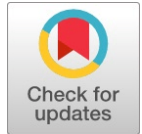# Object Detection Based on Faster R-CNN

M. Sushma Sri, B. Rajendra Naik, K. Jaya Sankar

Abstract: In recent years there is rapid improvement in Object detection in areas of video analysis and image processing applications. Determing a desired object became an important aspect, so that there are many numerous of methods are evolved in Object detection. In this regard as there is rapid development in Deep Learning for its high-level processing, extracting deeper features, reliable and flexible compared to conventional techniques. In this article, the author proposes Object detection with deep neural networks and faster region convolutional neural networks methods for providing a simple algorithm which provides better accuracy and mean average precision.

Keywords: Object Detection, Deep Learning, Neural Networks, Deep Neural Networks, Convolutional Neural Networks, Region Convolutional Neural Network, Faster Region Convolutional Neural Network.

I. INTRODUCTION

In present generations tracking an object playing an important role in video coding applications and to track any object the detection of an object is an important aspect for any moving or stationary conditions. This object detection involves identifying and classifying objects from a video scene and the image segmentation, image recognition, security areas and autonomous vehicles are all the applications relates to the Object Detection. The main process of classification and localization are also involved in Object Detection. In this regard the Object Detection are divided into two approaches namely: Machine Learning method and Deep Learning method. This Machine Learning method are considered for smaller datasets and Deep Learning methods are met for processing larger data sets. In present generations Object Detection based on Deep Learning plays a vital role in which this method employs convolutional neural networks based on region classification knowns as Region Convolutional Neural Network (R-CNN). In this article the author proposes Faster R-CNN based on deep learning method for detecting desired object from a data set. This Faster R-CNN is an extension for region convolutional neural networks technique used for detecting desired objects and extracts features based on Deep Neural Network [1] (DNN) with Convolutional Neural Networks (CNN) [2]. The main contributions of this article give review on Deep Learning, CNN and R-CNN techniques and the detection of object for CIFAR datasets gives better accuracy

M. Sushma Sri, Department of ECE, Osmania College of Engineering, Hyderabad, India. Email: hiremath.sushmasri@gmail.com
B. Rajendra Naik, Department of ECE, Osmania College of Engineering, Hyderabad, India. Email: rajendranaikb@gmail.com
K. Jaya Sankar, Department of ECE, Mahatma Gandhi Institute of Technology, Hyderabad, India. Email: kottareddyjs@gmail.com

and mean average precision compared to other method like LSTM structure. Recently researchers follow the growth of deep learning at image classification [3] and image description [4] and sensor-based activity recognition [5] [6]. In recent times deep neural networks for Object Detection based on Deep Learning and Neural Networks techniques are widely used for processing/ detecting/ tracking purposes. The major importance is given in this article is on region-based techniques known region convolutional neural network (R-CNN) [17], Fast R-CNN [23]. The R-CNN technique is based on region proposals for detecting the specified object. The main working model of DNN is shown in "Fig. 1". Conventionally the R-CNN techniques have developed in extraction of regions of desired objects but extended technique called Fast R-CNN gives better performance in detecting the desired objects.
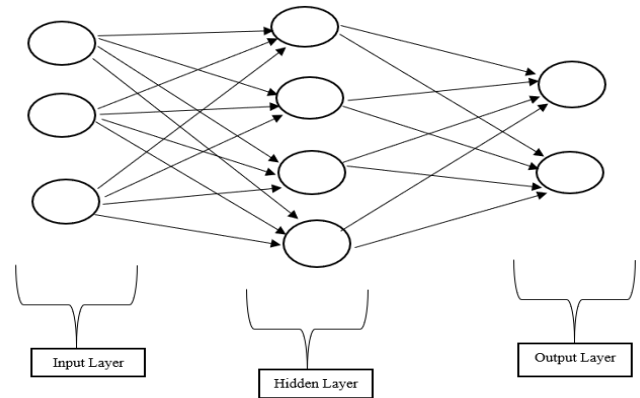
Fig. 1: Working model of Neural Networks

The Neural Network model consists of input layer, hidden layer and output layer in the input layer also known as first layer which consists of input signals / neuron and forwards the signals / information to the hidden layer. The main computations and feature extraction are processed in hidden layer which is also known as middle layer because it mediates between input and output layer. After processing the signals / information is fed to output layer used for generating output signals / neurons.

