DE-40 (DC&SE)

NUST COLLEGE OF

ELECTRICAL AND MECHANICAL ENGINEERING

Real Time Object Detection And Scene Understanding for Blind

A PROJECT REPORT

DE-40 (DC&SE)

Submitted by

PC AITZAZ BAKHT

NS OSAMA REHMAN

NS FURQAN SHAFIQ

BACHELORS

IN

COMPUTER ENGINEERING

YEAR

2022

PROJECT SUPERVISOR

DR. ARSLAN SHAUKAT

DR. USMAN AKRAM

COLLEGE OF

ELECTRICAL AND MECHANICAL ENGINEERING

PESHAWAR ROAD, RAWALPINDI

YEAR 2022

Real Time Object Detection And Scene Understanding for Blind

A PROJECT REPORT

DEGREE 40

Submitted by

PC AITZAZ BAKHT

NS OSAMA REHMAN

NS FURQAN SHAFIQ

BACHELORS

IN

COMPUTER ENGINEERING

Year

2022

PROJECT SUPERVISOR

DR. ARSLAN SHAUKAT

DR. USMAN AKRAM

COLLEGE OF

ELECTRICAL AND MECHANICAL ENGINEERING

PESHAWAR ROAD, RAWALPINDI

DECLARATION

We hereby declare that no portion of the work referred to in this Project Thesis has been submitted in support of an application for any other degree or qualification of this for any other university. If any act of plagiarism found, we are fully responsible for every disciplinary action taken against us depending upon the seriousness of the proven offence.

COPYRIGHT STATEMENT

ACKNOWLEDGEMENTS

First of all, Alhamdulillah, that our FYP is finally made and all Thanks to Allah for giving us the strength and moral to keep pushing forward and helping us on each and every step of the way.

Secondly, we would like to offer heartily thanks our supervisors, Dr Arslan Shaukat and Dr. Usman Akaram, who helped us a lot, tremendously, on each and every single issue, who’s help ad guidance became a source of strong determination for us. Thank You, sir`s you played a great role in our lives, one that we can never forget.

And lastly, we would like to thank our parents and friends, without whose unimaginable support and constant motivation, we might not have been able to complete our Final year project. They played an unparalleled role throughout our journey, and we are eternally thankful to them. Their constant support motivated us to do more than we ever realized and they inspired new hope in us, when we found none in ourselves.

ABSTRACT

Computer Vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs. There is lot of work available in literature that is based on the manufacturing of visual aid to assist the blind or visually impaired people. This project involves computer vision as detect objects in the environment in real time and produce an audio output. It uses deep neural networks that are trained to detect objects in the environment hence making the project involve machine learning. A blind person entirely depends on someone for his daily life tasks and always wonders how the world is around him. So, for someone who does not even know what is present in his surrounding this project would help to at least try to understand his/her surroundings. A mobile application has been developed for this purpose which will identify objects in the surrounding and make the user understand about the surroundings through detected objects.

The object detection model used is the Mobilenet SSD which is a Single Shot Detector model. YOLO was also tested for object detection but discarded due to very low FPS on mobile devices. TensorFlow Lite library is used for using object detection model on a mobile device. Android Text library is used for converting detected object output into voice. Open CV is used for preparing the object detection model and NumPy library has also been used for python programming part of the project. Some of the software's used include Android Studio for building the mobile application, PyCharm IDE for designing, training, and testing the object detection model. Using all these tools the project can detect objects and describing the environment to the user on all mobile devices as it has minimal requirements like camera access, sound etc. which are already present in all smart devices.

Table of Contents

List of Figures

List of Tables

Chapter 1: INTRODUCTION

The highly increasing ration of impaired people compelled the researchers all around the globe to invent different technologies and devices to assist them in carrying out their daily tasks like any other normal person. Therefore, this project is dedicated to the people who are visually impaired. A blind person cannot see anything. He/she entirely depends on someone for his daily life tasks and always wonders how the world is around him. So, for someone who does not even know what is present in his surrounding this project would help to at least try to understand his/her surroundings.

1.1 Overview

Today's world is a world of digitalization. The growing tech field and exponential development in the fields of transport, medicine, metropolitan cities have become the most influential developments in the lives of mankind. Visual reality, dimensioned reality and augmented reality are the commonly used methods for this purpose. The algorithms used for object detection are SSD (Single Shot Detector [1]), YOLO (You Only Look Once [2]), Fast R-CNN [3], Faster R-CNN [4] and MASK R-CNN [5]. The mobile application designed to detect the objects around and output will be received in the form of audio describing user about its surrounding environment.

1.2 Problem Statement

According to the statistics mentioned in the report of World Health Organization (WHO), over 285 million people worldwide are visually impaired. Therefore, this project is to help the blind people and give them an understanding of the world around them.

1.3 Proposed Solution

The major goal of our proposed solution is to develop a mobile application to make the users aware of their surroundings. The solution is based on Computer Vision and Machine Learning algorithms.

1.4 Objectives

1.4.1 General Objectives:

"To build an innovative state of the art software powered by Machine Learning (ML) and image processing techniques, providing a smart tool to aid visually impaired people."

1.4.2 Academic Objectives:

- ✓ Development of a smart and intelligent tool for blind people

- ✓ To implement Machine Learning techniques and simulate the results
- ✓ To increase productivity by working in a team
- ✓ To design a project that contributes to the welfare of society

Figure 1.1 Artificial Intelligence & Machine Learning [6]

1.5 Scope

This project finds its scope for visually impaired people. It is an innovating state of the art software powered by machine learning and image processing techniques, providing a smart administrative tool to aid blind people in understanding the nature objects present in their surrounding environment.

1.6 Deliverables

1.6.1 Object of interest:

It can detect the objects by using the same combination of image processing and machine learning techniques. By detecting the object of interest, we mean detecting the objects around the person who cannot be able to see them.

1.6.2 Audio Output:

The detected object will be known to the visually impaired person in the form of voice/audio.

1.7 Relative Sustainable Development Goals

The project aligns with following United Nation Sustainable Development goals:

- ✓ Sustainable Cities & Communities
- ✓ Responsible Consumptions

Figure 1.2 Aligned SDGs [7]

1.8 Structure of Thesis

Chapter 2 contains the literature review and the background and analysis study this thesis is based upon.

