Third International Conference on "Materials, Computing and Communication Technologies"
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Object Detection and Translation for Blind People Using Deep Learning

Dr. E. Punarselvam, Mr. R. Madhan Chakravarthy, Mr. T. Sudharsan, Mr. N. Visweshkannan, Mr. S.S. Yasvanth

[1]Professor and Head, [2]Student

Department of Information Technology, Muthayammal Engineering College (Autonomous), Rasipuram-637408, Tamil Nadu, India

Article Info
Volume 8, Issue 7
Page Number : 39-46

Publication Issue :
May-June-2022

Article History
Accepted: 01 June 2022
Published: 20 June 2022

ABSTRACT

There are million of people in India alone that are visually impaired. so, it's essential to understand for visually impaired people to recognize a product of their daily use so we made a system to identify products in their everyday routine by this system. There are many papers on this topic that will help a blind person. The paper helps a blind person in their daily use. The system consists of a camera, a speaker and an image processing system. The project tries to detect the object and transform that object into the audio form and inform blind person about those objects. The system consists of a box which has a portable camera and a system which will process that image. image are captured with a portable camera device with real-time image recognition on existing object detection models. after detecting an object that information is translate into audio

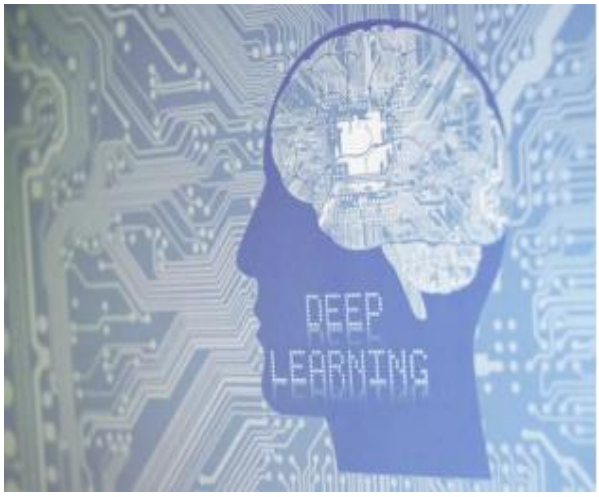
I. INTRODUCTION

DEEP LEARNING

Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features. Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features.

The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning. Deep learning excels on problem

domains where the inputs (and even output) are analog. Meaning, they are not a few quantities in a tabular format but instead are images of pixel data, documents of text data or files of audio data. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.



OpenCV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms,which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.Along withwell-established companies like Google, Yahoo, Microsoft,Intel,IBM, Sony,Honda,Toyota that employ the library, there are many startups such as Applied Minds, Video Surf, and Zeitera, that make extensive use of OpenCV. OpenCV's deployed uses span the range from stitching street view images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects atWillow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

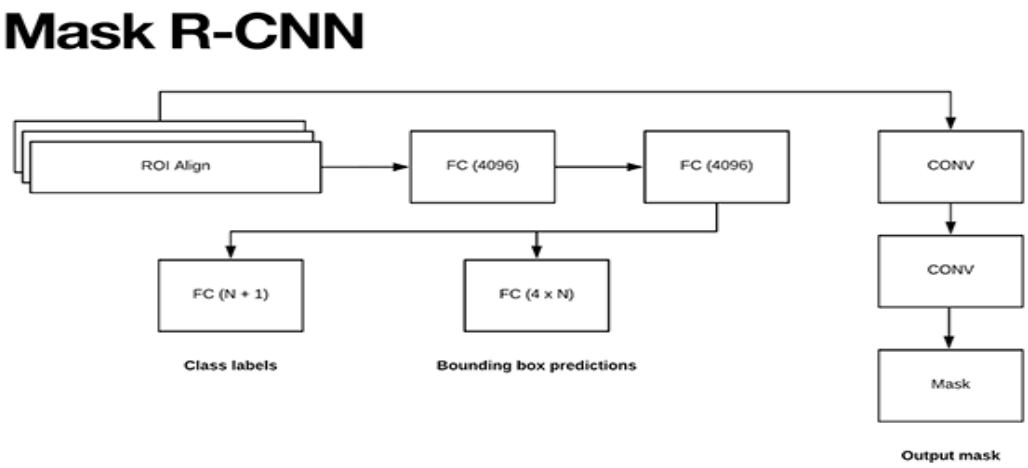
TENSORFLOW

TensorFlow is a free and open-source software library for data flow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google, TensorFlow is Google Brain's second-generation system. Version 1.0.0 was released on February 11, While the reference implementation runs on single devices, TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for

general-purpose computing on graphics processing units). Tensor Flow is available on 64-bit Linux, mac OS, Windows, and mobile computing platforms including Android and iOS. Its flexible architecture allows for the easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices.

