Intelligent headgear for visually Impaired in Museums

Duraimurugan N, Akash V, Akhileshwar M
Department of CSE, Rajalakshmi Engineering College, Chennai, India

***Abstract*— Visually impaired people have difficulty experiencing the simplest of things. Most visually impaired people are confined to a small list of places, like their place of work, their communities. They don't think about exploring outside their current lifestyle. Visually impaired people don't have a long list of recreational activities to alleviate their minds as we do. Museums are something that is considered as a visual experience. In this project, we shall try to help them feel what going to a museum is like. Through auditory enhancements, we are going to make a headgear which aids them by transcribing the artifacts, paintings, etc., and broadcasting them via headset using text to speech. Hence this helps visually impaired people to understand and appreciate many things with ease without the help of someone else.**

Keywords—object detection, yolov5, headgear for visually impaired, museums

I. Introduction

Life as such is very trying and difficult at times. Having to deal with sight loss or low vision is merely one of the challenges that the visually impaired are facing when living life. Blind individuals are just like anyone else but they just can't see. The lack of accessibility for the visually impaired is central to a number of the issues the blind or people with partial visibility . There are a limited number of inclusive/accessible activities for the visually impaired. Visiting a place as simple as a museum is long been a hurdle for the blind. Museums or often considered to be a leisure for the normal people.

Most museums as of now do not have any facilities to accommodate visually impaired people. In some museums the history of the artefacts are inscribed in braille along with English for the visually impaired. Currently when a visually impaired person wants to visit a museum he or she must hire a guide or have someone along with them to navigate through the museum at all times. They might have finding directions inside the museums if they go on their own.

In this project we create an intelligent headgear for the visually impaired. This headgear would help them navigate within the museums premise without the need of an additional person, by giving them directions via headphones. With the help of image processing the headgear identifies the artifact in front of them and creates a short description of the artifacts history and its appearance. The description is then converted to speech (audio) and then sent to the user via headset

II. Related works

Over the years various interfaces that use image processing for the blind had been created. There were other devices created for the same problem that was presented. A can which uses infrared was created [7]. But the downside was that it just told the user that an object was in front of them, without giving any descriptions. Another way imposed was to use microcontrollers and provide navigation for the visually impaired [12]. But the problems faced by the creators was again that they couldn't again determine what object is front of them. The device can only detect that objects are in front of them.

Another gadget created was that a portable machine attached to the belt like a "fanny pack" [13]. The disadvantage was that it can only detect 2D objects and not 3D objects, so the user has to take additional work and find whether it is 2D or 3D. A machine similar to our project was created [14], but due to a lower model of Raspberry Pi makes it slow in recognizing the objects.

There were machines created which involved canes [5, 15, and 7]. But some of them needed the user to hold the cane in a steady angle, which is not possible in upward or downward pathways [4]. In other devices, it is required the user to show the cane everywhere so that it can scan the entire path as the laser emitted covers only a straight line So considering all of the above creations, this project was implemented

III. METHODOLOGY

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

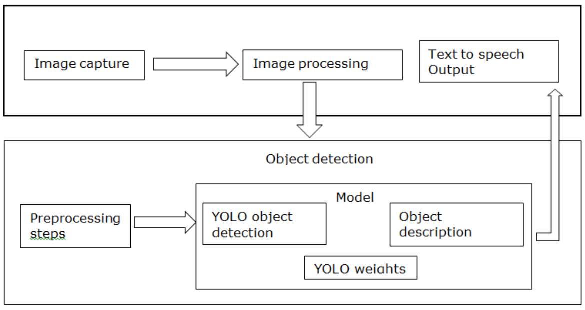


Fig.1: System Architecture

The system architecture shows the workflow of this project. Starting with image capture as shown in Fig 3.1. The captured image is sent to image processing where the image is divided into chunks which can be further utilized by the yolo algorithm. Using the weights file the yolo algorithm recognizes the object in front of the user and sends the label and the corresponding description as output. This then is converted from text to speech.

A. *Dataset collection and training*

The images for the model were collected from the official website for museums of Tamil Nādu. The number of images for each class was one. We had to populate the classes to form a dataset (Fig.1) that could be trained. We then labelled the datasets using an image labeller for object detection called labelling.

Fig.2: Dataset

We trained the model using the latest Yolov5 model. YOLO an acronym for 'You only look once', is an object detection algorithm that divides images into a grid system. Each cell in the grid is responsible for detecting objects within itself. Yolov5 is the latest release which is based on pytorch (PyTorch is an open source machine learning framework based on the Torch library, used for applications such as computer vision and natural language processing, primarily developed by Facebook's AI Research lab.). We produced a weight file with the accuracy of 99.5% as shownin the Fig.2.

