ZKG INTERNATIONAL

ISSN: 2366-1313

Real Time Object Detection and Tracking using Convolutional Neural Network

[1]R. Muruganantham, [2]K. Siri, [3]Harshita Jha, [4]S.Sai Venkata Anurag

[1]Professor, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,
muruganantham1979@gmail.com
[2,3,4]BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,
kokkusiri@gmail.com, harshitajhaa@gmail.com, anu.appe@gmail.com

Abstract: *Object detection is a computer vision technique for locating instances of objects in images or videos. Existing systems of object detection are Region-Based Convolutional Neural Network(R-CNN), Single Shot Detector (SSD). The biggest problem with the R-CNN family of networks is their speed they detect objects slowly. SSD unable to detect smaller objects. We proposed Real time object detection using Convolutional Neural Network (CNN) algorithm of You Only Look Once (YOLO)v3. This improves speed, real-time, accurate, precise identifications.*

Keywords: *Object detection, Region-Based Convolutional Neural Network, Convolutional neural network.*

I. INTRODUCTION

In the recent years, there has been an exponential progress in the field of machine learning and artificial intelligence which has led to improvement in accuracy, reduction in human efforts and failure rate. This development has played a commendable role in reducing processing time, which has further led to improvement in net productivity and corresponding reduction in the cost. To explore the application domain of machine learning systems, assume a situation of tracing our lost mobile in an untidy and messy house. It appears to be a cumbersome and frustrating task for anyone. It needs only a few milliseconds to track the Location of mobile. Well, this is precisely the power we can harness from these amazing object detection algorithms, which are at the bottom of heart the deep learning algorithms. The current research work focuses on proposing an object detection model that can take input from the web camera, find location of the object through webcam, and classify object on screen for its appropriate category. Eventually, the goal of the current work on object detection is to take raw images as inputs, find location of that object in the

ISSN: 2366-1313

given picture accurately and mask or classify object with appropriate categories [1].

Detecting and tracking objects are among the most prevalent and challenging tasks that a surveillance system has to accomplish in order to determine meaningful events and suspicious activities, and automatically annotate and retrieve video content. Under the business intelligence notion, an object can be a face, a head, a human, a queue of people, a crowd as well as a product.

In artificial vision, the neural convolution networks are distinguished in the classification of images. In this paper, an SSD and Mobile Nets based algorithms are implemented for detection and tracking in python environment. Object detection involves detecting region of interest of object from given class of image. Different methods are Frame differencing, Optical flow, Background subtraction. This is a method of detecting and locating an object which is in motion with the help of a camera. Detection and tracking algorithms are described by extracting the features of image and video for security applications [2].

YOLO Algorithm: When it comes to deep learning-based object detection, there are three primary object detectors you'll encounter. R-CNN and their variants, including the original R-CNN, Fast R-CNN, and Faster R-CNN, Single Shot Detector (SSDs), YOLO.

R-CNNs are one of the first deep learning-based object detectors and are an example of a two-stage detector. The problem with the standard R-CNN method was that it was painfully slow and not a complete end-to-end object detector, While R-CNNs tends to very accurate, the biggest problem with the R-CNN family of networks is their speed they were incredibly slow, obtaining only 5 FPS on a GPU. To help increase the speed of deep learning-based object detectors, both Single Shot Detectors (SSDs) and YOLO use a one-stage detector strategy. These algorithms treat object detection as a regression problem, taking a given input image and simultaneously learning bounding box coordinates and corresponding class label probabilities. In general, single-stage detectors tend to be less accurate than two-stage detectors but are significantly faster [3].

YOLO is a great example of a single stage detector. We'll be using YOLOv3 in this project, in particular, YOLO trained on the COCO dataset. We can use YOLO to perform object detection in input video files as well. Apply YOLO object detection to single images, video streams.

ISSN: 2366-1313

OBJECTIVE OF THE DETECTION

The aim of real time object detection and tracking Due to object detection's close relationship with video analysis and image understanding, it has attracted much research attention in recent years. Traditional object detection methods are built on handcrafted features and shallow trainable architectures.

