IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

Real Time Object Detection for Blind People Using Machine Learning

Shreeyash Pitke[1], Shubham Chitte[2], Prof. S. P. Deshmukh[3]
Student, Department of E&TC, NBN SINHGAD School of Engineering, Pune, India[1,2]
Associate Professor, Department of E&TC, NBN SINHGAD School of Engineering, Pune, India[3]

Abstract: *There are number of blind people in the society, who are suffering while exercising the basic things of daily life and that could put lives at risk while travelling. There is a necessity these days to provide security and safety to blind people. There have been few devices designed so far to help the blind. Blindness or visual impairment is a condition that affects many people around the world. The usage of the blind navigation system is very less and is not efficient. The blind traveler is dependent on other guide like white cane, information given by the people, trained dogs etc. So, here we are proposing a self-assistance system for blind people, which will be able to convey the person about direction and type of obstacle across his path.*

Keywords: CNN Algorithm, Raspberry Pi, Data Base Images or Videos, etc.

I. INTRODUCTION

Blindness, as well as low vision, are conditions where people have a decreased ability to see and visualize the outside world. This reduces their mobility and productivity in completing daily tasks. Blind people usually dependon experience, smart sticks or some other people to help them in walking and avoiding obstacles. They do not havea sense of sight which makes them highly dependent on their memory. Also, they cannot be aware of sudden changes in the surroundings which makes it almost impossible to react to an instantaneous situation. Understandingany of the visual aspects like color, orientation and depth of an object is not easy. Comprehending a three- dimensional object in a single go requires more time and effort than otherwise. However, in the recent past, technology has made much advancement for visually impaired human beings. Hands free devices work completelyon the audio input of users.

They do not require any visual or touch interaction which works as a boon for them. There are screen readers to help them read the screens on devices. However, these devices are not enough to make the personal and professional life of sight impaired people easy. They only take audio input and when users want to understand the images of their surroundings or texts, these are not very helpful. Research is still going on how to make mobility as easy as possible without any hurdle or danger on the road. It can be equally helpful indoors where it is effortless to get a rough idea of the place around and search for any item. Hence, here a solution is proposed on how to make them confident while travelling or otherwise. Object identification and detection are done using a camera as well as in a web application. As soon as the application launches, it starts capturing a live objects stream as an input from the camera. The objects are detected in the frame of the camera along with its approximate position which is conveyed to users via audio output. This will help users gauge the location of the object and the direction he or she should be moving in.

II. BRIEF LITERATURE SURVEY

[1] Real-Time Objects Recognition Approach for Assisting Blind People. [Jamal S. Zraqou Wissam M. Alkhadourand Mohammad Z. Siam, Multimedia Systems Department, Electrical Engineering Department, Isra University, Amman-Jordan Accepted 30 Jan 2017, Available online 31 Jan 2017, Vol.7, No.1] Blind assistance is promoting a widely challenge in computer vision such as navigation and path finding. In this paper, two cameras placed on blindperson's glasses, GPS free service, and ultra-sonic sensor are employed to provide. the necessary information about the surrounding environment.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. The two cameras are necessary to generate the depth by creating the disparity map of the scene, GPS service is used to create groups of objects based on their locations, and the sensor is used to detect any obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. The proposed method for the blindaims at expanding possibilities to people with vision loss to achieve their full potential. The experimental results reveal the performance of the proposed work in about real time system.

[2] Object Detection Combining Recognition and Segmentation [Fudan University, Shanghai, PRC, yfshen@fudan.edu.cn University of Pennsylvania,3330 Walnut Street, Philadelphia, PA19104 Liming Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen.] We develop an object detection method combining top-down recognition with bottom-up image segmentation. There are two main steps in this method: a hypothesis generation step and a verificationstep. In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust toobject deformation and background clutter.
The improved Shape Context is used to generate a set of hypotheses of objectlocations and figure ground masks, which have high recall and low precision rate. In the verification step, we first compute a set of feasible segmentations that are consistent with top-down object hypotheses, then we propose a False Positive Pruning (FPP) procedure to prune out false positives. We exploit the fact that false positive regions typically do not align with any feasible image segmentation. Experiments show that this simple framework is capable of achieving both high recall and high precision with only a few positive training examples and that this method can be generalized to many object4 - Microsoft COCO Common Objects in Context.

