iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33660

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: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

Third Eye: A Smart Assistant for Visually Impaired using Deep Learning

Chenthil. T. R[1], Meena. R[2], Megavarthini. G[3], Priyadharshini. R[4]
[1]*Assistant professor,* [2, 3, 4]*UG Students, Department of Electronics and Communication Engineering, Jeppiaar Engineering College, Chennai, Tamil Nadu, India.*

Abstract: ***Globally, one billion people have a vision impairment resulting in them unable to live a normal life. This project aims to develop a technological solution effectively enabling the visually impaired to identify text and objects in front of them. This Third Eye based on Raspberry Pi uses Deep Learning (DL) which tends to support in identifying the objects and faces that helps in orientation and navigation of these people in the unknown surrounding. It can also recognize text or any readable data and recite using earphones or speakers. The system will perform voice recognition from the user, when the user states object mode the camera is initiated enabling live streaming and a bounding box is applied to the object from which the distance and direction are calculated, then a voice note is given stating the name of the object, distance, and direction. If a face is detected, then it recognizes the face using the Detection Transformer (DETR) algorithm and then says the name of the person detected. When the user states text mode, the camera is initiated enabling live streaming and the text is captured and a voice is given stating the text. Thus, providing an effective technological solution for the blind to effectively identify objects and text in front of them. The proposed system which is fully voice-controlled uses the Raspberry Pi as a controller. The smart assistant is user-friendly and has a quick response which ensures real-time object detection for the visually impaired.***

Keywords: detection transformer, object detection, face recognition, convolutional neural network, OCR, audio output

I. INTRODUCTION

The impact of vision impairment severely impacts the quality of life. The serious problems encountered by blind people are mobility and dependency on others. This is because of the difficulty to recognize or detect their surroundings. Also, their visual impairment affects their interaction with others and social activities which can increase the rate of depression and anxiety. Many solutions were proposed in the past but still have limitations. These limitations may be caused by to lack of proposals that are suitable for a real-time environment. Researchers have made many kinds of research and proposals to provide a solution for visually impaired people to overcome obstacles and alert them. The latest research from the University of Oxford shows that blind people's brain improves other senses such as their hearing is more tuned to variations in frequencies so assisting devices and solutions can be provided in the form of voice-based. The implementation of our proposal has been done after referring to some previous innovations and assisting devices.

II. RELATED WORKS

Guang Chen et al[1] have proposed a novel method for improving the performance of generic object detection. Although great progress has been made in generic object detection, detecting small objects from images is still a difficult and challenging problem in the field of computer vision due to the limited size, less appearance, and geometry cues, and the lack of large-scale datasets of small targets. In this study, the small object detection networks are investigated along with a special focus on the modifications to improve the detection performance comparing to generic object detection architectures.

Peng Tang et al[2] proposed an object detection and recognition technique in high spatial resolution remote sensing imagery(HSRI) based on a convolutional neural network(CNN). As the robustness and applicability of generic object detection techniques are poor, this proposed model uses HSRI for the extraction and analysis of the image. The experimental results show that this approach has good object detection and recognition as the proposed CNN framework in HSSI is designated according to the suitable region of interest(ROI).

Waqas Aftab et al[3] proposed a model for detecting and tracking irregular and non-rigid bodies using the Spatio-temporal Gaussian process(STGP). This proposed model uses an extended model to get the augmented data which was further given a recursive filter and smoother to get filtered estimates of extended object tracking. This STGP method outperforms the previous Gaussian process extended Kalman filter approach in terms of accuracy and improvement in tracking of stimulated non-rigid objects.

©IJRASET: All Rights are Reserved

Yunhang Shen et al[4] proposed a novel scheme to perform weakly supervised object localization, termed object-specific pixel gradient (OPG). Weakly supervised learning refers to a learning framework where the model learns supervised models that are only partially or image-level labeled. The proposed model uses OPG which performs an iterative search for objects in the image effectively. The experimental results show that OPG with a pre-trained CNN model has fast and efficient object detection.
Dawei Du et al[5] proposed an efficient method for online deformable object tracking. This paper exploits higher-order structural dependences of different parts of the tracking target in multiple consecutive frames. Here, they first construct a structure-aware hyper-graph to capture such higher-order dependencies and solve the tracking problem by searching dense subgraphs on them.