Their performance easily stagnates by constructing complex ensembles which combine multiple low-level image features with high-level context from object detectors and scene classifiers. With the rapid development in deep learning, more powerful tools, which are able to learn semantic, high-level, deeper features, are introduced to address the problems existing in traditional architectures. These models behave differently in network architecture, training strategy and optimization function, etc. In this project, we provide a review on deep learning-based object detection frameworks.

Our review begins with a brief introduction on the history of deep learning and its representative tool, namely Convolutional Neural Network (CNN). Additionally, SSD have shown results with considerable confidence level, main Objective of SSD algorithm to detect various objects in real time video sequence and track them in real time. A set of default boxes is made to pass over several feature maps in a convolutional manner. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance. If an object detected is one among the object classifiers during prediction, then a score is generated. The object shape is adjusted to match the localization box. For each box, shape offsets and confidence level are predicted.

OBJECT DETECTION

Object Detection is the process of finding and recognizing real-world object instances such as car, bike, TV, flowers, and humans out of an images or videos. An object detection technique lets you understand the details of an image or a video as it allows for the recognition, localization, and detection of multiple objects within an image. It is usually utilized in applications like image retrieval, security, surveillance, and advanced driver assistance systems (ADAS).

DIGITAL IMAGE PROCESSING

Computerized picture preparing is a range portrayed by the requirement for broad test work to build up the practicality of proposed answers for a given issue. A critical trademark hidden the plan of picture preparing frameworks is the huge

ISSN: 2366-1313

level of testing and experimentation that Typically is required before touching base at a satisfactory arrangement. This trademark informs that the capacity to plan approaches and rapidly model hopeful arrangements by and large assumes a noteworthy part in diminishing the cost and time required to land at a suitable framework execution

II. LITERATURE SURVEY

Literature was reviewed from various sources, like from research papers, publications books, existing bibliographic information, and recommendations by the project panel. These research papers have provided us sufficient amount of data for the survey. The hierarchical approach is followed in the institutional organizations. Teachers, staffs and students have different privileges. So, for this system we have used access control method which suits the ranking that is the role-based access control method. Since there are large number of users present in an academic institution it is a prime requisite to grant certain privileges to each user according to their positions so that the sensitive information is not misused. The role-based access control makes it easy for the system to differentiate between its users which makes the system faster without any lagging. There are certain activities restricted to specific users so to avoid the violation of code of conduct fairness is maintained in the system.

In this study[4] by Wei Liu and Alexander C. Berg A method for detecting objects in images using a single deep neural network. Our approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape.

In this study [5 by Andrew G. Howard present a class of efficient models called MobileNets for mobile and embedded vision applications. MobileNets are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks. We introduce two simple global hyper-parameters that efficiently trade off between latency and accuracy. These hyper-parameters allow the model builder to choose the right sized model for their application based on the constraints of the problem.

In this study[6] by Adrian Rosebrock the YOLOv2 object detector divides an input image into an SxS grid where each cell in the grid predicts only a single object.If there exist multiple, small objects in a

ISSN: 2366-1313

single cell then YOLOv2 will be unable to detect them, ultimately leading to missed object detections.

In this study[7] by Sheheen Noor, Maria Waqas,this model device an efficient technique for an end-to-end object detection and tracking , which can then be used application like self driving cars. SiamMask requires semi-supervision in that it needs a bounding box to be drawn manually around the object that has to be tracked.We overcome this limitation by using state of art object detection algorithm.

In this study [8] by Mohammed Leesan and H.V.Ravish Aradhya objects are tracked based on colour, motion of single and multiple objects (vehicles) are detected and counted in multiple frames. Further single algorithm may be designed for object tracking by considering shape, colour, texture, object of interest, motion of object in multi direction.

