International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A Comparative Study of Object Detection using YOLO and SSD Algorithms

[1]Dinesh Suryavanshi M. Tech Final year student VITM Indore,

[2] Professor Vijay Bisen HOD (E & E Branch) of VITM Indore,

[3] Professor Neha Khatri Professor VITM Indore

Abstract- One of the most significant and difficult areas of computer vision is object detection, which is used in many aspects of daily life, including security monitoring, autonomous vehicle operation, and so on, to find instances of semantic objects belonging to a particular class. In this survey, we first look at the current approaches of typical detection models and discuss the benchmark datasets in order to comprehensively and deeply comprehend the major development status of the object detection pipeline. This project explains how convolutional neural network-based deep learning techniques are used for object detection. In this study, deep learning methods for cutting-edge object identification systems are evaluated. Any kind of visual medium is represented by a computer as a collection of numerical numbers. In order to inspect the contents of images, they need image processing techniques. In order to determine which is the quickest and most effective object detection algorithm, this research analyzes two popular algorithms: Single Shot Detection (SSD) and You Only Look Once (YOLO). The COCO (Common Object in Context) dataset is used in this comparative research to assess the effectiveness of the two algorithms and to analyze their strengths and weaknesses in terms of factors like accuracy, precision, and speed. Based on the findings analysis, it can be said that YOLO-v3 outperforms SSD in the same testing scenario.

Keywords- Object Detection, Deep learning, CNN, COCO, YOLO, SSD

I. Introduction:

Due to its numerous applications and recent technological advancements, object detection has recently been gaining a lot of attention. Numerous studies on this task are being conducted in academia and in practical settings, including security monitoring, autonomous driving, traffic surveillance, drone scene analysis, and robotic vision. The development of deep convolution neural networks and the computational power of GPUs are only two of the numerous elements and initiatives that have contributed to the swift advancement of object detection systems. The application of deep learning models in computer vision is currently widespread and includes both generic and domain-specific object detection. Most modern object detectors employ deep learning networks as their foundation and detection network to extract features from input photos (or videos), classify them, and then locate them.

Deep neural networks, which are comparable to humans in that they are formed of neurons, are utilized for object detection. Consequently, the term "object detection" refers to the identification and location of items in an image that fall under one of a number of established classes. Object detection (also known as object recognition) is a crucial subfield in computer vision because tasks like detection, recognition, and localization have broad applications in real-world contexts.

Object detection can generally be divided into two categories:

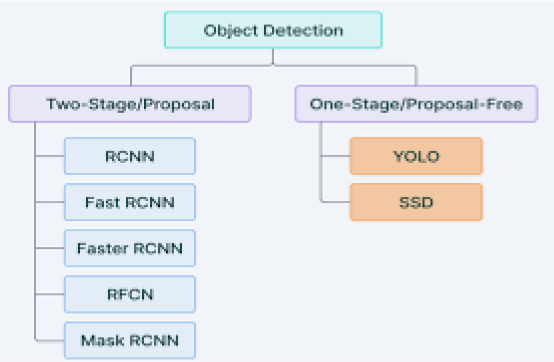

Figure 1 Flowchart for object detection 1

1. **One stage Detector** - where the process of object detection is a straightforward regression problem that takes an input and figures out the bounding box coordinates and class probabilities. The one stage detector includes devices like YOLO, YOLO v2, SSD, RetinaNet, etc. A neural network can predict items in a picture and identify them by drawing bounding boxes around them, which is an advanced method of image categorization.
2. **Two Stage Detector**- in which the detection is finished in two steps. In the first stage, regions of interests with a high likelihood of becoming an object are generated using a region proposal network. The second phase is object detection, during which objects are classified definitively and subjected to bounding box regression. Some examples of two stage detectors include RCNN, Fast RCNN, Faster RCNN, RFCN and Mask RCNN etc.

Our analysis' major goal is to contrast the operational performance and accuracy of the object recognition methods YOLO and SSD in various contexts and highlight some of the noteworthy features that set this study apart.

