Comprative Study and Analysis of Object Detection using R-CNN

Sneha Madane, Shruti Patil, Rohini Patil

Abstract**:** ***Object detection is a field that has been in the limelight a lot in the recent years. Computer vision and image processing are involved in this computer technology and are widely used. Along the path of harnessing the power of vision, numerous algorithms have been found from simple edge detection to pixel level object detection. In this paper we have studied the advancements in object detection algorithms like R-CNN and the latest one being we have studied papers based on types of R-CNN like Fast R-CNN, Faster R-CNN and Mask R-CNN. We have seen their applications in various fields, studied their efficiency, accuracy and limitations.***

Keywords: Region based Convolutional Neural Networks (R-CNN), Object Detection

I. INTRODUCTION

Humans are able to detect objects in their surroundings on a daily basis enabling them to perceive their surrounding environment well. In a bid to make machines autonomous and be able to navigate in the human world, it is imperative for the machine to perceive the environment in a similar manner to the way humans do. Object detection allows the machine to analyse the environment and detect objects around them. This can help the machine to recognise its surroundings and perform a multitude of tasks. Object Detection has found application in almost every field such as surveillance, vehicle navigation, autonomous robot navigation, face/people detection, etc. Basically, to locate objects in an image object detection techniques' approach is to make bounding boxes covering the objects. To identify such occurrences and identify them quickly algorithms like R-CNN, fast R-CNN etc. have been devised. The algorithms addressed and surveyed in this paper include Fast R-CNN, Faster R-CNN, and Mask R-CNN.

Issues of R-CNN are that classification of 2000 region proposals for each image takes up valuable training time of the network, approximately 47 seconds per test image. Moreover, learning is hindered to an extent as selective search is not a dynamic algorithm.

To solve the issues with R-CNN and basically fasten the R-CNN algorithm, Ross Girshick et al. came up with Fast R-CNN. In this algorithm, a convolutional feature map is created by feeding input image to the convolutional neural network instead of region proposals being fed. One can say it is similar to the R-CNN in some ways. The convolutional feature map thus created is studied to extract region of proposals from it.

Once identified, they are bent into squares with the use of Rol pooling layer and reshaped into fixed sizes. This is done to enable it to be fed into a fully connected layer. Softmax layer from the Rol feature vector is used to predict the class of the proposed region along with offset values of the bounding box. Fast R-CNN turns out to be faster than traditional R-CNN because it does not require 2000 region proposals to be fed to the CNN every time. On the contrary, the feature map is generated from the working of convolution operation once per image only [7], [8], [11].

The time-consuming and slow process of selective search is used to extract region proposals in both of the above algorithms(R-CNN and Fast R-CNN) which degrades the network performance. Shaoqing Ren et al. devised an object detection algorithm that can find out region proposals without using the selective search algorithm. This algorithm is quite similar to Fast R-CNN up to the point of providing image as input to a convolutional network and generation of feature map. The change comes in the next step where selective search was used on the feature map initially to extract region proposals. In this algorithm, region proposals are predicted with the assistance of a separate network. The resulting predicted region proposals are subject to refiguring by a Rol pooling layer which then classifies the image and offset values for bounding boxes are calculated. The above process yielded results faster than the previous algorithms and hence this algorithm was named Faster R-CNN [1], [9], [10].

The techniques discussed above used the bounding boxes approach. This can even be extended to locating pixels inside an image. Kaiming He, along with a team of researchers devised Mask R-CNN which allowed pixel level segmentation along with object detection. An extra branch is added to the Faster R-CNN which outputs a binary mask. This mask is used to detect if a given pixel is a part of the object. The procedure followed by the researchers initially led to some inaccuracies. But this problem was eliminated by adopting a method known as RoI Align which adjusts the RoI Pool in order to get more aligned. Once the masks are generated, they are combined with the bounding boxes from Faster R-CNN to get pixel level segmentations [2], [3], [4].

The rest of this paper is ordered as; section 2 will set the objectives of this survey, section 3 outline the methodology used to carry out the survey, section 4 will be a discussion about the performed survey in a tabulated format followed by section 5 which analyses the performance of the algorithms and the last section consists of the conclusion and future work.

Revised Version Manuscript Received on 05 May, 2019.

Sneha Madane, Department of Computer Science, Terna Engineering College, Navi Mumbai (Maharashtra)-400706, India.

Shruti Patil, Department of Computer Science, Terna Engineering College, Navi Mumbai (Maharashtra)-400706, India.

Rohini Patil, Department of Computer Science, Terna Engineering College, Navi Mumbai (Maharashtra)-400706, India.

