ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

A Novel Approach for Assisting the Visually Impaired to Recognize Objects in an Image

Sushma M P[1], Girija J[2]

Student, Department of Computer Engineering, Bangalore Institute of Technology, Bengaluru, India[1]

Associate Professor, Department of Computer Engineering, Bangalore Institute of Technology, Bengaluru, India[2]

ABSTRACT: Text and speech is the main medium for human communication. A person needs vision to access the information from the image. This paper proposes an image based assistive text reading to help visually impaired person in reading both text and objects present in the image. In this system the objects in the image is identified by the faster Regional Convolutional Neural Network (faster RCNN) object recognition algorithm. The proposed idea involves text identification from image using pytesseract in python and generates audio description about the objects in the image. This is a prototype for visually impaired people to recognize the products in real world by extracting the objects on Image and converting it into speech. Proposed method is carried out by using Deep Learning methods, python and OpenCV. This technology helps millions of people in the world who experience a significant loss of vision.

KEYWORDS: Object Recognition, CNN, RPN, pytesseract, Detection Network, back-propagation

I. INTRODUCTION

The environment around us is complex so we need several sensors such as vision, touch, smell etc. to survive in this world. All the creatures on Earth have a set of such sensors which help them searching food, water, and safety etc. Among all the sensors vision is critically important sensor because it give accurate and complex representation of the environment which can be processed to get valuable information The main advantages of sensor vision is compare to other are its large and wide range, ability to provide complex data which can be processed to extract information such as object color, shape etc. The most of the significant difficulties for visually impaired person is to read. Existing systems in computers and availability of digital cameras made it feasible to assist the visual impaired person by developing the camera based applications that combine computer vision tools with Optical Character Recognition(OCR) system.

This paper aims to build an efficient and accurate object recognition for an image. The idea involves using the object recognition algorithm faster Regional Convolutional Neural Network(faster RCNN) for object recognition and pytesseract for detecting the text in the image. The detected text is then converted to audio signals and to voice output. It is used for the visually impaired person for the daily activities purpose like shopping. This paper is organized as follows. Section II introduces the related work about the proposed system. Proposed methodology is then presented in section III. Implementations are explained in section IV. Section V concludes the paper.

II. RELATED WORK

Earliersystems helps to support the visually impaired persons. Most of the existing systems are built in MATLAB platforms. Algorithms used in earlier systems lack efficiency and accuracy.

The paper [1] presents a prototype for extracting text from images using Raspberry Pi. The images are captured using a web cam and are processed using Open CV and OTSU's algorithm. Initially the captured images are converted to gray scale color mode. The images are rescaled and cosine transformations are applied by setting vertical and horizontal ratio. After applying some morphological transformations OTSU's thresholding is applied to imageswhich is adaptive thresholding algorithm. After thresholding, contours for the images are generated using

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

special functions in Open CV. Using these contours, bounding boxes are drawn around the objects and text in the images. Using these drawn bounding boxes each and every character present in the image is extracted which is then applied to the OCR engine to recognize the text present in the image.

In [2], proposed a camera based assistive text reading framework to help visually impaired persons read text labels and product packaging from hand-held objects in daily life. The system proposes a motion based method to define a Region Of Interest (ROI), for isolating the object from untidy backgrounds or other surrounding objects in the camera vision. A mixture of-Gaussians-based background subtraction technique is used to extract moving object region. To acquire text details from the ROI, text localization and recognition are conducted. Then text regions from the object ROI are automatically focused. In an Adaboost model the gradient features of stroke orientations and distributions of edge pixels are carried out by Novel Text Localization algorithm. Text characters in localized text regions are binarized and recognized by off-the-shelf optical character identification software.

A bottom-up integration is performed in the information and merging pixels of similar stroke width into connected components to detect the text in natural scenes using SWT [3]. This allows to detect letters across a wide range of scales in the same image. Since it do not use a filter bank of a few discrete orientations, it can detect strokes (and, consequently, text lines) of any direction. This method carries enough information for accurate text segmentation and so a good mask is readily available for detected text. The need for integration over scales, orientations of the filter and, the inherent attenuation to horizontal texts are the limitations of this method. The linear features which are used in remote sensing and medical imaging domains are related with the definition of stroke. In road detection, the range of road widths in an aerial or satellite photo is known and limited, whereas texts appearing in natural image can vary in scale drastically. Additionally, roads are typically elongated linear structures with low curvature, which is again not true for text.

