IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Cricket and Football Detection Using YOLOV5 Algorithm

MD Shahnawaz Hussain[1], Rohan Jadhav[2], Rutvik Manthalkar[3], Uday Raj Kushagra[4], Prof. S. S. Peerzade[5]
B.E Students, Department of Computer Science[1,2,3,4]
Assistant Professor, Department of Computer Science[5]
Sinhgad College of Engineering, Pune, India

Abstract: *A cutting-edge technology in artificial intelligence and machine learning is called object detection. It is crucial to understand about object localization before moving on to object detection. For being able to localize object first, we need to know what the object is in the image. Then we need to assign bounding box to that specific object we want to detect in the image or video. Finding a single item's location in an image or video is called object localization. Finding numerous objects' locations in an image or video is called object detection. The use of Object Detection has been increased drastically within last decade. And, it is being used in many areas for refining efficacy in the task. object detection is being used in home automation, agriculture, Automated or self-driving cars, Surveillance industry, traffic tracking system, Activity Recognition, defense systems, sports, industrial work, automobile industries, robotics, aviation industry and many other fields. Object detection can be performed consuming various set of rules like R-CNN, Fast R-CNN, Faster R-CNN, Single Shot detector (SSD) and You Only Look Once(YOLO). For this project an assessment of R-CNN and YOLO algorithms will be performed and also their results as well as performance will be studied. The performance and accuracy should be supreme vital in examining the algorithms.*

Keywords: YOLO algorithms

I. INTRODUCTION

Computer vision has recently been employed in the industrial revolution's task. The automation, robotics, healthcare, and surveillance industries all utilize deep learning extensively [1]. Deep learning has been the most talked approach as a consequence of its discoveries, which are usually accomplished in applications like language processing, object identification, and classification. Outstanding growth is expected during the next several years, according to the market estimate. Strong Graphics Processing Units (GPUs) and a large number of datasets are both readily available, which is one of the primary explanations given for this. Both of these prerequisites are now readily available [1].

Image labelling and recognition are two of object detection's most important building blocks. Many datasets are available for use. One such popular picture categorization tool is Microsoft's COCO. A standard for evaluating object detection algorithms. The availability of a large dataset facilitates image recognition and classification [2].

The study of object detection is crucial to both AI and computer science. The process of object detection, which divides numerous visual components into different categories, is used to identify image and video fragments. Home automation, agriculture, automated or self-driving cars, surveillance, traffic monitoring, activity recognition, defence systems, sports, office work, the automotive and robotics industries, and many more disciplines use object detection. Artificial intelligence is the basic tenet of object detection. Computer vision is used to collect data. Then, these data will be included in state-of-the-art computer machine learning algorithms. An imaging gadget processes them after that. Now, based on the model's training process and the nature of the training data. The system is presented with the outcomes in the form of alerts, directions, or information. [3]

Object detection is a branch of computer vision and image processing that identifies objects of a specific type in digital images, such as people, bicycles, trees, dogs, cats, buildings, automobiles, and books [3]. As the need for object detection grows and its presentation improves over time, object detection has emerged as a major area of research.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Numerous techniques, including R-CNN, Fast R-CNN, Faster R-CNN, Single Shot Detector (SSD), and You Only Look Once, have been developed and are in use (YOLO). This system's assessment of the R-CNN and YOLO algorithms' performance, speed, and outcomes is being done. When examining the algorithms, the presentation and accuracy should be given the utmost importance.

II. LITERATURE REVIEW

Recent years have seen a lot of scientific interest in object detection. Effective learning technologies may be used to quickly locate and analyse deeper aspects. In order to compare different object detection methods and algorithms utilised by various researchers and derive useful conclusions for their usage in object detection, this initiative aims to gather data on these methods and algorithms. A literature review is meant to help readers understand our work.

The Fast R-CNN model was developed by Ross Girshick and published as an approach to identify objects [4]. It use the CNN approach to find targets. Girshick's technique is innovative in that it suggests a window extraction methodology, in contrast to the R-CNN model's more conventional sliding window extraction strategy. Support vector machines for classification and deep convolution networks for feature separation both get independent training [5]. They integrated feature extraction and classification to create a classification framework using the quick R- CNN approach [4]. R- training in a flash In comparison to the R-CNN period, the CNN period is nine times shorter. The region proposal network, a network template used in the faster R-CNN technique, is created by combining the proposal isolation area with a little amount of Fast R-CNN (RPN). Fast R-CNN and Faster R-CNN both produce accurate results. The research [5] asserts that the approach is a hybrid, deep learning-based object identification system that runs at 5-7 frames per second (fps).

