Lecture Notes in Networks and Systems 1023

Palaiahnakote Shivakumara
Saurov Mahanta
Yumnam Jayanta Singh *Editors*

Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology

NICEDT-2024, Volume 2

Springer

Lecture Notes in Networks and Systems

Volume 1023

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Palaiahnakote Shivakumara · Saurov Mahanta ·
Yumnam Jayanta Singh
Editors

Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology

NICEDT-2024, Volume 2

Springer

Editors
Palaiahnakote Shivakumara
University of Salford
Salford, UK

Saurov Mahanta
NIELIT
Guwahati, Assam, India

Yumnam Jayanta Singh
NIELIT
Guwahati, Assam, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-97-3603-4 ISBN 978-981-97-3604-1 (eBook)
https://doi.org/10.1007/978-981-97-3604-1

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

NICEDT-2024 is the second edition of the International Conference, hosted by the National Institute of Electronics and Information Technology (NIELIT) on February 16 and 17, 2024, in Guwahati, Assam, India. This conference was dedicated to exploring the latest developments in Information, Electronics, and Communication Technology, as well as allied fields, with the goal of attracting a diverse array of attendees from around the globe, comprising researchers, educators, industry experts, and eager students.

The conference maintained the same eight thematic areas as in the inaugural edition NICEDT-2023, covering almost all the areas of Digital Technologies. Out of the eight tracks, this volume [Volume-2] comprises a total of 40 full-length research or review articles selected from the four tracks: [1] VLSI and Semiconductors, Electronics System, IoT, Robotics and Automations; [2] Digital Technologies for Future; [3] Assistive Technology for Divyangjan (People with Disabilities); and [4] Strategy for Digital Skilling for Building a Global Future Ready Workforce. These papers were selected through a triple blind-review process and subsequently presented in the NICEDT-2024 conference.

During the inaugural session of NICEDT-2024, which took place on February 16, Dr. Yumnam Jayanta Singh, Executive Director of NIELIT, delivered the welcome address, emphasizing the conference's aim to promote collaboration between academia and industry within the NIELIT community. Dr. Singh outlined the conference's agenda, which encompassed eight thematic areas including Artificial Intelligence, Cyber Security, and Digital Skilling. Eminent individuals such as Dr. Karel Sterckx, Director, Bangkok University Center of Research in Optoelectronics, Communications and Computational Systems (BU-CROCCS) and Prof. Nitin Kumar Tripathi, Acting Vice President for Academic Affairs (VPAA), Asian Institute of Technology (AIT), Thailand, were acknowledged for their esteemed presence, and Shri Dhrubajyoti Borah commended NIELIT's efforts, underscoring the importance of academia–industry cooperation. Professor Pratap Jyoti Handique, Vice Chancellor of Gauhati University, emphasized the transformative role of communication technology and extended a warm welcome to attendees. Over 200 full-length articles were submitted, with approximately 80 set for publication in Springer's *Lecture Notes in*

Virtual Assistant and Navigation for the Visually Impaired Using Deep Neural Network and Image Processing

Palakolanu Harsha Vardhan Reddy, Bhumireddy Kashyap, Bhukhya Charan Naik, Prashant Pal, Shashank Kumar Singh, Saurabh Bansod, and Yogesh Kumar

Abstract The rapid advancement of leading-edge encourages people to employ available resources to simplify everyday duties and enhance the norm and caliber of living for individuals who are unsighted. This module gives real-time aural advice to enable safe and independent navigation and advises developing a product to assist blind in navigating their environment. The gadget features an inbuilt detection of barriers module utilizing artificial intelligence to recognize impediments and give the user portable, hands-free haptic feedback. Additionally, the device has a text identification mechanism which can transform any written content it recognizes to voice which the wearer can perceive through headphones incorporated into their glasses. This module uses image processing techniques to improve accuracy and speed. Users could not have accessed printed elements like menus and signs in this manner in the past. Additionally, the system features a module for recognizing sign boards that speak text on signs. Following a comprehensive evaluation that considered factors such as speed, accuracy, and user experience, the Virtual Assistant have the capacity to greatly improve the standard of life for individuals who are unsighted. This study advances assistive technology while demonstrating the amazing power of combining state-of-the-art technologies to solve practical problems. By integrating these qualities into wearable technology, the proposed method could enhance the quality of life and freedom for individuals who are blind.