In [4] describes the camera based text reading system for blind person. In this paper a binary image is created using global or local thresholding which can be decided from Fisher's Discriminant Rate (FDR). The technique is essentially based on OTSU's binarization method. It is an automatic threshold selection region based segmentation method. In this method when the characters are present on a frame, then-the local histogram has two peaks and this is reflected as a high value for the FDR. For quasi-uniform frames the value of the FDR is small and the histogram has only one peak. In the case of complex areas the histogram is dispersed resulting in higher FDR values, which are still lower than in the case of text areas. With a bimodal gray-level histogram the FDR is used to detect the image frames. When the image frames are of high FDR values, the local OTSU threshold is used for binarizing the image, frames with low FDR values.

III. PROPOSED METHODOLOGY

The proposed method is the prototype for the visually impaired person to recognize objects in the image. The method involves using the faster RCNN[8] object detection algorithm for object detection and pytesseract to detect text part of the image which is then converted into the voice output.

The proposed system is shown in figure 1 here first the training images are collected and they are trained for the object detection. The images are first passed through the resnet50 model where in this model it extracts the feature from the images using the filters and generates the feature maps for all the images. The feature maps generated from from the resnet50 is passed through the Regional Proposal Network (RPN) layer. The RPN layer gives the object proposals over the image. The detection network is used to generate final class labels and bounding box over the objects in the images. We train both the RPN and detection network and generate the weights for images and object detection. When the new test image is passed to system by using of the faster RCNN algorithm it identifies the objects in the images using of the weights previously calculated. This generates the bounding box over objects in the images this image is passed to the pytesseract module where it takes all the text part in the image and passes to the Google speech API where it generates the audio output for the objects detected.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

