IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 8, Issue.10, October 2020 E-ISSN: 2347-2693

A Comparative Study of Various Object Detection Algorithms and Performance Analysis

Anand John[1*], Divyakant Meva[2]

[1,2]Dept. of Computer Applications, Marwadi University, Rajkot, India

[*]*Corresponding Author: anand_john@yahoo.com, Tel.: +91-9428791354*

DOI: https://doi.org/10.26438/ijcse/v8i10.158163 | Available online at: www.ijcseonline.org

Received: 20/Oct/2020, Accepted: 24/Oct/2020, Published: 31/Oct/2020

***Abstract*—**Object finding is a fast-developing technique in the area of Computer Vision and Machine Learning. Computer vision is one of the principal tasks of deep learning field. Object detection is a technique that identifies the existence of object in an image or video. Object detection can be used in many areas for improving efficiency in the task. The applications for object detection are in home automation, self-driving cars, people counting, agriculture, traffic monitoring, military defence systems, sports, industrial work, robotics, aviation industry and many others. Object detection can be done through various techniques like R-CNN, Fast R-CNN, Faster R-CNN, Single Shot detector (SSD) and YOLO v3. A comparison of these algorithms is done and also their results as well as performance is analysed. The performance and exactness should be utmost important in analysing the algorithms.

***Keywords*—** Object Detection, Object Finding, R-CNN, Fast RCNN, Faster RCNN, Single Shot Detector, YOLO v3

I. INTRODUCTION

Object finding is an important part of computer vision phenomena. It is used to detect parts by differentiating the objects based on its types in pictures and videos. The applications of object detections are in face detection and recognition, home automation, self-driving cars, people counting, agriculture, traffic monitoring, military defence systems, sports, industrial work, robotics, aviation industry, medical imaging and many others. Object detection works by checking features from the test relative to the features taken from the training data. Built on the result of object detection the objects are classified and proper objects are identified and the names of the objects are displayed. Certainly, objects have to be detected even if the image is noisy and having different intensity in certain sections of the image. Appropriate identification of objects is necessary in object detection operations, so that the results are good even in brighter images. Object finding is the process of finding real-world object instances from a known class which may be cars, bikes, TV, remote, mobile, flowers, fruits and humans in images or videos. It recognizes, selects the object area, and detects multiple objects within an image which provides us with a description of what is there in the image. Even if the image contains very less numbers of object, there may be large number of possible locations of different size at which the object can be found and that need to be located anyhow. Object finding has become a very crucial research work going on nowadays as its requirements is increasing and its performance is getting better day by day. There are many methods which has been developed and used.

II. RELATED WORK

Object Detection Methods

The earlier techniques of object detection were using expert skills manually by using image operators, but by the evolution of convolutional neural networks (CNN), the task has been made much computationally better. CNN is the demonstrative model for object finding inside of an image by using deep learning [1]. The deep learning object detection techniques show better results than that of the earlier techniques. These convolutional neural networks (CNN) model is also called as VGG16. The layers of CNN are called feature maps. This feature map is a convinced mixed image and it has definite pixel value. The transformation of layer is directed by applying filtering and pooling. Filtering method transform the calculated weights with the values of an approachable ground of blocks and provides the values to a nonlinear function.

The object finding technique becomes more enhanced with the addition of regions with CNN characteristics (R-CNN), which actually performs in a totally different way from previous methods. The region with CNN features methods got amended drastically foremost with R-CNN, fast R-CNN and then faster R-CNN [2]. The regression classification related method includes SSD, YOLO. These methods have been designed to analyze the object detection problems and gives approximately correct interpretations.

III. METHODOLOGY

1) Region with Convolutional Neural Network (R-CNN): R-CNN was designed by Ross Girshick in 2014. Regions with convolutional neural networks is a deep learning approach, which combines box sized region proposals with convolutional neural network features. In R-CNN algorithm, it will first find regions in the image which are called region proposals that may contain an object. Then it

will calculate CNN features from the region proposals. In the last step, it will classify the objects using the extracted characteristics.