III. PROPOSED SYSTEM

This project aims to develop a technological solution effectively enabling the blind to perceive their surroundings. The input voice is recognized by the processing module with the help of the VGGish algorithm. When the user states object mode the camera is initiated enabling live streaming and a bounding box is applied to the object from which the distance and direction are estimated using x and y coordinates, then a voice note is given stating the name of the object, distance, and direction. When the user states text mode, the camera is initiated enabling live streaming, and the text is captured which is converted into readable data with the help of OCR (Optical Character Recognition), and a voice is given stating the text. Finally, when the user activates Face mode, the camera captures the person in front of the camera and uses the DERT object recognition algorithm, and gives a voice note stating the name of the person. If an unknown person is detected, the audio output would be an unknown person is standing in front of the camera. Thus, providing an effective technological solution for the blind to effectively identify objects and text in front of them.

A. Architecture Diagram of the Proposed System

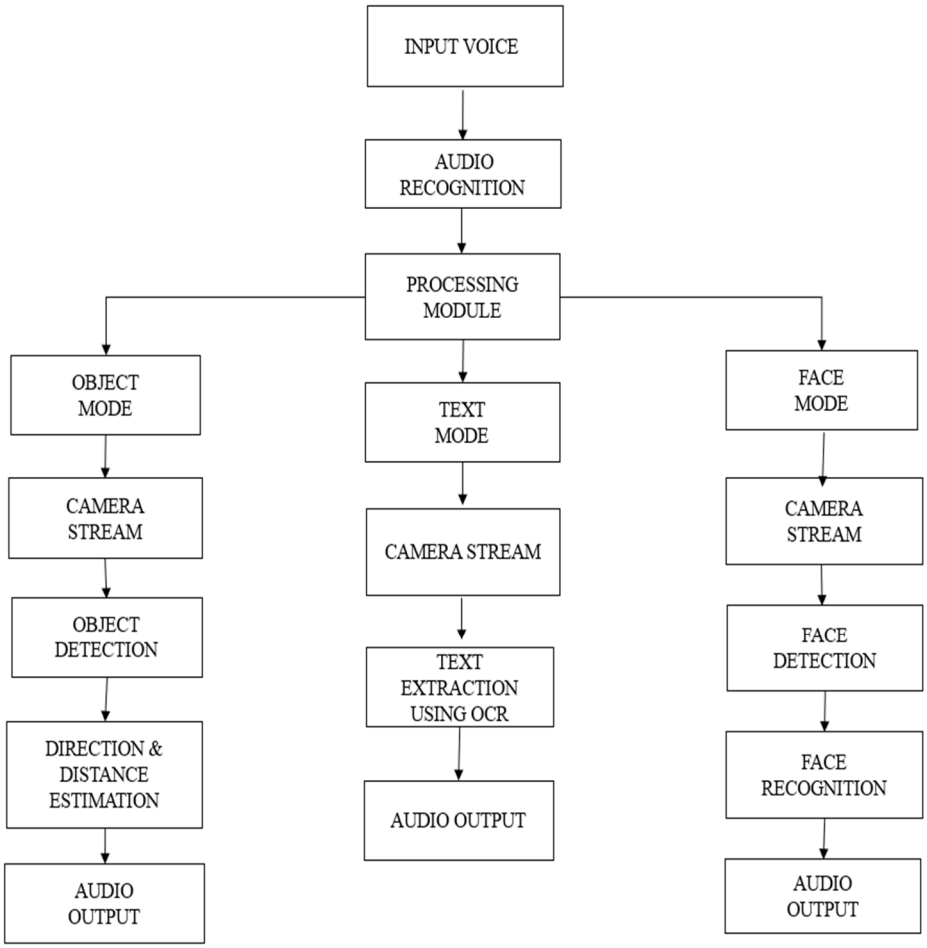


Figure 1: Architecture diagram of the proposed system

©IJRASET: All Rights are Reserved

The system consists of three modes: Object, Text, and Face mode. Each mode can be initialized by saying the name of the mode via the headphones.

1) *Object Mode:* The system will perform voice recognition from the user, when the user states object mode the camera is initiated enabling live streaming, and a bounding box is applied to the object from which the distance and direction are calculated, then a voice note is given stating the name of the object, distance, and direction.
2) *Text Mode:* When the user states text mode, the camera is initiated enabling live streaming and the text is captured and a voice is given stating the text. This is done with the help of OCR. In a text, if the sentence is a sequence of numbers and letters then the system vocalizes each character of that sentence.
3) *Face Mode:* When the user states face the face recognition module is initiated. The Detection transformer algorithm recognizes and eliminates no object classes and gives final audio output as this person (name of the person) is standing in front of the camera. As this project is an application specifically for the visually impaired, even when the user starts object mode and the camera detects a person, the system will recognize the person and gives the audio output.