Keywords Wearable · Navigation · Haptic feedback · Object detection · Text recognition module

P. Harsha Vardhan Reddy · B. Kashyap · B. C. Naik · P. Pal · S. K. Singh · S. Bansod (✉) · Y. Kumar
NIELIT Aurangabad, Aurangabad, India
e-mail: saurabhbansod@nielit.gov.in

P. Shivakumara et al. (eds.), *Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology*, Lecture Notes in Networks and Systems 1023, https://doi.org/10.1007/978-981-97-3604-1_29

1 Introduction

Visual impairment is a major problem that affects millions of individuals worldwide. Additionally, 217 million individuals suffered from moderate to severe visual impairment. There are several conditions that can cause blindness, but the most prevalent ones include age-related macular degeneration, cataracts, glaucoma, and uncorrected refractive errors.

Over the years, numerous techniques and devices have been developed to assist those who are unsighted in surroundings. While electronic navigation aids can provide more accurate information about obstacles than conventional methods, they are not always helpful for some obsolete and challenging tasks.

A white cane may not alert you to obstacles enough to avoid accidents, and its proper usage requires extensive training. Previously acknowledged devices with limitations include laser frameworks, sonar systems, guiding animals, and TGSIs. While aural feedback systems help identify nearby obstacles, and they are unable to give enough details about the hitch location or nature to allow for safe avoidance. Furthermore, navigation and text reading alter audio feedback.

This project, the foundation of the present work is an investigation of the possible coexistence of modern technology with unsighted people's requirements. The advancement of virtual assistants and deep neural networks has opened up new avenues for improving the lives of people with visual impairments. This publication focuses on developments that can provide text reading, navigation, and virtual aids for the blind.

By adopting smart glasses for object detection, the blind can obtain current environmental data. It acknowledges things and individuals using a camera along with advanced machine learning techniques, and they provide them with haptic or aural feedback.

Another option for a reading assist would be a set of smart glasses. Using OCR innovation, this text read by the glasses and converted to voice or braille from books, signs, and other printed items. As a result, the visually challenged might have greater freedom when reading and accessing information.

2 Review of Literature

This section evaluates related work in the realms of obstacle detection approaches and virtual reading aides since those are the main foci of this work.

Examining, modifying, and altering digital images to enhance their visual appeal or extract important information is known as image processing. Obtaining pictures processing before processing this may involve the use of methods like compression, segmentation, filtering, feature extraction, or object identification. For thresholding, edge detection, etc. following post-processing analysis (Fig. 1).

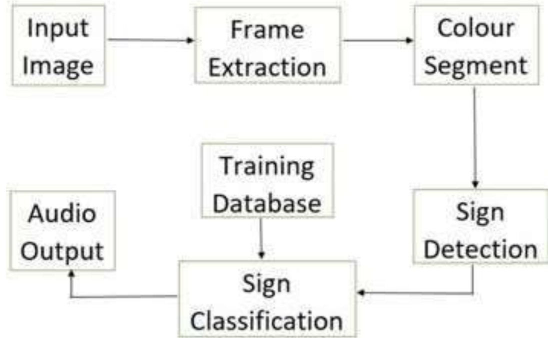


Fig. 1 Flow model of virtual assistance for sign detection

Sáez [1] He talked about how the TensorFlow framework was used to create the systems. Kandalan [2] used Python programming and the OpenCV package to create image processing algorithms for a Raspberry Pi hardware platform. The system receives input from the camera module and extract text from captured image. The suggested system methods are:

1. Image Pre-processing
2. Character Division
3. Identification of Characters
4. Voice-to-text Translation.

Text is extracted from digital photos using an open library called Tesseract OCR model. Using machine learning techniques, it examines images and detects patterns that correspond to letters, numbers, and other characters. The OCR engine in Tesseract operates in stages. First, the image is pre-processed to improve contrast, standardize text size, and lessen noise. Following that, certain characters are extracted from the image using character segmentation. After that, the Tesseract OCR engine looks at each character separately and extracts relevant features, like the character's height, breadth, and shape, using feature extraction techniques.

The Pi camera is used to record the content, and OpenCV is then used to analyse the picture to produce a higher quality version. Next, this Tesseract OCR system receives the edited picture and utilizes it to find the text. This identified content is then converted to speech using the eSpeak speech synthesizer, and the speaker plays the audio output Xiao [3]. The user takes pictures and starts the OCR process by pressing a tactile button on the control panel.

Object detection techniques frequently require multiple processing stages, such as pre-processing images, feature extraction, object proposal creation, object classification, and object location (Fig. 2).

