ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

Assistive Audio Vision for Visually Impaired: Using Machine Learning and OCR (Optical Character Recognition)

[1] Avantik Tiwari, [2] Avarna Gupta, [3] Vaishali Deshwal

[1] [2] [3] Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India
Corresponding Author Email: [1] avantik.tiwari.cs.2019@miet.ac.in, [2] avarna.gupta.cs.2019@miet.ac.in

Abstract*— In this paper, we propose a multi-task end to end optimised system which works on the object detection and OCR (Optical Character Recognition) models of machine learning and by using them our aim is to provide assistance to visually impaired people. Unlike traditional methods, used for training we do not create a dataset we used the already trained model for object detection and for OCR we used Tesseract tool which can convert the text in images into an actual text and then we convert text fetched from object detection model and OCR model into and audio. In first step of training, we used coco dataset for object detect model and to implement the process of object detection we used fast-R CNN and YOLOv5 which are the region-based object detection algorithms and by using these we can also train and test the model on our custom dataset also. In the second step, the output image given by the object detection model goes to the OCR process in which the text which shows what is present in the image is converted into an actual text using Tesseract tool and then that text is converted into the audio.

***Index Terms*—Assistive audio vision, optical character recognition, multi-task learning, optimised system for blinds.**

I. PROBLEM STATEMENT

Assistive audio vision for visually impaired is a project in which we as a contributor of this project aims to make visually impaired people independent to visualize the beauty of nature and surroundings and to read the letters, texts and paragraphs which are present on their computer screens or mobile devices as these letters cannot be touched by them so to make these easier our idea is to convert those textual words in the images into the audio format so that they can understand what was written in that image similarly if we have a picture in which some objects our present then our project help them to understand what is actually in that picture by finding the objects in that picture and then convert them into audio format.

The blindness in people or visual-impairment people in the world today is at least 2.2 billion reported by the WHO (World Health Organization) on 8 October 2019.Object detection and image processing are the popular techniques and which we are using in our project.

In this research we try to build an efficient model which can work as a guide for visually impaired people to analyze their indoor and outdoor environment, the prototype of the system is a software which takes images and outputs the audio which either read the text present in the image or describing the environment present in the image. Our paper is organized as follows: first is the introduction, literature review is presented in section 2, followed by the information about used datasets & research method in section 3after that results and discussio in section 4, and finally the conclusion and details of future works in section 5.

II. LITERATURE REVIEW

Assistive technologies for the visually impaired is a research field that is gaining increasing prominence owing to an explosion of new interest in it from disciplines. The idea of using an assistive technology is described in the paper by Alexy Bhowmick & Shyamanta M.Hazarika in which the description of growth of assistance for person with visually impairments. Various articles on mobile assistive technologies for the visually impaired were identified using a multi-layered systematic approach in our literature search. During the first stage, a Scopus search (covering databases including Web of Science as well as publishers including Elsevier and Springer) for the period up until December 2011 was carried out using various combinations of the following search terms: *vision loss, visual impairment, low vision, blind, assistive technology, IT.*

Mobile assistive technologies for visually impaired person, published by the LilitHakobyanBSc (Hons), JoLumsdenBSc (Hons), PhD Dympna O'Sullivan BSc (Hons), PhD Hannah Bartlett BSc (Hons) and MCOptom, PhD in which they provide a comprehensive overview of research and innovation to support people with visual impairment using the field of mobile assistive technologies. In their research they aimed at providing mobile phone assistance to visually impaired persons to lead more independent lives.

Wearable assistive technological devices for visually impaired, in this paper Ruxandra Tapu Bogdan Mocanu Titus Zaharia proposed portable and wearable assistive device to blind and visually impaired persons, which provide them a deep analysis of related strengths and limitations.

Assistive Technologies for visually impaired people using computer vision by Shankar Sivan and Gopu Darsan, in their research they describe how a computer vision is a developing area to create assistive technologies for the visually impaired individuals. They propose a system which incorporates various assistance features in a device which will act according to the need of a visually impaired person and which can be access by the common people, so their idea is to create a inexpensive assistive device.

III. DATASETS AND RESEARCH METHODS

The techniques which we are using to train our model is divided into two parts: Train the object detection model on coco dataset which help us to detect the objects in any test image for which we are using darknet code to create a weights file which help us to create our own object detector that used in Yolov5 model to detect the object present in the image. To create our custom object detector we require at least 4 GB GPU , so we build the object detector using darknet on Google Colaboratory: Colab 12 GB-RAM GPU.

