International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

YOLO vs CNN : A Comparative Study on Object Detection

Meha Shrivastava,
Research Scholar, Deptt of EC
UIT , RGPV
Bhopal (MP) , India
mehakhare@gmail.com,

Dr. Roopam Gupta
Professor, Deptt of IT
UIT, RGPV
Bhopal (MP) , India
roopamgupta@rgpv.ac.in

Abstract—With the availability of large amount of data and increasing demand of digitizing the visual systems, lot of research on object detection techniques has been the foremost requirement since last ten years. As the technological advancements are pacing at such a higher rate with increasing computational power and new architectures for CNN has already been proposed in a successive manner to enhance the detection techniques. For Real time Object identification, CNN has proposed several architectures but YOLO proves itself better than others based on the speed of identification. This paper presents the brief overview and comparative analysis of various object detection techniques like SSD, RCNN, YOLO and their subsequent versions for different applications. On comparing, it is found that single stage detectors outperform two stage detectors.

Keywords—Pascal VOC, MS COCO, YOLO, CNN, Object detection, Object Recognition, SSD.

1. Overview

1.1 Introduction:

With the rapid technical enhancements in the domain of computer vision, very frequent change is observed in the field of object detection as multiple machine learning and deep learning algorithms are developed to enhance the performance of various detection techniques. As detection of object includes object localization and object classification (Tausif Diwan et.al.,2022), it becomes very challenging task to detect the objects in real time dynamic environment. As real time scenarios are concerned demand of more effective technique in terms of accuracy is required on priority and two stage detection techniques somewhere lag in comparison with single stage detectors, so YOLO and its subsequent versions really proved better and performed very well in comparison with two stage detection techniques for object detection and recognition for varied applications. This motivates to write a review paper that will put some light on YOLO and its variants and analyze the differences in their architecture.

1.2 Challenges in Object Detection

Object detection is one of the important topics in the computer vision and nowadays varied algorithms based on machine learning and deep learning are used to enhance the performance in terms of efficiency. But still having several algorithms, it is quite challenging to properly detect and recognize the objects in real time applications like surveillance, mechatronics, dynamic applications like in dense traffic situations, public places like shopping malls. The points which must be taken care of includes:

- Multiple scale & aspect ratios
- Class Imbalance
- Detection of extremely small object
- Availability of large datasets.
- Detection of Multiple moving objects
- Dual priorities

1.3 Types of Object Detection Techniques

With the emerging trends in AI techniques, more focus will be given on deep learning-based object detection algorithms. Comparing with conventional algorithms, it is found that algorithms based on deep learning approaches detect accurate and more precise features of the target. The deep learning- based algorithms are classified in two categories:

1.) **Two-step target detection algorithms** such as Fast R-CNN (Wenyu Liu et.al., 2022) (regional conventional neural network), Faster R- CNN, Mask R-CNN, etc. In this algorithm, the detection done in two stages i.e., the region proposal network

(RPN) is created to extract the candidate target information, and then detection network is used to complete the prediction and recognition of the location and category of the target.