Advanced machine learning techniques, network-based solution for current time recognition of items is presented by Hunaiti [4] and Farhadi. The suggested approach, called YOLO, after dividing a picture into a grid. The system has a great degree of efficiency and accuracy when it comes to real-time object detection. Using the COCO dataset for testing and the PASCAL VOC 2012 dataset for training, they developed the YOLO model. A novel deep learning-based object recognition technique was introduced by Katzschmann [5]. They trained and tested their model on a computer

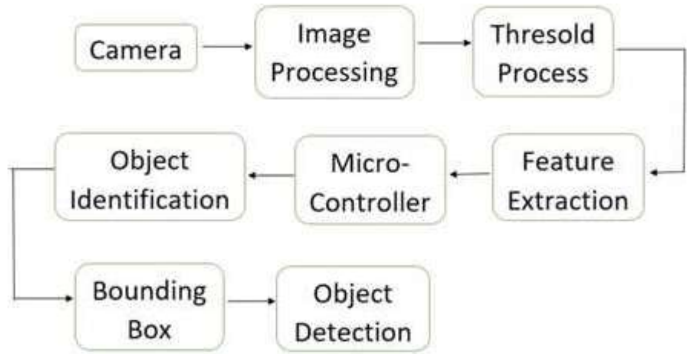


Fig. 2 Object detection flow model of virtual assistance

equipped with a GPU. Additionally, they made use of common deep learning tools like TensorFlow and Caffe.

For real-time object detection, Barontini [6] suggests the detection network receives the region proposals that the RPN creates and gives each suggestion an objectless score. The detection network uses the scores to perform bounding box regression and classification.

Aladrén [7] suggested a 12-core Intel Xeon CPU, and the Caffe deep learning framework. Even after several attempts, many modern systems are unable to fully integrate components such as computational resources and precision. For those who are unsighted, this model gives approachable choice. Additionally, offers enhanced efficiency, a better user experience, and contextual awareness in its object-detecting capabilities.

3 Proposed System

The creation of a simple, portable, Text-to-speech conversion and ETA model capabilities for common, daily inside and outdoor use was our main contribution to this project.

An electronic travel aid (ETA) is a gadget that, within a specific range, uses energy waves to identify things in the surrounding environment. After processing reflected data, the ETA notifies the user—typically through voice announcements, sounds, or vibrations. To increase users' mobility and safety, electronic travel aids can use sensor technologies, laser, sonar waves, or global positioning systems.

We introduce an innovative visual aid selection to individuals who are completely blind in this research. A number of unique characteristics prove the originality of the presented design.

A lightweight devices, portable, free of hands, and compact reading aid that can be placed on glasses for indoor as well as outdoor exploration.

Advanced algorithm execution in the entry-level setup.

Accurate, real-time distance measuring using cameras that saves money, optimizes design, and calls for fewer sensors.

This model consists of custom 3D printed lens holder, an Arduino board, an image sensor module, and smart eyewear. The camera module image-editing software is used to handle the images of printed content captured on the Raspberry Pi device.

Andoa [8] recovered text sections are then fed into an OCR machine, which transforms the text into a digital representation. A text-to-speech or Braille converting device then uses the digitized text to produce the corresponding output.

To raise the virtual assistant's accuracy and productivity, Duh [9], incorporate object identification through advanced machine learning techniques, to work, YOLO makes a grid out of the image and calculates the likelihood of every cell in the grid that have an object. It has capability of real-time object recognition in several images we train the YOLO system using a variety of printing commodities, such as novels, magazines, and publications, to allow the identification of specific products (Fig. 3).

An OCR system is used to transform the written content taken from collected text areas to electronic form. Tesseract is an open-source OCR engine that serves as the foundation for the OCR technology utilized by the reading aid. It is possible to instruct Tesseract to identify custom styles, accents, and a range of types of writing and fonts. In order to increase accuracy, the OCR system also uses post-processing methods including text normalization and spell checking (Fig. 4).

Integrated text-to-speech functions on the Raspberry Pi board are designed to provide a smooth user experience (Table 1).

