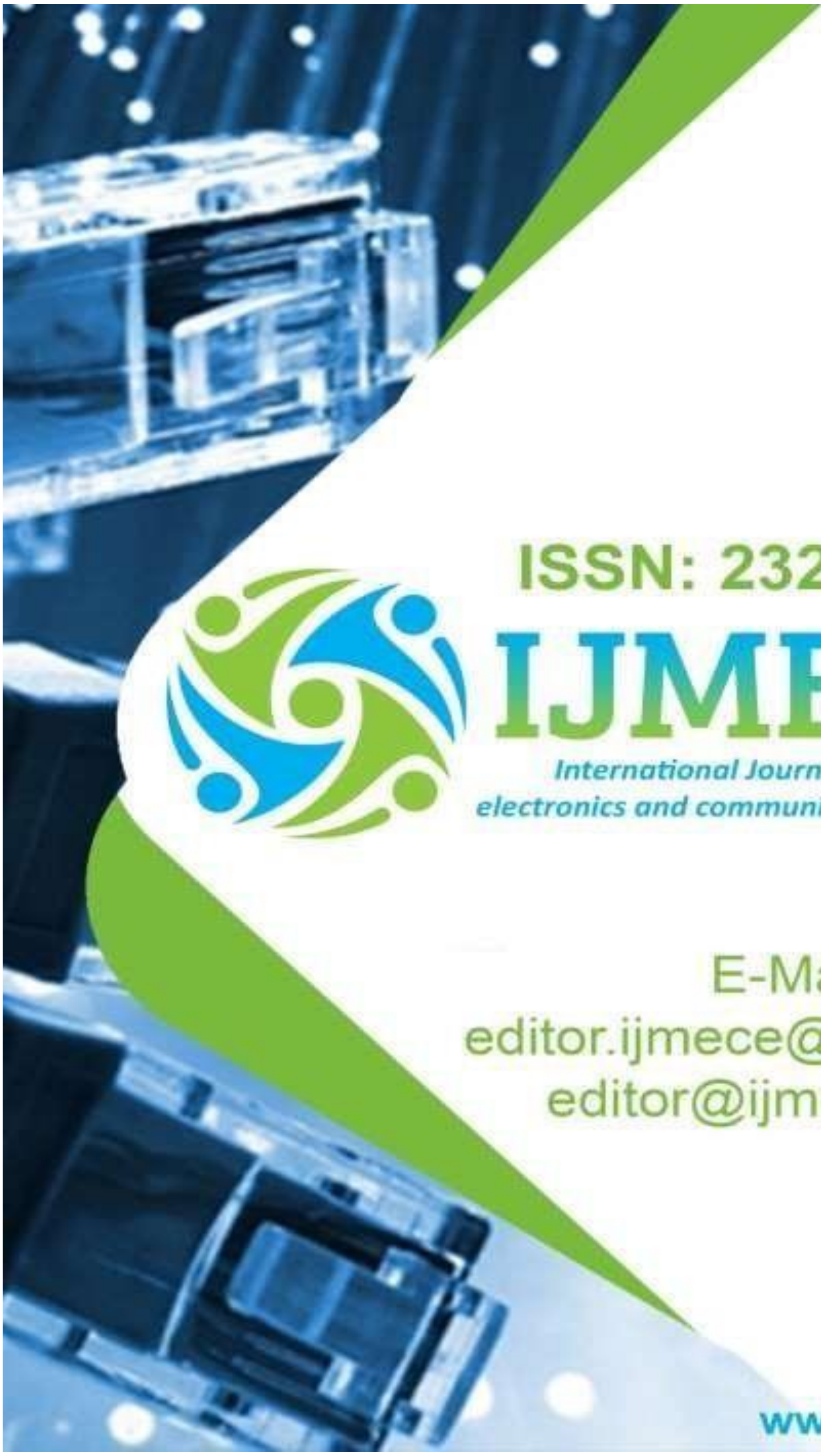

ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

REAL TIME OBJECT DETECTION IN PYTHON USING CNN ARCHITECTURE MODEL

Ms. B.Hemavathi, Assistant Professor, Department Of ECE SICET, Hyderabad
Donuru Poojitha Reddy, Anugu Ruchitha Balreddy, B Siri Chandra, A Avinash
UG Student, Department Of ECE, SICET, Hyderabad

ABSTRACT

Object detection is a computer vision technique for finding patterns of objects in images or videos. Existing object detection systems include region-based convolutional neural networks (R-CNN), single-shot detectors (SSD). The biggest problem of R-CNN family networks is their slow object detection speed. SSD cannot detect small objects. We proposed real-time object detection using You Only Look Once (YOLO) v3 Convolutional Neural Network (CNN) algorithm. This increases speed, real-time, precise and accurate identification.

