

ARTIFICIAL INTELLIGENCE EYE

A PROJECT REPORT

Submitted by

**S.BHARAT KUMAR [Reg No: RA1911004010181]
S J MOHAMMED FAREESTHA [Reg No:RA1911004010267]**

Under the guidance of

Dr.C.VIMALA

(Assistant Professor, Department of Electronics & Communication Engineering)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS AND COMMUNICATION ENGINEERING

of

COLLEGE OF ENGINEERING AND TECHNOLOGY



S.R.M. NAGAR, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report titled “**ARTIFICIAL INTELLIGENCE EYE**” is the bonafide work of “**SJ MOHAMMED FAREESTHA [Reg No:RA1911004010267]**” who carried out the project work under my supervision along with his batch mates **S.BHARAT KUMAR[Reg No:RA1911004010181]**. Certified further, that to the best of my knowledge the work reported herein does not form any other project report on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

SIGNATURE

Dr.C.VIMALA
GUIDE
Assistant Professor
Dept. of Electronics &
Communication Engineering

Signature of the Internal Examiner

SIGNATURE

Dr. SHANTHI PRINCE
PROFESSOR
HEAD OF THE DEPARTMENT
Dept. of Electronics &
Communication Engineering

Signature of the External Examiner

DECLARATION

We hereby declare that the Major Project entitled “**ARTIFICIAL INTELLIGENCE EYE**” to be submitted for the Degree of Bachelor of Technology is our original work as a team and the dissertation has not formed the basis of any degree, diploma, associate-ship or fellowship of similar other titles. It has not been submitted to any other University or institution for the award of any degree or diploma.

Place:

Date:

S.BHARAT KUMAR
[RA1911004010181]

S J MOHAMMED FAREESTHA
[RA1911004010267]

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to our Founder Chancellor Dr. T. R. Paarivendhar, for providing us the necessary facilities for the successful completion of our course.

We express our humble gratitude to Dr.C.Muthamizhchelvan, Vice-Chancellor, SRM Institute of Science and Technology, for their constant support and invaluable administration.

We extend our sincere thanks to Dr. T.V.Gopal, Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We are incredibly grateful to our Head of the Department, Dr. Shanthi Prince, Professor, Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

I would like to express my deepest gratitude to my guide, Dr.C.Vimala,Assistant professor,Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, her valuable guidance, consistent encouragement, personal caring, timely help and providing me with an excellent atmosphere for doing research despite the busy schedule.

We register our immeasurable thanks to our Project Co-ordinator, Dr.K.Harisudha, Assistant professor, Department of Electronics and Communication engineering, SRM Institute of Science and Technology, for giving valuable suggestions in each stage of our project .

We would like to express our gratitude towards Electronics and Communication Engineering Department for giving us an opportunity and encouragement which helped us in completion of this project.

SJ MOHAMMED FAREESTHA

ABSTRACT

Object recognition is a computer vision technique for identifying objects in image or videos. Object recognition is a key output of deep learning and machine learning algorithms. There are a few ways in which artificial intelligence can be used to help those who are blind or have low vision so in this way we can help poor people. Since the cost of the device is budget friendly and easily usable it can help a lot of people .

Artificial intelligence devices are developed in such a way that they are easily usable and affordable so this device is designed in such way to help the people that is by providing verbal feedback .These devices are able identify objects , navigate and perform other tasks with development of sensors in the future we can also incorporate lot more features such as we can identify the objects within particular distance or detect object in which speed it is approaching and even the angle of it. Artificial intelligence can also be used to help those are blind or have low vision is through the use of artificial intelligence-based applications that can provide information about the environment and help with tasks such as reading text or identifying objects. The goal is to teach a computer to do what comes naturally to humans : to gain a level of understanding of what an image contains.

In this project we are going to capture the image using ESP32 CAM which is widely used in various IOT applications. Then this information is run into an artificial intelligence system that we develop using python and for the working of the ESP32 CAM module we have used Arduino code for the generation of URL which serves as link between the Arduino and python programming. In the python programming the designated libraries have been installed for the identification of object . Then we are going to write a program for audio generation which we are going to incorporate it into the object detection programming so we can generate audio for the detected objects which are identified by the microcontroller. The camera module OV2640 model is having higher resolution for the detection purpose.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	i
ABSTRACT	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER 1 INTRODUCTION	1
1.1 Object detection	1
1.2 Image Processing	3
1.2.1 Phases of image processing	4
1.3 Internet of Things	6
1.3.1 Characteristics of IOT.....	6
1.3.2 Advantages of IOT	7
1.3.3 Disadvantages of IOT.....	7
CHAPTER 2 LITERATURE SURVEY	8
2.1 Application of Internet of things in image processing	8
2.2 Object recognition datasets and challenges : A review	8
2.3 Localization-aware channel pruning for object detection	9
2.4 Ascnet: 3D object detection from point cloud based on adaptive spatial Context features	10
2.5 Wi-Fi door lock system ESP32 CAM based IOT	11
2.6 Cloud based surveillance using EPS32 CAM	11
2.7 IOT based home security through digital image processing algorithm	12
2.8 Open CV for computer vision application	13
2.9 Two-stage multiple object detection using CNN and correlative filter For accuracy improvement	13

2.10 A comprehensive review of object detection with deep learning.....	14
CHAPTER 3 SYSTEM DESIGN	15
3.1 Image capturing by ESP32 CAM	16
3.1.1 Specifications of ESP32 CAM	18
3.1.2 Application	18
3.2 Image processing for object detection using artificial intelligence	
System developed in python	19
3.2.1 You Only Look Once (YOLO)	21
3.3 Audio generation for proposed system	24
CHAPTER 4 CODING,TESTING	25
CHAPTER 5 CONCLUSION	29
CHAPTER 6 FUTURE ENHANCEMENT	30
CHAPTER 7 REFERENCES	31

LIST OF TABLES

3.1 Comparison between Models	21
-------------------------------------	----

LIST OF FIGURES

1.1 Examples of Single and two stage detectors	2
1.2 Stages of Image processing	4
3.1 Stages in developing AI EYE	15
3.2 ESP32CAM AND OV2640 CAMERA	16
3.3 ESP32 CAM PINOUT DIAGRAM	17
3.4 Working of YOLO	21
4.1 Generation of URL	25
4.2 Laptop	25
4.3 Cell phone	26
4.4 Knife	26
4.5 Spoon	27
4.6 Fork	27
4.7 Two objects detected	28
4.8 Three objects detected	28

CHAPTER 1

INTRODUCTION

1.1 Object detection

Image processing and the Internet of Things have been used for a variety of applications. The individual applications of these technologies in industries have had some success, but the combination of these two technologies has made no recent progress. [1]. Object detection is one of the notable areas in deep learning and computer vision. Object detection is a technology that comprises computer vision and image processing and is used to detect objects in real-time applications or in stationary applications. The goal of object detection is to develop a computational model or set that comprises basic objects that sort out the basic problems of computer vision and, at the same time, can be used in real-time applications. e.g., if we take the self-driving car, it needs the help of both computer vision and image processing. That is, when the car is moving, we need the help of computer vision to see the objects, which can help us avoid obstacles, and at the same time, we need to know the distance between the moving objects, which can provide alerts and guidelines on how to cross the objects.

Object detection methods are of two types image processing techniques and deep learning methods. Image processing techniques do not need previous stored or detected images for object detection and the sorting of data are very poor in nature for this method. Open CV is a popular library for image processing. In this method the process in which the image is detected does not depend on the previously detected objects and the if they try to detect object they the model tries to train themselves to an extent in which it can make predictions. The object detection in this method is very complex as it involves many unwanted disturbances and when the model is trying to detect an object the or while processing the image we need to capture when there is proper lighting and no shades should appear while capturing the image. The reliability of data set is very high as object detection process is entirely depends on it[2]. Deep learning methods generally depend on supervised or unsupervised learning, with supervised methods being the standard in computer vision tasks. Deep learning method is the second method which can be used for object detection and it is

also the advance model compared to the image processing techniques. This method is designed in a superior way of recognising objects and it uses lot space and memory compared to the image processing method. In this method while capturing the image we can classify the objects in the image in very short time because of the data set which we are using comprises of lot images which are previously detected and also it can be identify the image even if it is partially visible in an image .

Object detection can be carried in single stage or double stage. Object detector can solve two simultaneous problems at same time that is to find the number of objects and then classify every single object with boundary box. The two stage object can performs the object detection in two steps (1)the computer vision needs to identify the region where the objects are located as it reduces the unnecessary loss of time for the next step as the scope of region is narrowed down to smaller size ,(2) since the region of search is narrowed down we can use the box to detect the objects so the trained AI model will display the label of the object and suppose if multiple objects are detected in the region then the boxes might display label of different colour and which might be easy for identification of object. Two-stage detectors are having very high accuracy compared to Single-stage detectors but the overall processing speed is slow compared to the Single-stage detectors because in Two-stage detectors the time consumption for each step it takes is the reason which it achieves the highest accuracy compared to other detectors . The various two-stage detectors include region convolutional neural network(RCNN),with evolutions faster R-CNN or Mask R-CNN. The latest evolution is the granulated R-CNN.

