May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

Speech Guidance using Real-Time Object Detection

Hardik Pathar[1], Naveen Kumar RV[2], Neha Yadav[3], NayazAhamed I S[4], Shilpa V[5]

[5]Assistant Professor, [1,2,3,4,5]Dept. Computer science, REVA University, Bangalore, India
[1]hardikpathar158@gmail.com, [2]naveenrv247@gmail.com, [3]nehayadavofficial775@gmail.com, [4]ahamednayaz49@gmail.com, [5]shilpa.v@reva.edu.in

Article Info
Volume 83
Page Number: 4531-4541
Publication Issue:
May - June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

There are 40 million visually impaired people in India, they find it very difficult to perform basic day to day commuting, as they are unable to read traffic signs and detect nearby objects, also they can't find the exact position of the vehicles and they often rely on other pedestrians to guide them to their destination. This project aims to provide a method to solve this issue through an application that contains an image recognition system that detects nearby boards and signs in surroundings. The camera is used to capture the image and the captured image is analyzed (object detection) and converted into text and then text to speech model converts text into audio format. This enables them to know their position and helps them to decide where to go without asking anyone else. It not only makes their travel easier but also saves a lot of time and enhances their safety on roads. It makes them independent.

Keywords: *visually impaired, TTS, Real-time objects, object detection, text to speech conversion.*

1. Introduction

As per recent surveys, 285 million people are visually impaired in the world among which 40 million are Indian. They find it extremely difficult to read and detect nearby objects.

In [1] Visual impairment is a major loss at a very high level some cases may emerge from diseases also major shocks and hereditary means which result in one possible problem for eyesight that cant me cured by medical approach at some conditions. In India we have broad definition of visual impaired people with disabilities and illiterate people who comes under same category. Some eye sight problems such as No vision or Zero Vision, It can also be referred as the level of Visual Acuity which is less than 6/60 also 20/200 in a normal eye having correct lenses and also due to an angle in a vision sight which happens to be 20 degree.

Vision is very important not only to view objects but also in dark conditions, contrast sensitivity, balance and color perception levels. In spite of losing all these required functions in eye, the impaired people rely on their senses to carryout day to day activities as well as employment to run the daily life smoothly.

In [2,40] The challenges faced by blind are very crucial and in current society people are not very well understanding towards blind .I would begin with the amount of difficulty when catching a right bus in right bus stop to knowing something displayed for public eye and some of the very common areas where they suffer are mentioned here. Roaming here and there - A noticeable heap of challenges for blind people, consider a person having the full and extreme condition, that is a loss of eyesight, which becomes a thorn in going from one place to a different one.

Let us say, impaired can go and look for objects easily in their homes or places they are familiar with, because its known to them from the experience the life has given to them. So for the persons or gentlemen who decides to help blind people must not indulge in their daily life activities and should refrain from doing their common chores in which they are best at.

If they think of going outside they will face difficulty in identifying the place and may result in accidents and

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

other unthinkable situations. Places like Commercial areas which are built with easy and known textile material to blind that is tactile tile and in this way its more helpful than normal tile. But, If I say honestly it is not followed in all the places ,I say most places do not have them, now it has become a great threat to impaired peoples who may go to these places and end up in trouble. [3, 41].

Second area is getting Literature works

I would say that blind peoples face difficulty in getting the right literary works which are in different format entirely. In India there are millions and millions more people are suffering because they are blind but there is no proper Braille works to read and daily novels are beyond their grasps. Internet, A giant network of huge information available to everyone regardless of any defects yet blind people do not have its access. A sufficient way for blind people to read on screen material is via any software but it would pose problem in deciphering texts from inadequately designed websites. In addition to this every content on the webpage must have a description to it, if not its complete waste of time for anyone to look at it.

Overly helpful Individuals may pose a problem in this manner,

In today's society there exist some people who stretch their helping hands to blinds also which results in this problem. let us say if a kind gentlemen is helping a blind not knowing the exact help required then its a waste of time from both sides. Another problem is if a blind individual is doing a work at his/her speed and the helper might do the same work at a high speed which results in problem of not getting the work done the way they wanted.

As said the individual who are approached by a blind person should respect and provide with utmost assistance possible, but unfortunately in recent years people have lost their humanity and do not help blind and ignore them.

Next problem as of our understanding is getting the independence in life even for a blind person.

A great priced accusation in a life of blind is to have an truly peaceful independence in his state. He/she may enjoy their life which has things that are accustomed to his/her needs and follow it for their life. There are very much availability of gadgets which are expensive yet not available to people and many are unknown to society.

A lot of different techniques have been used to solve this problem like Braille for visually impaired but a huge disadvantage of this is that a person has to first learn how to use a Braille that requires a lot of effort. Moreover, in the current world of digitalization, the text in the digital form cannot be read using Braille scripts. To solve the above-mentioned problem systems that scan the text written in books and papers and then convert into speech were developed.