II. YOLO (You Only Look Once):

II.I. **Exactly what is YOLO?**

You Only Look Once is what the expression "YOLO" means. Using a convolutional neural network (CNN), the YOLO algorithm detects objects in real-time. As the name implies, the approach only requires one forward propagation through a neural network in order to detect objects. This implies that a single algorithm run is used to conduct prediction throughout the full image. Several class probabilities and bounding boxes are simultaneously predicted using the CNN. The key characteristics of YOLO are its quickness, great accuracy, and capacity for learning.

• Speed - For real-time object detection, this technique is faster.

• High Accuracy - This method produces precise results with few background mistakes.

• Learning Capability - This algorithm has exceptional learning skills that allow it to learn object representations and use them in object detection.

Advantages and limitations of one-stage object detectors:

Table 1 Advantages and limitations of one-stage object detector1

Year	Methods	Advantages	Limitation
2016	**YOLO**	a. Fixed the limitation of yolov1. b. More efficient than SSD in real time application. c. Multi-scale training. d. More apt to detect small objects. e. Multi scale prediction.	a. Less accurate than SSD. b. Less efficient than RettinaNet.
2016	SSD	a. End-to-end training. b. Better accuracy than YOLO. c. Multiple sale feature extraction.	a. More time-consuming than YOLO. b. Less accurate than Faster R-CNN.

II.II How the YOLO algorithm works

YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Figure 2. YOLO Algorithm for Object Detection 1

Example of YOLO – YOLO is preferable when a minor inaccuracy may be overlooked. Examples include live traffic monitoring, life form detection in remote regions, monitoring of fruits and vegetables, self-driving vehicles, and cancer recognition techniques.

III. SSD (Single Shot Detector):

III.I **Exactly what is** SSD**?**

Instead of the two shots required by the region proposal network approaches discussed in the preceding section, the SSD enables a single shot to detect multiple objects in the image. Therefore, compared to region-based methods, SSD saves a lot more time. A VGG-16 [58] network is used to extract feature maps from an image as its input. Different filter sizes (1919, 1010, 55, 33, and 11) are used to add a number of convolutional layers. These are the feature maps that the 3x3 convolution filters will process for each cell to derive predictions, along with the 3838 feature map created by conv4_3 of VGG.

