IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TEXT/OBJECT RECOGNITION FOR VISUALLY IMPAIRED PEOPLE

R. KUMARAN1, M. BRINDHA2, P. HARUNYAA3, S. ARCHANA4
1 ASSOCIATE PROFESSOR, 2 ,3,4 UG STUDENT
INFORMATION TECHNOLOGY
V.S.B ENGINEERING COLLEGE, TAMILNADU, INDIA

ABSTRACT

Speech and text is the main medium for human communication. A person needs vision to access the information in a text. However, those who have poor vision can gather information from voice. This project has been built by using Node MCU microcontroller. It is controlling the peripherals like Camera and headset which act as an interface between the system and the user. Optical Character Recognitions (OCR) is implemented in this project to recognize characters which are then read out by the system through a headset. The camera is mounted on spectacles/head device; it captures a full view of the paper into the system. it is ensured that there are good lighting conditions. The content on the paper should be written in English and be of good font size. When all these conditions are met the system takes the photo, processes it and if it recognizes the content written on the paper it will announce on the Bluetooth headset speaker that the content on the paper has been successfully processed. After this it speaks out the content that was converted in to text format in the system from processing the image of the paper. In this proposed system, PC Based Reader for Blind helps a blind person to read a paper without the help of any human reader. It also achieves sound bite hearing system for both blind and deaf peoples, we will hear using vibrator when we bite it. It uses the Optical character recognition technology for the identification of the printed characters using image sensing devices and computer programming. It converts images of typed or printed text into machine encoded text. In this research these images are converted into the audio output (Speech) through the use of OCR and Text-to-speech synthesis. The conversion of printed document into text files is done using PC which again uses PyTesseract library and Python programming. The text files are processed & convert into the audio output (Speech) using GOOGLE Text-to-speech (gTTS) & python programming language and audio output is achieved. For visually and hearing impaired person, they can hear sound by using soundbite hearing system using vibrator. Object information using a deep learning object recognition technique. Object recognition algorithms are designed based on the You Only Look Once architecture, an object recognition deep learning model, to detect objects using a camera. In this dissertation, system analyzes accurate object information and obtain a location using a deep learning object recognition technique.

Keywords: MCU microcontroller, neural network and Optical Character Recognition

1. INTRODUCTION

According to the World Health organization (WHO), 285 million people are estimated to be visually impaired worldwide among which 90% live in developing countries [1] and forty-five million blind individuals world-wide [2]. Though there are many existing solutions to the problem of assisting individuals who are blind to read, however none of them provide a reading experience that in any way parallels that of the sighted population. In particular, there is a need for a portable text reader that is affordable and readily available to the blind community. In our world information is generally available in the form of books and documents. It is fully usable for the sighted people. From an ancient time, information is resembled in aural format as no other representation of it is founded in printing format. When an era has come of printing it facilitates the sighted people partially to acquire knowledge. A major problem for a blind or visually impaired person (BVI) to interact with the world to share knowledge. For them information has to be in a special tactile language or in voice format. The solution is rather simple; introduce a smart device with a multimodal system that can convert any document to the interpreted form to a blind. A blind can read document only by tapping words which is then audibly presented through text to speech engine. "Blind Reader" – developed for touch devices which is user friendly and effective interactive system for visionless or low vision people.

Visually challenged people and uneducated people face a lot of adverse challenges in their day to day life. Most of the time they are perplexed in a new environment or surrounding due to issues related to accessibility. So, this prevents them from experiencing the world in the same way as others do. Identifying and accessing things is something many of us may take it for granted. But the visually challenged people are curbed by their disability. Especially in a medicine taking scenario, it is difficult for them to find whether they have identified the medicine correctly or not. They will have to seek others help for it. Moreover, a mobile application will be easy to use and the hardware needed is very limited. In today's world, technology is growing at an alarming rate. It has found its way in every field of our life. But this technology is of no use if it couldn't provide itself to the aid of the disabled people. From the past few years, Mobile Phones have become a main source of communication for this digitalized society. We can make calls and text messages from a source to a destination easily. It is known that verbal communication is the most appropriate modem of passing on and conceiving the correct information. To help the people more effectively, engage with local and/or remote services text-to-speech (TTS) were first developed to aid the visually impaired by offering a computer-generated spoken voice that would "read" text to the user. In this project, we will take a look at text to speech conversion. Using Optical Character Recognition this kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction from color images is a challenging task in computer vision. Visually challenged people and uneducated people face a lot of adverse challenges in their day to day life. Most of the time they are perplexed in a new environment or surrounding due to issues related to accessibility. So, this prevents them from experiencing the world in the same way as others do. Identifying and accessing things is something many of us may take it for granted. But the visually challenged people are curbed by their disability. Especially in a medicine taking scenario, it is difficult for them to find whether they have identified the medicine correctly or not. They will have to seek others help for it.