II. RELATED WORK

In Real-Time Object Detection [7], the author couldn't explain parameters like accuracy or mean average precision which is important metric for object detection and concluded object detection using convolutional networks was overcome in this article by proposing a technique called Faster Region based Convolutional Neural Network. This technique is proposed for generating the detected object with better accuracy and mean average precision. CNN architectures [8], [9], [10], [11] are used for improving accuracy by adding more and more convolutional layers.

Object Detection Based on Faster R-CNN

The addition of more layers in convolutional neural networks is met for features extraction and as region are proposed to this convolutional neural network can be deployed at low cost and generates better performance of Object Detection.

The main important tasks in the area of computer vision are image segmentation, image restoration and object detection, of which Object Detection plays a dominant role. There are two approaches namely bottom-up and top-down approach in which the local feature extraction which plays a vital role in bottom – up approach and segmentation of an object is deployed in top-down approach [12],[13]. The provision of high and multi-level feature representations and the applications in correspondence to computer vision tasks such as Image Segmentation [14], Edge Detection [15] and Object Detection [16] are widely used. This object detection is most important task for tracking purposes. R-CNN by Ross Girshick [17] obtained mean Average Precision (mAP) of 53.3% on PASCAL VOC 2012 which was improved to 11% by using the proposed method explained in this article called Faster region convolutional neural network.

III. DEEP LEARNING NEURAL NETWORKS

Deep Learning are used for processing larger data sets and mapping input to output layers [18]. The Artificial Intelligence is emerging technique enables a machine to stimulate human behaviour by Machine Learning which in turn subset of Artificial Intelligence. If the data or information is present in the form of images are solved by CNN and if our input consists of time series data then R-CNN is used. In this article for Object detection the author used both Convolutional [19] and Region based Convolutional Neural Networks [20] for better efficiency and improved accuracy of data. The important aspect of Neural networks is to train the data to recognize the patterns and then predict the outputs for a new set of similar data [21]. It is combination of software and hardware that is designed for performing specific tasks. In this regard the Deep Neural Network (DNN) is defined as deep and wide Neural Network in which deep means a greater number of hidden layers and wide denotes many input/ hidden nodes. As shown in "Fig. 2",[22], the lots of data is converted to low-level, mid-level and high-level features. These features are extricated from learned features in a Convolutional Neural Network and as deeper inside the feature of an objects are extracted exactly and final extracted object is vehicle. İn this article 32x32x3 hidden layers are existed for detecting the objects.
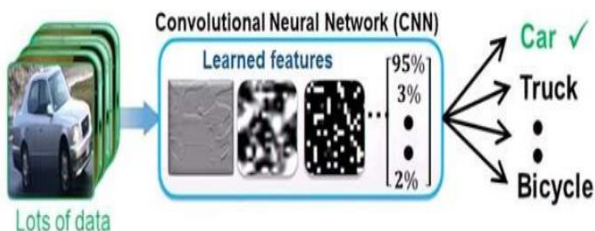

Fig. 2: Deep Neural Network

IV. REGİON CONVOLUTİONAL NEURAL NETWORKS

A Convolutional Neural Network plays a dominant role in neural networks which extracts features of an object in which it implements some challenges like using a huge amount of data for training, processing a huge power and developing

rapid change in applied field. In this article, the CNNs are applied to Deep Learning field for processing training data set based on R-CNN for regions to be extracted in detecting objects. The "Fig. 3" shows the architecture of R-CNN in which the input data set image is given to deep CNN and Region proposal for processing the objects and then the convolutional neural network features are extracted based on region which is the proposed method in this article and the final output image is a detected object. The main sections involved in R-CNN are Region Extractor, Feature Extractor and Classifier. The region extractor is used for regioning the input image which are extracted for creating some boundaries. The feature extractor the features are extricate by Deep Convolutional Neural Network for computations. The classifier as name suggests it is used for classifying the regions based on R-CNN.
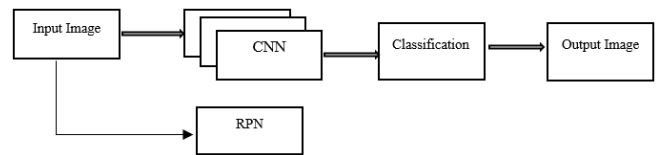

Fig. 3: Block Diagram of R-CNN Architecture

The main significance of R-CNN is to improve the quality in detecting objects and also extract high level features. This R-CNN method consists of Regional Proposal Network (RPN) used for generating regions for an object / image and CNN for extracting the features of an object. The classification is final stage for generating desired output.