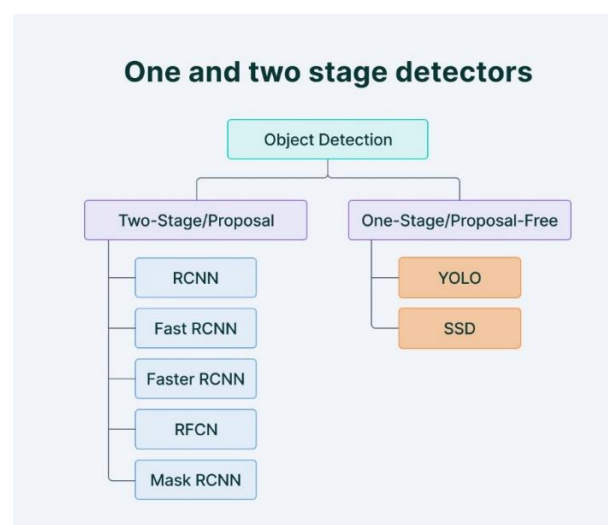


Figure 1.1: Example of Single stage and two stage detectors

Channel pruning methods are used to accelerate the convolutional network layers, which not only reduces the model size but also improves the inference speed [3]. Single-stage detectors are the detectors that are going to take a single step in the object recognition process. In this method, instead of narrowing down the scope of the search and identifying it while capturing the image, the detectors put the boxes in the image for all the objects available. This method uses less time, and therefore the recognition process is faster compared to the two-stage detectors. The processing speed of the detectors makes them suitable for real-time applications like face recognition, self-automation cars, etc. The most popular one-stage detectors include YOLO, SSD, and RetinaNet. The second important advantage of single-stage detectors in object detection is that the structural data set is simple compared to the complex two-stage detector, and the efficiency of the detectors is also much higher compared to two-stage detectors.

1.2 Image processing

Image is represented by its dimension based on the number of pixels. Image is generally two dimension and is represented in the form of (x,y). Pixels are sub-units of image which has specific characteristics such as colour, shade, opacity etc . Image processing is the process in which the captured image which is converted into digital form for processing information. Image processing is an important feature in which it can extract all important features in an image as it can be useful for various applications and it can also enhance the quality of the image if the image has to be restored that is to sharpen the edges of the object which are visible in the image. Tools for image processing like OpenCV, Matlab , TensorFlow ,CUDA,PyTorch etc. The most famous tool among them is the TensorFlow which is a free , open source software which is used for mathematical programming and is constantly used in the object detection process and in higher level computer vision application.

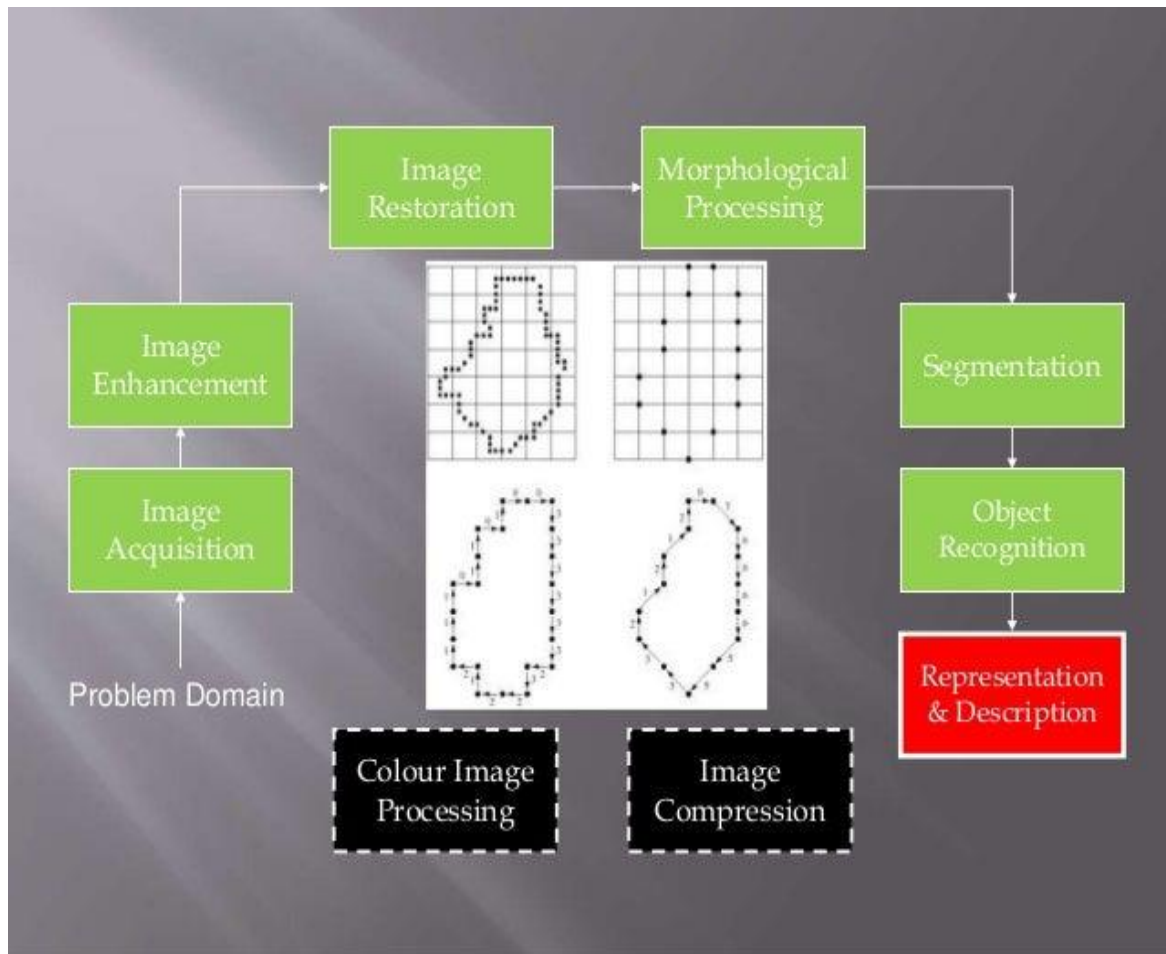


Figure 1.2: Stages of Image processing

1.2.1 Phases of image processing

1. Image Acquisition

The image is captured by the sensor (i.e camera) so we need to convert into a digital form to process it and we need analogue to digital converter to do the digitization process. Digitization is process of converting information into digital form.

2. Image Enhancement

Image enhancement is the process which is required for certain applications and they are manipulated in such way that they are better than the original image. Enhancing an image means to bring out the minute details of the image as it might be important to the application in which it is used.

3. Image Restoration

Image restoration is the process in which we need to enhance the quality of the image in such a way that the captured image might be blur or colour of the object might be faded so we need to process them with such proper precautions because it might be a crucial part of information .

4. Colour Image Processing

It is an area that has been gaining popularity because of the usage of images over the internet. It focuses on the colour of the image for the extraction of useful information and has some application in computer vision , biomedicine, telecommunication.

5. Image Compression

Image compression is the process is the process which deals with size of the image as the image needs to be of smaller in size of storage purpose as well as for the transmitting purpose as it save the storage capacity of the device in both ends and the quality of the image is also not compromised because of the size .

6. Morphological Processing

Morphological process is a process in which we need to focus on the shape of the object of the images as it provides information regarding the object and if the shapes of the object are not clear then we need to improve the edges of the object in the image by using tools dilation , erosion etc.

7. Image Segmentation

Image segmentation is one of the toughest phase in the image processing . Segmentation is the process of segmenting the image into its constituents parts. The more accurate is the segmentation the more accurate is the recognition process and the detection time is also low when the segmentation process is efficient .

8. Representation and Description

Representation is of two types (1) boundary representation which focuses on the corner of the images (2) regional representation which focuses in the texture or the skeletal shape of the image. These may

might be part of the solution which might be helpful in converting the raw information into a form which might be suitable for subsequent image processing process.

Description is otherwise called feature selection which are used to extract the useful attributes in the image for important process.

9. Object Recognition

Object recognition is the final phase of image processing and this process assigns the label to the detected object in the captured image provided by the description.

1.3 Internet of Things

The simple way of understanding IOT devices are that these devices are able to connect and operate other devices with the help of Wi-Fi commonly known as internet. The IOT devices have sensors which are embedded in them. The common examples of IOT devices are air conditioner which can be controlled by or phone, smart cars providing the shortest route, smart watch which is monitoring our daily activities like walking, jogging, heartrate etc.