All of the above-mentioned solutions neglected one very important aspect of people's lives and that is, travelling. Visually impaired find it very difficult to perform basic day to day commuting, as they are unable to read traffic warning signs and detect nearby vehicles, also they can't read the informatory signs to know their exact position and they often rely on other pedestrians to guide them to their destination.often pedestrians make fun of them and making them feel guilty of not able to see. Taking into consideration of real time issues and how to face them, we propose a system with easy to perceive model which asks the person using it to command for things and provide the person with the appropriate results.

In this project, we provide a method to solve this issue through an application that contains an image recognition system that detects nearby objects and individual objects based on users request in surroundings. The camera is used to capture the image from the nearby objects and the captured image is processed and the user will get to know object around him through speaker or any audio device.

By doing so, some problems faced by the impaired will be solved and impaired people will have faith in current technologies which can become a guide to them.

A. Organization of Paper

This paper follows this format. It starts with Introduction as part 1, System overview as part 2, Literature survey as part 3, Methodology as part 4, working as part 5 and Results and Future work as part 6 and 7.The References are in part 8.

2. System Overview

It involves the basics required in order to complete the task at hand and also to gain better understanding of the concepts.

A. What is Object detection?

Object detection a known area from computer vision technology which is being improved at massive scale. Regards to Deep Leaning technology![33] We see that there are lot of new algorithms that outperform its predecessor's and get good results every passing year[24].If we want to define Object detection I would put it in this easy way as "An approach to detect and identify the objects by class names", it can have some localization's in every turn of detection of object and proceeding to segmentation at later stage [25].

B. Super wise Learning

The various algorithms generate some set of functions which may map the given input to specific output. One known problem is a classification task, it requires the learner to learn mapping function which maps each vector to other corresponding classes by seeing the previous input and outputs provided [26].

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

"Supervised learning" lets take a quick overview of supervised learning method, It contains SVM's(support vector machines) and classifiers which are most common in research and development [27].

If we understand the clear definition as "its a machine learning task where a mapper function maps a given input to the specific output by the previous input and output pairs. Which is a labeled data or training examples, which fuels the supervised learning in getting right results[28]

C. Datasets

Datasets are one of the main building blocks of the Learning algorithms which acts as a starting place for researchers to work on and gather the more information and make it more useful.

- PASCAL VOC, a known dataset to all researchers.

The PASCAL VOC has very highly standardized images which are used daily and a tool set for accessing these images based on its labels and process each dataset[29] .It has optimized for 20 classes and works smoothly for the trained data. its challenge [30] is not available from 2012, but the quality of this is unrivaled and well maintained. The important fact of this dataset is its small size compared to the other various available datasets and its more suited in testing of network application programs.

D. Network of Neurons

The NN is connection of neurons, an AI based network which has multiple human made neurons that are named as nodes. A NN(neural network) is not a biological one but looks similar in case of structure and working and more suited in solving real time problems involving artificial intelligence [31].As the idea of project is concerned, we discuss more about the Convolutional Neural Network.

CNN (convolutional neural network),

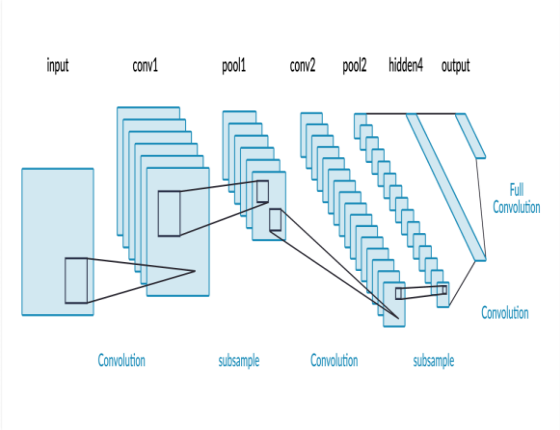

Figure 1: Typical representation of CNN,abbreviated as convolutional neural network in technological perspective.

It is a five layer network system, each singular layer includes a majority out of two dimensional plane, and each individual plane is made by a majority out of individual neurons [32]. In which 2 are convolution layers and 2 are sub-testing layers and 1 completely associated MLP layer [34,35].

E. YOLO Loss function Description

If we take the single grid then, it shows precognition, that is the bounding boxes required and loss calculation for each required box for the object given as input. The process of finalizing each BB (bounding box) depends upon the max output value of IoU obtained from given input object.

Following shows the list of loss functions used in YOLO for calculating the loss,

1. Classification loss
2. Localization loss
3. Confidentiality loss

Here, the error between the base value and for casted value is referred as loss in localization. The presence of object on the input acts as a confidence of object and if not present then it is referred as loss in confidentiality and if there is square in obtained results of class then it is referred as classification loss. The equation is shown below,

$$\sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Above Equation Showing class probability for each class.

It has $\mathbb{1}_i^{\text{obj}}$ results 1 if there is an object appearance else the value would be zero.

