iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62769

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Object Detection Using Image Processing for Blind Person

Devansh Srivastava[1], Ayush[2], Aditya Gandhi[3], Ambuj Tripathi[4]
Department of Electronics and Communication Engineering, ABES Engineering College Ghaziabad, India

Abstract: ***This research presents a fully autonomous assistive technology based on artificial intelligence that can distinguish various objects and provides real-time aural cues to the user, improving comprehension for visually impaired people. Multiple photos of items that are extremely useful for the visually impaired person are used to build a deep learning model. The learned model is made more robust by manually annotating and augmenting training photos. A distance-measuring sensor is incorporated in addition to computer vision-based object identification algorithms to enhance the device's comprehension by identifying barriers during navigation. The algorithms used to process images and videos were made to accept inputs from the camera in real-time. Deep Neural Networks were utilized to predict the objects, and Google's well-known Text-To-Speech (GTTS) API module was used to precisely detect and recognize the group or category of objects and locations contained in the anticipated voice message.***

Keywords: YOLOv3, Deep Neural Networks, GTTS, Object detection, Object recognition

I. INTRODUCTION

People with impaired vision, when traveling from one location to another, or when navigating, people encounter several difficulties. The ability to see allows humans to be aware of impediments in their path. To overcome the issues faced by blind individuals, a readily accessible solution is required. These kinds of tasks are difficult for the blind, and being able to recognize objects will be a necessary skill for them on a regular basis. They typically have trouble recognizing objects and navigating their environment, particularly when they're out on the streets. The majority of sighted individuals appeared to be fifty years of adulthood or beyond [1]. A few applications for the blind and visually handicapped have been made available. Even still, voice output and truly uninterrupted article recognition and object identification are lacking for visually impaired people. Certain applications make advantage of the Internet of Things to recognize obstacles in the user's vicinity and alert them through beeping or other means.

For a variety of reasons, users are required to carry a significant number of gadgets. For navigational support, devices such as smartphones, navigators, smart sticks with obstacle detectors, etc. are required. These devices may present problems for the user and are very expensive. These days, Braille text and tactile markers that are marked on the top of things for verification are among the methods that are widely used to assist the blind and visually challenged [4, 9].

Barcodes, Radio Frequency Identification Devices (RFID), and talking labels are examples of technologically advanced resources for locating objects nearby [5,9].

Individuals who are completely blind or have little vision usually struggle outside. Walking or driving on an overcrowded road might provide some challenges. Generally speaking, an individual who is visually impaired requires help from sighted friends or family members in order overcome unfamiliar environments and be safe. To get past all of these obstacles, they need canes. A cane, however, cannot identify the kind of object in front of it. Blind people, in their experience, are usually able to identify objects, which may cause injury or a collision if the object is not exactly what anticipated. They could come across difficult-to-avoid obstacles like stairs, dogs, parked cars, bicycles, etc. when walking along a path.

A group from South Korea created an Unmanned Underwater Vehicle (UUV) prototype in 2009. To find any obstacles in its route, the UUV had an imaging device and a beam of light placed to it. The laser marked on it, and the camera took pictures. It was thereafter turned into grayscale pictures.

It was able to identify any item ahead of the autonomous undersea vehicle by using the most advanced pixel and histogram [7]. This made obstacle detection more widely used.

In 2018, Gnana Bharathy demonstrated how to use video analytics in the cloud for object detection and classification, with precise and high-performing results. Here, the video stream analysis procedure is automated using a cascade classifier, which also provided the framework for the testing of several other types of video analytics algorithms [8].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

The main features of software designed to assist blind or visually impaired people were outlined in the study [9]. Reducing the need for distinct strategies for object identification and movement detection is the primary goal.

Because smartphones are used so often in daily life, these elements are implemented for the Android operating system. Two modules that are based on trainable artificial neural networks are object identification and motion detection. The image processing methods used to recognize objects and track their movements were covered in this study. The users in this system receive notifications in the form of spoken words.