This article examines a different investigation Kim et al. undertook. This work develops a system that uses CCTV cameras to identify and distinguish moving things by fusing CNN with background removal. It operates based on the background subtraction technique used for each frame [6]. In our work, we employed a similar architecture to the one in this study.

An other detecting network is YOLO. By Joseph Redmon et al., a single-time convolutional neural network has been proposed. It is employed to categorise a number of candidates and project the placement of the frame. This makes it possible to identify targets during their whole lifecycle. A regression problem is utilised to address the object detection problem [7]. Our YOLO-based bounding box prediction and feature extraction approaches used the strategy outlined in this paper as a reference.

Tanvir Ahmed et al. [8] have proposed a modified approach with a new inception model structure, a particular pooling pyramid layer, and enhanced performance. It utilises a YOLO v1 network model that has been optimised to reduce function loss. This study is used to support the sophisticated application of YOLO. A PASCAL VOC (Visual Object Classes) dataset is also used to undertake a thorough extended experiment. The network is an enhanced model that performs incredibly well [8]. The method outlined in this work was used to train the YOLO model using PASCAL VOC.

A revolutionary technique for object detection in video was developed by Wei Liu et al. utilising just one deep neural network. [9]. The Single Shot MultiBox Detector SSD is the name given to this technique. The team claims that SSD is a straightforward approach that needs an object proposition since its fundamental idea is the total elimination of the process that creates a proposal. Additionally deleted are the subsequent pixel and resampling processes. As a result, everything is completed in one motion. SSD is not only very easy to use, but it's also very easy to incorporate into the system and very easy to learn on. As a result, detection is made easier. The main characteristic of SSD [9] is the link between many feature maps and multiscale convolutional bounding box outputs. The work provided here served as an inspiration for the training and model analysis of the SSD model used in our investigation.

The framework of another paper is a more sophisticated SSD. The authors of the paper propose a one-shot detection deep CNN called Tiny SSD. To make real-time embedded object detection easier, TINY SSD was created. A stack of non-uniform SSD-based auxiliary convolutional feature layers and a non-uniform Fire subnetwork make up its significantly improved layers. Tiny SSD has the advantage of being 2.3 MB smaller than Tiny YOLO. According to the study's conclusions, a tiny SSD is a suitable choice for embedded detections [10].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Pathak et al study's [1] uses CNN for object detection to demonstrate how well deep learning works. In addition, the study employs a few deep learning methods for object identification systems. According to a recent study, deep CNNs work on the basis of weight sharing. It offers details on a few significant CNN points. CNN and their significance in deed learning were understood using the theory put out in this study.

In a recent work by Chen et al. [11], anchor boxes were used for face identification and a more accurate regression loss function. To solve the detection challenges related to different face scales, they released a face detector dubbed YOLO face, which is based on YOLOv3. According to the authors' analysis, their system outperformed earlier YOLO iterations and modifications [11]. The YOLOv3 was used in our study to contrast with other models.

Faster RCNN, cascade RCNN, R-FCN, YOLO and its variations, SSD, RetinaNet, CornerNet, and Objects as Point were all taken into account in a recent evaluation of deep learning-based detectors by Mittal et al. [12]. On these detectors, advanced assessment steps were carried out. The methodology and low-altitude datasets utilised for the relevant investigation are fully summarised in this publication [12]. Similar to the comparison conducted in this study, our comparative analysis, which employed coco metrics, was conducted in a similar fashion. The article also covers other comparative strategies that we took into account when doing our research.

Convolutional neural networks, sometimes known as CNNs, are algorithms for object detection. Another name for it is CompNet. [13] Artificial neural networks are commonly used for image analysis. The hidden layers known as convolutional layers are what make CNN a CNN. A convolutional layer receives input, modifies it similarly to previous layers, and then sends the modified data to the following layer. R-CNN stands for region-based convolutional neutral networks. R-CNN uses selective search methods to extract regions of interest, or ROI. from the input for video or pictures. To indicate an object's perimeter within an image, a rectangular box known as the ROI is employed. There could be a lot of ROIs in the image. A neural network accepts each ROI and then generates output features. A collection of support vector machine classifiers is used to determine the type of object that is present in the ROI [13].

