IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 3, March 2020

Object Detection using Deep Learning

Sunny Bajpai[1], Sushil Lokhande[2], Sahil Tandel[3], Maya Salve[4]

Student, Computer Technology, Bharati Vidyapeeth, Navi Mumbai, India[1,2,3]

Lecturer, Computer Technology, Bharati Vidyapeeth, Navi Mumbai, India[4]

Abstract: Object detection has recently become one of the popular technology around the world. It is based on Deep learning and neutral network which is a branch of artificial intelligence. Object detection is used to detect instances of semantics objects of certain class like pedestrians, vehicles, animals etc. There are various algorithm's that are used in object detection like CNN (Convolution Neutral Network), R-CNN (Region Based Convolution Neutral Network), faster R-CNN and YOLO(You only look once) Out of all YOLO perform outstanding performances it is fast and it is used in real time. As the name says it only look at image once and provide high accuracy.

Keywords: Object detection, YOLO algorithm, Deep learning, Neutral network

I. INTRODUCTION

Object detection is an computer based technology it is based on Deep learning and neutral network which is sub-branch of Artificial intelligence(AI).Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class these classes can be human, vehicle, road sign, animals etc. Object detection is widely used in real time processing .Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. Object detection is widely used in Image annotation, activity recognition, face detection, face recognition, video object co-segmentation. Object detection is also used in tracking objects specially in real time, For example tracking movement of a Vehicle and pedestrians here are various algorithm based on different approach like CNN which stands for convolution neutral network, R-CNN which is based on region of the image stands for Region convolution neutral network, Faster on R-CNN as the name suggest it is faster version and last one is YOLO stands for You only look once which look at image once.

II. LITERATURE SURVEY

You only look once is a state-of-the-art object detection algorithm which targets real-time applications, and unlike some of the competitors, it is not a traditional classifier purposed as an object detector. YOLO works by dividing the input image into a grid of $S \times S$ cells, where each of these cells is responsible for five bounding boxes predictions that describe the rectangle around the object. It also outputs a confidence score, which is a measure of the certainty that an object was enclosed. Therefore the score does not have any relation with the kind of object present in the box, only with the box's shape For each predicted bounding box, a class it's also predicted working just like a regular classifier giving resulting in a probability distribution over all the possible classes. The confidence score for the bounding box and the class prediction combines into one final score that specifies the probability for each box includes a specific type of object. Given these design choices, most of the boxes will have low confidence scores, so only the boxes whose final score is beyond a threshold are kept and scales. High scoring regions are considered for detections, but in YOLO we use different approach instead of looking image 100 or 1000 times In other word network does not look at the complete image instead parts of image which has high probabilities of containing object, Whereas YOLO look at image only once. We apply a single neutral network to full image, this network divides images into regions and then it predict bounding boxes and class probability. When compared with other algorithms it outperforms them It is extremely fast and It process image in real time in 45 fps (frame per second).

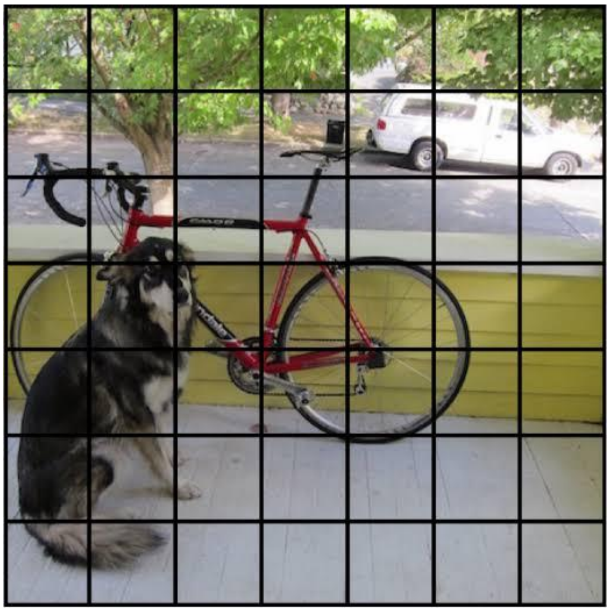
III. PURPOSED WORKING

Object detection is **one in every of** the classical problems in computer vision where you're employed to recognize what and where — specifically what objects are inside a given image and also where they're within the image. The matter of object detection is more complex than classification, which can also recognize objects but doesn't indicate where the thing is found within the image. Additionally, classification doesn't work on images containing over one object.

YOLO is popular because it achieves high accuracy while also having the ability to run in real-time. The algorithm

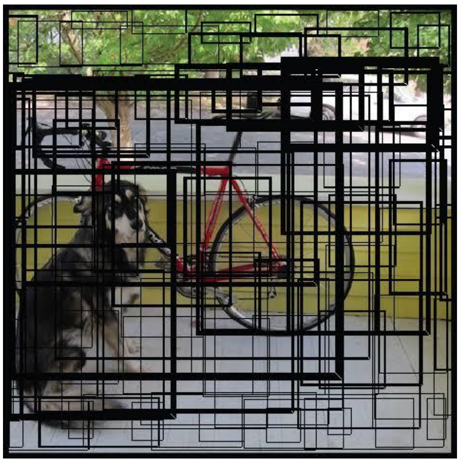
DOI 10.17148/IJARCCE.2020.9308

“only looks once” at the image within the sense that it requires only 1 forward propagation suffer the neural network to form predictions. After non-max suppression (which makes sure the thing detection algorithm only detects each object once), it then outputs recognized objects along with the bounding boxes.