IV. MODULE DESCRIPTION

A. *Processing Module*

In this project, Raspberry Pi 3B+ is the processing module that acts as a process for the input voice and acts as a controller. This Raspberry Pi model has dual-band Wi-Fi which makes it versatile and adaptable which comes in handy for fast and efficient object detection. This module processes the voice input of the user and gives the output.

B. *Camera Module*

The camera module here is a USB camera that captures the surroundings and gives live stream video for the Raspberry pi for further processing.

The camera used in this project has an image resolution of 25 megapixels with 6 light sensors which captures the image or video with high picture quality which further reduces the chances for incorrect or faulty detections. It is used to recognize an object with full of accuracy. It also enables night vision that allows images to be produced in levels of light approaching total darkness. So, this system can be used in an environment with minimum lighting too.

C. *Audio Recognition*

This project makes use of the VGGish model algorithm for effective audio recognition.

The input voice from the user is captured with the help of a microphone. Then it is given to the processing module which analyses the video and recognizes it with the help of the VGGish classification model. The below figure 2 shows the tensors from the Log Mel spectrogram which are given to an audio tagging system. This tagging system adds tags to audio data for further audio classification.

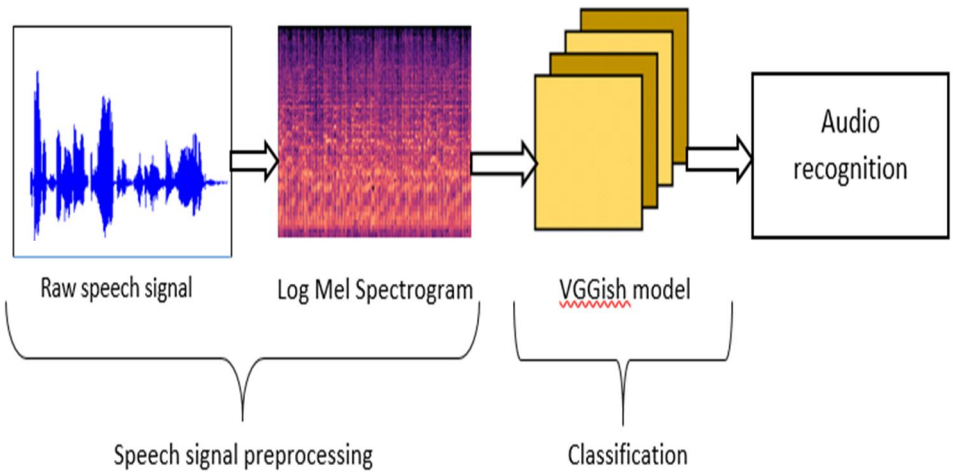


Figure 2: Sound classification and recognition

©IJRASET: All Rights are Reserved

D. OCR Module

OCR is that the use of technology to tell apart printed or written text characters within digital pictures of physical documents, like a scanned paper document. the fundamental method of OCR involves examining the text of a document and translating the characters into code that may be used for processing. OCR is usually conjointly mentioned as text recognition.

OCR systems area unit created from a mixture of hardware and code that's wont to convert physical documents into electronic text. Hardware, like AN optical scanner or specialized printed circuit, is employed to repeat or scan text whereas code generally handles the advanced process.

The process of OCR is most commonly used to turn hard copy legal or historic documents into PDFs. Once placed in this soft copy, users can edit, format, and search the document as if it was created with a word processor. Below figure 3 shows steps involved in the text to audio conversion using Optical Character Recognition.