$\hat{p}_i(c)$ it is showing the conditional class probability of individual class c.the measure and amount of errors from BB are localized.

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

In above Equation, loss from localization is shown. Where,

$\mathbb{1}_{ij}^{obj}$ results 1 if the j-th BB from cell 1 used or provides a zero as result.

λ_{coord}

It is responsible when there is an increase in the weights due to the loss obtained from BB. If there is no object

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

found inside the BB then we can relate it to loss in confidentiality of an object, it can be given as,

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2$$

In above Equation , Loss in confidentiality is shown.
Where, $\hat{C}_i \quad \mathbb{1}_{ij}^{obj}$
is the accuracy level of BB j in the cell i.
If the detected object is responsible for j-th BB then the result obtained is 1 else it would be zero.
If we find that there is no object at all then we can calculate the error as shown in following equation,

$$\lambda_{\text{noobj}} \sum_{i=0}^{S^2}\sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2$$

In above mentioned Equation, showing the confidence loss if our object is not found.
Where,
It is the tribute of $\mathbb{1}_{ij}^{noobj} \quad \mathbb{1}_{ij}^{obj}$
$\hat{C}_i$
the accuracy level of BB j in a cell i
λ_{noobj}
Loss of weights while detecting the object in background level.

3. Literature Survey

As referred from [4] , you only look once(YOLO), The brand new approach in field of object detection is presented, the specific object is detected by regression problem to spacial separated rectangular bounding boxes with a class label. It is very fast, it processes at 45 fps in high end computers which is a breakthrough. It detects the objects as follows, In this stage input images gets re-sized into 448 x 448 resolution and runs a convolutional network and thresholds resulting detection's by the models level of confidence.

Some handicap functionalities which were known after the models output are depicted here. It imposes spacial constraints on rectangular bounding box predictions. It faces difficulty with small objects that are visible in background such as small rocks, background wall group of birds in air among lot of trees.
It also faces difficulty in detecting new shaped object which is tested for first time as depicted in [5].

The loss function that approximates detection performance as well as error rate while dealing with small and large bounding box. It has more affinity towards localization error.

As per mentioned paper [4], In comparison with classifier proceed towards methods, Yolo directly corrects error from performance and increases accuracy .

In paper [6], to access the complete understanding of image we need to classify different images and to detect accurate location of objects in an image. It has various tasks like face detection, fundamental object detection and get valuable information from human behaviour and corresponding images

A way in object detection is by doing it in generic way [6].It aims to locate objects from the input image and provide a label for it within a rectangular box.

Following is another approach in object detection that is Salient detection, In this visual saliency task which is a prominent task in computerization, which aims at highlighting the most dominant regions of images.

In detection of faces preprocessing procedure is involved, it locates face region points which in vary large range of scales for different face images. Each face has unique object structural configurations and characteristics which leads in special attention.

We gather the facts on Object recognition from generic pipelines, that are used as base architecture in identification of objects. It gives particulars on expert systems which process the idea of object detection frameworks which are useful in solving problems, as clutter, low resolutions and occlusions in given object.

In paper [7], The Object identification includes category detection and also category recognition. Let us see about category detection, it deals with differentiating object with its background properties and category recognition does the classifying work, hence by doing so the process of specific object detection in digital video or image is specified. The process relies on the following sub processes like matching with existing data, pattern recognition has several algorithms which uses feature based techniques in detecting the object. As per above mentioned paper it has become very easy in implementing object detection model which is trained on a specific data set. The detection process includes the subtraction of the background layers of the image which is being detected and the final layer depicts the object.

In paper [8], this aims at detection objects using YOLO approach, it is a different approach compared to typical old convolutional network, a fast convolutional neural network predicts the BB using CNN and the probabilities for each class with high detection rate.
This method outperforms different existing strategies from natural pictures to different domains. This approach has an upper hand comparing to other classifier algorithms.

In paper [10], The computer vision field is mainly used in Expert learning tasks in real time object disclosure, we can see the use of PASCAL VOC, COCO and Image-Net data-sets which acts as a base for a deep learning model and neural networks such as Regions with convolutional neural network, Spatial Pyramid Pooling, Fast regions with CNN, Faster regions with CNN(convolutional neural network) which performs

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

prepossessing of images and provide object recognized with a class name. We got to know the significance of expert learning or AI learning based networks.

This [10] signifies a huge significance of machine learning concept, that is deep learning and its uses in current time and day. With the these technologies we are able to solve some major issues in real life and get an understanding of machine learning concepts.

It has become an effective implement in learning from the big chunks of data and achieve the success from day to day problems. Large amount of data is the epitome of the success of AI learning.

In paper [11], Analysis of real time images and videos are under taken as object detection as its base.
Deep learning is a subset of machine learning, and AI. It involves following set of algorithms and solutions, Deep Neural Network, Convolutional Neural Network, Time Delay Neural Network, Computer Vision etc to analyze and detect real time objects We can see convolutional neural network which is a multi-layer neural system in which majority are two dimensional planes and each plane is made of individual neurons. OpenCV, a open source computer vision and machine programming instructions. It is used in livening the utilization of AI, machine applications in real time business environments.