These huge amount of data are stored in common IOT platforms which make them easily processable at the same time the security of the data is also confidential about the particular user so that these devices are highly safe. The data which are processed by the common IOT platforms sort out the data of the particular user give an accurate analysis is given by extracting important information as per the requirement and the result is shared to other users for better user experience and better automation and improving the efficiencies of the device.

1.3.1 Characteristics of IOT

1. Connectivity is an important characteristics of IOT devices as it can be connected from any place and at any time.
2. Scalability the number of IOT devices it connect and manage at the capacity of withstanding a large number of workload.
3. IOT devices should be dynamic and self-adaptive in nature because it is capable of working in different environments.
4. Intelligence of IOT device depends on the data which it produces from the sensor of the device and it can be made to extract useful information.

5. The speciality of IOT device is that the self-upgradation as it will upgrade itself without the help of human .

6. The architecture of IOT devices are designed in such way that they should be able to operate with number of different protocols

7. The security of IOT devices are not compromised with the connection of number of devices and its privacy is well maintained.

8. Since the increasing number of devices the communication between devices is also sorted because for communications it is not necessary that both devices should be in network

9. The data gathered from devices are used for future analysis.

1.3.2 Advantages of IOT

The advantages of IOT devices are:

1. Controlling and automation of devices
2. Analysing of data for future use.
3. Real time application like surveillance of security camera
4. Save human resources
5. Scalability of devices
6. Connectivity of devices

1.3.3 Disadvantages of IOT

The disadvantages of IOT devices are:

1. Security concerns and potential for data leakage.
2. Privacy issues related to the data collection.
3. High network usage.
4. Complexity of device connection
5. Cost of establishment of device

CHAPTER 2

LITERATURE SURVEY

2.1 Application of Internet of things in Image processing

Narendra kumar and all proposed in the paper “ Application of Internet of things in Image processing ” the use of Internet of things in daily life has improved a lot for eg home security system which is build from IOT is a very important way of surveillance and these devices can be build in any type of location irrespective of temperature of region , location etc except that the network range might be limited in some devices and the network connectivity might be a problem in some rural area and the house can be frequently monitored from any part of the world and at any time so these technologies are very powerful tools in modern day life and other applications of IOT are like monitoring the environmental conditions of silkworms for optimal cocoon aggregation because as we know that the cocoon formation is during one of the hibernation stages of silkworm and it is stage where the silkworm is not safe subjected to any type of danger so that for the extraction of fibre from the cocoon artificially design an environment so that it can be properly monitored and the it can be extracted from it without any problem and at same time making the silkworm safe from all harmful situations, another major application of IOT is that it can be used in agriculture area in way that the for the device to detect which are ready to harvest and it can be used for daily purposes. The installation of these device in the particular field where the crop grows and we need to set the parameter values so that we need to see the fruit or smell the fruit to see whether they are ready for harvest are not so it reduces the amount of time it normally requires for labour work which need to be done during the harvest season[1].

2.2 Object recognition datasets and challenges : A review

Aria Salari and all proposed in the paper “ Object recognition datasets and challenges: A review” object recognition is one of the major function of computer vision as it serves number of application with it. The object recognition process is carried out with the help of the camera lens which is fixed and the model or the type od lens which is attached captures the image. The quality of image entirely depends on the camera lens and at the same time if the captured image is not clear than the object might not recognized. There are situations in which the image will be present in the dataset but due to the improper capturing of image it

might not be properly assessed and there might be another case in which the dataset in which the images are present are trained the dataset might be insufficient. So the artificial intelligence of the computer has to constantly updated otherwise the recognition of objects might be difficult in higher applications since it seems to be tedious process it can be made in such way that the it can be made self-learning or made to adapt in all different situations. The constant challenge with the object detection data set is that the discovery of new objects need constant updating of software or the dataset. The speed of object recognition is one of the important factor which might decide whether it can be used in real time application or not. Since a lot of devices are present with a large number of algorithms and we need the best one which can be used in any situations and at any time except the data handling capacity. Some of the important task of computer vision in our daily life like security surveillance as it is the most important way of ensuring the safety of the house when we are in abroad etc and at the same time we can also set intruder message to our phone so that when any attempt of theft happens we get can message immediately or we can develop a system in such away that it reaches the nearby police station of emergency purpose, self-driving car are developed by creating an sensor which is installed in the car and detect the object in the road and to detect the speed or the distance of the moving object which is near the car so that it can travel without any problem and the safety of the passenger is also not compromised even in medical field also it plays an very important role. Therefore the data set for these application are very important as these data are to be stored in the correct header [2].

2.3 Localization-aware channel pruning for object detection

Li Zhu and all at all proposed in the paper “ Localization-aware channel pruning for object detection ”channel pruning is one of the important model compression techniques used in the object detection for better performance and accuracy. Pruning techniques are of two types one is channel pruning and other is filter pruning since the usage of the channel pruning the detection time has reduced significantly compared to the filter pruning. Channel pruning is generally carried out in three steps (1) channel selection is first step in which select the channels which can give out the output with at most accuracy (2) winnowing is the second step after the channel selection of the after the selection of channel in the first step we need to see what other modification or changes need to be done so that it makes them easily processable for next step (3) weight and bias construction is the final step in which we perform some linear regression or some mathematical calculation for the weight reconstruction to get the accuracy for the object detection in the image. These steps are far

more superior and easy compared to the layer by layer pruning . However detection is much more easy and simple compared to the classification of objects. The ground work required for the object detection is very tremendous and the task is not easy because the data set which is required for the creation of the object detection is not easy because if we take for a particular image the each and every orientation of object should be taken into account because during image capturing if any part of the orientation of the object is missing despite of the object present in the data set than the computer vision might not recognize and the detection speed might also be affected despite the object present and the orientation of the object missing and usage of this method is emphasized compared to other methods because to reduce the computer resources and to save a lot of time [3].

2.4 ASCNet: 3D object detection from point cloud based on adaptive spatial context features

Guofeng Tong and all proposed in the paper “ASCNet: 3D object detection from point cloud based on adaptive spatial context features ” three dimension object detection is what is used in the real time application of the world so that the discovery of such software or the deep learning model is important which can be used for the higher application of computer vision. The most famous application of 3D object detection and is also the most realistic usage of it is self-driving cars. Its application is not easily achieved because to achieve the design of this AI model to serve the purpose of detecting an object in three dimensional surface was not easy and it has to overcome various problems so that it can be achieved. The software in which we are discussing is the ASCNet which is works based on the principle of the spatial awareness around the object which it has to detect so it going to detect with high accuracy. The object are considered as point in the software with respect to the surroundings which is being simulated in the software for the object detection. The three dimension object detection is should be carried out in any environment irrespective of the irregular surface or the area of the surface because there might be applications or situations in which the object detection should be carried. The projection of the objects are which are considered as pseudo images in the cloud so the neural networks which is used in the AI model is utilized effectively to reduce the detection time of the object with high accuracy and also the learning ability of the model can also be developed to reduce the manual effort required by the humans in updating the software so the development al these characteristic features might help the application of software to reach new heights and at the same time sensors which

help them in detecting the speed of the surrounding objects which are both the 2D and the 3D objects gives them the idea that these kind of applications are possible.[4].

2.5 Wi-Fi door lock system ESP32 CAM based on IOT

Dilip prathapagi and all proposed in the paper “ Wi-Fi door lock system ESP32 CAM based on IOT ” the security of the home or data security is important nowadays so combination of an Wi-fi door lock with ESP32 CAM module will enhance the security of the house. Digital door lock is indeed a great technology which has provided a great achievement in improving the safety of the house. Digital locks are of different types depending on the requirement of the people so for example we can have lock which can be open with the help of a card because the cards are encrypted with a particular type of code which is used as an key for the lock and without the card the we might be able open the door so that we ty to access the lock by using any other card it might get locked or it might give an intruder alert message, second we can use lock with our face recognition or fingerprint since all these locks are highly secure these locks are connected to the IOT devices with the help of wi-fi network and we might be able to receive notifications any time when the devices are connected and for even higher level of security we will be using retinal scan which might be the highest level of lock which is currently used for national security and with the help of the ESP32 CAM module we will be able to create a door lock with face recognition in a budget friendly way and that to in an customized way based on the requirements and set up intruder alert when someone not in your database tries to enter your house and along with that we can also set number password for double protection of safety purpose. In short all the locks of modern device are fully connected to our wi-fi network and easily notifiable about the intruder alert[5].