The stiff and unlearning Selective Search algorithm is quite bad. Sometimes, this can lead to the creation of subpar region recommendations for object detection. Considering the roughly 2000 proposed candidates. The network needs a lot of time to be trained. Additionally, we need to practise each step independently (CNN architecture, SVM model, bounding box regressor). Implementation proceeds exceedingly slowly as a result. Given that it takes roughly 50 seconds to evaluate an image with a bounding box regressor, R-CNN cannot be utilised for real-time object detection. Since all feature maps for the region proposed must be stored. Additionally, it raises the requirement for disc RAM for training [14].

III. METHODOLOGY

In this method, the item in the scene is detected using the YOLOv5 and RCNN algorithms. This section contains a thorough discussion of various algorithms.

3.1 YOLOv5

A company called Ultralytics released YOLOv5 in 2020. It immediately gained popularity after being added to a GitHub repository by Glenn Jocher, the founder and CEO of Ultralytics. The YOLOv5 object detection model was also available in the iOS App Store under the "iDetection" and "Ultralytics LLC" app names. Despite assertions that the Ultralytics version is the most sophisticated YOLO implementation currently on the market, the version's trustworthiness has remained in question in the community. The layout of the YOLOv5 object detection model is shown in Figure 1.

Since YOLO v5 is a single-stage object detector, it also has three essential parts, like other single-stage object detectors.

- YOLOv5 Backbone: To extract features from images made out of cross-stage partial networks, it employs CSPDarknet as the framework.
- YOLOv5 Neck: It uses PANet to construct a network of feature pyramids for feature aggregation and forwarding to Head for prediction.
- YOLOv5 Head: layers for object detection that use anchor boxes to generate predictions.
- Activation and Optimization: In YOLOv5, leaky ReLU and sigmoid activation are employed, with SGD and ADAM potential optimizers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

- Loss Function: Binary cross-entropy with logits loss is used.

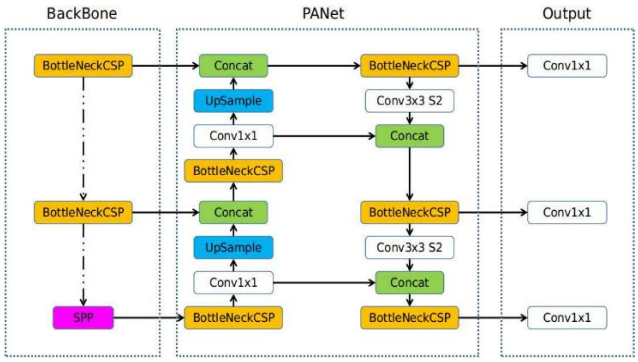

Figure 1. Architecture of YOLOv5 algorithm[15]

3.2 Faster RCNN

The first usable target detection model based on convolutional neural networks, the R-CNN [16] method, was reported by Girshick in 2014. mAP is 66% for the updated R-CNN model. The model starts by removing roughly 2000 area concepts from each image, which must be located via selective search, as illustrated in Figure 1. The retrieved picture attributes are then given to the SVM classifier for classification after being evenly scaled to a feature vector of a preset length. A trained linear regression model is then used to execute a bounding box regression process. When compared to the traditional detection strategy, the R-CNN does greatly improve accuracy, but it does so inefficiently and with more calculations. Second, the objects might be distorted if the region recommendation is directly converted to a fixed-length feature vector.

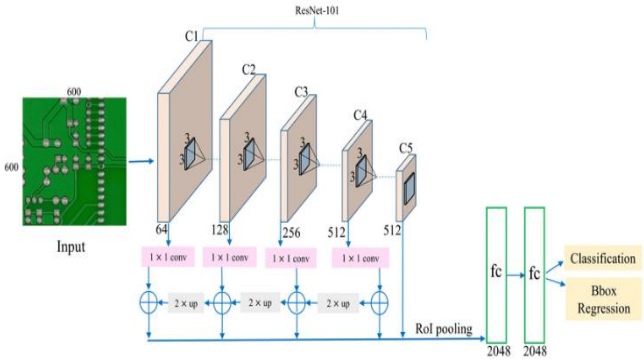

Figure 2. Architecture of RCNN algorithm[17]

