Object Detection Algorithms: A Review

Rhea Alex[1]

[1]*B.Tech Student, Computer Science and Engineering, GITAM (Deemed to be University), Visakhapatnam, India*

---***---

Abstract - *This paper aims to compares four different state-of-the-art object detection algorithms to obtain the most suitable algorithm. Object detection is one such field which is gaining wide recognition in the Computer Vision domain. Object detection algorithms is not only being implemented in applications such as self-driving cars, home automation etc. but also in outer space to identify the presence of water, various minerals, rocks in different planets. It often blows one's mind off to determine which algorithm to implement for a given problem at hand. Taking into consideration the speed and mean accuracy precision at which the algorithm performs we would like to obtain an algorithm which can perform real-time detection.*

Key Words*: *Object Detection, YOLO, Fast-RCNN, Faster-RCNN, SSD.

1. INTRODUCTION

Object detection deals with identification of objects in an image or a video. Today, many developers are setting their eyes on the Computer Vision industry. Many modern applications are using object detection algorithms to attract customers with new features. Most customers today are also looking for innovative technologies and object detection is one such field with a lot of scope for innovation. Computer Vision domain has revolutionized the way we live and interact with others.

Object detection algorithms are first trained using datasets which contains the images of the object we want the model to detect. Then we test the model by applying the model onto some real-world data. The algorithm will try to match the features from the training data to the real-world data. If the algorithm provides correct results, then the algorithm can be implemented in the application. However, detection of multiple objects in a single image or a single video frame is a daunting task.

Neural Networks are used to train machines to behave like humans. Convolutional neural networks are widely used for object detection. CNNs have become the standard against which all other algorithms are assessed. CNNs greatly improve the performance and accuracy of the system. This is because CNNs use less number of parameters by making use of the same parameter various times. One major advantage of neural networking for object detection is its ability to perform both semi-supervised and unsupervised learning.

Real-time object detection requires processing of each frame where multiple objects can be present in a single frame. Object localization is a huge challenge for object detection algorithms because not only do we want the algorithm to classify the object but also to identify the location of the object. Most algorithms struggle when group of objects are present together like a swarm of bees. Methods like non-maximum suppression must be done to overcome multiple bounding boxes from detecting the same object.

Object detection algorithms greatly improve in accuracy if the right dataset is used. In order to achieve a desired accuracy, the dataset must contain the right amount of images. Most algorithms improve in accuracy when more images are provided for training. The speed at which the algorithm processes a video is also an important factor to consider when choosing an algorithm for object detection. YOLO and SSD algorithms are faster when compared to Fast-RCNN and Faster-RCNN algorithms

2. OBJECT DETECTION ALGORITHMS

2.1 YOLO

YOLO stands for 'You only look once'. It is named so because in a single evaluation, the algorithm can predict both bounding boxes as well as the class prediction. Hence, YOLO is very fast and is implemented in most real-time applications. The main advantage of YOLO algorithm is that it uses logistic regression to perform detection. YOLO algorithm makes less background errors when compared to other algorithms. YOLO outperforms other algorithms for predicting the class of artistic images. However, this algorithm struggles to detect tiny objects. The backbone network for this algorithm is Darknet. Darknet is a customizable object detection framework created by Joseph Redmond for YOLO algorithm.

YOLO begins by dividing an input image into an S x S grid. Each of these grids must detect if an object is present or not. Each grid returns the bounding box, class probability and the confidence if an object is present or not. These values are then encoded and the final layer of the network returns the output. One limitation of this architecture is that each grid can predict only two bounding boxes and one class probability. Non-maximal suppression is used to refine the bounding boxes.

There are many version of this algorithm such as YOLO, YOLO v2, YOLO v3, YOLO-LITE. YOLO v3 has the highest mean accuracy precision (mAP) when compared to other versions. The YOLO v3 uses a Darkent-53 structure while YOLO v2 uses Darknet-19 structure. Therefore, YOLO v3 has 53 more convolutional layers which sums up to a total of 106 convolutional layers. Hence, YOLO v3 is slow when compared to YOLO v2 but achieves better mAP results. YOLO-LITE algorithm is used for performing detection for non-gpu computers.

2.2 FAST – RCNN

Fast –RCNN algorithm is an improvement over RCNN and SPPnet algorithms. The important change made in this algorithm is to pass a set of object proposals to the CNN layer rather than passing a single proposal for each image. RCNN algorithm is very slow and achieves less mAP than Fast-RCNN algorithm. Also, the training is expensive in RCNN algorithm because it requires a lot of space and time.

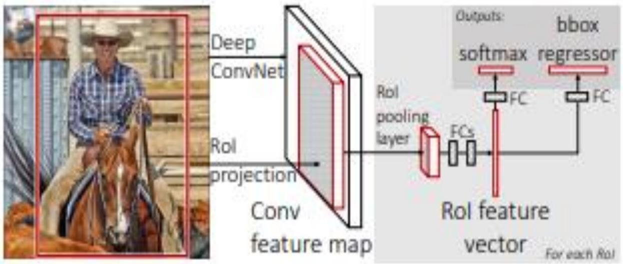

Fig 1: Fast-RCNN architecture

As, we can see from the fig 1, an image and few object proposals are given as input to the convolutional layer and max pooling layer. A convolutional feature map is created. The RoI pooling layer extracts a feature vector from the object proposals. These feature vectors are given as input to two fully-connected layers.

