Real-time object Detection using Deep Learning: A survey

Shubham Pal[1], Prof. Pramila M. Chawan[2]

[1]M.Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

***

Abstract - *Object Detection is related to Computer Vision. Object detection enables detecting instances of objects in images and videos. It identifies the feature of Images rather than traditional object detection methods and generates an intelligent understanding of images just like human vision works. In this paper, We revived begins the brief introduction of deep learning and object detection framework like Convolutional Neural Network(CNN), Recurrent neural network (RNN), faster RNN, You only look once (YOLO). Then we focus on our proposed object detection architectures along with some modifications. The traditional model detects a small object in images. We have some modifications to the model. Our proposed method gives the correct result with accuracy.*

Key Words: k means, OpenCV, CNN, R-CNN, Faster-RNN, and YOLO.

1. INTRODUCTION

Deep learning is part of machine learning. Too many methods have been proposed for object detection. Methods of object detection fall under deep learning. Object detection is a computer technology and widely used in Computer vision. Deep learning has been becoming popular since 2006.

1.1 A brief Overview of object detection

Object Detection is a Computer Vision technique. Object detection is a significant research area in Computer Vision. Which can be applied to many applications such as Driverless cars, security, surveillance, machine inspection, etc. Object Detection is used to identify the location of the object in an image, Face detection, medical imaging, etc. Invention and Evolution of Deep learning have changed the traditional ways of object detection and reorganization system.

Computer Vision identifies features present in images, Classifying Objects in the image, Classifying images along with localization, drawing a bounding box around object present in the image, Object segmentation or semantic segmentation, Neural style Transfer. Deep learning methods are the strongest method for object detection. To understanding images, we not only concentrate on classifying images but also try to estimate the concepts and locations of each object in images.

2. Literature Survey

Three are different approaches has presented by many researchers. An algorithm for the first face detector was invented by Paul Viola and Michael Jones 2001. The face had detected in real-time on Webcam feed. It was implemented by Opencv and Face Detection. This was not able to detect some orientation like upside down, titled, wearing a mask, etc. Due to the massive development of Object detection in Deep leaning, object detection classified model into (1) Model-based on region proposal; (2) Model-based on regression/Classification.

2.1 Model-based on Region

2.1.1. CNN: This network was introduced by Authors: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton in 2012. The network consists of five convolutional layers. It takes input as an image which is a 2D array of a pixel with RGB channel. Then Filters or features detector apply to the input image and get output features maps. Multiple convolutional are performed in parallel by applying the ReLU function. CNN works for only one object at a time so it does not work effectively in multiple objects images. CNN became a good standard for image classification after Kriszhevsky's CNN's performance We cannot detect objects which are overlapping and different backgrounds and do not classify these different objects but also do not identify boundaries, differences and relations in other.

2.1.2. RCNN: This network is introduced by Authors: Ross Girshick, Jeff Donahue, Trevor in 2013this network inspired by overfeat. This network includes three main parts, first is region extractor, second is feature extractor and final is classifier. It uses a selective search algorithm for object detection to generate region proposals. Extract 2000 regions for every image. Here 2000 convolutional networks used for each regions of the images. So have one Convolutional network required to process RCNN Region with CNN features divides the image into several regions. Run images through pre-trained AlexNet and finally apply the SVM algorithm.

2.1.3. Fast R-CNN: This network is an improved version of R-CNN which is introduced by Ross Girshick. The article claims that Fast R-CNN 9 times faster than previous R-CNN. Network select sets of bounding boxes then use feature extractor by CNN network then use classifier or regression for output the class of each boxes.

2.1.4. Faster R-CNN: This is an improved version of Fast R-CNN which introduced by Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun in 2015. Image is provided input to a convolutional network that provides convolutional map. To identify the regions here the separate network is used to predict the region proposals.

2.2. Model based on regression/Classification.

2.2.1. YOLO: YOLO (You only look once) at an image to predict what are those objects and where objects are present. A single convolutional network simultaneously predicts multiple bounding boxes and class and probabilities for those boxes. Treats detection as a regression problem. Extremely fast and accurate YOLO takes an image and split it into grids. Each grid cell predicts only one object. YOLO is extremely fast at test time and it requires single network evaluation and performs feature extraction, bounding box prediction, non max suppression, and contextual reasoning all concurrently. YOLO is not applicable for small objects that appears in groups such as flocks of birds. YOLO has several variant like fast YOLO. YOLO is a completely different approach. It looks just once but in clear ways. If a simple image gives through the convolutional network in a single pass and comes out the other end as a 13×13×125 tensor describing the bounding boxes for the grid cells. All you need to do then is compute the final scores for the bounding boxes and throw away the ones scoring lower than 30%.