Seyed Yahya Nikouei et.al [3] Human objects detection, behavior recognition and prediction in smart surveillance fall into that category, where a transition of a huge volume of video streaming data can take valuable time and place heavy pressure on communication networks. It is widely recognized that video processing and object detection are computing intensive and too expensive to be handled by resource-limited edge devices.
Inspired by the depth wise separable convolution and Single Shot Multi-Box Detector (SSD), a lightweight Convolutional Neural Network (L-CNN) is introduced in this paper. By narrowing down the classifier's searching space to focus on human objects in surveillance video frames, the proposed L-CNN algorithm is able to detect pedestrians with an affordable computation workload to an edge device. A prototype has been implemented on an edge node (Raspberry PI 3) using open CV libraries, and satisfactory performance is achieved using real-world surveillance video streams.

Mohannad Farag et.al [4] in this work the movement of SCARA robot is guided by deep learning-based object detectionfor grasp task and edge detection-based position measurement for place task. Deep Convolutional Neural Network (CNN) model, called KSSnet, is developed for object detection based on CNN Alexnet using transfer learning approach. SCARA training dataset with 4000 images of two object categories associated with 20 different positions is created and labeled to train KSSnet model. The position of the detected object is included in prediction result at the output classification layer. This method achieved the state-of-the-art results at 100% precision of object detection, 100% accuracy for robotic positioning and 100% successful real-time robotic grasping within 0.38 seconds as detection time.

Edward Rzaev et .al [5] In this review Object detection is one of the most active research and application areas of neural networks. In this article we combine FPGA and neural networks technologies to solve the real-time object recognition problem. The article discusses the integration of the YOLOv3 neural network on the DE10-Nano FPGA. Slightly worseindicators of the main metrics (mAP, FPS, inference time) when operating a neural network on a De10-Nano board in comparison with more expensive solutions based on GPUs, are offset by

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

differences in the cost and dimensions of the FPGA board used. Based on the results of the study of various methods for converting neural networks to FPGA.

Shaoqing Ren et. al [6] In this work, we introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectless scores at each position. The RPN is trained end-to- end to generate high-quality region proposals, which are used by Fast R-CNN for detection. We further merge RPN andFast R-CNN into a single network by sharing their convolutional features using the recently popular terminology of neural networks with 'attention' mechanisms, the RPN component tells the unified network where to look. For the very deep VGG-16 model.

Di Guo et,al [7] Grasp an object from a stack of objects in real-time is still a challenge in robotics. This requires the robotto have the ability of both fast object discovery and grasp detection: a target object should be picked out from the stack first and then a proper grasp configuration is applied to grasp the object. In this paper, we propose a shared convolutional neural network (CNN) which can simultaneously implement these two tasks in real-time. The processing speed of the model is about 100 frames per second on a GPU which largely satisfies the requirement. Meanwhile, we also establish alabeled RGBD dataset which contains scenes of stacked objects for robotic grasping.

Chandan G, et. al [8] in this review Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since past few years. Some of the popular object detection algorithms are Region-based Convolutional Neural Networks (RCNN), Faster- RCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance.

III. PROPOSED METHODOLOGY

1. Introduction

Object recognition is a kind of simple process for human beings but for computers it is not that easy task as it consists ofa step-by-step process of recognizing, identifying, and locating the objects with input with a given degree of precision. Recognition basically consists of classification and detection. Objects can be divided into their respective classes by performing three steps-feature extraction, localization, and classification on the objects. In classification, the algorithmrecognizes the class of the object with a degree of confidence. After classification, we know that the particular class of the objects from which this object belongs. Now, in detection, we put a bounding box around the object in the picture.

2. Block Diagram

The proposed method is to help blind person in detecting the obstacle in front of them i.e., car, person, traffic sign etc. as a camera based assistive object detection technique. The implemented idea involves person, car and traffic sign detection from image taken by camera and produced sound after detection of object.

3. Input Image

Camera captures image of note or object and is feed as input to the Python for further processing.

4. Pre-processing

Pre-processing images commonly involves removing low-frequency background noise, normalizing the intensity of the individual particle's images, removing reflections, and masking portions of images. Image

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

pre-processing is the techniqueof enhancing data images prior to computational processing.

5. Feature Extraction

Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. We are using Wavelet transform to extract features like RMS value, average, entropy of image.

6. Classification

Image classification refers to the task of extracting information classes from two or many class of image. Features extracted by wavelet transform and by using ultraviolet rays are feed to classifier so that classifier, here CNN algorithm,should be able to classify the note and detect object/obstacle.