2.6 Cloud based surveillance using ESP32 CAM

Abhishek Mehra and all proposed the paper “ Cloud based surveillance using ESP32 CAM ” surveillance by any means is only useful when it is live streaming. The main purpose of surveillance is that it can ensure the safety of the documents which might be present physically or digitally. The ESP32 CAM microcontroller along with the camera lens which is OV2640 which is having resolution of 1600×1200 which is very high so that the usage of it also of very much importance. Since the transmission of these videos might be transmitted or stored in a particular cloud service which makes them easily accessible so that videos might be high clarity and we can find every actions of people when we are not at home. Since the

videos are stored in the cloud service it can be easily retrievable and at the same time they cannot be erased easily because of its high protection and there are several features which block the unauthorised user to access because of its strong firewall and we can also see the live video transmission just by logging into the particular website all we need is the connectivity of the Wi-Fi network and to achieve this high end security it is developed using a famous backend software called Django because its special characteristics we have selected in our security usage. It has the ability to create an user interface with less programming and it also reduces the space required for the operation of this application and also the understanding of this programming is also easy and number of applications it can run at a time is phenomenal[6].

2.7 IOT Based home security through digital image processing

algorithm

A.B Dorothy and all proposed in the paper “ IOT Based home security through digital image processing algorithm ” automatic personal identification in access control has become popular instead of pattern , passwords and cards. IOT devices are growing rapidly in present days and the discovery of new IOT devices which have the ability to incorporate the features which might support the security of the house. Since the digital image processing is also used in the IOT device we need to be careful in selecting the IOT device which support them. The digital image processing all start with the capturing of image and need to convert them into digital form to process it for further detection and during these stages there phases of image processing in which we need to find the object hidden in the image or the find the shape of the image and if the image is blur then we can use software tools to improve the quality of the device and for detecting the object in the image they should be properly trained in such a way that we can process all the minute information for the comparison of data or analysis. Digital image processing can play an important role in the home security so we need to create an software or algorithm which can reduce the time of each phases which can be utilised for real time application. OpenCV is open source library which maintains all the dataset of the objects which are trained by AI and usage of this library function is crucial for the security purposes in designing the lock in an customised way [7].

2.8 OpenCV for computer vision application

Naveen Kumar and all proposed in the paper “ OpenCv for computer vision application” OpenCv is a library function which is used in many object detection process so that the application of this library function is very important .OpenCv library function is incorporated with many library function to perform many applications of computer vision. The computer designs the study of OpenCV to understand the content of the images which might be object , text etc. This library function has many dataset for recognition process either it might be a text or an object. The dataset for each process is designed and trained for maximum number of objects so we perform many changes to the given image like reducing the intensity of the image, filters and text to the object present in the image. The algorithms which we are using is highly advance so the object detection can be carried out in single step or two-stage detectors. If it focuses on the speed of the detection we can go for single stage detectors like YOLO,SSD etc ad if the algorithm focuses on the accuracy then we can use the two-stage detectors like RCNN, FastRCNN, Mask RCNN etc [8].

2.9 Two-stage multiple object detection using CNN and correlative filter for accuracy improvement

Fabiola Maria Teresa Retno Kinasih and all proposed in the paper “Two-stage multiple object detection using CNN and correlative filter for accuracy improvement” Convolutional Neural Networks is a type of algorithm which is used in the image processing. CNN is method which is used in the two-stage object detectors which focuses on the accuracy of the object detection than the speed of the object detection like the single stage detectors. The corporation of a two stage object detectors with a correlative filter will increase the accuracy of the image and at the same time the speed of the object detection will also improve along with it. Deep learning is a computer based process where the computer tries to learn all the recognition of objects, along with the shape, colour and all attributes with help of artificial intelligence.[9]

2.10 A comprehensive review of object detection with deep learning

Ravpreet Kaur and all proposed in the paper “A comprehensive review of object detection with deep learning” the computer has number of computer vision application and for each application the algorithm we need might be different. If the AI model requires an algorithm for object detection with high speed we might need the usage of single-stage detectors YOLO(You only look once) , SSD etc, and if the situation demands an AI model with high accuracy then we can use the two-stage detectors like CNN , Fast CNN etc. and the usage of each algorithms ahs number of application in vast areas like the agriculture , medical , transportation etc[10].

CHAPTER 3

SYSTEM DESIGN

In system design working of each phase is discussed in detail

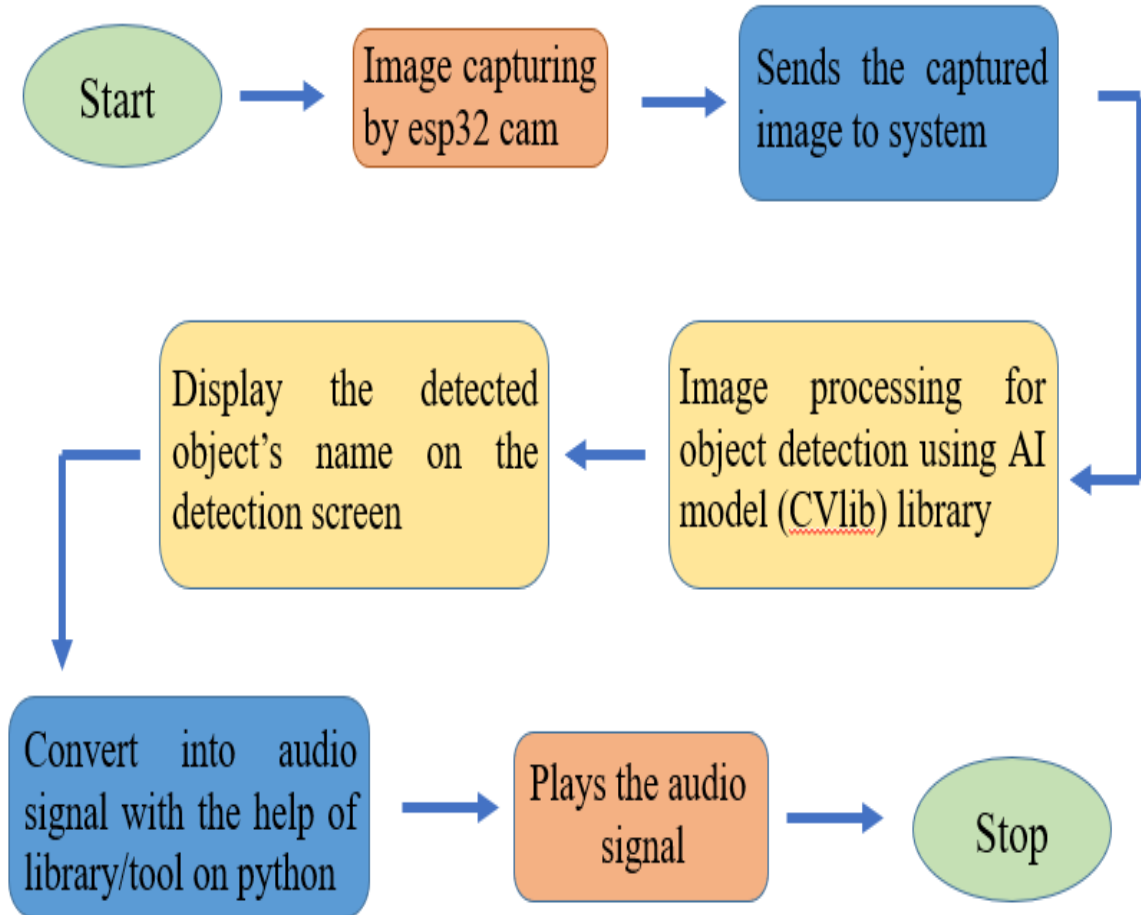


Figure 3.1: Stages in developing AI EYE

The 3 main phases of design methodology in developing an AI eye are :

- a) Image capturing by ESP32 CAM
- b) Image processing for object recognition using Artificial intelligence system developed in python.
- c) The proposed system generated audio with identified image hence any blind can understand.

3.1 Image capturing by ESP32 CAM

ESP32CAM is a microcontroller which has a 32 bit memory which is developed by the AI-thinker technology. The microcontroller has the ability to support wi-fi and connected to other device with the help of Bluetooth as the board which connects the microcontroller serves the purpose. The microcontroller supports ADC, DAC,USART etc. The storage capacity of the microcontroller is that in has 520KB space internally and in addition to it also has the capacity to store the memory space externally with the help of some storage cards with a capacity of 4MB. The EPS32CAM is one of the best microcontroller which can be used in the artificial intelligence eye because of the low heating of the microcontroller and the capacity of the information it can hold therefore the design of these pin are created in such a way so that they can easily be configurable according to the operation of the device and its function in which it can be supported.