Keywords: object detection, region-based convolutional neural network, convolutional neural network.

INTRODUCTION

In recent years, significant advances have been made in the field of machine learning and artificial intelligence, leading to increased accuracy, reduced human effort, and improved failure rates. This development played a commendable role in reducing processing time, leading to increased net productivity and reduced associated costs. To explore the application areas of machine learning systems, consider a situation where you want to track a lost cell phone in a cluttered house. Apparently,

A tiring and boring job for everyone. Tracking the location of your mobile phone only takes a few milliseconds. This is exactly the power you can use the amazing object recognition algorithms that are at the heart of deep learning algorithms. The current research work is focused on proposing an object recognition model that can receive the input of a web camera, locate the objects through the web camera, and classify the objects on the screen into appropriate categories. Finally, the goal of the current work in object detection is to take a raw image as input and find the location of that object in the image.

Carefully capture a given image and overlay or classify the objects into appropriate categories [1].

Object detection and tracking is one of the most common and difficult tasks surveillance systems must perform to determine meaningful events and suspicious activity and automatically annotate and retrieve One video content. In business intelligence concepts, objects are not only products, but also faces, heads, people, rows of people, and crowds.

In artificial vision, convolutional neural networks are distinguished in image classification. In this paper, SSD and mobile network based algorithms for detection and tracking are implemented in Python environment. Object recognition consists of recognizing regions of interest of objects from images of a certain class. Different methods include frame subtraction, optical flow, and background subtraction. It is a method that uses cameras to detect and locate moving objects. Detection and tracking algorithms are described by extracting image and video features for security applications [2].

YOLO Algorithm: When it comes to deep learning based object detection, there are three main object detectors used. R-CNN and its variants,

Includes Original R-CNN, Fast R-CNN, and Faster R-CNN, Single Shot Detector (SSD), and YOLO.

R-CNN is one of the first object detectors based on deep learning and is an example of two-stage detectors. The problem with the standard R-CNN method was that it was very slow and not a perfect end-to-end object detector. Although R-CNN is very accurate, the biggest problem with R-CNN family networks is their speed. They were incredibly slow, only getting 5fps on the GPU. To improve the speed of deep learning-based object detectors, both the single-shot detector (SSD) and YOLO use a one-step detector strategy. These algorithms consider object detection as a regression problem, taking an input image and simultaneously learning the bounding box coordinates and corresponding class label probabilities. In general, one-stage detectors are less accurate than two-stage detectors, but they are significantly faster [3].

Euro, me

Test surface and testing that is usually required before touching the base in a satisfactory arrangement. This trademark indicates that the ability to plan approaches and quickly model desired arrangements generally plays a significant role in reducing the cost and time required to implement a suitable framework.

II Literature Review The literature was reviewed from various sources, including research articles, publications, available bibliographic information, and project committee recommendations. These research papers provided sufficient data for the study. A hierarchical approach is adopted in the organization. Teachers, staff and students have different privileges. Therefore, in this system, we adopted a role-based access control method, which is a rank-based access control method. Due to the large number of users in academic institutions, it is necessary to grant specific privileges to each user based on their position to prevent misuse of sensitive information. Role-based access control makes it easier for the system to distinguish between users, making the system faster without lag. Therefore, certain activities are restricted to certain users

Violations of fair dealing rules are maintained in the system.

This study by Wei Liu and Alexander [4]

C. Berg How to identify objects in images using a deep neural network. Our approach, called SSD, discretizes the bounded box output space into a set of default boxes that have different aspect ratios and scales for each feature map location. During prediction, the network generates a score for the presence of each object class in each default box and adjusts the box to better match the shape of the object.

This work by Andrew G. Howard [5] introduces a class of efficient models called MobileNet for mobile and embedded vision applications. MobileNet is based on a simple architecture that uses depth separable convolutions to build lightweight deep neural networks. We introduce two simple global metaparameters that are efficient.

In the case of single cells, YOLOv2 will not be able to detect them and will eventually lose object detection.

In this study by Sheheen Noor and Maria Waqas [7], this model provides an efficient end-to-end object detection and tracking technique that can be used in applications such as self-driving cars. SiamMask requires semi-supervision as it requires manual drawing of a bounding box around the object to be tracked. We overcome this limitation by using the most advanced object detection algorithms.