II. METHODOLOGY

A. Object Discovery and Recognition

Finding and identifying entities in a clip or image is known as object detection, and it constitutes one of among the most crucial jobs in computer vision.

It integrates features from object localization and picture classification, allowing systems to not only identify an item's category but also create bounding boxes around it. Numerous contemporary applications, including augmented reality, facial recognition, autonomous driving, and security monitoring, are based on this technology.[2]

The technique of locating and recognizing items in a picture without first understanding the object categories is known as object discovery. This entails identifying every area in a picture that might contain an item, sometimes with the use of unsupervised or loosely supervised learning methods. When the item categories are not preset or the system must dynamically adapt to new objects, object discovery plays a critical role. Candidate object areas are often generated by methods like region proposal networks (RPNs), which can subsequently be subjected to additional analysis.[6]

On the other hand, object recognition entails both object detection and categorization into predetermined groups. A trained model that can correctly detect and label objects based on learnt characteristics is needed for this job. YOLO (You Only Look Once) and Faster R-CNN are two popular deep learning models and convolutional neural networks (CNNs) used for object identification. Large datasets are used to train these models, which helps them discover complex patterns and characteristics that differentiate between various object categories.

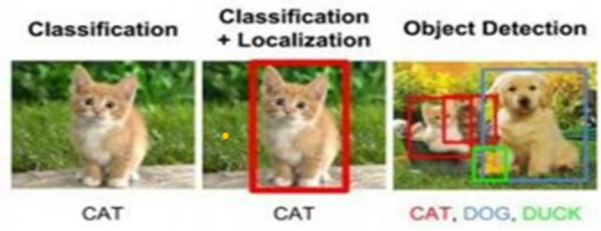


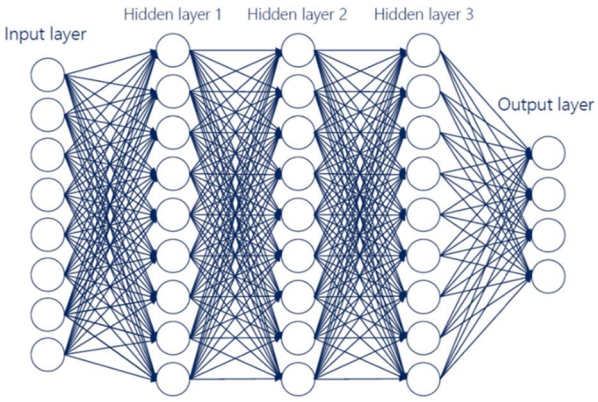
Fig. 1. Object Detection Process

B. Deep Neural Networks (DNNs)

Complex artificial neural networks with numerous layers, known as deep neural networks (DNNs), are able to recognize and analyze complicated patterns in data. An input layer, many hidden layers, and an output layer are the components of a typical DNN architecture. Using optimization methods such as gradient descent, DNNs are trained by putting them through forward propagation to produce an output and backward propagation to modify the weights depending on the mistake. The network may learn complicated functions by introducing non-linearity using activation functions like sigmoid, tanh, and ReLU. Overfitting is avoided and generalization is improved by using strategies like dropout, batch normalization, and weight regularization. Natural language processing (NLP), speech recognition, and picture identification are just a few of the industries that DNNs have transformed. One kind of DNN that is particularly good at object detection and picture classification is the convolutional neural network (CNN). Language translation and sentiment analysis are activities that are handled by NLP models like Transformers and Recurrent Neural Networks (RNNs). Deep neural networks (DNNs) are used in speech recognition to accurately translate spoken words into text, which powers voice assistants and transcription services.