Family R-CNN wasn't quick enough. It took longer to find the predicted region for the bounding box, train a model, find and categorise areas, and then check for advanced yields independently. When high levels of precision (like those provided by CNNs) are not necessary, it makes sense to rely on less precise but faster to train algorithms. Hence, YOLO's extraordinary ascent to fame. It initially increases detection times since it predicts objects in real-time. Second, YOLO presents trustworthy findings with minimal background errors. Additionally, the approach has excellent learning capabilities that allow it to understand item representations and use them in object identification tasks. The convergence of each of these characteristics explains why YOLO is so widely used [18].

Faster-RCNN improved the R-CNN model sequentially rather than using Selective Search, an expensive technique for creating bounding boxes and sending images to the CNN. Combining the SVM classification and bounds. Return to the first box. Additionally, the ROI (Region of Interest) pooling layer is implemented as a region of proposals to reshape from the convolutional feature map. such that they can be included into a layered relationship that works perfectly.

Faster RCNN consists of mainly four parts:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

- Conv Layers: Faster In order to detect targets for the CNN network, RCNN first extracts image feature maps using a series of core Conv+ReLU+pooling layers. The same feature maps are shared by following RPN layers and fully connected layers.
- Region Proposal Network: Region proposals are created through the RPN network. The anchors in this layer are corrected using bounding box regression after using softmax to detect whether they are in the foreground or background.
- ROI Pooling: Fully connected layer receives the feature maps and input suggestions from this layer, analyzes the data to choose the appropriate feature maps, and uses those feature maps to identify the target category.
- Classification: In the end, it makes use of the suggested feature maps.

Using the information from the network below, each convolution layer generates an abstract description. While the second layer takes up information about more complicated shapes and other things by observing pattern patterns in edge edges, the first layer often gathers knowledge about edge edges. After that, convolutional characteristics can be obtained. Although significantly shallower than the original picture, the spatial dimension is still less. The depth of the feature map is enhanced by the number of filters the convolutional layer has learnt, while the height and breadth of the feature map are decreased by the pooling layer between the convolutional layers.

The traditional detection approach generates a detection frame slowly. For instance, the Selective Search technique is used by R-CNN to create a detection frame. Instead of employing the conventional sliding window and SS technique, the Faster RCNN builds the detection frame immediately utilising the RPN. This is also another significant benefit of the Faster R-CNN, which might considerably speed up the construction of the detection frame.

Actually, the RPN network is split into two lines: one for the foreground and background, which are determined by the softmax classification anchors, and the other for determining the bounding box offset for the anchors in order to determine the precise proposal. By creating bounding box traction offsets and foreground anchors, the final proposal layer is in charge of gathering suggestions and eliminating those that are too small or beyond the bounds. Actually, the target placement function is performed here by the entire Proposal Layer network.

We can obtain a large number of object suggestions without a class score after RPN processing. The issue now is how to categorize them using these borderlines. Some of the simplest ways involve cropping each proposal, feeding it into a trained base network, and then extracting the features to train the classifier. Due to the need for computations for every 2000 suggestions, this is inefficient and performs slowly. It is more efficient to reuse the conv feature map by using ROI (region of interest) Pooling to extract a fixed-size feature map for each proposal.

The ROI Pooling layer must compute and put together the proposed feature maps before passing them to the following network. Two inputs are used by the ROI pooling layer: the original feature maps and the proposal boxes for the RPN output (different sizes).

The Classification determines which class each proposal belongs to using the proposal feature maps, the full connect layer, and the SoftMax function, and then returns the probability vector. The bounding box regression is once more used to concurrently acquire each suggestion. The position offset is used to provide a frame with a more precise object detection.

IV. IMPLEMENTATION

This system was implemented using the Python programming language on the Windows platform. The system's training procedure is carried out using the PyTorch library. While matplotlib is used to display the data, OpenCV is utilised to carry out image processing-related tasks. The model is trained using unique data consisting of annotated photographs of football. The dataset includes 190 photos for validation and 679 images for training. Using the internet platform Make sense, the data is annotated. The YOLOv5 and RCNN algorithms are used to train the data.