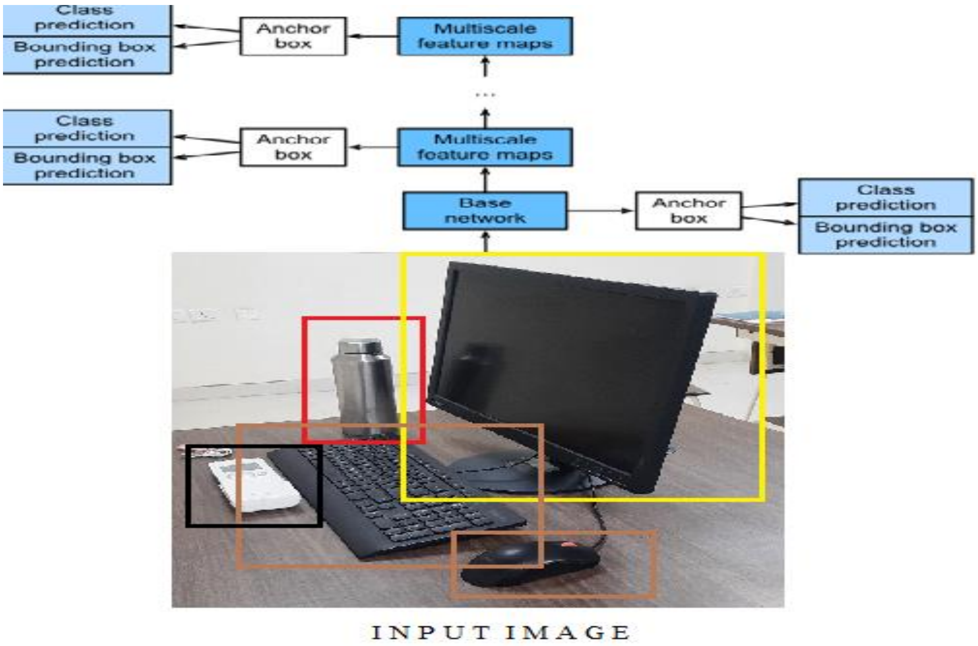

Figure 3. SSD algorithm for object detection 1

The SSD method is based on a feed-forward convolutional network that generates a fixed-size collection of bounding boxes and scores for the existence of object class instances in those boxes. The final detections are then produced by a non-maximum suppression step. Offset values (cx, cy, w, h) from the default box are contained in boxes. Scores include confidence values for each category of object, with the number 0 denoting the backdrop. Multi-reference and multi-resolution detection techniques are introduced by SSD. Multi-reference approaches establish a number of anchor boxes at various points in a picture, each with a unique size and aspect ratio, and then forecast the detection box depending on these references. The ability to recognize objects at various scales and network layer allows for multi-resolution approaches. An approach for identifying numerous object classes in photos is implemented by an SSD network by creating confidence scores for the existence of each item category in each default box. Additionally, it alters boxes to better fit the geometry of the objects. Due to the fact that it does not resample data for bounding box hypotheses, this network is appropriate for real-time applications.

Example of SSD – SSD is advantageous for more accurate object recognition. It is more suited for video forensics, legal investigations, landmark detections, and many more.

IV. COMPARATIVE ANALYSIS

For measuring object localization accuracy, various metrics have been suggested. It is customary to assess the precision of detections using the Intersection over Union (IoU), also known as the Jaccard Index. It can be calculated as the region of overlap of a predicted detection and its associated ground truth divided by the area of union of the expected detection and the ground truth. For binary or multi-class detection issues, the mean IoU for an image is calculated by averaging the IoUs for each class. To obtain an average IoU value, this can be applied to each image in the test collection. Another similar detection statistic is the F1-score (also known as the Dice Coefficient), which is calculated as the area of overlap divided by the sum of the pixels in the detected and ground truth regions. In terms of precision and recall measures, this measurement can be described. We can compute the average F1 score for all the photos in the test dataset and it can also be applied to all the target items that are visible in an image. IoU and F1-score metrics are connected and have a positive correlation for a given fixed ground truth. As a result, when comparing two models using IoU, the first model will always be preferred above the second model when using the F1 score.

Different item sizes, lighting conditions, image perspective, partial occlusion, complicated backgrounds, and scenes with several objects all may be handled by both detectors and yet yield respectable results. The virtually total elimination of FP situations in the SSD model, which is preferred in applications involving analysis, is one of its key advantages. However, YOLOv3 generates better overall outcomes. Although YOLO has trouble accurately localizing objects, SSD is faster than the prior progressive for single-shot detectors.

For real-time applications, efficiency is determined by speed and accuracy. Even up to YOLOv3, YOLO variations offer remarkable accuracy but call for expensive technology. This type would meet the necessary speed requirements for such devices. Similar to YOLOv5s, MobileNet-SSD V2 offers a speed that is somewhat faster, but it is less accurate. When the truth trade-off is very small and we have a propensity to square measure to run it on a video, SSD might be a better option. If exactness is important to you rather than wanting to move quickly, YOLO is a better choice. In light of the requirements of diverse applications, either model may be chosen.

V. CONCLUSIONS

A essential component of surveillance systems in many field applications is the real-time object detection and tracking on video feeds. In order to identify the things in the photos, our work compares two object identification systems employing CNN. For performance assessment in various settings, we investigated and examined the YOLO object detection model and MobileNet SSD model. Each of the models that have been compared has distinct qualities of its own and excels in the applications to which it has been used. In contrast to MobileNet SSD, which offers faster detection, YOLO offers more precision.

References

[1] I. Sabina N. CSE Department MGM College of Engineering and Pharmaceutical Sciences Valanchery, "Object Detection using YOLO And Mobilenet SSD: A Comparative Study," *International Journal of Engineering Research & Technology (IJERT)* , vol. Vol. 11 , no. 06-June-2022, pp. 134-138, 2022.

[2] D. M. C. D. P. P. H. P. S. A. P. a. H. A. P. P. D. o. C. S. S. S. V. c. V. P. E. K. Dawn Wilson1, "Analytical Study on Object Detection using Yolo Algorithm," *International Journal of Innovative Science and Research Technology* , vol. Volume 7, no. 8- August – 2022 , pp. 587-589, 2022.