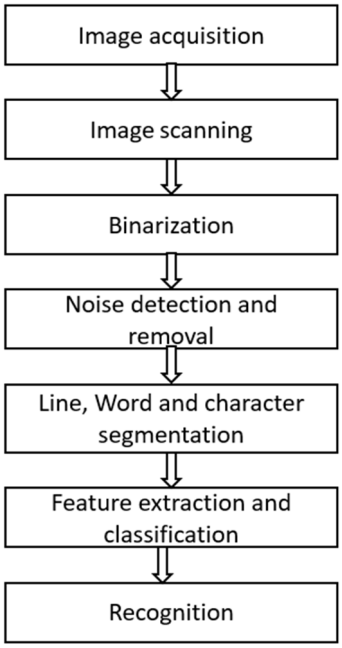


Figure 3: Optical Character Recognition

E. Object Detection

1) *Step 1:* First, the image through a convolution Neural Network Encoder because CNN works best with images. So after passing through CNN the image features are conserved. This is the higher-order representation of an image with many more feature channels.
2) *Step 2:* This enriched feature map of the image is given to a transformer encoder-decoder, which outputs the set of box predictions. Each of these boxes is consisting of a tuple. The tuple will be a class and a bounding box.
3) *Step 3:* Comparing and dealing with similar objects next to each other is another major issue and it is tackled by using bipartite matching loss.
4) *Step 4:* The direction and distance measurement is done in object mode. A bounding box is constructed on an object and the distance from the object is calculated based on the bounding box measurement taken along the x and y-direction.

The figure 4 below shows the Objection detection with transformers with bipartite matching losses for direct set prediction.

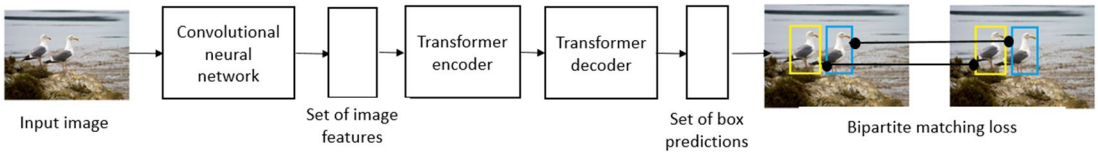


Figure 4: Object detection using DETR

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

V. SOFTWARE DESCRIPTION

A. Visual Studio

Microsoft Visual Studio is an IDE (Integrated Development Environment) from Microsoft. It is used to develop computer programs, as well as websites, web apps, web services, and mobile apps. Visual Studio uses Microsoft software development platforms as Windows API, Windows Forums, Windows Store. It can produce both native codes and managed code.

B. Python

In this project, python is employed because python is straightforward and easy to understand. The simplicity of the syntax permits you to subsume labyrinthine systems and make sure that all the weather has a transparent relationship with one another. Compared to different secret writing languages, like Java, Python encompasses a less restricted programming approach. It's multiple paradigms and might support a mess of programming designs, together with procedural, object-oriented, and practical ones. These areas unit the explanations for exploitation python during this project.

VI. RESULTS AND DISCUSSION

The below figures 5 and 6 shows the initiation of computer(server) and raspberry pi (client) respectively.

Figure 5: Server initiation

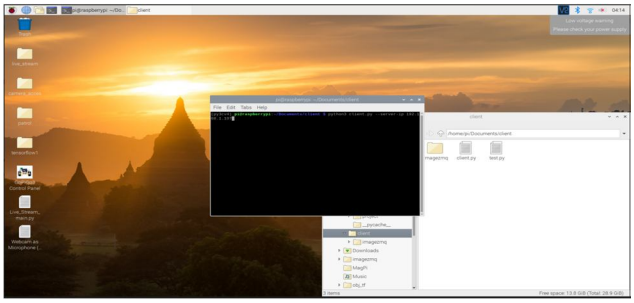


Figure 6: Raspberry Pi initiation

Once the live video stream is started recording and Raspberry pi is initiated, the input voice of the user is recorded followed by the audio recognition**.** Below figure 7 shows the screenshot of the audio recognition phase of the project. In this phase, the processing module recognized the audio with the help of the VGGish audio recognition model.

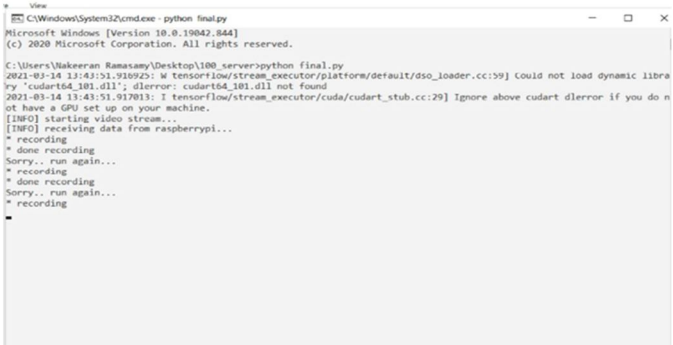


Figure 7: Audio recording initiation

©IJRASET: All Rights are Reserved

When the user initiates object mode the camera captures the picture and the objection detection phase starts. The below figure 8 shows the object detection with bounding box and confidence score.

