Eye for Blind

Latha L, Prrithivi R, Varnessha V, Priyadharshini B

***Abstract*— As Vision is one of the vital sense of human beings, that plays an important role in perceiving the surrounding environment. The blind's capability to move around a particular place, organizing daily activities are of vital importance for their well-being. Organizing daily activities can be especially difficult; it is too hard to distinguish different items, just by feeling them with their hands.Even though many papers have been published that propose a variety of vision related services, still there is a scope for improvement in developing new electronic aids for the blind [4].**

This paper proposesa system for identification of surrounding objects for blind people and also guiding them through voice. This method is actually based on the object detection technology and it uses YOLO algorithm to perform the object detection and classification of objects [4]. Thus, a visual substitution system based on feature extraction and matching is developed to recognize and locate objects and guide the blind people using voice control.

Keywords:

Object detection and classification, YOLO algorithm, Fully convolutional network and text-to speech conversion.

I. INTRODUCTION

Blind people face several problems inlife, one of the most important one is detection of obstacles and objects around them. World

Health Organization survey, points out that there are approximately 285 million people have visual impairments, 39 million are blind and 246 million have a decrease of Visual acuity[4].

As Visual impairment results in severe consequences linked to visual function, such as the activities like identifying objects, communication, reading and writing, it is proposed to develop a system to help the blind, and to tackle the primary difficulty of locomotion by knowing the surrounding around them. So an algorithm is developed to recognize and locate objects in a real time scenario and to inform about the location of objects to the blind through voice.

II. DRAWBACK OF EXISTING SYSTEM

As many applications require the need for detecting more than one object class, the processing speed and accuracy loss becomes higher. Also detecting those objects, where the objects and background pixels are interlaced with background pose again a problem.

Revised Manuscript Received on September 10, 2019.

Latha L, Professor, Department of Computer Science Engg, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India.

Prrithivi R, Student, Department of Computer Science Engg, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India

Varnessha V, Student, Department of Computer Science Engg, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India

Priyadharshini B, Student, Department of Computer Science Engg, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India

III. PROPOSED SYSTEM

The proposed system uses pre-trained dataset known as Common Objects in COntext. Images in the COCO dataset are obtained from the day to day scenes, and multiple objects can also be found in the same image. Generally, there are three stages in object detection.

- data preprocessing
- detection and localization
- Classification

Initially, detect the desired object in the respective image by localizing it. For that Opencv4 is used to detect objects or things in the particular given image. After objects are being localized in the image dimension level, it is gone through to next level for drawing bounding boxes around it.

After that, it is directed to do classifier to classify the objects inside the bounding boxes according to the already trained classes. The basic algorithm for detecting multiple objects at single shot we are using SSD algorithm (Single Shot Detection). This is the traditional method detecting multiple objects simultaneously at the same time.

Later then, YOLO has its own detection algorithm which is derived from SSD.Before the detected items in the image sent to the classification process, there is another intermediate method used to prevent unwanted classification process of objects having low threshold value(in our case default is 0.3 ie.,30%) Objects inside the bounding boxes which having a threshold value below 0.5 are omitted to the next level.

For further improved results, another parameter is used named Confidence (in our case default is 0.5) to prevent weak detection results.

The network type of YOLO is convolutional neural network, which is specifically called F-RCNN (Fast recurrent neural network). By using this network YOLO achieves faster classification of objects. The average time for detecting and classifying objects in a image depends on the system configuration. With proper CUDA aided machine the detection rate is 10 milliseconds/image. But in our case, its 1.5-1.6 seconds.

For detecting objects in videos, we are using a technique called Frame Extraction. Here the whole video is splitted into individual frames/images and then in turn each frame is sent it through the detection and classification phase and naming the class labels.

The average detecting rate of single frame is 1.2 - 1.8 seconds. The final video output time varies depends on the frames of given input video.

Finally a video writer is used to join the extracted frames into original video with object detection.

After labeling the classification results, the names are given to text to speech converter (Google API used for better results at a accuracy rate of 99%) and convert the labels into voice data. The format that we saved the voice data is .wav file extension.

Then using mpg321codec player we accessing the laptop speaker to hear the classified labeled names.

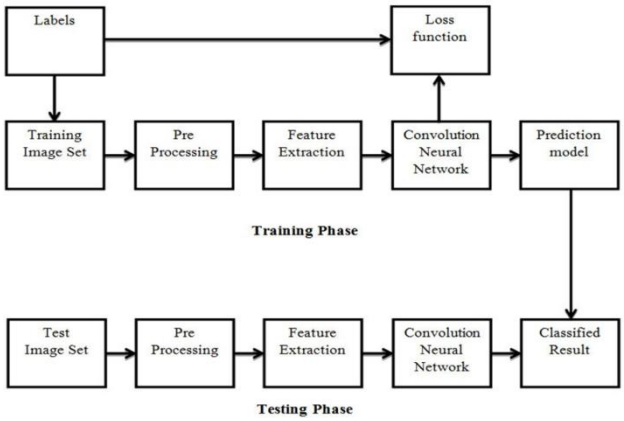

Fig.1 Our proposed system- an overall model

IV. WORKING OF THE SYSTEM

A fully convolutional neural network is applied to an input image. The network divides the images into grids/regions. A 13*13 grid is used in our system. It performs the detection and classification in almost all the grids. It also predicts 2 bounding boxes and its associated class probabilities for that grid.

