International Journal of Innovative Research in Engineering
Volume 3, Issue 2 (March-April 2022), PP: 45-51.
www.theijire.com **ISSN No: 2582-8746**

Object Detection, Convert Object Name to Text and Text to Speech

Prachi Tijare[1], Pranali Warkhede[2], Lina Godbole[3], Sayali Thakre[4], Rohini Dhakate[5], Ankit Mahule[6]
[1,2,3,4,5]*Computer science and Engineering, Nagpur Institute of Technology, India.*
[6]*Assistant Prof. Computer science and Engineering, Nagpur Institute of Technology, India.*
.

How to cite this paper: Prachi Tijare[1], Pranali Warkhede[2], Lina Godbole[3], Sayali Thakre[4], Rohini Dhakate[5], Ankit Mahule[2], "Object Detection, Convert Object Name to Text and Text to Speech", IJIRE-V3I02-45-51.

Abstract: *Object discovery is a computer technology related to visual perception and image processing that deals with the discovery of semantic objects of a certain category (such as people, structures, or animals) in photographs and electronic videos. Object Detection has applications in various areas of computer vision, including image detection and video monitoring. This is an android application built using tensorflow and open cv library in android studio. this app is able to see an object in real time. The yolo algorithm is used to identify and identify objects and reads the object name of blind people. saw an object and assigned rectangular boxes around it. identify each item and assign its tag to the item. This includes the accuracy of each diagnostic tool. Computer Vision is a branch of computer science and data processing system capable of visualizing and understanding images and scenes. the app is installed and written on Nakia, galaxy, MI Smartphones, this app can see many things in one frame. Colored images are converted to binary images using the lock method. Object acquisition is a challenge to recognition and is driven by the regional success proposal R-CNN.*

Keywords:- *Android, Morphological filters, Image tagging, Least, Significant Byte technique.*

I. INTRODUCTION

Object detection is the technique where in an image object is. The motive of object detection is to recognize and locate all known objects in a scene. intelligence to machines and making robots more and more autonomous and independent has been a sustaining technological dream for the mankind. It is our dream to let the robots take on tedious, boring, or dangerous work so that we can commit our time to more creative tasks. In real life, to achieve this goal, besides hardware development, we need the software that can enable robot the intelligence to do the work and act independently.

The searching or recognition process in real time scenario is very difficult. So far, no effective solution has been found for this problem. Despite a lot of research in this area, the methods developed so far are not efficient, require long training time, are not suitable for real time application, and are not scalable to large number of classes.

Object detection is relatively simpler if the machine is looking for detecting one particular object. However, recognizing all the objects inherently requires the skill to differentiate one object from the other, though they may be of same type. Such problem is very difficult for machines, if they do not know about the various possibilities of objects.

Mobile phones are becoming convergent platform for personal sensing, computing and communication. Integration of these facilities in the mobile will make communication more efficient and effective. Android is a development platform for the mobile applications which is having the maximum market share among the mobile technologies.

Detecting objects in a given image is the first step in image tagging applications. Tagging or labelling image is the first and most significant part of object detection. Tagging is indeed a very time-consuming process. The information to be associated with the image is tagged to the detected objects. Thus, user can get the information as and when required.

So far object detection is done by various methods for desktop applications. Identification of objects in an image on android platform is not fully developed. The present study is focused on detecting objects in an image which later can be used for various other purposes. Objects once detected can be saved and is available for further use. In the present study image processing algorithms are considered for detecting objects.

Tagging refers to attaching the identified information or data to the object. Image tagging refers to attaching the related text with the image which helps users as and when required. Though applications related to tagging in maps have been developed, there is still scope for tagging objects in images on mobiles.

Tagging should be done in appropriate place in the captured image. It should be done in such a way that the tagged information is available for users anywhere, anytime.

To accomplish this, objects in images should be identified and detected. Computer vision techniques are used for detecting and tagging objects in the image. Today, images and video are everywhere. Online photo sharing sites and social media platforms have billions. The field of theory research is dominated by machine learning and mathematics.

Using photos and videos to discover, classify, and track objects or events to "understand" the real-world scene. Computer programming and design algorithms to understand what is in this image is a field of computer vision. Computer view enables applications such as image search, robot navigation, medical image analysis, image management and much more. From

a computer perspective, i
image is a scene consisting of objects of interest and a background represented by everything else in the image.

