IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

REAL TIME OBJECT DETECTION AND CLASSIFICATION USING YOLO

[1]Manjunatha. H. R, [2]K. S. Vanishree
[1,2]Assistant Professor, Department of Computer Science, Government First Grade College, Shivamogga, Karnataka, India.
Email ID's: hrmanjunath.spr@gmail.com, vanishree.kss@gmail.com

ABSTRACT: Object detection is a very important aspect of multimedia applications. Object detection algorithm can help automatically detect cattle movements, traffic signals, and road lanes for self-driving vehicles to reach their destinations. The main goal of object detection is to scan digital images or real-life scenarios to locate instances of every object, separate them, and analyze their necessary features for real-time predictions Over the past two decades, computer vision has received a great deal of coverage. Object recognition and object detection are sub fields of computer vision, the task of giving computers the ability to perceive and respond to the world around them. This is a very useful technology and has many The recent innovations that are being deployed in the current era with the latest trends in technological advancements have made researchers and scientists develop systems that are capable of identifying objects using various machine learning and deep learning algorithms. Hence, in this work, real time object detection and classification using YOLO (You Look Only Once) is presented. YOLO is a powerful technique as it achieves high precision whilst being able to manage in real time. This paper explains the architecture and working of YOLOv3(version 3) algorithm for the purpose of detecting and classifying objects. The performance of presented algorithm is validated in terms of Accuracy, Sensitivity and Specificity.

KEYWORDS: Object detection, recognition, Deep Learning, You Look Only Once (YOLO).

I. INTRODUCTION

Object detection is the procedure of determining the instance of the class to which the object belongs and estimating the location of the object by outputting the bounding box around the object. Detecting single instance of class from image is called as single class object detection, whereas detecting the classes of all objects present in the image is known as multi class object detection.

Humans have a very special ability to detect and recognize person even in difficult scenes such as occlusions, poor lighting conditions, multiple objects, different poses, etc. The ubiquitous and wide applications like scene understanding, video surveillance, robotics, and self-driving systems triggered vast research in the domain of computer vision in the most recent decade. Being the core of all these applications, visual recognition systems which encompasses image classification, localization and detection have achieved great research momentum [1].

Quick, exact calculations for object detection would permit computer to drive vehicles without particular sensors, empower assistive gadgets to pass on constant scene data to human clients, and open the potential for universally useful, responsive automated frameworks. Object discovery includes identifying locale of interest of object from given class of picture. Visual object tracking is one of the most important areas of computer vision. Tracking objects is the process of tracking over time a moving object (or several objects) [2].

Researchers are trying very hard so that they could build a system which is as accurate as human vision, to detect and classify objects from images and videos. Object detection and classification has gained much importance among several computer vision (CV) tasks and is facing an exposure in applications like Object Character Recognition (OCR),

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

Autonomous Driving (AD), object tracking, face recognition etc

An object is having a specific structure, texture as well as some specific pattern. In natural environments, it is difficult to differentiate between the same types of objects because of high variation. The performance of an object detector reduces due to the lighting condition, change of appearance, and at what angle the object is facing towards the camera. Most object detector fails when some deformat ion happens to the object or changes in scale happens to the object [3].

Computer vision is one of the technologies that aim at digitally perceiving the real world at a higher level through digital images and videos. Object detection, a subset to computer vision is one of the prominent techniques in this area of research. Object detection is basically an algorithm based on either machine learning or deep learning approaches employed for classification of elements in diverse classes and localization in the image. Object detection and classification has gained much importance among several computer vision (CV) tasks and is facing an exposure in applications like Object Character Recognition (OCR), Autonomous Driving (AD), object tracking, face recognition etc.

Artificial Intelligence is a topic that is getting a lot of attention in technological advances and as mentioned earlier has many applications in many different fields. In this digital era, there is a tremendous growth in the area of artificial intelligence and machine learning. The aim of doing such experiments is to build machines that mimic the detection capability of humans. This growth has provided various options to the researchers. Also, the introduction of machine learning started using deep learning approaches to provide a big field for research. The fields like artificial intelligence, speech recognition, face recognition, object detection are the various areas of the applications of deep learning [5].

The field of artificial intelligence based object detection suffers with complexity due to the need of high accuracy and precision, especially in modern applications like autonomous cars, face detection, traffic sensing etc. With better resources provided, a computer can be trained as well for object detection and classification and obtain good results. Earlier, problem was handled by Support Vector Machines (SVM), scale-invariant feature transform (SIFT), Histogram of Oriented Gradients (HOG), etc.