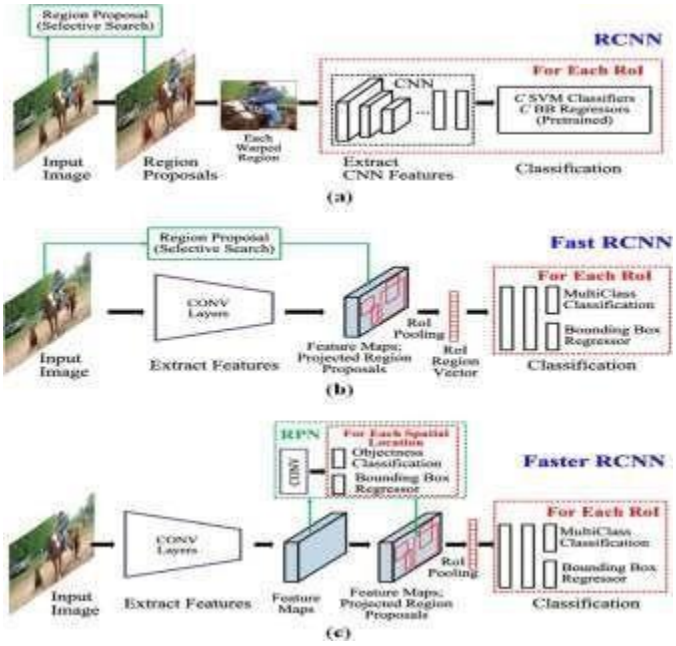

Figure 1: Two stage object detection

2.) Single step target detection algorithms, such as **SSD (Single Shot Multibox Detector), Yolo (You Only Look Once) and its subsequent versions** (Bobburi Taralathasari et.al., 2021) **.** In this algorithm RPN is not used, it directly generates the location and category information of the target through the network. It is an end-to- end target detection algorithm.

So, the single-step target detection algorithm leads two stage detection algorithms in terms of detection speed.

1.4 Datasets

The two main characteristics of object detection algorithm is detection and localization. Detection means to verify the class of the object and localization refers to locate the object through the bounding box. Different datasets for various applications are available which makes the machines to learn more about different object instances. Dataset is a basically a collection of images that is used for training machine learning models to perform tasks like object detection and recognition. Using challenging datasets (Bravo et.al.,2022) helps to set the benchmark for comparing the performance of different algorithms in same application. A large no. of datasets is available amongst which MS COCO (Mansi Mahendru et.al ,2021) and PASCAL VOC is most popular and is continuously upgrading its object collection. The table 1 shows the comparison of various datasets along with the various applications and object instances

Table 1: Comparison of various datasets with their applications and object included

Datasets	Applications	Object Instances
MS COCO	Object Detection & Recognition	330 k images with 92 categories out of which 82 are labelled.
PASCAL VOC	Object detection, semantic segmentation, and classification tasks.	1,464 images for training, 1,449 images for validation and a private testing set.
DOTA	Unmanned Aerial Vehicles	11268 satellite and aerial images
OPEN CV	Video image Stitching, Navigation, Medical analysis	1.9M images with 600 object categories
ImageNet	Image Classification and Object Detection	14,197,122 Annotated images
GTSDB	Traffic Identification Algorithm	43 Traffic Signs 51839 images

1.5 Organization of paper

This paper is divided into seven sections out of which I section comprises of introduction, challenges, types of detection techniques and varied datasets. II section describes about the CNN and its working. III section put highlights on work done.

Research gap and parameters to be evaluated is described in the section IV and V. Comparative analysis is given in section VI. Section VII presents the concluding remarks

.

2. Convolutional Neural Networks

A Convolutional Neural Network (CNN) (Upulie et.al.,2021) is a class of neural network which utilizes the deep learning approaches that can take an input image, assigns weights and bias to different objects in the image to differentiate one from the another. The pre-processing required in a Conv Net is found to be much lower when compared with conventional classification algorithms. The training of CNNs is done using enormous amount of dataset of labelled images. Once trained CNN can be used to classify, extract, and detect features of the images. A CNN comprises of mainly three layers: a convolutional layer, a pooling layer, and a fully connected layer arranged on top of one other.

2.1 Working of CNN:

The working of CNN starts with the input image selection after

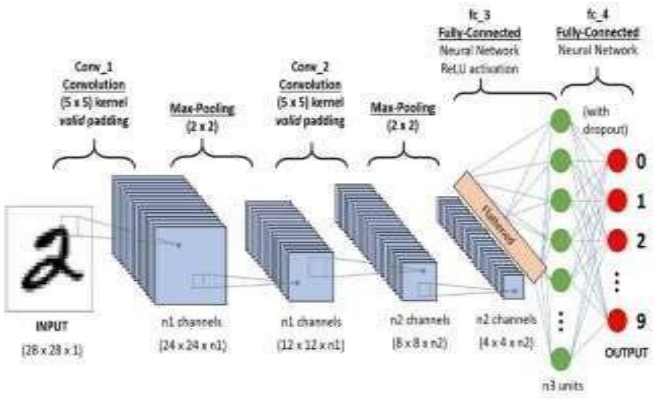

Fig 2: Structure of CNN

which the image is passed through the successive layers of

CNN for further processing. The CNN comprises of three layers:

- Convolutional Layer
- Pooling Layer
- Fully Connected Layer

CNN architecture is like human brain in which billions of neurons are arranged in such a way that will help in detection and recognition process of objects in a better way. The working of CNN is dependent on these artificial neurons which calculates the weighted sum of the inputs and gives the activation function values as the output.