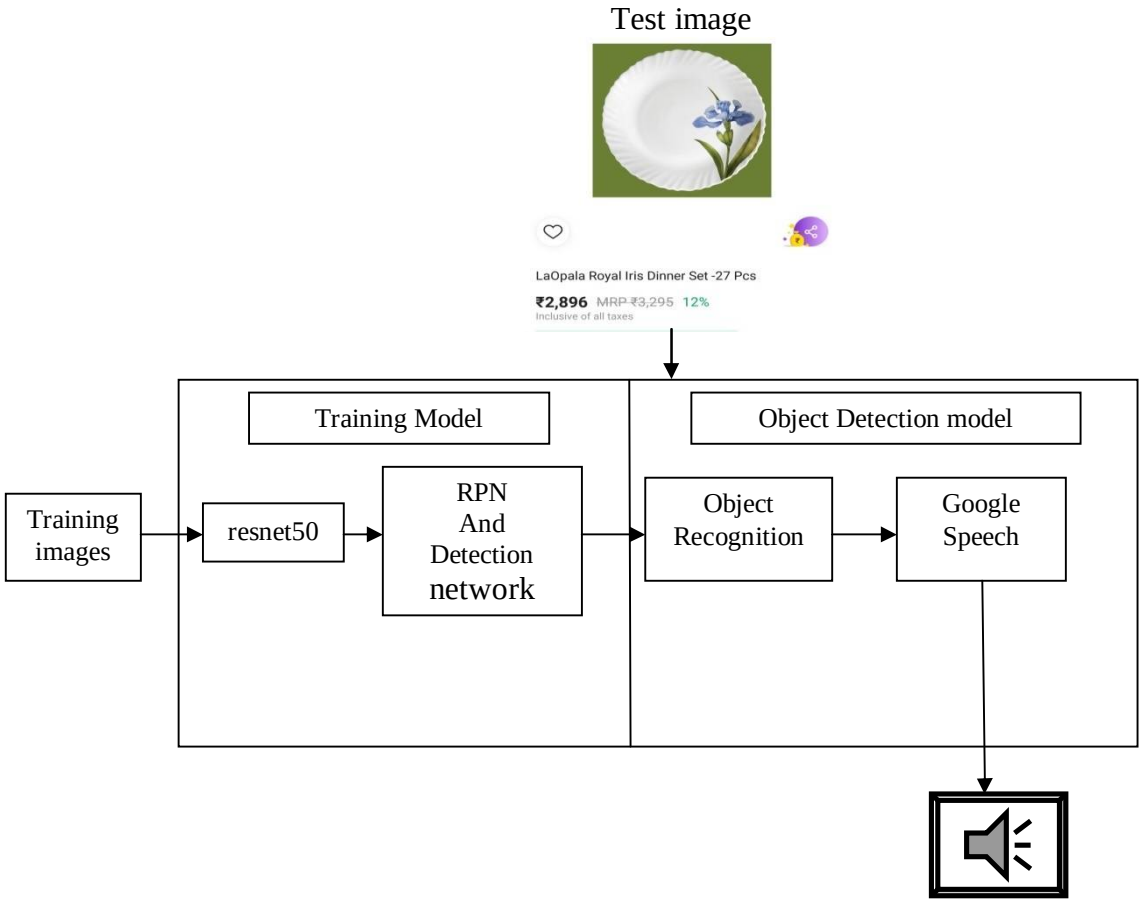

Fig. 1: System Architecture

IV. IMPLEMENTATION

As shown figure 1 in this system we have two models training model where we train the different images and calculate the weights and object detection detection module here we use object detection algorithms to detect the objects in the image. We need to train and test the RPN and Detection Networks on the training images for different scales. To train the RPN and the Detection network we apply the back-propagation algorithm[6] and follow the image-centric from [5] sampling strategy to train this network. For each image classifier loss and regression loss are calculated and total loss for all the images is calculated. We will train the images till the total loss reaches near to the zero and the weights calculated are saved for testing new images. When the new test image is passed we apply the faster RCNN algorithm where first the image is passed to the resnet50 model which generates the feature maps. Feature map generated is passed to the RPN where it gives the object propasals over the images using the weights calculated. Object Propasals generated over the image is passed through the Regional of Interest (ROI) layer where it uses the maxpooling strategy to give max pooled image. This maxpooled image is passed to the fully connected layer of the network which classifies the objects present in the image using softmax activation function. The bounding box for each object with class labels are generated this image is passed to the pytesseract which identifies the text in the image and passes to the Google speech API which generates the audio output for the image.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

V. RESULTS

We have collected different images of flower, dining plate ,bowl, pressure cooker etc to train the images which can be used to help the visually impaired person for shopping the items from the newspaper ads,online sites etc, and trained on these images.

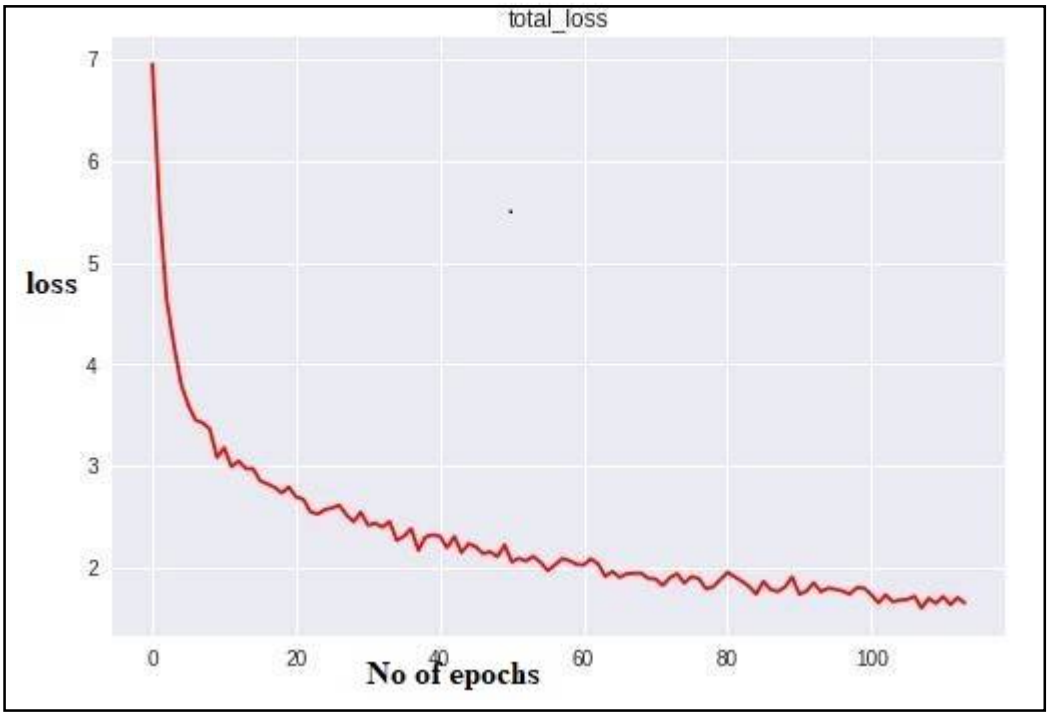

Fig. 2: A Graph showing the total loss calculation during the training model

The graph in figure 2 shows how the total loss calculated decreases as the the no of epochs is increased. The no of epochs is the number of times the each training data is trained in our system we have trained for the 110 times. First when we first started training the total loss was 7 it gradually decreased to 0.8. For each epoch trained the total loss is compared to previous loss if the current loss is less than the weights are updated in the neural networks, weights updated are saved. If the current loss is not decreased then weights are not updated.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