Fig.3: Test dataset

B. *Object detection module*

In this module we do a real time image processing of the place that the blind or the visually impaired person is currently looking at. Object detection is a computer vision task that involves identifying instances of an object of a certain class within an image or a video. We then find the object inside the frame with historical significance and isolate it from the rest of the chaos in that particular frame. This can be done using the help of OpenCV which is a deep learning model. OpenCV is a library of programming functions mainly aimed at real-time computer vision. The weights file from the dataset creation module is used to detect the object in the screen As shown in the Fig.2.

YOLO is a fast and accurate approach to object detection. The algorithm only looks at an image once and detects all the objects that are present along with their location.

YOLO works by splitting images into a grid where each grid cell identifies an object by itself. YOLO's architecture is shown in the Fig.3. The grid cells predict all the bounding boxes and give each of them a confidence score to determine the accuracy of each prediction. When it comes to performance, YOLO outperforms other object detectors by far. In real- time, it can process images at a rate of around 155 frames per second (fps), achieving double the mAP (Mean Average Precision - a popular evaluation metric for object detectors) of other object detectors like R-CNN. So, when the yolo algorithm is run we find all the artifacts within a frame with different confidence scores.

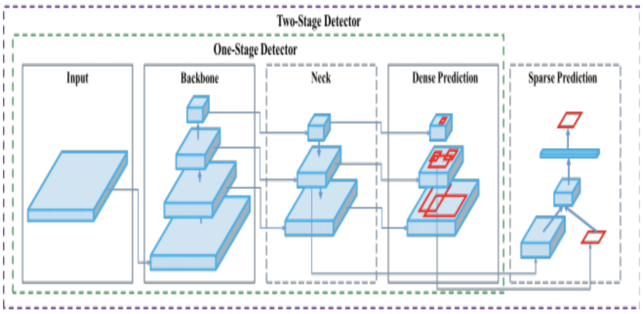


Fig.4: YOLOv5 architecture diagram

C. *Description Module*

After the image of the artifact is extracted from the frame, the historical data or facts about the object is identified. Since the project is for the visually impaired, they have some specific requirements in the description of the objects. So aside from the historical data we also need to give a description of the shape and texture of an object for them to register in their minds.

Image description generator is a process of recognizing the context of an image and annotating it with relevant captions using deep learning, and computer vision. It includes the labelling of an image with English keywords with the help of datasets provided during model training. The description is then converted to speech via Pyttsx3. Pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline, and is compatible with both Python 2 and 3. The speech output is transmitted via headphone.

D. COMPONENTS

A. *Raspberry Pi 4 Model B*

The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor TV, and uses a standard keyboard and mouse. It is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python. It's capable of doing everything you'd expect a desktop computer to do, from browsing the internet and playing high-definition video, to making spreadsheets, word-processing, and playing games.In our project This is used to run all the operations of this project (image detection, image description, text-to-speech conversion).

B. *Pi Camera Module – 5MP*

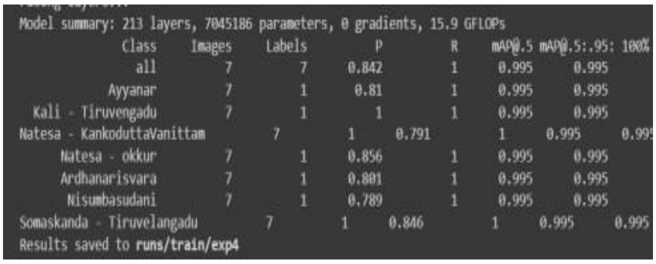
The Pi camera module is a portable light weight camera that supports Raspberry Pi. It communicates with Pi using the MIPI camera serial interface protocol.

C. *Headset with microphone*

Headset is used to hear the image description that has been converted to speech.

IV. RESULTS AND ANALYSIS

The object detection process is run with a confidence score of 0.3 So it could recognize objects even inn low lighting conditions. since higher confidence scores ignore some lowlit objects. Our model was created with an accuracy of 99.5%as shown in the Fig.7. We used 120 epochs (An epoch is a term used in machine learning and indicates the number of passes of the entire training dataset the machine learning algorithm has completed) and a batch size of 32.


Model summary: 213 layers, 7045186 parameters, 0 gradients, 15.9 GFLOPs

Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95: 100%
all	7	7	0.842	1	0.995	0.995
Ayyanar	7	1	0.81	1	0.995	0.995
Kali - Tiruvengadu	7	1	1	1	0.995	0.995
Natesa - KankoduttaVanittam	7	1	0.791	1	0.995	0.99
Natesa - okkur	7	1	0.856	1	0.995	0.995
Ardhanarisvara	7	1	0.801	1	0.995	0.995
Nisumbasudani	7	1	0.789	1	0.995	0.995
Somaskanda - Tiruvelangadu	7	1	0.846	1	0.995	0.995

Results saved to runs/train/exp4

Fig.7: Model summary

Initially we tried to train the model with just one image per class with which we couldn't even get a model with more than 50% accuracy (Fig.8). So, we had to manually create datasets from the existing images. The final result we got is completely satisfactory to our expectation. The raspberry pi did struggle a bit to run this heavy process. It also had delays which we cannot see innormal computer.