DNNs are important in assistive technologies for people with visual impairments. Real-time item localization and identification are provided by object detection methods such as YOLO (You Only Look Once), which also provide important ambient data. When used in conjunction with text-to-speech (TTS) systems such as Google's GTTS, these models provide audio descriptions of nearby items, improving blind people's autonomy and situational awareness. Through increased freedom, this integration greatly enhances accessibility and gives visually impaired people more control.[15]

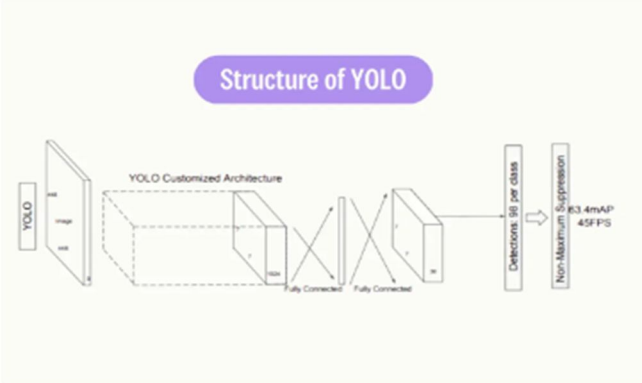
All things considered, DNNs have accelerated important developments in a number of fields, demonstrating their critical role in creating intelligent and easily accessed technology solutions. Through creative applications, their capacity to collect and analyze massive amounts of data has not only advanced technology but also enhanced human lives.



C. YOLOv3

In the field of object recognition, the YOLO V3 (You Only Look Once, Version 3) dataset is a mainstay, offering a strong basis for developing and testing state-of-the-art models. This dataset, which consists of a wide range of annotated photos, provides the foundation for the YOLO V3 model's exceptional speed and accuracy. Bounding boxes and associated class labels are carefully added to every image in the collection, capturing the spatial arrangement and objects' identities. The model is able to understand subtle spatial correlations and item features thanks to this painstaking annotation process, which makes precise and dependable identification possible in a variety of settings.[11]

Our default YOLO algorithm analyzes images in immediate succession at forty-five frames per second. Compressed version of the network called Quick YOLO processes an astounding 155 frames per second and gets doubling the mAP of conventional current time analyzers. Compared to the most sophisticated recognition algorithms, YOLO creates more localization errors but is less likely to predict false positives on background. Finally, YOLO recognizes very simple object abstractions. It outperforms different approaches to detection like DPM and R-CNN when extrapolated from genuine photos to other fields like artwork.[10]



Using the abundance of annotated photos in this dataset, YOLO V3 is trained to predict box boundaries and class probabilities with accuracy. Through training, the model improves its comprehension of the spatial configurations and visual properties of things, maximizing its real-time object detection and classification capabilities. Metrics like mean Average Precision (mAP) are commonly used to evaluate the trained model, offering a thorough analysis of its performance across various object classes.[12]

Applications for the YOLO V3 dataset may be found in a wide range of industries, such as assistive technology, autonomous driving, and surveillance. Systems may accomplish quick and accurate object recognition by using the YOLO V3 model, which was trained on this dataset. This allows for improvements in accessibility, safety, and security. The YOLO V3 dataset essentially serves as evidence of the critical significance that excellent annotated data play in fostering innovation and advancement in the fields of object recognition .[14]

D. GTTS

Google Text-to-voice (GTTS) is a ground-breaking technology that provides a fluid and easy method to engage with digital information by translating written text into voice that sounds natural. Google TTS, which is powered by cutting-edge neural network designs and machine learning algorithms, has completely changed how people interact and consume text on a variety of devices and apps.[13]

In order to provide a realistic audio experience, Google TTS primarily uses deep learning models to synthesis human-like speech from input text. These models are capable of catching minute details in intonation, cadence, and pronunciation. With the use of enormous volumes of training data, these models are trained to produce speech that closely mimics the organic rhythms and subtleties of human speech. This makes it possible for users to interact with material in a way that is more natural and immersive.

Google TTS's flexibility and versatility across several languages and dialects is one of its main advantages. The technology supports a broad spectrum of languages and accents, making it accessible to a wide variety of worldwide audiences and enabling smooth communication even in the face of linguistic difficulties. Furthermore, users may customize Google TTS's speech synthesis to suit their needs and tastes by adjusting characteristics like voice pitch, speed, and intensity.

Applications for Google TTS may be found in a variety of fields, from interactive voice response (IVR) systems in customer service and telecommunications to assistive technology for those with vision impairments.

