GALGOTIAS UNIVERSITY
(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Object Identification using Machine Learning Technique

A Report for the Evaluation 3 of Project 2

Submitted by

SHWETA SINGH

(1613101733 / 16SCSE101334)

in partial fulfillment for the award of the degree of

Bachelor of Technology

IN

Computer Science and Engineering

School of Computer Science and Engineering

Under the Supervision of

Dr. J. N. Singh, M.Tech., Ph.D.,
Professor

APRIL / MAY- 2020

TABLE OF CONTENTS

1. Abstract

Effective and precise object detection has been a significant theme in the headway of computer vision frameworks. With the appearance of profound learning procedures, the precision for object detection has expanded definitely. Object Classification is a rule task in picture and video handling. It is practiced over a large number of uses going from test and number arrangement to traffic observation. The crude AI ideas had given the platform to completing umber of picture handling assignments. These days necessity of location calculation is to work start to finish and set aside less effort to register. Constant location and order of items from video give the establishment to creating numerous sorts of systematic viewpoints, for example, the measure of traffic in a specific zone throughout the years or the complete populace in a zone. The assignment for the most part experiences moderate handling of order and identification or the event of incorrect recognition because of the consolidation of little and lightweight datasets. To beat these issues, YOLO (You Only Look Once) based discovery and arrangement approach (YOLOv2) for improving the calculation and preparing speed and at the equivalent time productively recognize the articles in the video. Classifier, for example, Haar course which utilizes Haar like highlights was crudely utilized for face identification. Besides, due to the ever-expanding request furthermore, extent of progress in the current fields, the crude techniques need a great deal of upgrade. Neural Networks have made the undertakings very plain cruising. Directly from the vanilla neural systems to Fast R-CNN and afterward Faster R-CNN, all models have contributed essentially in the space of PC vision.

The project expects to consolidate best in class procedure for object recognition with the objective of accomplishing high precision with a real time performance. A significant test in huge numbers

of the object detection frameworks is the reliance on other computer vision systems for helping the profound learning based methodology, which prompts moderate and non-ideal execution. In this project, we use a completely deep learning based approach to solve the problem of object detection in an end-to-end fashion. The system is prepared on the most testing openly accessible dataset (PASCAL VOC), on which an article discovery challenge is led every year. The subsequent framework is quick and exact, along these lines helping those applications which require object detection.

2. Introduction

(i) Overall description

This section portrays the ideas utilized in the undertaking Object Detection utilizing Machine Learning intricately. The venture goes under the domain Machine learning which is part of Artificial Neural Network. Machine learning ideas causes the framework to take in all alone from the encounters it gains, without the obstruction of the outside variables.

The YOLO (You Only Look Once) calculation utilizing Convolution Neural Network is utilized for the detection purpose. It is a Deep Neural Network idea from Artificial Neural Network. Artificial Neural Network is roused by the biological concept of Nervous System where the neurons are the hubs that structure the system. Thus, in Artificial Neural Network perceptrons act like the hubs in the system. Artificial Neural Network has three layers that are, Input Layer, Hidden Layer and the yield Layer. Profound Learning is the piece of the Artificial Neural Network that has different Hidden Layer that can be utilized for the Feature Extraction and Classification purposes.

(ii) Purpose

Convolution Neural Network (CNN) is the piece of Deep Learning that is utilized in examination of visual imagery. It has four various types of layers, they are, Convolution Layer, Pooling Layer, Activation Layer and Fully Connected Layer. Convolution Layer utilizes channel and walks to get the Feature Maps. These Feature Maps are the framework that is gotten after the Convolution Layer. It tends to be improved utilizing ReLU (Rectified Linear Unit) that maps negative qualities

to 0.The came about Feature Map is diminished by sending it into the Pooling Layer where it is decreased to the littler estimated lattice. This is the manner by which the highlights are separated. Toward the finish of the convolution neural system is the Fully Connected Layer where the real Classification happens.

