A Decisive Object Detection using Deep Learning Techniques

Padma P, Srinivasan Selvaraj, Ghayathri J, Suryakumar P

Abstract ***Object detection is one of the essential features of computer vision and image processing techniques. In today's world, the computer can replicate or outperform the operation that a human can do. One such thing is object detection, and In the case of it, the machines must be trained in such a way that it can recognize the object equivalent to the human does with maximum accuracy. Several object detection techniques are used to train the machine to detect the objects. Some of the most common object detection techniques are R-CNN, Fast R-CNN, Faster R-CNN) Single Shot MultiBox Detector (SSD), and You Only Look Once(YOLO),. Each of these techniques has a different way of approach and accuracy of detecting the objects in real-time. These techniques are differentiated based on their performances, i.e., speed and accuracy. Some techniques may be very accurate in detecting the objects but may lack in the time taken for detecting the objects, whereas, on the other hand, some techniques may be very fast in figuring out the objects but not with greater accuracy. We have trained an object detection model based on the YOLO technique which gave the best performance out of all other existing techniques, though the accuracy of the model is less, the speed of detection is extremely high. So based on our research we have figured out the best performance object detection techniques and also the most accurate technique. A well-trained object detection model must be very optimistic in terms of their speed and accuracy.***

Keywords : Object Detection, Deep Learning, R- Convolution Neural Network, Single Shot MultiBox Detector(SSD), You Only Look Once(YOLO).

I. INTRODUCTION

In general, the object detection is done in an image, video, or a real-time object in motion. To locate an object in the image is simple, but in case of locating an object in motion is a tedious task. It involves the FPS (Frames per second)of the video and how fast the object moves if the fps of the video is high, then the possibility and accuracy of detecting the object are less. To solve this problem, the algorithm called convolutional neural network (CNN) algorithm came into play. They are one of the best-suited algorithms so for in object detection techniques. This CNN extensively used in object detection, since it gives best results for recognizing patterns, refining edges(horizontally and vertically), colors, shapes, and textures. By running this CNN over the image, multiple objects can be detected on a single instance. The CNN consists of several layers of inputs, out of which least there are at least one hidden layer and oneoutput layer. The object detection problem can be categorized into two types. They are problems based on Classification, and Regression.

II. OBJECT DETECTION USING CLASSIFICATION TECHNIQUES

The classification technique is nothing but predicting the class of an single object in an image. This object is localized by drawing a bounding box around it and predicting the class. Consider figure(1).There is a sheep in the image, as a human, we can figure it out when we see it on first sight, but on computers perspective it too difficult for it to identify the object. This is one of the classification type problems in which the detected object is to be found as a goat or a sheep. There are several classed of objects trained using various classification algorithm — for example, a Decision tree algorithm or a support vector machine which trains the model based on several classes and predicts the output.

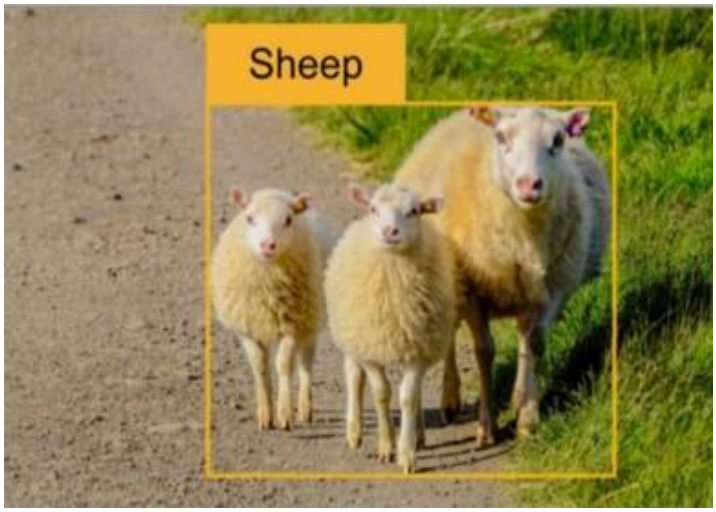


Figure 1: Object detected using classifier.

