Object Detection using Deep Learning with OpenCV and Python

Shreyas N Srivatsa[1], Amruth[2], Sreevathsa G[3], Vinay G[4], Mr. Elaiyaraja P[5]

[1-4]Student, Dept. of Computer Science Engineering, Sir MVIT, Karnataka, India
[5]Professor, Dept. of Computer Science Engineering, Sir MVIT, Karnataka, India

---***---

Abstract - *Computer Vision is a field of study that helps to develop techniques to recognize images and displays. It has different features like image recognition, object detection and image creation, etc. Object detection is used in face detection, vehicle detection, web images, and safety systems.*

The Objective is to distinguish of objects utilizing You Only Look Once (YOLO) approach. This technique has a few focal points when contrasted with other object detection algorithms. In different algorithms like Convolutional Neural Network, Fast-Convolutional Neural Network the algorithm won't take a gander at the image totally yet in YOLO the algorithm looks the image totally by anticipating the bounding boxes utilizing convolutional network and the class probabilities for these boxes and identifies the image quicker when contrasted with different algorithms.

Using these techniques and algorithms, based on deep learning which is also based on machine learning require lots of mathematical and deep learning frameworks understanding by using dependencies such as OpenCV we can detect every single object in image by the area object in a highlighted rectangular box and recognize every single object and assign its tag to the object. This additionally incorporates the exactness of every strategy for distinguishing objects.

Key Words: **YOLO, Convolution neural network (CNN), Fast-CNN, OpenCV**

1. INTRODUCTION

Object detection is perhaps the main exploration research in computer vision. Object detection is a technique that distinguishes the semantic objects of a specific class in digital images and videos. One of its real time applications is self-driving vehicles or even an application for outwardly hindered that identifies and advise the debilitated individual that some object is before them. Object detection algorithms can be isolated into the conventional strategies which utilized the method of sliding window where the window of explicit size travels through the whole image and the deep learning techniques that incorporates YOLO algorithm. In this, our point is to distinguish numerous objects from an image. The most well-known object to identify in this application are the animals, bottle, and people. For finding the objects in the image, we use ideas of object localization to find more than one object in real time. There are different techniques for object identification, they can be separated into two classifications, initial one is the algorithms dependent on Classifications. CNN and RNN go under this classification. In this classification, we need to choose the

interested areas from the image and afterward need to arrange them utilizing Convolutional Neural Network. This strategy is slow as we need to run an expectation for each selected area. The subsequent class is the algorithms dependent on Regressions. YOLO strategy goes under this classification. In this, we won't need to choose the interested regions from the image. Rather here, we predict the classes and bounding boxes of the entire image at a single run of the algorithm and afterward distinguish different objects utilizing a single neural network. YOLO algorithm is quicker when contrasted with other grouping algorithms. YOLO algorithm makes localization errors but it predicts less false positives in the background.

This document is template. We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace(copy-paste) the content with your own material. Number the reference items consecutively in square brackets (e.g. [1]). However, the authors name can be used along with the reference number in the running text. The order of reference in the running text should match with the list of references at the end of the paper.

2. LITERATURE SURVEY

In the year 2017 Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, BharathHariharan, and Serge Belongie proposed Feature Pyramid Networks for Object Detection. With the launch of Faster-RCNN, YOLO, and SSD in 2015, it seems like the overall structure an object identifier is resolved. Analysts begin to take a gander at improving every individual pieces of these networks. Highlight Pyramid Networks is an endeavor to improve the identification head by utilizing highlights from various layers to frame a feature pyramid. This feature pyramid thought isn't novel in computer vision research. In those days when highlights are still physically planned, feature pyramid is now a powerful method to recognize patterns at various levels. Utilizing the Feature Pyramid in deep learning is likewise not a groundbreaking thought: SSPNet, FCN, and SSD all showed the advantage of aggregating multiple layer highlights before classification. Nonetheless, how to share the feature pyramid among RPN and the region-based detector is still yet to be resolved.