The following metrics are used to assess how well the proposed system performs.

Precision

The percentage of all detection results that are accurately detected is known as precision.

$$Precision = \frac{TP}{TP+FP} (1)$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Recall

The accuracy of a positive prediction produced in the presence of a positive input is measured by recall. It only describes how well the model can recognize it.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

A number found to suit an object is called a TP (True Positive). False Positive (FP) indicates that it has been identified as an item of a different class. Or to put it another way, it's a false detection. False Negatives are defined as anything that should have been detected but wasn't, and True Negatives are defined as nothing that should not have been identified.

F1-Score

Instead of using the arithmetic mean, it is computed as the harmonic mean of recall and accuracy. The accuracy of detecting an item increases as the F1-score number increases from zero to one.

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} (3)$$

Mean Average Precision (mAP)

How accurate the projected result is indicated by the mean value of the AP (mean average precision), also known as the mAP (mean average precision).

$$mAP = \frac{1}{|Q_R|}\sum_{q=Q_R} AP(q) \ (4)$$

V. EXPERIMENTAL RESULTS

In this part, the system's outcomes are presented. Two algorithms, YOLOv5 and RCNN, are employed in this system to train and test the football photos. The effectiveness of a predictor is determined by the loss function that was used to classify the input data points in a dataset. The classifier makes more accurate predictions about the relationship between the input data and the output target as the loss value decreases. The consistent decrease in loss value after each session in Figure 3 represents the YOLOv5's gradual learning process. The curves created by the YOLOv5 loss function stabilize after five epochs.

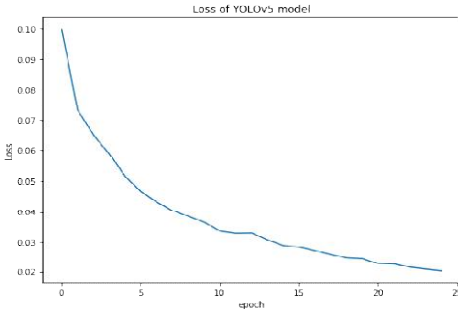

Figure 3. Training loss of YOLOv5 model

Using the PyTorch package, the YOLOv5 algorithm is implemented. 25 epochs are used to train the suggested system. Table I displays the model's loss after every 5 epochs.

Table I: Loss of the faster YOLOv5 model

Epoch	Loss	Execution time (sec)
5	0.04772	65
10	0.0337	65
15	0.02832	66
20	0.02324	64
25	0.02048	65

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

According to Fig.3, the YOLOv5 model's training loss is getting smaller with each passing epoch. Mean Square Error is used to calculate the model's loss (Error). We can infer from the graph that the YOLOv5 model produced superior outcomes. Figure 4 displays the YOLOv3 model's testing outcomes on the test image.

Figure 4. Qualitative analysis of the proposed system using the Yolo V5 model

From Figure 4, it is observed that the YOLOv5 algorithm detects the football and person correctly.The performance plot of the YOLOv5 model is present in Figure 5