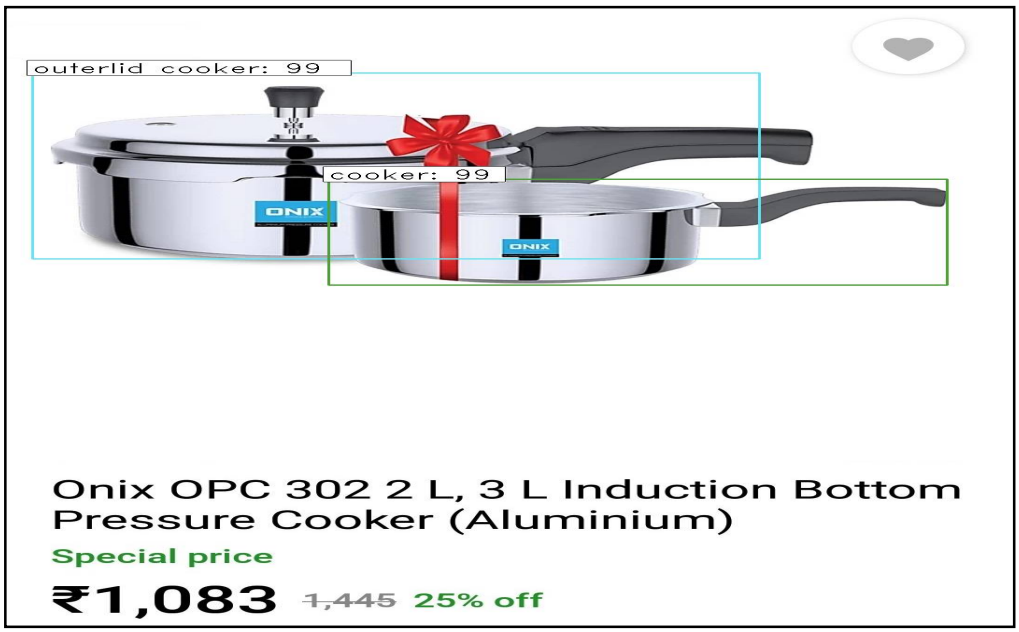

Fig. 3: An Example showing cooker objects detected

Fig. 4: An Example showing the dining plate object detected

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

In the figure 3 and 4 shows the objects detected in the image using the faster RCNN object detection algorithm, for each object the bounding box around that object is given with its label.This images are passed to the pytesseract which extracts text information and it passes it to the Google speech API which reads out the text data from the image.

VI. CONCLUSION

Thispaperpropose the system to help the visually impaired person to identify objects in an image. Our method enables a unified, deep-learning-based object detection system. In this system we use faster RCNN algorithm which is faster andaccurate when compared to the CNN, RCNN algorithms and small objects are detected accurately using this algorithm.

REFERENCES

1. Ms.Rupali, D Dharmale, Dr. P.V. Ingole, "Text Detection and Recognition with Speech Output for Visually Challenged Person", IJAIEM, Vol. 5, Issue. 1, pp.174-177, 2016.
2. Rajkumar N, Anand M.G, Barathiraja N, "Portable Camera Based Product Label Reading For Blind People",IJETT, Vol.10, pp.521-524 2014.
3. Boris Epshtein, EyalOfek, Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform", Microsoft Corporation. Pp.1-8.
4. Ezaki, Nobuo, et al. "Improved text-detection methods for a camera-based text reading system for visually impaired persons", Eighth International Conference on Document Analysis and Recognition (ICDAR'05) IEEE, 2005.
5. R. Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision (ICCV) IEEE, 2015.
6. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, Vol.1, Issue. 4, 1989.
7. J. Hosang, R .Benenson, P. Dollar, and B. Schiele, "What makes for effective detection proposals?" IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),Vol.38, Issue. 4, 2016.
8. ShaoqingRen, Kaiming He, Ross Girshick, Jian Sun, "Faster-RCNN: Towards Real-Time Object Detection with Regional Prposal Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, Issue. 6,2017.