In the year 2017 Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick proposed Mask R-CNN.In this paper Mask R-CNN is certainly not a commonplace object detection network. It was intended to settle a difficult example division task, i.e, making a mask for each object in the scene.

Nonetheless, Mask R-CNN indicated an incredible augmentation to the Faster R-CNN framework, and furthermore thusly motivated object location research. The fundamental thought is to add a binary mask prediction branch after ROI pooling alongside the current bounding box and characterization branches. Obviously, both perform multiple tasks preparing (division + detection) and the new ROI Align layer add to some improvement over the bounding box benchmark.

In the year 2017 NavaneethBodla, Bharat Singh, Rama Chellappa, Larry S. Davis proposed Soft-NMS – Improving Object Detection with One Line of Code. In this paper Non-maximum suppression (NMS) is broadly utilized in anchor-based object detection networks to diminish copy positive proposition that are close-by. All the more explicitly, NMS iteratively wipes out applicant boxes on the off chance that they have a high IOU with a surer applicant box. This could prompt some sudden conduct when two objects with a similar class are to be sure near one another. Soft NMS rolled out a little improvement to just downsizing the certainty score of the overlapped applicant boxes with a boundary. This scaling boundary gives us more control when tuning the localization execution, and furthermore prompts a superior exactness when a high review is likewise required.

In the year 2017 ZhaoweiCai UC San Diego, Nuno Vasconcelos UC San proposed Cascade R-CNN: Delving into High Quality Object Detection. While FPN investigating how to plan a superior R-CNN neck to utilize backbone highlights Cascade R-CNN examined an upgrade of R-CNN grouping and regression head. The basic assumption that is straightforward yet sagacious: the higher IOU rules we utilize while planning positive focuses on, the less false positive predictions the network will figure out how to make. In any case, we can't just increment such IOU threshold from regularly utilized 0.5 to more forceful 0.7, in light of the fact that it could likewise prompt all the more overpowering negative models during training. Cascade R-CNN's answer is to chain various recognition head together, each will depend on the bounding box recommendations from the past detection head.

In the year 2017 Tsung-Yi Lin PriyaGoyal Ross GirshickKaiming He Piotr Dollar proposed Focal Loss for Dense Object Detection. To comprehend why one-stage locators are typically not comparable to two-stage detectors, RetinaNet explored the frontal area foundation class unevenness issue from a one-stage detectors dense predictions. Take YOLO for instance, it attempted to predict classes and bounding boxes for all potential areas meanwhile, so the majority of the yields are coordinated to negative class during training. SSD tended to this issue by online hard model mining. YOLO utilized an objectiveness score to certainly prepare a closer view classifier in the beginning phase of training. RetinaNet thinks the two of them didn't get the way in to the issue, so it developed another loss function work called Focal Loss to assist the network with realizing what's significant. Focal Loss added a

power γ to Cross-Entropy loss. The α boundary is utilized to adjust such a focusing effect.

In the year 2018 Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, JiayaJia proposed Path Aggregation Network for Instance Segmentation. In this paper Occurrence division has a close relationship with object detection, so regularly another case segmentation network could likewise profit object recognition research in a roundabout way. PANet targets boosting data stream in the FPN neck of Mask R-CNN by adding an extra base up path after the first top-down path. To picture this change, we have a ↑↓ structure in the first FPN neck, and PANet makes it more like a ↑↓↑ structure prior to pooling highlights from various layers. Likewise, rather than having separate pooling for each element layer, PANet added an "adaptive feature pooling" layer after Mask R-CNN's ROIAlign to merge multi-scale features.