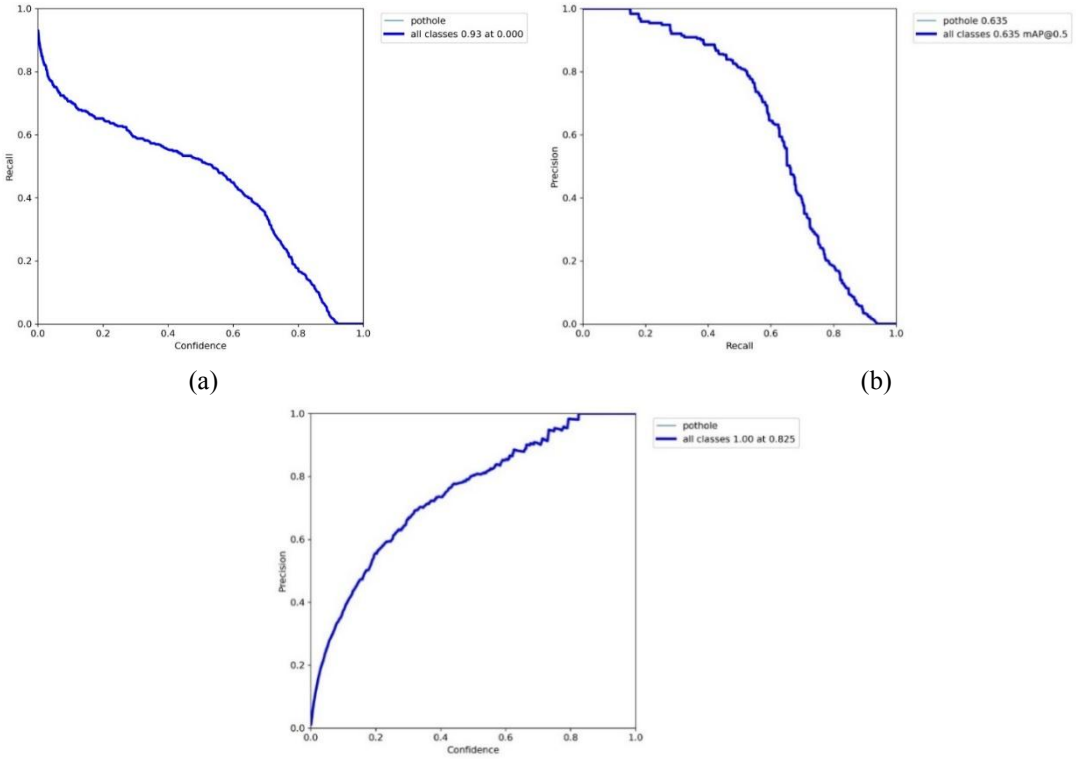

Figure 5. Performance plot of the YoloV5 model (a) Confidence Vs. Recall (b) Precision Vs. Recall (ROC) (c) Confidence Vs. Precision

Figure 5 illustrates how well the full-topology confidence score matches the observed recall for object detection in the dataset (a). Like the confidence scores, the entire confidence score should be interpreted under the assumption that the object was correctly identified.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Figure 5 displays the YOLOv5 training's accuracy vs. recall graph (b). Curves show how a statistical model with adjustable probability thresholds trades off true positive and false positive rates. Consider the accuracy during training as the area under the curve. The curve should ideally move from P=1, R=0 at the top left to P=0, R=1 on the bottom right in order to catch the entire AP (area under the curve). The threshold can be changed to run the model. Because of the huge region under the PR curve's observation, it will exhibit more accuracy.

Figure 5 illustrates how well the full-topology confidence score matches the dataset's observed object identification accuracy (c). Like the confidence scores, the entire confidence score should be interpreted under the assumption that the object was correctly identified.Using PyTorch and the detecto package, the faster RCNN algorithm is developed. 25 epochs are used to train the suggested system. Table II displays the model's loss after each 5 epochs.

Table II: Loss of the faster RCNN model

Epoch	Loss	Execution time (sec)
5	0.1551004589938703	983
10	0.157975669342645	982
15	0.155353497295456	978
20	0.15636332428973726	980
25	0.15698674087454532	977

The graphical representation of losses of the model per epoch is shown in Figure 6.

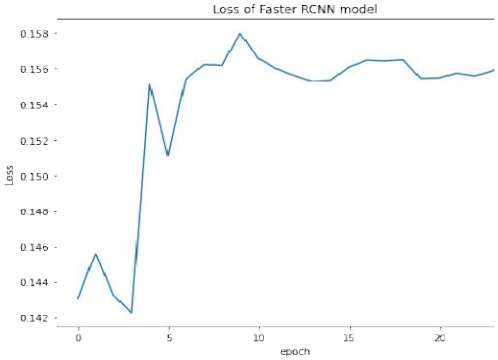

Figure 6. Performance plot of the Faster RCNN model using Loss parameter

Figure 7 displays the comparison of the loss of the YOLOv5 and RCNN models. It can be seen from the graphical analysis that the Faster RCNN outperforms the YOLOv5 model. In comparison to the YOLOv5 model, the loss of the Faster RCNN model is smaller.