Fig. 3 A prototype flow structure

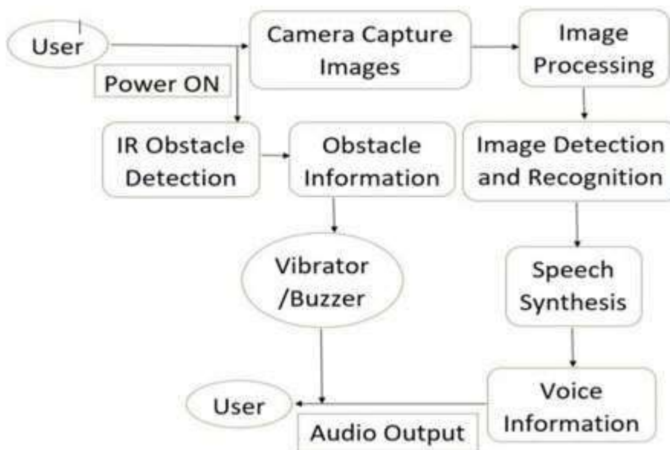


Fig. 4 Smart glasses used by different kinds

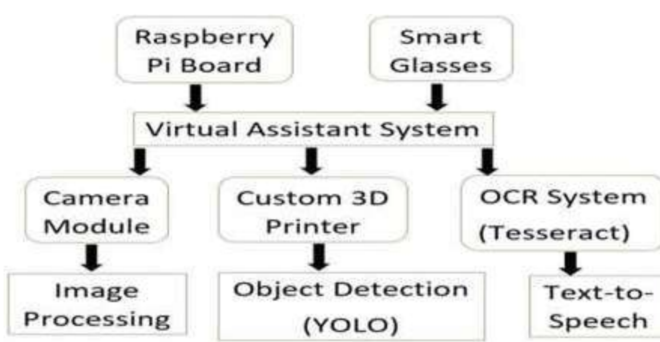


Table 1 Difference between proposed system and the existing system

Model	Existing system	Proposed system
Design	Typically, bulky and stationary [1]	Portable and hands-free
Object detection	Typically, not included [7]	Incorporates deep neural networks for detection of objects to allow for the identification of particular things
Text recognition	Depends on the technology of OCR, which comes with limitations in terms of accuracy and efficiency [4]	Employs image processing methods to increase precision and effectiveness
User interface	Frequently needs a physical interface, such as a keyboard or buttons [11]	Uses the glasses to provide a user-friendly, hands-free interface

All things considered, the proposed reading assistance for the blind provides significant improvements over the existing models with regard to portability, accuracy, and efficacy. Furthermore, because the suggested model incorporates deep neural network-based object detection and uses smart glasses for a hands-free interface, it offers blind individuals a more accessible and useful option.

4 Methodology

Before OCR and object detection, the gathered photographs are preprocessed to improve the algorithms' accuracy. Among the preprocessing techniques are:

- Image resizing: The acquired images are reduced in size to a set size to facilitate faster processing and need less memory.
- Image enhancement: The gathered photos are improved in order to improve the text sections' contrast and brightness.
- Noise removal: From the gathered images, noise and artefacts that can interfere with object recognition and OCR techniques are eliminated.
- Binarization: The obtained images are converted into binary format in order to facilitate segmentation and the extraction of the text parts.
- Usability: With simple controls and intuitive software, reading glasses should be easy to use.

The method is described as utilizing a Raspberry Pi camera to capture a text picture, followed by the speech synthesizer eSpeak, OpenCV, and the Tesseract OCR engine to convert the text to speech. Below is a detailed explanation of every action.

Finding the items: This is the initial action. Khan [10] Finding and identifying things in an inanimate or dynamic picture is called object identification. This important computer vision duty with multiple uses in security, robotics, and other fields. Among the widely used deep learning models for object recognition is the YOLOv5 model.

YOLOv5, a real-time object recognition framework, uses one convolutional neural network to identify objects in pictures and videos.

Pre-trained model loading is the first stage in using YOLOv5 for object detection. The model can be downloaded and loaded using the PyTorch library. Following the loading of the model, the input picture undergoes preprocessing that includes scaling it to a certain size and normalizing its pixel values so that they lie in the range of 0–1.Input of the image undergoes processing prior to being input into a YOLOv5 for output generation. Bounding boxes with confidence scores and class titles are the end outcome. The objects found in the original image are represented by these bounding boxes.

Bounding boxes, an associated class title, and a rating of trust are painted over the input image to visualize the outcome. To remove borders having inadequate trust ratings, apply the threshold amount. The final step is to display the created image or save it to a CD.

Optical Character Recognition (OCR): The Tesseract OCR engine will be used in the next phase to identify the text in the image. Tesseract is an open-source OCR engine developed by Google. It can identify text in a wide variety of languages from images and has an excellent accuracy rate. The way the engine works is by interpreting and evaluating the words and characters in the picture.

Text to speech conversion: The text is transmitted voice synthesis using the e-Speak voice synthesizer. Once the OCR engine has recognized it. e-Speak is a tiny, multilingual speech synthesizer that is available as open-source software. It can also be used to produce audio by entering text or phonemes. The method uses e-Speak to convert the recognized text to audio.

Input text: Emails, online pages, dictated or typed text, and more are all accepted sources of text input for the engine.