Deep learning technology has become a buzzword nowadays due to the state-of-the-art results obtained in the domain of image classification, object detection, natural language processing. The reasons behind popularity of deep learning are two folded, viz. large availability of datasets and powerful Graphics Processing Units. As deep learning requires large datasets and powerful resources to perform training, both requirements have already been satisfied in this current era. Unlike machine learning, deep learning proves to be a more compact process in which the classification and localization occurs in a single trail after the image is captured. The techniques like R-CNN, Fast R-CNN, YOLO have emerged eventually making the process faster and compact [4].

YOLO is one of the fastest algorithms out there to detect objects. Although it is no longer the most accurate algorithm for object detection, when you need real-time detection without losing too much precision, it is a very good choice. YOLO uses a single convolutional network to predict several bounding boxes and category probabilities for these boxes at the same time. Hence in this work, real

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

time object detection and classification using YOLO is presented. The remaining work is as follows: The section II describes the literature survey. The section III presents real time object detection and classification using YOLO. The section IV validates the result analysis and finally this work is concluded in section V.

II. LITERATURE SURVEY

Shrikant Jagannath Patro, Prof. Nisha V M et. al., [6] describes Real Time Video Analytics for Object Detection and Face Identification using Deep Learning. This paper proposed the model that is capable of performing video analytics at large scale and faster pace and generated the appropriate inference on time. It include the detail of the algorithm for the automation in video analytics personalized cameras, security and Surveillance system using deep learning. This include much optimized algorithm for identification of the faces. The second module is consisting of object identification using deep learning and libraries that are capable of identifying 3 million object i.e. COCOAPI. Video Summarization is a technique introduced to increase the speed of investigation. The Module produced much optimized summary of video.

Syed Mazhar Abbas, Dr. Shailendra Narayan Singh et. al., [8] describes Region-based Object Detection and Classification using Faster R-CNN The work focused on using Region Proposals Network(RPN) to extract region of interest in an image.RPN outputs an image based on the objectness score. The output objects are subjected to Roll Polling for classification. This research work focuses on training Faster R-CNN using custom based data set of images. This trained network efficiently detects objects from an image consisting of multiple objects. This network requires minimum GPU capability of 3.0 or higher.

Manisha Domale, Dr. Vijay Gaikwad et. al., [9] describes Robust Pedestrian Detection Framework using Harris Corner Detector and Kalman Filter. Multiple pedestrian detection in video sequences is proposed. Feature points in each and every frame are detected and then classified as foreground and background features. Foreground feature points then used to detect the pedestrian in the frame. These foreground feature points are then updated for every frame. Then foreground is extracted by background subtraction algorithm in combination with harries feature point detector to have better results. Moving pedestrian is obtained from above method is then tracked by Kalman Filter which will track the pedestrian with minimum bounding box. Experimental results show that the proposed method gives better result.

Raji priya.R, Saini Jacob Soman et. al., [10] describes Hybrid Approach for Object Detection from Dynamic and Static Backgrounds for Surveillance Systems This paper proposes a hybrid approach for object segmentation in both dynamic and static backgrounds. In this work, the objects from dynamic backgrounds are segmented by using a neural-fuzzy method. The neural stage is based on self organizing map like architecture. The fuzzy stage automatically computes and adjusts the main parameters required for segmentation without human interaction. Also the moving objects from static backgrounds are detected and tracked by using Kalman filter method and morphological operations.

III. OBJECT DETECTION AND CLASSIFICATION

In this work, real time object detection and classification using YOLO is presented. The workflow diagram of presented approach is shown in Fig. 1. The main goal of an object detection algorithm is to detect an individual object and its category

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

(e.g., pedestrian, vehicle). In this approach, CIFAR-10 dataset is used to train the YOLO. Real time images are captured through camera are collected and pre-processed to test the YOLO.

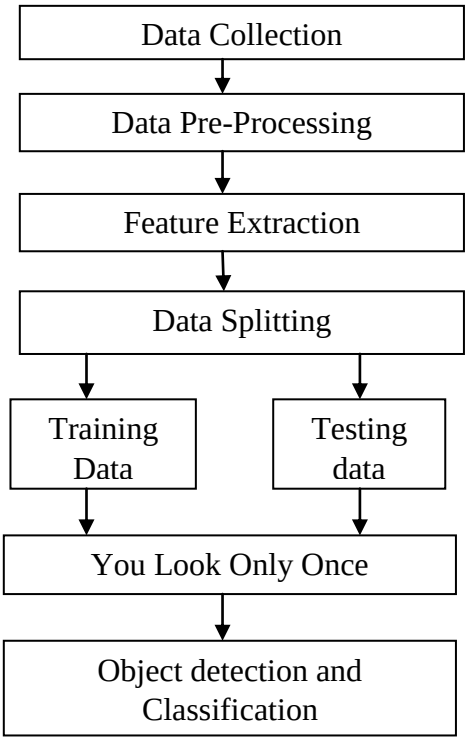

Fig. 1: workflow of Real Time object detection and classification using YOLO

The CIFAR-10 (Canadian Institute For Advanced Research) dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The dataset images include different classes which are as follows: airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck.The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class. The classes are completely mutually exclusive. There is no overlap between automobiles and trucks. "Automobile" includes sedans, SUVs, things of that sort. "Truck" includes only big trucks. Neither includes pickup trucks.