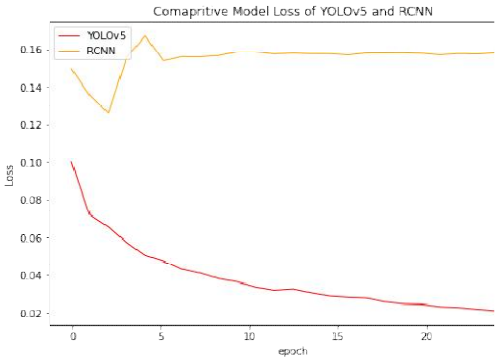

Figure 7. Comparative analysis of object detection usingYOLOv5 and Faster RCNN in terms of model loss

Figure 8 displays the qualitative analysis of the quicker RCNN model on the unidentified image.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Figure 8. Qualitative analysis of the proposed system using the Faster RCNN model

Football detection using a faster RCNN algorithm yielded encouraging results on a custom dataset.

VI. CONCLUSION

The football and cricket ball object detection using the YOLOv5 and Faster RCNN algorithms has been presented in this method. Football and cricket balls are included in the online dataset collection. Making use of the MakeSense platform, the dataset is first labelled. With the help of a faster RCNN algorithm and optimised training parameters, the data is trained. The model is evaluated using an unidentified image sample, and the findings are provided. Yolov5 has lesser loss than the quicker RCNN algorithm, according to the qualitative and quantitative analysis. By applying the system to more things in the future, the suggested system can be enhanced. By hyper-tuning the model's parameters, the loss can be reduced.

REFERENCES

[1]. Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. Procedia Comput Sci. 2018; 132:1706–17.

[2]. Palop JJ, Mucke L, Roberson ED. Quantifying biomarkers of cognitive dysfunction and neuronal network hyperexcitability in mouse models of Alzheimer's disease: depletion of calcium-dependent proteins and inhibitory hippocampal remodeling. In: Alzheimer's Disease and Frontotemporal Dementia. Humana Press, Totowa, NJ; 2010, p. 245–262.

[3]. "Object Detection Explained: Tensorflow Object Detection: AI ML for Beginners: Edureka." Available online: https://www.edureka.co/blog/tensorflow-object-detection-tutorial/

[4]. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2016;39(6):1137–49.

[5]. Ding S, Zhao K. Research on daily objects detection based on deep neural network. IOP Conf Ser Mater Sci Eng. 2018;322(6):062024.

[6]. Kim C, Lee J, Han T, Kim YM. A hybrid framework combining background subtraction and deep neural networks for rapid person detection. J Big Data. 2018;5(1):22.

[7]. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016, pp. 779–788.

[8]. Ahmad T, Ma Y, Yahya M, Ahmad B, Nazir S. Object detection through modified YOLO neural network. Scientific Programming, 2020.

[9]. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. Ssd: single shot multibox detector. In: European conference on computer vision. Cham: Springer; 2016, p. 21–37.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

[10]. Womg A, Shafiee MJ, Li F, Chwyl B. Tiny SSD: a tiny singleshot detection deep convolutional neural network for real-time embedded object detection. In: 2018 15th conference on computer and robot vision (CRV). IEEE; 2018, p.95101

[11]. Chen W, Huang H, Peng S, Zhou C, Zhang C. YOLO-face: a real-time face detector. The Visual Computer 2020:1–9.

[12]. Mittal P, Sharma A, Singh R. Deep learning-based object detection in low-altitude UAV datasets: a survey. Image and Vision Computing 2020:104046.

[13]. "Yolo: Real-Time Object Detection Explained." V7, Available Online: https://www.v7labs.com/blog/yolo-object-detection

[14]. "R-CNN: Region Based CNNS." Geeks for Geeks, Pawangfg, 1 Mar. 2020, Available Online: https://www.geeksforgeeks.org/rcnn-region-based cnns/

[15]. How to Use Yolo v5 Object Detection Algorithm for Custom Object Detection. Available online: https://www.analyticsvidhya.com/blog/2021/12/how-to-use-yolo-v5-object-detection-algorithm-for-custom-object-detection-an-example-use-case/

[16]. Girshick, R., Donahue, J., Darrel, T.,Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: Computer Vision and Pattern Recognition. Columbus.2014, pp. 580-587.

[17]. Ding, Runwei& Dai, Linhui& Li, Guangpeng& Liu, Hong. (2019). TDD-Net: A Tiny Defect Detection Network for Printed Circuit Boards. CAAI Transactions on Intelligence Technology.4. 10.1049/trit.2019.0019.

[18]. The evolution of YOLO: Object detection algorithms. Available online: https://blog.superannotate.com/yolo-object-detection/