Chapter 3 contains the design and development of the project.

Chapter 4 contains the conclusion of the project.

Chapter 5 highlights the future work needed to be done for the commercialization of this project.

Chapter 2: Background and Literature Review

A new product is launched by modifying and enhancing the features of previously launched similar products. Literature review is an important step for development of an idea to a new product. Likewise, for the development of a product, and for its replacement, related to visual aid for blind people, a detailed study regarding all similar projects is compulsory. Our research is divided into the following points.

- Industrial Background
- Existing solutions and their drawbacks
- Research Papers

2.1 Industrial Background:

Initially, Pakistan Industries were barren. Then, these started exporting under liberal policies resulting in increase in industrial growth due to the rapid expansion of domestic demand and encouragement for export. Despite of declination of growth, Pakistan managed to make progress and growth in the new century. And now industries are inclining towards smart industries, automation, based on new technologies (Internet of Things (IOT), Machine Learning (ML)

techniques, Artificial Intelligence (AI)). Hence, smart glasses provide good market growth and impacts economy directly as it is a fully automated system.

2.2 Existing Solutions & Drawbacks

Different solutions are previously being provided to aid blind people, but every product has some pros and cons. Following are some solutions which are already being prepared and being implemented

.

- **Smart guiding glasses for visually impaired people in indoor environment**

To overcome the travelling difficulty for the visually impaired group, a smart guiding device in the shape of a pair of eyeglasses is developed for giving these people guidance efficiently and safely [8]. Different from existing works, a novel multi-sensor fusion-based obstacle avoiding algorithm is proposed, which utilizes both the depth sensor and ultrasonic sensor to solve the problems of detecting small obstacles, and transparent obstacles, e.g. the French door. For totally blind people, three kinds of auditory cues were developed to inform the direction where they can go ahead. Whereas for weak sighted people, visual enhancement which leverages the AR (Augment Reality) technique and integrates the traversable direction is adopted. The prototype consisting of a pair of display glasses and several low-cost sensors is developed, and its efficiency and accuracy were tested by a few users. The experimental results show that the smart guiding glasses can effectively improve the user's travelling experience in

complicated indoor environment. Thus, it serves as a consumer device for helping the visually impaired people to travel safely.

- **Med Glasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients**

Today, with the arrival of an aging society, the average age of the population is rising. It is known that the physiology of a person degrades with age [9]. There are approximately 285 million visually impaired people in the world, of whom 140 million are elderly people over the age of 50, and 110 million of these visually impaired elderly people suffer from multiple chronic diseases. In the case of multiple medication usage, these 110 million vulnerable people will be more likely to take the wrong medicines or forget to take their medication. To solve this problem, a wearable smart-glasses-based drug pill recognition system is developed using deep learning, named Med Glasses, for visually impaired people to improve their medication-use safety. The proposed Med Glasses system consists of a pair of wearable smart glasses, an artificial intelligence (AI)-based intelligent drug pill recognition box, a mobile device app, and a cloud-based information management platform. Experimental results show that a recognition accuracy of up to 95.1% can be achieved. Therefore, the proposed Med Glasses system can effectively mitigate the problem of drug interactions caused by taking incorrect drugs, thereby reducing the cost of medical treatment, and providing visually impaired chronic patients with a safe medication environment

.

- **Intelligent Smart Glass for Visually Impaired Using Deep Learning Machine Vision Techniques and Robot Operating System (ROS)**

The Smart Glass represents potential aid for people who are visually impaired that might lead to improvements in the quality of life [10]. The smart glass is for the people who need to navigate independently and feel socially convenient and secure while they do so. It is based on the simple idea that blind people do not want to stand out while using tools for help. The Smart glass consists of ultrasonic sensors to detect the object ahead in real-time and feeds the Raspberry for analysis of the object whether it is an obstacle or a person. It can also assist the person on whether the object is closing in very fast and if so, provides a warning through vibrations in the recognized direction. It has an added feature of GSM, which can assist the person to make a call during an emergency. The software framework management of the whole system is controlled using Robot Operating System (ROS). It is developed using ROS catkin workspace with necessary packages and nodes. The ROS was loaded on to Raspberry Pi with Ubuntu Mate.

- **Lid Sonic for Visually Impaired: Green Machine Learning-Based Assistive Smart Glasses with Smart App and Arduino.**

There is a large body of work focused on the development of mobility assistive devices for visually impaired (VI) people [11]. However, none of them seems to satisfy the needs of VI

people, which might suggest that these requirements have not been considered during the development process. In this sense, this study aimed to develop a novel assistive system based on the opinions provided by a group of VI persons who also participated in the performance assessment stage. Two ultrasonic sensors and one infrared sensor were combined to estimate the proximity and height of an obstacle in front of the user who acquired that information via audio messages and vibrating alerts. The proposed system was tested by twelve VI participants, who were asked to provide suggestions for improvement. Our prototype performed well indoors and achieved overall positive feedback when detecting obstacles at different heights, although it was unable to provide directional information. Future research endeavors in this field might be benefited from more collaborative participation between end-users, researchers, and institutes for VI people.

2.3 Background of Used Models

The project mainly works on the principles of image processing amalgamated with machine learning algorithms. The project is divided into different modulus and every module is inter-woven with the next module. The object detection algorithms are as follow:

- Fast R-CNN
- Faster R-CNN
- MASK R-CNN
- Histogram Oriented Gradients
- Region Based Fully Convolutional Network (R-FCN)

