Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Real-Time Object Detection

Dr. N Lakshmi Prasanna[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Ch Janaki Annapurna[2], **G Yeshwanth**[3], **G Shaik Shabina**[4], **B Prema Tejalingam**[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1,2,3,4,5] nlakshmi@vvit.net, chjanaki01@gmail.com, bintu.yeshwanth@gmail.com, shabinaanju@gmail.com, prematejaballa@gmail.com.

DOI:10.48047/IJIJFANS/V11/I12/207

Abstract

Object detection is a pivot and prime process in various applications and procedures such as surveillance, classification, recognition and prediction including image retrieval, computer visioning, video streams and many more. Real-time object detection requires identification of different kinds of objects specified in images, videos or live feed streams. It is basic and important to maintain the level of accuracy along with quick inference. This paper proposes an algorithm to perform the real-time object detection typically leverage machine learning, deep learning to produce effective results. The goal is to power machines to identify defined objects in a live feed, videos or images and achieve desired outcomes. The algorithm used is YOLO version3 (YOLOv3). It presents a fast and accurate object detection method with higher performance. To create reliable applications for resolving practical problems, computer vision techniques like tracking and counting are combined. It is the improved proposal to many machine learning algorithms like CNN, RNN, YOLO v1 and v2. Some of the applications are traffic surveillance, vehicle detection, face detection and recognition, number-plate detection, overspeed tracking and more. The factors included are segmentation, accuracy, precision, fastness, performance and efficiency. It is useful to meet the needs of growing technology.

Keywords: YOLO, Object detection, Real-time, Machine Learning, Deep Learning.

Introduction

The growth and expansions of the technological development is at rapid rate. It is necessary to always keep up and be ahead of the pace. Every activity in society is interlinked with technology. Technology provides ease and comfort by way of use. Improvements in technology results in individual and economic developments in a competitive world. It is important to adapt and work for faster, quicker, more efficient techniques of the existing technologies. One such process is object detection in real-time. It is considered as a prior work in many applications including recognition, detection, classification, prediction and more.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Real-time object detection

In real-time digital pixel images and videos, it is a computer technology used for computer vision and image or video processing that finds instances of semantic items of particular classes. Classes can be defined to accomplish this. Applications range from smart assistants to face security to surveillance.

Fig.1. object detection in real-time

The Fig.1 shows an example by detection of persons, car, traffic light, handbag, backpack in real-time scenario.

YOLO algorithm

A well-known and widely used real-time object recognition tool is the YOLO (You Only Look Once) algorithm, which can locate objects in an image or video frame with high accuracy and quick inference [1]. Between 2016 and 2023, it has several working versions, ranging from version 1 (v1) to version 7 (v7). In order to anticipate bounding boxes, sometimes referred to as anchor boxes, and class probabilities for each grid cell, YOLO divides the input into a grid. The YOLO algorithm is a potent tool for object detection that has a number of essential qualities.

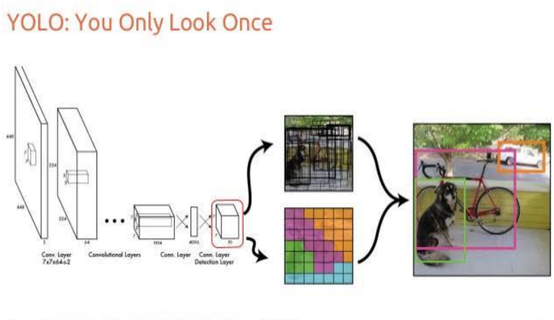


Fig.2. YOLO (You Only Look Once)

The Fig.2 shows the algorithm YOLO in technical manner from convolutions to gridding and then the detection.

Single-stage detection

Compared to other algorithms like RNN, R-CNN and fast R-CNN it performs object detection and classification in a single stage which leads in faster and fewer false positives in results [6].

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Grid-based approach

The input is divided into a grid by YOLO, which then forecasts bounding boxes (also known as anchor boxes) and class probabilities for each grid cell. It is helpful for pixelated digital inputs like pictures or videos. This contributes to more accurate detection of items with various sizes and ratios [9].