Text analysis: To recognize words, phrases and the appropriate Stress behaviors and intonation, engine uses natural language processing algorithms to examine the text.

Speech synthesis: The engine uses a speech synthesis technique to convert the prosody and the sounds into a wave form of digital audio.

Output Audio: The final step is to play the audio output through speakers. The audio output is received by the speaker on the Raspberry Pi board, which then plays it back. The user can then hear the recognized text as speech.

By pushing a tactile button, the user may control the machine to take an image and initiate the OCR process. The system recognizes text with Tesseract OCR, converts it to voice, utilizes OpenCV to process the image, and emits the audio through the speaker when the user presses the button.

5 Simulation Results

The different parts of the architecture can be described mathematically by a set of equations that make up the YOLOv5 model. Here is a basic summary.

Input: YOLOv5 model takes as its input an image that is represented as a tensor of size (C, H, W), where C is the number of channels (three for RGB images, for illustration), H is the image's height, and W is its width.

Using most recent YOLOv5 model, a custom dataset has been included in the object detection project. I generated my own dataset, by manually tagging about 200 images

```
# Load YOLOv5s model
model = torch.hub.load('ultralytics/yolov5', 'yolov5s', pretrained=True)
```

and video frames from the web using python platform, which is user friendly. Since my dataset is significantly small, I will narrow the training process. By giving an empty string (''), the weights are initialized randomly. The model name will be sent to 'weights' parameter, initializes our YOLOv5 model.

For every object in the picture, there is one bounding-box (B-Box) annotation in the text file. The annotations fall between 0 and 1 after being normalised to the image size. The following format serves as their representation.

Class of object—center X—center Y—Width of Box—Height of Box.

```
# Draw the bounding boxes on the frame
    for detection in results.xyxy[0]:
        x1, y1, x2, y2, confidence, class_idx = detection.tolist()
        label = model.names[int(class_idx)]
        color = (0, 255, 0)
        cv2.rectangle(frame, (int(x1), int(y1)), (int(x2), int(y2)), color, 2)
        cv2.putText(frame, label, (int(x1), int(y1)), cv2.FONT_HERSHEY_SIMPLEX, 1, color, 2)
```

Convolutional Layers: The inserted picture will be processed via sequencing convolutional layers in order to extract features from the picture. Each convolutional layer is represented by the equation below:

$$\boldsymbol{Y_i} = f(\boldsymbol{W_i} * \boldsymbol{x_i} + \boldsymbol{b_i}) \tag{1}$$

where put tenser is denoted by x_i, , the activation function (such as ReLU) is represented by f, and the weight tensor for the i-th convolutional layer is denoted by W_i.

Down sampling: The characteristic charts of convolutional layers are down sampled by utilizing combining sheets in order to lower geometrical measurements of characteristic charts.

The down sampling procedure is represented by the following equation:

$$\begin{aligned} \boldsymbol{y\_(i, j, k)} = \mathbf{max}(\boldsymbol{x_i}, \boldsymbol{i} = 1, \ldots, \boldsymbol{N}, \boldsymbol{j} * \boldsymbol{stride}_{\boldsymbol{h} \le i \le (j+1)} * \boldsymbol{stride_h}, \\ \boldsymbol{k} * \boldsymbol{stride}_{\boldsymbol{w} \le i \le (j+1)} * \boldsymbol{stride_w}) \end{aligned} \tag{2}$$

where x_i is the input tensor, $y_{i,j,k}$ is the output tensor, N is the number of input channels, $stride_h$ and $stride_w$ are the vertical and horizontal strides, respectively, and max is the maximum pooling operation

Input: YOLOv5 model takes as its input an image that is represented as a tensor concatenated with the feature maps produced by the corresponding convolutional layers to produce a set of high-resolution feature maps that capture both low-level and high-level features of the input image. The equation that follows represents the concatenation operation:

$$\boldsymbol{Z_i} = \text{concat}(\boldsymbol{y_i}, \boldsymbol{x_i}) \tag{3}$$

where output tensor of the down sampling layer is denoted by y_i and output tensor of the corresponding convolutional layer is denoted by x_i.

$$\boldsymbol{Y_i} = f(\boldsymbol{W_i} * \boldsymbol{x_i} + \boldsymbol{b_i}) \tag{4}$$

$$\boldsymbol{t_x} = \text{sigmoid}(\boldsymbol{y_i}, \boldsymbol{t_x}) \tag{5}$$

$$\boldsymbol{t_y} = \text{sigmoid}(\boldsymbol{y_i}, \boldsymbol{t_y}) \tag{6}$$

$$\boldsymbol{t_w} = \exp(\boldsymbol{y_i}, \boldsymbol{t_w}) \tag{7}$$

$$\boldsymbol{t_h} = \exp(\boldsymbol{y_i}, \boldsymbol{t_h}) \tag{8}$$

$$\boldsymbol{t_c} = \text{sigmoid}(\boldsymbol{y_i}, \boldsymbol{t_c}) \tag{9}$$

where f is the activation function (e.g., ReLU), t_x, t_y, t_w, t_h, and t_c are the predicted coordinates and confidence score for the detected object, sigmoid is the sigmoid activation function, exp is the exponential function, and y_i is the input tensor.