- Spatial Pyramid Pooling

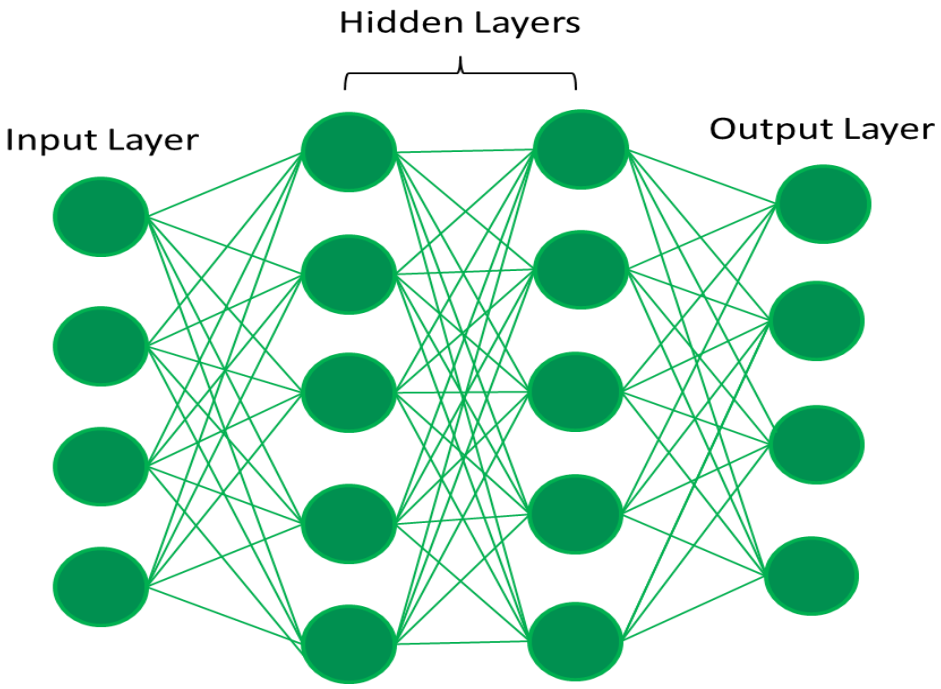

Figure2.1 Deep Learning Neural Networks [12]

2.3.1.1 Fast R-CNN

The image is only supplied to the underlying CNN once in Fast R-CNN, while the selective search is performed on the other hand as usual. The region ideas obtained by Selective Search are then projected onto the CNN feature maps. This is known as ROI Projection (Region of Interest).

2.3.1.2 Faster R-CNN

Faster RCNN is the modification of Fast R-CNN. The main difference is that Fast R-CNN generates Regions of Interest via selective search, whereas Faster RCNN employs "Region Proposal Network," or RPN. RPN accepts picture feature maps as input and outputs a set of object proposals, each with an objectless score.

2.3.1.3 Mask R-CNN

Faster R-CNN is extended by Mask R-CNN to tackle instance segmentation challenges. This is accomplished by adding a branch for anticipating an object mask alongside the existing branch for bounding box recognition. Mask R-CNN is an intuitive extension of Faster R-CNN in theory, but properly designing the mask branch is important for good results.

2.3.1.4 Histogram Oriented Gradient

Region-based Fully Convolutional Networks or R-FCN is a region-based detector for object detection. Unlike other region-based detectors that apply a costly per-region subnetwork such as Fast R-CNN or Faster R-CNN, this region-based detector is fully convolutional with almost all computation shared on the entire image.

R-FCN consists of shared, fully convolutional architectures as is the case of FCN that is known to yield a better result than the Faster R-CNN. In this algorithm, all learnable weight layers are convolutional and are designed to classify the ROIs into object categories and backgrounds.

2.3.1.5 Region Based Fully Convolutional Network (R-FCN)

Region-based Fully Convolutional Networks or R-FCN is a region-based detector for object detection. Unlike other region-based detectors that apply a costly per-region subnetwork such as Fast R-CNN or Faster R-CNN, this region-based detector is fully convolutional with almost all computation shared on the entire image.

R-FCN consists of shared, fully convolutional architectures as is the case of FCN that is known to yield a better result than the Faster R-CNN. In this algorithm, all learnable weight layers are convolutional and are designed to classify the ROIs into object categories and backgrounds.

2.3.1.6 Pyramid Spatial Pooling

Spatial Pyramid Pooling (SPP-net) is a network structure that can generate a fixed-length representation regardless of image size/scale. Pyramid pooling is said to be robust to object deformations, and SPP-net improves all CNN-based image classification methods. Using SPP-net, researchers can compute the feature maps from the entire image only once, and then pool

features in arbitrary regions (sub-images) to generate fixed-length representations for training the detectors. This method avoids repeatedly computing the convolutional features.

Chapter 3: EXPERIMETNAL METHODOLOGY

3.1 Block Diagram:

- First the mobile application performs scene capturing with the help of the mobile camera.
- The surrounding objects are then classified with the help of the object detection model used.
- The detected and classified objects output is then converted into an audio output with the help of the text to speech library used.

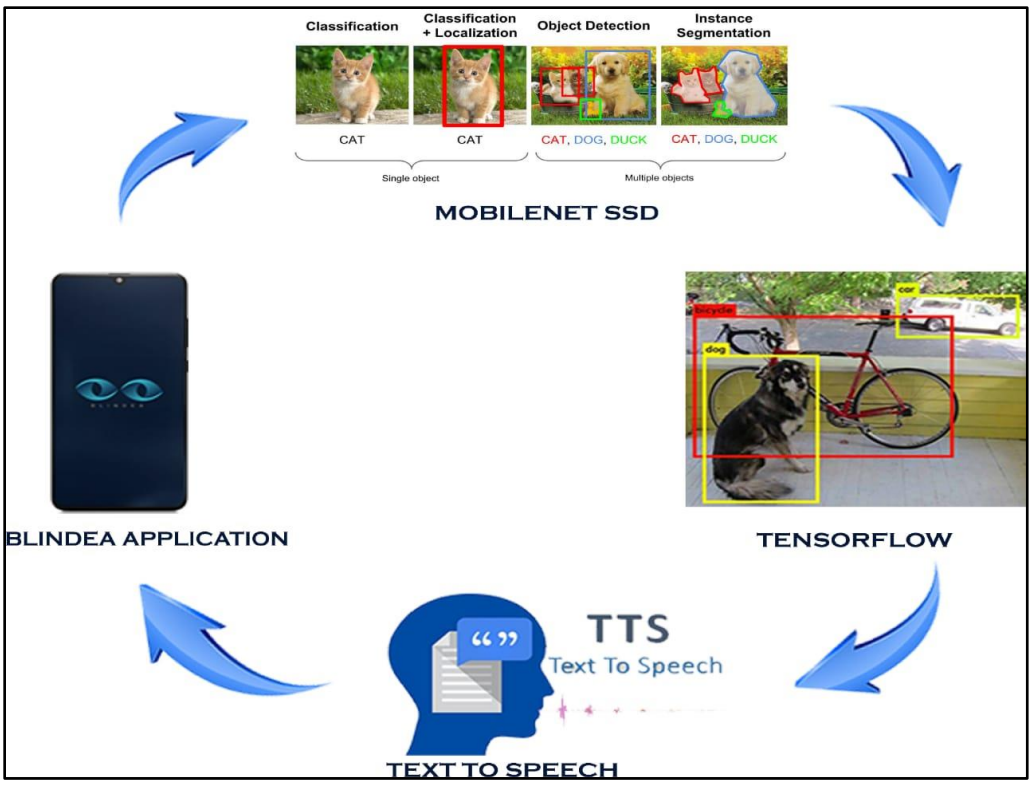

Figure 3.1 System Level Diagram [13]

3.2 COCO Dataset:

Dataset refers to a compilation of instances that share a mutual attribute. Dataset is used to sculpt a machine learning algorithm going forward. The more data is provided, the better is the efficiency.