2. PROBLEM STATEMENT

India is home to the largest number of visually impaired people in the world, about 40 million, which accounts for 20% of the world's blind population. Moreover, more than 90% of these people have little to no access to the necessary assistive technologies.

Independent travel is a well-known challenge for blind or visually impaired persons and also the increasing availability of cost efficiency, high performance and portable digital imaging devices.

Speech and text is the main medium for human communication.

A person needs vision to access the information in a text. However, those who have poor vision can gather information from voice.

The blind people and illiterates are facing difficulty in understanding the content they have.

This leads to manipulations and scams and health issues.

There is no technology avail till date for blind and deaf to hear voice/sound

3. OBJECTIVE

To detect the text from any document, it can recognize any signboard, recognize medicine name/text book

To extract and recognize text from scene images effectively using computer vision technology and to convert recognized text into speech so that it can be incorporated with hardware to develop Electronic Travel aid for visually impaired people.

For blind and deaf people, this system provides audio output by using OCR and soundbite hearing system

4.LITERATURE SURVEY

Xi Hai Xie Chen Zhao Fan et all proposed "Varible Step Size Multi-Layer Neural Network Blind Equalization Algorithm" IEEE-2022

In communication systems, they have developed the ISI (Inter-Symbol Interference) and ICI (Inter-Carrier Interference) seriously affect the quality of communication. Blind equalization techniques can effectively solve this problem. In this paper, the fixed step size multilayer neural network blind equalization algorithm is introduced. The steady-state error and convergence speed of the blind equalization algorithm for fixed step multilayer neural network are limited. To solve this problem, an adjustable step size blind equalization algorithm for multilayer neural network is proposed, it is mainly based on the blind equilibrium algorithm of multilayer neural network, and the convergence speed and convergence accuracy of the system are improved by adding the variable step size algorithm and adjusting the parameters of the network structure, which increases the step size at the beginning to speed up the convergence speed and decreases the step to improve the convergence. Simulation results show that the improved algorithm has faster convergence speed and higher convergence accuracy.

N. Yu. Liberovskiy V. S. Priputin et all proposed "Fourth Cumulant Blind Source Separation Efficiency Evaluation in the Task of Cognitive Radio" IEEE-2022

They have developed the Blind source separation algorithms as a part of cognitive radio can be actively used in the task of building a smart transport system with a large number of licensed and unlicensed users. In this paper, the efficiency of the algorithm for blind separation of two complex signals is evaluated. The algorithm uses as a criterion for source separating the system of equations, which simultaneously nullify output signals covariance and fourth-order mixed cumulant. Unlike iterative methods, the considered algorithm is performed in a finite number of arithmetic operations. The paper investigates the similarity limit of input signals, separation efficiency depending on the sample size and signal-to-noise ratio. It is shown that the proposed algorithm makes it possible to effectively separate linear combinations of independent signals with a difference in the signal-to-interference ration in the input signals of at least 1 dB. It is shown that the proposed algorithm performs efficient separation of signals when the size of the sample used to calculate the statistics of the second and fourth orders is at least 10000. Compared to the Fast ICA algorithm, the proposed algorithm requires three times less samples to detect FSK-2 signals. It is shown that the proposed algorithm performs effective signal separation when the signal-to-noise ratio of the input signals is at least 24 dB.

Minsik Kim Alan E. Willner et all proposed "Turbulence Resilient Free-Space Optical Communication Using Iterative Blind Equalization"IEEE-2022

They have developed in the free-space optical communication, the orthogonal multiple data-carrying beams can increase the transmission rate, but there are inherent cross-talk and inter-symbol interference due to atmospheric turbulence. The conventional blind equalization can mitigate interference, but the equalizer output symbols still have residual interference in the case of strong turbulence. To improve the error performance, we consider an iterative blind equalization that finds the channel response and the demodulated symbols. Numerical results show that the iterative blind equalization significantly improves the bit error performance.