In this work by Mohammad Lisan and H.V.Ravish Aradhya [8], objects are tracked based on color, and the movements of single and multiple objects (vehicles) are detected and counted in several frames. In addition, a single algorithm can be designed to track objects by considering shape, color, texture, target object, and object movement in different directions.

THE PROPOSED METHOD

YOLO algorithm for detection and tracking is implemented in Python environment. Object recognition consists of recognizing regions of interest of objects from images of a certain class. Different methods include frame subtraction, optical flow, and background subtraction. It is a method that uses cameras to detect and locate moving objects.

Detection and tracking algorithms are demonstrated by extracting image and video features for security applications. Features are extracted using CNN and deep learning. Classifiers are used to classify and count images. YOLO-based algorithm with GMM model provides excellent accuracy for feature extraction and classification using deep learning concepts.

Just watch it once

YOLO proposes the use of an end-to-end neural network that predicts bounding boxes and class probabilities at once. YOLO significantly outperformed other real-time object detection algorithms and achieved advanced results.

While algorithms such as Faster RCNN work by using a region recommendation network to identify potential regions of interest and perform detection on those regions individually, YOLO completely performs all predictions using a single connected layer.

The methods that use the proposed area networks are repeated several times

Those boxes. These confidence scores indicate how confident the model is that the box contains the object and how accurate it thinks the predicted box is.

YOLO predicts multiple bounding boxes for each grid cell. During training, we want each object to be responsible for only one bounding box projection. YOLO designates a forecaster as "responsible" for forecasting the object based on which forecast has the highest flow IOU with the ground truth. This provides specialization between bounding box projections. Each predictor improves the prediction of a specific size, aspect ratio, or class of an object and improves the overall recall score.

One of the key techniques used in the YOLO model is Non-Maximum Suppression (NMS). NMS is a post-processing step used to improve the accuracy and efficiency of object detection. Object detection typically creates multiple bounding boxes for a single object in an image. These bounding boxes may overlap or be in different positions, but they all represent the same object. NMS is used to identify and remove redundant or incorrect bounding boxes and output a bounding box for each object in the image. YOLOv3 is the third version of the YOLO object recognition algorithm and aims to:

It improves the accuracy and speed of the algorithm.

One of the key improvements in YOLO v3 is the use of a new CNN architecture called Darknet-53. Darknet-53 is a variant of the ResNet architecture specifically designed for object detection tasks. It has 53 convolution layers and can achieve advanced results in various object detection criteria. In addition to these improvements, YOLO v3 can now handle a wider range of object sizes and aspect ratios. It is also more accurate and stable than previous versions of YOLO.

SYSTEM ARCHITECTURE

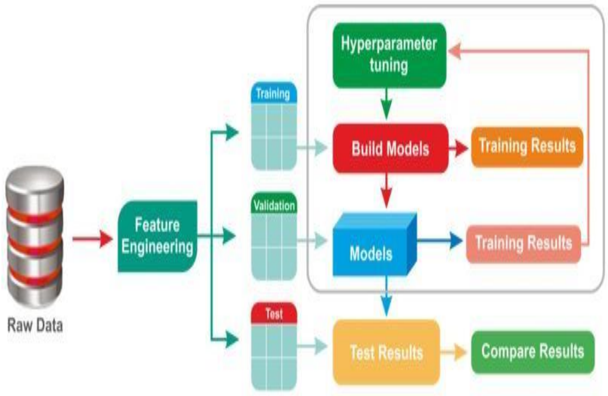

Fig.1 System architecture

IMPLEMENTATION

The followed method is for real object tracking in videos which consists of Objectdetection and tracking using YOLO . The whole implementation is done in python.

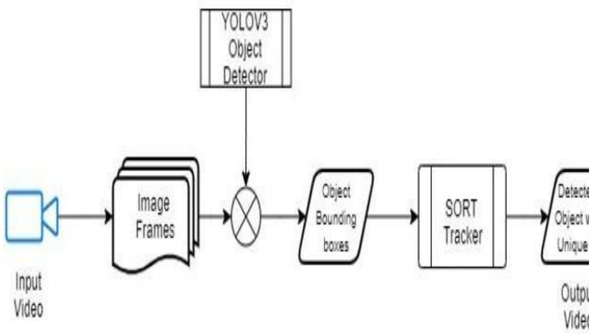

Fig.2 Flowchart representation

COLLECT RAW DATA

A custom dataset of 800 images with 6 classes: people, cars, trucks, buses, bicycles, and motorcycles is used to train YOLOv3, which is pre-trained on the MS COCO dataset with 80 classes. This model was trained over 320 cycles using Google Colab. All 800 images were manually annotated using the Label Img tool. The dataset was trained using the PyTorch library.

