IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

A Design Thinking based Object detection Technique using Yolo v5

S. Rajasulochana1, R.K. Naveen kumar2, D. Yogeshwaran3, J. Keethitha4

Assistant Professor, Department of Information Technology, SNS College of Technology, Coimbatore, TamilNadu[1]

UG Scholar, Department of Information Technology, SNS College of Technology, Coimbatore, TamilNadu[2]

UG scholar, Department of Information Technology, SNS College of Technology, Coimbatore, TamilNadu[3]

UG scholar, Department of Information Technology, SNS College of Technology, Coimbatore, TamilNadu[4]

Abstract: Object detection has been used in many of the field now and it has become the main reason for the development of many applications of the auto driving cars, Statistics and etc. In this paper we will see how the YOLO algorithm works and how it is more efficient than other object detection algorithms using the comparison graphs with the various versions of the YOLO algorithm and other algorithms such as Convolutional Neural Networks, Fast-CNN, etc., In this algorithm, the dataset used for object detection can predefined dataset or dataset manually generated according to the use cases. The experimental data has been taken for the testing of the YOLO algorithm and the dataset is trained and tested with given dataset. Here the image is converted into bounding boxes to which a particular value is given so that it is faster in detecting the images than other object detecting algorithms.

Keywords: Object detection, Fast – Convolution Neural Network, Bounding Boxes, YOLO

I. INTRODUCTION

Object detection is a recent technologythat detects the objects or real time entity indigital images and videos. One of its real-time applications is self-driving cars. The task of this algorithm is to detect multiple objects from an image. The most common real time entity to detect in thisapplication is the car, motorcycle, and person etc., and for locating the objects in the image object localization.

The object detection techniques are classified into two types as object detection based on Classification, such as CNN, Fast R-CNN and the object detection based on Regression. YOLOmethod falls under the second category. In YOLO algorithm, the regions are not selected from the image. Instead, bounding boxes of the whole image are predicted ata single time and multiple objects are detected using a single neural network. YOLOalgorithm is as fast as compared to other classification algorithms for object detection. In real time, the algorithmcan process 40 to 45 frames per second. YOLO algorithm makes object localization errors but predicts less false positives in the background during object detection.

A. Object detection

Object detection is a computervision technique that focuses on identifying and labelling the objects within the images, videos or even live videos. Object detection models are trained with a huge amount of pre-defined visuals in order to carry out this process with new data. It then becomes as simple as giving input (i.e.,images) and receiving a fully marked-up output visual. A key component being the object detection bounding box. They are tagged by a label of the object, whether it be a person, a car, or a dog to describe the target object. By tagging, the neural network is trained to detect and obtain the idea of what the object actually is so it can identify the object or particular product.

B. Object detection using neural networks

The employment of the neural networks has achieved significant performance than the deep learning.The neural networks mimic the neural architecture of the human brain. They consist of an input layer, hidden inner layers, and an output layer. The learning for these neural networks can be supervised, semi-supervised, and unsupervised according to the amount of annotated data. Deep neural networks for object detection is the quickest and most accurate results for single and multiple object detection.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

II. METHODS AND ALOGORITHMS FOR OBJECT DETECTION

Object detection cannot be done by designing a model for handling the task. These object detection models are trained with hundreds of thousands of visual content to increase or optimize the detection accuracy. Training and refining models are made by creating a dataset for a given field of the environment or through help of readily available datasets like COCO (Common Objects in Context) to help give you a headstart in scaling your annotation pipeline.

Some of the known algorithms are:

1. R-CNN
2. Fast R-CNN
3. Faster R-CNN

A. R-CNN

The first largely successful family of methods was R-CNN (Region-Based Convolutional Neural Network). It surpassed its previous versions 2,000 regions from the image, which were referred to as region proposals, instead of an exceedingly large number of regions prior to thisthe input image is selected, of which 2,000 region proposals are extracted. Next, the features would be extracted from each individual region, which would then go on to be classified as one of the known classes. The primary shortcoming of R-CNN lies in the fact that although it extracted 2,000 region proposals, it was nonetheless a lengthy process. This paved the way to the new and improved Fast R-CNN.