IV. METHDOLOGY

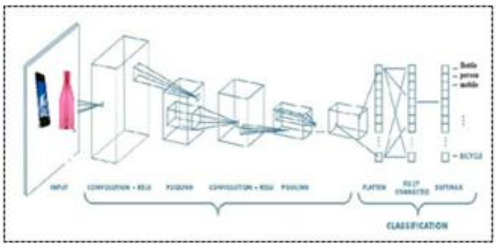

Figure 4.1: CNN Algorithm

CNN is a type of neural network model which allows us to extract higher representations for the image content. Unlike the classical image recognition where you define the image features yourself, CNN takes the image's raw pixel data, trains the model, then extracts the features automatically for better classification.

1) Principles of CNN:

Convolution:

A convolution sweeps the window through images then calculates its input and filter dot product pixel values. This allowsconvolution to emphasize the relevant features.

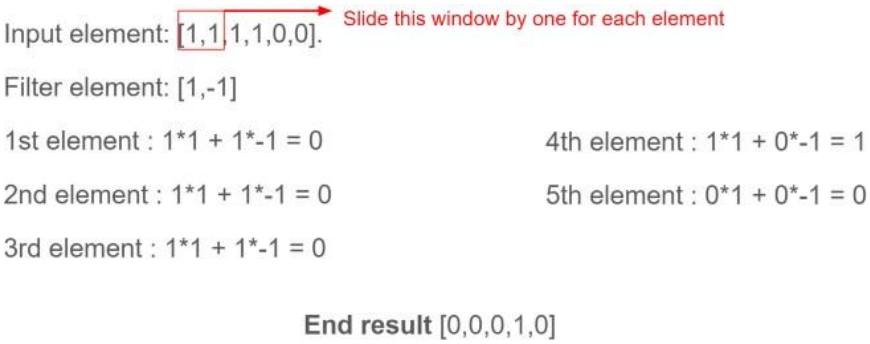

Figure 4.2: Convolution Operation with features(filter).

Look at this input. We will encase the window elements with a small window, dot multiplies it with the filter elements,and save the output. We will repeat each operation to derive 5 output elements as [0,0,0,1,0]. From this output, we can know that the feature change (1 becomes 0) in sequence 4. The filter has done well to identify the input values. Similarly,this happened for 2D Convolutions as well.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

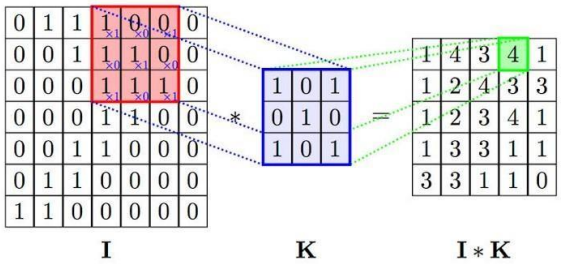

Figure 4.3: D Convolution Operation with features(filter)

With this computation, you detect a particular feature from the input image and produce feature maps (convolved features) which emphasizes the important features. These convolved features will always change depending on the filter values affected by the gradient descent to minimize prediction loss. Furthermore, the more filters deployed, the more features that CNN will extract. This allows more features found but with the cost of more training time. There is a sweetspot for the number of layers, usually, I will put 6 for 150 x 150 size of image.

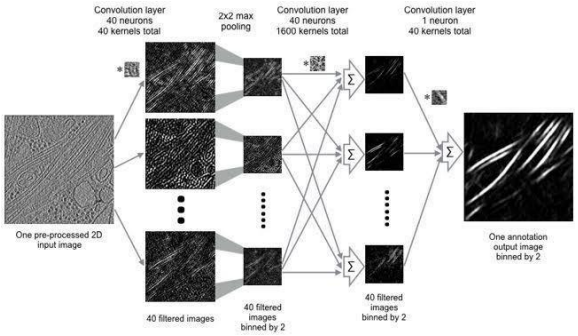

Figure 4.4: Feature map in each layer of CNN

However, what about the corner or side values. They do not have enough adjacent blocks to fit the filter. Should we remove them? No, because you would lose important information. Therefore, what you want to do instead is padding; you pad the adjacent feature map output with 0. By inserting 0 to its adjacent, you no longer need to exclude these pixels. Essentially, these convolution layers promote weight sharing to examine pixels in kernels and develop visual context to classify images. Unlike Neural Network (NN) where the weights are independent, CNN's weights areattached to the neighboring pixels to extract features in every part of the image.