ALGORITHM

RCNN- Regional Based Convolutional neural network



ALGORITHM

```
layers = [
imageInputLayer([r c d])
convolution2dLayer(3,60,'Padding','same')
batchNormalizationLayer
reluLayer
maxPooling2dLayer(2,'Stride',2)
convolution2dLayer(3,30,'Padding','same')
batchNormalizationLayer
reluLayer
maxPooling2dLayer(2,'Stride',2)
convolution2dLayer(5,30,'Padding','same')
batchNormalizationLayer
reluLayer
maxPooling2dLayer(2,'Stride',2)
fullyConnectedLayer(13)
softmaxLayer
classificationLayer];
```

R-CNN

S-Convolutional Neural Network (CNN) operates from a mathematical perspective and is a regularized variant of a class of feed forward artificial network (ANN) known as multilayer perceptron's that generally means fully connected networks in which every neuron in a layer is connected to all neurons in the further layers

Regularization applies to objective functions in ill-posed optimization problems and adds on the information in order to solve an ill-posed problem or to prevent over fitting

Now, we'll see how CNN trains and predicts in the abstract level so, When it comes to programming a CNN, it usually takes an order 3 tensor as input with shape (no. of mages) x (image width) x(image depth) that sequentially goes through a series of processing like convolutional layer, a pooling layer, a normalization layer, a fully connected layer, a loss layer, etc, that makes abstracted images to a feature map, with the shape (no. of images) x (feature map width) x (feature map Channels) [7,16]. Here, tensors are just higher-order matrices and below we have given layer by layer running of CNN in a forward pass:

$$x^1 \rightarrow w^1 \rightarrow x^2 \rightarrow \ldots \rightarrow x^{L-1} \rightarrow w^{L-1} \rightarrow x^L \rightarrow w^L \rightarrow z^L$$

Where usually an image (order 3 tensor). It goes through the processing in the first layer, denotes parameters involved in the first layer's processing collectively as a tensor . The output of the first layer is , which also acts as the input to the second layer processing and the same follows till all layers in the CNN have been finished, which outputs . To make a probability mass function, we can set the processing in the (L-1) th layer as a Soft Max transformation of (cf. the distance metric and data transformation notes) last layer is the loss layer. Let us suppose here t that is the corresponding target (ground-truth) value for the input , then a cost or loss function can be used to measure the discrepancy between the CNN prediction and the target t, for which a simple loss function can be given as follows:

$$\frac{p}{1-p} = b^{\beta_0+\beta_1 x_1+\beta_2 x_2}$$

$$p = \frac{b^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{b^{\beta_0+\beta_1 x_1+\beta_2 x_2}+1} = \frac{1}{1+b^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}}$$

By the above formula when are fixed, either the probability that Z=1 for a given observation or the log-odds that Z=1 for a given observation can be easily computed. In a logistic model [17,18], the main use-case is to be given the probability p that Z=1 and an observation (.x1,x20)

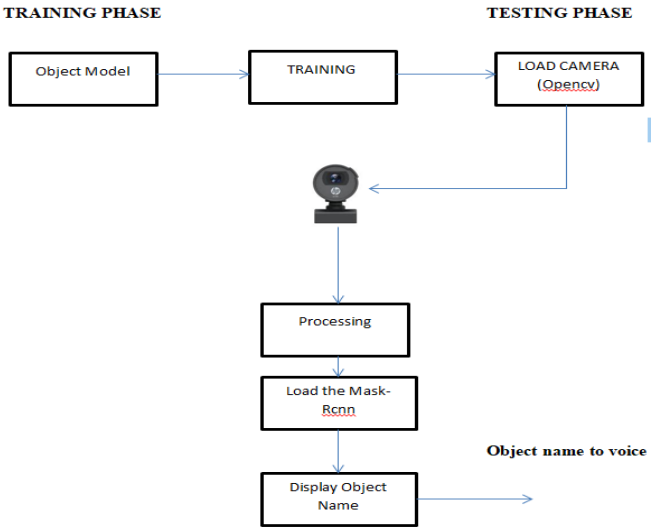
II. PROPOSED SYSTEM

This project tries to detect the object and transform that object into the audio form and inform blind person about those objects. Our system consists of a box which has a portable camera and a system which will process that image. image are captured with a portable camera device with real-time image recognition on existing object detection models. after detecting an object that information is translate into audio.