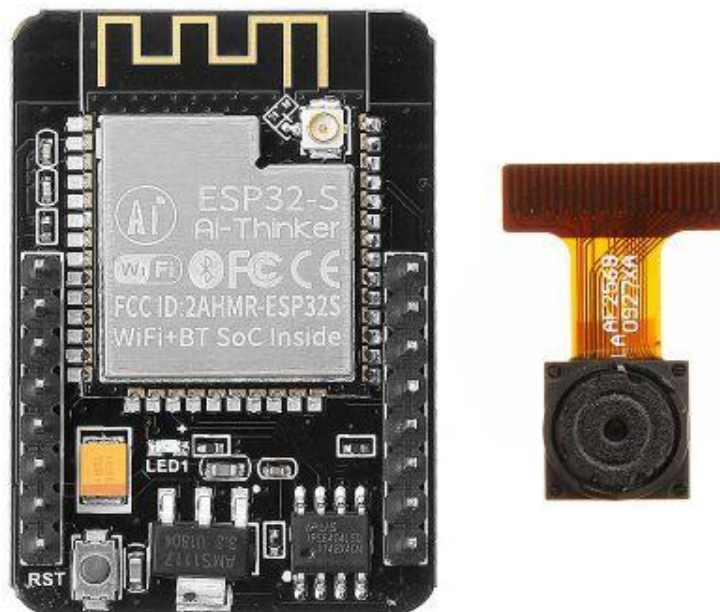


Figure 3.2: ESP32CAM AND OV2640

The microcontroller and the camera lens which is the OV2640 lens combined constitutes the artificial intelligence eye. The lens of the camera can be made to work in three different

resolution among them the highest resolution of the camera is the 1600×1200. The pins of the microcontroller is connected to the board. Suppose if the dataset has to be manually trained in that case we have the advantage of an SD card port which can support up to 4GB memory space so all the orientation of the image is stored in that card.

The OV2640 is an excellent camera which can be utilised for the capturing of image and the feature it support in the setting of the camera are brightness, saturation etc. The image captured by the OV2640 camera does not occupy much space compared to the normal image which is captured by other camera so the processing speed of the image is much more better the capturing of image is also very accurate compared to the other camera lens so it is also an advantage which is not present in any other camera lens so the process of converting an image into other format for the dataset creation will become much more simple process .

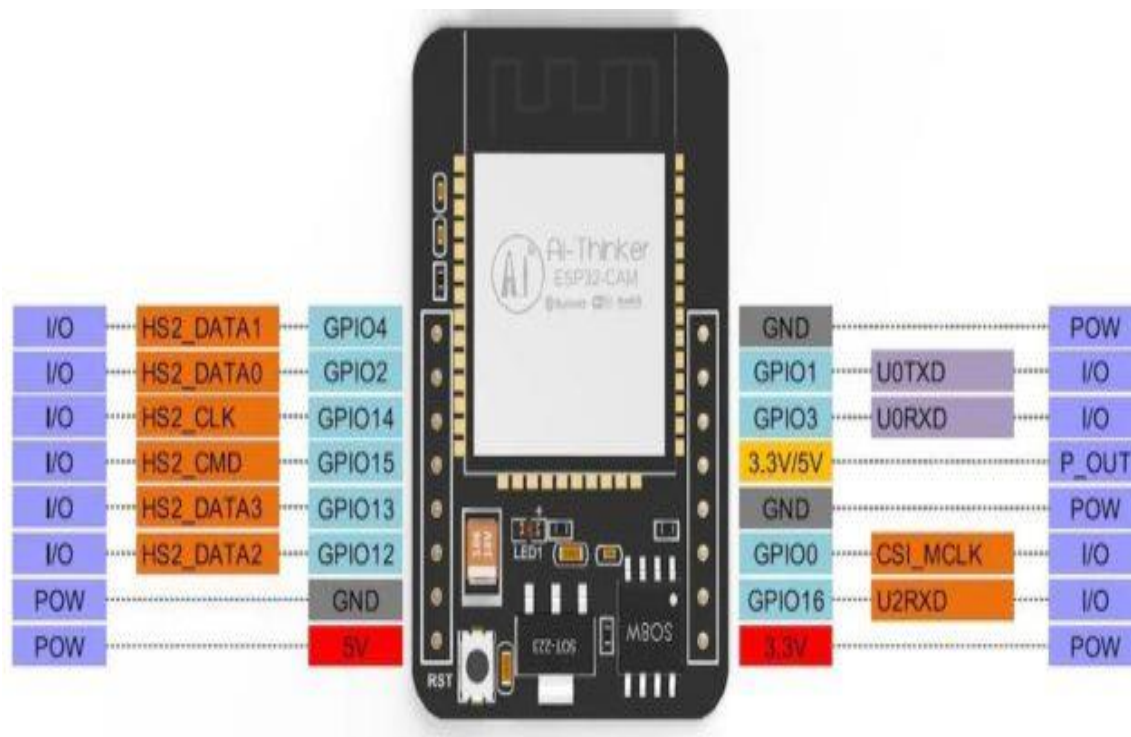


Figure 3.3: ESP32CAM PINOUT DIAGRAM

3.1.1 Specifications of ESP32 CAM

The specifications of ESP32CAM are:

1. Image format : JPEG(OV2640 support only), BMP , grayscale
- 2.Support TF card upto : 4GB
- 3.Wi-Fi protocol : IEEE802.11 b/g/n/e/i
- 4.Power supply: 5V
- 5.Image sensor format:1/4

3.1.2 Applications

The specifications of ESP32CAM are:

- 1.Surveillance camera
- 2.Smart attendance
- 3.Safety lock using facial recognition
- 4.Smart parking system
- 5.Medical field

The ESP32 CAM module can be connected to the bluetooth board instead of FTDI module to reduce the complexity of wires and easy understanding. The first phase is the image capturing which is done by the ESP32 CAM module which has the OV2640 module with it. In this phase we are going to use the Arduino IDE software for the capturing of images as well as the generation of URL.The ESP32 libraries can be downloade in the arduino software itself and othe libraries have to downloaded seperately from the other source.

The required libraries for the programs are:

- 1.WebServer.h- For the microcontroller to act as an server
- 2.WiFi.h- To access the URL which is generated
- 3.esp32cam.h- For accessing all the ESP32 CAM module features

In this phase we are going to code for the generation of URL and the image capturing by ESP32 microcontroller and most importantly the ESP32 CAM module is connected to the laptop with the help of an USB cable. The camera module can be adjusted in three different resolution which is mid, high and low respectively. The connection of the ESP32 CAM module with the laptop which has wi-fi acts as an IOT device and the microcontroller acts as an server when the URL is created for the particular module it can be accessible to both the laptop as well as you mobile. The microcontroller acts as an server because we are giving our SSID and password of our Wi-Fi and when the compilation of the coding is done we have to reset the microcontroller and wait for the serial monitor to display if it displays “CAMERA OK” then it is successful along with the URL in which we can see the image and when it fails it displays “CAMERA FAIL”. The URL screen has some settings in which the captures image is screen with high resolution, low lighting or with shades etc. This process can be used to capture the images and store it in the microcontroller or with the help of the SD card so that it is possible to create an dataset for own real time surveillance as it is the best way and should be able to get message we an unidentified person enters [6].

3.2 Image processing for object detection using artificial intelligence system developed in python

In the second phase we are going to code for the object detection and the type of algorithm we are using for object detection and connecting the arduino code with the python code using the URL.

In the object detection code we are using the URL which was generated in the arduino so that the python program can access the images for the object detection.

The libraries used in the python code are:

1. OpenCV- OpenCV is an python library function which is used for the image processing and the library function has the ability to perform many function with the help of other library function and trained data set are all present in this library function.
2. Numpy- It is an library function which is capable of performing mathematical operations or mathematical manipulation with an array more that one dimension.

3. cvlib- It is the most important library function in case of object detection as it is the library function which contains file on object detection , face recognition, gender detection etc.
4. Matplotlib- Matplotlib is an library function which is capable of plotting graphs, pie chart , and changing the background of the workspace and the changing the colour of the graph in case when multiple plotting of graphs have been done.
5. Concurrent.futures- The function of this library function is that two output windows can run at a time .
6. urllib.request- The function of this library is to get the orientation of the image from the particular URL.

Object recognition are used in many day to day applications like number plate recognition, face detection and recognition , object tracking, self driving cars, robotics [10] etc. The object recognition process is can be updated depending upon the situation it demands. The usage of Two-stage object detectors along with a particular filter can enhance the object recognition process so we need always depend on the old methods. Since the utilisation of filter the object can be easily recognised so we need to just use the appropriate filter that can use reduce the time and the improve the accuracy [9]. The deep learning algorithms are created in such a way that instead of manually training the data set and updating the algorithm we can design in such a way that the computer can learn all the new , understand and extract the information which needs to be utilised for the further recognition process.

In object recognition process there is the need of multiple library because there is inter dependency of each library on one other. COCO is the data set which is used in the in the particular model and the algorithm which we use id YOLOv3.The choice selection of this algorithm is that YOLO algorithm has number of advantages compared to other object detection algorithms in terms oof speed of object identification and the precision of object detection in dull or blur region in the image because of this the usage of these algorithm are very commonly used in the real time application like self-automation car and many more . OpenCv plays an crucial role when it comes to computer vision and it plays an important role such as drawing boxes, resizing image , rotating image etc [8].

