International Journal of Innovative Research in Engineering
Volume 3, Issue 3 (May-June 2022), PP: 550-553.
www.theijire.com
ISSN No: 2582-8746

Analysis of Object Detection Algorithms

Rishabh Tiwari[1], Prof. Vijay Mane[2], Ravindra Chaugule[3]
[1,2]Electronics Department, Vishwakarma Institute of Technology, Pune, Maharashtra, India.
[3]Principal Software Engineer, Renishaw Metrology Systems Ltd, Pune, Maharashtra, India.

How to cite this paper: Rishabh Tiwari[1], Prof. Vijay Mane[2], Ravindra Chaugule[3] ," Analysis of Object Detection Algorithms", IJIRE-V3I03 550-554.

Abstract: *Object recognition is a computer vision and imaging technology that recognizes and locates semantic artefacts of a specific class (such as people, structures, and automobiles) in automated objects and observations in computerized images and recordings. Continuous object tracking and identification is a large, active, yet unclear and challenging area of PC vision. It has become a well-known module for a variety of important applications such as video surveillance, self-driving, face recognition, and so on. The issue investigated as part of the overview includes various quality measurements, calculations, speed/size trade-offs, and preparation methods.*

Key Word: *Object Detection, YOLO, SSD, RCNN*

I. INTRODUCTION

Object detection is the recognition of an object in an image that is constrained and ordered. Item limitation refers to the process of identifying the condition of at least one object inside an image (or video) and drawing a bounding box around it. Image ordering is a method of grouping or anticipating the classification of a certain object in an image. Object detection combines these two methods, limiting and arranging at least one object in a picture. Detection is further refined, indicating what the image's "main subject" is. Object detection can discover many objects, arrange them, and locate them in the image.

Each object class has its own unique characteristics that aid in class organization - for example, all circles are circular. These exceptional features are used in object class detection. The object detection models anticipate the bounding boxes. For each object, the model predicts bounding boxes and categorization probability. When it comes to object detection, it's reasonable to expect too many bounding boxes. Each box also has a confidence score, which shows how likely the model believes the image contains an object. Finally, any boxes with a score below a certain threshold are deleted (called non-maximum suppression).

There are two types of object detectors: one-stage detectors, such as Faster R-CNN or Mask R-CNN area proposal networks, and two-stage detectors, such as Faster R-CNN or Mask R-CNN area proposal networks, which create first-stage regions of interest and submit region of proposals down the pipeline for object classification and bounding box regression. These types of models achieve the highest levels of accuracy, but are typically slower. R-CNN's difficulty is that training the network still takes a long time because it has to classify 2000 regional proposals for each image, making real-time implementation impossible. As a result, no prior knowledge is required. This could lead to the establishment of unsatisfactory regional ideas. Then there are single-stage proposals like YOLO and SSD, which treat artefact detection as a simple regression problem by taking a picture and learning probabilities of the class and bounding box co-ordinates. These models are less accurate than two-stage object detectors, but they are significantly faster.

II. COMPARISON OF SINGLE STAGE AND TWO STAGE OBJECT DETECTION

A. Two Staged Detection

The accuracy of detection is prioritized in this technique. The detection and posture assessment processes are separated in the two-stage technique. Identified items are clipped and processed by a different network for present assessment after object detection. This necessitates resampling the image at least three times: once for region suggestions, once for detection, and once for current evaluation. That works well, but it is time consuming because the detection and classification portions of the model must be run multiple times. Fast R-CNN, Faster R-CNN, and Mask R-CNN are the three basic models.

B. Single Stage Detection

The speed of inference is prioritized in this technique. The proposed method, on the other hand, does not require resampling of the image and instead uses convolutions to recognize the item and its location in a single forward projection. Because the image is not re-sampled and the detection and position assessment calculations are shared, this results in a significant speedup. That is substantially speedier and more appropriate for cell phones. YOLO, SSD, SqueezeDet, and Detect Net are some of the most well-known one-stage object detectors. The MSCOCO dataset is the most often used benchmark. Models are frequently evaluated using the Mean Average Precision metric.