When it comes to accessibility, Google TTS is essential since it gives visually impaired people access to digital content through aural input, enabling them to freely explore and engage with online information. Additionally, Google TTS makes it easier to create audio versions of text-based materials in educational settings, which improves learning outcomes for students with a range of learning requirements. In summary, Google Text-to-Speech is a potent tool for producing high-quality speech output from text input and marks a revolutionary development in natural language processing and human-computer interaction. In the connected world of today, Google TTS continues to influence how we interact with and consume digital material thanks to its wide language support, adaptable settings, and variety of applications.

E. Pyglet

Pyglet is a robust Python library made specifically for playing audio and video in multimedia applications. Pyglet offers developers a rich feature set and an easy-to-use toolbox for producing immersive audio experiences on a variety of devices.[3] Pyglet, at its heart, makes it easier for Python applications to import, manipulate, and play audio files. Pyglet's extensive collection of audio manipulation tools makes it simple for developers to import and dynamically edit audio files in a variety of formats, such as WAV, MP3, and OGG. With the help of these technologies, developers can easily construct rich and interactive audio experiences by enabling operations like volume adjustment, panning, and audio effect application.

Platform independence is one of Pyglet's main advantages as it enables programmers to design cross-platform audio apps that function flawlessly across many operating systems. Pyglet streamlines the development process and maximizes compatibility by offering a uniform and dependable audio playing experience across a variety of platforms, whether it's a desktop program, web-based project, or mobile app.

Pyglet is a versatile tool that may be used for a multitude of purposes since it supports sophisticated features including positional audio, real-time synthesis, and streaming audio. Pyglet gives developers the ability to produce dynamic and captivating audio material that captivates audiences and improves user experiences, for everything from immersive gaming experiences to interactive multimedia presentations and instructional aids.

Moreover, Pyglet is usable by developers of all experience levels because to its clear and simple API, copious documentation, and vibrant community. Pyglet offers the materials and tools you need to realize your audio projects,
regardless of experience level with Python.

To sum up, Pyglet is a flexible and strong Python audio player library that gives programmers a complete toolbox for producing dynamic and immersive audio experiences on a variety of devices. Pyglet's platform freedom, sophisticated capabilities, and intuitive interface enable developers to develop dynamic and captivating audio apps that capture users and improve user experiences. Pyglet opens up a world of possibilities for audio creation.

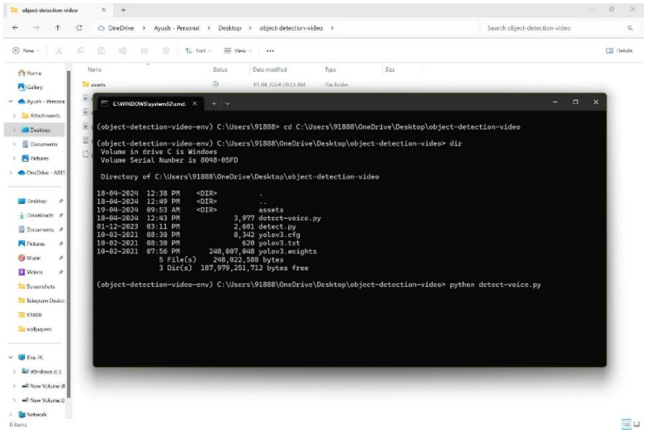
III. RESULTS AND DISCUSSION

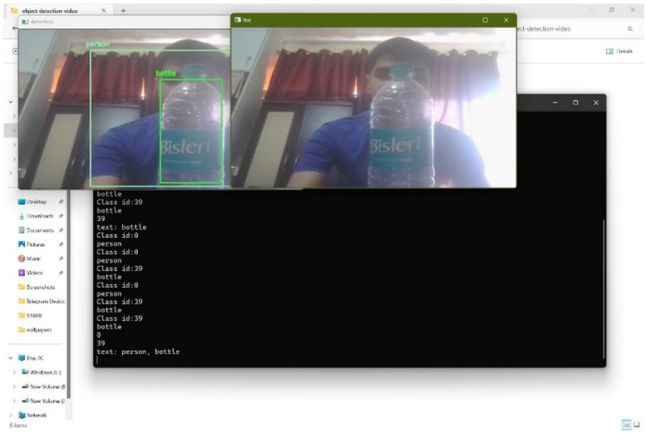
We use Open CV to create our real-time object detector based on deep learning, which requires quick access to our webcam or video stream, with each frame utilizing object detection. When we first start, we often grab a frame from the stream and adjust its size. After that, the frame is transformed into a blob with the aid of the DNN module.

