A Review: Object Detection using Deep Learning

Zinal K. Naik
Chhotubhai Gopalbhai Patel Institute of Technology
Uka Tarsadia University
Bardoli, India

Monali R. Gandhi
Chhotubhai Gopalbhai Patel Institute of Technology
Uka Tarsadia University
Bardoli, India

ABSTRACT
Identifying and detecting the different objects in an image is important skill in computer vision. Fast and reliable object detection is significant approach for interacting with one's environment. Human uses a process called as visual attention to quickly decide which location of an image need to be processed in detail and which can be ignored. But for the machine it is difficult task to identify object and exact location of the object in an image. To overcome this difficulty machine learning and deep learning have made great progress with the help of algorithm series based on R-CNN. This algorithm series have given amazing results with the help of some well-known datasets like Image-Net, Pascal voc, coco etc.

Keywords
Deep learning, Neural Network, dataset, faster Region with Convolutional neural network, ROI, RPN

1. INTRODUCTION
Object detection is the issue of finding and characterizing a numerous number of objects in a image. Object detection is developing from the single object detection to the multiple-object detection. The meaning of the first is that we identify one single object from the image, it can be said that it is a problem of classification, and the

meaning of the later is not only identify all the objects in an image, but also identify location[1] of the objects. With the rapid development of deep learning, a number of research areas have succeed good results, and accompanied by the continuous improvement of convolution neural networks. With the help of deep learning algorithm series based on R-CNN (region with convolution neural network) give better results.

In this paper, we have summarized some algorithms based on R-CNN for object detection. Here section I discussed about introduction of object detection, section II discussed about algorithm series for object detection and section III Discussed about conclusion and future scope.

2. ALGORITHM SERIES OF R-CNN
A .*Neural Network*

For deep learning dataset and neural network are important parts. Deep learning used by the network has been constantly improving, in addition to the changes in the network structure, the more is to do some tune based on the original network or apply some trick to make the network performance to enhance.[1] here is the algorithm series for object detection based on R-CNN.

1) R-CNN
This algorithm will take an input image and Concentrates around 2000 region proposals, and then with the help of CNN, computes features for each proposals [2], and then linear SVM classifies each region.

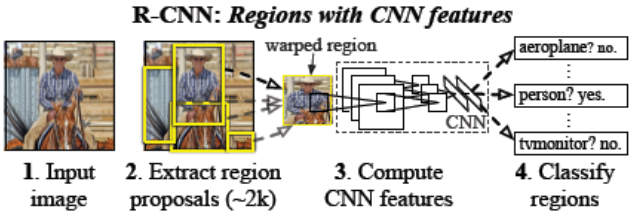

Figure 1: Object detection system for R-CNN[2]

The goal of R-CNN is to take an image, and properly identify where the key objects (via a bounding box) in the image. R-CNN proposes a bunch of boxes in the image and see if any of them actually match to an object**.** When proposals are done, R-CNN combine the region to a standard square size and passes it to the reformed version of Alex-Net. On the final layer of the CNN, R-CNN adds a SVM that simply classifies the object [2].

2) SPP-NET
SPP-Net is an improvement based on the R-CNN with faster speed. SPP-Net proposed a spatial pyramid pooling (SPP) layer that removes limitations on network fixed size. SPP-Net only needs to run the convolution layer once (the whole image, regardless of size), and then use the SPP layer to extract features, compared to the R-CNN, to avoid replication of convolution operation. The candidate area, reducing the number of convolution times. In the whole image will passed to the convolutional layer that will saves a lots of time because same area is not calculated multiple times. So major chunk of time (90%) [5] is consumed on the convolutional layers in R-CNN that will resolve in SPP-NET, it will reduces the computation time drastically. Figure 2 describe the network architecture of SPP-NET.

After passing the image in to convolution layers, with the help of selective search algorithm independent feature are calculated for regions. That can be perform by pooling type of operation on just that section of the feature maps of last convolution layer that corresponds to the region.