Images were tagged with YOLO format. A total of 200 images were used for validation. All images will be marked with .txt after annotation in YOLO format. You can also tag images with the Pascal version of VOC.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

Education

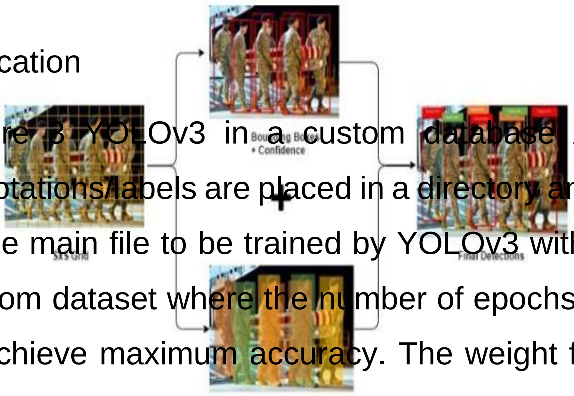

Figure 3 YOLOv3 in a custom database After the labeling is complete, the images and annotations/labels are placed in a directory and all this information is sent as parameters or code in the main file to be trained by YOLOv3 with the help of the PyTorch library. . will be done. A custom dataset where the number of epochs is determined by the size of the dataset and tries to achieve maximum accuracy. The weight file is the final output after training and is used to recognize objects in the model. The incoming video passes through the system and first the total number of frames is extracted and transmitted.

Object detector, in this case YOLO. YOLO, an object detector, generated bounding boxes with a class ID and confidence level for each bounding box.

test

In several videos, the proposed system is tested. The test is divided into two parts: object detection and tracking. The design of the project is based on Python and has been evaluated on five different video sequences running at powerful FPS.

After training, the weight file is used to recognize objects in videos. The input video file is divided into the total number of frames, each image is sent to the trained object detector, and the bounding box is displayed after the detection is completed.

information is passed onto algorithm and object tracking performed.

Fig.4 Labelling images using Labelling tool

RESULTS

We tested our object detector for few images to check how well it was trained and the following precision and recall graph was obtained.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

Table.1 Quantitative Analysis of the Proposed System

Video#	Total Frames	Accuracy	Precision	Recall
1	812	0.794	0.843	0.934
2	930	0.851	0.958	0.93
3	1160	0.75	0.9	0.9
4	835	0.781	0.833	0.892
5	590	0.444	0.5	0.889

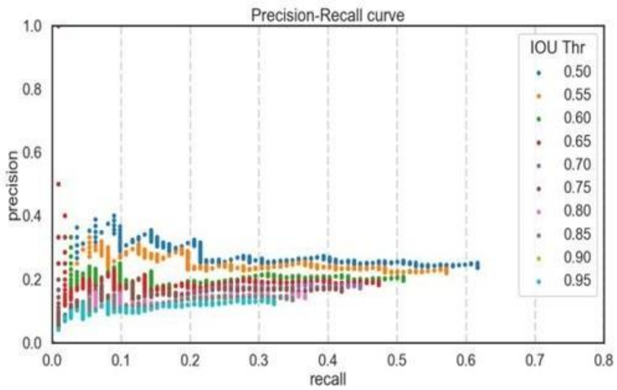

Fig.5 Precision Recall graph for custom dataset.

PERFORMANCE ANALYSIS

The quantitative analysis is performed these parameters True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

TP: Where the model correctly predicts a positive object class

FP: Where the model incorrectly predicts a positive object class

FN: Where model incorrectly predicts a negative object class

TN: Where the model correctly predicts a negative object class

Here, TP = a, TN = b, FP = c, FN = d.