Object detection can be further divided into soft detection, which only detects the presence of an object, and hard detection, which detects both the presence and location of the object.

II. RELATED WORK

There are many previous works that find items based solely on colour characteristics. The authors used the colour separation method using thresholding, to find flowers at the scene. They relied on the colour of the flower as it is considered to be the basic commodity of an agricultural product. They work in the HSV colour field and with histogram analysis, the colour of the flower is defined. After that, the flowers were separated from the background and obtained from photographs by a process of separating the image. Some researchers are finding things based on physical characteristics. Researchers detect two-dimensional shapes and typing shapes in an image. All known shapes are identified by dividing the images into separate corresponding regions and determining the shape element and using it to identify the type of shape.

Function presented in the received circle, square objects and triangle in the picture. Their method uses the amount of stiffness from the inserted image and the binary image is obtained by summarizing using the Otsu method. The sound is removed by a central filter and the edges are obtained using a Sobel operator. Unnecessary edge pixels are removed by minimizing them. Their method maintained 85% accuracy when tested on the selected site. Many other researchers have found objects based on both colour and shape characteristics. Identified authors of basic geometric shapes and key colours in a two-dimensional image using image processing techniques with the help of MATLAB. Key shapes are square, circle, triangle and parallelogram. The process involves converting an RGB image to a gray scale image to black and white.

The binding region of the binding parallel program is calculated regardless of the rotation angle and the relative relation of this space to the object area is calculated and compared with the predefined quantity ratios to reflect the given type of object. thing. The contribution of outstanding colour pixels helps to determine the colour of the object. The research work discussed a process related to digital image processing and geometric concepts to determine the two-dimensional shapes of objects such as a square, a circle, a rectangle and a triangle, and the colour of an object. Steps that include RGB image to black and white image conversion, colour pixel separation of object and background, region-based filtering and the use of a compound box and its metrics mathematical features.

Object metrics compared to preset values are the properties of the selected object. The discovery of the shape of objects is built consistently on their evolution. Object colours are determined by analyzing the RGB information for all pixels within each object. Their algorithm was developed and modified in MATLAB. One of the best ways to see and experience something in a scene is to learn the location features of both images. Different key points are geographically reliably under various image modifications, such as translation, rotation, scaling and many ideas that can be extracted and used to find similar key points between images, and then select the location of the object of interest. To find an object based on local factors, many researchers use local features such as SURF (Strong Ascending Factors), SIFT (Scale Invariant Feature Transform), and BRSK (non-volatile binary key points).

International Journal of Multimedia & Its Applications (IJMA) Vol.9, No.1, February 2017. On this page the Android app can detect objects in an image uploaded to a mobile gallery, based on its colour, texture. or local features of BRSK are being upgraded.

There are several possible ways to remove the service-based component of a desktop application. Ommar et.al analyzed contour-based object acquisition and used conceptual representation of concerts to determine the object and also introduced non-parametric representations of contours, curvatures and junctions to allow for your precise local performance. ChengEn lu and others have studied and argued that the discovery of a contour-based object can also be made in dense areas.

When a container model is given, it begins to rot in pieces. It is then grouped into sections and finds similarities between these piles. Based on the same thing found.

Jammie shotten and others have learned that object detection can be done using local location-based features. Here it is proposed to build two levels of supervised reading. Entering information in various formats, there is a great deal.

various forms of steganography. Darshan Mistry conducted a detailed study of the LSB method. Embedding Method Least Significant Bit (LSB) suggests that data be hidden in small parts of the cover image.

This process can be used to enter text in a 24-Bit, 8-Bit, Gray scale format. Different embedding techniques are algorithms used for the purpose of detection and extraction. After the object has been removed each item is listed for marking using the LSB method of steganography.