3.2.1 You Only Look Once (YOLO)

You only look once is the most ideal algorithm for real time applications because of its training methods for dataset and the accuracy in detecting the objects. There are several versions of YOLO architecture which are based on the update of various versions. YOLO algorithm uses convolutional neural networks for object detection. Unlike the outdated method of object detection the YOLO algorithm needs only the image for the recognition process and the bounding boxes are its characteristics features in it are used in such a way that the object is bounded by boxes present in the image. The detection process is in such a way that it can detect multiple objects at the same time and the probabilities of a particular class detected and are displayed along with the label of the detected objects .

3.2.1.1 Working of YOLO

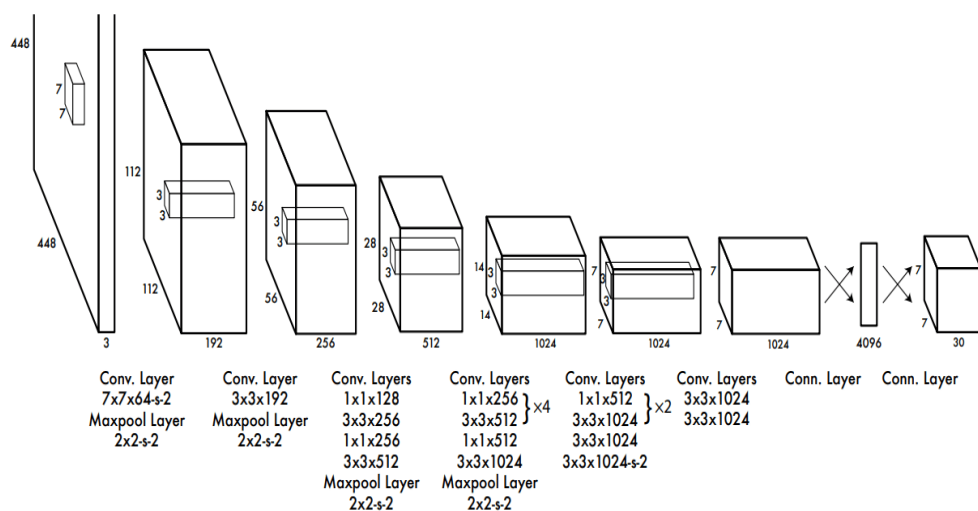


Figure 3.4 : Working Of YOLO

Model	mAP	combined	gain
Fast-RCNN	71.8	-	-
Fast-RCNN(2007 data)	66.9	72.4	.6
Fast-RCNN(VGG-M)	50.2	72.4	.6
Fast-RCNN(cafe-Net)	57.1	72.1	.3
YOLO	63.4	75.0	3.2

Table 3.1: Comparison between models

The YOLO architecture utilise the convolutional neural networks and object detection is carried at three different layers. Each convolutional layers are followed by normalization. Since YOLO algorithm utilises convolutional layer in has the ability to not to miss out any leakage of data because of its ability the smallest of smallest objects are not missed out on the image even though the image which is captured might be of low quality or the brightness of the image is low or the image is blur or the image captured might be of low saturation. etc

The YOLO algorithm works by first capturing the image with high quality even though the image is of low quality it can be used for the detection process. So the main way to approach to the detection process is to first divide the image into number of grids. The grids are now the main sub-units which is are helpful in detecting the objects in each grid in the following manner suppose the in the grid if there is an centre of an object than that grid can be used for that particular object detection now we use the boundary boxes in the grid to find the probabilities of the class of the object in which it is present this is the way in which the algorithm works in an accurate and efficient way .

3.2.1.2 Advantages OF YOLO

1. The number of datasets which the algorithm can store is of very large number so the storage capacity of the system is not a problem and at the same time it can be customised so that it can be used to many applications.
2. YOLO algorithm accuracy rate is very high because the algorithm uses the boundary boxes for the object detection which makes them easily recognisable in very aspects to not miss every single information.
3. The grid technique which is used for the image segmentation is an excellent way to reduce the time required for the object recognition for each grid and the grid alone is responsible for object detection.

3.2.1.3 Limitations of YOLO

1.The main drawback of the YOLO algorithm is that the objects present in the image are detected by the concept of boundary boxes but the it is does not has the ability to detect the small objects present in the image .

2.The second important issues of the YOLO algorithm is that is does not have the ability to detect the multiple objects are placed in cluster because while detecting the AI technology might miss out an object or by mistake it might recognise some other object.

3.2.1.4 Applications of YOLO

YOLO algorithm can be applied in the following fields:

1. **Autonomous driving:** YOLO algorithm is used in the real time application which is the automation car because it has the ability to store an large number of objects in the dataset which makes it possible to detect the objects in the road and avoid them without any collision .
2. **Wildlife:** The usage of this algorithm in the wildlife sanctuaries is very helpful because it helps them to keep record of new species of animals in the forest and then store it in the dataset for object detection for future purpose at the same time the keep the record of the animals present in the forest.
3. **Security:** YOLO algorithm can be also be used in the security purpose because the dataset can be customised in such away that it stores the image of the person with particular name or employee id and when an unauthorised person tries to enter then we can suddenly get the intruder message.

In this phase of coding we are going to use python software where we are going to define two functions. One of the function we are using it for live transmission of objects and the other function we are using it for detection purpose. In the function which we are using it for detection we have used label and box for proper detection and in case of multiple objects detected the boxes are displayed with different colours to avoid confusion and when the compilation of code is done we are going to get two screens live transmission and detection and when the camera module captures the image it compares the image with the database of previously acquired images to display the output.

3.3 Audio generation for proposed system

In this phase we are going to incorporate the audio to the system so that the blind people could hear them. The audio can be given with the help of program or with the help of library function .

The library function used is this code is pyttsx3 which is used to convert the text to speech and Engine is an instance tool which is very easy to use tool which helps to convert the text to speech.

engine.say() we have to type the label so that the audio is generated and engineAndWait() is helpful when multiple objects are detected in the screen the computer takes some time to analyse and generate audio.

CHAPTER 4

CODING,TESTING

```
10:04:20.935 -> load:0x40078000,len:13924
10:04:20.935 -> ho 0 tail 12 room 4
10:04:20.935 -> load:0x40080400,len:3600
10:04:20.935 -> entry 0x400805f0
10:04:21.652 ->
10:04:21.855 -> CAMERA OK
10:04:24.457 -> http://192.168.43.215
10:04:24.457 -> /cam-lo.jpg
10:04:24.457 -> /cam-hi.jpg
10:04:24.457 -> /cam-mid.jpg
```