III. DRAWBACKS OF TWO STAGED DETECTOR

- It allots a significant amount of time and effort to train the network required to group 2000 regional ideas per image.
- Because each test image takes roughly 47 seconds to generate, it cannot be updated in real time.
- The specific search algorithm is pre-programmed; therefore, no learning is required at that point.
- They completely lose all intrinsic information about the object's position and orientation, and the information is sent to similar neurons that are unlikely to be able to handle it.
- A CNN produces predictions by looking at an image and then validating whether or not particular portions of the image are there. If they are, the image is rearranged properly. R-CNN, Fast R-CNN, and Faster R-CNN were designed to overcome obstacles. By comparing its output with the COCO data-set, the Faster R-CNN was the best algorithm out of all the above in terms of accuracy and training time.
- Faster R-impediment CNNs were created to circumvent this restriction. YOLO (you only live once) mistake analysis.
- Faster R-impediment CNNs were created in attempt to circumvent this constraint. Using Quick R-CNN to analyses YOLO faults reveals that YOLO commits multiple locale errors. However, this reduced the SSD's accuracy when compared to the Faster R-CNN. Additionally, using darknet frames, YOLO object detection algorithms have been developed; in terms of accuracy and inference time, the current version of, for example, the V3 from YOLO has outperformed the Faster R-CNN and SSD

IV.SINGLE STAGE OBJECT DETECTION

A. YOLO

YOLO employs a highly unorthodox strategy for object detection in real time: it is a CNN. The technique uses a single neural network to process the entire image (or frame of video) and then isolates the image into areas, predicting bounding boxes and probabilities for each. YOLO is notable for requiring only one advance propagation over the neural network, and these bounding boxes would evaluate the intended probability with great precision while still being able to execute in real time. It returns accepted items with bounding boxes after non-max suppression (which assures that each object is recognized exactly once). YOLO takes an image as information and divides it into a S X S grid with m bounding boxes inside each grid.

For each bounding box generated, the network generates a yield 'a' class probability and balances esteems. The bounding box with a probability class greater than the threshold value is chosen and utilized to locate the object within the image. YOLO is faster than other object identification algorithms in terms of order of size (45 FPS). The YOLO algorithm's limiting and disadvantageous feature is that it has difficulty recognizing a smaller object, which is due to the YOLO algorithm's spatial restrictions. Unlike sliding window and area proposal-based techniques, YOLO recognizes objects of interest in images since it is used to see the full image throughout training and testing time, giving it every insight into the entire image and object and its appearance. The algorithm divides the image into grids and performs the image classification and NMS algorithm on each grid cell. It calculates scores on all grids and forecasts N bounding boxes. The certitude score represents the precision of the bounding box for that class. As a result of some of these Unwanted bounding boxes or things can be avoided by setting a threshold because boxes have low safety ratings.

B. SSD

The SSD architecture employs an algorithm for detecting various object classes in a image by assigning confidence scores to the instance of each object category. It also causes the shape of the things in the boxes to shift. Because it does not re-evaluate bounding box assumptions (like Faster R- CNN does), this is suited for real-time applications. The SSD architecture is CNN-based, and two stage detector since it uses two stages to recognize 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 image classification issues, such as noise robustness, transformations, and impediments, have been retained, but new challenges, such as detection of various artefacts, overlapping images, and determining their positions within a picture, have been added. SSD achieves a better balance between speed and accuracy. Once an image is inputted, it only runs a standard network and outputs a function diagram.

V. DIFFERENCE BETWEEN YOLOV3 AND SSD (RESULTS)

Sl.no.	Parameters	YOLO	SSD
1	Model name	You Only Look Once	Single Shot Multi-Box Detector
2	Speed	Low	High
3	Accuracy	80.3% High	72.1% Low
4	Time	0.84~0.9 sec/frame	0.17~0.23 sec/frame
5	Frame persecond	45	59
6	Mean Averageprecision	0.358	0.251

Table1: Difference between YOLO and SSD

Table 1 compares YOLO with SSD in terms of speed, accuracy, time, frame per second (FPS), Mean Average Precision (mAP), and whether they are suitable for real-time applications. The table above clearly illustrates that YOLO outperforms the SSD method, which has lower accuracy but higher FPS. YOLOv3 runs at 416 X 416 in 29 ms @ 31.0 mAP, which is almost as precise as SSD but 2.2 times faster. It is apparent that a careful tradeoff was made in order to reach this speed. Even with a low mAP, YOLO has a suitable mAP for real-time applications, and it is clear that it is the best algorithm in its class.

