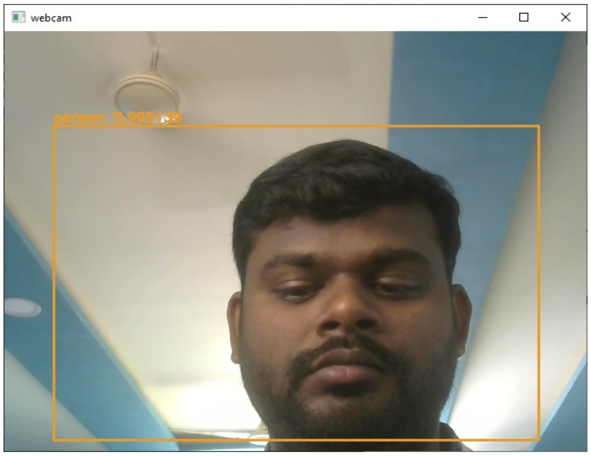
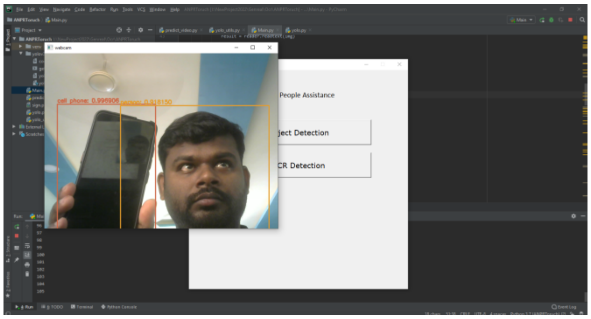


Fig 2: Person detection



```
result = reader.readtext(img)
```

Fig 3: Cell phone detection

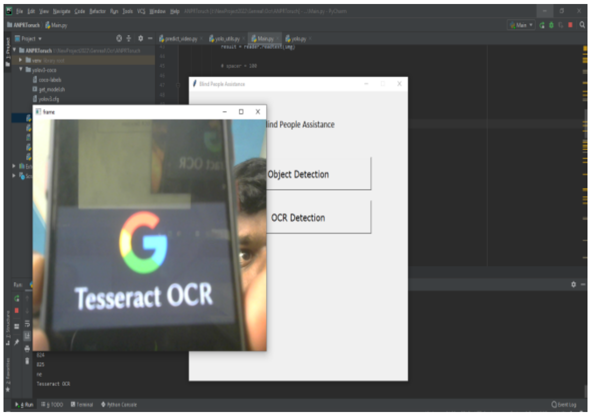


Fig 4: Text recognition

This suggests an application is being developed to assist visually impaired individuals using object and optical character recognition technologies.

The phone's screen is being captured and displayed within the IDE, indicating the application is likely under testing or development.

VI. CONCLUSION

We display a vision system for blind people utilizing object-like images and a video scene in this project. For object recognition, this system employs machine learning. In order to identify some objects under various conditions. Detection is concerned with detecting objects within an image or video. The Object Detection API, which is based on machine learning, makes it simple to create and use a object detection model. Blind people have little data on self-velocity objects and direction, both of which are necessary for travel. The navigation systems are expensive and out of reach for most blind people. As a result, the primary goal of this project is to assist blind people. This method effectively distinguishes the target object from the back story or other objects in the camera's field of view.We proposed a new object localization method based on mathematical models of injury alignment and edge distributions to retrieve object areas from complex backgrounds. From captured images, the corresponding charts estimate the overall structural feature of the object. To design the feature point layouts of an enhanced image into a feature vector, block patterns are defined. This system is based on real-time camera analysis and interpretation, and it can assist blind people in recognising objects in their surroundings.

REFERENCES

[1] Ashiq, Fahad, et al. "CNN-based object recognition and tracking system to assist visually impaired people." IEEE access 10 (2022): 14819-14834.

[2] Wang, Wenguan, et al. "Salient object detection in the deep learning era: An in-depth survey." IEEE Transactions on Pattern Analysis and Machine Intelligence 44.6 (2021): 3239-3259.

[3] Mahrishi, Mehul, et al. "Video index point detection and extraction framework using custom YoloV4 Darknet object detection model." IEEE Access 9 (2021): 143378-143391.

[4] Masud, Usman, et al. "Smart assistive system for visually impaired people obstruction avoidance through object detection and classification." IEEE access 10 (2022): 13428-13441.

[5] Kang, Junhyung, et al. "A survey of deep learning-based object detection methods and datasets for overhead imagery." IEEE Access 10 (2022): 20118-20134.

[6] Mahadevkar, Supriya V., et al. "A review on machine learning styles in computer vision—techniques and future directions." Ieee Access 10 (2022): 107293-107329.

[7] Wang, Xufei, and Jeongyoung Song. "ICIoU: Improved loss based on complete intersection over union for bounding box regression." IEEE Access 9 (2021): 105686-105695.

[8] Zafar, Sadia, et al. "Assistive devices analysis for visually impaired persons: A review on taxonomy." IEEE Access 10 (2022): 13354-13366.

[9] Kadam, Kalyani, et al. "Detection and localization of multiple image splicing using MobileNet V1." IEEE Access 9 (2021): 162499-162519.

[10] Zheng, Wenfeng, et al. "Improving visual reasoning through semantic representation." IEEE access 9 (2021): 91476-91486.

[11] Li, Zhichun, et al. "Enhancing Revisitation in Touchscreen Reading for Visually Impaired People with Semantic Navigation Design." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6.3 (2022): 1-22.