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used. Data preprocessing is the required task for cleaning the data and making it suitable for the classifier model which also increases the accuracy and efficiency of classification algorithm.

Feature extraction identifies the most discriminating characteristics, which a machine learning or a deep learning algorithm can more easily consume. Training machine learning or deep learning directly with raw signals often yields poor results because of the high data rate and information redundancy. Feature extraction plays a key role in image processing. Along with other tools, this technique is used to detect features in digital images such as edges, shapes, or motion. Once these are identified, the data can be processed to perform various tasks related to analyzing an image. The detection stage input consists of high-level features. These features may be computed by a deep feature extractor such as Resnet, Inception or MobileNet.

Then the data is divided into testing (20%) and training (80%) data. The testing and training data is applied to YOLO algorithm. Images were labeled in the YOLO format.. All the images have a specified .txt associated to them after annotation were done in the format of YOLO. You only look once or YOLO is an algorithm for object detection, which has features much different from the

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

previous algorithms. YOLO is an algorithm which operates on the basis of probability of objects in bounding box and grids. The proposed algorithm, YOLO version-3 consists of total 106 layers. The architecture is made up of 3 distinct layer forms. Firstly, the residual layer which is formed when activation is easily forwarded to a deeper layer in the neural network. In a residual setup, outputs of layer 1 are added to the outputs of layer 2. Second is the detection layer which performs detection at 3 different scales or stages. Size of the grids is increased for detection. Third is the up-sampling layer which increases the spatial resolution of an image. Here image is up sampled before it is scaled. Also, concatenation operation is used, to concatenate the outputs of previous layer to the present layer. Addition operation is used to add previous layers.

This algorithm deploys bounding boxes and classifies the image hence predicting possibilities for these boxes using a unit convolutional network. YOLO works on input image by splitting it into an SxS grid, in each of which "m" bounding boxes are marked. In these bounding boxes, the network predicts class probability and offset values. The bounding boxes exceeding the threshold value for class probability are selected and are instrumental in locating the object inside the image. YOLO processes the image as fast as 45 FPS than the rest algorithms for object detection. This YOLO algorithm detects and classified the objects as birds, cat, dog, vehicles, cars and trucks, etc.

IV. RESULT ANALYSIS

In this section, real time object detection and classification using YOLO is implemented. The result analysis of presented approach is evaluated here. The performance of presented approach is validated in terms of Accuracy, Sensitivity and Specificity.

Accuracy: Accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. It is defined as the ratio of number of correctly classified instances to the total number of detected instances.

Sensitivity: It is also known as True Positive Rate (TPR) or Recall. It is measured as the ratio of number of correctly detected positive instances to the total positive instances.

Specificity: described as the algorithm/model's ability to predict a true negative of each category available. It is defined as the ratio of ratio of correctly detected negative instances to the total number of negative instances.

The table 1 indicates the performance evaluation.

Table 1: Performance Evaluation

Metrics/Methods	Naïve Bayes Classifier	Presented YOLO algorithm
Sensitivity (%)	80.23	93.34
Specificity (%)	78.15	94.32
Accuracy (%)	82.34	94.78

Compared to naïve bayes classifier, presented approach has better results in terms of sensitivity, specificity and accuracy. The Fig. 2 shows sensitivity and specificity comparison.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

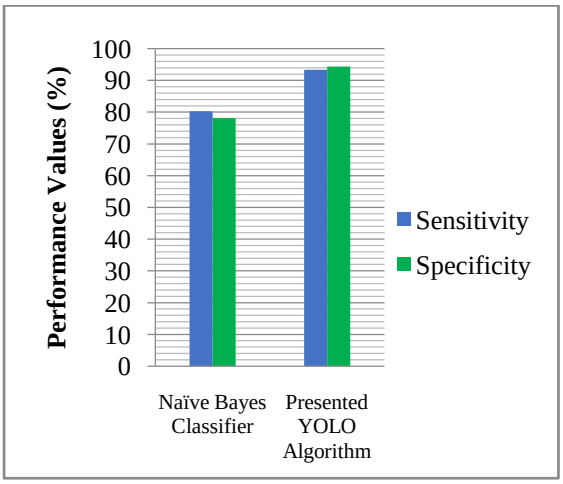

Fig. 2: Performance Metrics Comparison

In fig. 2, the x-axis represents different algorithms and y-axis represents performance values in terms of percentage for object detection and classification. Presented YOLO algorithm has high sensitivity and specificity than previous naïve bayes classifier approach. The Fig. 3 shows the accuracy performance comparison.

