Object Detection And Classification Using Deep Neural Network For Blind People

Dr.L.Murali[1], R.Priyadharshini[2], S.Saranya[3], N.Suvetha[4]
[1, 2, 3, 4] Dept of ECE
[1, 2, 3, 4] P.A.College of Engineering and Technology, Pollachi.

Abstract- *Object detection primarily deals with identification of objects that is present in real world such as people, animals and objects of suspense or threatening objects. Object detection algorithms use a good vary of image processing application for extracting the objects desired portion. It is usually employed in application such as retrieval of images, defence, medical field, and security purposes. During this project it deals with real time object detection and classification for blind people. Image has been detected and classified using the Deep Neural Network algorithm. Once after the detection of the image, the output is given in the form of audio to the blind person through earphones or speakers without any misclassification. This project has been implemented using Python Language.*

Keywords- Object detection, Deep neural network, classification.

I. INTRODUCTION

Human can easily detect and identify objects present in an image. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects and detect obstacles with little conscious thought. With the availability of large amount of data, faster GPUs, better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy. Image processing is a technique to enhance raw images received from cameras or sensors. Based on satellites, space probes and aircraft pictures taken in normal day-to-day life for various application.Various techniques have been developed in Image processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned space crafts, space probes and military reconnaissance flights. Image processing systems are becoming popular due to easy availability of powerful personnel computers, large size memory devices, graphics software etc., Image processing is used in various applications such as Remote Sensing, destructive Evaluation, Forensic studies, Textiles, Material science, Military, Film industry, Document processing, Graphic arts, Printing Industry.

II. OBJECT DETECTION USING BACKGROUND SUBTRACTION

A.IMPLEMENTATION OF BACKGROUND SUBTRACTION

The background subtraction is the process of separating out foreground objects from the background in the sequence of video frames as shown in figure 3.1. It is widely used approach for detecting moving objects from static cameras. It should segment objects of interest when they first appear (or reappear) in the scene.

Background objects that are inserted or removed from the scene that become/was the part of the background; If an object that was part of the background that is moved, both it and part of the background that was behind the object appear to change; if a foreground objects pixel characteristics are almost the same as the background. There are various problem in the background modeling that must be addressed or handled.

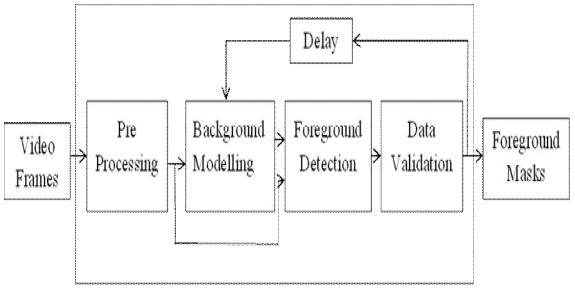

Figure 3.1 Flow diagram of Background Subtraction algorithm

The background subtraction is that the method of separating out foreground objects from the background within the sequence of video frames. It's wide used approach for detection moving objects from static cameras. It ought to phase objects of interest after they 1st seem (or reappear) within the scene. Foreground detection compares the input video frame with the background model and identifies the candidate foreground pixels from the input image. This system cannot be won't to enhance distinction in bright pictures like an out of doors scene beneath serious fog.

The foreground regions area unit fully grown from robust foreground pixels by together with the neighboring pixels with absolute distances larger than smaller threshold.

B. RESULT USING BACKGROUND SUBTRACTION

The frame differencing method is to detect a target, as this only requires the camera to be kept static for a while. Frame differencing is the simplest method for moving object detection, because the background model is simply equal to the previous frame. However, sometimes the blob contains too many background pixels when the target is moving very fast, or the blob may lose part of the target information when the target is moving slowly.

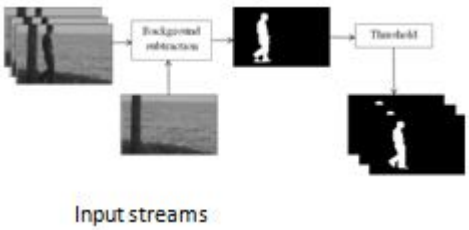

Figure 2.2: Foreground detection process

The Figure 2.2 shows the foreground detection process in a frame differencing method, where the target detection result using the frame differencing method.

It is impossible to obtain pure foreground pixels when using the frame differences as the background model, but by using the following method, we can remove the background pixels and retrieve more foreground pixels, on the condition that the color of the foreground is not similar to the pixels of the nearby background.

C.DRAWBACKS:

The background image isn't fixed however must adapt to camera oscillations,illumination changes, sudden changes, motion changes, high frequency background objects(like branches in tree, waves and similar), changes in the background geometry. It ends up in misclassification, less accuracy, no multiple object detection.