(iii) Motivations and scope

In Artificial Neural Network perceptrons act like the hubs in the system. Artificial Neural Network has three layers that are, Input Layer, Hidden Layer and the yield Layer. Profound Learning is the piece of the Artificial Neural Network that has different Hidden Layer that can be utilized for the Feature Extraction and Classification purposes. The crude AI ideas had given the platform to completing umber of picture handling assignments. These days necessity of location calculation is to work start to finish and set aside less effort to register. Constant location and order of items from video give the establishment to creating numerous sorts of systematic viewpoints, for example, the measure of traffic in a specific zone throughout the years or the complete populace in a zone. The assignment for the most part experiences moderate handling of order and identification or the event of incorrect recognition because of the consolidation of little and lightweight datasets.

3. Literature Survey

Real-time detection and classification of objects from video give the establishment to producing numerous sorts of expository perspectives, for example, the measure of traffic in a specific region throughout the years or the absolute populace in a territory. By and by, the assignment normally experiences moderate handling of characterization and discovery or the event of mistaken identification because of the consolidation of little and lightweight datasets. To beat these issues, YOLO (You Only Look Once) based discovery and order approach (YOLOv2) for improving the calculation and preparing speed and simultaneously productively recognize the items in the video. Classifier, for example, Haar course which utilizes Haar like highlights was crudely utilized for face discovery. In addition, due to the ever-expanding request and extent of progress in the current

fields, the crude strategies need a great deal of upgrade. Neural Networks have made the assignments very plain cruising. Directly from the vanilla neural systems to Fast R-CNN and afterward Faster R-CNN, all models have contributed altogether in the area of PC vision. This paper for the most part centers in discovery and arrangement running from single class articles to multi class objects. The Haar course classifier was prepared on a bunch of positive and negative examples which were later sewed together to frame a vector document lastly structure the xml record. Then again, COCO dataset utilized for executing YOLOv2 and R-CNN calculation because of the nearness of related model in it. Moreover, utilization of GPU (Graphics Processing Unit) to speed up and forms at 40 edges for every second.

4. Problem Statement

The project "Object Detection System using Machine Learning Technique" distinguishes objects effectively dependent on YOLO calculation and applying the calculation on image information and video information to detect objects. Numerous issues in computer vision were immersing on their precision before 10 years. Be that as it may, with the ascent of deep learning techniques, the exactness of these issues radically improved. One of the serious issue was that of image identification, which is predicting the class of the image.

5. Proposed Model

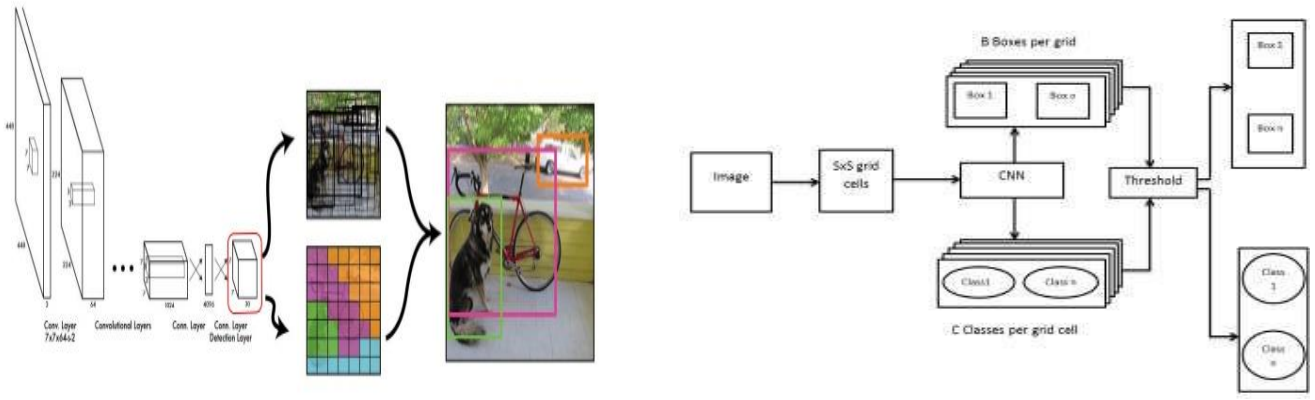

Fig-1: Data Flow Diagram of the System and YOLO Architecture

The Figure shows the Architecture Diagram of the Proposed YOLO Model. Images are given as the input to the system .If Video can also be taken as input as it is nothing but a stream of images. As the name recommends You Only Look Once, the information experiences the input goes through the