2.2.2. SSD: SSD (Single Shot MultiBox Detector) Objective of localization and classifications are done in a single forward pass of the SSD network. The first advantage of the network is fast with good accuracy. it runs a convolutional network on input images only one time and computes a features map. Histograms of Oriented Gradients are invented by Navneet Dalal and Bill Triggs invented in 2005. We want to look at each pixel that directly surrounding it. Here compare current pixel to every surrounding pixel. It failed in more generalized object detection with noise and distractions in the background.

3. PROPOSED SYSTEM

3.1PROBLEM STATEMENT

“To implement object detection and recognition in an images and videos using deep learning.”

3.2 PROBLEM ELABORATION

The main objective is to detection and recognition Objects in Real-time. We require rich information in real life. We have to observe the objects which are moving respect to the camera. It will help to recognize objects interaction. We focus on accuracy in this paper.

3.3 Proposed Methodalogy

This model includes feature extractor with Darknet-53 with feature map upsampling and concatenation. Proposed Model includes various modification in object detection techniques.

3.3.1 Darknet 53:

This proposed system uses a variant of Darknet which has originally 53 layers network and trained on Imagenet. For the detection more 53 layers are using onto it, a totally of 106 layers of convolutional underlying for proposed system. This is the reason the proposed system becomes slow.

3.3.2 Detection of three scales

This model makes detection at three different scales. Here Detection is generated by applying 1 x 1 detection kernels on feature maps for three various sizes on three various palces in the networks 1 x 1(M x (5 +N)) is the shape of the detection kernel. Here M is the number of bounding boxes on the feature map and N is the number of classes. Feature map generated by this kernel has the same height and width of the previous feature map also detect attributes along with depth. Three different scales are used. The first detection is created by the 82nd layer. The first 61 layers of the image are sampled by the network. If we have an image X416 then the feature map will be of size 13 x 13. Detection is made by using the 1 x 1 kernel, and the resultant feature map will be 13 x 13 x 255. The second detection is created by the 94th layer of the model and the resultant feature map will be 26 x 26 x255. Then final detection is made by the 106th layer and yielding Feature map size 52 x 52 x 255

3.3.3 Detecting smaller objects

In model 3 layer has different responsibilities, whereas 13 x 13 layer detects a large object, 52 x 52 layer is responsible for detecting smaller objects with the help of 26 x 26 layer detect medium objects.

3.3.4 Choice of anchor boxes

This model total uses 9 anchor boxes for the detection of an object. We are using k-means clustering to generate 9 anchors. For clustering arrange all anchors in descending order according to the dimensions and assign large anchors for the first scales after three anchors for the second scale, and the last three anchors for the third scale.

This model predicts more bounding boxes. This model predicts boxes at 3 different scales, for the images of 416 x 416, the number of predicted boxes are total 10647

In Class Prediction, Softmax is not used. Independent logistic classifier and binary cross entropy loss are used

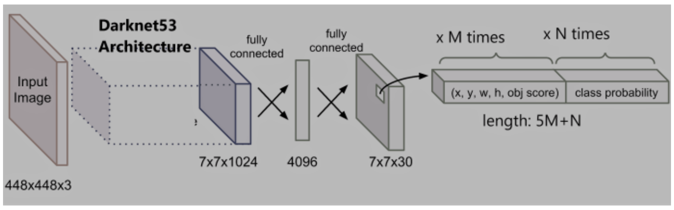

Fig -1: Object Detection model

4. CONCLUSION

This paper provides a detailed review of many of the models in this paper related to object detection such as R-CNN, YOLO SSD, etc. Then we have introduced the limitations of each technology. This proposed model focus on accuracy than speed. The previous models are not accurate when images have small object Small object in images need to be detected.

REFERENCES

[1] The Object Detection Based on Deep Learning Cong Tang 1,2,3 , Yunsong Feng 1,2,3 , Xing Yang 1,2,3 , Chao Zheng 1,2,3 , Yuanpu Zhou 1,2,3 2017 4th International Conference on Information Science and Control Engineering

[2] Moving object detection and tracking Using Convolutional Neural Networks Shraddha Mane Prof.Supriya Mangale Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3

[3] Pedestrian Detection Based on YOLO Network Model Wenbo Lan ; Jianwu Dang ; Yangping Wang ; Song Wang 2018 IEEE International Conference on Mechatronics and Automation (ICMA)

[4] Bones detection in the pelvic area on the basis of YOLO neural network Zuzanna Krawczyk ; Jacek Starzyński 19th International Conference Computational Problems of Electrical Engineering

[5] Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO. 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)

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

[7] You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi in 2016 https://arxiv.org/abs/1506.02640