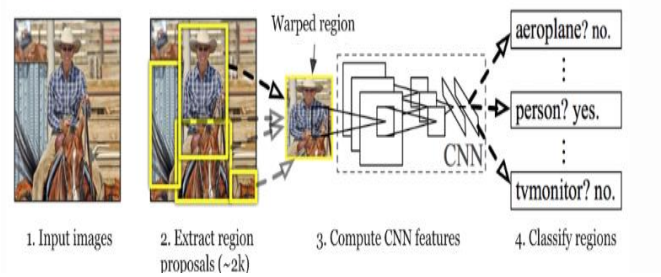

Fig. 1: The architecture of R-CNN

Regions with convolutional neural networks uses selective search for finding regions in an image and it creates 2000 region proposals for each image, i.e. we get the region of interest (RoI). During the classification of object in the image, R-CNN will only see small regions and regions having good output. The region proposals are cropped out of the image and all regions are reshaped into a fixed size. By applying bounding box regressor all regions are classified with special class specific linear support vector machines (SVMs). The region proposals are determined by checking total of positive and negative values. For each object area the bounding box regression is applied to the counted regions and then it is filtered with a non-maximum suppression (NMS) to create the bounding boxes.

Using R-CNN for object detection, there were major developments and significance over manual mathematical methods, but still there are limitations.

(1) It has to extract 2k regions for each image using selective search algorithm which is very time consuming.
(2) Using an R-CNN model is costly and slow. It takes much larger time to execute even a small work set for testing and it requires large storage memory requirement by its characteristics.
(3) Using an R-CNN model is a multi-phase procedure. In the beginning, a convolutional network is passed on region proposals. Then the classifier is replaced by SVMs to adjust with the attributes of Convolutional network. Lastly, bounding-box regressors are applied to classify the object. So, it becomes a long procedure.
(4) Moreover, additional boxes are required to attain the results in this model.

2) Fast R-CNN:

Girshick removed the limitations of R-CNN by introducing classification as well as bounding box regression and designed a unique CNN structure called fast R-CNN [3]. The fast RCNN method has the following gain:

(1) It can find object better precisely than R-CNN.
(2) This method is one-step process with little loss of different tasks.
(3) This method changes the whole network altogether.
(4) Extra memory is not required for storing the calculation.

In the model fast R-CNN, the image is processed with deep convolutional network and max pooling layers to produce a convolutional image layer with different region proposal. Then, for each region proposal the pooling layer extracts a fixed-length object's characteristics from the convolutional layer. Each object's characteristics is placed into an order of fully connected layers that gives two common output layers. One output layer that gives softmax probability between the object and background values and second output layer produces four selection box positions values for each of the object.

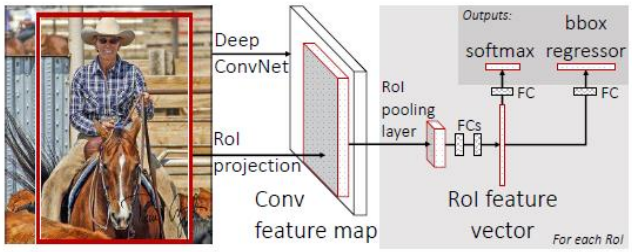

Fig. 2: The architecture of Fast R-CNN

Fig. 2 demonstrates the fast R-CNN structure which combines different parts such as convolutional network, region of interest pooling, and classification layer in a single structure. The Region of interest uses max pooling to transform the object's characteristics into a smaller attribute values which makes the calculation faster.

To make Fast R-CNN quick, two insight is needed. If region of interest is created from variant images, back-propagation using the spatial pyramid pooling (SPP) layer is not effective [4]. Fast R-CNN algorithm arranges hierarchically all the region of interests tested in each image. Apparently, calculation is shared by region of interests created from similar image in both the propagation. Further, in forward pass the time for calculating the fc layers are very high. To speed up the detection process, shorten singular value decomposition can be applied to change bigger layers into smaller layers. Although, fast R-CNN has region proposal, all its layers are trained with a multi-job loss in a one-step process. It improves the accuracy and testing time with the saving of extra expense in the storage space. However, the progress is not effective as the region proposals are created separately by additional process, which is expensive.