It composes of 2500 figurines, which aims at recognizing faces, understand figures, track camera activities, track real time moving objects and make a high assured object. This library is extensively used in associations, street view pictures, understanding interfaces from a real time video input and helping robots in analyzing objects. We saw the use of neural network algorithms and purpose of deep learning at a major scale.

In paper [12], the image recognition technology with speech synthesis which are cost effective in nature, user friendly and has a conversation system which are built on top of Raspberry Pie. It includes a camera and a lens which scans for input text on any medium and converts them into audio format using text to speech engine.

Optical character Recognition, its character recognition module which scans images and out puts the characters via audio stream. The scanned images from OCR undergo a prepossessing stage and gets segmented as in [13]. After segmentation process characters from the original document are split into individual words and sent to text to speech converter.

Image processing,normally characters and words are on written medium and will be on either paper or any software document, the objective is to detect those words and detect the directional axis by using feature map technique

Optical character recognition, These system are used in manipulation of scanned text or sum characters from any input medium and output the characters in a textual format. Main purpose of using OCR is in doing pattern recognition and artificial intelligence and computer vision. TTS(text-to-speech), A system capable of reading out the texts and its level of volume is adjusted by the person using as depicted in [14] , if the text is in the form of input stream from the computer or for the paper. If we understand its working principle then it might give an idea about these systems. It operates on the corpus speech analysis principle, which gives high quality natural speech output, in above mentioned paper [13], Blinds are not in a place to feel disappointment from these systems. The image preprocessing part allows for complete extraction of text from given input stream.

4. Methodology

The methodology consists of start and end of the stepwise process performed during the model in detection of objects in real time environment.

It consists of Start, which signifies the start of flow diagram, Frame input from camera will have either a image or a real time video in the frame from inbuilt webcam, Yolo and Dark net framework performs the segmentation, preprocessing and most of the recognition work and the object which is recognized is spoken out by the assistant and its class name with the level of accuracy of the object is displayed on the screen. Finally stop specifies the end of process.

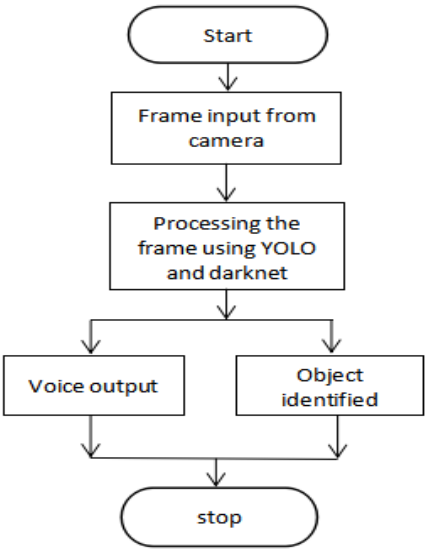

Figure 2: Flow diagram of the model

The above figure is explained in following steps
1. The application opens the camera of the device and captures the nearby images.
2. The inputs from camera are fed in the form of frames to the application.
3. The frames are processed using YOLO and darknet
 Image is first divided into grids. Each cell predicts bounding boxes and calculates probability. The boxes below a certain threshold probability are deleted.

The remaining boxes will form boundary of detected objects by eliminating duplicates. The most exact images

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

are detected. Darknet framework is used to train neural network and serves as a base for YOLO. Predefined classes of datasets are used for training YOLO using darknet.
4. Once the object is identified, the label on the object is converted into voice.
5. The application successfully detects objects and converts their names into voice.

5. Implementation

The detailed working of the model is explained in the architectural design diagram of the model.

Architectural Design

The below design diagram consists of data input where images and videos are given as input and in object identification, feature extraction and object generation is given as input and in training process we provide feature extraction and object generation and its results to train and its processed through neural network and indexed in order and passed to the output stage.

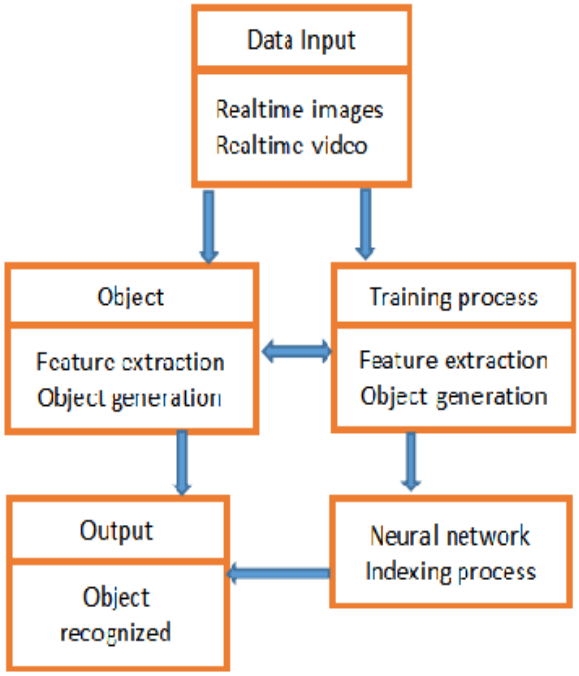

Figure 3: High Level Diagram

1. Data Input

It is the main component of the model where the images and videos are taken via webcam in real-time and it's forwarded to object identification and training process.