As for as the future of this project is concerned, we are planning to utilize a better raspberry pi with more processing power and more compact than what it is now. Also, we need more datasets and to get even better precision and classes to detect.

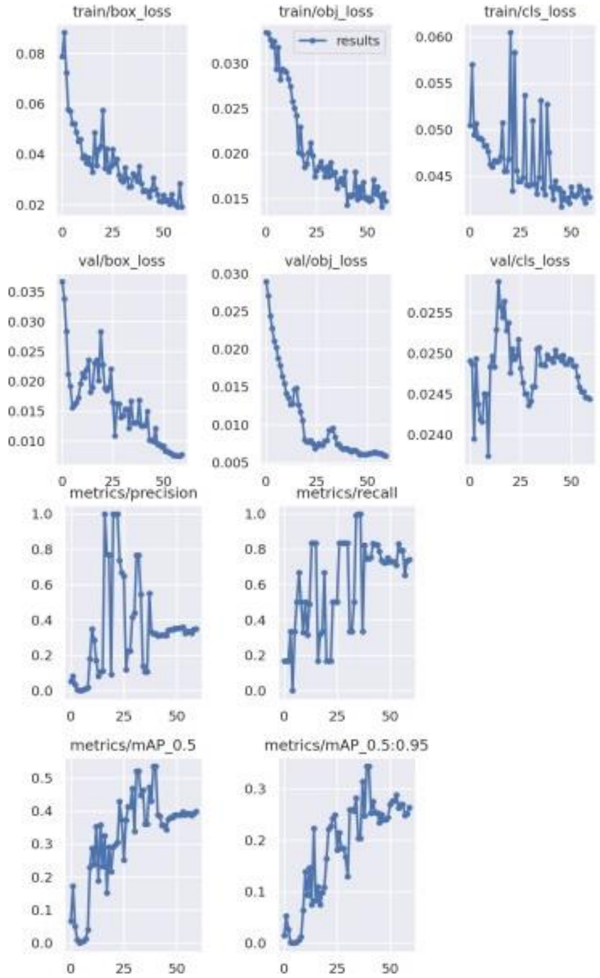


Fig.8: Initial model precision data

V. CONCLUSION

This paper introduced the intelligent headgear for visually impaired in museums project. Our main idea was to create a simple headgear that can detect the artifacts. We have implemented our idea using image processing, image description and text-to-speech modules. Since there weren't any proper datasets available, a dataset from scratch had to be created. For the creation of a dataset, information had to be collected from several museum websites. This module does not have 100% accuracy, hoping that future researchers can get 100% accuracy. We also hope that people in the future make devices that help people with disabilities by understanding our project.

REFERENCES

[1] "Vision-Based System for Assisting Blind People to Wander Unknown Environments in a Safe Way", Andrés A. Díaz-Toro ,1 Sixto E. Campaña-Bastidas ,1 and Eduardo F. Caicedo-Bravo (2021)

[2] "An ultrasound sensor-based body area network for guiding blind people", António Pereiraab, Nelson Nunesa, Daniel Vieiraa, Nuno Costaa, Hugo Fernandesc, João Barrosoc

[3] "Third Eye for Blind", Azra Batool, Shaheen Naz

[4] "Real-Time Object Detection with Yolo", Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam (2019)

[5] "GUIDECANE", Iwan Ulrich, Johann Borenstein (2003)

[6] "Unified, Real-Time Object Detection", Joseph Redmon , Santosh Divvala, Ross Girshick, Ali Farhadi (2015)

[7] "THE LASER CANE", J. Malvern Benjamin, Jr., M.S .E.E

[8] "Eye for blind people", M.Navaneethanath, S.R.Naveen Prasath, M.Azeem Ahamed, S.Diwakaran (2019)

[9] "Object Detection and Recognition using one stage improved model", Pranav Adarsh; Pratibha Rathi; Manoj Kumar (2020)

[10] "A Remote Guidance System for the Blind", Przemyslaw Baranski, Maciej Polanczyk, Pawel Strumillo (2010)

[11] "YOLO-LITE: A Real-Time Object Detection Algorithm", Rachel Huang, Jonathan Pedoeem, Cuixian Chen (2018)

[12] "Navigation Tool for Visually Challenged using Microcontroller", Sabarish.S(2013)

[13] "Navbelt", Shraga Shoval , Iwan Ulrich, Johann Borenstein (2003)

[14] "Smart Helmet for Visually Impaired", Salman Shaikh, Saiyed Farhan Jafar, Karan Sosa, Pratap Nair (2019)

[15] "Dynamic Environment Exploration Using a Virtual White Cane", D. Yuan and R. Manduchi