Figure 4.1: Generation of URL

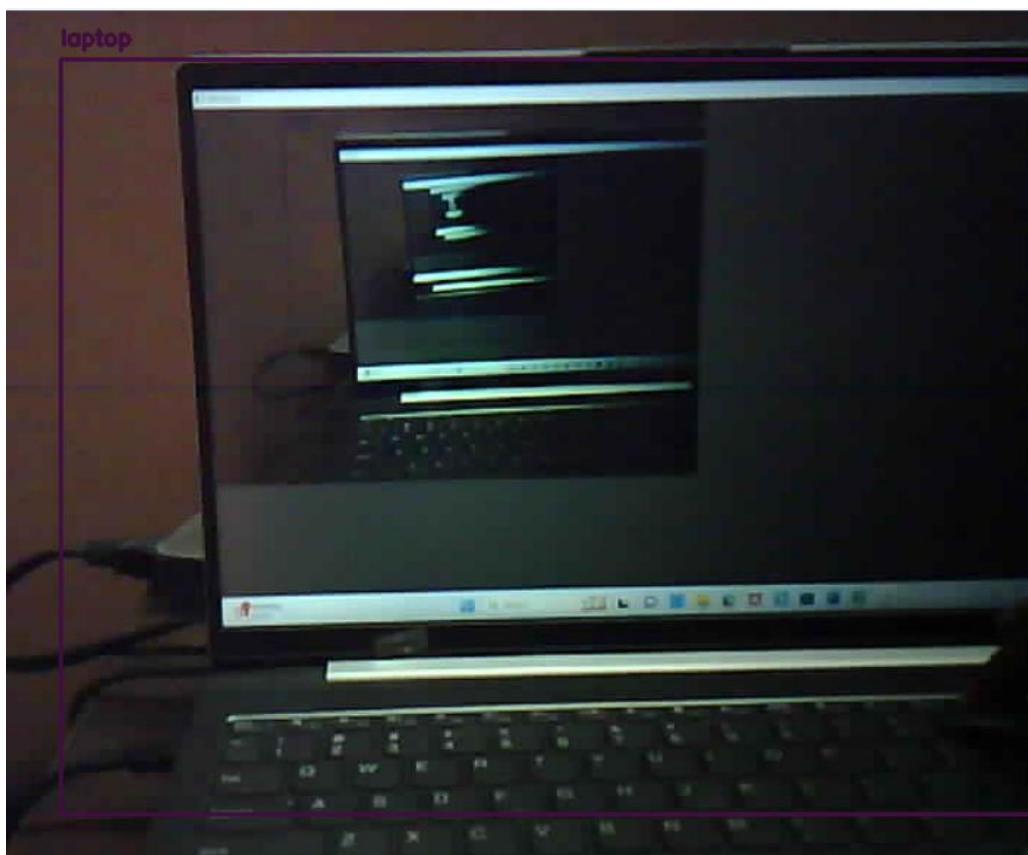


Figure 4.2 : Laptop

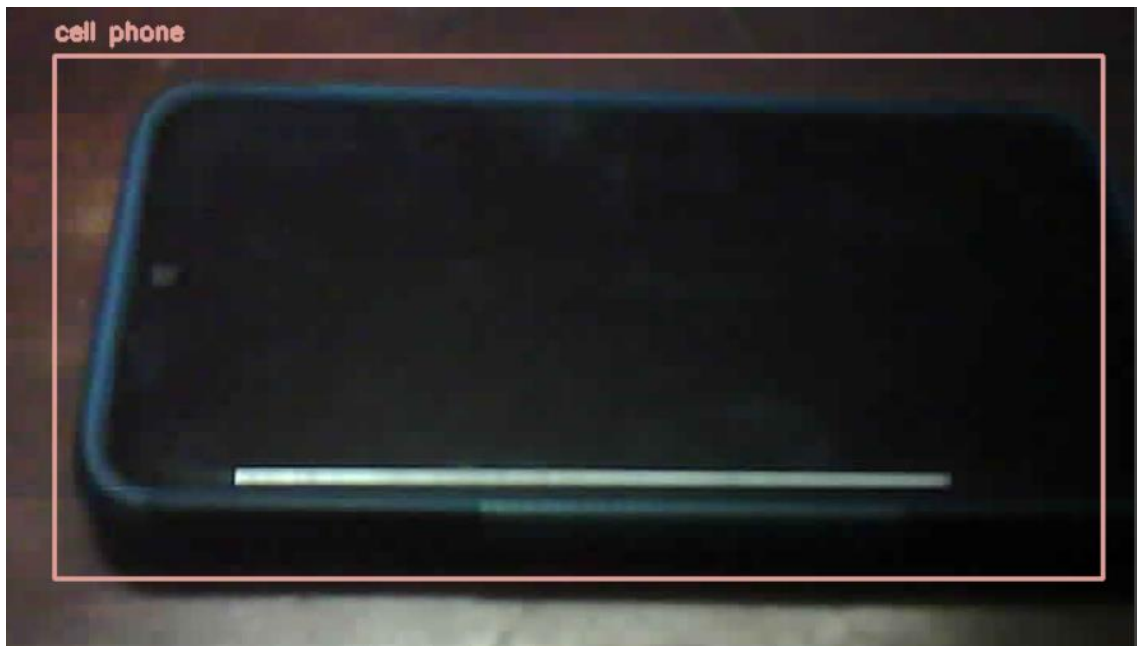


Figure 4.3 : Cell phone



Figure 4.4 : Knife

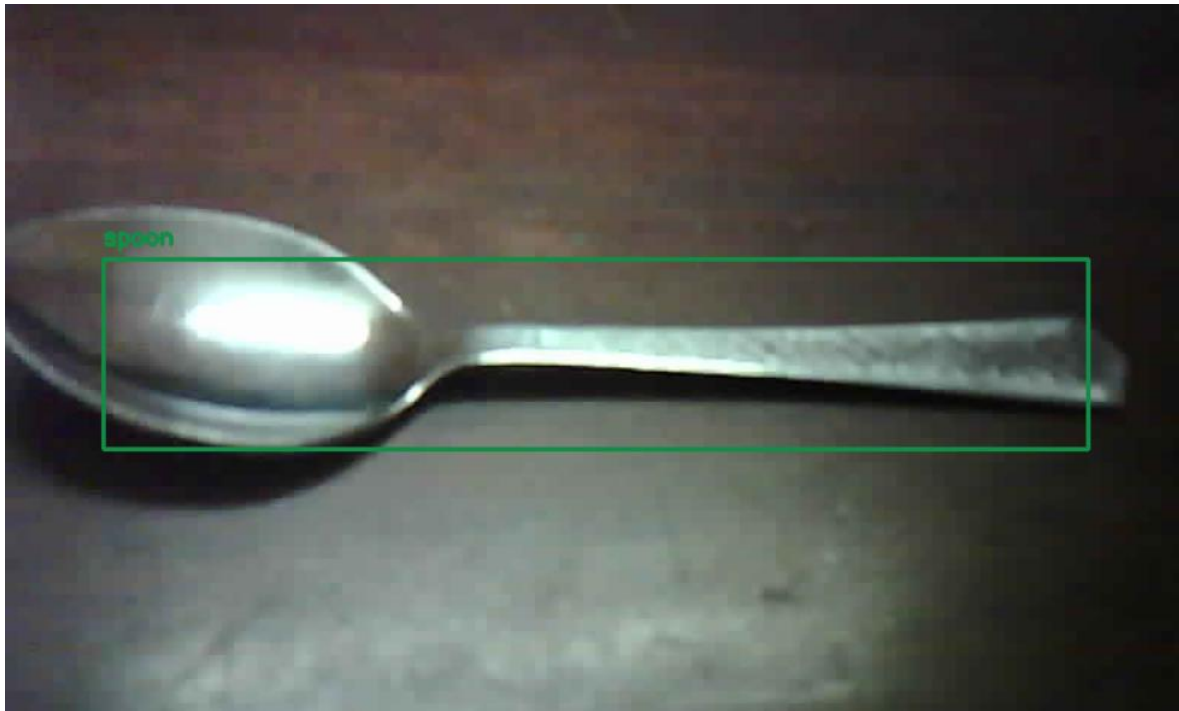


Figure 4.5 : Spoon



Figure 4.6 : Fork

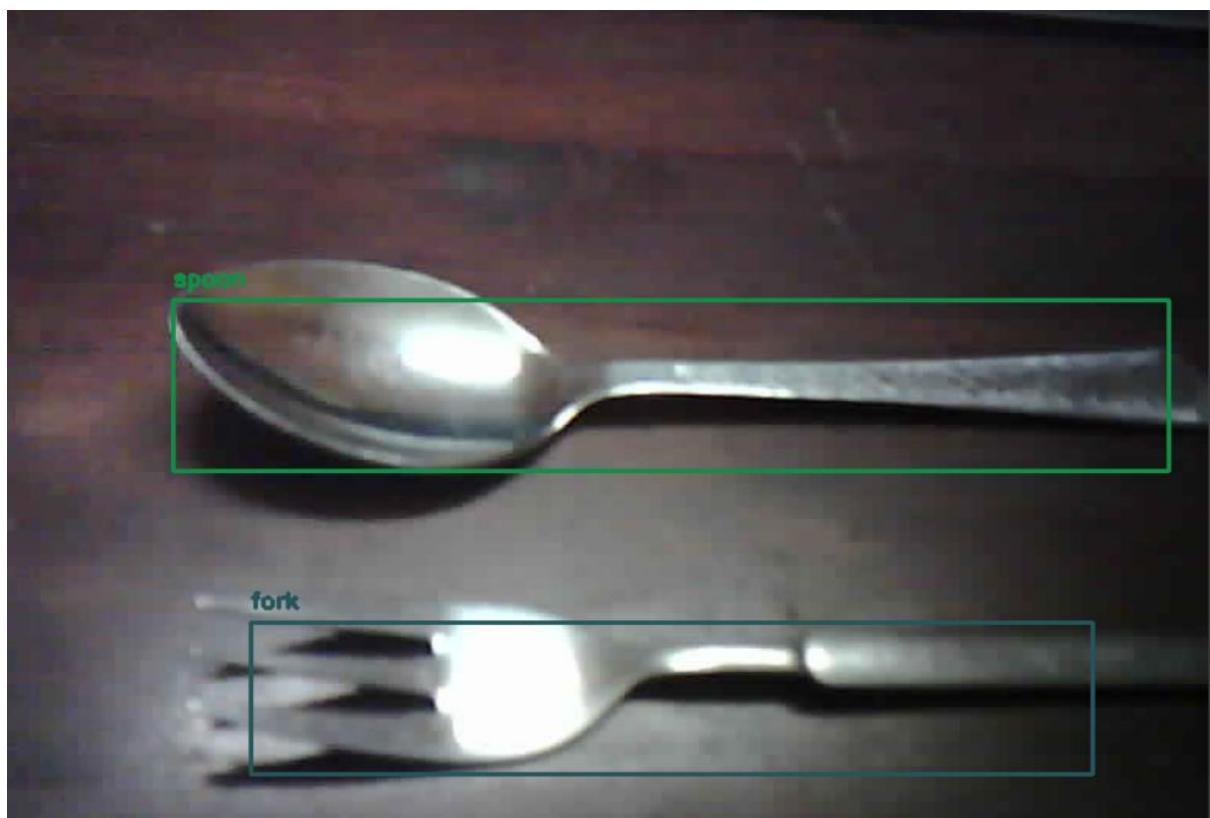


Figure 4.7 : Two objects detected

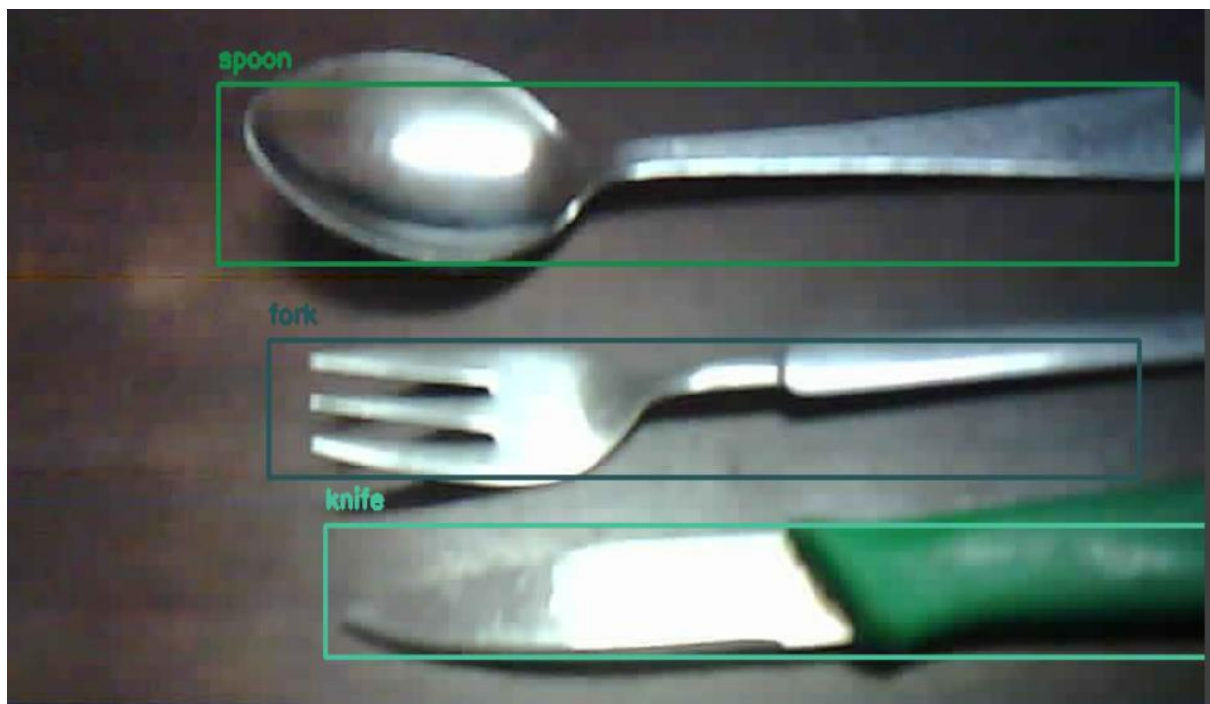


Figure 4.8: Three objects detected

CHAPTER 5

CONCLUSION

The artificial intelligence eye is developed with the help of ESP32 CAM microcontroller and OV2640 camera lens. The software used are arduino and python. The arduino software is used for checking the camera lens and the generation of URL. The python software is used for object detection with the libraries like cvlib, OpenCV, Matplotlib, Numpy, which helps in importing all the necessary dataset. The object detection algorithm used is YOLO because of its high processing speed the data set for object detection can also be customised based on the need of the person. The audio which is generated helps the blind people in their day to day activities. The artificial intelligence eye is developed in a budget friendly way and in an economic way.

CHAPTER 6

FUTURE ENHANCEMENT

- 1.The future plan of this project is that by using neuro imaging techniques the blind brain can see the images the with the signal from AI.
- 2.Addressing ethical and safety considerations: This involves ensuring that the technology is safe and secure, and that it is developed and used in a manner that respects privacy and individual rights.
- 3.Real-time Processing: Enhancing real-time processing capabilities to ensure that the AI Eye can operate in real-world conditions and respond quickly to changing conditions.

CHAPTER 7

REFERENCES

- [1] Narendra Kumar, Krishna kumar ,“ Application of Internet of Things in Image processing ", 2022 Institute ofElectrical and Electronics Engineers Delhi section conference ,pp 1-5, February2022.
- [2] Aria Salari, Abtin Djavadifar, Xiangrui Liu, Homayoun Najjaran ,“Object recognition data sets and challenges: A review”,Neurocomputing,vol 495, pp 129-152, july 2022.
- [3] Zihao Xie,Lin Zhu,Lin Zhao , "Localization-aware channel pruning for object detection “ , Neurocomputing, vol 403, pp 400-408, August 2020.
- [4] Guofeng Tong, Hao peng, Zheng Li ,“ASCNet: 3D object detection from point cloud based on adaptive spatial context features”, Neurocomputing, vol 475, issue 12,pp 89-101, February 2022.
- [5] Dilip Prathapagiri, “Wi-Fi door lock system ESP32 CAM based on IOT”, The International journal of analytical and experimental model analysis, vol 13, issue 7, pp 2001-2003, july 2021.
- [6] Abhimanyu Jadli, Abhishek Mehra ,“Cloud based surveillance using ESP32 CAM", International conference on intelligent technology,pp 1-5,November 2021.