Soon-Young Kwon Ji-Hyeon et all proposed "Kim SVR-based Blind Equalization on HF Channels with a Doppler Spread" IEEE-2022

They have developed A transmission signal through a high-frequency (HF) channel is usually reflected by the ionospheric layers and become a multipath signal, resulting in inter-symbol interference (ISI). To remove ISI, a receiver recovers the multipath-faded signal by using channel equalization. Among various channel equalization methods, blind equalization that does not use training sequences draws an interest because it may increase bandwidth efficiency. The HF signal needs to be equalized with a small number of symbols due to a Doppler spread. Therefore, to equalize the HF channel signal, a batch method based on support vector regression (SVR) can be used. In this respect, we applied an SVR-based batch blind equalization to HF channels and then analyzed its performance.

Liang Wang Hewen Wen et all proposed "Soft Decision Adjusted Modulus Algorithm for Blind Equalization" IEEE-2022

They have developed the Enhanced decision-adjusted modulus algorithm (EDAMA) is excellent to reduce the residual errors of constant modulus algorithm (CMA) for blind equalization problems. However, EDAMA is unsuitable when processing time-varying channel and high order quadrature amplitude modulation (QAM) signals even for low ISI levels due to slow convergence rate and high misestimation. To improve convergence rate and reduce misestimation while maintaining low residual error, a new blind equalization algorithm is proposed by introducing soft decision algorithm into EDAMA. The simulation results show that the new algorithm has a better convergence rate than EDAMA.

5. SYSTEM DESIGN

5.1 EXISTING SYSTEM

In this existing system, they have provided a detailed description of this system IPCD dataset. The primary objective of this work is to propose a BVIP compatible and efficient automated system for recognizing Indian paper currency denominations. The underlying model requires training on a range of estimable currency images with varied diversity and classes to build an effective automated currency recognition framework. The existing system IPCD dataset consists of a wide variety of currency images, including new denominations and old denominations of 10, 20, 50, and 100 banknotes. In addition to this, the new set of 500 denominations, including the newly introduced 200 and 2000 denominations, are also included in the dataset. The existing dataset is among the most diverse datasets in terms of number of images, denomination classes, illuminations, and background variations. The training images should be composed of real-life BVIP usage scenarios to develop an efficient and generalizable network. However, this critical aspect is often overlooked in existing Indian currency datasets. A brief comparison of existing datasets. Using smaller datasets to train and evaluate the currency, classification approaches steer to non-viable and biased processes. Even the recent datasets involve lesser images as well as lack domain-specific scenarios creating vagueness about the viability of solutions.

5.2 PROPOSED SYSTEM

This proposed system consists of two modes of operation for visually impaired person and hearing impaired person. For visually impaired person, the proposed system consists of two main modules the image processing and voice processing module. The image processing module captures image using camera converting the image into text. Voice processing module changes the text into sound and processes it with specific characteristics so that the sound can be understood. At first the image processing module, where OCR converts .jpg to .txt form 2nd is voice processing module which converts .txt to speech OCR is very important element in this module OCR or Optical Character Recognition is a technology that automatically recognizes the character through the optical mechanism, this technology intimates the ability of the human sense of sight. Before feeding the image to OCR it is converted to binary image to increase the image recognition accuracy. The proposed method is to help blind person in reading the text present on the text labels, printed notes and products as a camera based assistive text reader. The implemented idea involves text recognition from image taken by camera on spectacle and recognizes the text using OCR. Conversion of the recognized text file to voice output by eSpeak algorithm. A prototype was developed which uses a camera, PC and ESP 8266-12E NODE MCU controller that