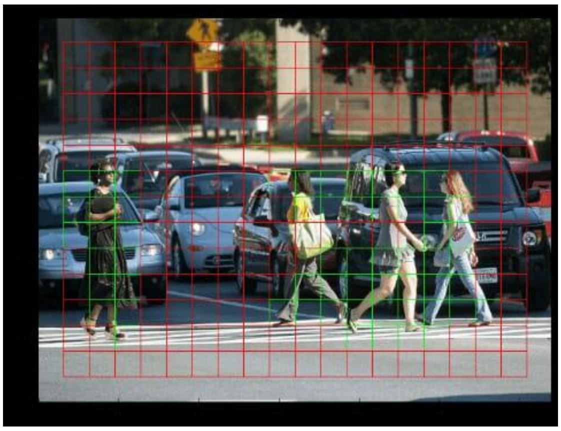


Fig.3. Griding of the input

The Fig.3 shows the grid structuring of the input which is basis for further procedure in the approach.

Non-maximum suppression

In YOLO, non-maximum suppression (NMS) is useful to remove redundant bounding boxes (anchor boxes) and improves the accuracy. NMS compares the confidence scores of overlapping bounding boxes (anchor boxes) and removes the ones with lower scores [5].

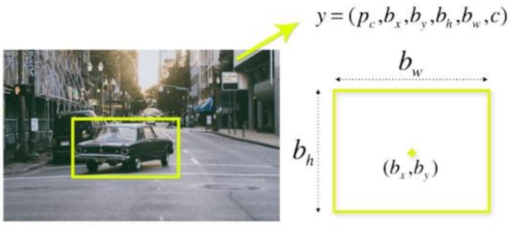


Fig.4. Boundary box of detected object

The Fig.4 depicts the bounding box formation of the detection with given dimensions.

Pretrained models

Pretrained models have the ability to detect a wide variety of items since they have been trained on huge datasets. Many technologies, including robotics, surveillance systems, and self-driving cars, can benefit from the YOLO approach. It is favored by researchers, developers, and practitioners and can be utilized to produce performance that is state of the art [5].

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

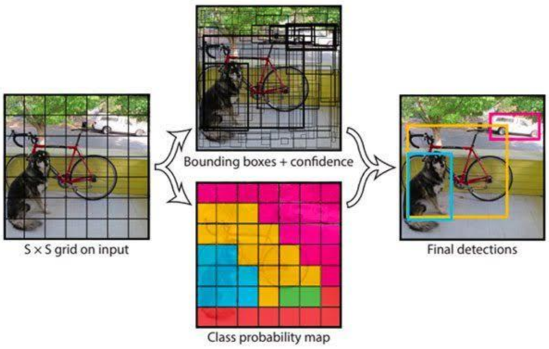


Fig.5. YOLO process

The Fig.5 describes how the YOLO approach gives the final detections from gridding, boundary boxes with confidence score, probabilities.

Literature Survey

The table presents the comparison of real-time detectors and less than real-time. It is evident that the real-time detectors are more precise and has better frame rate than the less than real-time detectors.

Real-Time Detectors

Algorithm	mAP	FPS
100Hz DPM	16.0	100
Fast YOLO	52.7	155
YOLO	63.4	45

Less than Real-Time Detectors

Algorithm	mAP	FPS
Fast DPM	30.4	15
R-CNN	70.0	0.5
YOLO	66.4	21

The Fig.6 shows the mean Average Precision (mAP) of various methods and a plot graph is presented over the values. This provides the comparison of precision of detection of objects. It has been plotted over COCO dataset among the YOLOv3 and other algorithms. Higher mAP shows the higher accuracy and precision of object detection [8].

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

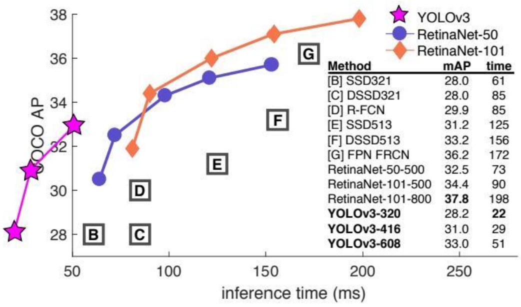


Fig.6. Average precision graph

Problem Identification

With the fast technological and software enhancements, it is important to develop even more sophisticated techniques. One such domain is object detection in real-time. The pivot task of the object detection in real-time is to detect, identify, determine, where and which objects are located in the input feed either images or videos or live stream and which category the specific object belongs to. It is defined as object localization and object classification. It has to scan digital (pixeled) inputs or real-life scenarios to perform localization, analyzation and prediction. The object detection is a prior task and part of data architecture.