COCO dataset helps in preparing dataset to get appropriate dataset to do the object detection.

Figure 3.2 COCO Dataset [14]

3.3 Object Detection Models

The object detection algorithms used in this project are

3.3.1 Mobile net SSD (Single Shot Detector)

The mobilenet-ssd model is a Single-Shot multibox Detection (SSD) network intended to perform object detection. This model is implemented using the Caffe* framework.

The model input is a blob that consists of a single image of 1, 3, 300, 300 in BGR order, also like the densenet-121 model. The BGR mean values need to be subtracted as follows: [127.5, 127.5, 127.5] before passing the image blob into the network. In addition, values must be divided by 0.007843.

The SSD model is made up of 2 parts namely

1. The backbone model
2. The SSD head.

The Backbone model is a typical pre-trained image classification network that works as the feature map extractor. Here, the image final image classification layers of the model are removed to give us only the extracted feature maps.

SSD head is made up of a couple of convolutional layers stacked together and it is added to the top of the backbone model. This gives us the output as the bounding boxes over the objects. These convolutional layers detect the various objects in the image.

The SSD is based on the use of convolutional networks that produce multiple bounding boxes of various fixed sizes and scores the presence of the object class instance in those boxes, followed by a non-maximum suppression step to produce the final detections

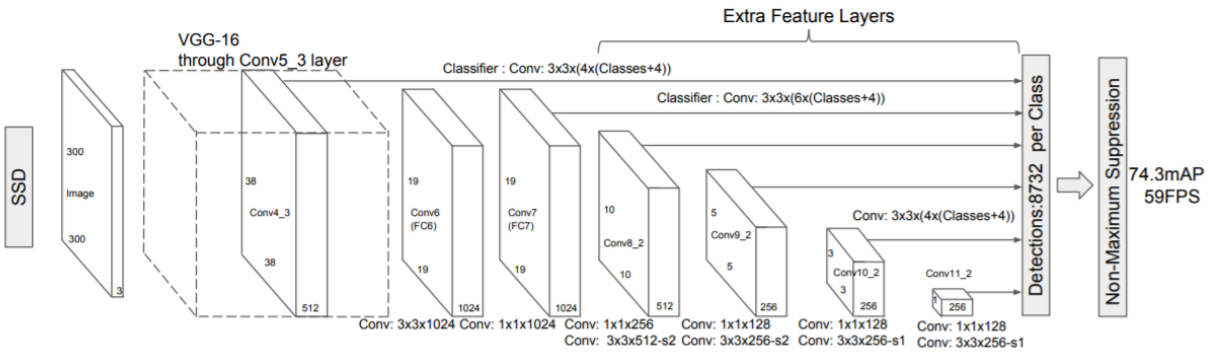

Figure 3.3 SSD Architecture [15]

The SSD model works as follows, each input image is divided into grids of various sizes and at each grid, the detection is performed for different classes and different aspect ratios. And a score is assigned to each of these grids that says how well an object matches in that particular grid. And non maximum supression is applied to get the final detection from the set of overlapping detections. This is the basic idea behind the SSD model.

Here we use different grid sizes to detect objects of different sizes, for example, look at the image given below when we want to detect the cat smaller grids are used but when we want to detect a dog the grid size is increased which makes the SSD more efficient.

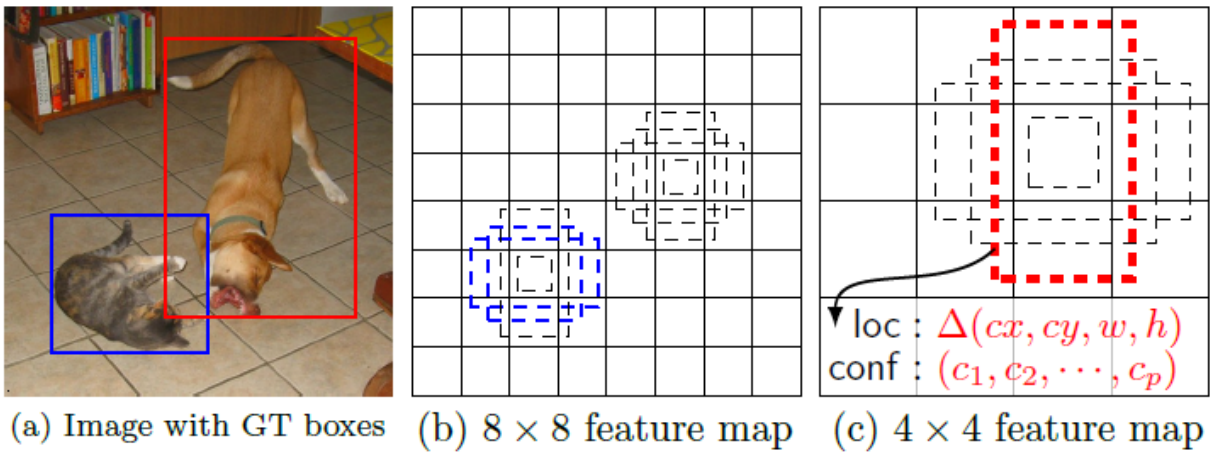

Figure 3.4 SSD Feature Map [16]

3.3.2 YOLO (You Only Look Once)

This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images.

YOLO algorithm employs convolutional neural networks (CNN) to detect objects in real-time. As the name suggests, the algorithm requires only a single forward propagation through a neural network to detect objects.

This means that prediction in the entire image is done in a single algorithm run. The CNN is used to predict various class probabilities and bounding boxes simultaneously.