VI. APPLICATIONS

A. Advertising Detection

Fig 2. Advertising Detection

In both the virtual and real worlds, detecting advertisement boards has important uses. Google Street View, for example, may use it to update or personalize the advertising that is shown on street photos.

B. Animal Detection

Fig 3. Animal Detection

The YOLO model can be used to detect a variety of creatures. From photos and real-time video feeds and recordings, the YOLO model is capable of distinguishing horse, sheep, cow, elephant, bear, zebra, and giraffe.

C. Object Detection

Fig 4. Object Detection

Object detection is the process of detecting and characterizing a large number of objects in a photograph. The main distinction is the "variable" portion. In comparison to issues like classification, the yield of object detection is variable since the number of distinguishable items varies from picture to picture. The YOLO model can be used to classify a variety of things.

D. Activity Recognition

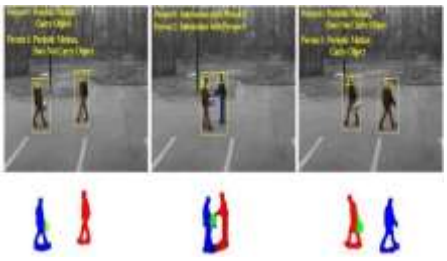

Fig 5. Activity Recognition

The goal of activity recognition is to recognize one or more person' actions and objectives from a set of observations of specialist activity and surrounding circumstances. Many networks have been considered in this area of research because it can provide personalized support for n number of applications and its relationship to a variety of other domains, such as the interaction between the human and machine and the humanistic method.

E. People Counting

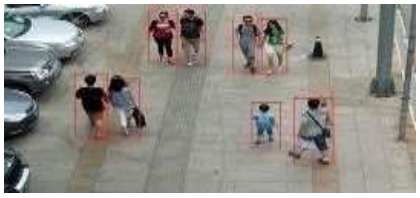

Fig6:People Counting

Object detection can also be used for people checking, as well as for breaking down store performance and crowd measurements during events. These will be more difficult in general, since people will be moving out of the picture at a faster rate.

F. Others

Logo detection and video object detection are two further real-world uses. The detection of logos in web-based business systems is a hot topic of research. When compared to generic detection, the logo event with a clear non-rigid change is substantially less.

VII. CONCLUSION

Due to its strong learning capacity and interest in handling constraint, scale transformation, and context shifts, deep learning-based object detection has become a prominent focus in research in recent years. As a result of the preceding debate, we may conclude that using the YOLO Model in real life can greatly benefit many organizations.

Yolo is a generalized algorithm that outperforms numerous techniques in natural and various areas from object detection. As we are certainly aware, Yolo would have a tremendous impact in commercial and industrial sectors as one of the most promising models. The algorithm's goal is to categories artefacts that employ a single neural network. The algorithm may be quickly displayed and trained on an entire image. The regional techniques outlined above limit the classifier to a single region. YOLO predicts the full picture when it comes to boundaries. Furthermore, it predicts fewer positive outcomes in backgrounds. Other classification methods are far more difficult to utilize in real time than this one.

References

[1]. *Kanishk Wadhwa, Jay Kumar Behera, "Accurate Real-Time Object Detection using SSD" SRM Institute of Science and Technology,*
[2]. *Chennai,2020*
[3]. *Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, "Object Detection with Deep Learning: A Review".*
[4]. *Joseph Redmon, Ali Farhadi, "Yolov3:An incremental improvement", arXiv preprint arXiv:1804.02767, 2018.*
[5]. *Joseph Redmon, Ali Farhad, "YOLO9000: Better, Faster, Stronger", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.*
[6]. *J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition", IJCV, 2013.*
[7]. *Erhan, D., Szegedy, C., Toshev, A., Anguelov, D, "Scalable object detection using deep neural networks". In: CVPR (2014).*
[8]. *Guei-Sian Peng "Performance and Accuracy Analysis in Object Detection" CALIFORNIA STATE UNIVERSITY SAN MARCOS.*
[9]. *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.2016.91*
[10]. *Lin, T.Y., Marie, 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, September 2014*
[11]. *Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C; "SSD: Single Shot Multi Box Detector". In Proceedings of the European Conference on Computer Vision, Amsterdam, the Netherlands, 11–14 October 2016.*
[12]. *K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition Computer Vision and Pattern Recognition", 2014.*