For the hard work, pass the blob through the network that provides us with our detections and set it as the input into our Neural Network model. At this stage, we have detected items in the input frame. At that stage, we examine confidence values and choose whether or not to surround the object with a box and label.

Since we can identify numerous items in a single image, we start by iterating over our detections. Make sure you regularly assess the level of confidence linked to every identification. We call this confidence value "probability."

When the confidence level is high, meaning it is over the threshold, the prediction will be shown in the terminal along with a colored bounding box and an outline of the image with text.





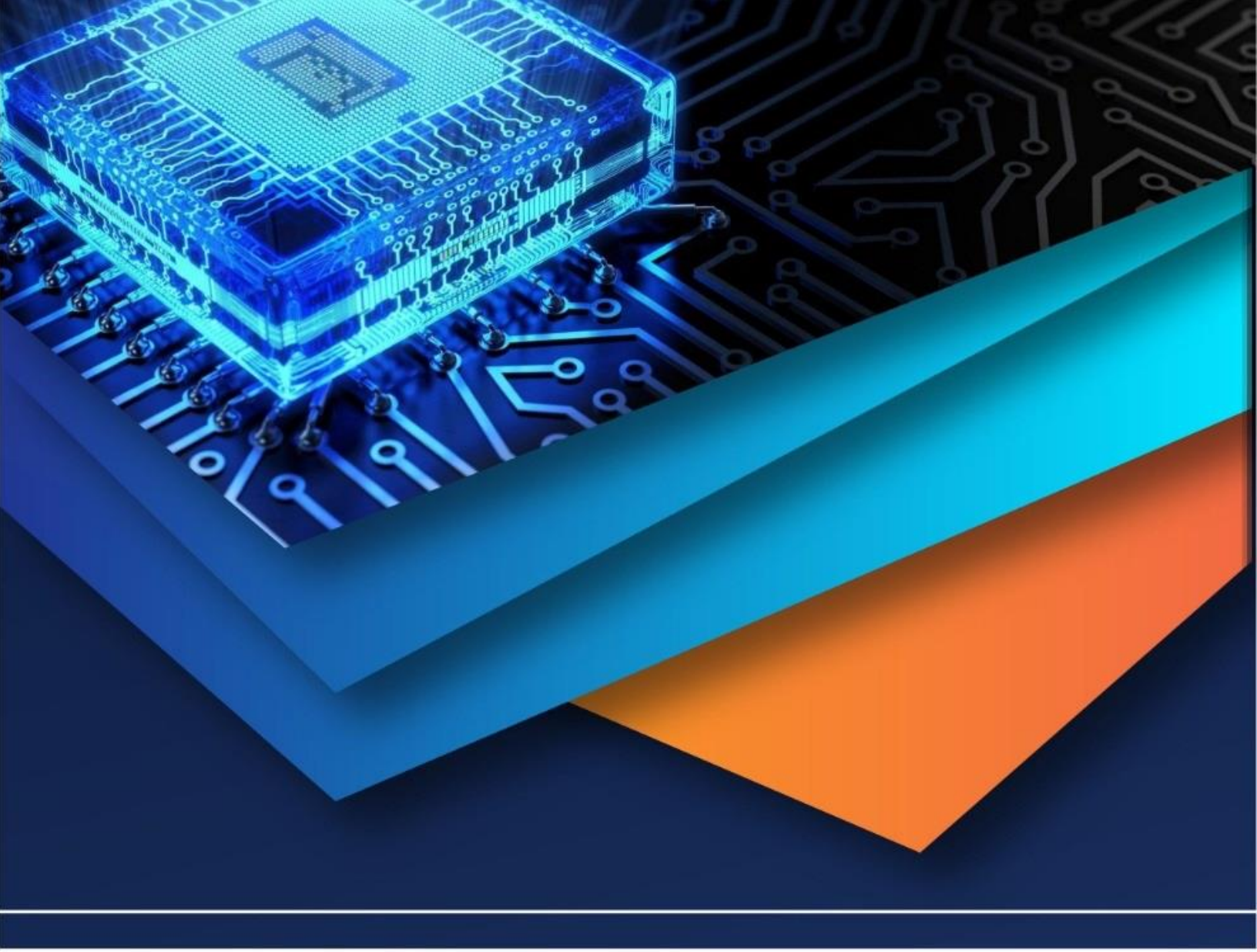
IV. CONCLUSION

Across several sectors, object detection's future has great chances. Processing of videos and objects Recognition techniques are suggested using current resources and devoted to users that are blind or visually impaired. Individuals with vision impairments have significant obstacles when strolling around and over hurdles in their day-to-day existence. In daily life duties, this software can assist blind and visually impaired people. consumers. The outcomes of the CNN, SVM, YOLO2, and YOLO3 were contrasted using accuracy and producing frames per second. Yolo3 accuracy surpasses other methods by 46.8% and Yolo3 performed better by generating frames up to 18 per second. It will decrease the mobility issue and object identification using a little solution that eliminates the requirement for Bring whatever extra equipment you want to use for it.

REFERENCES

[1] Mayur Rahul, Namita Tiwari, Rati Shukla, Devvrat Tyagi and Vikash Yadav (2022), A New Hybrid Approach for Efficient Emotion Recognition using Deep Learning. IJEER 10(1), 18-22. DOI: 10.37391/IJEER.100103.