[3] 2. L. N. 3. D. K. V. R. 1Peddinti Mounika, "A COMPARATIVE STUDY OF YOLO AND SSD OBJECT DETECTION ALGORITHMS," *IJCRT* , vol. Volume 9, no. 11 November 2021, pp. 821-827, 2021.

[4] D. S. (. M. V. B. (. M. N. K. (3), "Object Detection Using Deep Learning," *TIJER*, Vols. Volume 10, Issue 9, no. September 2023, pp. a59-a68, 2023.

[5] E. Güney, C. Bayilmiş and B. Çakan, "An Implementation of Real-Time Traffic Signs and Road Objects Detection Based on Mobile GPU Platforms," *IEEE Access (Volume: 10),* pp. 86191 - 86203, 16 August 2022.

[6] G. Li, Z. Ji, X. Qu, R. Zhou and D. Cao, "Cross-Domain Object Detection for Autonomous Driving: A Stepwise Domain Adaptative YOLO Approach," *IEEE Transactions on Intelligent Vehicles,* pp. 603 - 615, 06 April 2022 .

[7] D. G. Lema, O. D. Pedrayes, R. Usamentiaga, P. Venegas and D. F. Garcí, "Automated Detection of Subsurface Defects Using Active Thermography and Deep Learning Object Detectors," *IEEE Transactions on Instrumentation and Measurement,* p. 4503213, 22 April 2022.

[8] S. Noor, M. Waqas, M. I. Saleem and H. N. Minhas, "Automatic Object Tracking and Segmentation Using Unsupervised SiamMask," *IEEE Access (Volume: 9),* pp. 106550 - 106559, 28 July 2021 .

[9] Y. Dai, W. Liu, H. Wang, W. Xie and K. Long, "YOLO-Former: Marrying YOLO and Transformer for Foreign Object Detection," *IEEE Transactions on Instrumentation and Measurement (Volume: 71),* p. 5026114, 15 November 2022.

[10] A. Koirala, M. Jha, S. Bodapati, A. Mishra, G. Chetty and P. Kishor, "Deep Learning for Real-Time Malaria Parasite Detection and Counting Using YOLO-mp," *IEEE Access (Volume: 10),* pp. 102157 - 102172, 21 September 2022 .

[11] N. Zarei, P. Moallem and M. Shams, "Fast-Yolo-Rec: Incorporating Yolo-Base Detection and Recurrent-Base Prediction Networks for Fast Vehicle Detection in Consecutive Images," *IEEE Access*

(Volume: 10), pp. 120592 - 120605, 14 November 2022 .

[12] Z. Chen, K. Wu, Y. Li and M. Wang, "SSD-MSN: An Improved Multi-Scale Object Detection Network Based on SSD," *IEEE Access (Volume: 7)*, pp. 80622 - 80632, 14 June 2019.

[13] T. Ye, Z. Zhang, X. Zhang and F. Zhou, "Autonomous Railway Traffic Object Detection Using Feature-Enhanced Single-Shot Detector," *IEEE Access (Volume: 8)*, pp. 145182 - 145193, 10 August 2020.

[14] X. Lu, J. Ji, Z. Xing and Q. Miao, "Attention and Feature Fusion SSD for Remote Sensing Object Detection," *IEEE Transactions on Instrumentation and Measurement* , p. 5501309, 22 January 2021 .

[15] Z. Lyu, H. Jin, T. Zhen, F. Sun and H. Xu, "Small Object Recognition Algorithm of Grain Pests Based on SSD Feature Fusion," *IEEE Access (Volume: 9)*, pp. 43202 - 43213, 17 March 2021 .

[16] K. B. Lee and H. S. Shin, "An Application of a Deep Learning Algorithm for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels," *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, no. 21 October 2019, pp. 7281-7291, 26-28 August 2019.

[17] R. Kim, G. Kim, H. Kim, G. Yoon and H. Yoo, "A Method for Optimizing Deep Learning Object Detection in Edge Computing," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, no. 21 December 2020, 21-23 October 2020.

[18] 2. A. 3. T. 4. S. 5. D. 6. S. 7. P. 1Anushka, "Object Detection using Deep Learning: A Review," *Journal of Physics: Conference Series*, 2020.