The YOLO algorithm consists of various variants. Some of the common ones include tiny YOLO and YOLOv3.

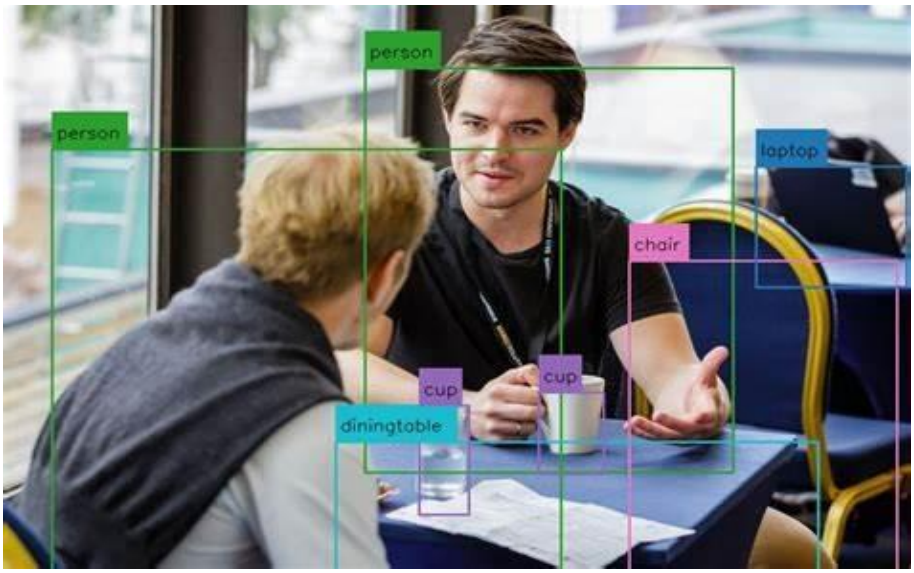

Figure 3.5 YOLO Object Detection [17]

YOLO makes use of only convolutional layers, making it a fully convolutional network (FCN). It has 75 convolutional layers, with skip connections and upsampling layers. No form of pooling is used, and a convolutional layer with stride 2 is used to downsample the feature maps. This helps in preventing loss of low-level features often attributed to pooling.

Being a FCN, YOLO is invariant to the size of the input image. However, in practice, we might want to stick to a constant input size due to various problems that only show their heads when we are implementing the algorithm.

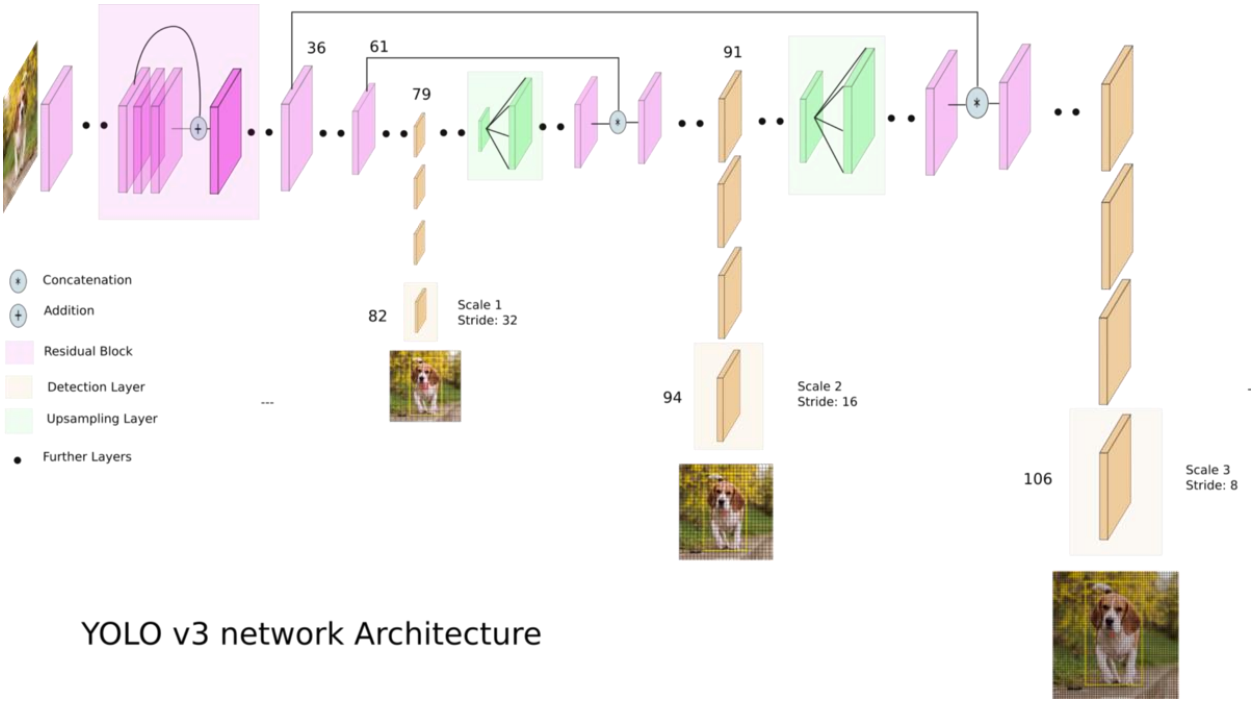

Figure 3.6 YOLO Architecture [18]