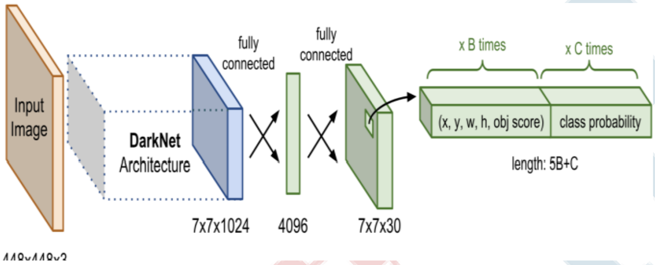
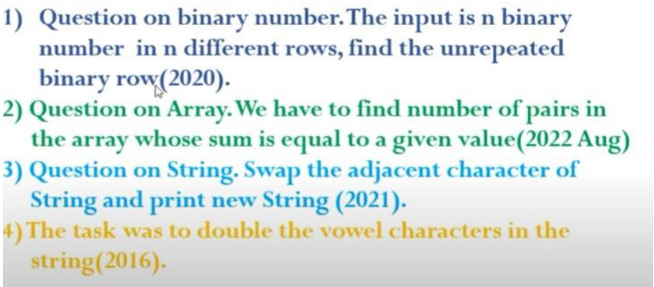


Figure 1. The Algorithm of Yolov5

To train the model we require the data that should be in a Yolo format and for this we used labelImg tool to annotate the data for our custom classes, it is beneficial for us because with the help of these we can add as many classes as possible in our current dataset. The second technique which we are using is to convert the text in an image to an actual text and for this we create a custom function using tesseract tool which can analyze the text present in the image. As our model is built for analyze the two kinds of images: The one in which objects arre present and the one in which the textual content is present, so we require above two techniques to complete our task and only require coco dataset for object detection and we also create a skewing function so that the machine can apply OCR(Optical Character Recognition) without showing any error in the console. Finally we, are using an API(Application Programming Interface) by google known as gTTS(Google Text-to-Speech) which converts our text into an A.I audio.

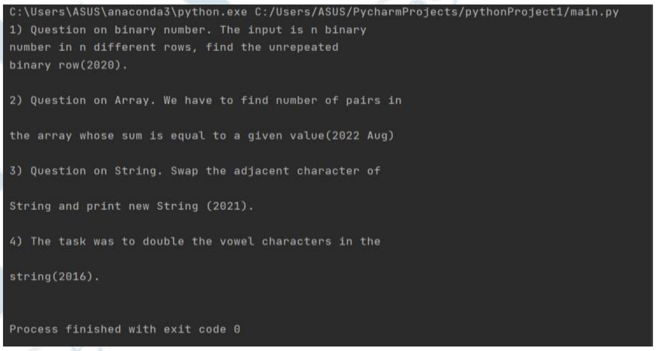
IV. RESULTS AND DISCUSSION

From our proposed methodology explained above, after creating our own custom object detector using darknet and then train our Yolov5 model with the coco dataset, we get an object detector which can convert the text present in the image into a speech format and can also recognized the object present in the image.We can't show the audio output in this paper but we can show the output of object detector and our OCR(Optical Character Recognition) model.



1) Question on binary number.The input is n binary number in n different rows, find the unrepeated binary row(2020).
2) Question on Array.We have to find number of pairs in the array whose sum is equal to a given value(2022 Aug)
3) Question on String. Swap the adjacent character of String and print new String (2021).
4) The task was to double the vowel characters in the string(2016).

Figure 2. The test image



```
C:\Users\ASUS\anaconda3\python.exe C:/Users/ASUS/PycharmProjects/pythonProject1/main.py
1) Question on binary number. The input is n binary
number in n different rows, find the unrepeated
binary row(2020).

2) Question on Array. We have to find number of pairs in

the array whose sum is equal to a given value(2022 Aug)

3) Question on String. Swap the adjacent character of

String and print new String (2021).

4) The task was to double the vowel characters in the

string(2016).


Process finished with exit code 0
```

Figure 3. The output in console

These are the example of the OCR model which we used to convert the text present i the image into an actual text and after that it will be converted into speech so that a person who are visually impaired can understand, what is typed in that image.