III. METHODOLOGY

Object detection is a challenge in computer vision it is driven by the success region od proposal methods R-CNN. The project aims to incorporate state-of-the-art technique for object detection with the goal of achieving high accuracy with a real-time performance. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning-based approach, which leads to slow and non-optimal performance. In this project, we use a completely deep learning-based approach to solve the problem of object detection in an end-to-end fashion. The network is trained on the most challenging publicly available dataset (PASCAL VOC), on which a object detection challenge is conducted annually. The resulting system is fast and accurate, thus aiding those applications which require object detection. We present YOLO, a new approach to object detection. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is single network, it can be optimized end-to-end directly

on detection performance.
The structural element is an important part of mathematical morphology. Defined as a pixel configuration where the source is defined. A morphological filter that is used to measure each pixel of an image is used using a structural element. If the base of the editing element corresponds to a given pixel and all pixels that are opposite the structural element are included in the set. The editing feature has a simple shape like a square, a circle etc. Erosion and elasticity are two important morphological users. Erosion replaces the current pixel with the smallest pixel found in the set pixel set there, as Dilation replaces the current pixel value with the higher pixel value found in the specified pixel set. Since a binary image contains only black (0) and white (255) pixels, each pixel is replaced by a white or black pixel. Erosion and elastic filters are used in the cascade to produce other filters which are open and closed filters etc.

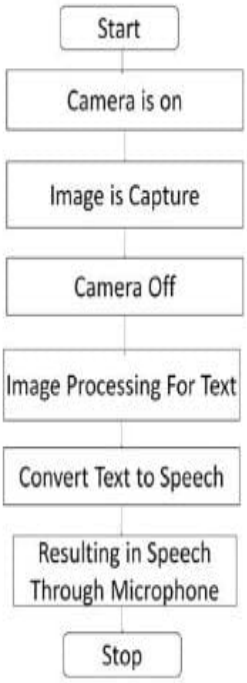

Fig. Block diagram of proposed work

The model is provided with two types of model training data. Image data is entered into the model, which reads and predicts the output accordingly. Another type of data is label data, which is given at the end of the model to compare with the predicted output.
1. Convolution with 64 different filters in size 3 * 3.
2. Multiplication of 2 (collection is usually done).
3. Maximum compound by 2.
4. Convolution with 256 different filters in size.
5. Maximum compound by 2.

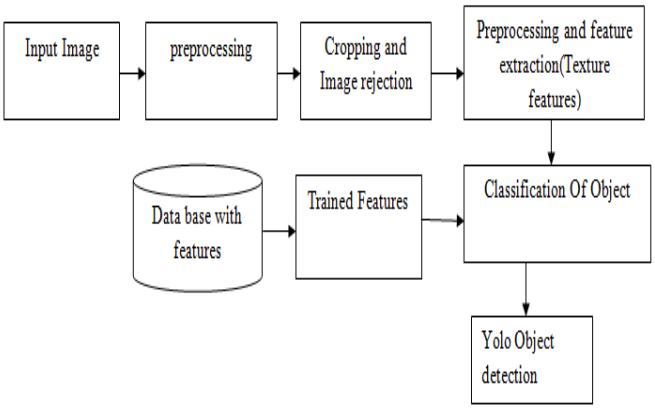

Fig. Data flow diagram

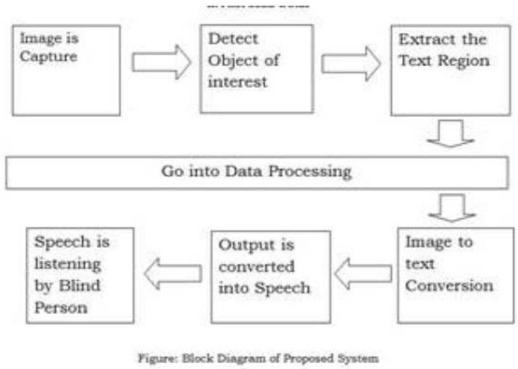

Figure: Block Diagram of Proposed System

- **Techniques for Object Recognition**:

1. **Template Matching**: It's is used to recognize small parts of an image which is then match a template image. It is a simple and straight-forward process**.**
2. **Colour Based:** Colour based object detection is also significantly used and it provide simple to implement method.
3. **Shape Based**: Shape detection provides great importance in object detection or reorganization they recognize the object diagrammatically.

