International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

A Comparative Study of Widely Used Image Detection Algorithms

[1]Harsh Zota, [2] Manoj Dhande

[1]Student,Department of Computer Science Engineering, Shah & Anchor Kutchhi Engineering College , Chembur, Mumbai, India
[2]Professor, Department of Computer Science Engineering, Shah & Anchor Kutchhi Engineering College , Chembur, Mumbai, India

---

Abstract - *In today's revolutionary world where everything is at just one tap of a finger. In such a technology prone world image classification plays a major role. A normal human eye has the greatest kind of vision but yet fails to detect a small detail in a big image or even any sort if object among a hundred other objects. It is a strain to the human eye to locate such differences or objects in bigger images. There are multiple images that are a conundrum to the eye .Object detection can not only be used in daily life but can also be used in fighting crime, this is done by detecting the image in a video or a larger image and matching faces. But for all this a major requirement is the appropriate selection of the algorithm. The choice of algorithm most suitable for your project is of great importance. In this paper we will describe 6 widely used image detection algorithms and in-turn give the pros and the cons for the particular algorithm. The paper covers You only look once Algorithm (YOLO), Histogram Oriented Gradients Algorithm (HOG), Fast-Region Based Convolution Neural Network Algorithm (Fast R-CNN), Region Based Convolution Neural Network Algorithm(R-CNN), Region Based Fully Convolution Network Algorithm(R-FCN), and lastly Single Shot Detector Algorithm (SSD),*

Key Words: Technology prone, Image classification, Object Detection, Facial recognition, SSD, R-FCN, R-CNN, Fast R-CNN, HOG, YOLO.

1. INTRODUCTION

In this modern evolutionary day and time, machine learning is the most popular and upcoming fields to the new times. The basics today shall become the roots for the greater technology coming in the future. The simple algorithms today, evolving, will design greater technologies tomorrow. Image classification or object detection falls under a wide scope of progress. Object detection is basically detection of a particular face or an object in a larger image or a video frame. Object detection has seen a major revolutionary change in the computer vision field. It helps us figure out which category does the object fall under.

Object detection has been a field of study and interest for over many decades. The first ever designed algorithm, Viola-Jones object detection framework and today we have come to YOLO. There have been many algorithms in between and we will study majorly used ones in this research paper.

Object detection algorithms used predefined images called annotated images to train the algorithms for the implementation of the algorithm. For image annotations there are multiple tools like VGG Image Annotator or LabelImg and many more. The results of the algorithm are highly based on the kind of images used and the accuracy of the annotations, better the resolution better the outcomes.

The primary purpose of the paper is to research different used algorithms from the object detection and help reader find the most suitable algorithm to his own use. We see multiple times that people with a low dataset and a very simple model use high end algorithms where the same results can be obtained from a much lesser complex algorithm and also the other way around. Hence the paper helps you choose the best algorithm most suitable for your project.

2. LITERATURE REVIEW

1. YOLO- YOU ONLY LOOK ONCE

YOLO is a clever convolutional neural network (CNN) for doing object detection in real-time. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. [1] YOLO is the latest algorithm in the domain of object detection. It is the best in terms of speed and accuracy together when it comes to training big datasets.

YOLO is a real-time object detection algorithm and can detect objects in moving frames or videos. The framework on which YOLO can be implemented is called Dark-net developed by Joseph Redmond for YOLO [2]. YOLO divides the frame in a number of grid and then it tries to detect the object in the particular grid. It uses a Non-max suppression filter to keep only

the most dominating bounding boxes in a particular grid and hence gives most optimum results. There have been many upgrades since the primary YOLO algorithm was proposed and they were named as YOLO v2 and YOLO v3 [3]. Fig 1 explains the multiple layers in YOLO and how with each layer the algorithm refines the image and ignores the background and hence giving the most accurate results.