3) Fast-RCNN
A Fast R-CNN network takes an entire image as an input and a set of object proposals. The primary undertaking of system is forms the entire picture with a few convolutional and max pooling layers to deliver a convolution feature map.

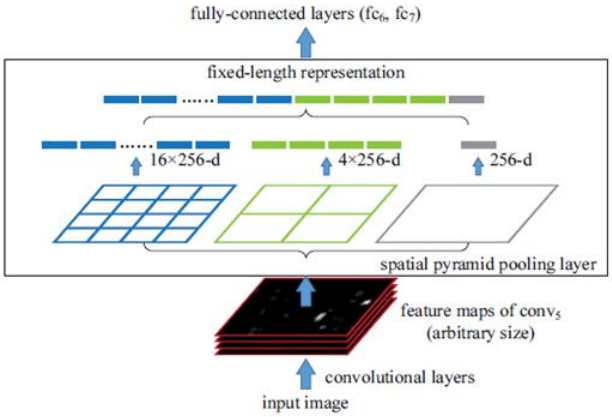

Figure 2: Architecture of SPP-NET [5]

Then, for each object proposal a ROI pooling layer extracts a fixed-length feature vector from the feature map [6]. At that point each feature vector is nourished into a succession of fc layers that at last branch into two sibling layers.[6]

An input picture and numerous ROI are contribution to a FC network.[6] Every ROI is pooled into a fix size feature map and after that mapped to an feature vector by Fc layer.

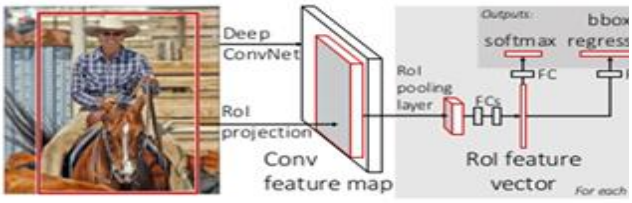

Figure 3: Fast R-CNN architecture [6]

A ROI pooling layer

ROI pooling layer uses a max pooling layer to convert the features inside any valid ROI into a insignificant feature map with a fixed spatial extent of H × W, H and W indicated for layer hyper-parameters.[6] Every ROI is distinct by a four-tuple (r, c, h, w) that identifies its top-left corner (r, c) and its height and width (h, w). ROI max pooling works by separating the h × w ROI window into an H × W grid of sub-windows of approximate size h/H × w/W and max-pooling the values in each sub window into the corresponding output grid cell[6]. Also, ROI layer is just the extraordinary instance of the SPP layer utilized as a part of SPP-nets in which there is just a single pyramid level.[6] fast R-CNN have a single-stage training using a multi-task loss and training will modernize all network layers and fast R-CNN have higher detection quality than R-CNN.

4) Faster-RCNN

Faster R-CNN basically work on Region Proposal Networks (RPN).It will shares convolutional layers with object detection networks by sharing convolutions at test time, the minimal cost for computing proposals is less.[8]

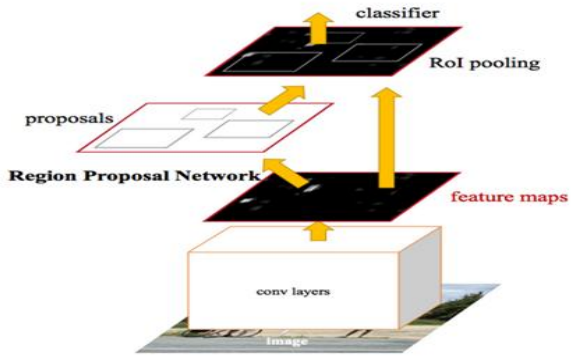

Figure 4: Architecture of Faster R-CNN [8]

RPN takes an image as input and outputs a set of rectangular object proposals. RPN works through passing a sliding window over the CNN feature map and at each window, and it will get k potential bounding boxes and scores for how good each of those boxes is expected to be. k boxes represent such common aspect ratios [8] call anchor boxes**.** To produce region proposals, slide a small network over the convolutional feature map output by the last shared convolutional layer. This small network takes as input an n × n spatial window of the input convolutional feature map. Each sliding window is mapped to a lower-dimensional feature.