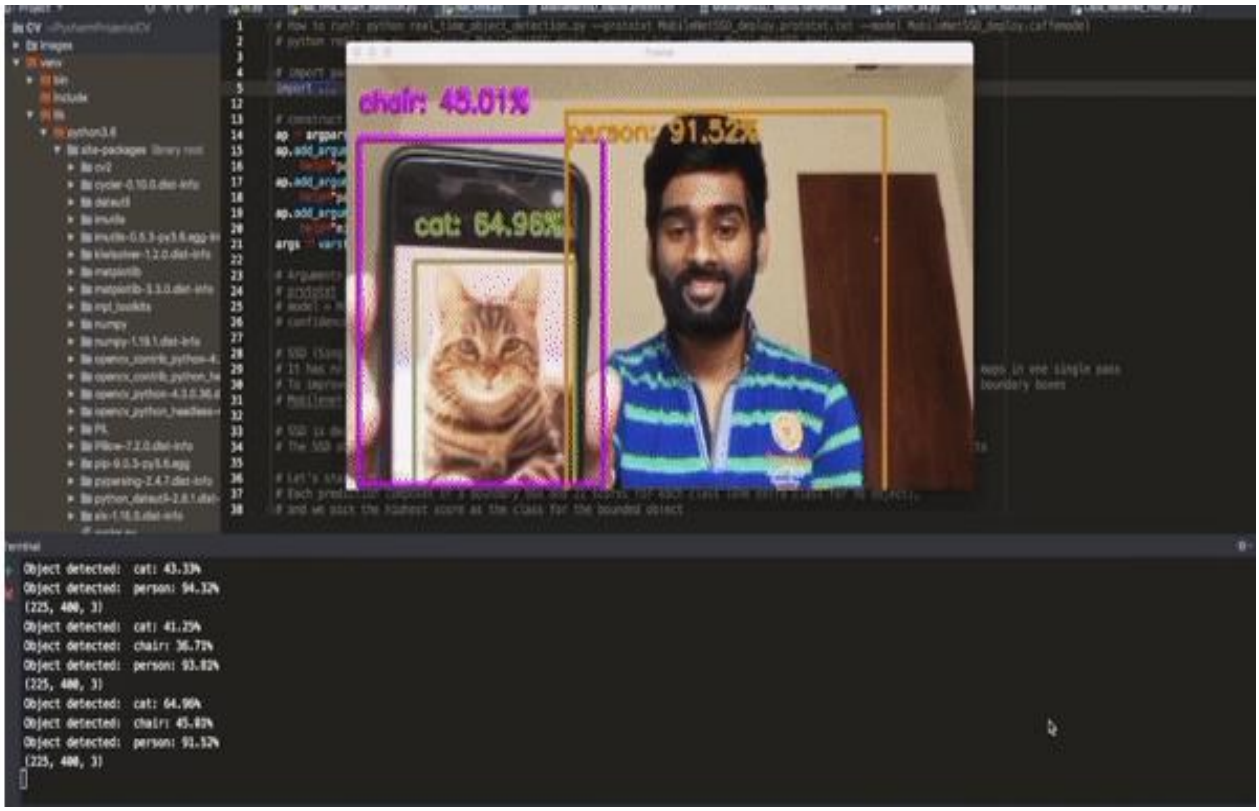

Figure 3.7 Real-time Object Detection with YOLO v3 [19]

3.4 MODEL PARAMETERS

Parameters	SSD MobileNet	YOLO
Batch/epoch	20,000	1,000
Batch size	64	64
Momentum	0.9	0.9
Weight decay	0.95	0.0005
Learning rate	0.004	0.001

Table 3.1 Model Parameters [20]

3.5 TEXT TO SPEECH LIBRARIES

- **ANDROID TEXT TO SPEECH**

The TTS engine that ships with the Android platform supports a number of languages: English, French, German, Italian and Spanish. Also, depending on which side of the Atlantic you are on, American and British accents for English are both supported.

The TTS engine needs to know which language to speak, as a word like "Paris", for example, is pronounced differently in French and English. So, the voice and dictionary are language-specific resources that need to be loaded before the engine can start to speak.

Although all Android-powered devices that support the TTS functionality ship with the engine, some devices have limited storage and may lack the language-specific resource files. If a user wants to install those resources, the TTS API enables an application to query the platform for the availability of language files and can initiate their download and installation.

The TTS engine manages a global queue of all the entries to synthesize, which are also known as "utterances". Each TextToSpeech instance can manage its own queue in order to control which utterance will interrupt the current one and which one is simply queued. Here the first speak() request would interrupt whatever was currently being synthesized: the queue is flushed and the new utterance is queued, which places it at the head of the queue. The second utterance is queued and will be played after myText1 has completed.

3.6 Building Mobile Application

For each frame, the class description of things observed will be a set of sequence images, such as "table." The image's various positions, such as top, left, right, center, and bottom, were also received, and they will be added to the class predictor "auto." Using the ATTS package, the text notion may then be delivered to the Android Text-to-Speech API.

Chapter 4: Results and Discussion

4.1 OUTPUTS

- After testing the mobile application in real time and performing real time object detection to detect objects the mobile application gives following results in which several objects have been detected like person, laptop, chair etc.
- The mobile application consists of a button with sound icon when pressed it converts the detected object output in the form of audio which can tell the user what is present in the environment.