Methodology

YOLOv3

The algorithm YOLOv3 (You Only Look Once version3) is suggested in this research for the real-time object detection process. It uses deep convolutional neural network features, a more accurate version than the original techniques, to learn and recognize the items. Deep learning libraries such as Keras or OpenCV are used to do this [3]. Several artificial intelligence (AI) applications use object classification systems to recognize defined items in designated classes. Sorted and grouped together are the objects that share similar qualities while the others are disregarded [7]. The Convolutional layers learn the characteristics that are then passed on to classifiers, allowing for detection and prediction. Because it employs 1x1 convolutions, the name fits. This suggests that the prediction map's size is the same as the feature map's size.

The Fig.7 depicts the process of YOLO version3 approach which includes the convolutions network as of CNN.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

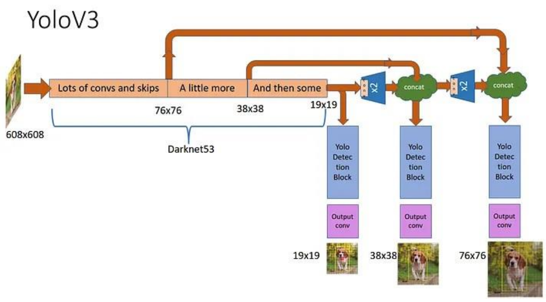


Fig.7. YOLOv3 process

Convolutional Neural Network

YOLO is a Convolutional Neural Network (CNN) that primarily functions as a real-time object detector [2]. These systems use classifiers to process the input as structured arrays of data and identify patterns therein. As a result, the model's predictions can take into account the entire input when put to the test. Regions are graded according to how closely they resemble particular classes. Positive detections are areas that scored highly. The convolutions of various types, sizes, and filters are shown in the Fig.8.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
	Convolutional	32	1 × 1	
1×	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
	Convolutional	64	1 × 1	
2×	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig.8. various Convolutions

Architecture

The YOLOv3 algorithm divides the input into grids at first, then predicts a random number of boundary boxes (anchor boxes) around the items that perform well in the established classes for each grid. Only one object is detected by each boundary box, which has a corresponding confidence value based on how accurate the forecast should be. The ground truth boxes' dimensions from the original dataset are clustered to find the most typical

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

forms and sizes before being used to create the border boxes. In contrast to other systems, YOLO can conduct classification and bounding box regression simultaneously (single-stage). Speed, accuracy, precision, frame rate and class specificity have all improved.

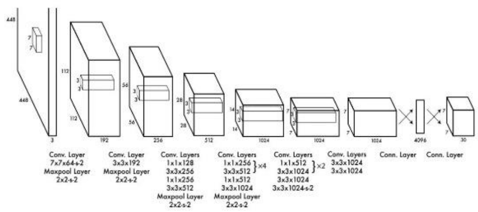


Fig.9 YOLOv3 architecture

The Fig.9 shows the architecture from the internal site of YOLOv3.

Speed

YOLOv2's backbone feature extractor was Darknet-19, whereas YOLOv3's is Darknet-53[3]. Because there are 53 convolutional layers instead of 19, it is more effective. Compared to ResNet101, it is 1.5 times faster. Hence, YOLOv3 performs better without the requirement for model retraining while also being faster and more accurate in terms of mean average precision (mAP) and intersection over union (IOU) values. As a result, it operates significantly more quickly than earlier detection methods.

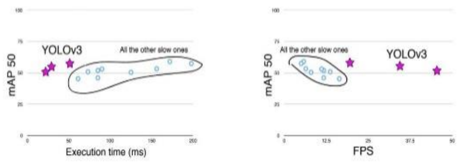


Fig.10. YOLOv3 speed

The Fig.10 shows improved speed, execution time, frame rate of YOLOv3.