network only once and the consequence of distinguished object with Bounding Boxes and Labels are obtained. The images are separated into SXS grid cells before sending to the Convolution Neural Network (CNN). B Bounding boxes per grid are created around all the distinguished articles in the picture as the consequence of the Convolution Neural Network. Then again, the Classes to which the articles have a place is additionally ordered by the Convolution Neural Network, giving C Classes per grid. At that point a limit is set to the Object Detection. Right now have given a Threshold of 0.3. Lesser the Threshold esteem, progressively number of bouncing boxes will show up in the yield bringing about the ungainly yield. When the info is chosen Preprocessing is done, where the SXS grids are framed. The resultant in this manner framed with the matrices is send to the Bounding Box Prediction process where the Bounding Boxes are drawn around the distinguished items. Next the outcome from the past procedure is sent to the Class Prediction where the Class of the item to which it has a place is anticipated. At that point it is sent to the identification procedure where a Threshold is set so as to lessen awkwardness in the yield with many Bounding Boxes and Labels in the last Output. Toward the end an image or a surge of images are created for image and video or camera input individually with Bounding Boxes and Labels are obtained as the Output.

6. Implementation

Following is the algorithm for detecting the object in the Object Detection System.

1. The input image is divided into SxS grid.

2. For every cell it predicts B bounding boxes. Each bounding box contains five components: (x, y, w, h) and a crate certainty score

3. YOLO identifies one article for every grid cell just paying little mind to the number bounding boxes

4. It predicts C conditional class probabilities.

5. In the event that no items exists, at that point certainty score is zero Else certainty score ought to be more prominent or equivalent to edge esteem

6. YOLO then draws bouncing box around the identified articles and predicts the class to which the item has a place.

7. Results and Analysis

We utilized pre-trained dataset of COCO which had 80 classes. The reason behind 80 classes is that because a greater number of classes resulted in the incompleteness of the data. Following area will depict the distinctive Test Cases and the outcomes acquired.

7.1 Test Cases

The following table indicates the different Test Cases, the Expectations well as the Test Result.

Test Case ID	Test Conditions	Expected Result	Test Results
TC1	When image is considered as input	Image with bounding box along with the objects and predicted class	SUCCESSFUL
TC2	When video is considered as input	Video with bounding box along with the objects and predicted class	SUCCESSFUL
TC3	When camera is considered as input	Objects identified in the real time with bounding box, confidence score and predicted class	SUCCESSFUL
TC4	When black and white image is considered as input	Image with bounding box along with the objects and predicted class	SUCCESSFUL
TC5	Image withimage objects is considered asinput	Image with detected objects	UNSUCCESSFUL
TC6	When image with overlapping objects is	Image with bounding box around the objects and	SUCCESSFUL

	considered as input	predicted class	
TC7	When image with distant objects is considered as input	Image with detected objects	UNSUCCESSFUL

7.2 Results

This part shows different results that were obtained by giving various Test Cases described above.