Figure 8: Object detection

On the input audio from the user being recognized to be the text the camera module is initiated which detects the text in front of it. The below figure 9 shows as the user states the text mode, the text captured in the image is given in the form of a voice note and the text can be read in the terminal console too.

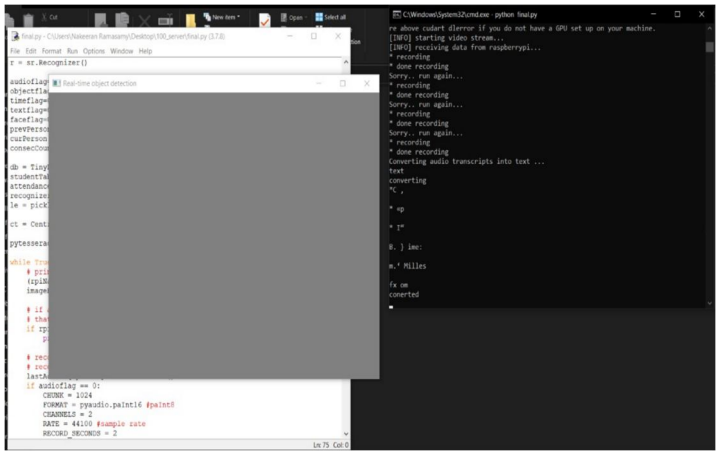


Figure 9: Text reading

. The below figure 10 shows the overall setup of the proposed system.

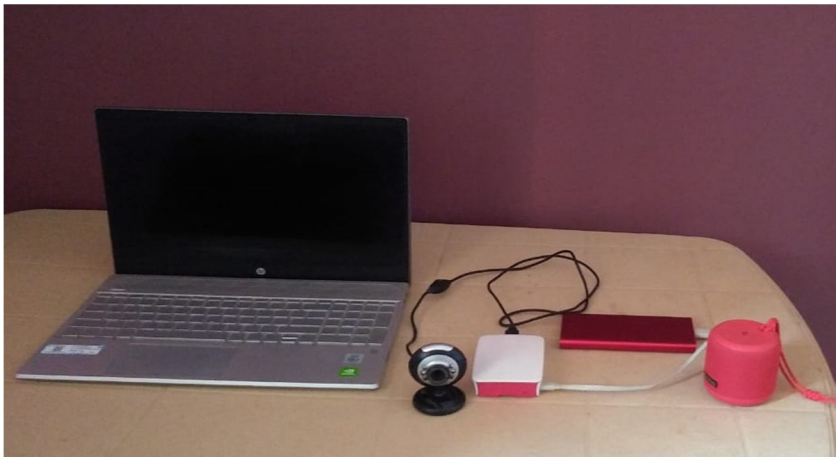


Figure 10: Overall Setup of the proposed system

©IJRASET: All Rights are Reserved

VII. CONCLUSION

This project is used to effectively improve the livelihood of visually impaired people. The system successfully performs object detection, text reading, and face recognition and gives an audio output of the results. So, the system has a very large scope in terms of helping blind people. In the coming future, this project for object detection can be improved by reviewing its application in various other fields for text identification and effective object detection. In the homecare and NGOs, they are more chance to develop or convert this project in many ways. Thus, this project has an efficient scope in the coming future where manual predicting can be cheaply converted to computerized production.