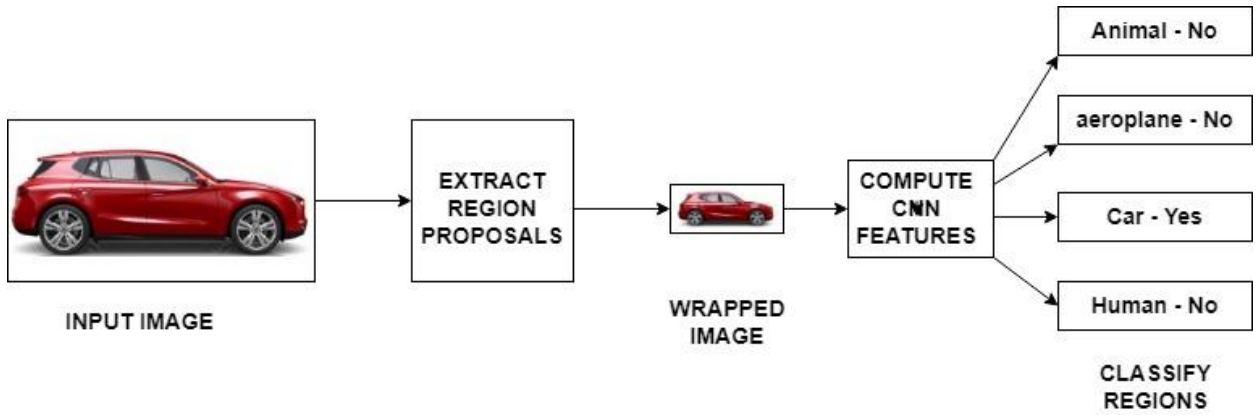

Fig. 1 Working of R-CNN algorithm

B. Fast R-CNN

The approach is similar to the R-CNN algorithm. But, instead of feeding the region proposals to the CNN, we feed the input image to the CNN to generate a convolutional feature map. From the convolutional feature map, we identify the region of proposals and warp them into squares and by using a RoI pooling layer and then it is reshaped into a fixed size so that it can be fed into a fully connected layer. From the RoI feature vector, a softmax layer is used to predict the class of the proposed region and also the offset values for the bounding box.

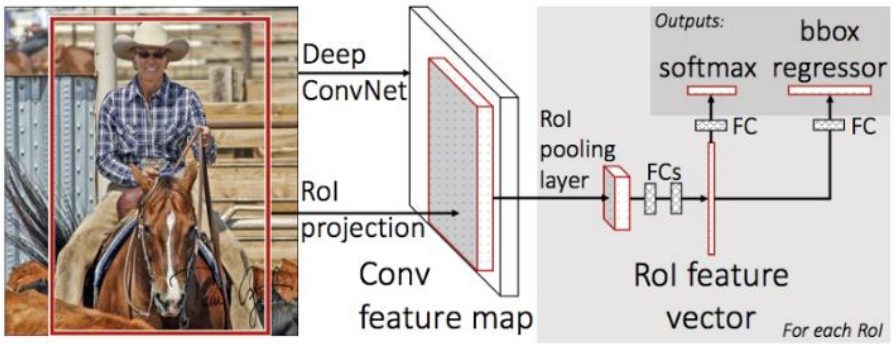

Fig. 2 Working of Fast R-CNN algorithm

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

The reason "Fast R-CNN" is faster than R-CNN is because you don't have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation is done only once per image and a feature map is generated from it.