III. OBJECT DETECTION USING DEEP NEURAL NETWORK

A Deep neural network(DNN) is an neural network with multiple layer between the input layer and output layer as shown in fig1.1. Each mathematical calculationis considered a different layer, and complex DNN have more number of layers.

Earlier versions of neural network like the first perception where shallow, composed of one input output layer and almost only one hidden layer in between. More than three layers (including both input layer and output layer) qualify as deep neural network.

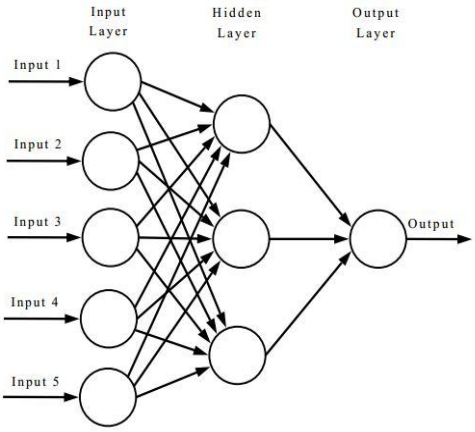

Figure 3.1: Deep neural network with multiple layers.

Typically, and neural network is first of all educated or fed big quantities of statistics. Training includes imparting input and telling the community what the output out to be. For example, to build a network to become aware of the faces of actors, preliminary education is probably a sequence of photographs of actors, non-factors, masks, animal faces and so forth. Each enter is accomplished via the matching identification,consisting of actors names, not actor or not human statistics. Providing the solutions allows the model to regulate its internal weightings to discover ways to do its tasks higher.

To overcome the drawbacks in the background subtraction method, we have proposed a new system which would help in better detection and classification of the objects. The camera is used to capture the scene that the blind people encounter in their daily life. The captured image was then compared with the trained image in the data base. After that the system generated a voice synthesis of the detected object to assist the blind people to identify the object easily. The steps in developing this system and software shown in figure 1.3

A. Obtaining the image dataset:

A standardized dataset called COCO (common object in context) dataset is an excellent object detection dataset with

80 classes, 80,000 training images and 40,000 validation images

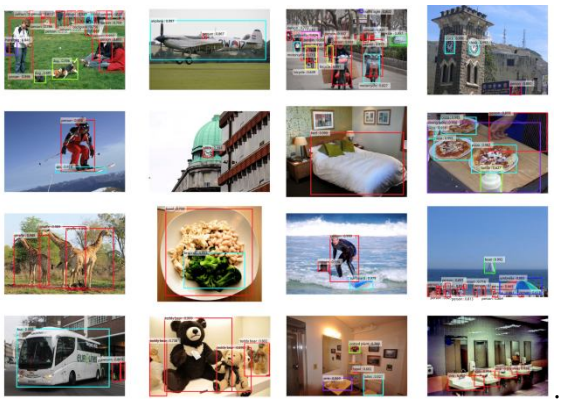

Figure 3.2: Example of dataset COCO

Each image has a size of only 32*32 pixels. The small size makes its sometimes difficult for humans to recognize the correct category, but it simplifies things for our computer model and reduce the computational load required to analyse the images.

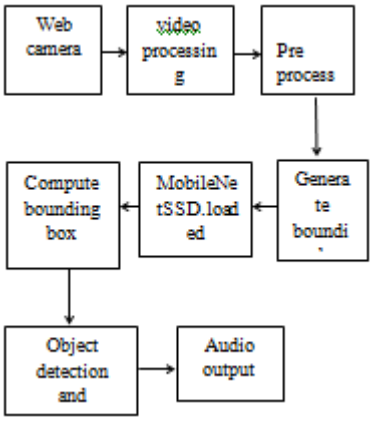

Figure 1.3: Block diagram for SSD detection

B.GENERATE BOUNDING BOX:

In geometry, the minimum or smallest bounding or enclosing box for a point set(S) in N dimensions is the box with the smallest measure (area, volume, or hyper volume in higher dimensions) within which all the points lie as shown.

CONDITIONS:

- In the two-dimensional case it is called the minimum bounding rectangle.
- MBRs are frequently used as an indication of the general position of a geographic feature or dataset, for either display, first-approximation spatial query, or spatial indexing purposes.
- The degree to which an "overlapping rectangles" query based on MBRs will be satisfactory (in other words , produce a low number of "false positive" hits) will depend upon the extent to which individual spatial objects occupy(fill) their associated MBR.
- If the MBR is full or nearly so(for example, a map sheet aligned with axis of latitude and longitude will normally entirely fill its associated MBR in the same coordinate space), then the "overlapping rectangles" test will be entirely reliable for that and similar spatial objects.
- Then most of the MBR will be empty and an "overlapping rectangles" test will produce a high number of false positives. One system that attempts to deal with this problem, particularly for patchy data, is c-squares.