Precision

Greater accuracy is produced by higher average precision (AP). YOLOv2 had unmatched precision for small objects with an AP of 5.0 when compared to other systems like RetinaNet (21.8) or SSD513 (10.2). YOLOv3 achieved an AP advancement of 13.3 percent [2].

Specificity of classes

In order to forecast classes during training, YOLOv3 utilizes independent logistic classifiers and binary cross-entropy loss [4]. This enables the usage of challenging datasets like the Open Pictures Dataset (OID). OID typically has overlapping labels in datasets. The multilabel technique used by YOLOv3 allows for more detailed classes and numerous bounding boxes per individual. The mathematical function known as a softmax, which was employed by YOLOv2, transforms a vector of numbers into a vector of probabilities, with the probability of each value being proportional to the relative scale of each value in the vector. These restrictions, which require each bounding box to belong to a single class, are inefficient, especially when using OID datasets.

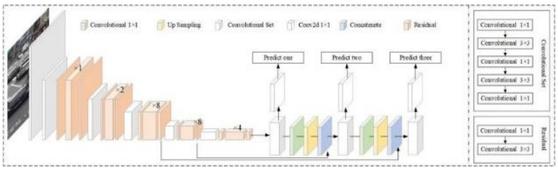


Fig.11. YOLOv3

The Fig.11 shows the working of YOLOv3 approach.

Implementation

This paper proposes the YOLOv3 algorithm for system implementation for performing the task of real-time object detection.

Prerequisites

For the working of this paper proposed algorithm or system implementation, the following are necessary requirements for proper execution:

• Python installed on system (preferred Python 3).

• Anaconda installed on system.

• Jupyter notebook with python or Anaconda supported kernel.

• OpenCV, pip and dependencies installation.

• Datasets and input files containing the predefined classes as COCO names, configuration file of YOLOv3, YOLOv3 weights.

System implementation

Step 1: Install the packages required.

Step 2: Install numpy and argparse.

Step 3: Download the datasets and input files along with configuration files.

Step 4: Run the file with commands or run orderly cells in Jupyter for execution.

Procedure

• Execution is done when all requirements have been imported and installed.

• Create the arguments and the argument parsing.

• Load the COCO (Common Object in Context) class labels that were used to train the YOLO model.

• Create a list of colors at the beginning to represent each potential class label (random).

• Assign paths to the model configuration and YOLO weights.

• Load the COCO dataset to load the YOLO object detector (predefined 80 classes).

• Take the input's spatial dimensions after loading it.

• Choose the output layer names that YOLO requires.

• Create a blob (Binary Large Object) from the input and use the YOLO object detector's forward pass, yielding bounding boxes and related probabilities.

• Create appropriate initial lists of confidences, class IDs, and bounding boxes that were discovered.

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

• Repeat after each layer's output.

• Iterate through each detection.

• Get the confidence (probability) and class ID for the current item detection.

• Eliminate faulty forecasts by making sure that the detected probability exceeds the minimal probability.

• Rescale the enclosing box coordinates in accordance with the image's dimensions. YOLO returns the bounding box's width and height along with the center (x, y)-coordinates.

• To determine the top and left corners of the bounding box, use the center (x, y)-coordinates.

• Update the class IDs, confidences, and bounding box coordinates list.

• Use non-maxima suppression to get rid of overlapping, weak bounding boxes.

• Verify that there is at least one detection.

• Repeat the index loop.

• Extract the coordinates for bounding box.

• Create a rectangular bounding box and label it on the image.

• Display the results.

Results

The proposed algorithm performs the real-time object detection in images, videos and also live feed. The objects are boxed with colors and also provide the accuracy and precision of identification. The outcomes of the proposed technique are depicted in both images and videos. Even multiple object detection of different kinds of defined classes occurred in a single input.

Accuracy table of various algorithms on COCO (Common Object in Context) dataset is provided.

ALGORITHM	ACCURACY
Faster R-CNN	21.9
R-FCN	31.5
SSD300	23.2
SSD512	26.8
YOLOV3	33
RetinaNet	40.8
FPN	33.9

YOLOv3 is better than CNN, Faster R-CNN, R- FCN, SSD and the previous versions of YOLO in terms of accuracy but RetinaNet and FPN gives improved accuracy over YOLOv3.