Detection layer: The output of the detection layers in order for the removal of overlapping spots and chooses the major trustworthy perceptions of every item of the group.

Performances can be evaluated over the training, validation. The final output for the object detection is as follows.

```
# Display the resulting frame
   cv2.imshow('Object Detection', frame)

   # Exit the loop when 'q' is pressed
   if cv2.waitKey(1) & 0xFF == ord('q'):
       break
```

Convolutional, down-sampling, concatenation, and detection layers are some of the methods used by the YOLOv5 model, an advanced deep neural network, to identify and locate objects in an input image (Figs. 5, 6 and 7).

These are few comparison between different object detection models (Graph 1).

An object detection accuracy data called Significance Norm Accuracy at a threshold of 0.5. The SNA threshold range is in between 0.5 to 0.95 and represented by value SNA@0.5:0.95 (Graph 2 and Table 2).

The following is the mathematical expression for an OCR model:

$$y = f(x;\theta) \tag{10}$$

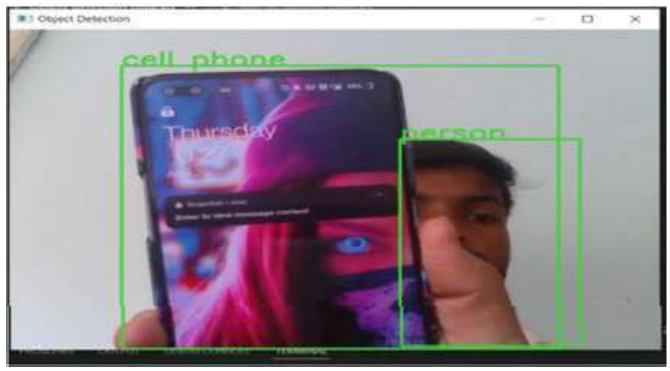
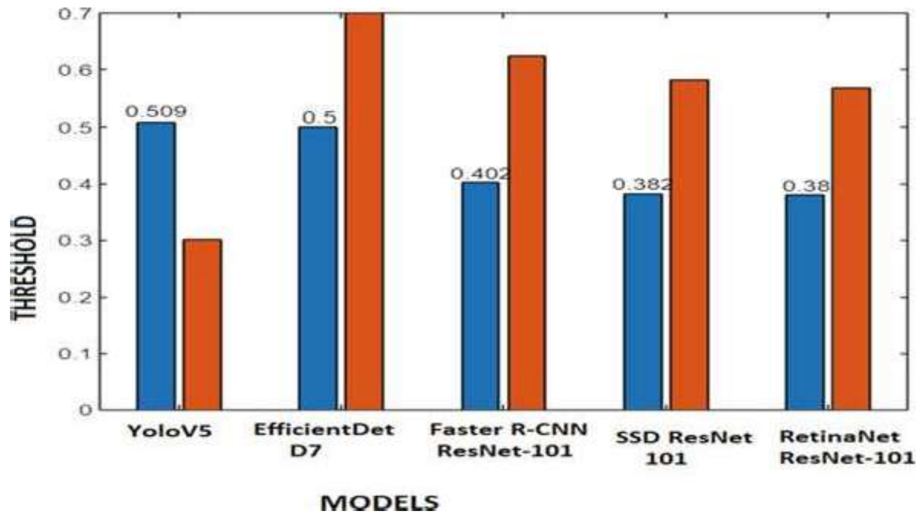


