International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 08 – Issue 06 || June 2023 || PP. 22-26

A study on Object Detection using Deep Learning

Prof. Hrishikesh Mogare[1], Prof. P. V. Mitragotri[2], Mr. Sujit Kumbhar[3]
[1](Department of MCA, KLS Gogte Institute of Technology, VTU, INDIA)
[2](Department of MCA, KLS Gogte Institute of Technology, VTU, INDIA)
[3](Department of MCA, KLS Gogte Institute of Technology, VTU, INDIA)

Abstract: The paper, a study on object detection using deep learning allows us to detect and label the objects present in the frame. There are various categories available like pedestrians, Two-Wheeler, Three-Wheeler, Car, Standard Truck, Minibus, Transport vehicle etc. which aims to annotate the objects present in the frame automatically with the bounding boxes along with respective label of category which is makes the task easy for the data scientists, researchers to build the any AI model in efficient way. The primary users of the model are the data scientists, researchers, and the education purposes.

Keywords: Convolutional Neural Network, Fast R-CNN Faster R-CNN YOLO (You Only Look Once) SSD (Single Shot Detector)

1. Introduction

Computer vision is a branch of AI that helps machines to collect and comprehend information from media files. Machines can group items and react by employing machine learning (ML) techniques to media, such as identifying the vehicles in data from traffic [1].ML is a branch of AI which trains machines to think like human beings while feeding an ML algorithm with dataset to predict future outcomes. Detection of objects is a huge, lively and dense area of computer vision. Picture identification, object detection, image synthesis, image super-resolution, and many other aspects of computer vision are included. In this paper, we're employing SSD ResNet50 V1 FPN 640x640 (RetinaNet50) which is very accurate object detection algorithm and approach. We would be able to detect every object in an image using appropriate algorithms, which are derived on deep learning and machine learning. Dependencies such as Tensor Flow, OpenCV etc help in identifying each and every object and assign a tag to the object. As we have used the SSD model, SSD dependent on predetermined regions and grid points on the input image at each anchor point. SSD draws perfect boundary box also when the objects are numerous and even if they overlap each other. SSD is mainly built for the real application that is the reason why it is widely being used.

2. Convolutional Neural Network (CNN)

To accomplish the task of object detection, one of the most widely used artificial neural networks is employed and is known as Convolutional Neural Network (CNN). CNN uses the concept of weight-sharing. Convolution is a technique to help in integration which shows function overlapping. The layered architecture of CNN for object recognition is shown below in the figure 1. Feature maps are created by convulsing the images. This procedure is done twice. The required filters are applied, and feature maps are created as a result. They are eventually processed.

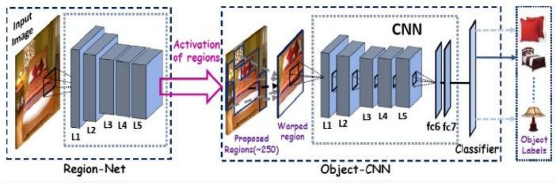

Fig 1: Use of CNN for object detection.

Convolutional neural networks work similar to the classic supervised learning approaches. Input images are taken and their features are detected. Features, on the other hand, are learned automatically. All of the arduous work of extracting and describing features is done by the CNN itself.

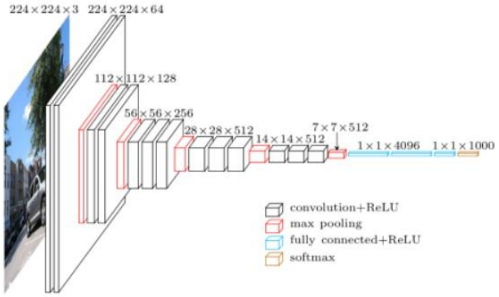

Fig 2: Architecture of CNN

There are multiple layers present in the convolution neural network that help in detecting the objects more precisely. To identify any objects there are several types of image annotation methodologies available

2.1 Bounding boxes: The most frequent type of annotation used in computer vision is the bounding boxes. Bounding boxes are boundaries that identify where the target object is situated as shown in Figure 3. The x and y axis coordinates in the upper-left corner and in the lower-right corner of the rectangle can be used to determine them. Bounding boxes are frequently employed for object recognition and localization tasks. [2]

Fig 3: Bounding Box

2.2 Polygonal: Objects don't necessarily have to be rectangles. Polygonal segmentations, based on this concept, are a sort of data marking in which intricate polygonal shapes are used in place of rectangular shapes to distinguish objects with greater accuracy as shown in Figure 4.