V. PROPOSED METHOD

The proposed model seen in "Fig. 4" proposes that the CIFAR-10 data set is loaded which consists of 60,000 color images in 10 different classes and each class contains 6000 images of 32 x 32 dimension. In this regard larger data set the 50,000 are training images and remaining images are test data set. In this article the author uses 300 images of dataset for processing object detection. After loading a vehicle detector image is displayed and applied to CNN for extracting features and for detection purposes the CNN analyses smaller parts of image, so that input size is resized. The proposed technique detects the image into four steps called training options configuration. In this configure training options the first two steps are used for region proposal and detection processed in Faster R-CNN. The last two steps are met for combining the processing images from the first two steps so that a single processing image is detected. The final output image processed by Faster R-CNN is evaluated with evaluation metrics.
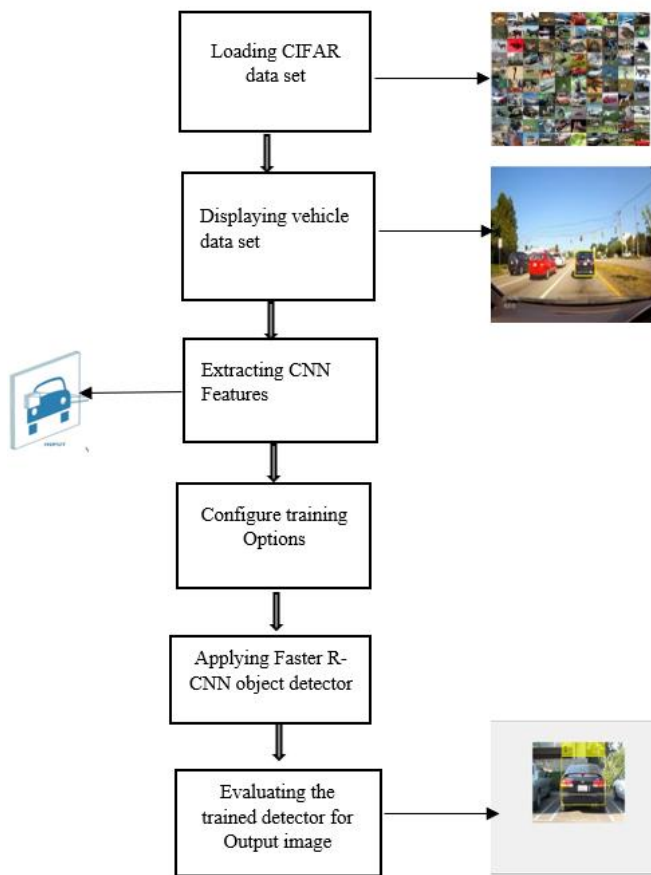
73

Fig.4: Proposed Model

A. Faster R-CNN Deep Learning

Object Detection uses "Faster Region Convolutional Neural Networks" (R-CNN) in which firstly the Convolutional Neural Networks (CNN) are used for classifying input image regions and R-CNN detects the regions which are likely to be detected so that the computational cost is reduced and increases the accuracy. Additional to this many region proposals are generated for better performance in detecting desired output. In this article Faster R-CNN object detector is used to detect a particular thing. The main advantage of requiring training images is to reduce the time and "Fig. 5" shows the block diagram for Faster R-CNN model architecture for object detection.
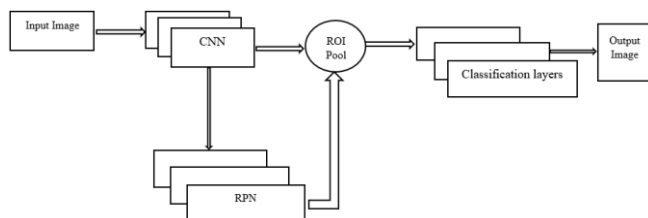

Fig.5: Block Diagram of Faster R-CNN Architecture

The Faster R-CNN is addition to R-CNN model produces region proposal network instead of acquiring from any algorithm. The run time performance of Faster R-CNN is better than the R-CNN. This technique allows does not support to custom region proposal networks. For object detection purposes the Region Proposal Network (RPN) is used for locating objects and generating regions. The importance of RPN as it classifies features of full objects / images of CNN and also it is low-cost region proposal. The

both CNN and RPN are fed to the Region of Interest (ROI) pooling which widely used for object detection purposes using CNN. The main important of ROI pooling to reshaping and then classifying desired regions. The desired regions are obtained classification to generate final output.