II. OBJECTIVES

In this study, we will review different object detection techniques over various applications that were conducted in the past three years. Papers that were published from 2016 till date will be reviewed in this study with the motivation to find out the magnitude of usage of such techniques, the accuracy with which they perform and their limitations if any.

III. METHODOLOGY

In this study, a number of electronic databases were used to search on the topic. Only the most recent papers and articles (past 2-3 years) were considered relevant for performing the study. The searches were performed on the topic of Object Detection Algorithms with a major focus on the evolution of R-CNN techniques in the field of Object Detection. Popular Electronic Databases that were used as a source for study included IEEE Xplore, Google and Science Direct. Different keywords were used in order to perform an extensive search on the topic. The below table lists the databases used for performing literature search and the names of the websites for the same.

IV. DISCUSSION

The studies and papers selected above are summarized below in a tabular format. Methodology, application, efficiency, accuracy, research results and limitations and gaps. The methodology column describes the method/algorithm used in the paper while the efficiency denotes how the proposed algorithm affects performance. Accuracy helps in determining the false positives/negatives and limitations and gaps are the problems identified with the method. Figure 1 refers to the various object detection techniques we have covered in this survey paper.

Method	Data	Efficiency	Detection Rate	Output	Issues
A two-step Fast-RCNN which consists of a convolution network that extracts feature and a RoI network. [7]	Object Detector sturdy to various conditions like Occlusions, Deformations and Illuminations	Partially addressed in comparison to OHEM models	Addressed, through tables	Learn invariances in object detectors like occlusions and deformations. Also, boosts detection performance on VOC and COCO significantly	Not addressed
A Scale-Aware Fast R-CNN that uses two sub networks to detect small and large pedestrians and passed through a gate function for fusion. [8]	Pedestrian Detection **Dataset:** Caltech, INRIA and ETH, KITTI dataset	Addressed through means of comparison graphs	Addressed	Shared convolutional layers were used and two sub networks were unified into a single architecture to detect different instances of pedestrians	Detects only pedestrians from the available background
A Fast R-CNN method where features are pooled from the last convolutional layer for each bounding box proposal[11]	Using Pooling that is dependent on Pooling and Layer wise Cascading rejection classifiers **Dataset:** PASCAL, KITTI dataset	Addressed	Addressed, elaborately discussed with a detailed representation of improvement	The Scale Pooling increases accuracy of identifying small objects and the Cascading reject the negative object proposals	Not Addressed

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Modifying Faster R-CNN specifically for vehicle detection on KITTI dataset. [9]	Vehicle detection **Dataset:** KITTI dataset	Not addressed	Extensively addressed, graphically represented	Better performance on easy examples of KITTI dataset, worse performance on moderate and hard examples	Accuracy on moderate and hard examples is worse
Faster R-CNN that has two modules the first concerns Regional Proposal Network and the second is Fast R-CNN to refine proposals. [10]	Face detection **Dataset:** Face Detection Dataset and Benchmark (FDDB), WIDER face dataset,	Partially addressed, sharing of convolutional layers	Addressed, graphically represented for different datasets	Effective face detection performed	Special patterns of human faces not considered

V. PERFORMANCE ANALYSIS AND COMPARISON

R-CNN belongs to the state-of-the-art CNN based deep learning object detection techniques category. Fast, Faster and Mask R-CNN are modifications of this approach based on different applications and requirements.

Fast Region based Convolutional Neural Network is a modification of R-CNN. However, in R-CNN the region proposals are fed to the CNN, where as in Fast R-CNN the input image is fed to the CNN to generate a convolutional feature map. This feature map is used for the identification of region proposals which are then warped into squares by using a Region of Interest pooling layer. They are then reshaped into fixed size to be fed to the fully connected layer. From the RoI feature vector we make use of a softmax layer for predicting the class of the region that is proposed and also the offset value for the bounding boxes. The reason Fast R-CNN is better than R-CNN is because a large number of region proposals are not required to be fed to the CNN every time. Instead, a feature map is generated from the convolutional operation performed only once per image. R-CNN and SPPNet first trains the CNN for softmax classifier, then uses the feature vectors for training the bounding box repressor. Thus, R-CNN and SPPNet are not end-to-end training. On the other hand, Fast R-CNN improves training and testing speed and detection accuracy. The main advantages include: Fast R-CNN trains the very deep VGG-16, 9 times faster than R-CNN and 213 times faster at test time. Also, Compared to SPPNet, it trains VGG-16 3 times faster and tests 10 times faster, plus it's more accurate.