- [7] A.Beartice Dorothy, S.Britto Ramesh Kumar ,“ IoT Based Home Security through Digital Image Processing Algorithms ”, world congress on computing and communication Technologies , pp 345-356, February 2017.
- [8]Mahamkali ,Naveenkumar & Ayyasami Vadivel ,“OpenCv for computer vision application”, National conference on Big data and cloud computing,pp 52-56,March 2015.
- [9]Fabiola Maria Teresa Retno Kinasih,Carmadi Machbub,"Two-stage multiple object detection using CNN and correlative filter for accuracy improvement", Heliyon,vol 9,issue 1,pp 1-18,January 2023.

- [10] Ravpreet Kaur, Sarbjit Singh, "A comprehensive review of object detection with deep learning", *Digital signal processing*, vol 132, pp 132-140, December 2022
- [11] S. Rokhade, G.M K, M.M S, S. Banu, J.S N and, T.D, "smart sericulture system based on IOT and image processing technique", *International conference on computer communication and informatics (ICCCI)*, pp 1-4, April 2021.
- [12] N.A Anugraheni, A. Suhendi and H. Bethanigtyas, "Image processing of IOT Based cherry Tomato growth monitoring system", *International conference on instrumentation*, pp 207-210, November 2019.
- [13] A. Sharma and N. kumar, "Encryption of text using fingerprints as inputs to various algorithms", *International journal of science and research*, vol 3, issue 4, pp 418-421, April 2014.
- [14] Ahmad, Faizan, Najam, Aaima, Ahmed, Zeeshan, "Image-based face detection and recognition" state of the art", *International Journal of computer science*, vol 9, issue 6, pp 1-4, November 2012.
- [15] Kumar, Ashu & Kaur, Amandeep & Kumar, Munish, "Face detection techniques: A Review", *springer*, vol 42, issue 2, pp 927-948, August 2019.
- [16] Donghai Zhai, Xun Li, Yuxin Zhou, "Object detection methods on compressed domain videos: An overview, comparative analysis and new directions", *Measurement*, vol 207, pp 123-145, February 2023.
- [17] Zubair Ahmad Lone, Alwyn Roshan Pais, "Object detection in hyperspectral images", *Digital Signal Processing*, vol 131, pp 1-14, November 2022.
- [18] Zhenyu Wu, Haibin Yan, "Adaptive dynamic networks for object detection in aerial images", *Pattern Recognition Letters*, vol 126, pp 8-15, February 2023.
- [19] Wenwen Qi, "Object detection in high resolution optical image based on deep learning technique", *Natural Hazards research*, vol 2, pp 384-392, December 2022.
- [20] Siyuan Ren, Xiao Pen, Bo Han, "Dynamic graph transformer for 3D object detection", *Knowledge based system*, vol 259, pp 345-356, January 2023.

[21]Xue-song Tang, Xianlin Xie, Dawei Li," A line-segment-based non-maximum suppression method for accurate object detection ",Knowledge based systems, vol 251, pp 678-689, September 2022.

[22] Jin Yuan ,Weili Guan, Meng Wang ,"Gated CNN: Integrating multi-scale feature layers for object detection ", Pattern Recognition, vol 105, pp 12-24, September 2020