Speech based Object Identification using Region Proposal Faster RCNN Algorithm

Samiksha Choyal, Ajay Kumar Singh

Abstract: ***This paper describes the applications of object detection and demonstrates the methodology along with the results of one of the object recognition models that is Faster Regions with Convolutional Neural Network (Faster R-CNN). This experiment of object detection has been conducted on a new proposed dataset of everyday objects. The implementation has been done with the help of tensorflow object detection models. The results obtained after testing and training the images with this model are depicted in the form of a graph. Further, the final output of the recognized object is shown to the user in the form of a speech.***

Index Terms: Convolutional neural network, Deep learning, Faster RCNN, Object Recognizition.

I. INTRODUCTION

Object detection is a part of computer vision whose objective is detecting instances of objects of classes which are certain in digital videos or images. The system helps in identifying what objects are present and the location of these objects in a given image. Since there are abundant changes in aspect, position, obstruction and also the lighting condition it is problematic to appropriately succeed in object detection. Therefore, it draws attention and has been a major research field [1]. Computer vision applications face various challenges while detection and classification of objects in real world like large variation in appearance [2]. Deep learning has attained excellent achievement in different tasks such as natural language processing, image classification and object detection. The basis of deep learning technologies is that humans do not design the feature layers rather general purpose procedure of learning is used to learn from the data [3].

The object recognition can also be summed up as the model or the systems which detect objects and require differentiating between objects into their object classes respectively [4]. After the search for identical features and whether the image contains the image given as input the output is displayed. CNN has bought momentous advancements in detection of objects over conventional methods or techniques. But the drawback of CNN, which led to introduction to new or advanced techniques, was when the object had multiple objects. The object detection based on CNN can be categorized into two, such as region proposal based algorithm like Region Convolutional Neural Network (RCNN), Fast RCNN and Faster RCNN and another one is a regression based algorithm that is You Only Look Once (YOLO) and Single Short Multibox Detector (SSD) [5].

This paper shows the implementation with a deep learning algorithm that is Faster RCNN on a dataset of images collected from Google images. The graphs of each trained classes have been recorded and displayed in the result section. Thus, an audio output is produced of the identified object.

II. OBJECT DETECTION APPLICATION

A. Biometric recognition: This technology uses physical or behavioral traits of human for recognition of an individual authentication and security. The identification of an individual is based upon distinguishable a biological feature that is hand geometry, fingerprints, iris and retina patterns, DNA, etc. are known as biometrics. For the analysis of biometric template matching an object recognition technique can be used.

B. Surveillance: The recognition and tracking of objects can be done for numerous video surveillance systems. The suspected people or vehicles can be tracked with the help of object recognition.

C. Industrial Inspection: The machinery parts recognition can be done with object recognition and then monitoring of working or damage could be observed.

D. Content-based image retrieval: CBIR can be referred when the recovery is dependent upon image content. The content based image retrieval and automatic keyword glossary is provided by ‘ontoPic’ which is based on supervised learning.

E. Robotic: An important issue for research in recent years is autonomous robots. One of the famous competitions is humanoid robot soccer. When there is unreliable and vital environment the soccer players which are robots rely heavily upon their vision systems. The various environment information is collected as terminal data by robots with the help of a vision system for finishing the functions such as robot tactic, localization, avoiding barriers. The computational effort can be decreased for recognizing the crucial objects in the field by features of objects that can be obtained by techniques of object recognition.

F. Medical analysis: The detection of tumor in MRI images, skin cancer is a few medical imaging examples for object detection.

G. Optical Character/digit recognition: The recognition techniques can recognize characters in scanned documents.

Revised Manuscript Received on December 22, 2018.

Samiksha Choyal, Department of Computer Science and Engineering, Mody University of Science and Technology, Lakshmangarh, India.

Ajay Kumar Singh, Department of Computer Science and Engineering, Mody University of Science and Technology, Lakshmangarh, India.

H. Human Computer interaction: The system can store human gestures for interacting with humans during real time and thus recognizing them. Any application on the phone or interactive games can be the system.

Intelligent vehicle systems: The detection and recognition of traffic sign, especially for detecting and tracking a vehicle these intelligent vehicle systems are required. In the phase of detection a scene can be scanned for the establishment of regions of interest with the segmentation method based on color. The Haar wavelet features that are acquired from AdaBoost training help in detecting the sign candidates in ROI. Then, for sign recognition SURF is applied. The local invariant features are found by SURF in signatures of candidate and then matching of these features takes place with the existing templates in the dataset. The maximum numbers of matches are found out in the template image and thus recognition is performed.

III. ALGORITHM USED

A network for detection of objects takes image as an input and the output is in the form of bounding boxes across the objects of interest [7]. Faster RCNN was proposed by Ren et al in 2016 [8]. The instinctive solution is the integration of CNN and the region proposal algorithm. The Faster RCNN fundamental structure is depicted below in figure 1 [9].