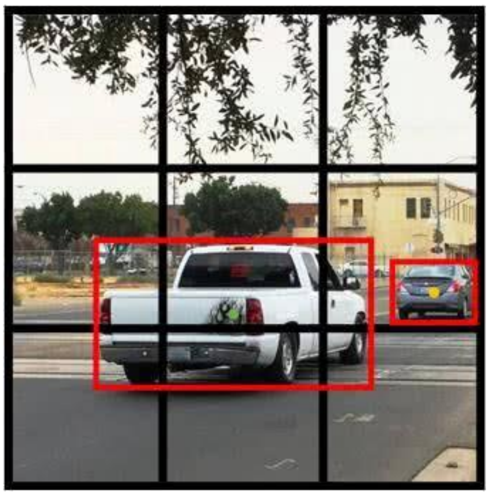
The YOLO model is that the very first attempt at building a quick real-time object detector. Because YOLO does not undergo the region proposal step and only predicts over a limited number of bounding boxes, it is able to do inference super-fast. YOLO divides input image into an S X S grid. We can divide image into any number of grids its depends on the image. Once image is divided into grid than each grid cell is responsible for predicting bounding boxes m. A bounding box describes the rectangle that encloses an object .For example an image may look like this as below As shown each grid is divided into number of grid. Each cell predicts bounding boxes and confidences. Confidences means if there is object or not if confidence score is higher than it means predicted bounding box contain some object. It doesn’t tell what kind of object it is just if shape of box is good. If there is no proper object is found in the grid than the bounding box value will be zero and if proper object is found than bounding value is 1.Each bounding box also predicts a class than we combine them and thus object is predicted.

Confidence score and class are combined which tell us about specific type of object. YOLO only look at image once thus providing faster result compared to other algorithm.



IV. METHODOLOGY

YOLO (You only look once) is one of the useful framework but how it different from other and how it Outperformed other algorithms and how it is easy to understand and use we can understand by an example. Suppose an input image as show below than YOLO framework then divides input into grid (3 X 3 grid) just as example



Than each cell predict bounding boxes and class probability for object. If case there is total 3 class in which we want object to be classified into these classes can be car, any vehicle and pedestrian. For each grid cell Y will be an eight dimensional vector.

$$Y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

As show above Y has 8 unit here each unit represents something. Pc represent whether object is present in grid or not where as bx,by,bh,bw specific bounding box and c1,c2c,3 represent classes as shown in image here car is an class. Suppose there is no object in a grid like shown below than confidence value will be 0 and if it has object it will be represented by 1.

Since in above cell there is no object its pc will be 0 and since there is car present in other cell it pc will be 1 and object is car c2 will be 1 and c1 and c3 will be 0 Whereas bx,by,bh,bw will be calculated relative to this grid.bx,by is x and y coordinates, bh is ratio of height of the bounding box. Each grid can only identify one object in case of multiple objects we use concept of Anchor Boxes. In our example each cell predict confidence and class probability Likewise these happen to all cell present than these are Confidence score and class are combined which tell us about specific type of object. YOLO only look at image once thus providing faster result compared to other algorithm. Another important thing in yolo function is Loss function. YOLO predicts multiple bounding boxes per grid cell. To compute the loss for the true positive, we only want one of them to be responsible for the object. For this purpose, we select the one with the highest IoU (intersection over union) with the ground truth. This strategy leads to specialization among the bounding box predictions. Each prediction gets better at predicting certain sizes and aspect ratios.

V. CONCLUSION

The proposed system is used to detect the object and motion of the object using various algorithms. Out of all algorithms present like R-CNN, CNN, faster CNN YOLO has demonstrated significant performance gains while running at real-time performance. Object detection is based on computer technology, It is used to detect object of various classes. It is based on deep learning and neutral networks which is also branch of artificial intelligence. Using these methods 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 TensorFlow, OpenCV, imageai etc., we can detect each and every object in image by the area object in an highlighted rectangular boxes and identify each and every object and assign its tag to the object. This also includes the accuracy of each method for identifying objects. YOLO has outperformed all other algorithms by providing accurate object and processing image faster compared to other algorithms. YOLO can carry out multi-scale predictions. Comparing to other algorithms this algorithm is efficient and provide relatively less error. As day by day technology is improving object detection is necessary and is very important for robotic.

DOI 10.17148/IJARCCE.2020.9308

VI. ACKNOWLEDGEMENT

I respect and thank **Ms. Maya Salve** for providing us an opportunity to do the project work in Object detection and giving us all support and guidance which made us complete the project duly. I am extremely thankful to her for providing such a nice support and guidance, although she had busy schedule managing the corporate affairs. I owe my deep gratitude to our project guide **Ms Maya salve**, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

REFERENCES

[1]. J. Schmidhuber, "Deep learning in neural networks: An overview,"*Neural networks*, vol. 61, pp. 85–117, 2015. 1

[2]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587. 1

[3]. C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in neural information processing systems*, 2013, pp. 2553–2561. 1

[4]. L.Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik,J.Terwilliger, J. Kindelsberger, L. Ding, S. Seaman *et al.*,"Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behaviour and interaction with automation," *arXiv preprint arXiv:1711.06976*, 2017. 1

[5]. O. Akgul, H. I. Penekli, and Y. Genc, "Applying deep learning in augmented reality tracking," in *Signal-Image Technology & Internet- Based Systems (SITIS), 2016 12th International Conference on*. IEEE, 2016, pp. 47–54. 1

[6]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826. 1

[7]. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *AAAI*, vol. 4, 2017, p. 12. 1

[8]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 1

[9]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv: 1409.1556*, 2014. 1

[10]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779– 788. 1, 2, 6, 7

DOI 10.17148/IJARCCE.2020.9308