In the year 2018 ChengjiLiu, Yufan Tao, JiaweiLiang, Kai Li, Yihang Chen proposed Object Detection Based on YOLO Network. In this paper YOLO v3 is the latest form of the YOLO versions. Following YOLO v2's convention, YOLO v3 acquired more thoughts from past exploration and got a powerful incredible one-stage finder like a beast. YOLO v3 adjusted the speed, exactness, and execution unpredictability really well. Also, it got truly mainstream in the business as a result of its quick speed and basic parts. Basically, YOLO v3's success comes from its all the more impressive backbone include extractor and a RetinaNet-like identification head with a FPN neck. The new spine network Darknet-53 utilized ResNet's skip connections with accomplish a precision that is comparable to ResNet-50 yet a lot quicker.

In the year 2020 Mingxing Tan, Ruoming Pang, Quoc V Le proposed EfficientDet: Scalable and Efficient Object Detection. In this paper EfficientDet indicated us some all the more energizing advancement in the object detection area. FPN structure has been end up being an amazing technique to improve the identification network performance for objects at various scales. Popular detecting network, for example, RetinaNet and YOLO v3 all received a FPN neck before box regression and arrangement. Afterward, NAS-FPN and PANet both showed that a plain multi-layer FPN structure may profit by more plan enhancement. EfficientDet kept investigating toward this path, in the end made another neck called BiFPN. Essentially, BiFPN highlights extra cross-layer associations with energize include aggregation to and fro. To legitimize the proficiency part of the network, this BiFPN additionally eliminated some fewer valuable associations from the first PANet plan. Another creative improvement over the FPN structure is the weight feature fusion. BiFPN added extra learnable loads to highlight aggregation so the network can get familiar with the significance of various branches. Besides, much the same as what we found in the image characterization network EfficientNet, EfficientDet likewise acquainted a principled path with scale an object identification network. The φ parameter in the above formula controls both width

(channels) and depth (layers) of both BiFPN neck and detection head.

3. METHODOLOGY

3.1 YOLO Loss function:

The loss function plays a major role in reducing the error in prediction of the framework. If we take the single grid then, it predicts many bounding boxes and in the process of algorithm of the loss we make use of one of the bounding boxes for specified objects the process of choosing the bounding box depends upon the greater value of IoU. There various available loss functions such as Classification, Confidence and Localization losses.

Where, Localization loss is for the error between the ground truth values and deduced value, it is the quantifying of errors in the deduced boundary boxes locations and the dimension measure, box which is in charge for the object is the only considered. Confidence loss is a measure of how sure is the model about the object detected belonging to that class. Classification loss is the standard squared error of class category probabilities.

3.2 Finding Bounding Box of an Object:

In the Classification and Localization, the data normally that comes out of the framework in a presentable general way as (X, y). bx, by, bw and bh [7] as shown in Figure 4 below, where,

Where,

X = input image data matrix,

y = is an array of all the class labels that corresponds to image X,

bx = in the detection's box the x coordinate,

by = in the detection's the y coordinate,

bw = in the detection's the width,
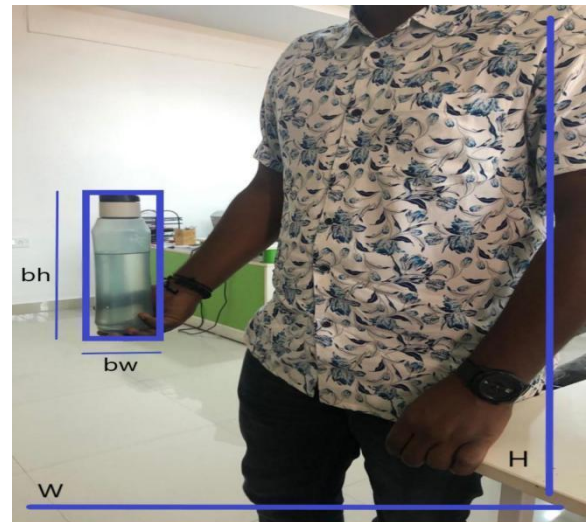
bh = in the detection's the height,

Figure 1: Finding the width of an object

The image is divided into boxes to do object localization tasks so the convent's in place here. Then a different output layer will be responsible to predict the bounding box coordinates and do the required alterations to the loss function. Then the input image is passed on in the pipeline to the framework which then divides into grids in a single pass. The process of Image objects classification and determination of object location on each of the grids present. Then predicting the rectangular bounding box and its corresponding class Id and class probability for objects in the box [5].