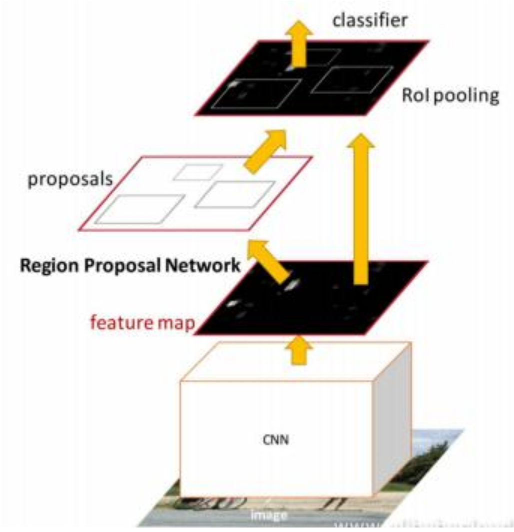

Fig. 1. Faster RCNN fundamental structure.

The division of Faster RCNN is split in four important elements that is Convolutional layer, Region proposal network, RoI Pooling and Classification. This algorithm uses Region Proposal Network which generate Regions of Interest (RoI) depending upon the input image. The three convolutional layers and a proposal layer which is a new layer sums up together to build RPN.

This algorithm can be understood in brief using 4 steps. First, an input image is passed to the Convolutional Network that will return the feature maps for the image. Second, get the object proposals by applying the RPN on the feature maps obtained in step 1. Third, all the proposals have to be of the same size so for this ROI pooling layer is applied. Fourth, the classification and prediction of the bounding boxes for the image is obtained by passing the proposals to fully connected layer [8].

IV. METHODOLOGY

The methodology used in the implementation of the Faster RCNN algorithm on everyday life objects. The steps described below were involved in performing the object detection. The basic structure of the object identification is presented in figure 2. The implementation steps are as follows:

A. Annotating Images: With the help of LabelImg the images or dataset collected is labeled.

B. Create Label map: In this mapping of each label is done to integer values. Both the processes, training as well as detection use this label map.

C. Create Tensorflow Records: In this the labeled images in xml format have to be converted to csv format. After this the conversion from csv format to record format is done.

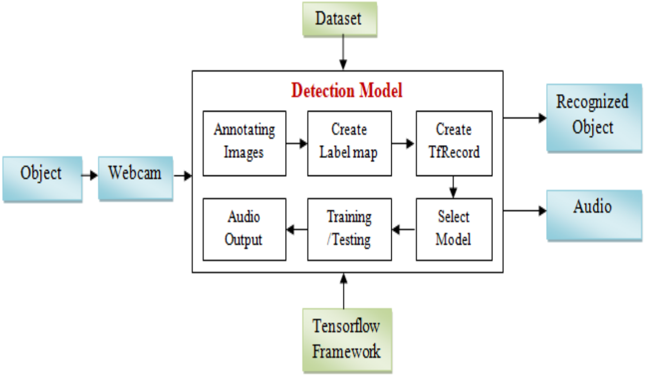

Fig. 2. Methodology used for object detection.

D. Selecting the Model: The model is selected with which training is to be done. Faster RCNN is the model used in this project.

E. Training and Testing: After the selection of the model the images are trained and then tested to know if the object identified is correct or not.

F. Text of the Recognized Object: The object that is recognized is stored in a text file which is then passed on for getting the output in the form of speech.

G. Audio Output: This is the last step which uses python text to speech conversion libraries.

V. RESULT AND DISCUSSION

The training and testing of this project was done for classes 4,10,15,20 and 25 objects. The images taken for training and testing were 90 and 10 percent of the total images respectively. The total loss is the loss that is recorded while training and mean average precision is the metric for measuring the accuracy. These two terms, graphs have been recorded with the help of TensorBoard for each interval that was trained. These graphs also help in knowing how much more training is needed or it is sufficient and need to be stopped. The Total Loss stabilizes quickly

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

because of the use of a pre-trained model. The results obtained after testing 25 classes are depicted with the help of graph below in figure 3 and 4 respectively.

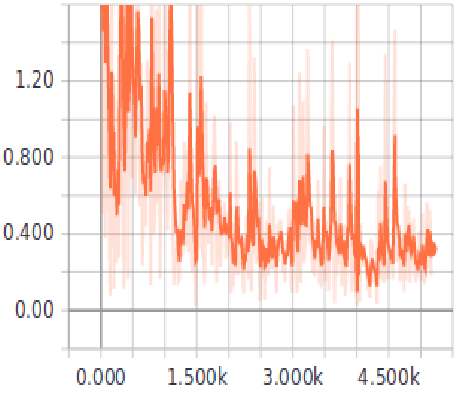

Fig.3. Total Loss for 25 classes.

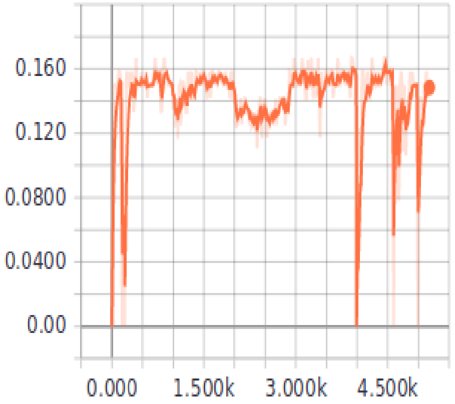

Fig. 4. Mean Average Precision for 25 classes.