The Figure 2 shows the steps for the classification of the image. The first step is getting the digital image of acceptable size and transformed into a matrix by its pixel values. Then subsequently, the features are extracted, that is if an object to be detected is a sheep, then its features such as color class, area, perimeter, aspect ratio, shape, number of legs, and so on are extracted

Revised Manuscript Received on November 22, 2019.
* Correspondence Author
Padma P, Department of Information Technology,Sri Sai Ram Engineering College, Chennai, India Email: : padma.it@sairam.edu.in
Srinivasan Selvaraj, Department of CSE, RMD Engineering College, Chennai, INDIA. Email: ssn.cse@rmd.ac.in
Ghayathri J, Department of Information Technology,Sri Sai Ram Engineering College, Chennai, India
Suryakumar P, Department of Information Technology,Sri Sai Ram Engineering College, Chennai, India

from it. In order to train the model, a suitable algorithm is used, such as decision trees and operated in different classes of data. The prediction is made by locating the object using a bounding box around it and a label. Finally the accuracy of the algorithm is calculated.

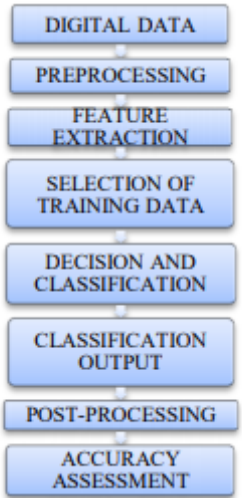


Figure2: Process involved in image classification

III. THE PROPOSED SYSTEM

The regression always returns a number instead of a class. The regression is a more advanced and effective method for object detection. The classification technique only classifies the object based on their features and class, whereas the regression model predicts the different objects in a single image or a video and returns a numerical confidence value instead of being categorical. There are several regression techniques used in object detection , some are Region-based convolutional neural network, Fast region-based convolutional network, Faster region-based convolutional network, -based fully convolutional network, You only look once, and Single-shot detector.

Model	mAP	FPS	Real Time speed
Fast YOLO	52.7%	**155**	Yes
YOLO	**63.4%**	45	Yes
YOLO VGG-16	66.4%	21	No
Fast R-CNN	70.0%	0.5	No
Faster R-CNN VGG-16	73.2%	7	No
Faster R-CNN ZF	62.1%	18	No

Figure 3: Performance of various regression techniques

As seen in the above figure, Fast YOLO is very much faster when compared to all other techniques. It could detect a real-time moving object up to 155 fps.

Though the speed is exceptional, the accuracy of detecting the object is very less. The YOLO model takes the image an input then this image is divide into SxS grids. Each grid cell consists of bounding box regression + classification heads that share the same convolutional features from the layer below. Each grid has BB bounding box regression heads that predicts a bounding box each with 5 values and confidence of an object being there. Where (x,y)(x,y) is the center of the bounding box relative to the grid cell. Each grid cell also contains a classifier which predicts CC conditional class probabilities as well. So the overall final output from a YOLO network is tensor of size:

$S \times S \times (5*B+C)$

This YOLO model consists of 24 convolutional layers, along with the2 fully connected layers. Most of these layers are pre-trained using Imagenet dataset.

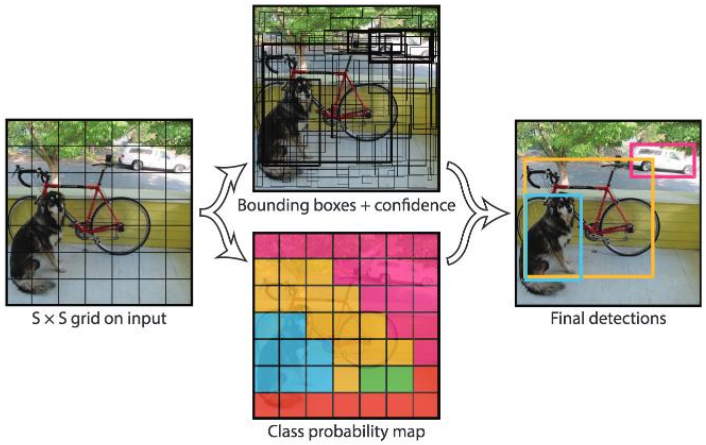
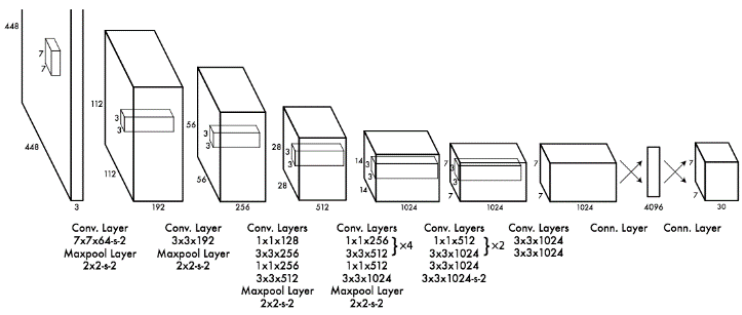