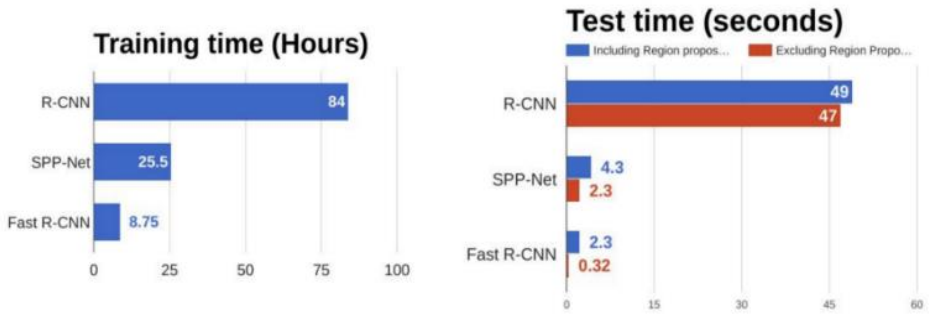

Fig. 3 Comparison between Training and Test time

From the above graphs, you can infer that Fast R-CNN is significantly faster in training and testing sessions over R-CNN. When you look at the performance of Fast R-CNN during testing time, including region proposals slows down the algorithm significantly when compared to not using region proposals. Therefore, region proposals become bottlenecks in Fast R-CNN algorithm affecting its performance.

C. Faster R-CNN

In R-CNN and Fast R-CNN, they both uses selective search to find out the region proposals. Selective search is a slow and time-consuming process affecting the performance of the network.

Similar to Fast R-CNN, the image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals. The predicted region proposals are then reshaped using a RoI pooling layer which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes.

III. YOLO v5 ALGORITHM

First, an image is taken and YOLO algorithm is applied. In our example, the image is divided as grids of 3x3matrixes.We can divide the image into any number grids, depending on the complexity of the image. Once the image is divided, each grid undergoesclassification and localization of the object.

The object ness or the confidence score of eachgrid is found. If there is no proper object found in the grid, then the object ness and bounding box value of the grid will be zero or if there found an object in the grid then the object ness will be 1 and the bounding box value will be its corresponding bounding values of the foundobject. The bounding box prediction is explained as follows. Also, Anchor boxes are used to increase the accuracy of object detection which also explained below in detail.

A. Bounding box predictions

YOLO algorithm is used for predictingthe accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and classprobabilities. Both image classification and object localization techniques areapplied foreachgrid of the image and each grid is assigned with a label. Then the algorithm checks each grid separately and marks the label which has an object in it and also marks its bounding boxes. The label soft he girds without object are marked as zero.

Consider the above example, an image is takenand it is divided in the form of 3 x 3 matrixes. Each grid is label led and each grid undergoes both image classification and objects localization techniques. The label is considered as Y.Y consists of 8 values.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

Y=	pc
	bx
	by
	bh
	bw
	C1
	C2
	C3

Pc – Represents whether an object is present in the grid or not. If present pc=1else0.

bx, **by**, **bh**, **bw** – are the bounding boxes of the objects (if present).

c1, **c2, c3**–aretheclasses.If the object is a carthenc1andc3will be 0 and c2 will be1.

IV. ACCURACY IMPROVEMENT

A. Anchor box

Fig. 4 Anchor box

Consider the above picture, in that both the human and the car's midpoint come under the same grid cell. ,In this case, we use the anchor box method. The red colour grid cells are the two anchor boxes for those objects. Any number of anchor boxes can be used for a single image to detect multiple objects. In our case, we have taken two anchor boxes.

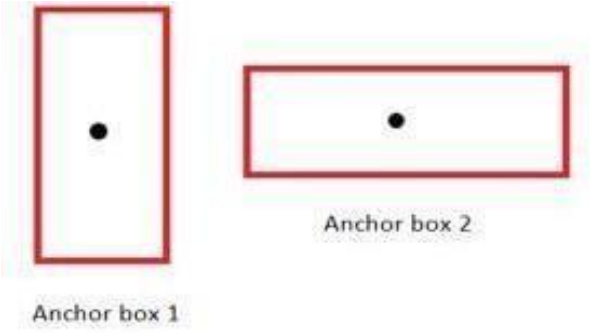

Fig. 5 Anchor box image retrieval

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

The above figure represents the anchor box of the image considered. The vertical anchor box is for thehuman and the horizontal one is the anchor box of the car. In this type of overlapping object detection, the label Ym contains 16 values i.e, the values of both anchor boxes.