ADVANTAGES

- Low cost system is used.
- No special external hardware is needed.
- Software based system leading to low chances of complete system failure

ARCHITECTURE



MODULES

- Generation of Training and Testing
- Generation of Training Model
- Testing the Mode
- Detect Object name
- Convert to voice

MODULES DESCRIPTION

TRAINING PHASE

- **Generation of Training and Testing**

This is the first module of this system, (Training the model on the coco.name and yolo.weight.cfg and yolo3.cfg using Tensorflow & Keras)

TRAINING PHASE

- **Testing the Mode**

Open camera, Load the Model, Identify the face mask wear or not

This the second module of this system, (Loading the trained model and applying detector over live video stream with the help of camera to detect the object

- **Detect Object name**

To identify the object with the help of model using coco.name and yolo.cfg and yolo.weights.cfg

- **Convert to voice**

To convert text to voice with the help of GTTS library in python to make voice output in the headphones.

III. CONCLUSION

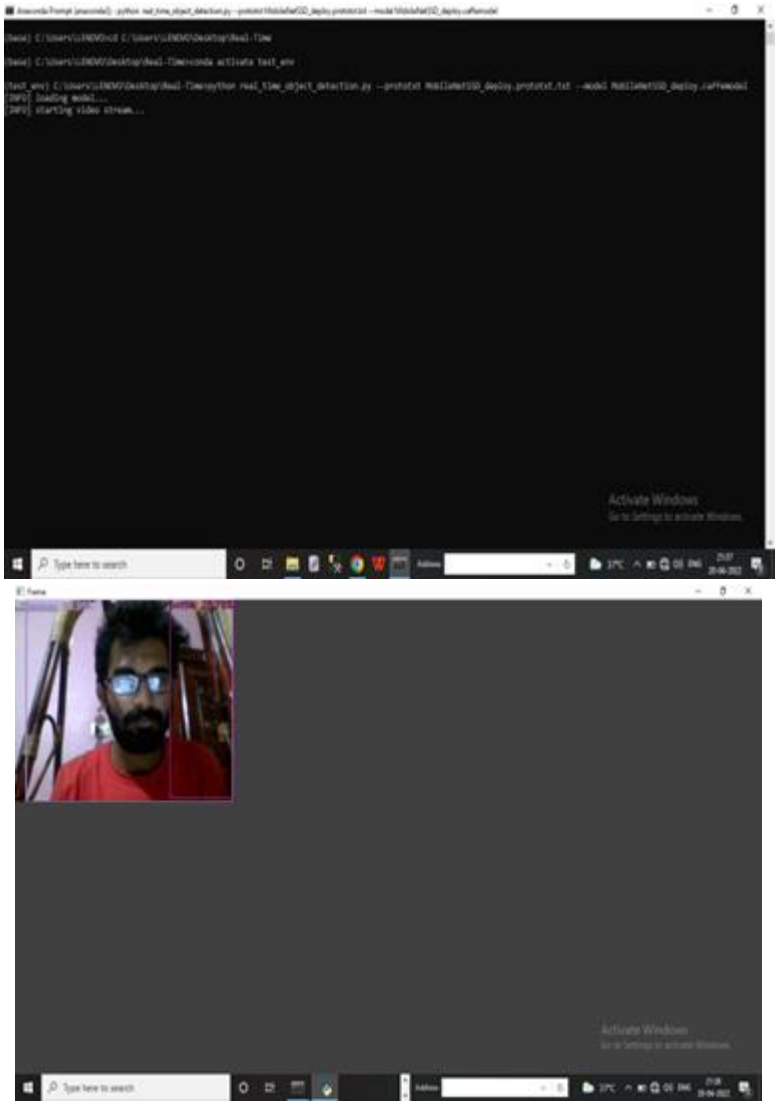
The Project entitled “Object Detection and Translation for Bind People Using Deep Learning” has been developed and this satisfies all proposed requirements. The system is highly usable and user friendly. All the system objectives have been met. All these phases of development are done according to methodologies. The