Figure 4: Detection of objects using YOLO.

The reason why the YOLO algorithm has better performance is that YOLO runs the convolutional neural network all over the image instead of a specific part of it. It makes the YOLO work faster when compared to other regression models. At the same time since it a pre-trained model, it can predict only a few objects, for example, animals, pedestrians, table, chair, and more. It looks the image only once and detects the object, so the accuracy is very less when compared to other techniques. The YOLO cannot deal with the images with less resolution, so it is one of the significant disadvantages of using it. Also, it is the fastest detection method so far which exist in the earth.



Figure(5):YOLO architecture

IV. SINGLE SHOT DETECTOR(SSD)

The single-shot detector is similar to the YOLO, and it is used to predict maximum object in a single shot by using bounding boxes. Here the model takes an image as input and passes convolutional layers of different size filters. Then the feature maps are extracted, and it is used to predict the bounding boxes. Each of the bounding boxes has four parameters that are height, width, the center coordinates.The SSD initiates with a VGG model, which is converted to into fully convolutional network. Then some extra convolutional neural layers are added, that help to handle bigger objects. The output of the VGG network is generally a 38x38 feature map. The added layers produce will produce 19x19, 10x10, 5x5, 3x3, 1x1 feature maps. These feature maps are used for predicting the bounding boxes at different scales.

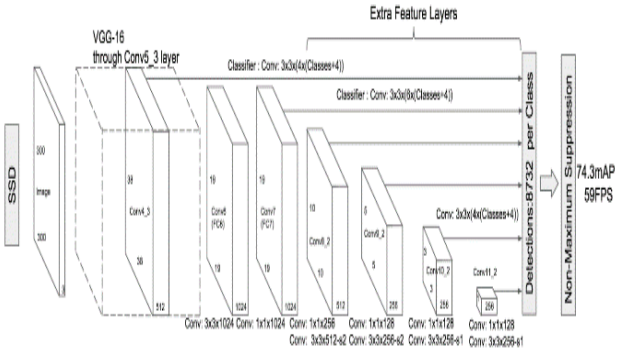


Figure 6 : SSD architecture

V. APPLLICATIONS OF OBJECT DETECTION

The object detection system is extensively used in day to day life. The facial recognition of a person while unlocking the phone is widely used in many mobile cameras. It is widely used in many surveillance systems for security purposes and also used in traffic surveillance cameras for detecting the person who is violating the rules. This technology is also used in the medical industry for identifying the cancer cells and many more.

Figure 7: Object detection in traffic camera surveillance system

VI. CHALLENGES FACED IN OBJECT DETECTION:

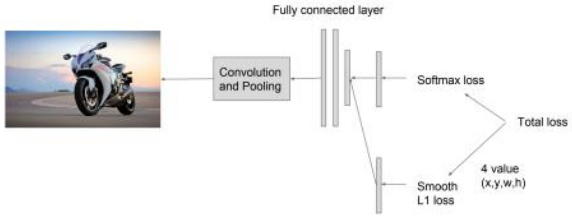
The major challenges that are faced in the object detection are that not all the objects in an image or a video can be detected; only particular objects are detected. Look at the **Figure 8.** Here the objects such as a watch, pencil, monitor are not detected.