works in real time. The system captures the frame and checks the presence of text in the frame. The captured image is first converted to gray scale and then filtered using a Gaussian filter to reduce the noise in the image. Here adaptive Gaussian thresholding is used to reduce the noise in the image. The filtered image is then converted to binary. The binarized image is cropped so that the portions of the image with no characters are removed. The cropped frame is loaded to the Tesseract OCR so as to perform text recognition. The output of the Tesseract OCR will be text file which will be the input of the e-Speak. The e-Speak creates an analog signal corresponding to the text file given as the input. The analog signal produced by the e-Speak is then given to a headphone to get the audio output signal. For visually impaired person and hearing impaired person, Optical Character Recognition (OCR), Text to Speech Synthesis (TTS) and soundbite hearing system are integrated, which enables the blind and deaf person to hear an audio output through vibration motor. Soundbite hearing system processes sound and wirelessly transmit the sound signals via the teeth. It is an alternative to surgical bone conduction which require surgical implantation into the skull conduct sound Soundbite sensor uses the teeth instead of the implanted component and eliminates the need for surgery. Conventional hearing aid which an amplify sound can cause distortion for these patient. This sound vibration travel through the medium and sound is heard when sound waves travel through medium of bones to arrive at the inner ear it transmits suitable vibration. Sound waves arriving to the head can cause the skull bones to vibrate. These vibrations can be transmitted to the inner ear through the temporal and frontal bones as well as through the jaw and soft tissue. The receive signals from speaker convert them into sound vibrations although the vibrations are strong enough too picked up cochlea with help of amplifier, driver relay circuit and vibrator. This way, the sound is transported from your impaired ear directly to your hearing ear. This hearing device will be fitted to the upper left or right teeth in the back of your mouth. This doesn't require any of your teeth to be altered, and the device can be inserted and removed easily. This hearing device is a flat piece (in Rea l-Time Product) that contains a sealed rechargeable battery, and electronics and wireless capabilities that can pick up sound transmissions from the behind-the-ear microphone. The system is good for portability. The portability allows the user to carry the device anywhere and can use at any time.

6. HARDWARE DESCRIPTION

Power Supply
Transformer
Bridge Rectifier
Regulator
Slide Switch
Driver Circuit
Darlington Pair
Relays
Vibration Motor