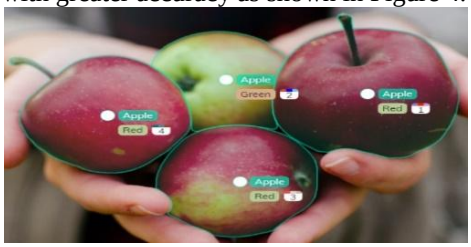

Fig 4: Polygonal labeling

2.3 3D cuboids: Similar to bounding boxes, 3D cuboids contain additional in-depth knowledge about the object. Hence, these 3D cuboids can be used to produce a 3D representation of an object, allowing the system to identify various properties such as its volume and position in a 3D space. An example is shown in Figure 5.

Fig 5: Cuboid Shape

3. Different Methodologies

There are some algorithms available to detect any object.
Fast R-CNN Faster R-CNN YOLO (You Only Look Once) SSD (Single Shot Detector)

3.1 Fast R-CNN: Fast R-CNN, is a coaching methodology for detection. This algorithm deals with flaws of RCNN and SPPNet while increasing its pace and precision. [3]. The architecture is shown below in Figure 6.

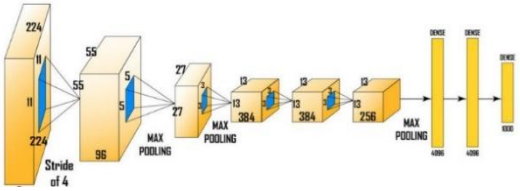

Fig 6: Architecture of R-CNN

3.2 Faster R-CNN: Faster RCNN is a analogous algorithm. This uses RPN which is less expensive than the others for full image convolutional features. An RPN is a fully convolutional network which is trained to foresee limits of objects. [4]

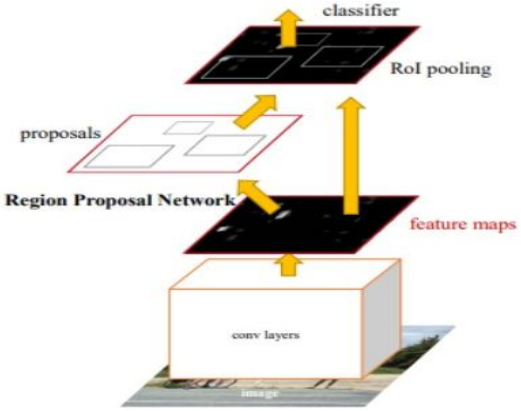

Fig 7: Architecture of Faster R-CNN

3.3 YOLO: YOLO utilizes and different approach. A single Neural Network is employed on the entire image in this approach. It divides the image into sections and generates boundaries and predicts the possibilities. The estimated possibilities are used to evaluate the boundaries.

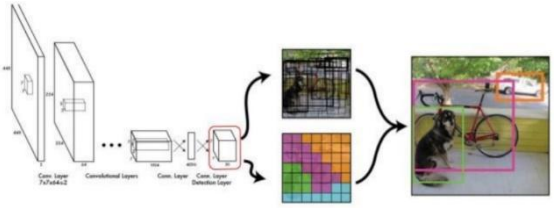

Fig 8: YOLO Object Detection

The YOLO neural network has 24 convolutional layers. Each layer is noteworthy and the layers are characterized by their functionality. The working of YOLO is depicted in Figure 8.