Faster R-CNN is an extension to Fast R-CNN. In both the techniques (R-CNN and Fast R-CNN), region proposals are found using selective search. Selective search is a slow process and affects the performance of the network. Keeping the rest of the algorithm same, Faster R-CNN eliminates the selective search to identify region proposals and uses a separate network instead. As a result, the testing time in Faster R-CNN is much less as compared to its predecessors and can be used for Real-Time Object

Detection. The learning efficiency of the training dataset makes Faster R-CNN models more suitable for identifying classified moving objects. It can be deemed superior to ordinary CNN algorithms due to its high accuracy in labelling correct classes during the validation and testing of dataset. Figure 2 shows the testing speed of different object detection techniques. The X-axis refers to the speed and the Y-axis refers to the various techniques under survey.

Mask R-CNN is completely based on the architecture of Faster R-CNN. It has two major additions. First and foremost, a more accurate Region of Interest Align module replaces the Region of Interest Pooling module. Secondly, an additional branch out of the Region of Interest Align module is inserted. The additional branch is used to accept the output of the ROI Align which is then fed to the two Convolution layers. The final output that we get from the two Convolution layers, is the mask itself. Mask R-CNN inputs the CNN feature map and outputs a matrix with 1's for pixel that belong to the object and 0's elsewhere.

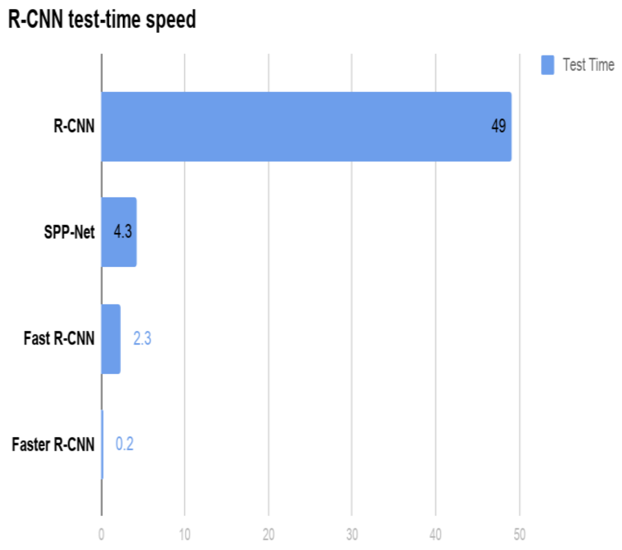

Figure 1

VI. CONCLUSION

In this paper, we have reviewed the techniques used for Object Detection, particularly the different versions of R-CNN. In a short span of three years, the field of Object Detection has seen numerous algorithms, each overcoming the limitations of the previous one. We observed that the basic algorithm of R-CNN was too slow as the proposed regions in each image overlapped and the CNN computation had to be run again and again. This was overcome by a new version, the Fast R-CNN. This technique was then speeded up by Faster R-CNN which replaced the selective search algorithm by using the classifier results for getting region proposals. Pixel Level Segmentation was then added by Mask R-CNN the fastest algorithm for Object Detection was finally proposed which significantly reduced the time taken for detection. These algorithms were successfully used in a number of applications such as face detection, inshore ship detection, airplanes detection on the ground.

REFERENCES

1. Minh-Tan Pham and Sébastien Lefèvre, “Buried Object Detection from B-Scan Ground Penetrating Radar Data Using Faster-RCNN”, IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6804 - 6807, 2018.
2. S. Nie, Z. Jiang, H. Zhang, B. Cai, and Y. Yao, “Inshore Ship Detection Based On Mask R-CNN”, IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 693 - 696, 2018.
3. K. Malhotra, A. Davoudi, S. SIegel, A. Bihorac, and P. Rashidi, “Autonomous detection of disruptions in the Intensive Care Unit Using Deep Mask R-CNN”, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1944 - 19442, 2018.
4. Y. Zhang, Y. You, R. Wang, F. Liu, and J. Liu, “Nearshore vessel detection based on Scene-mask R-CNN in remote sensing image”, 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), pp. 76 - 80, 2018.
5. Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, and M. Li, “Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO”, 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Vol. 02, pp. 562 - 565, 2016.
6. V. Kharchenko and I. Chyrka, “Detection of Airplanes on the Ground Using YOLO Neural Network”, 2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET), pp. 294 - 297, 2018.
7. X. Wang, A. Shrivastava, and A. Gupta, “A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection”, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3039 - 3048, 2017.
8. J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, “Scale-aware Fast R-CNN for Pedestrian Detection”, 2018 IEEE Transactions on Multimedia, pp. 985 - 996, 2018.
9. Q. Fan, L. Brown and J. Smith, “A closer look at Faster R-CNN for vehicle detection”, 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 124 - 129, 2016.
10. H. Jiang and E. Learned-Miller, “Face Detection with the Faster R-CNN”, 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 650 - 657, 2017.
11. F. Yang, W. Choi and Y. Lin, “Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2129 - 2137, 2016.