2. Object identification

It consists of Feature extraction and object generation.

- **Feature extraction**

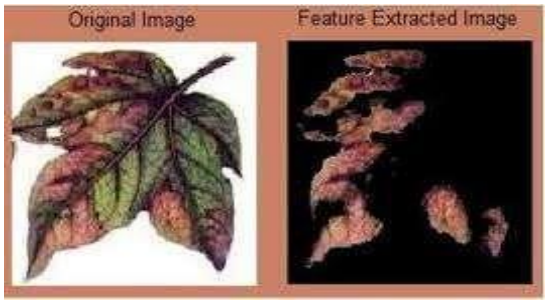

Figure 4:Example image showing feature extraction [44].

Main objective is to take only important features and eliminate the rest. It contains the numerous convolutional layers and each layer has different channels. The channels present in principal convolutional layers identify the straightforward features. Max pooling performs the operation of dividing information picture into non covering rectangles and choosing most extreme point inside every cell [36].

- **Object formation**

It is the process of training the Convolutional neural network with dataset and use the parameters obtained from training to further processing.

3. Neural network and Indexing

Convolutional neural network is a five layer neural system singular layer is made by a majority out of two dimensional plane, and each individual plane is made by a majority out of individual neurons [37].In which 2 are convolution layers and 2 are sub-testing layers and 1 completely associated MLP layer [38,39].
If we see the layers of CNN, that is Convolutional layer as first, Pooling layer as second and a completely connected layer as third[42, 43]. The process of feature extraction is made possible in the initial and foremost layers of neural network and if we focus on the prediction and probability then they are made possible by completely connected layers [44, 45]. at the end it composes of 24 primary layers and 2 completely connected layers as from start to end.
Indexing refers in ordering of the processed images with the id value and its class label is assigned in parallel [46].

4. Output

The object detected is conveyed to the person via audio and the level of confidence and degree at which the object is present is displayed on the screen as logs.

A. Data Pre Processing Stage

For processing, we should normalize the images and the object will get a bounding box around it with a label

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

name. The labeled images are trained to YOLO V2 model and gets recognized.

Figure 5: Example image showing the data preprocessing [43].

As the process is continued the mean of images and its result is being used in SD(standard deviation) to get the processed data from the raw input image.

B. YOLO Processing

It's fully based on YOLO V2 deep learning model. This model uses pre-trained weights and load them to Open CV model.

A text -to-speech engine will speak out the name of object detected.

YOLO: A realtime object detection module

In a general view, the aim of YOLO is to look at an image and distinguish the object it scanned and provide a recognition to it via several steps.

YOLO is abbreviated as "You only look once". Its fast and accurate framework which is perfect option for real-time object detection. The architecture of YOLO is simple, it has a 3x3 pooling and 2x2 max pooling which are based on kernel pooling.

Layer	kernel	stride	output shape
Input			(416, 416, 3)
Convolution	3×3	1	(416, 416, 16)
MaxPooling	2×2	2	(208, 208, 16)
Convolution	3×3	1	(208, 208, 32)
MaxPooling	2×2	2	(104, 104, 32)
Convolution	3×3	1	(104, 104, 64)
MaxPooling	2×2	2	(52, 52, 64)
Convolution	3×3	1	(52, 52, 128)
MaxPooling	2×2	2	(26, 26, 128)
Convolution	3×3	1	(26, 26, 256)
MaxPooling	2×2	2	(13, 13, 256)
Convolution	3×3	1	(13, 13, 512)
MaxPooling	2×2	1	(13, 13, 512)
Convolution	3×3	1	(13, 13, 1024)
Convolution	3×3	1	(13, 13, 1024)
Convolution	1×1	1	(13, 13, 125)

Figure 6:above image showing the standard layers present in YOLO

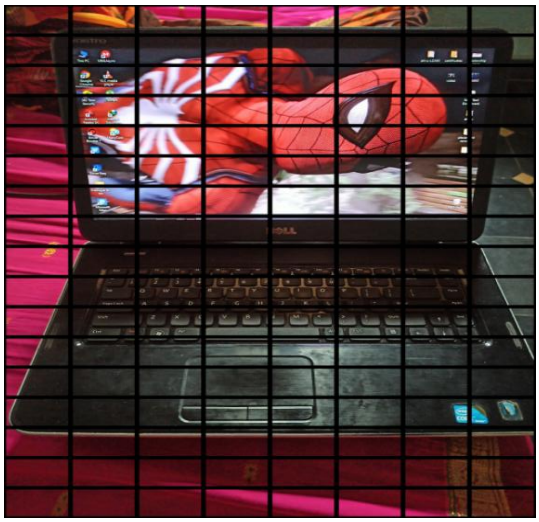

Figure 7: Example Image divided into cells

1. Grid formed in first stage.
2. Each cell formed is of same size.

In above figure, YOLO scans the image just one time and the cells of same size is being put that is the image being split into several cells and it plays a significant role getting the BB's. Each of these cells play a very important role in providing the required confidence in detecting the object. The BB's are figures with rectangular shape with the detected object class on top of the respective bounding boxes.