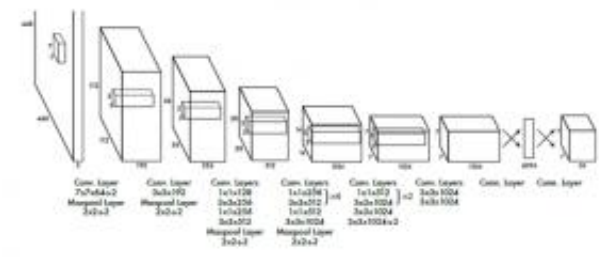

FIG 1

2. HOG- HISTOGRAM ORIENTED GRADIETS

HOG algorithm primarily focuses on the features of the images. It concentrates more on the borders and the shape of the image to extract its features individually from the background. HOG not only does extract the features of the image but also the direction of the edge. It does so by extracting the gradient and the orientation of the edges [4]. These orientation are called localized portions, this tells us that image is broken down to smaller edges and these edges are extracted. With the help of the gradient it creates a histogram and hence it has its name as Histogram Oriented Algorithm.

For resized images HOG calculates the difference in the gradients. The gradients are then converted in form of histograms. It divides the image into 8x8 cells. Each cell in-turn provides the magnitude and the direction of the edge. This algorithm due to the sharpness till the smallest pixel also captures the curves of a bicep of a person [5].

3. FAST-RCNN – FAST REGION BASED CONVOLUTION NEURAL NETWORK

This algorithm overcomes the drawbacks of the RCNN model and is faster compared to it. It unlike YOLO does not look at the image as a whole but passes the whole image to the CNN. It is 9 times faster than R-CNN[6].

The basic approach of the RCNN and Fast RCNN remain the same, as we can see in FIG 2, in Fast RCNN we put the images in the Convolutional Neural Network layer and hence generate a feature map. With the help of this feature map we identify different areas of requirement and warp them in form of smaller squares. Then the squares are made of a fixed size with help of a RoI pooling layer. The softmax layer then helps us to predict the class for the required region and also the offset values of the bounding box [7].

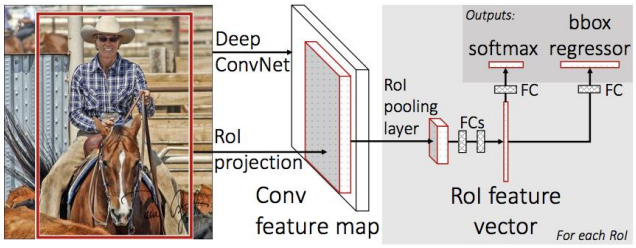

FIG 2

Fast R-CNN has an average prediction time of about 2 seconds as it only passes the features extracted from the feature map to the convolution neural network.

4. R-CNN REGION BASED CONVOLUTION NEURAL NETWORK

As seen in the models before R-CNN the algorithms used a huge number of regions but here that problem is resolved. In R-CNN it takes only 2000 regions from the whole image which is called image proposal. Now instead of distinguishing large number of different regions we work with a limited 2000 regions [8].

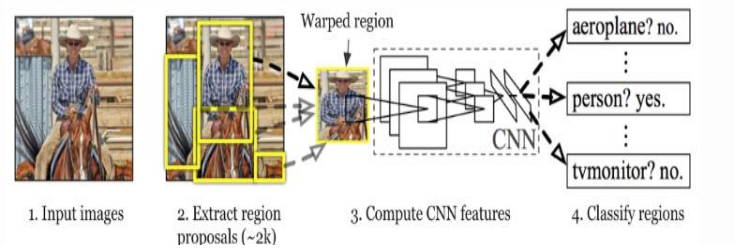

FIG 3

As we can see in FIG 3 there are 2000 regions proposed and then the image is reduced to those 2000 regions and those regions are fed to the Convolutional neural network. This hence gives us the results based upon are classifications.

5. R-FCN - REGION BASED FULLY CONVOLUTION NETWORK

R-FCN is for detecting the smallest detail in the image. It uses multiscale training for feature extraction. Here there is a 9-region based feature map created. Each region points towards the top-left, bottom-right top-center and so on parts of the image. These feature maps are called position-sensitive sore maps as each map detects a sub region. The sub regions can be further divided into regions of 3x3 or 5x5 and more details can be extracted.