7. SOFTWARE DESCRIPTION

Sketch Ide – ESP Node MCU Programming Idle

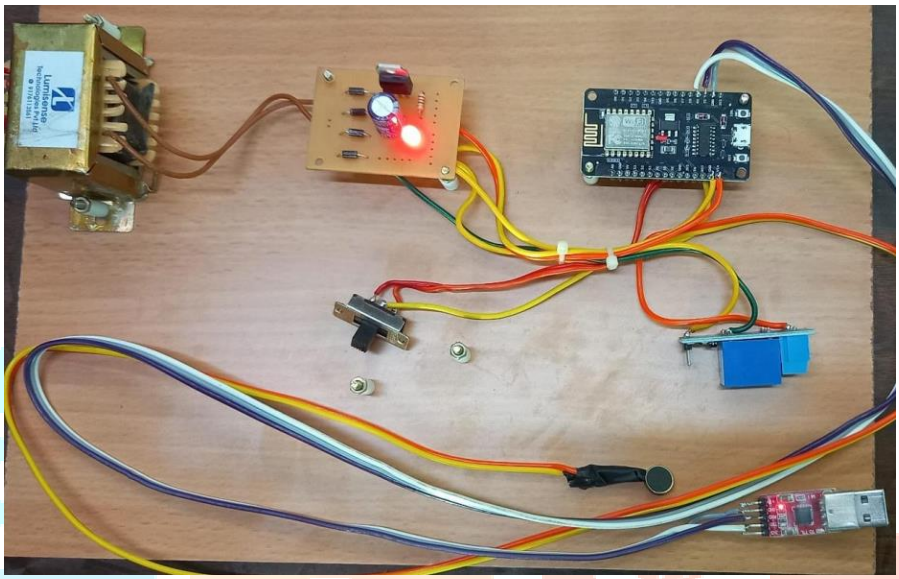
Arduino

Python

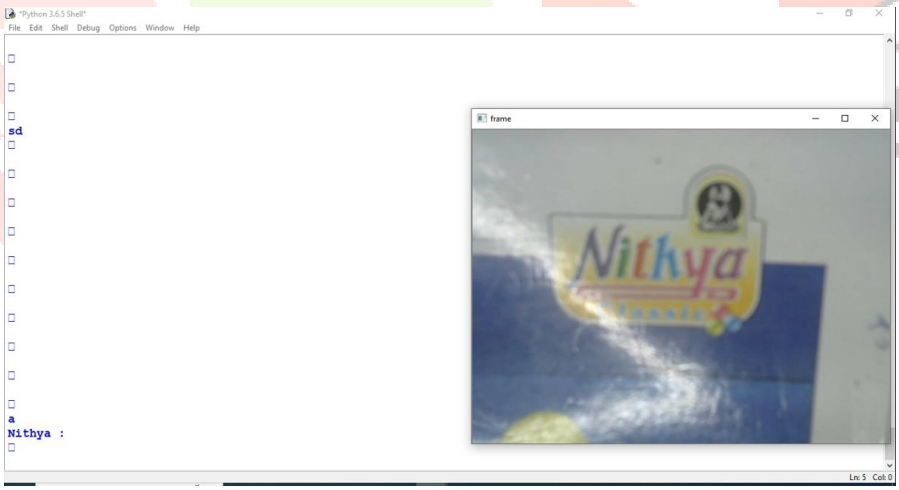
8. RESULTS AND DISCUSSION

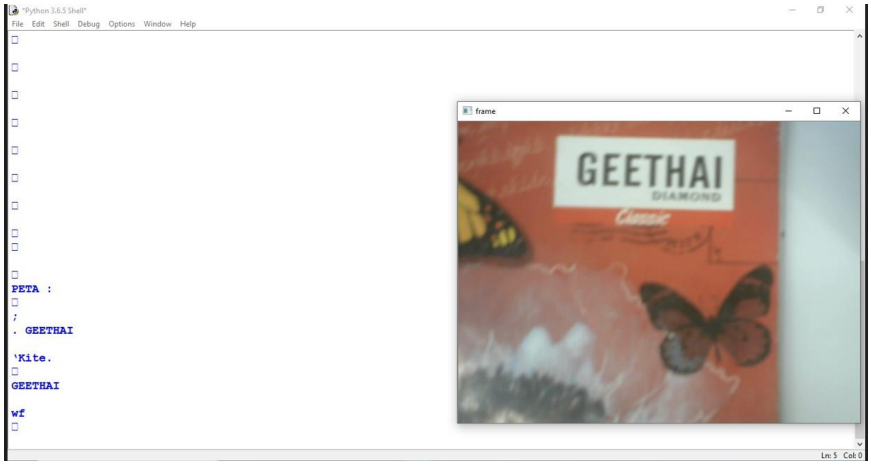
The image which contains the text is placed at a considerable distance from the camera so that the image is clear enough with proper illumination. The Fig. 13 shows the original captured image by the camera. The original image is put through certain image processing techniques in order to obtain an enhanced image for accurate text extraction. The Tesseract OCR engine then converts the enhanced image into machine readable text and stores it in .txt file. The Fig. 14 shows the converted text in .txt file. The text stored in the .txt file is converted into an audio file using Python based TTS Synthesizer. The Speaker or Microphone is employed to read the audio file. The results obtained from this system indicates the image captured using the camera on spectacle, the preprocessed image which is given to tesseract OCR engine to extract the text from the image and

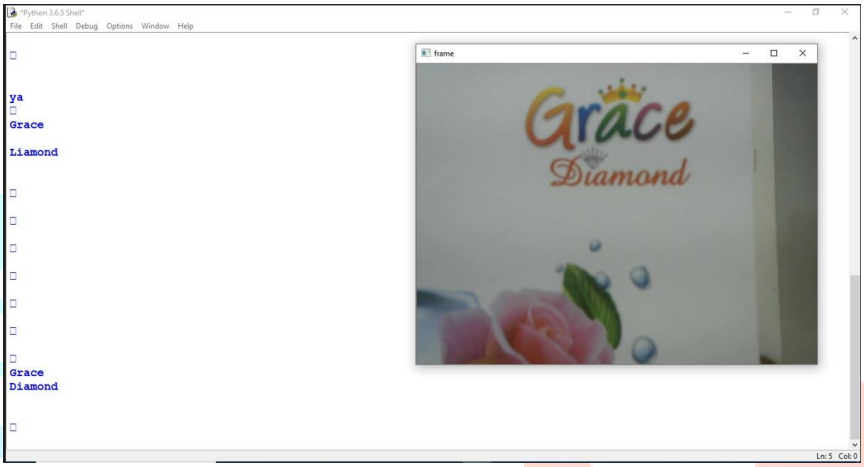
the output from the tesseract OCR engine. The accuracy can be improved by making use of a HD auto focus camera. Bone conduction devices are used in wide range of applications such as communication system, language development approach, audiometric investigation and finally most importantly in hearing rehabilitation. This review is focused on hearing rehabilitation, where the common indications are mixed hearing loss and also single sided deafness. These bone conduction devices can be semi- implantable, where some part of the devices is implanted. Yolo detected object in this bounding boxes are equal to the real boxes of the objects. The final detection will consist of unique bounding boxes that fit the objects perfectly detect the name of object are (Person, dog, chair, bottle) in this voice to headset.