In addition to the above mentioned cells, It also provide a accuracy level known as confidence level of object within the BB. The higher the score of an object inside the BB higher the confidence. The below image depicts the BB's drawn and scaling the for accuracy.

Figure 8: Example Image showing Predicted intermediate bounding boxes

1. Probable bounding boxes around the object.
2. More the bounding boxes around will yield in high confidence level.

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

The bounding boxes predicted by the grid cells,If we consider for each individual BB, then it has class label and its pretty much alike to the classifier. It also gives distribution of probability over each class it has obtained.

It is the standard version available to researchers, developers in doing the research in machine learning and neural networking areas, the main advantage is its size and its optimization which is far greater with less available data.

In the end of object labeling step the level of accuracy obtained and the predicted results before hand are compared and to be precise the object is identified on the above obtained result from the confidence level from each BB. For example, yellow box is sure that it contains the object "laptop". These classified objects are stored in a data file. This data file is fed to the python text to speech module which communicate these objects to the users.

Figure9: Example Final bounding box with the class label.

1. A yellow line enclosing the object with is class label at top left.
2. The object is accurate and probable bounding boxes are merged to one final bounding box.

C. Text to Speech Model

Text -to-speech model is trained model that synthesises the text into speech that is given as input [48]. This model is used to convert the detected objects and characters into speech [49].

D. Speech Recognition

Speech-Recognition, As it is mentioned it is just an application that lets you take the input from the user by a mouse click or by any processing step leading to speech recognition as input. With help of this the user can convert his given audio input into set of instructions to the machine to process further or to make it as document [50].

6. Results

Model is tested against different inputs for the correct results.
Following steps are considered to get the desire result
1. Run model
2. Model introduction.
3. The user can find individual object as well as multiple objects with the command.
4. Give input to the model via microphone.
5. Model recognizes object and gives the confidence level and time taken, approximate frame rate.

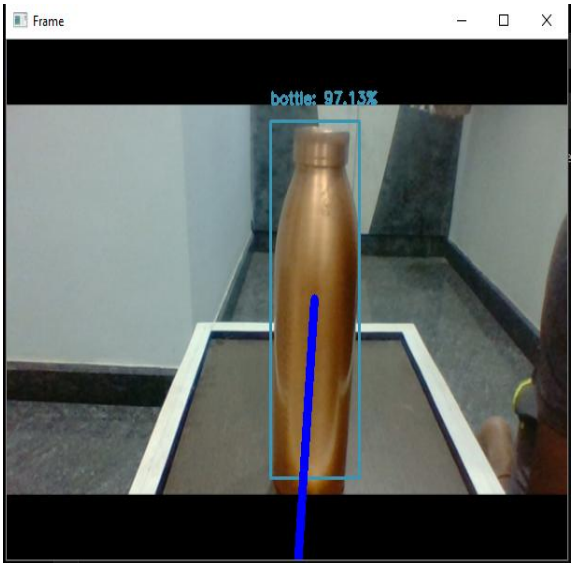

Figure 10: Object found

1. Approximate FPS : 15 fps upto 20
2. Elapsed time : 4.43s
3. Objects level of accuracy: 97.13%

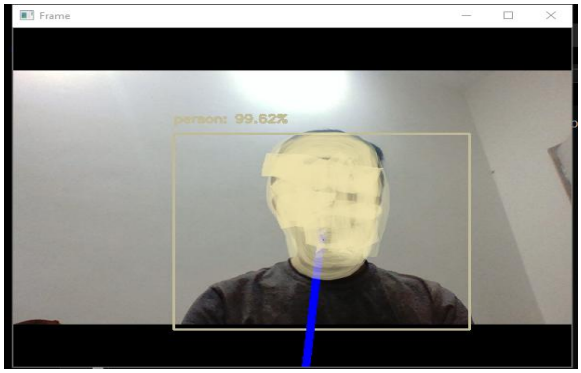

Figure 11: Object found.