B. Evaluation Metric

The accuracy is used to measure, predict or estimate the object from an image. İt is the ratio between number of true predictions to the total number of predictions is seen in "equation (1)".

$$Accuracy = \frac{Number\ of\ true\ predictions}{Total\ number\ of\ predictions} \tag{1}$$

Mean Average Precision (mAP) is a popular metric in measuring accuary of object detctors like Faster R-CNN, Single Shot Detector. The mAP computes the precison value for a recall value in which precision is used measure the ratio between True Positive (TP) to the sum of TP and False Positive (FP) based on predictions seen in equation 2 and recall measures the ratio between TP to the sum of TP and Fasle Neagtive (FN) is seen in "equation (3)".

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The true positive and true negative are number of correct predictions of instances in positive and negative respectively and false positive and false negative are number of incorrect predictions of instances negative and positive respectively.

The final formula of mean Average Precision is seen in "equation (4)".

$$mAP = \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

VI. RESULTS

The "Fig. 6" shows the input image taken from data set in which the author detected a particular vehicle and it is further trained for detecting desired object.

Fig. 6: Input Image

Object Detection Based on Faster R-CNN

The proposed technique known Faster R-CNN is applied to desired object for the training the labelled particular vehicle i.e., our desired object. The final verified object is shown in "Fig.7".
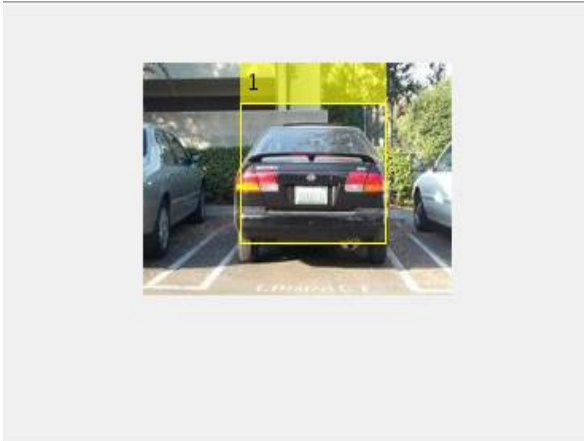
Fig.7: Detected Image

To predict that the detected object is desired object here evaluation of metrics is applied which is seen in "Fig. 8". For training larger data sets the deep learning is used and to detect the object from a data set Faster R-CNN technique is used. Here the training data set is CIFAR data set which consists of 300 images and each image contains one to two labelled instances of vehicles.
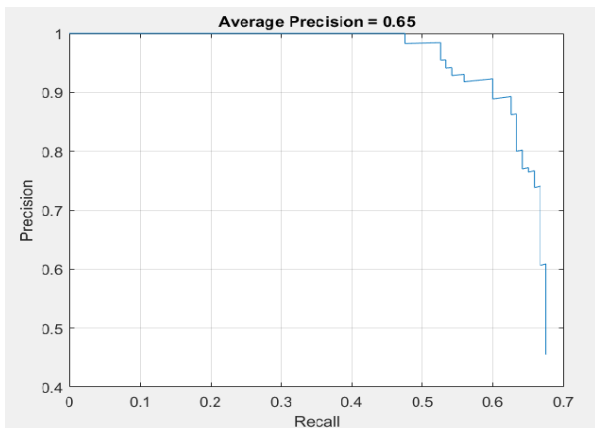
Fig.8: Graph of detected image

The Table I shows the accuracy and mean average precision evaluation metrics for image based on LSTM structure and proposed method known Faster R-CNN technique is improved by 0.33 (33%) in mean Average Precision and accuracy is improved by 5.6%. The preliminary results of both images use different datasets namely CIFAR 100 data set and CIFAR 300 dataset. The CIFAR 100 dataset is based on conventional technique based on LSTM structure and CIFAR 300 dataset is based on Deep Neural Network and Faster R-CNN.

Table I: Comparative study of Accuracy and mAP metrics for object detector image.