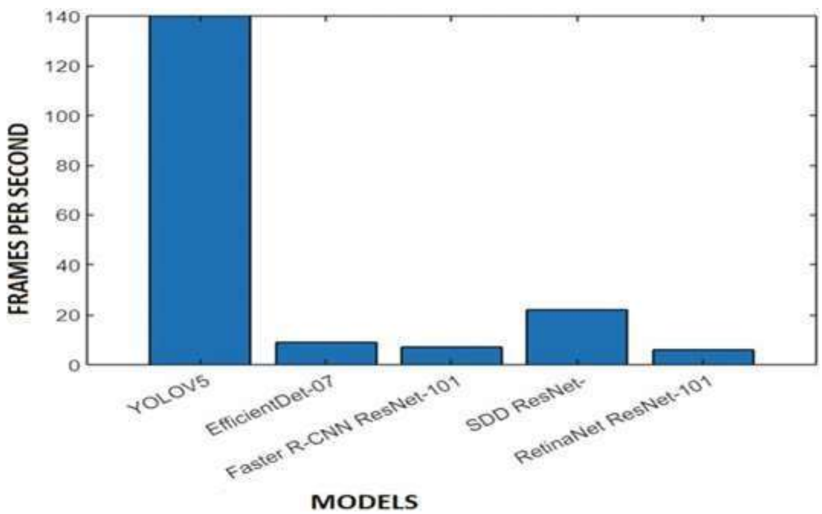
Fig. 5 Mobile and human being spotting

Fig. 6 Water bottle and human being spotting

Fig. 7 Wrist watch and human being spotting



Graph 1 Comparison of different models with threshold



Graph 2 Evaluation of multiple models with frames per second

Table 2 Uses fps to represent the SNA intersection over union

Model	SNA@0.5	SNA@0.5:0.95	FPS
YOLOv5 (proposed system)	0.509	0.301	140
EfficientDet-D7	0.5	0.7	9
Faster R-CNN ResNet-101	0.402	0.624	7
SSD ResNet-101	0.382	0.582	22
Retina Net ResNet-101	0.38	0.567	6

where x denotes the input image, y denotes the output text, and θdenotes the OCR model parameters. An input image and an output text are mapped using the function f.

Many mathematical equations are used in optical character recognition (OCR) to describe the various parts of the OCR system.

Image Preprocessing: To improve picture quality and eliminate noise, OCR begins with preprocessing the input image. Scaling, binarization, and noise reduction are a few of the mathematical procedures employed in picture preprocessing. The following equations can be used to express these operations:

$$I' = \text{scale}(I) \tag{11}$$

$$I' = \text{binarize}(I) \tag{12}$$

$$I' = \text{reduce noise}(I) \tag{13}$$

where I' is the picture that has been preprocessed and I is the input image.

Segmentation: The preprocessed picture is split up into distinct letters or words. The segmentation operation can be represented by the following function:

$$S = \text{segment}\left(I'\right) \tag{14}$$

where S is a set of segmented words or characters.

Feature Extraction: To depict the shape and texture of each segmented character or word, features are taken from it. The following function can be used to symbolize the feature extraction process:

$$F = \text{extract\_features}(S) \tag{15}$$

where F is the collection of feature vectors connected to every segmented word or letter.

Classification: Utilizing a machine learning approach, the attribute points are grouped into appropriate letters and terms. The following function can be used to represent the classification operation:

$$C = \text{classify}(F) \tag{16}$$

where C represents the group of permitted letters or words.

Post-processing: To improve accuracy of identified text and fix errors, the OCR engine's output is post-processed. Spell, grammar, and formatting checks are instances of post-processing operations. The following function can be used to represent the post-processing operation:

$$T' = \text{postprocess}(T) \tag{17}$$

where the corrected text is T′and the recognized text is T.

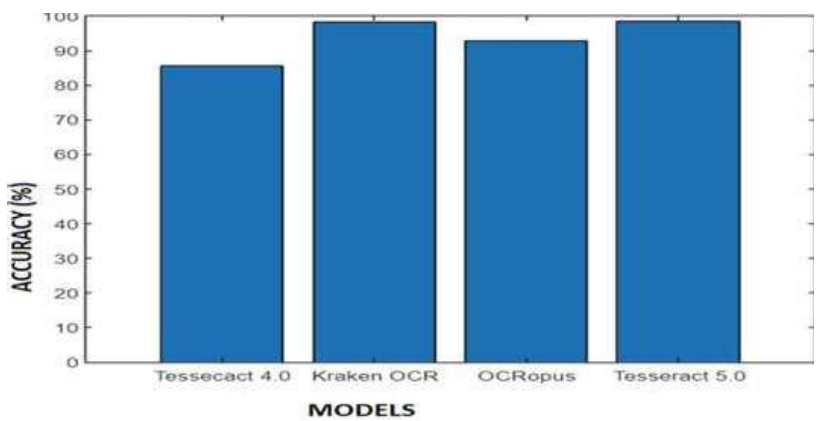
To identify text and quality of the input picture, OCR generally combines post-processing, machine learning, feature extraction, and image processing methods (Figs. 8, 9 and 10).