There are two output layers. The first layer which performs softmax probability estimates for all the classes. The second layer which returns four real-values which represents the bounding box coordinates. In this algorithm, stochastic gradient descent batches are sampled hierarchically and the RoIs from a single image share the space and memory in forward and backward passes.

2.3 FASTER-RCNN

Despite all the changes made in Fast-RCNN, the algorithm is still quite slow. This is because the Fast-RCNN algorithm uses selective search for generating object proposals. Selective search method is very slow in generating proposals. A new technique called Regional Proposal Network (RPN) to perform detection. The main observation in this paper is the fact that the convolutional feature map in Fast-RCNN could be used for generating object proposals as well.

In this algorithm, first region proposals are generated and then these proposals are given as input to the Fast-RCNN detector. A regional proposal network takes an input image and generates rectangular proposals along with their objectness scores using the convolutional feature map. Backpropagation and Stochastic gradient descent are used to train the RPN. Joint training and alternate training methods are used in order to share the convolutional layers between Fast-RCNN algorithm and RPN.

2.4 SSD

Single shot detector eliminates the need for proposals and does not resample the pixels. This makes the algorithm faster and is as accurate as regional proposal based methods such as Faster-RCNN. The main observation in this algorithm was to use small convolutional filters applied to feature maps to predict the box offsets and category scores. Detection at multiple scales is possible using this algorithm as it uses separate filters for different aspect ratios.

SSD is a feed-forward convolutional network that takes an image and ground truth box as input. A small set of default boxes are evaluated. Each default box predicts the shape offset and confidence for all the objects. The default boxes are matched with the ground truth boxes during training. Non-maximum suppression is used in the final stage to remove any imprecise bounding boxes.

3. METHODOLOGY

The research carried around in this paper is obtained by studying various object detection algorithms. The focus is to obtain a clear understanding about the mean accuracy precision achieved by different algorithms. This research provides a comprehensive study about the most suitable algorithm for object detection.

4. RESULTS

The result achieved by different algorithms on the PASCAL VOC 2007 + 2012 dataset is described. As we can see in table 1, the result achieved by Fast-RCNN and Faster-RCNN algorithm is described. Faster-RCNN on the ResNet backbone achieves 76.4 mAP but is not as fast as Faster-RCNN on the VGG-16 backbone. The input resolutions of the images were approximately 1000 x 600.

Table 1: Results achieved by Fast-RCNN and Faster RCNN algorithms

Method	mAP	FPS
Fast-RCNN	70	0.5
Faster-RCNN VGG-16	73.2	7
Faster-RCNN ResNet	76.4	5

As we can see in table 2, the result obtained by different versions of YOLO algorithm is described. Fast YOLO processes video with the highest speed but achieves the lowest accuracy. YOLO v2 achieves the highest accuracy. At 155 fps, Fast YOLO achieves a great mAP. On increasing the size of the input image, the mAP increases but the speed decreases.

Table 2: Results obtained by different versions of YOLO

Method	mAP	FPS
Fast YOLO	52.7	155
YOLO	63.4	45
YOLO V2 (416x416)	76.8	67
YOLO v2 (480x480)	77.8	59

As we can see in table 3, the result obtained by different versions of SSD algorithm is given. SSD 512 achieves 76.8 mAP. The frames per second can be increased by increasing the batch sample size.

Table 3: Results obtained by different versions of SSD algorithm

Method	mAP	FPS	Batch Size
SSD 300	74.3	46	1
SSD 512	76.8	19	1
SSD 300	74.3	59	8
SSD 512	76.8	22	8

If the speed of the algorithm is the main priority, then Fast YOLO is an excellent choice for the application. If both accuracy and speed is required, then YOLO v2 is an excellent choice.

5. CONCLUSION

This paper reviews four popular object detection algorithms. The research paper describes the pros and cons of different algorithms. These algorithms are widely used in the Computer Vision domain. These algorithms can be implemented in different applications depending on the requirement of the application. This paper gives a subjective knowledge about the different object detection algorithms.

REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, **2016.**

[2] Joseph Redmon, Ali Farhadi, University of Washington, Allen Institute for AI, YOLO9000: Better, Faster, Stronger, **2016**

[3] Joseph Redmon Ali Farhadi, YOLOv3: An Incremental Improvement, 2018

[4] Ross Girshick. Fast R-CNN, **2015.**

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**, 2016.**

[6] Wei Liu , Dragomir Anguelov , Dumitru Erhan , Christian Szegedy , Scott Reed , Cheng-Yang Fu , Alexander C. Berg, SSD: Single Shot MultiBox Detector **, 2016**