3.4 SSD: SSD is used for live detection of objects. SSD is faster than Faster RCNN because it eliminates the requirement for RPN. [5] These improvements allow SSD to match the accuracy of the Faster

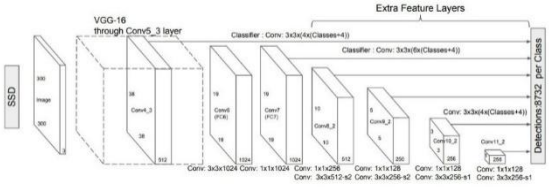

Fig 9: Single Shot Multi Box Detector

The SSD detection of objects comprises of two things:

3.5 Extract feature maps, and Apply convolution filters to detect objects. The SSD training method is derived from the Multi Box approach as shown in Figure 9, but it can handle a variety of object classes. Assume that there is a marker formatching the ith to the jth ground truth box of class. Using the pairing method described above we can write, the weighted sum of the localization loss (loc) and the confidence loss as follows

$$L(x,c,l,g) = \frac{1}{N}\left(L_{\text{conf}}(x,c) + \alpha L_{\text{loc}}(x,l,g)\right),$$

Where N is the number of combined default boxes.

Small item detection relies on higher-resolution feature maps. As a result, as compared to other detection algorithms, SSD typically performs poorly for small objects. If this is a problem, we can improve that using the higher-resolution images.[6]

$$L_{\text{loc}}(x,l,g) = \sum_{i\in\text{Pos}}^{N}\sum_{m\in\{cx,cy,w,h\}} x_{i,j}^{p}\,\text{smooth}_{L1}\left(l_i^m - \hat{g}_j^m\right),$$

$$\hat{g}_j^{cx} = \frac{g_j^{cx} - d_i^{cx}}{d_i^w},$$

$$\hat{g}_j^{cx} = \frac{g_j^{cx} - d_i^{cx}}{d_i^w},$$

$$\hat{g}_j^{w} = \log\left(\frac{g_j^w}{d_i^w}\right),$$

$$\hat{g}_j^{h} = \log\left(\frac{g_j^h}{d_i^h}\right).$$

The loss of confidence is the soft max on multi classes confidence (c)

$$L_{\text{conf}}(x,c) = -\sum_{i\in\text{Pos}}^{N} x_{i,j}^{p}\log\left(\hat{c}_i^p\right) - \sum_{i\in\text{Neg}}\log\left(\hat{c}_i^0\right),$$

Where,

$$\hat{c}_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)},$$

4. Time and Accuracy

The average MAP of each meta–architecture is represented in the scatter plot (Figure 10). The plot shows the average running time of each image. The SSD meta-architecture is faster but is less accurate, whereas the faster R-CNN meta-architectures are more precise but take more time.

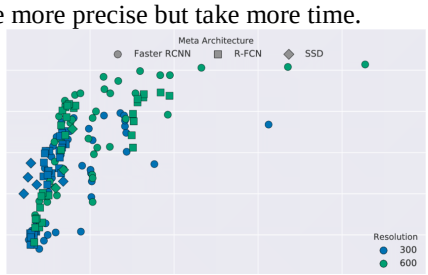

Fig 10: Time and accuracy

5. Conclusion

The first phase in the implementation of autonomous vehicles and bots is detection of objects. In this work the role of several algorithms for object detection is decoded. The paper also discourses the different deep learning frameworks and services available for object detection. Appropriately detecting an object in a video surveillance is an essential aspect in computer vision research. Processing the image obtained from a

surveillance camera is difficult due low image resolution, changing light conditions, moving items in the background, and minor changes in the background such trees etc. An overview of these things have been presented here.

References

[1]. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, ``Microsoft COCO: Common objects in context," in Computer Vision_ECCV, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740_755.

[2]. https://towardsdatascience.com/image-data-labelling-and-annotation-everything-you-need-to-know-86ede6c684b1.

[3]. Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. (2017). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE transactions on pattern analysis and machine intelligence.

[4]. S. Ren, K. He, R. Girshick, and J. Sun, ``Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137_1149, Jun. 2017.

[5]. https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06.

[6]. J. Jeong, H. Park, and N. Kwak, ``Enhancement of SSD by concatenating feature maps for object detection," 2017, arXiv: 1705.09587. [Online]. Available: https://arxiv.org/abs/1705.09587.