Here are the comparison of accuracy of the several Text-to-Speech models (Graph 3 and Table 3).

Fig. 8 Assistance of warning signal

Fig. 9 Assistance of road sign posts

Fig. 10 Assistance of public notice boards



Graph 3 The accuracy of various models graphically

Table 3 Picks certain models and compares their accuracy

Model	Accuracy (%)
Tesseract 4.0	86.75
Kraken OCR	97.02
OCRopus	93.56
Tesseract 5.0 (proposed system)	97.98

6 Conclusion

The study's findings imply that the recommended virtual assistant and navigation system might be useful for visually impaired individuals. Because the technology can do exact distance measurement, object detection, and picture processing, it can improve the virtual assistant's efficacy and precision. The virtual assistant can now detect objects thanks to deep neural networks and image processing algorithms, giving the user more freedom and mobility. Thanks to technology developments, image processing methods and virtual assistants may find application in the healthcare and transportation sectors, among other areas.

The OCR technology of the reading aid is based on text recognization in a wide range of fonts, sizes, and languages. Many benefits are offered by text-to-speech engines, includes improving ease of access for those with vision disability, with expanded creativeness when reading long documents, and enhanced interaction with digital content including customer support systems, virtual assistants, and educational materials.

Several datasets and evaluation criteria are useful for the evaluation of the system's execution; the outcome reveals high object identification, text extraction, and conversion efficiency. The system's auditory signals which include buzzer noises and visual augmentation techniques, further improve usability and efficiency for those with varying levels of vision impairment.

Overall, as it enables visually impaired individuals a fundamental, movable, hands-free ETA instance for regular, day-to-day interior and exterior usage, the recommended virtual assistant and navigation system has an ideal remedy for various problems to the persons with destitute of vision which they frequently face. Furthermore, it may greatly improve their level of living.

References

1. Sáez Y, Montes H, Garcia A, Muñoz J, Collado E, Mendoza R (2021) Indoor navigation technologies based on RFID systems to assist visually impaired people: a review and a proposal. IEEE Lat Am Trans 19(8):1286–1298. https://doi.org/10.1109/TLA.2021.9475859
2. Kandalan RN, Namuduri K (2020) Techniques for constructing indoor navigation systems for the visually impaired: a review. IEEE Trans Human-Mach Syst 50(6):492–506. https://doi.org/10.1109/THMS.2020.3016051
3. Xiao J, Joseph SL, Zhang X, Li B, Li X, Zhang J (2015) An assistive navigation framework for the visually impaired. IEEE Trans Hum Mach Syst 45(5):635–640. https://doi.org/10.1109/THMS.2014.2382570
4. Hunaiti Z, Garaj V, Balachandran W (2009) An assessment of a mobile communication link for a system to navigate visually impaired people. IEEE Trans Instrum Meas 58(9):3263–3268. https://doi.org/10.1109/TIM.2009.2022374
5. Katzschmann RK, Araki B, Rus D (2018) Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. IEEE Trans Neural Syst Rehabil Eng 26(3):583–593. https://doi.org/10.1109/TNSRE.2018.2800665
6. Barontini F, Catalano MG, Pallottino L, Leporini B, Bianchi M (2021) Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: a user-centered approach. IEEE Trans Haptics 14(1):109–122. https://doi.org/10.1109/TOH.2020.2996748
7. Aladrén, López-Nicolás G, Puig L, Guerrero JJ (2016) Navigation assistance for the visually impaired using RGB-D sensor with range expansion. IEEE Syst J 10(3):922–932. https://doi.org/10.1109/JSYST.2014.2320639
8. Ando T, Tsukahara R, Seki M, Fujie MG (2012) A haptic interface "force blinker 2" for navigation of the visually impaired. IEEE Trans Industr Electron 59(11):4112–4119. https://doi.org/10.1109/TIE.2011.2173894

9. Duh P-J, Sung Y-C, Chiang L-YF, Chang Y-J, Chen K-W (2021) V-Eye: a vision-based navigation system for the visually impaired. IEEE Trans Multimedia 23:1567–1580. https://doi.org/10.1109/TMM.2020.3001500
10. Khan S, Nazir S, Khan HU (2021) Analysis of navigation assistants for blind and visually impaired people: a systematic review. IEEE Access 9:26712–26734. https://doi.org/10.1109/ACCESS.2021.3052415
11. Katzschmann RK, Araki B, Rus D (2018) Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. IEEE Trans Neural Syst Rehabil Eng 26(3):583–593