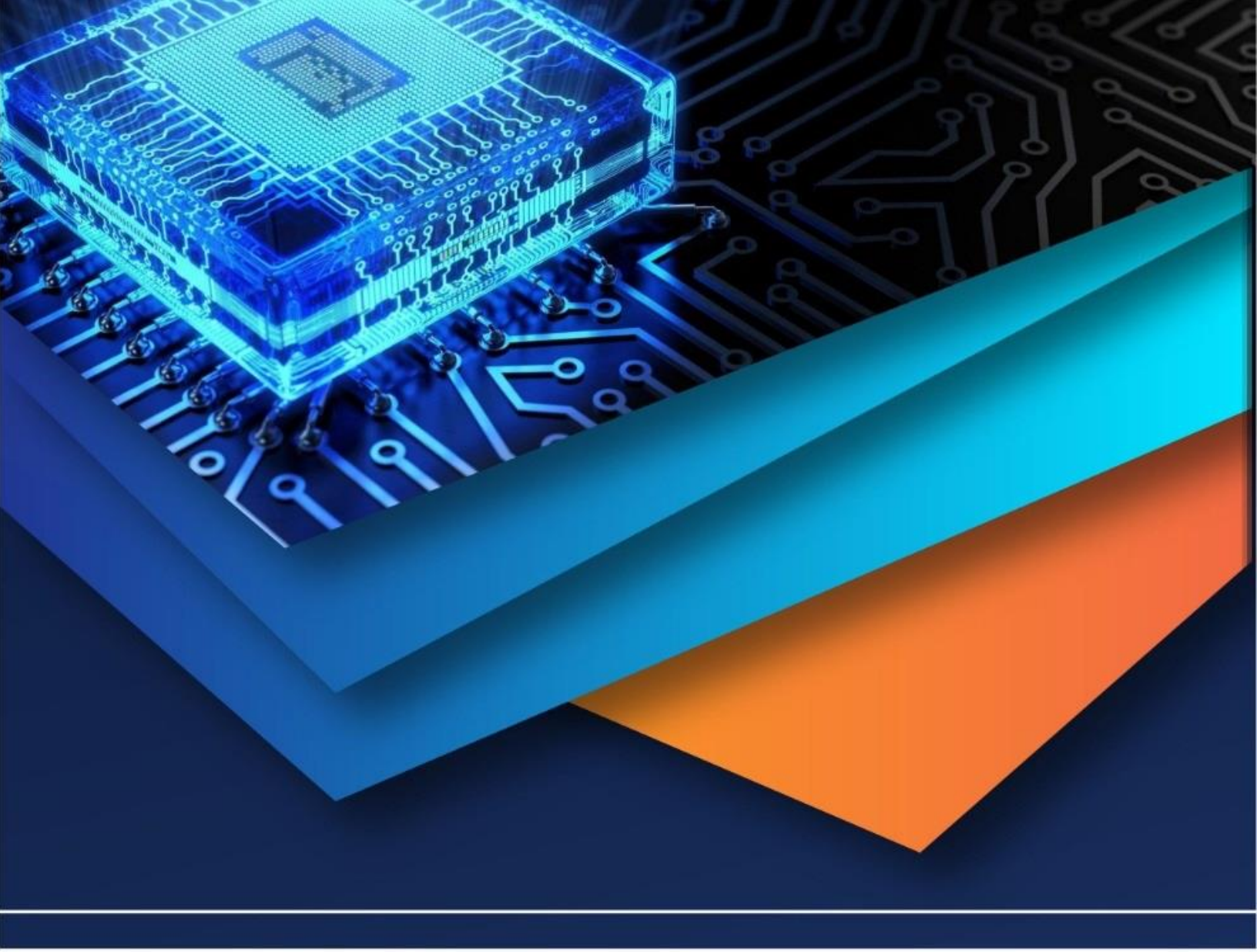
VIII. ACKNOWLEDGEMENT

We would like to thank the Jeppiaar Engineering College (Department of Electronics and Communication Engineering) for providing the resources and guidance. We also express our sincere gratitude to Mr.T.R.Chenthil, Assistant Professor, Department of Electronics and Communication Engineering, Jeppiaar Engineering College, Chennai for guiding and providing the solution and appropriate references.

REFERENCES

[1] Dawei Du, Hong gang Qi, Wenbo Li, Longyin Wen, Qingming Huang, and SiweiLyu “Online Deformable Object Tracking Based on Structure-Aware Hyper-graph” [2016, Vol. No: 1057-7149].
[2] Guang Chen, Haitao Wang, Kai Chen, Zhijun Li, Zida Song, Yinlong Liu, Wenkai Chen, and Alois Knoll “A Survey of the Four Pillars for the Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution and Region Proposal” [2020].
[3] Peng Tang, Chunyu Wang, Xinggang Wang, Wenyu Liu, Wenjun Zeng, and Jingdong Wang “Object Detection in High-Resolution Remote Sensing Imagery Based on Convolutional Neural Networks with Suitable Object Scale Features” [2019, Vol No: 0196-2892].
[4] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, Philip H. S. Torr “Deeply Supervised Salient Object Detection with Short Connections” [2018, Vol. No: 0162-8828].
[5] Waqas Aftab, Roland Hostettler, Allan De Freitas, Mahnaz Arvaneh, and Lyudmila Mihaylova “Spatio-temporal Gaussian Process Models for Extended and Group Object Tracking with Irregular Shapes” [2018, Vol.No:0018-9545].

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)