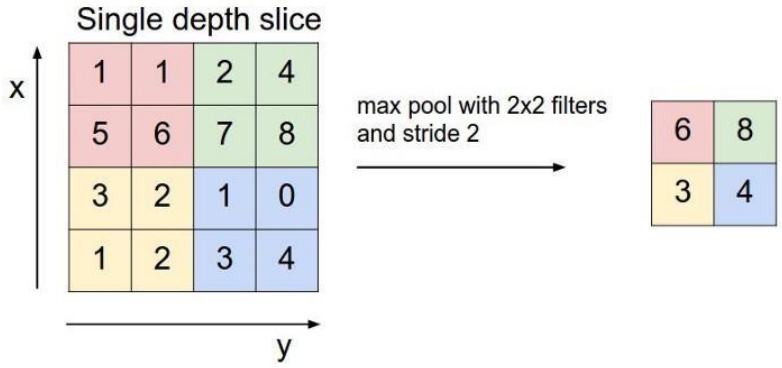

Figure 4.4: We take the maximum max pooling slices of each 2x2 filtered areas

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

CNN uses max pooling to replace output with a max summary to reduce data size and processing time. This allows youto determine features that produce the highest impact and reduces the risk of overfitting. Max pooling takes two hyperparameters: stride and size. The stride will determine the skip of value pools while the size will determine how bigthe value pools in every skip.

2) The CNN Big Picture + Fully Connected Layer

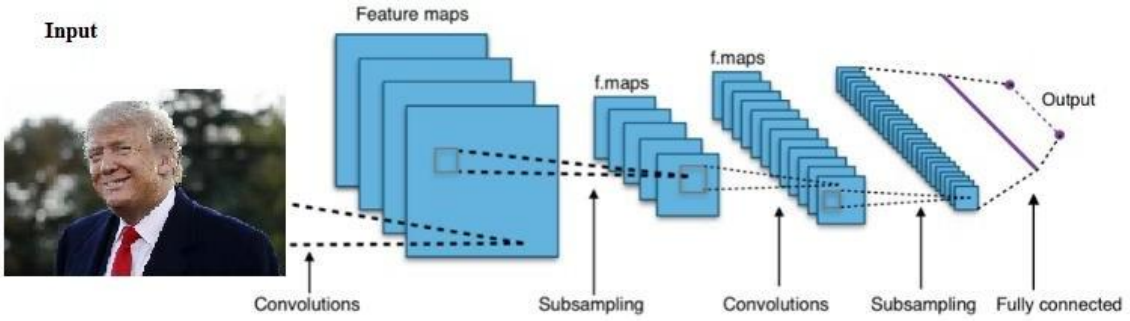

Figure 4.5: CNN architectures with convolutions, pooling (subsampling), and fully connected layers for SoftMaxactivation function.

Finally, we will serve the convolutional and max pooling feature map outputs with Fully Connected Layer (FCL). We flatten the feature outputs to column vector and feed-forward it to FCL. We wrap our features with SoftMax activation function which assign decimal probabilities for each possible label which add up to 1.0. Every node in the previous layeris connected to the last layer and represents which distinct label to output.

3) Using test set as the validation set to test the model

Even though we do not use the test set to train the model, the model could adjust the loss function with the test set. Thiswill base the training on the test dataset and is a common cause of over fitting. Therefore, during the training, we need touse validation sets then ultimately test the finished model with the unseen test set.

Dataset is relatively small

When dataset is small, it is very easy to specialize onto a few set of rules and forget to generalize. For example, if your model only sees boots as shoes, then the next time you show high heels, it would not recognize them as shoes. Therefore,in the case of small training data set, you need to artificially boost the diversity and number of training examples. One way of doing this is to add image augmentations and creating new variants. These include translating images and creatingdimension changes such as zoom, crop, flips, etc.

4) Over Memorization

Too many neurons, layers, and training epochs promote memorization and inhibit generalize. The more you train your model, the more likely it becomes too specialized. To counter this, you could reduce the complexity by removing a fewhidden layers and neurons per layer.

Alternatively, you could also use regularization techniques such as Dropout to remove activation unit in every gradient step training. Each epoch training deactivates different neurons. Since the number of gradient steps is usually high, all neurons will averagely have same occurrences for dropout. Intuitively, the more you drop out, the less likely our model memorizes.