This feature is served into two sibling FC layers box-regression layer and a box-classification layer.[8] At each sliding-window location, it concurrently predict multiple region proposals, where the number of maximum possible proposals for each location is denoted as k. So the regression layer has 4k outputs encoding the coordinates of k boxes, and after that classification layer will outputs 2k scores that estimation possibility of object or not for each proposal. So because of RPN network training time will increased and it will gives good mean average precision value.

The comparative analysis of the above discussed algorithms is as follows:

Table1: Advantages and Disadvantages of Algorithm series

Algorithm	Advantages	Disadvantages
R-CNN	Effectively train small amount of data.	R-CNN passes all the regions independently through CNN and therefore it is very slow algorithm.
SPP-Net	SPP allows the whole image to be passed through the convolutional layer only once.	There is no regional proposal method.

Fast R-CNN	Higher detection quality than R-CNN and SPP-Net because of ROI pooling layer.	Fast R-CNN takes too long time for regional Proposal.
Faster R-CNN	Uses deep neural network to compute a regional proposal box.	Do not work for pixel level Segmentation.

The above table shows that faster R-CNN will beneficial compare to R-CNN, SPP-NET, and Fast R-CNN. For object detection there are various datasets are used like image Net, COCO, Pascal VOC, oxford- IIIT pet, KITTI vision. Here is in below table comparative analysis is given on the basis of Pascal voc dataset 2007 and VGG 16 model. And parameters like MAP (mean average precision), training time and testing time.

Table 2: Parameter wise Comparative analysis of Algorithm series [1][3][6]

Algorithm / Parameter	RCNN	SPP-NET	Fast-RCNN	Faster-RCNN
MAP(Mean Average Precision) (Pascal voc07)	66%	63.28%	66.9%	73.2%
Test time/image (Pascal voc07)	50 s	2.3 s	2 s	0.2 s
Train time (Pascal voc07)	84 h	25 h	9.5 h	4.5 h

The above table shows the comparative analysis of algorithm series. here the mean average precision value shows the detection quality, testing time will be calculated in second and training time is calculated in hour. There are different type of datasets are used for object detection like Image-Net, COCO and Pascal Voc dataset. In Image-Net there are 14 million images with 20000 classes. In COCO dataset there is 300000 images and 80 classes. In Pascal Voc (2012) there are a 12000 images and classes.

3. CONCLUSION

Deep learning technique is the faster approach for detecting multiple objects from image based on algorithm like R-CNN, SPP-Net, Fast-RCNN and Faster RCNN. All algorithm are discussed with parameter like mean average precision (Map), training time and testing time. Hence, Faster R-CNN gives better execution time as compare to the other one.

In future it is possible to get more accurate results in object detection with the help of pixel level segmentation.

4. REFERENCES

[1] Xinyi Zhou, Wei Gong, WenLong Fu, Fengtong Du, "Application of Deep learning in Object Detection": in ICIS, Wuhan, China, pp. 24-26, 2017.

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation": IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, June 23-28, 2014, pages: 580-587, 2014.

[3] Wang Zhiqiang, Liu Jun, "A Review of Object Detection Based on Convolutional Neural Network": Proceedings of the 36th Chinese Control Conference IEEE July 26-28, 2017.

[4] Jawadul H. Bappy, Amit K. Roy Chowdhury, " CNN based region proposals for efficient object detection": IEEE International Conference on Image Processing, Phoenix, USA, September 25-28, pages:3658-3662,2016.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, NO. 9, September 2015.

[6] Ross Girshick, "Fast R-CNN". International Conference on Computer Vision 27 Sep 2015.

[7] P. N. Druzhkov, v.d. kustikova, "A Survey of Deep Learning Methods and Software Tools for Image Classification and Object Detection." Pattern Recognition and Image Analysis, 2016, Vol. 26, No. 1, pp. 9–15. © Pleiades Publishing, Ltd., 2016.

[8] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.