The bounding boxes are characterized by five attributes. They are x , y - denoting the centre of the bounding box, w , h - Width and Height of the box with respect to the center and bounding box, and c - Confidence of the class probabilities.

Non Zero Suppression Algorithm:

An initial threshold value for confidence level is set. The bounding boxes with confidence less than the threshold is eliminated and other boxes refers to the presence of an object. Non Zero Suppression Algorithm is applied to get the needed boundary boxes.

Then the image is passed into 28 CNN sets. Where each set has 3 layers namely Convolution Layer**,** splits the image into 3 layers namely the RGB split. Then a filter of dimension similar to the image's dimension is applied on the RGB layer. In mathematical term the image is represented as a matrix. Thus in this step a matrix multiplication between the image and the filter is done. As a result each grid gets a value and is passed on to the max pooling layer where the grid with values obtained from the previous step is obtained. Then based on a fixed dimension say 2*2 the maximum number from those grid are obtained throughout the image.

Finally given to the resultant layer where the image matrix from the max pooling layer has obtained with a reduction in dimension. This resultant image matrix is passed to the next set.Similarly it happens through all the 28 sets and the result is a FCNN a 1D matrix with values.When the image is passed from one set to another the dimension reduces thus to facilitate the max pooling process, Padding is carried out.

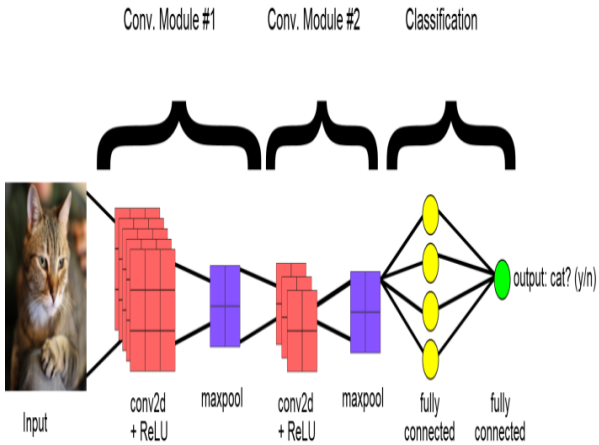

Fig.2 Working of F-CNN

The final 1D matrix is finally passed to a **S**upport **V**ector **M**achine (SVM) for classification.The matrix values are plotted in a graph, then based on the separation vectors, the class label for each object are determined.The **C** parameter controls the tradeoff between smooth decision boundary and classifying training points correctly.

Text To Speech Conversion

Using YOLO, the output is obtained in text format. Using the google API, the Text-to-Speech conversion is done, the resulting text is converted into speech. The text obtained as a result of classification is converted into UNICODE format (ASCII value).Then passed to the Google API.There the Unicode is matched with ones Google API contains.Then the respective speech is formulated.

V. RESULTS & DISCUSSION

Experimental results for object detection using the proposed approach have been produced for several image sequences**.**Asingle neural network is used that divides the input image into regions and predicts bounding boxes and probabilities for each region. This model performs better thanclassifier-based systems, as it looks at the whole image at the test time, which makes its predictions by including the global context in the image. The output is converted into voice from text using Google API for visually challenged. Thus, a system is designed for visually challenged people for identifying the objects around them.

VI. REFERENCES

1. JayanEledath, Luke McDowell, Michael Hansen, Lambert Wixson, Art Pope, Gary Gendel,"Real-Time Fixation, Mosaic Construction and Moving Object Detection from a Moving Camera",1998.
2. Mohamed Dahmane, Jean Meunier," Real-time moving object detection",May 2005.
3. Arnab Roy, SanketShinde and Kyoung-Don Kang," An Approach for Efficient Real Time Moving Object Detection", 2009.

4. Liming Wang, Jianbo Shi, Gang Song, and I-fan Shen,"Object Detection Combining Recognition and Segmentation",May 2009.
5. Kalpesh R.Jadav, M.A.Lokhandwala,A.P.Gharge," Vision based moving object detection and tracking", May 2011.
6. Balakrishnan, Nagaraj, Arunkumar Rajendran, and Karthigaikumar Palanivel. "Meticulous fuzzy convolution C means for optimized big data analytics: adaptation towards deep learning." International Journal of Machine Learning and Cybernetics (2019): 1-12.
7. Shamshad Shirageri1, Dr.G.R.Udupi, Prof.G.A.Bidkar," Design and development of Optical flow based Moving Object Detection and Tracking (OMODT) System",April 2013.
8. Mahesh C. Pawaskar, N. S.Narkhede and Saurabh S. Athalye," Detection Of Moving Object Based On Background Subtraction",June 2014.
9. DipaliShahare and RanjanaShende," Moving Object Detection with Fixed Camera and Moving Camera for Automated Video Analysis",2014.
10. Asim R. Aldhaheri and Eran A. Edirisinghe," Detection and Classification of a Moving Object",2014.
11. Mohamed Hammami, Salma KammounJarraya, and Hanêne Ben-Abdallah," A Comparative Study of Proposed Moving Object Detection Methods",May 2011.