Fig -2: **Image with Detected Object**

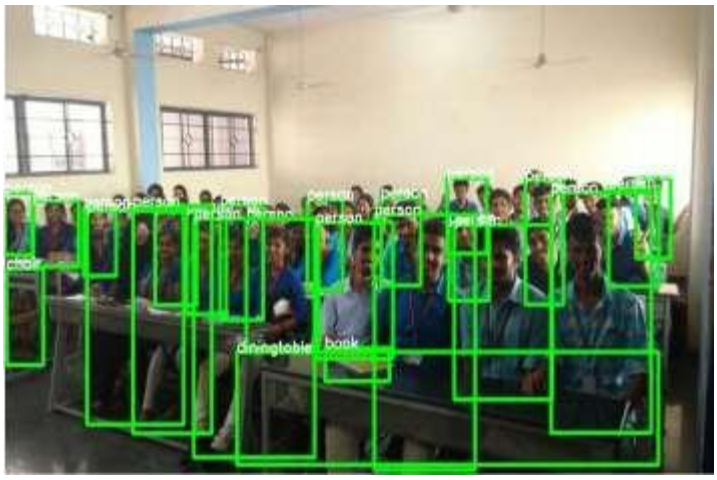

Fig-3:Image with Overlapping Objects

Fig-2 outlines the yield of the Object Detection System. Jumping Boxes are drawn around the Objects distinguished. Fig-3 delineates the yield got when items are covering. This shows halfway noticeable articles will likewise be identified by drawing bouncing box around it

alongside the name demonstrating the class to which it has a place. In theFig-3 a few people are in part obvious in the picture of a jam-packed study hall. The framework can recognize each individual noticeable in the picture.

Fig -4: Output obtained with Video Input

The output created when Video is given as the information is appeared in Fig-4. The video that will be given as contribution to the framework ought to be in .avi position. Fig-5 represents the yield when Camera is utilized to recognize the article. Fig-6 shows the yield created when a haze picture is given as the info. Arbitrary jumping boxes are drawn with no distinguished item. This is one of the drawbacks of the project which gives ineffective test outcome.

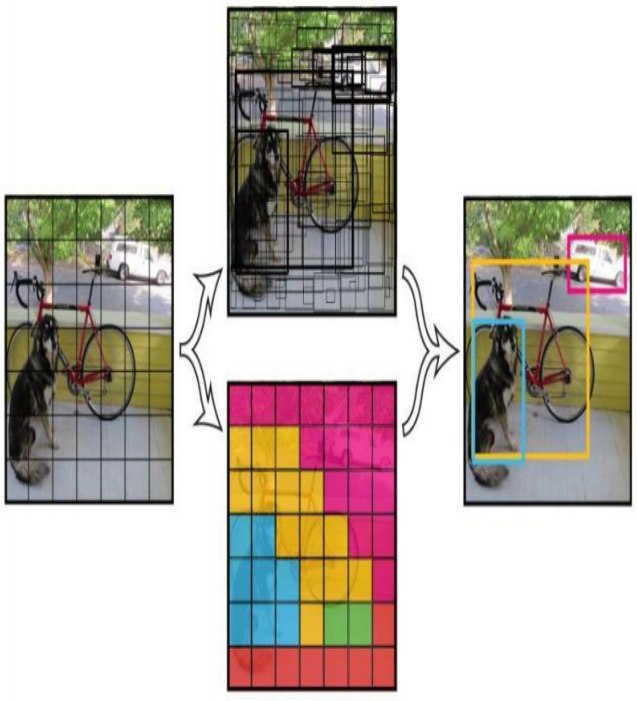

Fig-5: Unified Method

Fig -6: Output obtained in Real-Time

8. Conclusions and Future Work

The project is created with target of recognizing continuous objects in image, video and camera. Bounding Boxes are drawn around the detected articles alongside the mark showing the class to which the item has a place. We have utilized CPU for the handling in the undertaking. Future improvements can be centered by executing the venture around the framework having GPU for quicker outcomes and better precision.

9. References

[1] Lele Xie, Tasweer Ahmad, Lianwen Jin , Yuliang Liu, and Sheng Zhang ,“ A New CNN-Based Method for MultiDirectional Car License Plate Detection”, IEEE Transactions on Intelligent Transportation Systems, ISSN (e): 1524-9050, Vol-19, Issue-02, Year-2018, pp. 507-517.

[2] L. Carminati, J. Benois-Pineau and C. Jennewein, “Knowledge-Based Supervised Learning Methods in a Classical Problem of Video Object Tracking”, 2006 International Conference on Image Processing, Atlanta, GA, USA, ISSN (e): 2381-8549, year-2006.