There are 8 best algorithms for object detection:

1. **Fast R-CNN**

Fast R-CNN architecture an input image and multiple regions of interest (RoIs) are input into a fully convolutional network each RoI is poled into a fixed-sized feature map & then mapped to a feature vector by fully connected FCS. The network has two output vectors pet RoI softmax, probabilities and per-class bounding box regression offsets.

2. **Faster R-CNN**

Faster R-CNN is an object detection algorithm that is similar to R-CNN. Faster R-CNN is a single, unified network for object detection. The RPN module serves as the attention of this unified network RPN trained for region proposal task. Train fast R-CNN using the trained RPN use the detector network to realize RPN training fine-tune layer unique to RPN network.

3. **Histogram of Oriented Gradients (HOG)**

The histogram of oriented gradients (HOG) is a feature definition used to obtain objects in image processing and other computer recognition techniques. Histogram oriented gradients descriptor technique incorporates the appearance of gradient shapes in local parts of the image, such as the discovery window, interest region (ROI), among others. One advantage of features like HOG is their simplicity, and it's easy to understand the information you need to carry.

4. Regional Convolutional Neural Networks (RCNN)

The Regional-based Convolutional Network (RCNN) approach is a combination of regional proposals with Convolution Neural Networks (CNNs). R-CNN helps to do local things with a deep network and trains a high-level model with only a small amount of data to find annotations. It achieves excellent object acquisition accuracy by using in-depth ConvNet to separate object suggestions. R-CNN has the ability to block thousands of classes of objects without using limited methods, including hashing.

5. Single Shot Detector (SSD)

Shot Detector (SSD) is a way of finding objects in images using a single deep neural network. The SSD method divides the output space of the binding boxes into a set of automatic boxes over a limited feature. After selecting, the method estimates each location of the map element. The Single Shot Detector network includes predictions from multi-featured maps with different configurations to naturally capture objects of various sizes.

6. Spatial Pyramid Pooling (SPP-net)

Spatial Pyramid Pooling (SPP-net) is a network structure that can produce fixed representation regardless of image size / size. Combining pyramids is said to be powerful in object design, and SPP-net develops all CNN-based image classification methods. Using SPP-net, researchers can calculate feature map features from the whole image only once, and then combine features in specific regions (images below) to produce presentations of fixed length training for finders. This approach avoids making the computer repeatedly convolutional features.

7.YOLO (You Only Look Once)

YOLO is a shortened form of "You Only Look Once", and it uses convolutional neural networks for object detection. YOLO can detect multiple objects on a single image it means that YOLO applies a single neural network to the whole image. This neural network divides image regions and produces probabilities for every region after that YOLO predicts number of bounding boxes that cover some region on the image and chooses the best ones according to their probabilities.

8. R-CNN has test time per image 50secounds, speed up of 1x and mAP(voc-2007) is 66.0%. Fast R-CNN is having test time per image as 2secounds, speed up of 25x and mAP(voc-2007) is 66.9%. Faster R-CNN has test time per image 0.2secounds, speed up is 250x and mAP(voc-2007) is 66.9%.

IV.EXPERIMENTAL SETUP

This object detection technique is used to detect object convert it to text and then into speech. All the captured are detected. It is also used to identify individual object in image. But fail to detect specific object from image its difficult and long process.

In the present approach it has been attempted to capture a image with help of camera the detected image is names as its actual name and then convert it to speech. It is android application.

Android does not allow linking of other languages. However, all other computer vision techniques are implemented in OpenCv. OpenCv functions can be called from android. After the application is developed it can be ported from laptop to mobile using USB chord. A mobile device driver is to be installed on android mobile to make it compatible with laptops.

V.RESULTS AND DISCUSSION

To upgrade the proposed Android application, OpenCV3 Library software tools are used. The app is installed and tested on any Smartphones.

Figure 1 shows the app Icon User Can Identify an Object with a location feature. To improve this application, we have used Java Planning Language. However, all other computer recognition techniques are used in OpenCV. OpenCV function can be called from Android. Once the system is upgraded it can drive from Android. Once the system has been upgraded it can be moved from the laptop to the cell phone using a USB chord. The mobile device driver will be installed on Android Mobile to make it compatible with laptops.