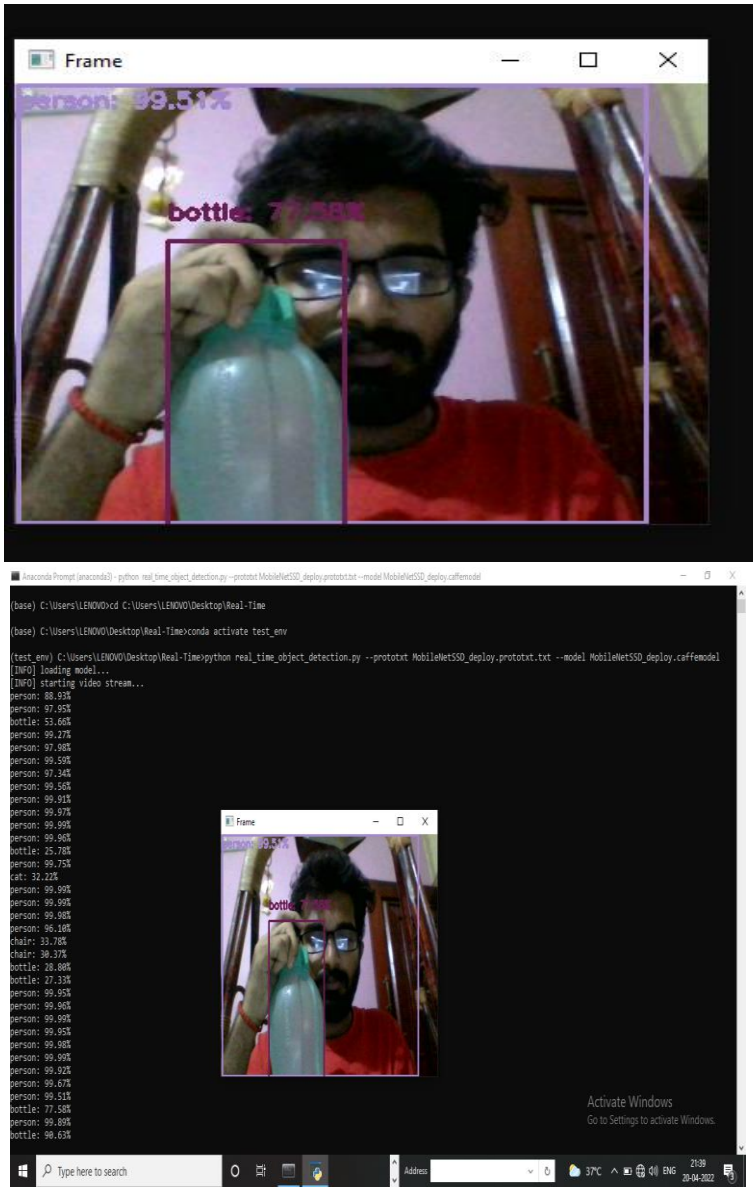
application will execute successfully by fulfilling the objectives of the project. Further extensions to this application can be made as required with minor modifications.

IV. FUTURE ENHANCEMENT

Discovering of object is a very time exhausting process to draw large quantities of bounding boxes manually. To release this burden, semantic prior unsupervised object discovery multiple instance learning and deep neural network prediction can be integrated to make the best use of image-level supervision to cast object category tags to similar object regions and improve object limits. Furthermore, this model is loaded into android, and objects are detected in mobile camera, and those object names are spelled out by the voice assistant API in the app which is helpful for blind people to navigate oneself.

V. RESULTS





VI. REFERENCES

[1]. Jinsu Lee, Sang-Kwang Lee, seong- Yang.”An Ensemble Method of CNN Modelsfor Object Detection”,IEEE,2018.

[2]. SebastianHandrich,”„FaceAttributeDetectionwith MobileNetV2 and NasNet- Mobile”, IEEE,2018.

[3]. Amith verma, Toshanal Meenpal, Asthosh Balakrishnan, “Facial Mask detection Using Semantic /segmentation”, ICCCs, 2019.

[4]. Md.saabbir Ejaz, Md. Rabiul Islam, Md. Sifutullah, ananya sarker, “Implementation of Principal Compenent analysis on Masked and Non-Masked Face Recognition”, IEEE, 2019.

[5]. Rafiuzzaman Bhuiyam, Sharun akter Khsushbbu, Md.Sanzidul Islam,”Adeep Learning Based assestive system to classify COVID-19 Face Mask for Human Safety with YOLOv3,IEEE,2019

[6]. Aniruddha Srinivas Joshi, shreya Srinivas Joshi, Goutham Kanahasabai, Rudraksh Kapil,"Deep Learning Framework to detect Face Mask from video footage",International Conference on computer Intelligence and communication Networks,2020

[7]. Pathasu doungmala, Katanyoo Klubsuwan,"Half an dFull Hlemet wearing detection in Thailand using Haar Lke Feature and Circle Hough Transformation on Image Processing",IEEE, 2016.

[8]. Yiyang Zou,Lining Zhao, shanxng Qin, Mingyag pan,"ship target detection and identification based on SSD_Mobilenet V2",IEEE,2016.

[9]. Rohith C A, shilpa A Nair, Parvathi saniNair,Nithin Prince John,"an efficient Helmet detection For MVD using Deep learning",IEEE, 2019.