Image	Accuracy (%)	mAP
CIFAR 100 dataset [25]	93.9	0.32
Proposed Method	99.5	0.65

VII. CONCLUSION

The Object Detection plays an important role in ranging from the navigation of autonomous vehicles to video data compression and biomedical applications. In this regard the author proposes Deep Learning technology for detecting the objects for larger data sets and this data set are trained on different object detection models but the proposed method which uses faster region convolutional neural networks based on deep learning techniques is used to detect the desired object with 99.5% accuracy and mean average precision with 0.65.

ACKNOWLEDGMENT

REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks," in NIPS, 2012.
2. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation,"
3. in CVPR, 2014.
4. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.
5. Donahue, J., Anne Hendricks, L., Guadarrama, S., et al. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2625-2634), 2015.
6. Lane, N.D., & Georgiev, P. Can deep learning revolutionize mobile sensing? In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (pp. 117-122). ACM, 2015.
7. [6] Li, X., Zhang, Y., Li, M., et al. Deep Neural Network for RFID Based Activity Recognition. In Wireless of the Students, by the Students, and for the Students (S3) Workshop with MobiCom, 2016.
8. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779 -788.
9. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with
10. convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
11. K. Greff, R. K. Srivastava, and J. Schmidhuber, "Highway and residual networks learn unrolled iterative estimation," arXiv preprint arXiv:1612.07771, 2016.
12. J. G. Zilly, R. K. Srivastava, J. Koutnik, and J. Schmidhuber, "Recurrent highway networks," arXiv preprint arXiv:1607.03474, 2016.
13. Y.-H. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energyefficient reconfigurable accelerator for deep convolutional neural
14. networks," IEEE Journal of Solid-State Circuits, vol. 52, no. 1, pp. 127–138, 2017.
15. W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in CVPR, 2016.
16. J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 3, pp. 576–588, 2017.
17. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015.
18. S. Xie and Z. Tu, "Holistically-nested edge detection," in ICCV, 2015.
19. R. Girshick, "Fast r-cnn," in ICCV, 2015.
20. Girshick R, Donahue J, Darrell T, Rich Feature Hierarchies for Accurate "Object Detection and Semantic Segmentation" [J]. 2013:580-587.

21. Ren S, He K, Girshick R, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137.
22. Y. Li, B. Sun, T. Wu, and Y. Wang, "face detection with end-to-end integration of a convnet and a 3d model," in ECCV, 2016.
23. X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: An improved faster rcnn approach," arXiv:1701.08289, 2017.
24. C. Guo and L. Zhang, "A simple method for detecting salient regions Pattern Recognition", 42(11):2363–2371, 2009.
25. Training deep learning models with transfer learning video-MATLAB.
26. Zhong-Qiu Zhao,Peng Zheng, Shou-tao Xu, "Object Detection with Deep Learning: A Review", arXiv:1807.05511v2 [cs.CV], 16 April, 2019.
27. A Comprhensive Guide to Convolutional Neural Networks-the EL15 way.
28. Xinyu Li, Yanyi Zhang, Shuhong Chen, "Concurrent Activity Recognition with Multimodal CNN-LSTM Structure", research gate, Feb-2017.

AUTHORS PROFILE

I M. Sushma Sri is a Research Scholar from Osmania University, Hyderabad. And I have ten years of teaching experience. Her areas of interest include, Image processing, Wireless Networks, Signal Processing, Communications and Deep Learning.

Dr. B. Rajendra Naik is a Professor and HOD in Osmania University, Hyderabad. He is currently Nodal Officer for Visvesvaraya Ph.D. Scheme and Dean Students Affairs OU, Hyderabad. He received Bachelor Degree (ECE) from Nagarjuna University and Master's Degree in Digital Systems and Doctorate Degree from Osmania University. He has published over 60 research papers in National, International Conferences and Journals. He has 10 years of teaching and 12 years of research experience.

Dr. K. Jaya Sankar is a Professor and Principal in Mahatma Gandhi Institute of Technology, Hyderabad. He received Bachelor Degree (ECE) from SVU, Tirupati and Master's Degree (Microwave & Radar Engg) from Osmania University, Hyderabad. He received Ph. D from Osmania University. He publications of 50 International Journals, 45 International Conferences, 3 national Journals and 9 National Conferences. He has 28 years of teaching and 15 years of research experience in teaching and guided many of students at his research.