IMAGES:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022**

Fig.12. Detection in traffic

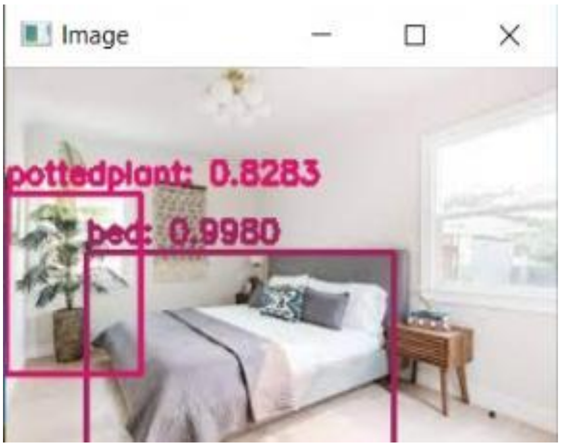
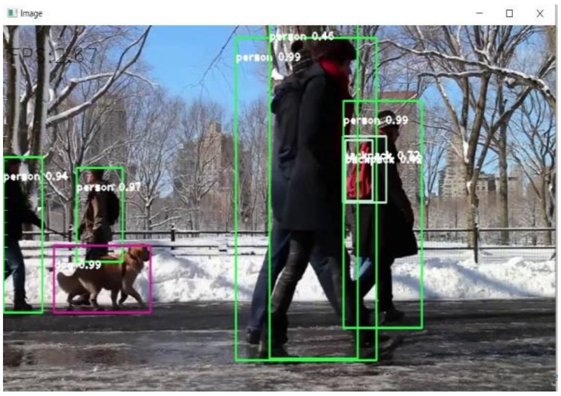


Fig.13. Object Detection in a room

Fig.14. Detection on road

We can observe multiple kinds of objects detected in the presented input through boundary boxes over the objects in Fig.12, Fig.12, Fig.13, Fig.14, Fig.15.

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022



VIDEOS: Fig.15. Real-time Object Detection in video

Conclusions

So, the proposed paper provides the suitable solution algorithm-YOLOv3 for the defined problem statement. Real-time objection is performed with the predefined classes to determine and identify the objects in the input. Also, the performance is better with greater speed, accuracy, precision which improves effectiveness and efficiency along with reliability. Finally, system identifies the specific object in the feed.

Future scope

The domain of object detection is very wide. Especially in real-time, the growing advancements require paced up developed technologies. The newer versions of YOLOv3 include YOLOv4, YOLOv5, YOLOR, YOLOv7 and many more. They are developed by using Pytorch unlike YOLOv3 used Darknet [6]. There is always an open to developments of existing ones. To achieve better outcomes with more speed and performance also overcoming the limitations of the previous.

References

[1] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[2] J. Redmon, A. Farhadi, "YOLO9000: Better, Faster, Stronger",2017.

[3] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement",2018.

[4] Wu, Jianxin "A scalable approach to activity recognition based on object use", 11th International Conference on Computer Vision, IEEE 2007.

[5] Oza, Poojan; Sindagi, Vishwanath A.; VS, Vibashan; Patel, Vishal M., "Unsupervised Domain Adaptation of Object Detectors: A Survey",2021.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[6] Khodabandeh, Mehran; Vahdat, Arash; Ranjbar, Mani; Macready, William G., “A Robust Learning Approach to Domain Adaptive Object Detection”,2019.

[7] Soviany, Petru; lonescu, Radu Tudor; Rota, Paolo; Sebe, Nicu., “Curriculum self-paced learning for cross-domain object detection”, Computer Vision and Image Understanding,2021.

[8] Menke, Maximillian; Wenzel, Thomas; Schwung, Andreas, “Improving GAN-based Domain Adaptation for Object Detection”, 25th International Conference on Intelligent Transportation Systems (ITSC), IEEE 2022.

[9] Menke, Maximillian; Wenzel, Thomas; Schwung, Andrea, “AWADA: Attention-Weighted Adversarial Domain Adaptation for Object Detection”,2022.