EXPERIMENTAL SETUP

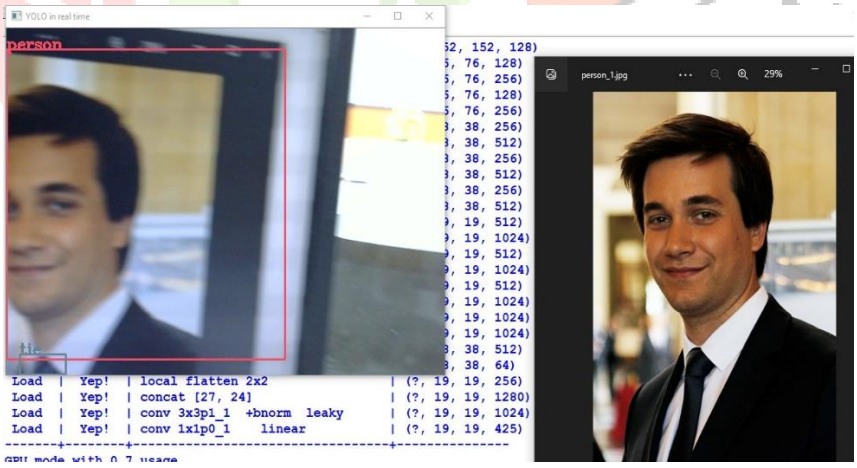




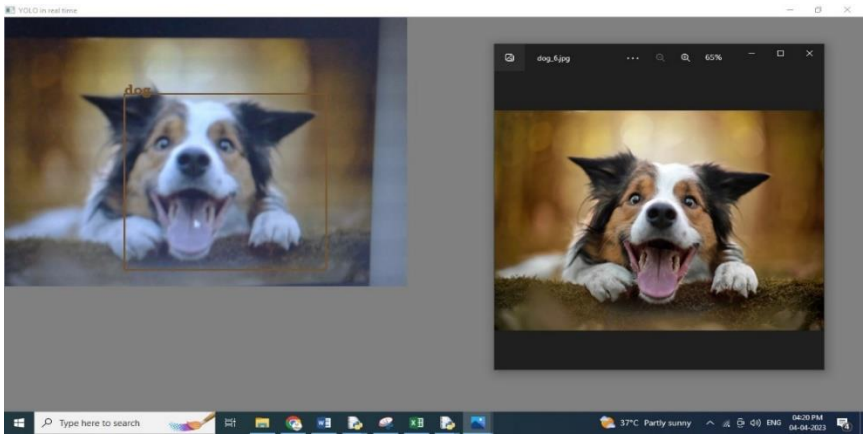


RECOGNIZED THE CHARACTER USING OPTICAL CHARACTER RECOGNIZATION

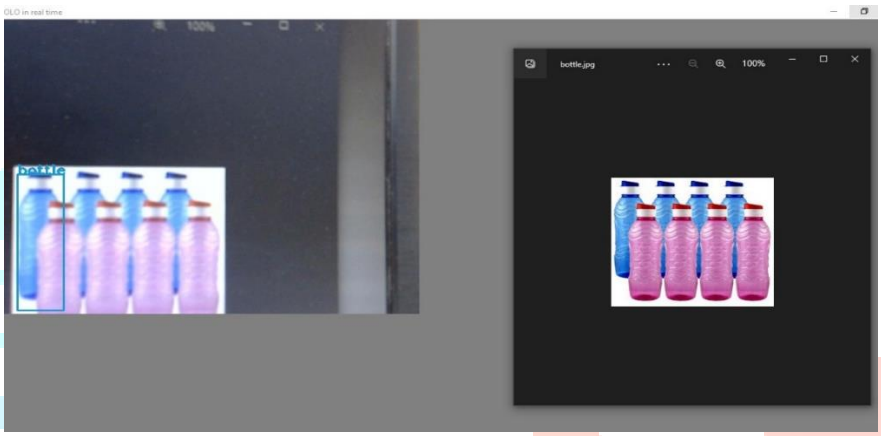
8.1 YOLO (OBJECT DETECTION)