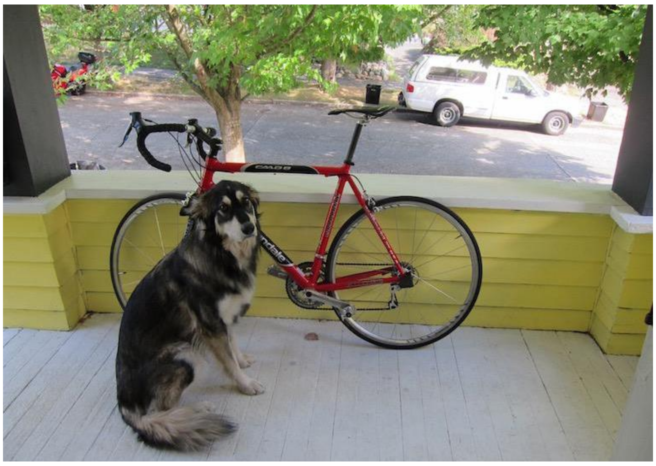


Figure 4. Image in real situation

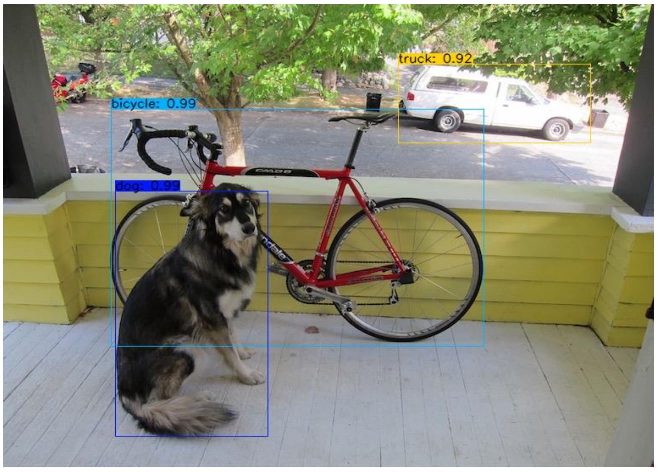


Figure 5. Output of the object detection model

V. CONCLUSION AND FUTURE WORKS

In this paper, we described about our project which consists of an object detector model, OCR model and a text to speech API which combine to form an assistive audio vision for visually impaired people. For creating and implementing these three models we use tools which includes Yolov5, labelImg, darknet, pytesseract, tensorflow and gTTS. Our project can be used for the images which have text present in it and images in which only objects are present and it can be work as a assistive tool for visually impaired persons and help them to enjoy listen their text messages and also help them to enjoy their surroundings. In our future works, we try to run this model on the voice command by the user, so that it can become more easy for any visually impaired person to use this model as an app or as wearable glasses.

REFERENCES

[1] World Health Organization, “Blindness and vision impairment,” [Online], Available: https://www.who.int/newsroom/fact-sheets/detail/blindness-and-visual-impairment.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. of the IEEE, vol. 86, no. 11, 1998, pp. 2278-2324, doi: 10.1109/5.726791.

[3] Viola, P., and Jones, M. J., "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137-154, 2004.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84-90, May 2017, doi: 10.1145/3065386.

[5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D et al., "Going deeper with convolutions, " in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[8] Martinez-Martin, Ester, Felix Escalona, and Miguel Cazorla. "Socially assistive robots for older adults and people with autism: An overview." Electronics 9, no. 2 (2020): 367.

[9] Gong, Hang, Tingkui Mu, Qiuxia Li, Haishan Dai, Chunlai Li, Zhiping He, Wenjing Wang et al. "Swin-Transformer-Enabled YOLOv5 with Attention Mechanism for Small Object Detection on Satellite Images." *Remote Sensing* 14, no. 12 (2022): 2861.

[10] Hakobyan, Lilit, Jo Lumsden, Dympna O’Sullivan, and Hannah Bartlett. "Mobile assistive technologies for the visually impaired." *Survey of ophthalmology* 58, no. 6 (2013): 513-528.

[11] Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia. "Wearable assistive devices for visually impaired: A state of the art survey." *Pattern Recognition Letters* 137 (2020): 37-52.

[12] Sivan, Shankar, and Gopu Darsan. "Computer vision based assistive technology for blind and visually impaired people." In *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*, pp. 1-8. 2016.