The SSD does not use the Transfer Suggested Network instead, it resolves in a very simple way. Spot Both location and class points use a small conversion filter. After removing the map features, the SSD uses 3 * 3 convolution filters in each cell to perform Prediction.

Fig.1 Application Icon

Result: Object Detection is a computer vision system that works to identify and locate objects within an image or video. Specifically, the discovery of an object draws a binding box next to these finds, which allows us to find it.
where said objects are in a given sence. Image detection assigns a lable to an image.

A picture of a Dog receives the label “Chair”. A picture of two Chairs, still receives the label “Chair”. On the Other hand, draws a box around each dog and labels the box “Chair”. The Model predicts Where each Objects is and what labels Should be applied. In that way, Object detection provides more Information about an image than recognition.

Fig.2 Detected image

Here is an example of how this difference looks in performance: In the current study object detection is made on Android using that yolo algorithm. The Binary image found in the color image is provided as an input to the open morphological function and closure. Front and rear objects are classified based on their strength values. After image splitting, all the main objects are removed from the scene using contour-based reading techniques.

YOLO (Watch Only Once) is an object recognition program that is very different from the regional based algorithms seen above. At YOLO one Convolutional network predicts connecting boxes and class opportunities in these boxes.

Illustration: Is a simple photo where a chair is taken in the backyard on the plain. Limits have been found. After Obtaining the Name of the Object is displayed above the Borders. After image splitting, all the main objects are removed from the scene using contour-based reading techniques.

Fig.3 Detected image

The below picture is of detected picture of apple and banana hence this application can also detect a multiple-objects at the same time. Again, image of tea cup and cell phone is detected in below images. After detection it will show the name of that particular object and covert it to speech. We can hear the sound using speaker as this is android app it can be open in all android phones. It is very useful for blind people they can easily use it with help of smart phones.

Some experimental images of following Application:

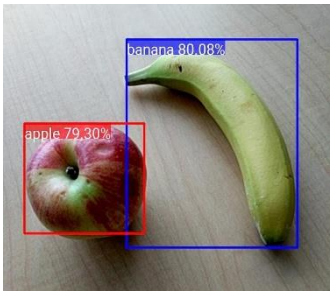

Fig.4 Detected image

The image detected is of fruits i.e, apple and babana.

Fig.5 Detected image

The image detected is of cell phone

Fig.6 Detected image

The image detected is of cup.

Fig. multiple object detection

VI.CONCLUSION AND FUTURE WORK

To solve the common problem of blind consumer testing, we have developed a camera-based product learning framework to help blind people learn product information in hand-held objects in their daily lives. This project helps blind people to see a product in the market place. A single neural network predicts binding boxes and class possibilities directly in full images in a single test. Since every acquired pipeline is a single network, it can be upgraded from one end to the acquisition function.

The Future of Object Recovery: This project defines object acquisition, one of the most complex applications available in recent years. Object discovery in photography and video has received a lot of attention from computer vision and pattern recognition communities in recent years. We have made great strides in the field, processing one image used to take 20 seconds per image and today it takes less than 20 milliseconds.

Download image to process text conversion. For a more complex background give the correct and correct output. Open World Learning and Practical Vision. Discovery of multiple models. Pixel Level Recovery (Segment) and Backgrounds. Sophisticated optical filtering and Real Improvement Dislikes for taxpayers we see. Traffic data collection and Medical imaging and Increasing industrial applications and spatial testing and detection of fingerprints and retina detection. Stereography Development. Processing a morphological image.

References

1. *CFar es Jalled, Moscow Institute of Physics & Technology, Department of Radio Engineering & Cybernetics Ilia Voronkov, Moscow Institute of Physics & Technology, Department of Radio Engineering & Cybernetic.*
2. *https://citlprojects.com/ieee-based-python-projects/image-processing-video-processing/object-detection-convert-object-name*
3. *Sandeep Kumar, Aman Balyan, Manvi Chawla Computer Science & Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi, India.*
4. *Jason Brownlee's Gentle Introductory Introductory Introduction on May 22, 2019 in Deep Learning for Computer Vision.*
5. *Step-by-step presentation of Basic Object Detection Algorithms Pulkit Sharma - October 11, 2018.*
6. *www.upgrad.com/blog/ultimate-guide-to-object-detection-using-deep-learning.*