C.MOBILE NET SSD:

Mobile Nets are a new family of convolution neural networks that are set to blow your mind, and today we are going to train one on a custom data set. SSD is a popular algorithm in object detection. Mobile net is a neural network that is used for classification and recognition whereas the SSD is a framework that is used to realise the multi box detector. Multi box is the name of the technique for bounding box regression. Only the combination of both can do object detection. Single shot multi box detector is used for real time processing as shown in figure 3.3.Well-researched domains of object detection includes face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

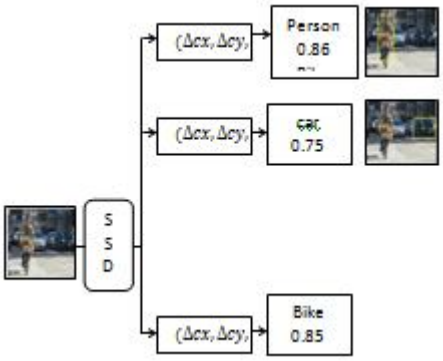

Figure 3.3: Flow diagram of SSD

E. OBJECT DETECTION USING DNN:

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such

as humans, buildings, or cars) in digital images and videos. It is widely used in computer vision task such as face detection, face recognition, video object co-segmentation. It is also used in tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, tracking a person in a video.

Every object class has its own special feature that helps in classifying the class for example all circles are round. Object class detection uses these special features. For example, when looking for circles, objects that are at a particular distance from a point (i.e. the centre) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed. A similar approach is used for face identification where eyes, nose, and lips can be found and features like skin colour and distance between eyes can be found.

Deep learning algorithm is also known as deep structured learning it is the part of broader family of machine learning methods based on learning data representations as opposed to task specific algorithms. Deep learning architectures such as deep neural network has been applied to the fields including computer vision, speech recognition, etc..

1. Use a cascade of multiple layers of non linear processing units foe feature extraction and transformation. Each successive layer uses the output from the previous layer as input.
2. Learn in supervised (e.g., classification) and/or unsupervised
   (e.g., pattern analysis) manners.
3. Learn multiple levels of representation that correspond to different levels of abstraction; the level from a hierarchy of concepts.

TRACKING:

A tracking system is used for observing of persons or objects on the move and supplying a timely ordered sequence of location data for further processing. Tracking had come to be a tedious task to use in sophisticated due to the facts items are improperly segmented. Some primary issues of erroneous segmentation are long shadows, full and partial occlusion of gadgets with every other. At segmentation level and at tracking stage coping with shadows and occlusions is vital in strong algorithm. With the regular emphasis on safety issues, thousands of surveillance cameras are established at numerous locations, such as colleges, visitor's junctions, banks and different public places. A big quality of video information is captured each minute.

F. GOOGLE TEXT-TO-SPEECH

Google Text to speech is a screen reader application developed by Google. The position of objects are detected and send it as a text string to gTTS. The class prediction of the objects detected in every frame will be a string. We will also obtain the co-ordinates of the objects in the image and append the position like

"top"/ "mid"/ "bottom"
"left"/ "centre"/ "right"

An audio signal is a representation of sound, typically as an electrical voltage for analog signals and a binary number for digital signals. Audio signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz (the limits of human hearing). Audio signals may be synthesized directly, or may originate at a transducer.

IV. RESULT AND DISCUSSION

In this result multiple objects has been detected successfully and accuracy has been displayed for better classification of objects. The objects such as water bottle and person have been detected with an accuracy of 99.34% and 88.63% respectively as shown in figure (b).


Figure(a) Figure (b)

The figure (a) shows multiple objects such as remote, cup, scissors have been detected with an accuracy of 70%, 96%, and 99% respectively.Likewise several objects can be trained and accuracy can be increased with Deep Neural Network. The chances of misclassification can be avoided and also multiple objects can be detected at the same time.

A. PREDICTION OF OBJECTS WITH POSITION

We will also obtain the coordinates of the boundary box of every object detected in our frames overlay the boxes on the objects detected and return the stream of frames as a video playback. For example "bottom left cellphone" – meaning a cell phone was detected on the bottom left of the camera view. The output window shows the position of the object with text string. If the position shows 'Mid center

person' which indicates a person was detected on the mid center of the camera view. We will also schedule to get a voice feedback on the first frame of each second.

B. AUDIO FEEDBACK

The Google Text to Speech converter is used to produce verbal description of objects detected in real time on the webcam which is more important since the person is blind. Here the annotated text is converted into voice response and give the basic positions of the objects in the person/camera's view.