3) Faster R-CNN:

In both the method, R-CNN and fast R-CNN the object is detected by performing selective thorough search. This thorough search is a long process and takes larger time affecting the performance of the object detection method. This selective thorough search technique was a bottleneck in the efficiency of the object finding algorithm. Hence, from Microsoft a research team consisting of Shaoqing Ren, Kaiming He, Ross Girshick and Jian San

came up with an object detection algorithm called faster R-CNN that removes the selective search algorithm in 2015. In faster R-CNN, Shaoqing Ren and his team introduced an additional region proposal network which takes convolutional characteristics from the convolutional network and computes it to find the object, instead searching again for the object in the image. Faster R-CNN consist of two sections. The first section is having a fully convolutional neural network that creates region proposal networks, and the second section is having the fast R-CNN detector that calculates the proposed regions for object classification. This whole algorithm is a combined process for finding the object. The structure of faster R-CNN is shown in Fig. 3 [5].

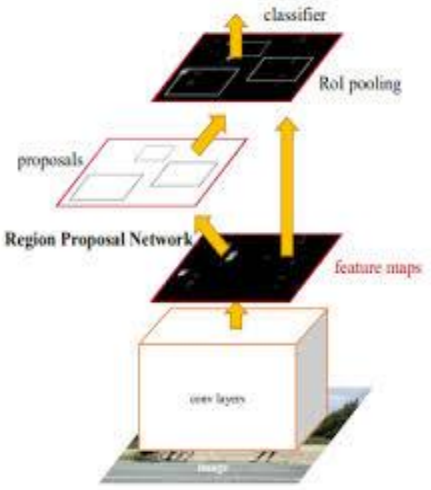

Fig. 3: Faster R-CNN design having combined network

1) Region Proposal Networks

Region Proposal Network (RPN) passes the image into the fully convolutional network to form a number of region proposals providing each an effective value as identical to [6]. These object proposals each having an effective value is passed to the fast R-CNN for further calculation. To generate these region proposals, a slight network is placed over the convolutional characteristics via the previous combined layer. The network creates a window from the convolutional characteristics. For each sliding window a minor dimensional characteristic is achieved. This characteristic is placed into two connected layers, a bounding box regression layer and a classification layer. Rectified linear units (ReLU) is used to decrease variations in the output.

For different window, it predicts any n number of regions. So, the regression layer has 4n outputs providing the pixel values of n boxes, and the classification layer produces 2n values which evaluates likelihood of object in each region. The n regions each having a box are given values which are called as anchors. An anchor is considered to be in the center of the sliding window having its specific values (Fig. 3). The graph shown in Fig. 4 shows 9 anchors with position (320, 320) for the image with pixel size (600, 800).