[12] Afif, Mouna, et al. "Recognizing signs and doors for Indoor Wayfinding for Blind and Visually Impaired Persons." 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, 2020.

[13] Bhole, Swapnil, and Aniket Dhok. "Deep learning based object detection and recognition framework for the visually-impaired." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.

[14] Dahiya, Dhruv, Hardik Gupta, and Malay Kishore Dutta. "A Deep Learning based Real Time Assistive Framework for Visually Impaired." 2020 International Conference on Contemporary Computing and Applications (IC3A). IEEE, 2020.

[15] Chang, Wan-Jung, et al. "MedGlasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients." IEEE Access 8 (2020): 17013-17024.

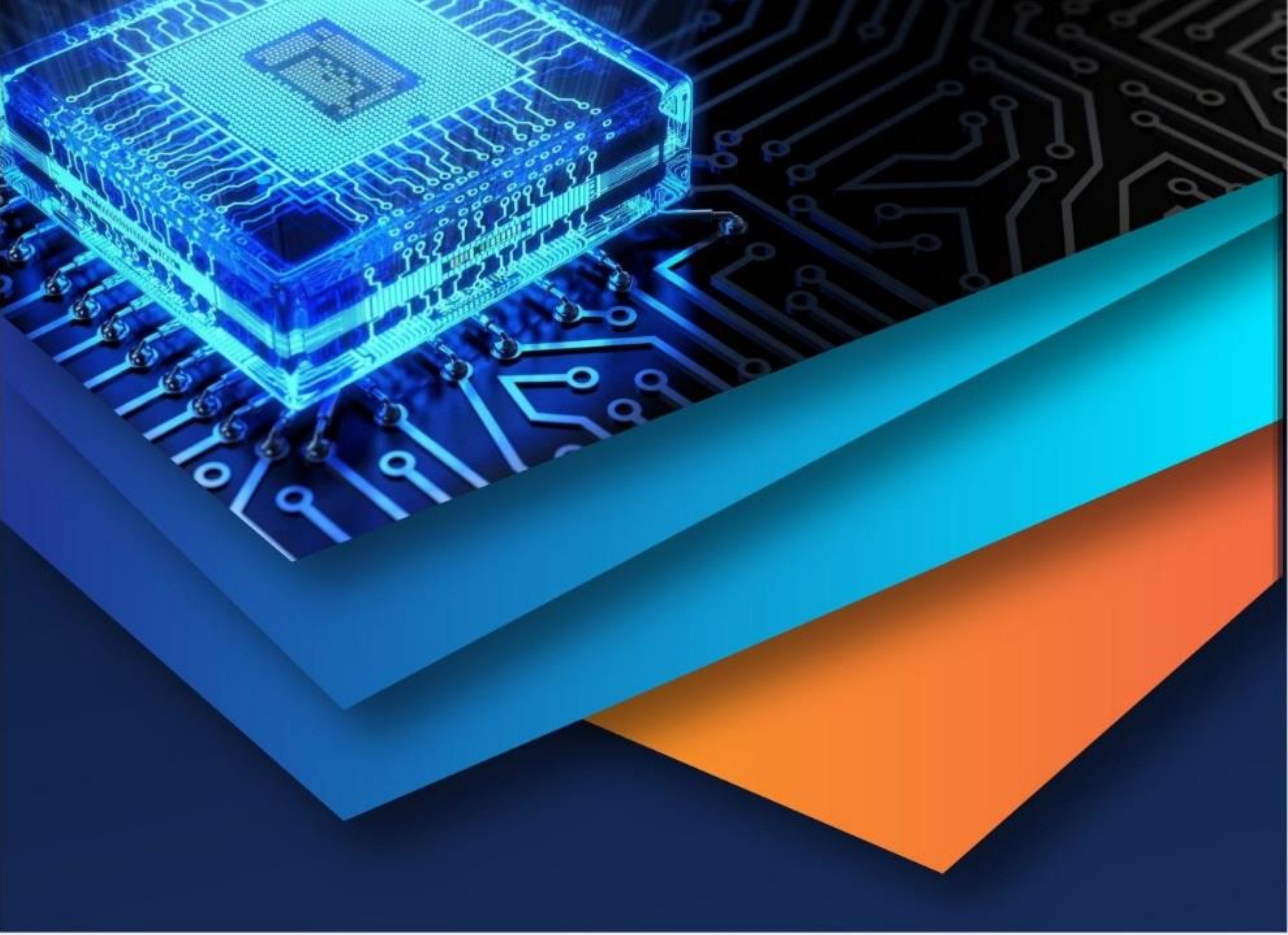
[16] Chang, Wan-Jung, et al. "A deep learning based wearable medicines recognition system for visually impaired people." 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE, 2019.

[17] Sufri, N. A. J., et al. "Vision based system for banknote recognition using different machine learning and deep learning approach." 2019 IEEE 10th Control and System Graduate Research Colloquium (ICSGRC). IEEE, 2019.

[18] Svaigen, Alisson Renan, Lailla M. Siqueira Bine, and Linnyer Beatrys Ruiz Aylon. "An Assistive Haptic System Towards Visually Impaired Computer Science Learning." Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference. 2018.

[19] Tange, Yutaka, Tomohiro Konishi, and Hideaki Katayama. "Development of vertical obstacle detection system for visually impaired individuals." Proceedings of the 7th ACIS International Conference on Applied Computing and Information Technology. 2019.

[20] Rahman, Sami ur, Sana Ullah, and Sehat Ullah. "A mobile camera based navigation system for visually impaired people." Proceedings of the 7th international conference on communications and broadband networking. 2019.



doi crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

SJIFactor
TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI