Object Detection Classifier using TensorFlow

[1]Ajay Talele, [2]Varun Lale, [3]Jitesh Patil, [4]Tejal Meshram
[1]Guide, [2]Student, [3]Student, [4]Student
[1]Department of Electronics,
[1]Vishwakarma Institute of Technology, Pune, India

Abstract : Recently, engineers, scientists and developers are getting fascinated by the words- objects or things etc. These keywords are being used in various engineering fields. An object is nothing but an entity. The entity can be a living or a non-living thing. In the field of computer vision or image processing, whenever an image is needed to be processed on, it needs some information which classifies the different objects in an image with the help of their classes. The image may contain several objects, which can be of similar types or which may vary from each other by a long margin. In this paper, we discuss about detection of objects of certain classes (such as humans, cars or fruits) in digital images and videos and how it is done in real time. This paper also discusses about the different algorithms which are required for the detection of objects in the domain of computer vision and deep learning and choosing the correct algorithm for our application. We are taking the example of two to 3 types of fruits which are distinct from each other in terms of their overall visual and physical characteristics.

IndexTerms **- Objects, Computer Vision, Classifier, R-CNN, TensorFlow.**

I. INTRODUCTION

All of us are aware of the fact that the world around us is changing. New topics such as Machine Learning, Artificial Intelligence, Internet of Things and Computer Vision are now trending. The evolution of technologies is in full swing and we must be in accordance with it. In the current scenario, machines are being trained to work smartly and with less human intervention. Image processing is a technique which we use for the analysis and manipulation of an image so as to improve the quality and clarity of the image. The field of image processing is being developed such that multiple objects in the image can be detected and recognized as well.

When a typical image is taken into consideration, it may contain different types of objects such as cars, buildings, humans, fruits, vegetables, landscapes etc. These objects are evidently different from each other given their physical characteristics and aspects are quite distinct. In applications like security, manufacturing facilities, vehicle detection, people counting and face detection, object detection has become very important as each entity which needs to be manufactured, analyzed or is under surveillance, needs to be recognized as a distinct entity.

With advancements in the fields like Machine Learning and Computer Vision, the applications in the above mentioned fields are coming more and more into reality and are easier to develop than ever before. Object detection techniques are widely used in applications like security and manufacturing plants where a large number of data sets is required. The data set may contain a number of images which are processed and analyzed which results in detection of a recognizable object. The advanced techniques used in the current situation such as SIFT or HOG or CNN are making significant advancements in the field of performance day by day and thus are ready to play a role in real time applications. People tend to make a mistake while distinguishing between image classification and object detection.[1-3] When an image is needs to be classified into some category, image classification is done. However, whenever we need to know the location of objects or number of objects of a certain category/class in an image, we use object detection. There is, however, some overlap between these two scenarios. If we want to classify an image into a certain category, it could be a possibility that the object or the characteristics that are required to perform categorization are too small with comparison to the full image. In that case, we would achieve better performance with the help of object detection instead of image classification even if we are not interested in the exact location or number of objects.[4] With an image classification model, we can generate image features (through traditional or deep learning methods) of the full image. These features are the main essence of an image which is to be classified. These features contain the unique information which distinguishes other aspects of an image with the one we require for classification. We use feature extraction for a more conspicuous region-based classification of objects. In order to do that, we need a significant amount of data and it must be labelled to fit into a model.. In order to improve the model however, it is advised to experiment with different approaches. Images with their corresponding bounding box coordinates and labels can be termed as labelled data. That is, the bottom left and top right (x,y) coordinates + the class.

II. R-CNN AND FASTER R-CNN

A bounding box is very important in terms of the algorithms that are used for object detection. When we are trying to detect an object in the image, we will draw a bounding box around it. There could be more than just one bounding boxes in an image representing more than one objects. The major reason why we cannot continue with the problem of drawing bounding boxes by building a standard convolutional network followed by a fully connected layer is that, the length of the output layer is continuously changing and is not constant at any point of time, this occurs due to the uncertainty of the occurrences of the desired objects which may vary[5]. We can use a basic approach to solve this problem by taking different regions of interest from the image, and use a CNN to classify the presence of the object within that region. Another problem with this approach is that the objects of interest within the image might have different spatial locations within the image and different aspect ratios. Hence, we would have to select a huge number of regions and this could computationally be a huge problem as the number of computations would swell by a huge amount. Thus, some algorithms like R-CNN and YOLO were introduced to find these incidences in a very fast way.

2.1. R-CNN :

To bypass the problem of selecting a huge number of regions, we use selective search approach to extract just 2000 regions from the image and we call these regions as region proposals. Therefore we can just work with 2000 regions rather than taking chunk of regions which complicates the computations. The 2000 region proposals are generated using the selective search algorithm which is as follows : generating initial sub-segmentation, we generate many candidate regions. We can make use of greedy algorithm to gather similar regions of interest and show all of them as a larger region. These generated regions are shown as proposals for final candidate region proposals.[6-8] A 4096-dimensional feature vector is given as the output when the above mentioned candidate regions are warped in to a square and are provided to a convolutional neural network. The CNN acts as an extractor which extracts features and the output dense layer consists of the features extracted from the image and the extracted features are fed into a Support Vector Machine to classify the presence of the object within that candidate region proposal. The algorithm predicts four values which are offset values to increase the precision of the bounding box which is in addition to deciding the presence of the object. The problem with R-CNN is that it still takes a huge amount of time to train the network as we would have to classify 2000 region proposals per image.[9] Each test image takes around 47 seconds and so, it cannot be implemented in real time. The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage. Bad candidate region proposals are a result of such fixed algorithm.

2.2 .FASTER R-CNN :

The performance of the network is affected by selective search which is a slow and time-consuming process.[10] Faster R-CNN algorithm eliminates the selective search algorithm and lets the network learn the region proposals. A convolutional network is provided with an image as an input which gives convolutional feature map as the output which is our requirement. A separate network is used to predict the region proposals on the feature map. This network is different than the selective search algorithm. RoI pooling layer helps to reshape the predicted region proposals which also helps in classifying the image within the proposed region and also predicting the offset values for the bounding boxes. The region boxes can also be called as anchors. We have used Faster R-CNN approach in our implementation. Figure 1 shows how the faster R-CNN works in a sequential way. The feature maps play a huge role in Faster R-CNN.[11]

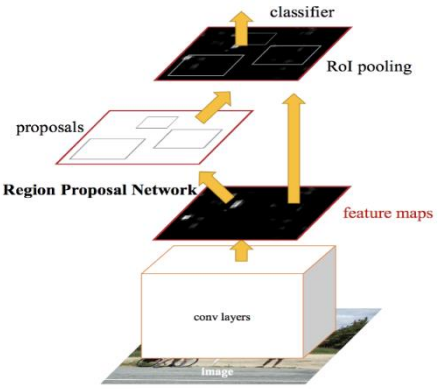

Fig. 1 Faster R-CNN

III. PROCEDURE

3.1 Coding :

Setting up the Object Detection directory structure by writing suitable Python codes. The Python codes are written for the Faster R-CNN implementation which includes feature extraction and mapping as well as Region Proposal Network and RoI pooling operations. A model is created for each and every process. For example, when we need to take an image under observation, we need to import it as a test data (the data which is to be tested). The actions of importing an image is done with the help of Python codes.

3.2 Gathering and labelling pictures

We created a huge database by gathering different images of bananas containing the objects we are requiring for our application. Labelling these pictures is also an important task. The gathered images were of either fresh bananas or rotten bananas. Since our focus was to detect rotten bananas, the number of images with a rotten banana were higher than the images with fresh bananas.[12] Figure 2 shows the labelling process for the training data.

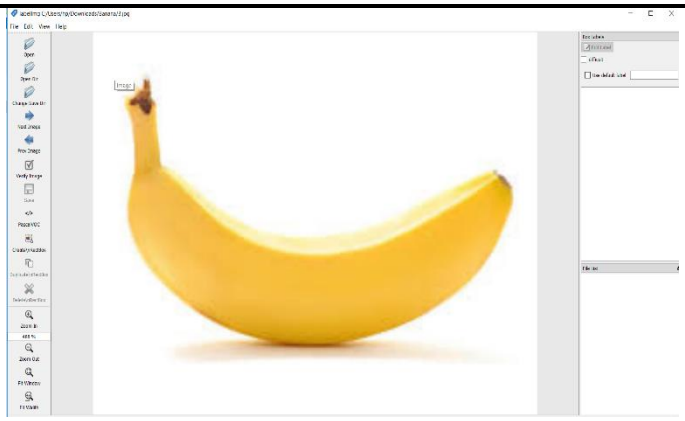

Fig. 2 Labelling the images

3.3 Generating training data

3.4 Creating a label map and configuring training

The object detection training pipeline must be configured. The models created for various purposes such as Region of Interests pooling or feature extraction etc. will be checked and appropriate model will be taken into consideration. It is very important to trace back the labelled images. A label map is created for the same purpose.