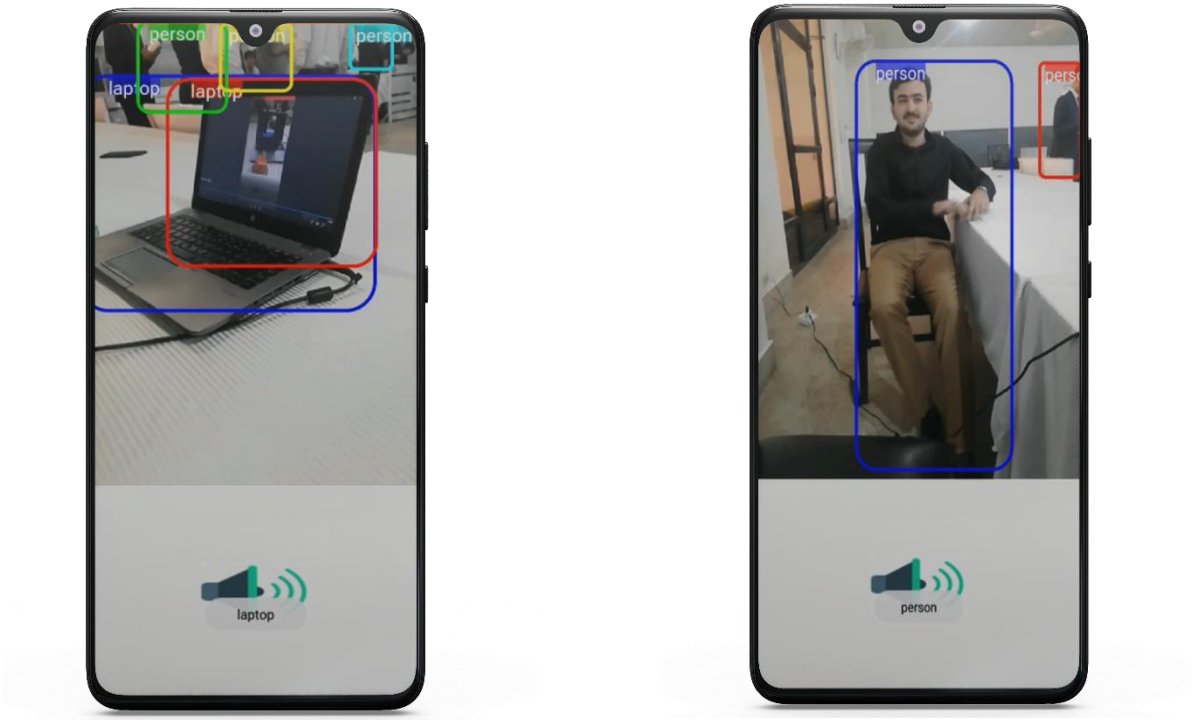

Figure 4.1a Real-time OUTPUTS

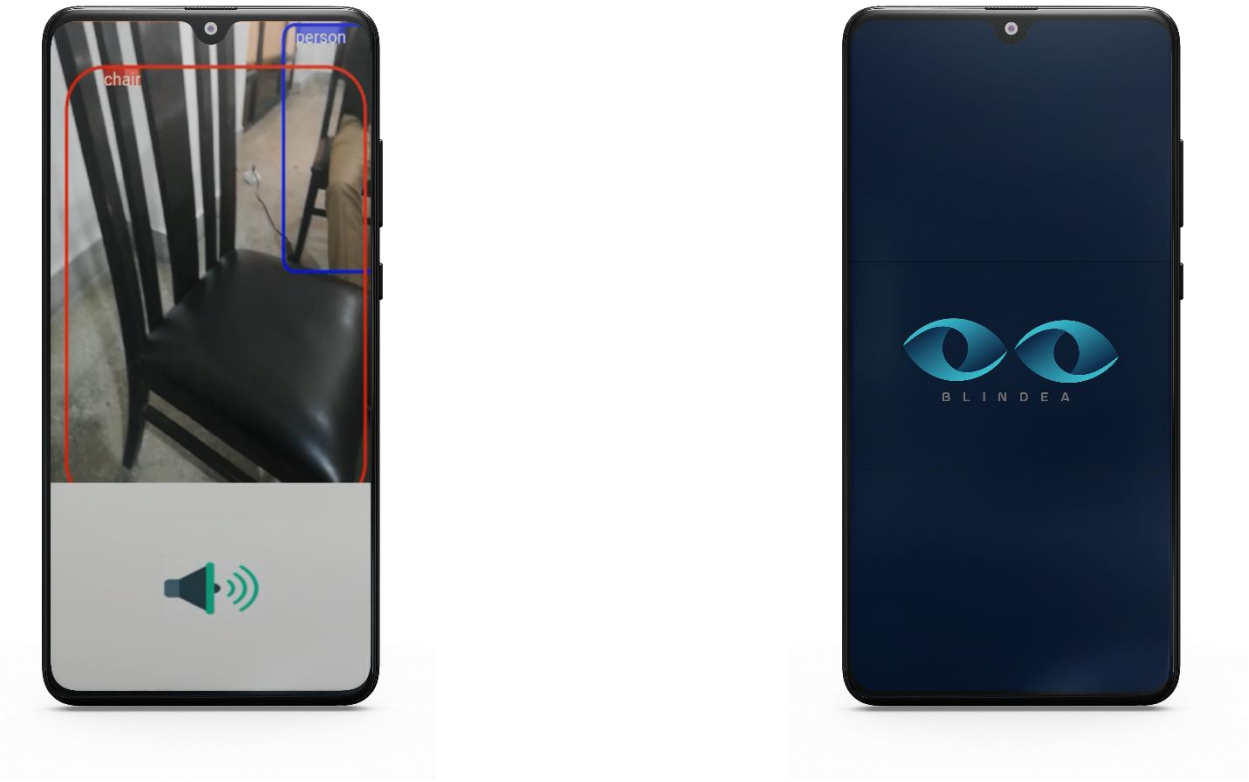

Figure 4.1b Real-time OUTPUTS

4.2 TESTING ACCURACY

- **SSD**

  The trained network parameters were adopted for the MobileNet-SSD defect detection network. The test set image contained four different types of defect samples, each of which had 30 images obtained through resampling. Each sample involved one or more defects. The detection results of the trained MobileNet-SSD defect detection network on the four kinds of defect samples

Defect Type	Sample Number	Successful Detection Number	Leakage Number	Error Detection Number	Positive Rate (%)
Breach	30	30	0	0	100.00
Dent	30	27	2	1	90.00
Burr	30	28	1	1	93.33
Abrasion	30	39	1	0	96.67
Total	120	115	4	2	95.00

Table 4.1 Detection Result Of Trained Mobilenet SSD Model

It can be seen from the above table that the surface defect detection network completes the defect marking of 120 defect samples with a 95.00% accuracy rate. There were missing and false samples in dent and burr defects and missing samples in abrasion defects. This is because the notches are more obvious than the other defects, and related to the image quality and subjective feelings of humans.

Number	Model	Positive Rate	Training Time (Day)	Detection Time (s)
1	SqueezeNet	85.83%	2	0.31
2	Faceness-Net	90.83%	3	0.59
3	MobileNet	90.83%	<1	0.54
4	MTCNN	91.67%	3	0.64
5	PVANet	94.17%	2	0.25
6	MobileNet-SSD	95.00%	<1	0.12

Table 4.2 Correct detection rate, training time and the detection time per image of each network

As shown in the table, the MobileNet-SSD surface defect model is fast and stable, thanks to the improved SSD meta-structure of the feature pyramid. In general, the proposed algorithm outperformed the contrastive algorithms in detection rate, training time and detection time. The final detection time of our algorithm was merely 120 milliseconds per piece, which meets the real-time requirements of the industrial filling line.