V. CONCLUSION AND FUTURE WORK

The object detection and classification has been successfully implemented and objects have been classified and then the final output will be in the form of audio. This experiment can be implemented in the presence of light only. Then the results shows that the objects have been detected without any misclassification. Using Deep Neural Network chances of misclassification can be avoided and also multiple objects can be detected at the same time. Also Deep Neural Network has increased the accuracy with the help of the multiple layers present in it.

This system can be further developed by making it possible to detect the objects even in the absence of light and also detection time can be reduced. In future the accuracy of the overall project will get increased and also training of multiple object classification in Deep Neural Network.

REFERENCES

[1] Chris Stauffer and Eric Grimson, (2017) 'Adaptive background mixture models for real time tracking', in Computer Vision and Pattern recognition and Institute of Electrical and Electronic Engineers Computer Society Conference, Vol.2, pp.246-252.

[2] Gene Gindi, S. and James Duncan, (2003) 'A Complete Object recognition System as a Computer Vision Course Project', Institute of Electrical and Electronics Engineers, Vol.8.

[3] Hanxi Li, ChaunhuaShen, and Quinfeng Shi, (2011) 'Real time visual tracking using compressive sensing', in Computer Vision and Pattern Recognition, Institute of Electrical and Electronics Engineers Conference on, pp. 1305-1312.

[4] HariharaSantosh and Krishna mohan, (2017) 'Multiple Object Tracking using Extended Kalman filter, GMM, Mean shift algorithm', Institute of Electrical and Electronics Engineer International Conference on advanced communication control and computing technologies, pp.1484-1488.

[5] HendrikLensch, Martin Butz, Jeannette Bohg, Daniel Kappler, (2015) 'Object detection using Deep Learning', Institute of Electrical and Electronics Engineers.

[6] Jae KyuSuhr, Ho Gi Jung, Gen Li and Jaihie Kim, (2015) 'Mixture of Gaussian based background subtraction for Bayer pattern image sequence', Institute of Electrical and Electronics Engineers Transaction on circuits and systems for video technology, Vol.21, No.3, pp.365-370.

[7] Kalyan Kumar Hati, Pankaj Kumar Sa and BanshidharMajhi, (2016) 'Intensity range based subtraction for effective object detection', Institute of Electrical and Electronic Engineers Transaction on Signal processing, Vol.20,No.8, pp.759-762.

[8] Kyungnan Kim, ThanaratChalidabhongse, David Harwood and Lary Davis, (2017) 'Background Modelling and Subtraction by codebook construction', in Image Processing International Conference Vol.5, pp.3061-3064.

[9] Lljoo Beak, Albert Davies, Geng Yan, Ragunathan Raj Rajkumar (2018) 'Real time tracking, detection and classification of moving and stationary objects using multiple fish-eye images', Institute of Electrical and Electronics Engineers, Vol 8.

[10] Linhao Li, Quinghua and Xin Li, (2019) 'Moving object detection in video via hierarichalmodelling and alternating optimization', Institute of Electrical and Electronic Engineers Transaction on Image processing Vol.28, No.4, pp. 2021-2036.

[11] Malay Shah and RupalKapdi, (2017) 'Object detection using deep neural networks', on International Conference on Intelligent Computing and Control System,Vol.24, No 5, pp.5386-2745.

[12] Pierre-Luc St- Charles, Guillaume-Alexander Bilodeau, and Robert Bergevin, (2016) 'Universal background subtraction using word for consensus models', Institute of Electrical and Electronic Engineers Transaction on Image Processing, Vol.25, No.10, pp.4768-4781.

[13] Reagen Galvez, L. ArgelBandala, A. Elmer Dadios, P. Jose Martin, Z. and Maningo, (2018) 'Object Detection using Convolution Neural Network', Institute of Electrical and Electronics Engineers, pp.292-297.

[14] Sowmyavarma and Sreeraj, M., (2013) 'Object detection and classification in surveillance system', Institute of Electrical and Electronics Engineers Recent Advances in Intelligent Computational System, Vol 8, pp-356-424.

[15] Wenrui Hu, Yehui Yang, Wensheng Zhang, and Yuan Xie, (2017) 'Moving Object Detection using tensor-based low-rank and saliently fused-sparse decomposition', Institute of Electrical and Electronic Engineers Transcations on Image Preoceesing, Vol.26, no.2, pp.724-737.

[16] Wong, Y.C. Lai, J.A. Ranjit, S.S.S. Syafeeza, A.R. Hamid, N.A., (2019) 'Convolution Neural Network for object detection system for blind people' in the Journal of Tele Communication, Electronic and Computer Engineering Vol.11 No.2.

[17] Yuanhong Mao, Zhanzhuang He, Zhong Ma, Xuehantang and Zhuping Wang, (2019) ' Efficient Convolution Neural Networks for Object Tracking Separable Convolution and Filter Pruning', Institute of Electrical and Electronics Engineers Vol.7.