Figure(8):Incomplete detection of objects

The other major challenge is that the problem arises in the variable dimensions of the output, which are caused by the carriable number of objects that is present in the given input image. If the model is more complex, the more time it requires for inference; and the less complex the model is, the less is the accuracy. Thus a trade-off between accuracy and the performance is chosen as per the requirements of the application. In general, any machine learning or deep learning model, it requires a fixed resolution of the input and the output to train the model in many cases. This makes the work tedious in resizing the image dimensions for proper input.Consider an example of training a machine for detecting a glass bottle on a shelf. There are so many other utensils kept on the shelf along with the glass tumbler. The searching or recognition process in such a scenario is very difficult. So far, no effective solution has been found for this problem.

VII. RELATIONSHIP BETWEEN CLASSIFICATION AND REGRESSION

Generally in any image output, the bounding boxes are predicted using the regression method and the classes that are inside each bounding boxes are identified using the classification method.The overview of the design is shown in the **Figure 9.**



Figure(9): Design overview.

VIII. LIMITATION OF OBJECT DETECTION

The deep neural network tends to outperform very well on benchmarked datasets but can fail to a critical level on images in the real world outside the datasets. Also, these deep neural networks are overly sensitive to changes in the image which would not fool a human observer as shown in **Figure 10.**

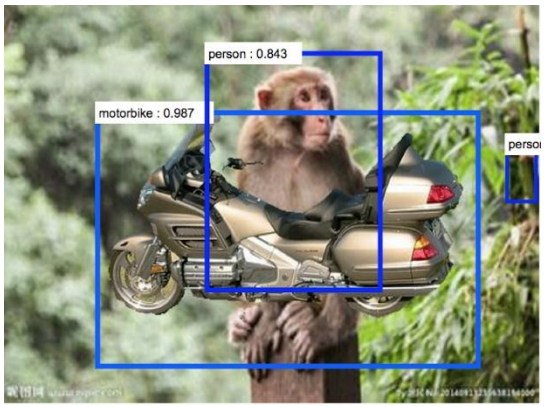


Figure 10: The occluding motorbike turns a monkey into a human

To create a better model, there must be good data sets available. However, unfortunately, to detect the real-world object, the data set needed is plenty, and it cannot be fulfilled to meet the real-world objects. So this is one of the major drawbacks of any machine learning or deep learning techniques.

IX. CONCLUSION

In this paper, we have dispensed various techniques of real-time object detection. Although there are lots of object detection techniques available, depending on the usage, specific techniques should be chosen to achieve a better-trained model. Always an excellent data set and a well-written algorithm combinedly makes a well-defined model. So as far as we arJe concerned we have fond the better results on the YOLO object detection technique in terms of better performance and Fast R-CNN in terms of their accuracy.