1. Approximate FPS : 15 fps up to 20 fps
2. Elapsed time : 3.36s

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

3. Objects level of accuracy: 99.62%

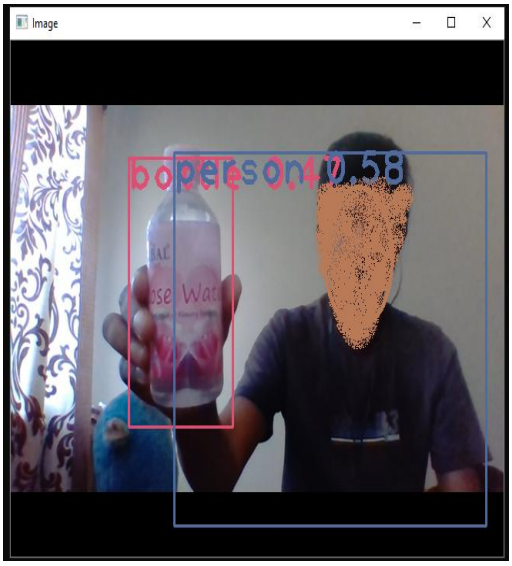

Figure 12: Multiple Objects found.

1. Approximate FPS : 2 fps upto 3 fps
2. Elapsed time : 2 - 3 s

From these pictures we can clearly see the recognized object and its labels are being displayed and conveyed to the person via text to speech engine. We can see the level of confidence of an object at top left of recognized object which shows the accuracy of object detected and recognized in percentage.

The confidence level may vary based on the object and the picture quality. comparision of yolo with other object detection models is as follows

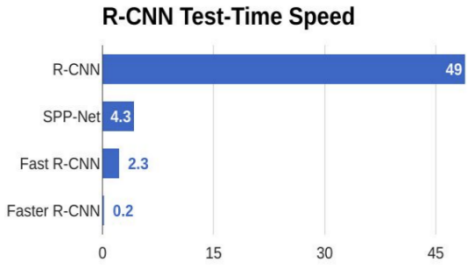

Figure 12: comparison between faster R-CNN (YOLO) and other models

By the above comparison , we can tell that the model that we implemented is taking less time to detect objects and also is very accurate due to Faster R-CNN technique.

7. Conclusion and Future Enhancements

As we have explored up until now, we got the significance of dataset that is involved and how the machine is capable of helping blind via using existing technologies. Technologies such as Image Recognition, Image Classification ,Face Identification and Real life detection's are done and processed through the Deep learning applications and had been a great help to mankind.

The objects around can now be heard by the visually impaired and this enables them to know their position and helps them to decide where to go without help from anyone else. It not only makes their travel easier but also saves a lot of time and energy and enhances their safety on roads. It makes the life of visually impaired and illiterate better by making them independent. As the matter is discussed, we proposed a simple, yet unique and easy to configure system that would help the blind in some ways which would have been helped from long back. The aforementioned system can be used by anyone through zero knowledge about it.

This paper provides the system that implements real-time object detection for detecting the nearby objects and speaking out the detected objects. This will be helpful for the blind to roam around without help of others. This system currently works only on windows based systems with high computation power. As everyone today uses android smartphones this system can be enhanced to work on android environments with medium computation power.This enhancements can be very use full for users as they don't need to carry heavy laptops wherever they go.

Also this system can be implemented using IoT smart goggles that use internet and cloud technology that have high computation power and can improve detection's. Using dedicated cameras can also boost the detection success rate.

References

[1] AnuArora, Anjali Shetty "Common Problems Faced By Visually Impaired People",2014

[2] "Visual Challenges in the Everyday Lives of Blind People " Erin Brady, , Yu Zhong, Samuel White, Jeffrey P. Bigham ,Meredith Ringel Morris 2013

[3] Lalit Kumar, Daily Life Problems Faced by Blind People 2018. https://wecapable.com/problems-faced-by-blind-people/

[4] Real-Time Object Detection,You Only Look Once:UnifiedSantoshDivvala, Ross Girshick, Ali Farhadi University of Washington, Allen Institute for AI, Facebook AI Research,Joseph Redmon, 2016

[5] YOLO9000:Better, Faster, Stronger Ali Farhadi University of Washington, Allen Institute for AI, Joseph Redmon, 2016

[6] Peng Zheng, Shou-taoXu, "Object Detection with Deep Learning: A Review",Zhong-Qiu Zhao, Xindong Wu 2019

[7] Sandeep Kumar, AmanBalyan, ManviChawla "Object Detection and Recognition in Images", 2017