[3] Jinsu Lee, Junseong Bang and Seong-II Yang, “Object Detection with Sliding Window in Images including Multiple Similar Object”, 2017 IEEE International Conference on Information and

Communication Technology Convergence (ICTC), Jeju, South Korea, ISBN (e): 978-1-5090-4032-2, December-2017.

[4] Qichang Hu, Sakrapee Paisitkriangkrai, Chunhua Shen, Anton van den Hengel and Faith Porikli,"Fast Detection of Multiple Objects in Traffic Scenes with Common Detection Framework", IEEE Transactions on Intelligent Transportation Systems, ISSN (e): 1558-0016, Vol-17, Issue-04, Year-2016, pp. 1002-1014.

[5] Haihui Xie, Quingxiang Wu and Binshu Chen, "Vehicle Detection in Open Parks Using a Convolutional Neural Network", 20015 6th International Conference on Intelligent Systems Design and Engineering Applications(ISDEA), Guiyang, China, ISSN (e): 978-14673-9393-5, August-2015.

[6] Malay Shah and Rupal Kapdi, "Object Detection using Deep Learning Networks", 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, ISBN (e): 978-1-53862745-7, June-2017.

[7] Shoji Kido, Yasusi Hirano and Noriaki Hashimoto, "Detection and classification of lung abnormalities by use of convolutional neural network(CNN) and regions with CNN features(R-CNN)", 2018 International Workshop on Advanced Image Technology, Chiang Mai, Thailand, ISBN (e): 978-15386-2615-3, December-2017.

[8] Kai Han, Muyi Sun, Xiaoguang Zhou, Gaunhong Zhang, Hao Dang and Zhicai Liu, "A New Method in Wheel Hub Surface Defect Detection Algorithm Based on Deep Learning", 2017 IEEE International Conference on Advanced Mechatronic Systems, Xiamen, China, ISBN (e): 978-1-5386-2602-3, December-2017.

[9] Syed Mazhar Abbas and Dr. Shailendra Narayan Singh, "Region based Object Detection and Classification using Faster R-CNN", 2018 4th International Conference on Computational Intelligence and Communication Technology(CICT), Ghaziabad, India, ISBN (e): 978-15386-0886-9, February-2017.

[10] Jinzhuo Wang, Wenmin Wang and Wen Gao, "Multiscale Deep Alternative Neural Network for Large Scale Video Classification", IEEE Transactions on Multimedia, ISSN (e): 1941-0077, Vol-20, Issue-10, Oct-2018, pp. 25782592.

[11] Widodo Budiharto, Alexander A S Gunawan, Jarot S. Suroso, Andry Chowanda, Aurello Patrik

and Gaudi Utama, “Fast Object Detection for Quadcopter Drone using Deep Learning”, 2018 3rd International Conference on Computer and Communication System(ICCCS), Nagoya, Japan, ISBN (e): 978-1-5386-6349-3, April2017.

[12] Pushkar Shukla, Beena Rautela and Ankush Mittal, “A Computer Vision Framework for Automatic Description of Indian Monuments”, 2017 13th International Conference on Signal Image Technology and InternetBased Systems(SITIS), Jaipur, India, ISBN (e): 978-15386-4283-2, December-201.

[13] M. Buric, M. Pobar and Ivasic-Kos, “Object Detection in Sports Videos”, 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics(MIPRO), Opatija, Croatia, ISBN (e): 978-953-233-095-3, May-2018.

[14] Yin-Lon Lin, Yu-Min Chiang and Hsiang-Chen Hsu, “Capacitor Detection in PCB Using YOLO Algorithm”, 2018 International Conference on System Science and Engineering(ICSSE), New Taipei City, Taiwan, ISBN (e): 978-1-5386-6285-4, December-2017

[15] Romain Vial, Hongyuan Zhu, Yonghong Tian and Shijian Lu “Search Video Action Proposal with Recurrent and Static YOLO”,2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, ISBN (e): 978-15090-2175-8, September-2017.