The multiple layers present in CNN architecture is responsible for detecting different features of an input image.

At each successive layer the complexity pattern increases and the first layer detect the dimensional features and passes to the next layer to detect the diagonal features and at last that is at fully connected layer full object recognition done.

3. Major Contributions

Tausif Diwan et al. (2022) explored 2 stage object detection with its applications. Review has been done on YOLO technique, architecture, advancements underlying pretrained architectures,Loss functions etc.

Akansha Bhatija et.al. (2019) explained moving object detection using YOLO detector and SORT tracker utilized custom dataset consisting of 800 images for 6 specific classes but the system is limited to pedestrians and vehicles.

Shailendra Kumar et al. (2021) proposed improved YOLO v3 algorithm for small target detection and confirmed improvement in terms of accuracy, recall and average accuracy.

Upulie H.D.I et al. (2021) reviewed the structure of CNN algorithms and gave the overview of YOLO for real time object detection. It is confirmed that YOLO is having better performance over other algorithms in real time object detection.

Abhinu C G et.al. (2021) proposed multiple objects tracking system using YOLO v5 with Pytorch that can detect objects which were already trained and can track and take count of

objects in each frame. It can also detect objects on CPU GPU.

Ming Li Zhang et al. (2022) put some highlights on the application of CNN in traffic sign detection and classification as really traffic sign detection and recognition is a difficult task in dynamic scenarios.

Peiyuan Jhang et al. (2022) reviewed all versions of YOLO and concluded that different versions of YOLO have many differences but some common features also but every version is improved with its previous one in terms of accuracy, recall and mAP.

Joseph Redmon et el. (2016) introduced new technique for objectdetection as YOLO (You Look Only Once) for real time objectdetection scenario.

Bobburi Taralathsari et al. (2021) developed a console-based application in which input image is taken and object names are written on the top of bounding box which are drawn around theimage. The training of dataset are done using Google Colab and through latest version of YOLO, more accurate output is taken.

4. Research Gap

From the survey of literature, following points have been revealed:

1. Efficiency of most of the object tracking methods degradesunder the presence of multiple moving objects in a scene.
2. There is trade-off between accuracy and inference time in twostage and single stage detection process as two stage detection methods outperform single stage detectors in terms of accuracy with complex architecture while single stage detection method is having simpler design but comparatively better inference time.

5. Evaluation Parameters

- **Mean Average precision:** It is defined as the meanof average precision (AP) of all classes.

$$mAP = \sum_{q=1}^{Q} \frac{Ave(P, Q)}{Q}$$

Where, q= no of queries

Ave (P, Q) =average precision for that given query

mAP may be calculated by taking mean of AP.

- **Precision:** It is defined as the ratio of true prediction to total no. of predictions.

$$Precision = \frac{True\ Positive}{(True\ positive + False\ Positive)}$$

- **Recall:** It is defined as the ratio of true prediction to the total no of objects present in an image.

$$Recall = \frac{True\ Positive}{(True\ positive + False\ Negative)}$$

- **F1 Score:** It is a measure to predict whether a model is accurate or not. Basically, it is defined as harmonic mean of precision and recall. It decides whether a model is accurate or not.

$$F1\ score = \frac{2*(precision * Recall)}{(precision + Recall)}$$

6. Comparative Analysis

6.1 Various Object Detection Techniques

The object detection technique is classified into two types viz: two stage and single stage depending upon the stages involved in the detection process. The table 2 summarizes the comparison of various object detection techniques with respect to some parameters like region proposal, backbone network, no. of stages involved, speed, accuracy, and their costof computation etc. This will help in analyzing and selecting the suitable method for specific application.