3.5 Training

Once the data is gathered, configured and analyzed, we must train the classifier to work on our database.

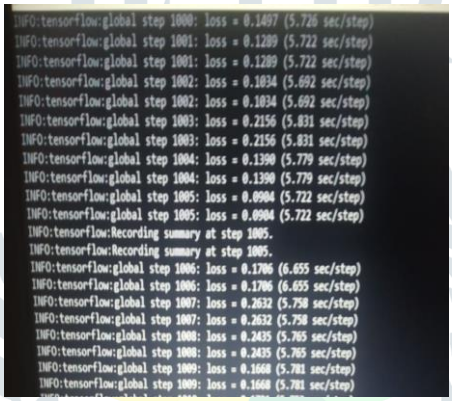

Fig.3 Training of the images

3.6 Exporting the inference graph

3.7 Testing and using the newly trained object detection classifier.

IV. PERFORMANCE AND RESULTS

Whenever an image is provided to the classifier after training it to classify between a fresh banana and a rotten banana, the given image will be classified as either a fresh/rotten based on its appearance. If the image is having a banana which is fresh in appearance, the classifies it as a fresh one with a percentage of the appearance that matches with the number of images provided to the classifier. Same is the case with the rotten bananas as well. If the appearance of given image is rotten i.e. black, reduced size or distorted size etc., then the classifier classifies the given image as a rotten one with the percentage of its appearance that matches with the given number of images.

Fig. 4 Image classified as a fresh banana

Fig. 5 Image classified as a rotten banana

V. RELATED WORK

We have referenced the work of Stephen Gould, Tianshi Gao and Daphne Koller in the field of region based segmentation and object detection. Their work introduces a new system which handles multi-class image segmentation. We have also looked into the work of Christian Szegedy, Alexander Toshev, and Dumitru Erhan in the field of deep learning in which they have also made use of object detection to show their work.

VI. CONCLUSION

We have successfully implemented an object detection classifier with the help of a Faster R-CNN Algorithm and now the classifier can be used to distinguish between a fresh and a rotten banana. Eventually, the same classifier can be used to classify different types of fruits and vegetables in the food products industry to detect the quality of the fruit or vegetable that will be used to produce the required products.

REFERENCES

[1] Guo, L., Liao, Y., Luo, D. & Liao, H., 2012. Generic Object Detection Using Improved Gentleboost Classifier. Phys. Procedia 25, 1528–1535.

[2] Elhariri, E., El-Bendary, N., Hassanien, A.E. & Snasel, V., 2015. “An Assistive Object Recognition System for Enhancing Seniors Quality of Life”. Procedia Comput. Sci. 65, 691–700.

[3] C. Tang, Y. Feng, X. Yang, C. Zheng and Y. Zhou, "The Object Detection Based on Deep Learning," 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, 2017, pp. 723-728..

[4] P. Ahmadvand, R. Ebrahimpour, P. Ahmadvand, "How popular CNNs perform in real applications of face recognition", 2016 24th Telecommunications Forum (TELFOR), pp. 1-4, 2016..

[5] Tran Thi Trang, Cheolkeun Ha, "Irregular moving object detecting and tracking based on color and shape in real-time system", Computing Management and Telecommunications (ComManTel) 2013 International Conference on, pp. 415-419, 2013.

[6] X. Zhou, W. Gong, W. Fu and F. Du, "Application of deep learning in object detection," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 631-634.

[7] N. Singla, "Motion detection based on frame difference method", International Journal of Information & Computation Technology, vol. 4, no. 15, pp. 1559-1565, 2014.

[8] W. Ouyang, X. Wang, X. Zeng et al., "Deepid-net: Deformable deep convolutional neural networks for object detection", 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2403-2412, 2015.

[9] A. Borji, M.M. Cheng, H. Jiang et al., "Salient object detection: A benchmark", IEEE Transactions on Image Processing, vol. 24, pp. 5706-5722, Dec 2015.

[10] P.F. Felzenszwalb, R.B. Girshick, D. McAllester et al., "Object detection with discriminatively trained part-based models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1627-1645, 2010.

[11] Stephen Gould, Tianshi Gao & Daphne Koller, “Region-based Segmentation and Object Detection,” 23rd Annual Conference on Neural Information Processing Systems, pp. 655-663, 2009.

[12] Christian Szegedy, Alexander Toshev & Dumitru Erhan, “Deep Neural Networks for object detection,” Advances in Neural Information Processing Systems, pp. 26, 2013.