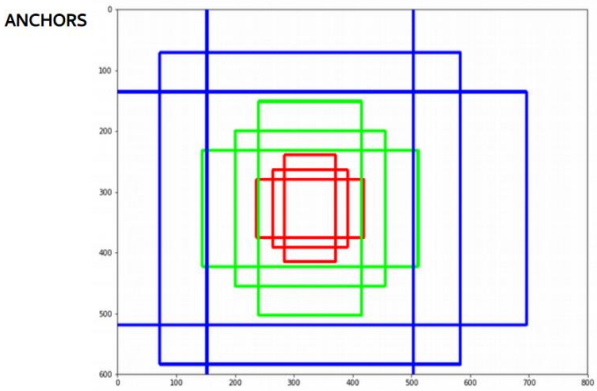

Fig. 4: Anchors with different colours showing different measurements

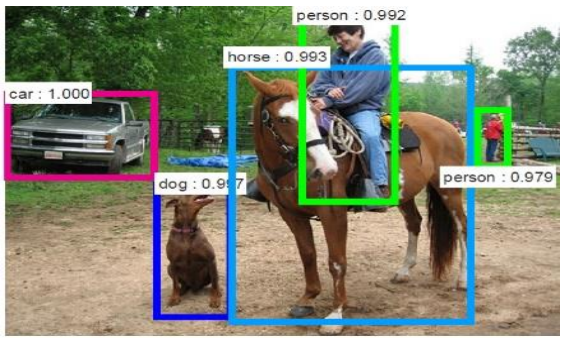

Fig. 5: Object detection using RPN

Fig. 5 shows how object detection is done using RPN by drawing the bounding box on each object. Thus, faster RCNN has region proposal with CNN features which can be used for object detection by back-propagation method.

4) SSD: Single Shot MultiBox Detector:

Single Shot MultiBox Detector (SSD) is one among the leading object finding procedures proposed by Christian Szegedy, which is presented in a distinct way in the SSD paper [7]. The swiftness and accuracy of SSD for object finding method, came about 74.3% mean average accuracy at 59 frames per second making execution time 3 times faster and more accurate than faster R-CNN which is tested on typical datasets like Pascal VOC and common object in context (COCO). SSD also improved the insufficient detection method of YOLO series in object finding and made it more accurate and efficient.

Table 1. PASCAL VOC2007 test detection results

System	data	car	cat
SSD (299*299) [7]	07+12	65.7	75.2
SSD (443*443) [7]	07+12	76.4	78.6

SSD's structure as shown in Fig. 6 is built on the VGG-16 structure, but it eradicates the fully connected layers. SSD takes on the idea of anchor in faster R-CNN. The bounding box is determined by the earlier boxes with varying values and measurement, then the characteristics is extracted by CNN and then classification and regression are done correctly. The whole method is executed only in a single

process. SSD network selects the priors which contains the object of interest and find the coordinates that match exactly the object's shape. These fully convolutional bounding box are same as anchor boxes used in Region Proposal Networks. The method is simplified by changing the 3*3 convolution layers to 1*1 layers to predict the values of the particular object and so an intermediate layer is not required.

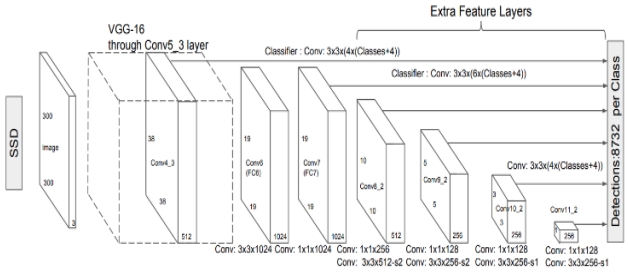

Fig. 6: The architecture of SSD

Significantly, SSD design is the first work which joins different values from multiple features and with different resolutions classifies the object and improves quality of object detection. SSD design is the first technique to use bounding box values with characteristics of differing resolution in the network. Also, SSD detects multiple objects without sharing convolutional layers. SSD shares a same objective as MultiBox, yet, it can detect multiple categories in a single shot calculation unlike the two-stage method. After that, hard negative mining is performed where the highest prior values are taken which leads to faster performance of the network. Ending results are got by performing Non-Maxima Suppression (NMS) on multi-scale bounding boxes.

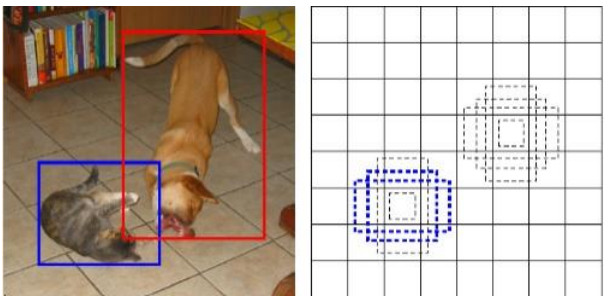

Fig. 7: Object detection using SSD

Yet, SSD is not efficient in finding some little objects correctly, but it can be developed by using better feature extractor model like ResNet101 and using better network like Stem Block and Dense Block [8]. Deeply Supervised Object Detector (DSOD) give better result than SSD on all levels with increased performance, which requires only half constraints compared to SSD and 1/10 constraints compared to Faster R-CNN.