Table 2: Comparison of various object detection techniques

Parameters	SSD	R CNN	Fast R CNN	Faster R CNN	YOLO
Region Proposals	Single Convolutional Network	Selective Search Algorithm	Selective Search Algorithm	Selective Search Algorithm	Single Convolutional Network
Feature Extraction (Back Bone Network)	Light Weight & faster	Heavy Weight & Time Consuming	Heavy Weight & Time Consuming	Heavy Weight & Time Consuming	Light Weight & Faster
No of Stages and their role	01	02	02	02	01
Speed & Accuracy	Fast & Accurate	Slow & More Accurate	Slow & More Accurate	Slow & More Accurate	Fast & Accurate
Computational Cost	Less Expensive	Expensive	Expensive	Expensive	Less Expensive

6.2 Comparative Analysis of Subsequent versions of YOLO

From the literature surveyed, it is made clear that single stage detectors worked better that two stage detectors and single stage object detector are mostly suitable for real time object detection however they lag in performance metrics (Bravo et.al.,2022). The table 3 summarizes the use of different versions of YOLO algorithms and each successive variant of YOLO works better than previously one.

Table 3: Summary table based on different versions of YOLO

S. No.	Reference	Algorithms	Findings
1	Peiyuan Jiang et. al. (2022)	Yolo and its subsequent versions	Developments of yolo algorithms were discussed. Yolo to Yolo v5 versions were reviewed.
2	Tausif Diwan et.al. (2022)	Yolo and its subsequent versions	Two stage as well as single stage objects detection techniques were discussed and found that yolo technique are performing better
3	Upulie H.D.I et al (2021)	CNN Architectures with YOLO V1 and YOLO V2	Fundamental structure of CNN were discussed and introduction to yolo technique is given.
4	Joseph Redmon et.al. (2016)	Yolo	Yolo technique is introduced in this paper and found that it can be trained for full images.
5	Bobburi Taralathasri et. al (2021)	Basic Yolo Algorithm	Object Detection system using deep learning technique detects object efficiently using yolo algorithm.
6	Mansi Mahendru1 et al. (2021)	Yolo	Comparison between Yolo and Yolov3
7	Upesh Nepal et. al.(2022)	Yolo v3, Yolo v4, Yolo v5	YOLO techniques were compared for landing spot detection and it is found that yolov5 outperforms better than yolov3 and yolo v4.
8	Srivastava et al. (2021)	Yolo v3, Faster R CNN, SSD	SSD, Faster R CNN, and YOLO were compared and it is found that in identical testing environment, YOLO performs better that all.
9	Ming Li, Li Zhang et.al. (2022)	Yolo	Yolo model was proposed for traffic sign recognition.
10	Wenyu Liu et.al. (2022)	Image Adaptive Yolo	IA Yolo approach was proposed for normal and adverse weather conditions.
11	Diego Met. et.al. (2021)	CNN	Mot Techniques in dense traffic scenario were discussed.
12	Akansha Bathija et.al. (2019)	Yolo, SORT	Moving object detection is done using YOLO detector and SORT tracker.
13	Shailendra Kumar et. al. (2021)	Yolo v4, Deep Sort, Tensor Flow	Identification of objects using Kalman filter and SORT algorithm is used.
14	Zhang Gongguo et.al. (2021)	Yolo v3	Improved yolov3 method is proposed in this and applied for small target detection.
15	Abhinu C G et. al. (2021)	Yolo for MOT	Object detection, counting and tracking is done using YOLO with the help of Pytorch.

7. Conclusion

This paper gives a review on varied object detection techniques. It is clear from the comparative analysis that two stage detectors performs better in terms of accuracy with complexity in architecture but is time consuming. Instead, single stage detector consumes less time with simpler architecture. Also, comparison has been done with the variants of CNN and YOLO

Data Availability Statement:

The data that support the findings of this study are openly available in Coco Dataset

at https://cocodataset.org/#download

Conflict of Interest statement:

There is no conflict of interest amongst authors.

Funding: No fundings granted from anywhere for this research.