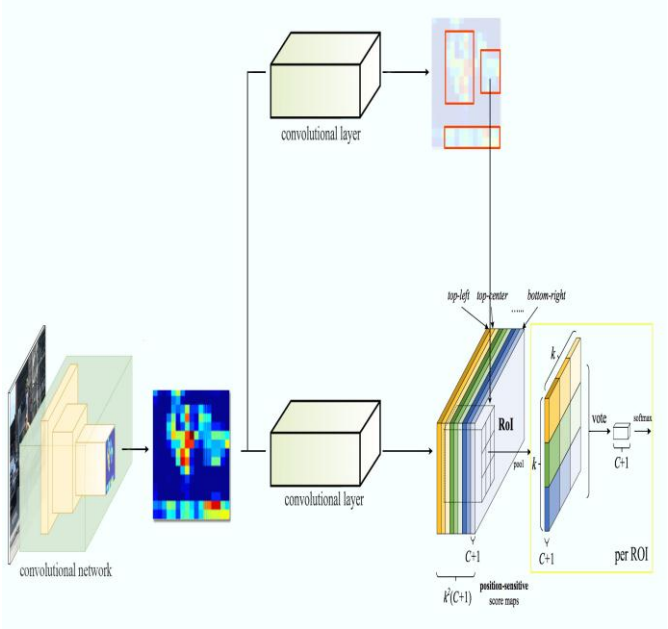

FIG 4

As seen in FIG 4 that a selected sub-layer is further divided into a region of 3x3. This process of maps in maps is called position sensitive-RoI Pooling. R-FCN has shown to be 20 times faster than Faster R-CNN [9].

6. SSD- SINGLE SHOT DETECTOR

Single Shot Detector as the name suggests like YOLO gives the results in a single shot. It does not choose proposed regions to process in the CNN. It undergoes a certain number of feature extractions and hence a feature layer of a particular size is obtained. The convolution layer is then applied on this extracted feature layer [10].

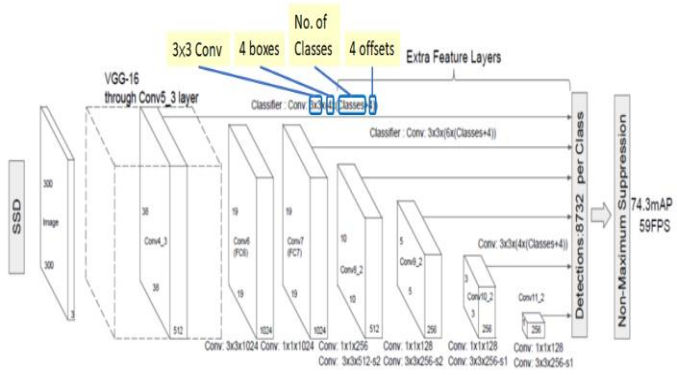

FIG 5

As we can see in the figure the feature layer goes through small convolutions of 3x3 for more accurate results. It is a faster method compared to Faster R-CNN.

3. METHODOLOGY

For this paper there was careful and precise research done over all given 6 object detection algorithms. After careful examining of each algorithm in terms of speed, accuracy and complexity the results were given. The focus was to help reader understand all different algorithms and give its pros and cones

4. RESULTS

Following were the results obtained for each algorithm upon research

4.1 YOLO

PROS	CONS
The biggest advantage of YOLO is its amazing speed, it has a processing rate of 45 frames per second [1].	It happens to struggle with small objects that appear in groups like a flock of tiny birds [1].
It is able to understand the generalized object representation. Due to this when trained even on real world the images and predictions work fairly on artwork too [1].	It yet struggles to generalize objects in new or unusual aspect ratios or configurations.

4.2 HOG

PROS	CONS
The major advantage of HOG is that it provides more global information, In smaller scales, and hence provide more detailed information [5].	It faces problems in the end because the final descriptor vector grows thus it takes more time to extract and hence takes more time to train using the classifier [5].
	It is a machine learning algorithm and thus proves to be less efficient compared to other deep learning algorithms [5].

4.3 FAST R-CNN

PROS	CONS
The training takes place in a single stage and uses multi-task loss. It can also update all network layers.	It uses selective search to find all regions of proposal, in doing so it wastes a lot of time and is

	hence slow
It has a higher detection quality which is measured in (mAP) than R-CNN ,SPPnet.	It takes only 2 seconds to detect objects compared to R-CNN, but when we consider large data-sets it doesn't prove to be that fast [7].