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

[8] P. Chokkalingam ,Duraimurugan, S. "Real-Time Object Detection with Yolo"Geethapriya. S,,2019
[9] Pratik Kalshetti, AshishJaiswal, NamanRastogi, PrafullGangawane "Object Detection", 2018
[10] XinyiZhou,Wei Gong, WenLong Fu, Fengtong Du "Application of Depp Learning in Object detection", 2017
[11] Manish Kumar, HarshalKanchan, Himani Singh, KumariNeha "Deep Learning through Image Analysis of Real -Time videos" ,2018
[12] AbhijithShaji, AbhisheAravindan, NishamRafeeque, Naveen K K "READING ASSISTANT FOR VISUALLY IMPAIRED PEOPLE",2018
[13] Sidra Abid Syed, TahaMushtaqShaikh, Fatima Aijaz, NimraMehmood, Shujaat Ahmed "Blind Echolocation Device with Smart Object Detection" 2019
[14] Hao Yang, Hao Wu, Hao Chen "Detecting 11K Classes: Large Scale Object Detection without Fine-Grained Bounding Boxes", 2019
[15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016
[17] An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera , Yifeng Yang, Liaoyuan Zeng, JianwenChen,Wenyi Wang, Sean McGrath ,Rumin Zhang,, 2018
[18] Alexander Womg, Mohammad JavadShafiee, Francis Li, Brendan Chwyl "Tiny SSD: A Tiny Single-Shot Detection Deep Convolutional Neural Network for Real-Time Embedded Object Detection", 2018
[19] JianwuDang , Yangping Wang , Song Wang "Pedestrian Detection Based on YOLO Network Model",2018
[20] Real-time face detection based on YOLO**Wang Yang ,Zheng Jiachun,2018**
[21] Joseph Chet Redmon, "Survival Strategies for the Robot Rebellion", Avaiable :https://pjreddie.com
[22] BytePace, "What is Tesseract and how it works?" Jun 10, Available: https://medium.com/@Bytepace/what-is-tesseract-and-how-it-works-dfff720f4a32
[23] NanoDano, "Text-to-speech in Python with pyttsx3" 2018, Avaiable: https://www.devdungeon.com/content/text-speech-python-pyttsx3
[24] Prince Grover " Evolution of Object detection and localization algorithms",2018
[25] https://towardsdatascience.com , Object detection Simplified, Prakhar Ganesh,"https://towardsdatascience.com,2019
[26] Vladimir Nasteski , An overview of the supervised machine learning Methods,2017
[27] AakankshaSharma . A review of supervised machine learning algorithms.Narina Thakur, In Computing for Sustainable Global Development ,Amanpreet Singh2016
[28] Supervised Learning Wikihttps://en.wikipedia.org
[29] Van Gool, C. K. I. Williams, , and A. Zisserman. The PASCAL Visual Object Classes Challenge. M. Everingham, L,J. Winn, 2010.
[30] S. M. Ali Eslami,The PASCAL Visual Object Classes Challenge: A Retrospective, Luc Van Gool Christopher K. I. , Andrew Zisserman, Williams John Winn,2015
[31] Neural Networks Wikipedia, https://en.wikipedia.org/wiki/Neural_network
[32] Montserrat, Daniel Mas, et al. "Training object detection and recognition CNN models using data augmentation." Electronic Imaging 2017.10 (2017): 27-36.
[33] and Dennis Sng. "Deep learning algorithms with applications to video analytics for a smart city: A survey." arXiv preprint arXiv:2015,Wang, Li,.
[34] Y.A. Bachtiar, Convolutional Neural Network and Maxpooling Architecture on ZynqSoC FPGA,T. Adiono,2019
[35] J. Redmon , Real-time grasp detection using convolutional neural networks. CoRR, A. Angelova. 2014.
[36] YOLO, "Understanding of Object Detection Based on CNN Family", Juan Du1, China.
[37] DumitruErhan, Christian Szegedy, Alexander Toshev, "Scalable Object Detection using Deep Neural Networks", 2014
[38] Yi Li, Kaiming He, "R-FCN: Object Detection via Region-based Fully Convolutional Networks"Jifeng Dai,, 2016, Jian Sun.
[39] Karen Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition", AndrewZisserman,
[40] "Vision Impairment and Blindness", https://www.who.int/news-room/fact sheets, 2019
[41] G. Hagargund, S. V. Thota, M. Bera, and E. Fatima, "Image to Speech Conversion for Visually Impaired," Vol. 03, No. 06, pp. 7, 2017.
[42] F. Benzarti,"Object detection and identifification for blind people in video scene," H. Jabnoun, H. Amiri, 2015.
[43] "Rich feature hierarchies for accurate object detection and semantic segmentation."Girshick, Ross, 2014

May - June 2020
ISSN: 0193-4120 Page No. 4531-4541

[44] "Faster R-CNN: Towards real-time object detection with region proposal networks." Ren, Shaoqing,2015.
[45] "Improving neural networks by preventing co-adaptation of feature detectors." 2012,Hinton, Geoffrey E.
[46] Alex, Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks."Krizhevsky, IlyaSutskever, 2012.
[47] YufanTao,YihangChen.Object Detection Based on YOLO Network,Chengji Liu , Jiawei Liang , Kai Li ,2018
[48] JieZhu,Text-To-Speech quality evaluation based on LSTM Recurrent Neural Networks, Meng Tang ,2019
[49] Vangiebeal, TTS - text to speech https://iewww.webopedia.com/TERM/T/TTS.html
[50] SpeechrecognitionLibrary, https://pypi.org/project/SpeechRecognition/