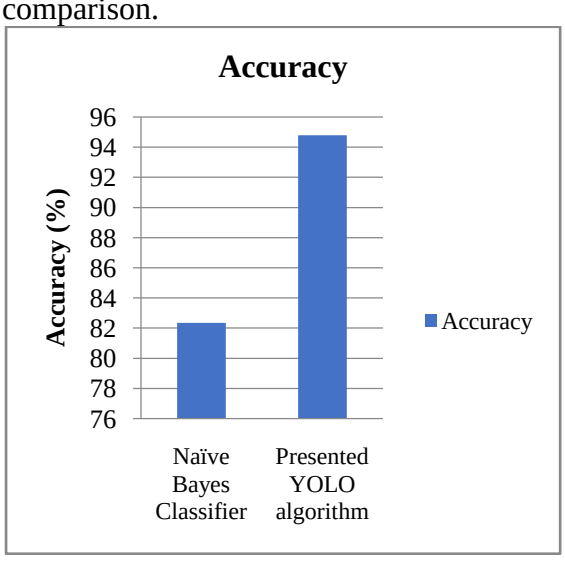

Fig. 3: Accuracy Comparison

In fig. 3, the x-axis represents different algorithms and y-axis represents Accuracy performance values in terms of percentage. Compared to naïve bayes classifier, presented YOLO algorithm has better accuracy. Hence presented approach has accurately detected various real time objects like car, dog, birds, cats, vehicles, etc. hence this approach will be used for real time applications.

V.CONCLUSION

An object detection algorithm can help automatically detect cattle movements, traffic signals, and road lanes for self-driving vehicles to reach their destinations. This, in turn, eliminates the need for drivers for logistic errands. To achieve this, real time object detection and classification using YOLO is presented in this work. In this approach, CIFAR-10 (Canadian Institute For Advanced Research) dataset is used which contains 60000 32x32 colour images in 10 classes, with 6000 images per class. YOLOv3 is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. This algorithm deployed bounding boxes and classified the images as different classes. The performance of presented approach is evaluated in terms of accuracy, specificity and sensitivity. compared to previous algorithms, presented approach has better results in terms of Accuracy, Sensitivity and Specificity. Presented algorithm has effectively detected and classified various real time objects like car, dog, birds, cats, vehicles, etc. hence this approach is used for real time applications.

VI. REFERENCES

[1] Chisulo Mukabe, Nalina Suresh, Valerians Hashiyana, Titus Haiduwa,William Sverdlik, “Object Detection and Classification Using Machine Learning Techniques”, DSMLAI '21', August 9–12, 2021, Windhoek, Namibia , 2020 Association for Computing Machinery, ACM ISBN 978-1-4503-8763-7/20/06, doi:10.1145/3484824.3484895

[2] M. Sushma Sri, B. Rajendra Naik, K. Jaya Sankar, “Object Detection Based on Faster R-CNN”, International Journal of Engineering and Advanced Technology

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 7, 2022

(IJEAT) ISSN: 2249-8958, Volume-10 Issue-3, February 2021
[3] D. K. Behera and A. Bazil Raj, "Drone Detection and Classification using Deep Learning," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 1012-1016, doi: 10.1109/ICICCS48265.2020.9121150.
[4] Shifa Shaikh, Vrushali Karale, Gaurav Tawde, "Assistive Object Recognition System for Visually Impaired", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 09, September-2020, ISSN: 2278-0181
[5] B. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.
[6] Shrikant Jagannath Patro, Prof. Nisha V M, "Real Time Video Analytics for Object Detection and Face Identification using Deep Learning", International Journal of Engineering Research & Technology (IJERT), Vol. 8 Issue 05, May-2019, ISSN: 2278-0181
[7] Praahas Amin1, Anushree B. S.2, Bhavana B. Shetty3, Kavya K.4, Likitha Shetty, "Object Detection using Machine Learning Technique", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05, May 2019
[8] Syed Mazhar Abbas, Dr. Shailendra Narayan Singh, "Region-based Object Detection and Classification using Faster R-CNN", International Conference on "Computational Intelligence and Communication Technology" (CICT 2018), pp. 1-6, doi: 10.1109/CIACT.2018.8480413.
[9] Manisha Domale, Dr. Vijay Gaikwad, "Robust Pedestrian Detection Framework using Harris Corner Detector and Kalman Filter", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 6 Issue 02, February-2017
[10] Raji priya.R, Saini Jacob Soman, "Hybrid Approach for Object Detection from Dynamic and Static Backgrounds for Surveillance Systems", International Journal of Engineering Research & Technology (IJERT), IJERTIJERT, ISSN: 2278-0181, IJERTV3IS090564, Vol. 3 Issue 9, September- 2014