4.4 R-CNN

PROS	CONS
The CNN algorithm overcomes all its previous algorithms and also detects important features without the presence of any human to supervise.	The size of every region is not fixed and hence can cause image distortion as the size of region proposed has to be uniform
It is computationally efficient. It performs pooling and also has parameter sharing	It has also proven to be inefficient for convolutions on every regional proposals.

4.5 R-FCN

PROS	CONS
It is 20 times faster than the Faster R-CNN algorithm	The disadvantages seen in F-RCN are very low due to its high speed and accuracy
R-FCN uses residual network strikes, this helps in getting high speed and accuracy.	

4.6 SSD

PROS	CONS
SSD can be trained for better results. With more training it can give optimum results [10].	It requires a large number of images to provide the most optimum results.
SSD makes more predictions and it has a good control over location, scale and aspect of even low resolution images.	

From the above analysis about the six different algorithms. We can infer the following. Different cases can be deduced. The given algorithms can be divided into one-stage and two-stage detectors. YOLO and SSD are one-stage detectors. Fast R-CNN, R-CNN and R-FCN are two-stage detectors. HOG is a feature extraction algorithm.

CASE 1: In a case where you have to detect an object in a moving frame or a video YOLO will be the best option as it has a frame rate of 45 frames per second. In a moving frame the YOLOv3 is capable of detecting multiple objects in a single frame.

Let us consider a case of a traffic signal where there are hundreds of objects in a single frame. YOLO is capable of detecting all the images at high accuracy with the help of multiple bounding boxes. For such cases and projects YOLOv3 is the best possible algorithm.

CASE 2: Suppose you have a large dataset but the resolution of the images in of a highly degraded quality, here we can use SSD. In comparison to other algorithms this will give the best possible result as it is has good control over location, scale and aspect of low resolution images.

In case of images that have to be rebuilt or you have to train your model from images over decades old with a very low resolution then SSD should be your choice. It will require a large number of images but the model will give results as per requirement.

CASE 3: In a scenario with a single image and you require the minute details of the face HOG will be the best option. HOG helps to extract the smallest details of the image. Other algorithms are not fit to extract such minute details of the image.

When you have a single large image of an object, let us for example say a car , for such images only the outlines of the image is enough for the rough recognition of the image. HOG will be the best choice. This is because it is a machine learning algorithm not a deep learning algorithm so can give optimum results on a regular machine also.

CASE 4: In a small project with limited number of images we can use R-CNN algorithms. It is feasible for small amount of data and with the large objects in the Image as it chooses a random region to propose for the CNN.

From the above analysis we can infer that for all scenarios YOLO will be the best choice of algorithm in any scenario.

5. CONCLUSION

In the paper we successfully understand about six different object detection algorithms. Along with that we get to know basic pros and cons of a particular algorithm. This will help the reader to not only choose well a model for his project but also help to learn about multiple other object detection algorithms. The paper gives subjective knowledge about all algorithms.

REFERENCES

[1] ODSC, Open Data Science, (2018, September 25), Overview of YOLO object detection Algorithm. https://medium.com/@ODSC/overview-of-the-yolo-object-detection-algorithm-7b52a745d3e0#:~:text=YOLO%20is%20a%20clever%20convolutional,and%20probabilities%20for%20each%20region.

[2] Joseph Redmon, University of Washington, Yolov3: An Incremental Improvement.

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

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[5] Aishwarya Singh, 2019, September 4, Feature Engineering for images: A valuable Introduction of the HOG feature descriptor. https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] Rohit Gandhi, R-CNN, Fast R-CNN, Faster-RCNN, YOLO Object detection algorithms, https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

[8] Pulkit Sharma, October 11 2018 ,A step-by-step Introduction to object detection Algorithms, https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/

[9] Jonathan Hui, March 28 2018, Understanding Region Based Fully Convolution Neural Networks (R-FCN) for object Detection, https://medium.com/@jonathan_hui/understanding-region-based-fully-convolutional-networks-r-fcn-for-object-detection-828316f07c99

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

[11] Sultana, Farhana, F.S, March 2019 , A review of object detection models based on convolution Neural Networks, Reserachgate, DOI: 10.13140/RG.2.2.26529.25447