Moreover, a better SSD object detection algorithm is given by Sheping Zhai, Dingrong Shang, Shuhuan Wang, and Susu Dong in their paper which is specified in [9]. By keeping SSD as main algorithm, the specific extraction network DenseNet-S-32-1 is given which replaces the VGG-16 network. To detect multiple object a different mechanism called fusion is presented to combine low and high-level characteristics. At the end, the location is recognized before the object calculation to increase the algorithm speed. The DF-SSD algorithm is tested from the beginning. The consequences show that the algorithm DF-SSD with a particular input image achieves 81.4% mean access precision on PASCAL VOC 2007 datasets [9]. The detection accuracy of DF-SSD got better by 3.1% mean access precision. DF-SSD requires only 0.50 constraints to SSD and 0.11 constraints as compared to faster R-CNN.

5) YOLO v3 (You Only Look Once)

YOLO developed by Joseph Redmon, is a combined architecture model and is extremely fast. YOLO algorithm detects objects in an image at 45 frames per second. It is better than other detection methods, including DPM and R-CNN. But YOLO v3 is faster than previous YOLO. YOLO v3 runs in 22 ms at 28.2 mean access precision as accurate as SSD but three times faster than SSD. It has similar performance as DSSD. YOLO v3 is also good at detecting small objects in an image.

Different thoughts can be obtained from different paper, especially from many different research persons. In YOLO 9000, the model finds four constraints for each detection box and it can detect over 9000 different categories and has 19.7% mean access precision [10].

YOLO has YOLO-based convolutional neural network group of algorithms to find the object. It's most recent version is YOLOv3 [11]. YOLO is considered as fully convolutional network because it makes use of only convolutional layers. The YOLOv3 technique proposes finding the object a regression problem. The method is designed in DarkNet based on VGG which was previously developed in GoogleNet. The softmax loss in YOLO v2 is replaced by a logistic loss, which got better in finding small object [12].

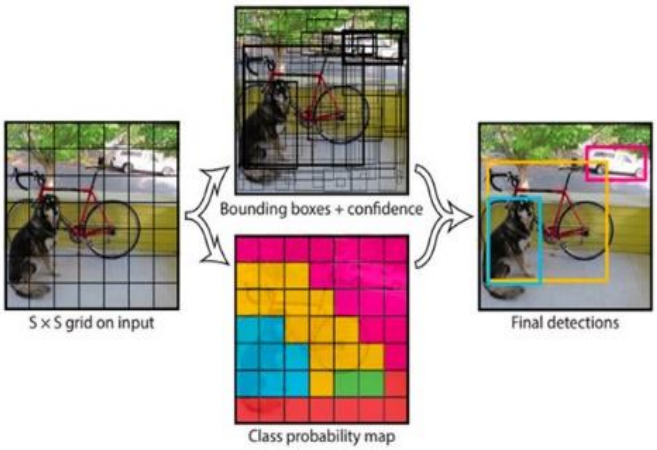

Fig. 8: Object detection using YOLO

Darknet-53 has 53 layers and ReLu activation making a 106 total convolutional layer. The Keras library is to be

added in the tensorflow for the implementation of the YOLOv3 algorithm which also has many open source libraries [13]. The learned weights are used in YOLOv3 to find the object in the images given.

YOLO v3 removes region proposal method and joins every process in one network so as to form an accurate object detection algorithm. It finds the exact object from the different types of object. YOLO v3 method uses three anchors and so three bounding boxes for each cell is generated. The YOLO v3 method divides the image into a small net of cells and so each cell will give a selection box offsets and classify the objects through forward convolution [14]. The bounding boxes are combined to detect the object after a post-processing process by the algorithm. If a grid cell is center of an object, then that grid cell is considered in finding the object. Each grid cell is used to find the exact area of the selection boxes.

YOLO v3 is similar as compared to YOLOv2. The YOLO v3 network predicts four coordinates values for each bounding box [15]. Only one bounding box is selected for each of the object. Also, for predicting three different scales i.e. three different anchors are used. More number of images makes the technique more precise, but it makes the technique slower to start division of different objects [16]. In YOLO v3 technique, k-means clustering is used for finding improved bounding box prior. The k-means clustering chooses arbitrarily as beginning cluster centers k combination of width and height values which is mandatory in order to identify the object.

IV. RESULTS AND DISCUSSION

The proposed method, yolo v3 model is implemented and, in the model, the learned weight file is loaded in the tensorflow, and then the detail debug information is displayed about what was loaded, as given in the below two lines.
loading weights of convolution #104
loading weights of convolution #105

After that the description of the objects which are detected are displayed by the model and their confidence. Using the YOLO v3 model, by testing the photograph in the tensorflow we get the results [17]. It can be seen that the model has found three zebras displaying the name and the estimation all above 90% likelihood.
[(1, 13, 13, 255), (1, 26, 26, 255), (1, 52, 52, 255)]
zebra 94.91063356399536
zebra 99.86327290534973
zebra 96.87087535858154

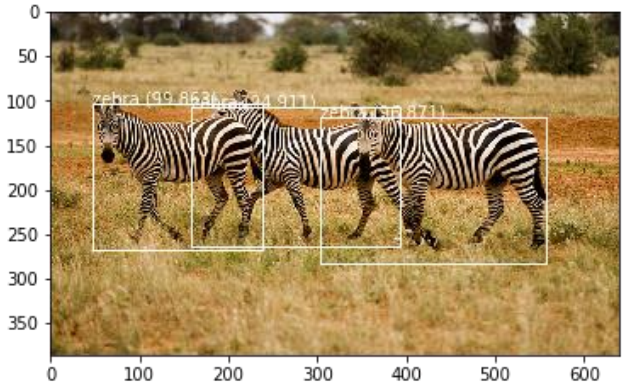

Fig. 9: Output of yolo v3 model detecting three zebras

In YOLO v3 model, the proposed method detects objects with correct output and approximately in the correct position showing the bounding box all around the objects.

V. CONCLUSION

Several object detection techniques like R-CNN, fast R-CNN, faster R-CNN, single shot detector (SSD), YOLO v3 etc. are being discussed and compared. From the discussions, it is found that as the model was developed the speed and accuracy has been improved and increased. Fast R-CNN is improved than RCNN but Faster R-CNN is much improved than fast R-CNN. Also, single shot detector is better than faster R-CNN, while YOLO v3 is better than single shot detector. Earlier, till YOLO v3 was not developed SSD was the best. But now, the best technique found latest is YOLO v3 which is much better than SSD also and much faster than SSD. YOLOv3 is extremely fast and accurate. Hence, using YOLO v3 model, we can detect multiple objects faster using tensorflow and add our own images and labels in the datasets. This YOLO v3 model is beneficial as it can detect object directly and all objects are detected single time only in this model.

ACKNOWLEDGMENT

I thank the Almighty God for his grace and mercy in writing this paper. I also thank my guide for helping me and guiding me in every area. I also thank all the professors of Faculty of computer Applications, Marwadi University for guiding me in my paper.

REFERENCES

[1] Y. LeCun, Y. Bengio, G. Hinton, "*Deep Learning*", Nature, Vol.**521,** pp.**436-444**, **2015**.
[2] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich Feature *Hierarchies for Accurate Object Detection and Semantic Segmentation*", IEEE Conference on Computer Vision and Pattern Recognition, Columbus, **USA**, pp. **580-587, 2014.**
[3] R. Girshick, "*Fast R-CNN*", *IEEE* International Conference on Computer Vision *(ICCV)*, Santiago, **USA** pp.**1440-1448, 2015.**
[4] Z. Zhao, X. Wu, S. Xu, P. Zheng, "*Object Detection with Deep Learning: A Review*", IEEE Transactions on Neural Networks and Learning Systems, Vol.**30,** Issue.**11**, pp.***3212-3232*, 2019.**

[5] S. Ren, K. He, R. Girshick, J. Sun, "*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.**39**, Issue.**6**, pp.**1137-1149, 2017.**

[6] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers and A.W.M. Smeulders, "*Selective Search for Object Recognition*", International Journal of Computer Vision, Vol.**104**, pp.**154–171, 2013.**

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A.C. Berg, "*SSD: Single Shot Multibox Detector*", Springer, Vol.**9905**, pp.**21-37, 2016.**

[8] Z. Shen, Z. Liu, J. Li, Y. Jiang, Y. Chen, X. Xue, "*DSOD: Learning Deeply Supervised Object Detectors from Scratch*", IEEE International Conference on Computer Vision (ICCV), Venice, **Italy**, pp.**1937-1945, 2017.**

[9] S. Zhai, D. Shang, S. Wang, S. Dong, *"DF-SSD: An Improved SSD Object Detection Algorithm Based on DenseNet and Feature Fusion*", IEEE Access, Vol.**8**, pp.**24344-24357, 2020.**

[10] J. Redmon, A. Farhadi, "*YOLO9000: Better, Faster, Stronger*", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, **HI**, pp.**6517-6525, 2017.**

[11] J. Redmon, A. Farhadi, "*YOLOv3: An Incremental Improvement*", ResearchGate, **2018**.

[12] J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, B. Cai, "*An Improved Yolov2 for Vehicle Detection*", Sensors, Vol.**18**, Issue.**12**, pp.**427**2, **2018**.

[13] A. Agrawal, A.N. Modi, A. Passos, A. Lavoie, A. Agarwal, A. Shankar, I. Ganichev, J. Levenberg, M. Hong, R. Monga, S. Cai, "*Tensorflow Eager: A Multistage, Python-Embedded DSL for Machine Learning*", Proceedings of Machine Learning and Systems 1 (MLSys 2019), Stanford, **California**, pp. **178-189, 2019**.

[14] J. Y. Lu, C. Ma, L. Li, X.Y. Xing, Y. Zhang, Z.G. Wang. J.W. Xu, "*A Vehicle Detection Method for Aerial Image Based on YOLO*", Journal of Computer and Communications, Vol.**6**, Issue.**11**, pp.**98-107, 2018.**

[15] L. Zhao, S. Li, "*Object Detection Algorithm Based on Improved YOLOv3*", Electronics, Vol.**9,** Issue.**3,** pp.**537**, **2020**.

[16] N. Raviteja, M. Lavanya, S. Sangeetha, "*An Overview on Object Detection and Recognition*", International Journal of Computer Sciences and Engineering, Vol.**8**, Issue.**2**, pp.**42-45, 2020.**

[17] A. Kaur, D. Kaur, "*Yolo Deep Learning Model Based Algorithm for Object Detection*", International Journal of Computer Sciences and Engineering, Vol.**8**, Issue.**1**, pp.**174-178, 2020.**

AUTHORS PROFILE

Mr. Anand John pursued DOEACC B-LEVEL, DOEACC Soceity, India in 2004 and Master of Computer Science from Punjab Technical University in year 2009. He is currently pursuing Ph.D. in Faculty of Computer Applications, Marwadi University, India and currently working as Programmer in Department of Computer Applications, Christ College, Saurashtra University, India since 2007. He is also available online. His main research work focuses on Image Recognition and Computational Intelligence based education. He has 13 years of teaching experience and 2 years of Research Experience.

Dr. Divyakant Meva pursued Master of Computer Applications from Saurashtra University, India in 2002 and Ph.D. from Saurashtra University, India in year 2015. He is currently working as Associate Professor in Faculty of Computer Applications, Marwadi University, India since 2010. He is a member of CSI since 2013. He has published more than 21 research papers in reputed national and international journals. His main research work focuses on Biometrics, Blockchain, Cyber Security and Artificial Intelligence. He has 18 years of teaching experience and 5 years of Research Experience.