PERSON DETECTED



DOG DETECTED



BOTTLE DETECTED

9. CONCLUSION

Thus this paper proposes an intelligent text reader based on Python which converts the text present in an image into speech efficiently. The product delivers a mobile and inexpensive way of converting text in an image to speech. Since certain image processing techniques are used on the original image, it results in an enhanced and an improved quality image, which is then converted into text using Tesseract OCR engine. The converted text is thus 99% accurate. Further, the text is converted to an audio file by making use of a Python based TTS Synthesizer which results in a natural and pleasant audio output. Thus this device guarantees high precision and it is also consistent. Presently, the device is compatible with English language. Future research can focus on making it multilingual so that it can work well in different languages too, also to add other functions such as voice changing option and provide an alert whenever a bad picture is taken. Today and in coming future there will be huge demand of TTS and audio assistance. In this paper we have attempted to extract text from text documents, images as well as handwritten text Also the model works with satisfactory accuracy. By this approach text and images from a word document, Web page or e-Book can be read and can generate synthesized speech through a computer's speakers. We have developed a soundbite hearing system for hearing impairment person. We used the bone conduction principle which provide hearing support to the patients. We implemented signal processing strategy for restoration of cross-channel suppression in hearing impaired listeners. Sound bite hearing system is a non-surgical bone conduction prosthetic device that transmits sound via the teeth. It is very useful for hearing disability person. This device is sure to help the deaf people to communicate effectively in both direct and mobile communication. This paper shows the implementation of the project 'Reader afor Blind'. It was developed to aid blind in everyday life and help them to be independent. The project aimed to cover a broader aspect of life and hence we incorporated both the parts into one. This project aims to assist the blind people in reading the printed text on pamphlets, books, magazines and other printed material. One can be assisted in reading their everyday newspaper with the help of this device. Our project also achieves sound bite hearing system for both blind and deaf people. The YOLO (You Only Look Once) algorithm is a popular object detection algorithm that operates by dividing an input image into a grid of cells and predicting bounding boxes, class

probabilities, and objectless scores for each cell. This allows YOLO to detect objects in real-time with high accuracy and speed.

10.REFERENCES

[1] Pandian, A. P. (2019),” Artificial Intelligence Application in Smart Warehousing Environment for Automated Logistics”, Journal of Artificial Intelligence, 1(02), 63-72.
[2] Nirmal, D. "Artificial Intelligence Based Distribution System Management and Control. “Journal of Electronics 2, no. 02 (2020): 137- 147.
[3] V. Veeramsetty, G. Singal, and T. Badal, ‘‘CoinNet: Platform independent application to recognize Indian currency notes using deep learning techniques,’’ Multimedia Tools Appl., vol. 79, pp. 22569–22594, Aug. 2020.
[4] W. Sun, X. Zhang, and X. He, ‘‘Lightweight image classifier using dilated and depth wise separable convolutions,’’ J. Cloud Comput., vol. 9, no. 1, pp. 1–12, 2020
[5] Ms. Kavya. S, Ms. Swathi, Mrs. Mimitha Shetty, 2019, Assistance System for Visually Impaired using AI, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) RTESIT – 2019 (VOLUME 7 – ISSUE 08.
[6] V. Mane. (Sep. 2020). Indian Currency Notes. [Online]. Available: https://www.kaggle.com/vishalmane109/indian-currency-note-imagesdataset-2020
[7] M. Han and J. Kim, ‘‘Joint banknote recognition and counterfeit detection using explainable artificial intelligence,’’ Sensors, vol. 19, no. 16, p. 3607, 2019.
[8] “WLU6331 Wi-Fi Adapter RF Exposure Info (SAR) Raspberry Pi Trading” Dec. 24, 2014. Accessed on: Aug. 1, 2020. [Online]. Available: https://fccid.io/2ABCB-WLU6331/RF-Exposure-Info/RFExposure-Info-SAR-pdf-2489354
[9] C. Park, S. W. Cho, N. R. Baek, J. Choi, and K. R. Park, ‘‘Deep feature based three-stage detection of banknotes and coins for assisting visually impaired people,’’ IEEE Access, vol. 8, pp. 184598–184613, 2020.
[10] T. D. Pham, C. Park, D. T. Nguyen, G. Batchuluun, and K. R. Park, ‘‘Deep learning-based fake-banknote detection for the visually impaired people using visible-light images captured by smartphone cameras,’’ IEEE Access, vol. 8, pp. 63144–63161, 2020.