If there is an object located in a grid, it will take the midpoint of the grid where there are objects and that corresponding detection data would be put to the grid which consists of the center point of the detected objects and their class ID, names for the middle grid will be assigned. Even in some cases if an object might be present in multiple grids, it will only be put to a single grid which are good strong confidence in which its midpoint is located. X coordinate of the detection's box and y coordinate of the detection's box will always lie in between of 0 and 1 both inclusive as the middle point will always be present inside of the grids, but width of detection's box and height of detection's box can exceed 1 in some-cases, when the measurements of the rectangle or bounding box are exceeding the dimensions of the grids.

4. CONCLUSIONS

In this paper, we have applied and proposed to utilize YOLO algorithm for object recognition in light of the fact that of its favorable circumstances. This algorithm can be actualized in different fields to tackle some real-life issues like security, checking roadways or in any event, helping outwardly debilitated people with help of input. In this, we have made a model to distinguish different number of objects.

REFERENCES

[1] Chengji Liu1, Yufan Tao1, Jiawei Liang1, Kai Li1, Yihang Chen1 "Object Detection Based on YOLO Network" in 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), https://ieeexplore.ieee.org/document/8851911.

[2] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia," Path Aggregation Network for Instance Segmentation" in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/document/8579011

[3] Kaiming He, Georgia, Gkioxari Piotr, Dollar Ross, Girshick, "Mask R-CNN" in 2017 IEEE International Conference on Computer Vision (ICCV), https://ieeexplore.ieee.org/document/8237584

[4] Zhaowei Cai, Nuno Vasconcelos, "Cascade R-CNN: Delving into High Quality Object Detection" in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/document/8578742/authors#authors

[5] Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis, "Soft-NMS – Improving Object Detection with One Line of Code" in 017 IEEE International Conference on Computer Vision (ICCV), https://ieeexplore.ieee.org/document/8237855

[6] Liguang Yan, Baojiang Zhong Weigang Song, "REGION-BASED FULLY CONVOLUTIONAL NETWORKS FOR VERTICAL CORNER LINE DETECTION" in 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), https://ieeexplore.ieee.org/document/8266465

[7] Koen E. A. van de Sande, Jasper R. R. Uijlingst, Arnold W. M. Smeulders, "Segmentation as Selective Search for Object Recognition" in 2011 International Conference on Computer Vision, https://ieeexplore.ieee.org/document/6126456

[8] Mingxing Tan, Ruoming Pang, Quoc V. Le, "EfficientDet: Scalable and Efficient Object Detection" in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/9156454

[9] Andrew Edie, Johnson and Martial Hebert, "Recognizing Objects by Matching Oriented Points" in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/abstract/document/609400

[10] Tsung-Yi, Lin Priya Goyal, Ross Girshick Kaiming, He Piotr Dollar, "Focal Loss for Dense Object Detection" in 2017 IEEE International Conference on Computer Vision (ICCV), https://ieeexplore.ieee.org/document/8237586

[11] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger" in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/8100173

[12] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature Pyramid Networks for Object Detection" in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/8099589

[13] Chengcheng Ning, Huajun Zhou, Yan Song, linhui Tang, "INCEPTION SINGLE SHOT MULTIBOX DETECTOR FOR OBJECT DETECTION" in 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), https://ieeexplore.ieee.org/document/8026312

[14] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection" in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://ieeexplore.ieee.org/document/7780460

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" in IEEE Transactions on Pattern Analysis and Machine Intelligence, https://ieeexplore.ieee.org/document/7485869