V. RESULTS AND DISCUSSION

Fig. 6 Output screenshots

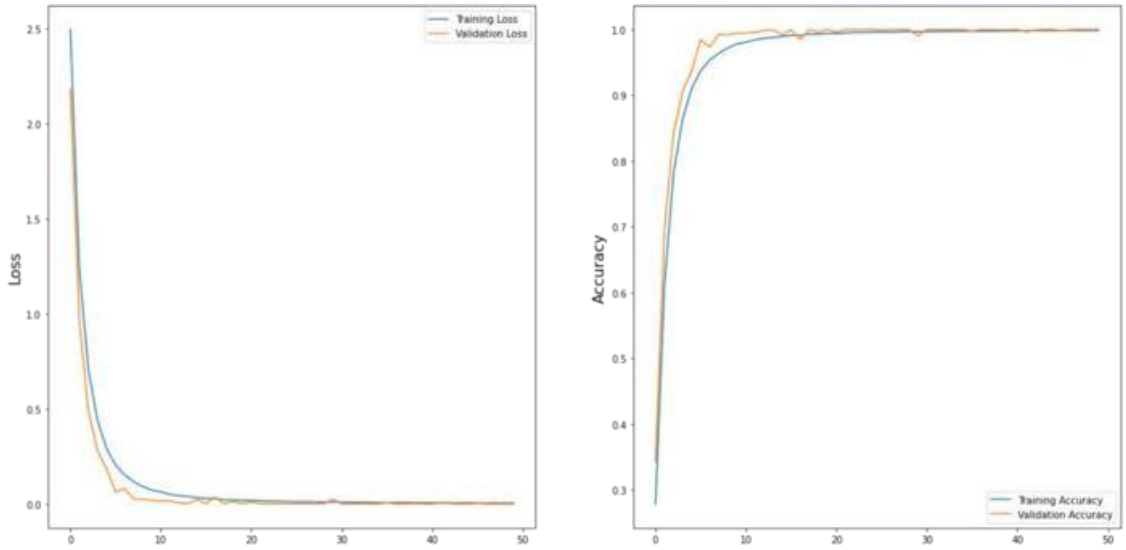

Fig. 7 Accuracy graph

VI. CONCLUSION

In this paper, YOLO algorithm is used for the purpose of detecting objects using a single neural network. This algorithm when generalized, it outperforms different strategies to differentiate from natural pictures to different domains. The algorithm is simple to build and can be trained directly on a complete image. Region proposal strategies limit the classifier to a particular region. YOLO accesses to the entire image in predicting boundaries. And also it predicts fewer false positives in background areas.Comparing to other classifier algorithms this algorithm is much more efficient and fastest algorithm to use in real time.

In this project, we concluded that the object detection that is used for training computers and AI's in the field of computer vision can be optimized for various purposes. And it has a vast area of applications in various industrial verticals like smart city, crime suspecting, laboratories, augmented reality, virtual reality and so many more industries that are applying object detection. We just have showed how the object detection technology can be used for detecting various objects by changing the dataset of the objects. This will help to change the view of the usage of the object detection technology.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 1, January 2023

DOI: 10.17148/IJARCCE.2023.12124

REFERENCES

[1]. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755. J. Mar. Sci. Eng. 2022, 10, 377 13 of 14

[2]. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. Int. J. Comput. Vis. 2010, 88, 303–338.

[3]. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 23–28 June 2014; pp. 580–587.

[4]. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv 2015, arXiv:1506.01497.

[5]. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2019, 43, 1483–1498.

[6]. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.

[7]. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

[8]. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

[9]. Shim, S.; Cho, G.C. Lightweight semantic segmentation for road-surface damage recognition based on multiscale learning. IEEE Access 2020, 8, 102680–102690.

[10]. Yuan, Y.; Islam, M.S.; Yuan, Y.; Wang, S.; Baker, T.; Kolbe, L.M. EcRD: Edge-cloud Computing Framework for Smart Road Damage Detection and Warning. IEEE Internet Things J. 2020, 8, 12734–12747.

[11]. R Sabitha, A Aruna, S Karthik, J Shanthini (2021), Enhanced model for fake image detection (EMFID) using convolutional neural networks with histogram and wavelet based feature extractions, Pattern Recognition Letters 152, 195-201.