$$ccuacy = \frac{a+}{oa}$$

$$cso = a$$

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

$$a + c$$
$$a$$
$$ca =$$
$$a +$$

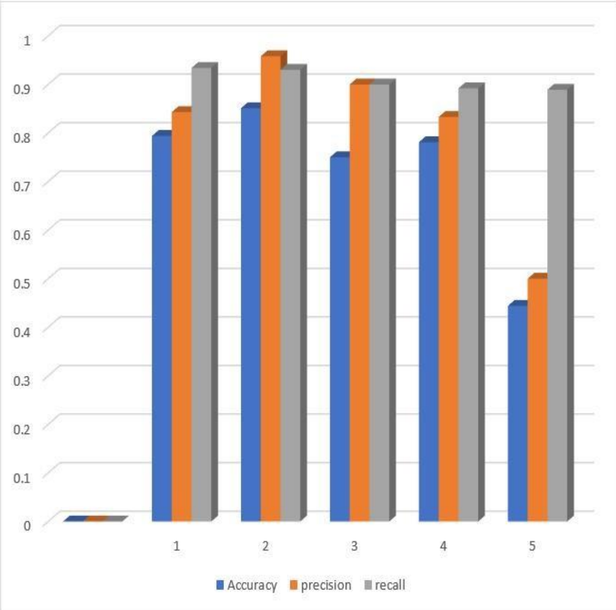

Fig.6 Final performance result of proposedwork

While performing visual object detection and tracking task, video is broken down into frames and each frame as well as a video output is saved with detection and tracking information obtained for each input video after using YOLO for object detection and tracking respectively.

Below are output screens of videos tested,which provide output as bounding boxes with class name and confidence scores.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

Fig 7 multiple objects are detected as persons

Fig.8 Bottle and glass objects are detected

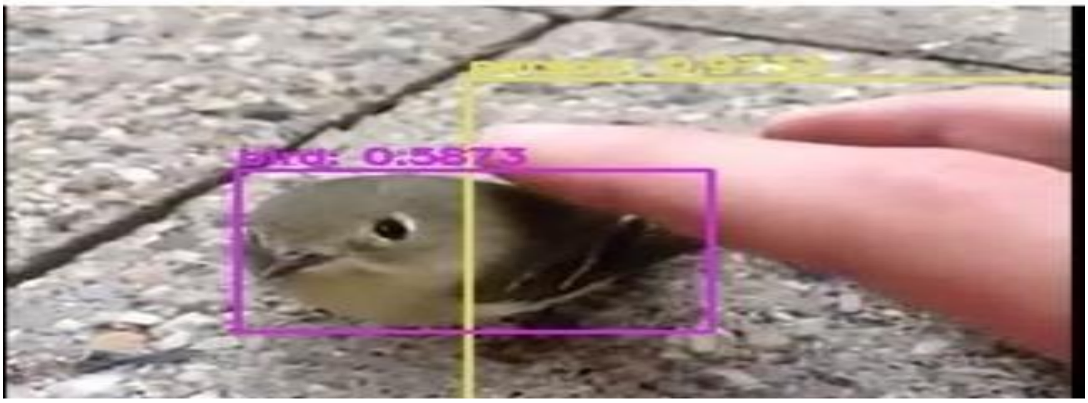

Fig 9 Bird object detected

Different objects are identified in Figures 7, 8 and 9.

From Figure 7, multiple objects are identified as individuals with confidence scores.

From Figure 8, objects as bottles and glasses and from Figure 9, objects as birds and people are recognized by their confidence scores.

CONCLUSION

Objects are detected using the YOLO algorithm in a real-time scenario. In addition, YOLO has shown results with considerable confidence that detecting different objects in real-time video sequences and tracking them in real time is the main goal of the YOLO algorithm. This model showed excellent detection and tracking results for the trained objects. The mobile networks model is more suitable for portable and embedded vision-based applications where there is no process control. The main goal of MobileNet is to optimize the delay.

Build a small neural network at the same time. In addition, it can be used to detect, track and respond to specific objects of interest in video surveillance in certain scenarios.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

REFERENCES

[1] Wei Liu and Alexander C. Berg, "SSD: Single-Shot Multi-Box Detector", Google Inc., December 2016.

[2] Andrew J. Howard, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Google Inc., 17 April 2017.

[3] Justin Lai, Sidney Maples, "Ammunition Detection: Development of a Real-Time Weapon Detection Classifier,"

Stanford University, February 2017

[4] Shriamesh Kamat, "UAV: Application of Object Detection and Tracking Technology to Unmanned Aerial Vehicles", Texas A&M University, 2015.

[5] Adrian Rosebrook, "Object Recognition with Deep Learning and OpenCV", pyimagesearch.

[6] Mohana and H.V.R., "Efficient and elegant algorithms for real-time object detection, counting and classification for video surveillance applications from a fixed camera", 2016 International Journal of Circuits, Control, Communication and Conference (I4C). , Bangalore, 2015, pp. 1-7.

[7] Akshay Mangawati, Mohana, Mohammad Lisan and H. V.

[8] Apoorva Raghunandan, Mohana, PA.