The figure 5 and 6 shows the results obtained through webcam for multiple objects in real time.

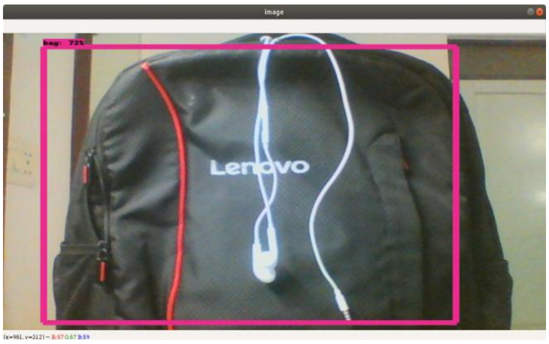

Fig. 5. Bag detected through webcam

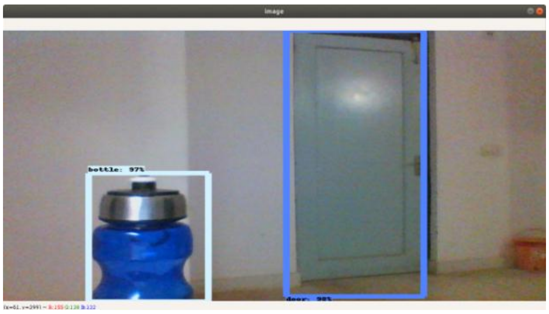

Fig. 6. Bottle and door detected through webcam.

The table 1 shows the results obtained in terms of total loss, mean average precision, fractions in each batch of 12 for different classes trained. Mean average precision is the average of the maximum precisions at different recall values.

Table 1. Comparison based on different parameters.

Classes Parameters	4	10	15	20	25
Mean Average Precision	0.124	0.145	0.15	0.148	0.147
Total Loss	0.14	0.25	0.18	0.03	0.34
Fraction in a batch	1.00	0.100	1.00	1.00	1.00

VI. CONCLUSION

This paper shows the implementation of the Faster RCNN algorithm. The graphs were obtained for mean average precision and total loss. With the increase in number of epochs the loss decreases gradually. The time taken to train the model is more as it is trained on a CPU system. The results could be improved with a better quality of the camera. The loss increases are the number of objects increases. The mean average precision is almost near 0.14 for all the classes. The batch fraction remains same for all the intervals. In future, more number of objects could be added and then the efficiency and performance could be improved and comparison with regression based algorithms would justify which algorithms are good for the recognizition purpose.

REFERENCES

1. Y. Ren, C. Zhu, and S. Xiao, “Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures,” Mathematical Problems in Engineering, vol. 2018 ,2018.
2. Singh, A.K., Shukla, V.P., Tiwari, S. and Biradar, S.R., “Wavelet based histogram of oriented gradients feature descriptors for classification of partially occluded objects”, International Journal of Intelligent Systems and Applications, 7(3), p.54, 2015.
3. B. Zhao, J. Feng, X. Wu, and S. Yan, “A survey on deep learning-based fine-grained object classification and semantic segmentation,” International Journal of Automation and Computing, vol. 14, no. 2, pp. 119–135, 2017.
4. K. U. Sharma and N. V Thakur, “A review and an approach for object detection in images,” Int. J. Computational Vision Robototics, vol. 7, no. 2, pp. 196–237, 2017.
5. J. Du, “Understanding of Object Detection Based on CNN Family and YOLO,” Journal of Physics Conference Series vol. 1004, no. 1, 2018.

6. R. Girshick, J. Donahue, T. Darrell, J. Malik, and UC Berkeley, “Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587, 2014.
7. M. Ave, A. Carpenter, and M. St, “Applying Faster R-CNN for Object DeSStection on Malaria Images”, pp. 56-61, 2017.
8. Q. Zhang, C. Wan, and M. Jiang, “Multiple Objects Detection based on Improved Faster R-CNN,” Proceedings of 9th International Conference on Signal Processing System - ICSPS 2017, pp. 99–103 ,2017.
9. J. Li, D.Zhang, J.Zhang, “Facial Expression Recognition with Faster R-CNN,” Procedia Computer Science, vol. 107, no. Icict, pp. 135–140 ,2017.

AUTHORS PROFILE

Samiksha Choyal 1980.is a Master of Technology student in Department of Computer Science and Engineering, Mody University of Science and Technology. She received her Bachelor of Technology (CSE) from Mody University in 2017. Her research interest include image processing, Deep Learning .

Ajay Kumar Singh was born in India, in 1980. He received his B.E. (Computer Sc. & Engineering) in 2001, M.Tech. (Information Technology) in 2006 and PhD. in Computer Sc. & Engg. from the MITS Lakshmangarh in 2015. He has joined as a Asst. Prof. in Mody Institute of Technology & Science, Deemed University Lakshmangarh in 2009.He has published over 23 papers in refereed journals and conference proceedings. His current research interest includes Image Processing, Image classification and their applications in computer vision.