V. RESULT AND DISCUSSION

This project's output shows the detected objects with a rectangle box around them and a label on top showing the object'sname and thus the accuracy with which it was detected. It can reliably extract any number of items present during a singlepicture.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

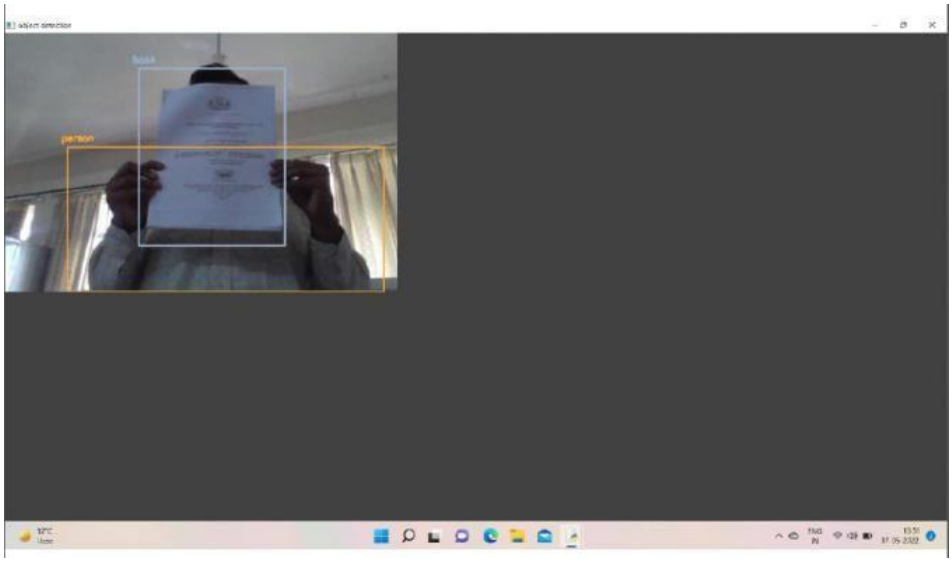

Figure 5.1: This result page detected book with person

Figure 5.2: This result page detected person with cell phone

Figure 5.3: This result page detected currency with their names

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

Figure 5.4: This result page detected bottle, person & laptop with their names in Bounding boxes

Figure 5.5: This result page detected person in their bounding boxes with their name

In the above images, such as the result of our design system "Real time object detection for blind people," we have observed that we have collected various output results by giving a real-time object and there are 99.6 percent of detected objects with their specific names, such as: person, person with cell phone detected, bottle, currency, bottleWe achieve 99.6 percent accuracy in detecting real-time objects by using CNN classifier. Object detection is a computer vision technology that locates items in pictures or movies.

VI. CONFUSION MATRIX

Accuracy		Precision	Recall	F1score
Class 1	99.6	0.99	1	1
Class 2	99.6	1	0.99	1

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 8, June 2022
Impact Factor: 6.252

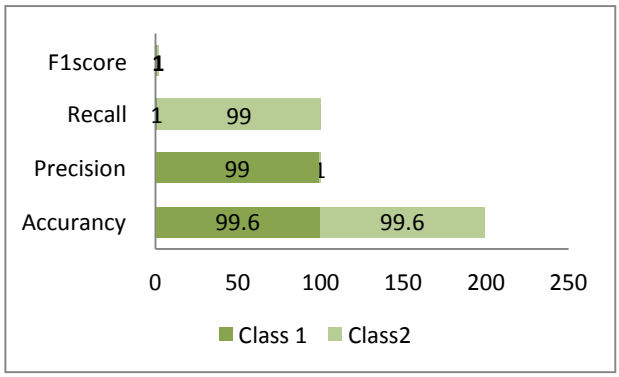

Figure 5.6: Graphical Representation of confusion matrix

Above confusion matrix it will be show that for class 1 real time object image total no 0302 images will be given As database it will be show Result with Accurate detected 302 images and Also it gives a not labeled images with totalno of 200 images to detected The total no 2 images will be rejected so overall Accuracy of our system is 99.6% A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

VII. CONCLUSION / FUTURE WORK

Blind people are weak person due to their disabilities. They face several problems in their life, one of these problems that is the most important one is detection the obstacles. So, we have proposed the system using which we can clearly identified the obstacle on the path as well as the currency using the implemented algorithm. The result shows the configurable accuracy of the system then the existing system. The system has several advantages like Accuracy 99.6 %of the system is very high. System complex city is very low, Processing time is less. In future we tried to add staircase detection and object recognition as well as fake note detection to make it complete system. Due to covid-pandemic, it will be helpful to blind person to maintain the safe distance.