[2] Object discovery and representation networks Olivier J. H´enaff, Skanda Koppula, Evan Shelhamer, Daniel Zoran, Andrew Jaegle, Andrew Zisserman, Joˆao Carreira, and Relja Arandjelovi´c DeepMind, London, UK Matusiak, K., Skulimowski, P., Strurnillo, P.,
[3] Pyglet documentation, piglet.readthedocs.io/en/latest/
[4] https://github.com/pyglet/pyglet/releases
[5] mis, Digital map and navigation system for the visually impaired, Department of Psychology, University of California, Santa Barbara, 1985.

[6] Small Object Discovery and Recognition Using Actively Guided Robot,Sudhandhu Mittal BITS, Pilani, M. Siva Karthik,Robotics Research Centre IIIT-H,Suryansh Kumar Robotics Research Centre IIIT-H,K. Madhava Krishna Robotics Research Centre IIIT-H
[7] Neha Bari, Nilesh Kamble, Parnavi Tamhankar, Android based object recognition and motion detection to aid visually impaired, International Journal of Advances in Computer Science and Technology, Vol. 3, No.10, pp. 462-466, 2014.
[8] Jason Yip, Object Detection with Voice Feedback YOLO v3 + gTTS, https://towardsdatascience.com/object-detection-with-voice-feedback- yolo-v3-gtts-6ec732dca91.
[9] Samkit Shah, CNN based Auto-Assistance System as a Boon for Directing Visually Impaired Person, 3rd International Conference on Trends in Electronics and Informatics, 2019.
[10] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.
[11] ODSC-Open Data Science, Overview of the YOLO Object Detection Algorithm, https://www.medium.com/@ODSC/overview-of-the-yolo- object-detection-algorithm-7b52a745d3e0.
[12] Joseph Redmon, Ali Farhadi, YOLO9000: Better, Faster, Stronger, IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517- 6525, 2017.
[13] gTTS,https://gtts.readthedocs.io/en/latest/
[14] https://www.datacamp.com/blog/yolo-object-detection-explained
[15] https://www.tutorialspoint.com/python_deep_learning/python_deep_learning_deep_neural_networks.htm



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY