iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42968

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Image/Object Detection Using Shot Multi-Box Detector (SSMBD) Algorithm

Vikash Singh[1], Shalvi Sainath[2], Tarun Choudhary[3], Rishabh Jain[4], Rochak Bajpai[5]
[1, 2, 3, 4]*B.Tech Students, Department of Electronic and Communication Engineering, KIET Group of Institutions, Ghaziabad, UP*
[5]*Guide*

Abstract: ***Object identification, is one of most important roles in computer vision, has been a hotspot for research and application over the past 20 years. Its purpose is to recognise and locate a large number of items in a given environment that fall into specific categories rapidly and consistently. It has gained a lot of study attention because of its tight association with image and video analysis.***
Its purpose is to discover and locate a large number of things in a given image that belong to specified categories rapidly and consistently. More sophisticated tools that can learn semantic, high-level, richer aspects are being offered to address the existing issues as deep learning advances.
There are several types of algorithms. Based on the model training approach, the algorithms can be divided into two categories: single-stage detection algorithms and two-stage detection algorithms. Our investigation begins with an overview of deep learning and its most prominent tool, the Convolutional Neural Network (CNN). We'll also look at a common traditional object detection framework, along with some variations.
Other tasks such as face detection and object tracking, as well as other important characteristics, would be implemented. Thus, future work in both object detection and relevant neural network-based learning systems should adhere to these guidelines and be useful.
Important terms: SSD, Convolution Neural Network, YOLO

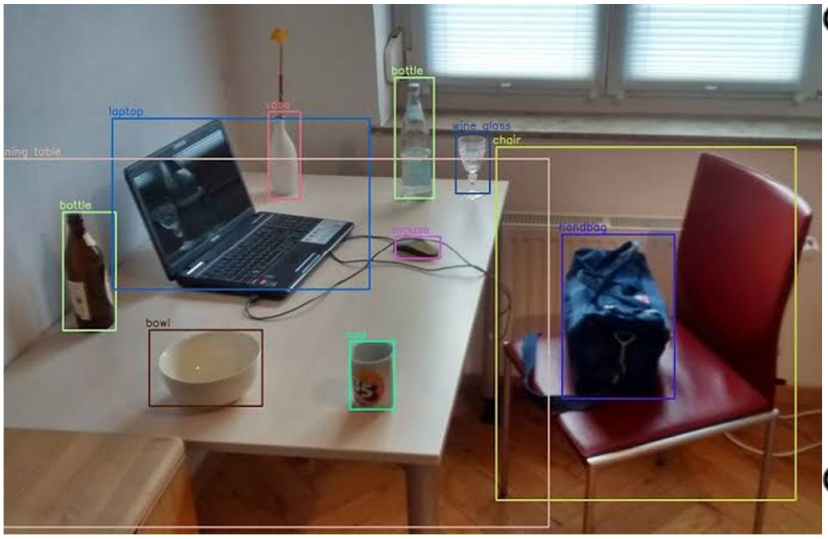
I. INTRODUCTION

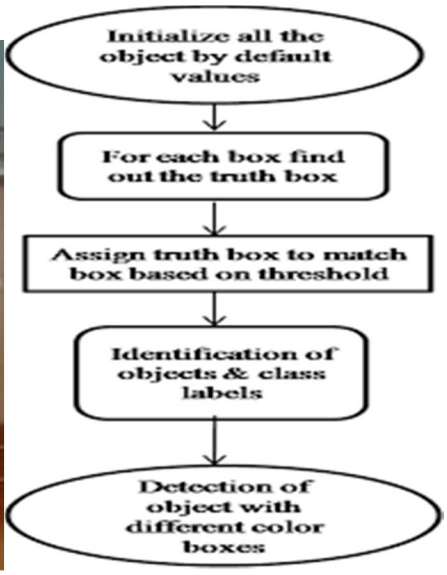
Object detection may be a basic research direction within the fields of computer vision, deep learning, artificial intelligence, etc. it's a very important prerequisite for more complex computer vision tasks, like target tracking, event detection, behaviour analysis, and scene semantic understanding.
It aims to locate the target of interest from the image, accurately determine the category and provides the bounding box of every target. It's been widely employed in vehicle automatic driving, video and image retrieval, intelligent video surveillance, medical image analysis, industrial inspection and other fields. Traditional detection algorithms on manually extracting features mainly include six steps: pre-processing, window sliding, feature extraction, feature selection, feature classification and post processing and usually for specific recognition tasks.
Its disadvantages mainly include small data size, poor portability, no pertinence, time complexity, window redundancy, no robustness for diversity changes, and good performance only in specific simple environments. Deep learning has become the most talked-about technology owing to its results which are mainly acquired in applications involving language processing, object detection and image classification.
The market forecast predicts outstanding growth around the coming years. The main reasons cited for this are primarily the accessibility of both strong Graphics Processing Units (GPUs) and many datasets. In recent times, both these requirements are easily available. This review article aims to make a comparative analysis of SSD, Faster-RCNN, and YOLO. The first algorithm for the comparison in the current work is SSD which adds layers of several features to the end network and facilitates ease of detection. The Faster R-CNN is a unified, faster, and accurate method of object detection that uses a convolutional neural network. While YOLO was developed by Joseph Redmon that offers end-to-end network.

A. SSD (Single Shot Detection)

The SSD architecture uses an algorithm to recognise numerous item classes in a photograph by assigning confidence scores to the existence of each class of items.

©IJRASET: All Rights are Reserved





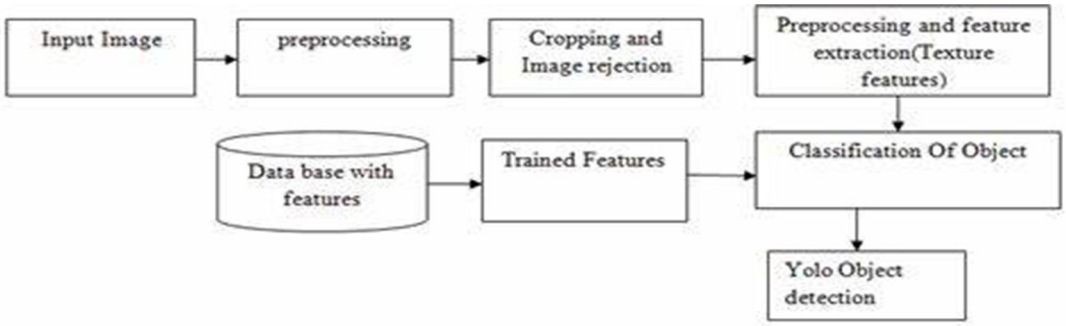
It also causes the shape of the things in the boxes to shift. Because it does not re-evaluate bounding box assumptions (as Faster RCNN does), this is ideal for real-time applications. The SSD architecture is CNN-based, and it uses two stages to recognise the target classes of objects: extracting feature maps and using convolutional filters to detect the objects. Object detection and pattern recognition are still problems in computer vision. The main picture classification issues, such as noise resilience, transformations, and barriers, have been carried over, as well as new ones, such as identification of moving objects.

The estimation of traffic density is a critical component of an automated traffic monitoring system and can also be done by SSD. Traffic density estimation can be utilised in a variety of traffic applications, ranging from congestion detection to macrotraffic control in urban areas. We used Single Shot Detection (SSD) and MobileNet-SSD to estimate traffic density in this paper. SSD can handle objects of various shapes, sizes, and viewing angles.

II. OTHER ALGORITHM FOR OBJECT DETECTION

A. YOLO

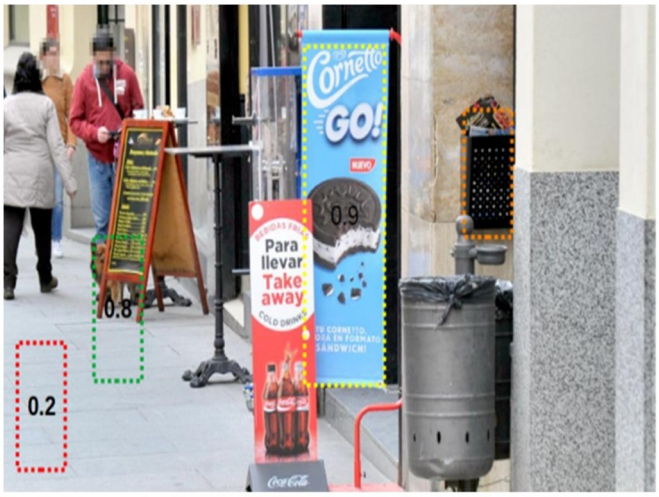
It employs a very unexpected strategy for real-time object detection: it is a CNN. The procedure uses a single neural network to process the entire image then divides the image into regions and predicts boundary boxes. For each region, there are boxes with probability. These encircling boxes would use high probability to evaluate the intended probabilities YOLO is an acronym that stands for "you only live once." It is well-known for requiring only one advance propagation across the internet. Neural network After non-max suppression (which takes a long time), ensuring that each object is recognised exactly once) it returns boundary boxes for recognised objects YOLO functions by recognising an image as data and dividing it into S X Inside each grid, create a S grid with m boundary boxes. Other object detection techniques are slower than YOLO because of the YOLO algorithm's spatial restrictions, it struggles to discern a tiny item. Unlike sliding window and area proposal-based systems, YOLO recognises artefacts in images very effectively since it is trained to observe the full image during training and testing time to obtain every insight into the entire image and object and its appearance. The method divides the image into grids and performs the image classification and limitation algorithm on each grid cell. It scores all grids and forecasts N bounding boxes.



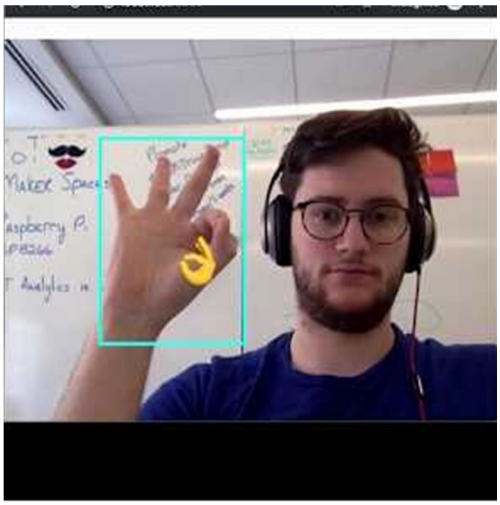
III. APPLICATIONS

A. *Advertising Detection*

In both the real and virtual worlds, detecting picture advertisement boards has critical uses. Google Street View, for example, may use it to update or personalise the advertising that shows on street photos as in (i)



(i)



(ii)

B. *Activity Recognition*

The goal of activity recognition is to identify one or more operators' behaviors and intentions from a collection of observations of specialized activity and ambient circumstances. Many informatics networks have been considered in this area of research because it provides personalised support for many applications and its relationship to a variety of domains, such as the interaction between the human machine and the humanistic method. We can see in image (ii).

IV. CONCLUSION

In this paper, we propose a novel single shot object detector named Detection with Enriched Semantics (DES). To address the problem that low level detection feature map does not have high level semantic information, we introduce a segmentation branch, which utilize the idea of weakly supervised semantic segmentation, to provide high semantic meaningful and class-aware features to activate and calibrate feature map used in the object detection. We also utilize a global activation module to provide global context information and pure channel-wise learning. Our method is flexible and simple, and does not require too much modifications to the original detection framework SSD. Quantitative evaluation shows our method excels in both accuracy and speed. Our method can also be applied to other two stage or single shot object detectors, with stronger backbone, and we remain this as future work.

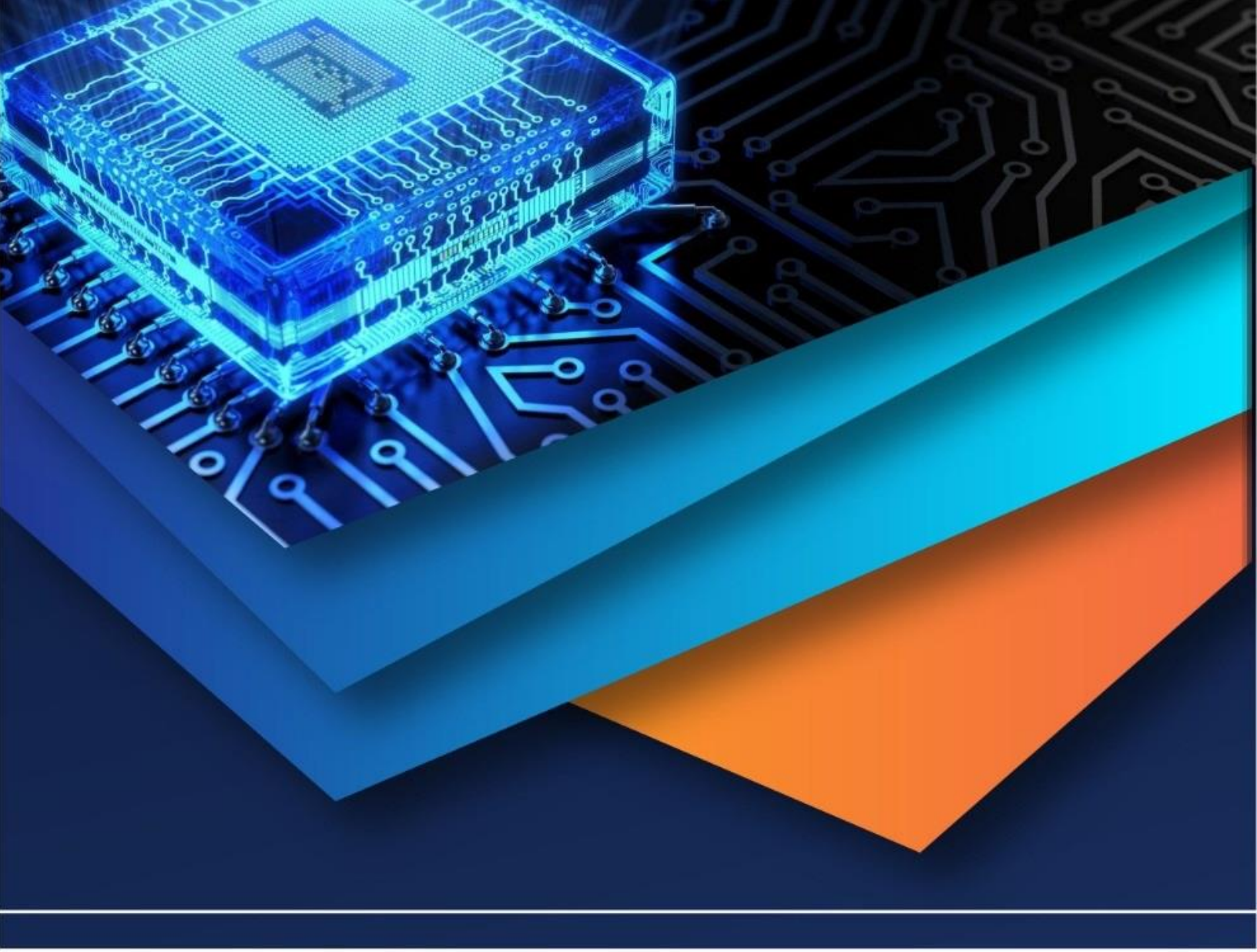
We have developed an object detector algorithm using deep learning neural networks for detecting the objects from the images. The research work uses a single shot multi-box detector (SSMBD) algorithm along with Faster CNN to achieve high accuracy in real time for detection of the objects. The performance of our algorithm is good in still images and videos. The accuracy of the proposed model is more than 75%. The training time for this model is about 5-6 hours. This model uses convolutional neural networks to extract feature information from the image and then perform feature mapping to classify the class label. Our scheme uses separate filters with different default boxes to tackle the difference in aspect ratio and also used multi-scale feature maps for object detection. The prime objective of our algorithm is to use truth box to extract feature maps. For checking the effectiveness of the scheme, we have use Pascal VOC and COCO dataset. We have compared the values of different metrics such as mAP and FPS with other previous model, which indicates that the algorithm achieves a higher mAP and uses more frames to gain good speed.

REFERENCES

[1] Guei-Sian Peng “Performance and Accuracy Analysis in Object Detection” CALIFORNIA STATE UNIVERSITY SAN MARCOS.
[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, “You Only Look Once: Unified, Real-Time Object Detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.9
[3] Sciencedirect.com/science/article/pii/S1877050918308767
[4] Journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00434-w#Sec31
[5] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition Computer Vision and Pattern Recognition”, 2014.

[6] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, "Object Detection with Deep Learning: A Review"
[7] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd Int. Conf. Big Data Anal. ICBDA 2017, pp. 721–724, 2017, doi: 10.1109/ICBDA.2017.8078730.
[8] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," J. Phys. Conf. Ser., vol. 1004, no. 1, 2018, doi: 10.1088/1742-6596/1004/1/012029.
[9] Mauricio Menegaz, "Understanding YOLO – Hacker Noon," Hackernoon. 2018. [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016, doi: 10.1109/CVPR.2016.91.
[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Lerning Applied to Document Recognition," proc. IEEE, 1998, [Online]. Available: http://ieeexplore.ieee.org/document/726791/#fulltext-section.
[11] T. F. Gonzalez, "Handbook of approximation algorithms and metaheuristics," Handb. Approx. Algorithms Metaheuristics, pp. 1–1432, 2007, doi: 10.1201/9781420010749.
[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)