References

1. Abhinu C G et. al., Multiple Object Tracking using Deep Learning with YOLO V5: (IJERT) ISSN: 2278- 01 NCREIS - 2021, Volume 9, Issue 13.

2.Ajeet Ram Pathak et. Al. Application of Deep Learning for Object Detection, Procedia Computer Science 132 (2018) 1706–1717 International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

3.Akansha Bathija, et.al, Visual object detection and tracking using YOLO and Deep Sort, IJERT, ISSN 2278- Volume 8 Issue 11, November 2019.

4. Alexey Bochkovskiy et.al, YOLOv4: Optimal Speed and Accuracy of Object Detection arXiv:2004.10934v1 [cs.CV] 23 Apr 2020.

5.Andrew G. Howard et.al., Mobile Nets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861v1 [cs.CV] 17 Apr 2017.

6.Bobburi Taralathasri et.al, REAL TIME OBJECT DETECTION USING YOLO ALGORITHM, International Journal of Computer Science and Mobile Computing, Vol.10 Issue.7, July- 2021, pg. 61-67.

7.Chien-Yao Wang et.al., YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, arXiv:2207.02696v1 [cs.CV] 6 Jul 2022.

8.Diego M Jimenez Bravo et.al., Multi Object Tracking in Traffic Environments: A systematic Literature Review., [cs.CV]27 Sep 2015.

9.Joseph Redmon, et.al, You Only Look Once: Unified, Real-Time Object Detection, May 2016, arXiv:1506.02640v5 [cs.CV].

10.Mansi Mahendru1 et.al., Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3, 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence 2021).

11.Ming Li, Li Zhang, et.al, Yolo based Traffic Sign Recognition Algorithm: Hindawi Computational Intelligence and Neuroscience, Volume 2022, Article ID 2682921.

12.Peiyuan Jiang et al., A Review on YOLO Algorithm developments: Procedia Computer Science199 (2022) 1066–1073.

13. Rekha B.S. et.al, Literature Survey on Object Detection using YOLO, IRJET, Volume 07, Issue 06, June 2020.

14.Rodrigo Verschae et. Al. Object Detection: Current and Future Directions, Frontiers in Robotics & AI, November 2015, Volume 2, Article 29.

15. Ross Girshick, Fast RCNN, arXiv:1504.08083v2 [cs.CV] 27 Sep 2015.

16.Ross Girshick, Rich feature hierarchies for accurate object detection and semantic segmentation, 2014 IEEE conference on Computer Vision and Pattern Recognition, 23-28 June 2014.

18.Sankar K. Pal et.al, Deep learning in multi-object detection and tracking: state of the art, Applied Intelligence https://doi.org/10.1007/s10489-021-02293-7., 2021.

19.Shailendra Kumar, Vishal, Pranav Sharma, Nitin Patel, Object Tracking and Counting in a Zone using YOLOv4, Deep SORT and TensorFlow, ICAIS 2021, ISBN 978-1-7281-95377.

20.Shaoqing Ren et. al., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, arXiv:1406.4729v4 [cs.CV] 23 Apr 2015.

21. S. Manjula et. Al., A STUDY ON OBJECT DETECTION, IJPT| Dec-2016 | Vol. 8 |Issue No.4 | 22875-22885.

22.Srivastava et.al., Comparative Analysis of Deep Learning detection Algorithms, Journal of Big Data (2021) 8:66/Springer 2021.

23.Tausif Diwan et al., Object detection using YOLO: challenges, architectural successors, datasets, and applications: Springer August 2022.

24. Upesh Nepal and Hossein Eslamiat *, Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs: Sensors 2022 464 https://doi.org/10.3390/s22020464.

25. Upulie H.D.I et al., Real-Time Object Detection using YOLO: A review, DOI: 10.13140/RG.2.2.24367.66723, Research gate.

26.Wenyu Liu, et.al, Image Adaptive YOLO for object detection in Adverse Weather Conditions, Thirty Sixth AAAI conference on Artificial Intelligence (AAAI-22).

27. Zhang Gongguo et.al, An improved small target detection method based on YOLO v3, 2021 InternationalConference on Electronics Circuits and Information Engineering (ECIE).