Chapter 5: Conclusion and Future Work

In this thesis, we discussed a smart system that can provide assistance to visually disabled people smartly and more efficiently than the typically used available systems. Our proposed system has an advantage over other traditional systems due to the latest algorithms used for the detection of objects of interest. Techniques used in our proposed system; Image Processing techniques to process the video, frame by frame, and Machine Learning included algorithms such as YOLO and SSD, which were used to detect the objects of interest; are also briefly explained included their working and importance. The purpose of increasing productivity and overcoming problems in existing solutions is being achieved by using the modern techniques. Additionally, the objectives; smooth traffic and providing special privileges, are attained. Hence saves time as well as the prestigious lives of patients.

Let's conclude the whole process of developing the "Google Glasses for Visually Impaired People" in the following steps:

5.1 Training the Data

The dataset for training is taken from the COCO i.e., Context Common Objects. And for the purpose of programming, You Only Live Once (YOLO) v3 is used.

5.2 Model

Convolution Neural Architecture is implemented through the YOLO algorithm. And to setup the module, python language is implemented on PyCharm.

5.3 API – Application Programming Interface

For each frame, the class description of things observed will be a set of sequence images, such as "table." The image's various positions, such as top, left, right, center, and bottom, were also received, and they will be added to the class predictor "auto." Using the ATTS package, the text notion may then be delivered to the Google Text-to-Speech API.

5.4 Output

For the further assistance of visually impaired people, audio output is preferred so that they can heard the objects around them to get the better idea of the surrounding environment.

Our proposed system, Google Glasses for Blind People, is cost-effective as it is purely made for the core purpose of serving Pakistan, any system that could be beneficial to its citizens. Else, similar solutions provided by other countries are very costly. Moreover, it provides an ease to adoption which can be adopted by beneficiaries, mass deployment can also be done and most importantly no product training is required.

5.5 FUTURE WORK

The visually impaired people and even people with major eyesight issues are increasing with each passing day. Therefore, the need of smart equipment and devices is there to assist them.

Future milestones that need to be achieved to commercialize this project are the following:

- Using HD-cameras to record real-time footage.
- Linking the recorded video stream with the manufacturers to watch in order to provide the accurate description of the surrounding environment.
- Incorporating the knowledge of distance between the user and the obstacle.

5.5.1. Giving it a Hardware Form

The next step is to develop glasses that can actually be wearable by the people who need it just like the Google Glasses made by the Google. It is luxury for most of the people but a need for visually impaired people.

5.5.2. Improved Quality

The quality of the product/ service can be improved by using HD cameras to get real-time data for processing. This will help in achieving the correct result even when the person is in moving state. Companies, like eSight, are currently working on it.

Figure 5.1 eSight-Smart Glasses for the Blind [21]

5.5.3. Improved Accuracy

Connecting the smart glasses with live customer support is a great idea. The recorded video can be sent to the technical agents, and they can guide the user about his environment accurately and timely. Aira is currently working on this idea.

Figure 5.2 Smart Glasses by Aira [22]

References

1. Smart guiding glasses for visually impaired people in indoor environment https://ieeexplore.ieee.org/document/8103374

2. MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients https://ieeexplore.ieee.org/document/8962044

3. Intelligent Smart Glass for Visually Impaired Using Deep Learning Machine Vision Techniques and Robot Operating System (ROS) https://link.springer.com/chapter/10.1007/978-3-319-78452-6_10

4. LidSonic for Visually Impaired: Green Machine Learning-Based Assistive Smart Glasses with Smart App and Arduino. https://www.mdpi.com/2079-9292/11/7/1076

5. Lan, F., Zhai, G., & Lin, W. (2015, November 1). Lightweight smart glass system with audio aid for visually impaired people. IEEE Xplore. https://doi.org/10.1109/TENCON.2015.7372720

6. (PDF) Smart Glasses for the Visually Impaired People. (n.d.). Www.researchgate.net. https://doi.org/10.1007/978-3-319-41267-2_82

7. Chen, L., Su, J.-P., Chen, M., Chang, W.-J., Yang, C., & Sie, C.-Y. (2019). An Implementation of an Intelligent Assistance System for Visually Impaired/Blind People. 2019 IEEE International Conference on Consumer Electronics (ICCE). https://doi.org/10.1109/ICCE.2019.8661943

8. Suresh, A., Arora, C., Laha, D., Gaba, D., & Bhambri, S. (2018). Intelligent Smart Glass for Visually Impaired Using Deep Learning Machine Vision Techniques and Robot Operating System (ROS). Robot Intelligence Technology and Applications 5, 99–112. https://doi.org/10.1007/978-3-319-78452-6_10
9. Busaeed, S., Mehmood, R., Katib, I., & Corchado, J. M. (2022). LidSonic for Visually Impaired: Green Machine Learning-Based Assistive Smart Glasses with Smart App and Arduino. Electronics, 11(7), 1076. https://doi.org/10.3390/electronics11071076
10. Bai, J., Lian, S., Liu, Z., Wang, K., & Liu, D. (2017). Smart guiding glasses for visually impaired people in indoor environment. IEEE Transactions on Consumer Electronics, 63(3), 258–266. https://doi.org/10.1109/tce.2017.014980
11. Karthikayan, P. N., & Pushpakumar, R. (2021, November 1). Smart Glasses for Visually Impaired Using Image Processing Techniques. IEEE Xplore. https://doi.org/10.1109/I-SMAC52330.2021.9640715
12. Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232. https://doi.org/10.1109/tnnls.2018.2876865
13. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Cv-Foundation.org, 779–788. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html
14. GlobalData Healthcare. (2019, March 25). Future of smart glasses: how new technology help the visually impaired. Verdict Medical Devices. https://www.medicaldevice-network.com/comment/future-of-smart-glasses/

15. https://www.ijert.org/comparison-of-yolov3-and-ssd-algorithms
16. https://www.google.com/search?q=yolov3+architecture
17. https://www.google.com/search?q=SSD+Architecture&sxsrf=ALiCzsZArTFl4SuzeP46o-lTSu2bAousIQ:1655291589053&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiW2qjZqa_4AhU8SvEDHcjrBAsQ_AUoAXoECAEQAw&biw=1366&bih=649&dpr=1#imgrc=-YkUFkm5Gom_rM
18. https://doi.org/10.1109/TENCON.2015.7372720
19. https://ieeexplore.ieee.org/document/8103374
20. https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b
21. https://doi.org/10.1109/tce.2017.014980
22. https://lowvisionmd.org/aira/#:~:text=Aira%20is%20monthly%20subscription%20service,whatever%20situation%20they're%20in.