REFERENCES

1. Xinyi Zhou,Weigong,WenLongFu,Fengtong Du, “Application of deep learning in object detection”.published in 2017 IEEE/ACIS 16th International Confernece on computer and Information Science(ICIS).
2. Zhong-Qiu Zhao, Peng Zheng, Shou-Tao Xu,XindongWu,”Object Detection with Deep Learning:AReview”in 28 january 2019.DOI:10.119
3. Joseph Redmon, Santosh Divvala, Ross Girshich,AliFarhadi,”You Only Look Once:Unified, Real-Time Object Detection”,,IEEE Explore,12 December 2016.
4. M. B. Blaschko, C. H. Lampert, "Learning to localize objects with structured output regression", Computer Vision,EECV -2008.
5. T. Dean, M. Ruzon, M. Segal, J. Shlens, S. Vijaya-narasimhan, J. Yagnik et al., "Fast accurate detection of 100000 object classes on a single machine",IEEE conference 2013.
6. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, "Object detection with discriminatively trained part based models",IEEE transaction on pattern analysis and machine intelligence,Vol.32,published on 2010.
7. Haroon Idrees, Khurram soomro , Mubarak shah , fellow , “detecting humans in dense crowds using locally – consistent scale prior and global occlusion reasoning”, IEEE transactions on pattern analaysis and machine intelligence,published on oct 2015.
8. T. Serre, A. Oliva, T. Poggio, “A feedforward architecture accounts for rapid categorization,” Proceedings’ in 2015 2nd International Conference on Computing for Sustainable Global Development , IEEE 2015
9. F. Fleitis, H. Wang,S. Chen, “Enhancing Product Detection With Multicue Optimization for TV Shopping Applications,” IEEE Transaction on Emerging Topics in Computing, published on June 2015.
10. B. Leibe, A. Leonardis, and B. Schiele, “Robust object detection with interleaved categorization and segmentation,” Int. J. Comput. Vispublised on May 2008.
11. A. Ayvaci,S. Soatto, “Detachable object detection: Segmentation and depth ordering from short-baseline video,” IEEE Trans. Pattern Anal. Mach. Intell., published on Oct. 2012.
12. 30-43. 12. A. Torralba, K. P. Murphy, W. T. Freeman, ``Using the forest to see the trees: exploiting context for visual object detection and localization," Commun. ACM, published on Mar. 2010
13. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Computer. Society. Conf. Comput. Vis. Pattern Recognition, vol. 1. Dec. 2001.
14. A.Savich, M. Moussa, and S. Areibi, “The impact of arithmetic representation on implementing MLP-BP on FPGAs: A study,” IEEE Trans. Neural Netwpublished on Jan. 2017
15. J. Mutch and D. G. Lowe, “Multiclass object recognition with sparse,localized features,” in Proc. IEEE Conf. Computer Vision Pattern Recognit. Published on eJun. 2006.
16. T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman,,T.Poggio, “A Theory of Object Recognition: Computations andCircuits in the Feedforward Path of the Ventral Stream in PrimateVisual Cortex,” AI Memo 2005-036/CBCL Memo 259, published on 2005.
17. Bashirahamed.F.Momin,M.Mujawar,”Vehicle detection and attribute based search of vehicles in video surveliancesystem,proceedings international conference on circuit,power and computing technologies”,IEEE,published on 2015.
18. Kewen Yan ,Shaohui Huang , Yaoxian Song , Wei Liu , NengFan,”FaceRecogonition based on Convolution neural network”,IEEE xplore,11th September 2017.
19. UngseonYun , Byeungwoo Jeon, ”Implementation of real-time high speed convolution usinf DSP”,IEEE xplore,17 June 2019.
20. ZdenkaBabic, P. Mandic Danilo, "A Fast Algorithm for Linear Convolution of Discrete Time ime Signals" in TELSIK on september2001.

AUTHORS PROFILE

Mrs.P.Padma, B.E, M.E, has been working as an Assistant professor in the Department of Information Technology since 2010.She is doing her research in Cloud Security. She has participated and presented paper in various conferences in the subject of Cloud Security, Big data and Data mining. Active member of IAENG.

.**Srinivasan Selvaraj**, has been working as an Associate Professor in the Department of Computer Science and Engineering since March 2012. He obtained his B.E(CSE) from Mohammad Sathak Engineering College and M.E(CSE) from Mepco Schlenk Engineering College. He obtained his Ph.D in Information and Communication Engineering at Anna University, Chennai in 2011. He has been in the teaching profession for the past 14 years. He has been handling courses to both UG and PG students. His areas of interest include Distributed Systems, Cloud Computing, Operating Systems and Data Mining. He has published 11 papers in International Journals and 12 papers in National and International conferences.. He is a life time members of Computer Society of India.

J.Ghayathri received the Masters degree in Computer Science and Engineering from Anna University, Coimbatore. Currently, she is working as an Assistant Professor in the Department of Information Technology at Sri Sairam Institute of Technology, Chennai. She is having a teaching experience of more than 12 years. Her interests include Data Science, Cloud computing and Artificial Intelligence. She is an active member of professional bodies such as ISTE and CSI.

P.Suryakumar, pursuing B.Tech in the field of information technology in Sri Sairam Engineering college, Chennai, and currently working on several pieces of research related to Machine learning and deep learning algorithms, hoping to make some new inventions and discoveries in the field of Computer vision and Robotics. He is an active member of CSI.