III. PROPOSED METHODOLOGY

YOLO algorithm is implemented for detection and tracking in python environment. Object detection involves detecting region of interest of object from given class of image. Different methods are Frame differencing, Optical flow, Background subtraction. This is a method of detecting and locating an object which is in motion with the help of a camera. Detection and tracking algorithms are described by extracting the features of image and video for security applications. Features are extracted using CNN and deep learning. Classifiers are used for image classification and counting. YOLO based algorithm with GMM model by using the concepts of deep learning will give good accuracy for feature extraction and classification.

YOU ONLY LOOK ONCE

YOLO proposes using an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once. YOLO achieved state-of-the-art results, beating other real-time object detection algorithms by a large margin.

While algorithms like Faster RCNN work by detecting possible regions of interest using the Region Proposal Network and then performing recognition on those regions separately, YOLO performs all of its predictions with the help of a single fully connected layer.

Methods that use Region Proposal Networks perform multiple iterations for the same image, while YOLO gets away with a single iteration. YOLO divides an input image into an S × S grid. If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object. Each grid cell predicts B bounding boxes and confidence scores for

ISSN: 2366-1313

those boxes. These confidence scores reflect how confident the model is that the box contains an object and how accurate it thinks the predicted box is.

YOLO predicts multiple bounding boxes per grid cell. At training time, we only want one bounding box predictor to be responsible for each object. YOLO assigns one predictor to be "responsible" for predicting an object based on which prediction has the highest current IOU with the ground truth. This leads to specialization between the bounding box predictors. Each predictor gets better at forecasting certain sizes, aspect ratios, or classes of objects, improving the overall recall score.

One key technique used in the YOLO models is non-maximum suppression (NMS). NMS is a post-processing step that is used to improve the accuracy and efficiency of object detection. In object detection, it is common for multiple bounding boxes to be generated for a single object in an image. These bounding boxes may overlap or be located at different positions, but they all represent the same object. NMS is used to identify and remove redundant or incorrect bounding boxes and to output a single bounding box for each object in the image.

YOLOv3 is the third version of the YOLO object detection algorithm, aiming to increase the accuracy and speed of the algorithm.

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

SYSTEM ARCHITECTURE

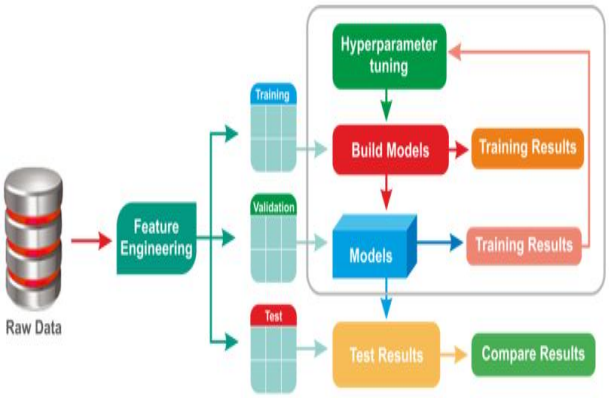

Fig.1 System architecture

IV. IMPLEMENTATION

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

ISSN: 2366-1313

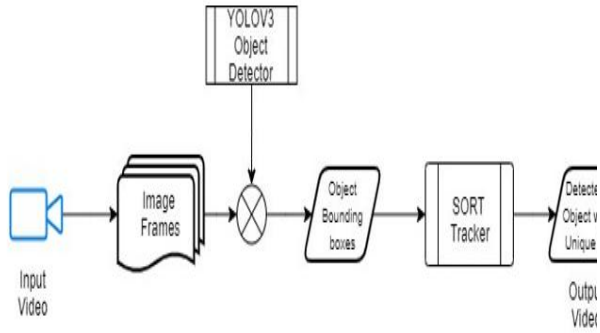

Fig.2 Flowchart representation

Raw data gathering

Custom dataset consisting 800 images having 6 classes: Person, Car, Truck, Bus, Bicycle and Motorbike was used for training YOLOv3 which was already pre-trained for MS COCO dataset consisting of 80 classes. Model was trained for 320 epochs using Google Colab. All the 800 Images were annotated manually using Label Img tool. Dataset was trained with help of PyTorch library.

Images were labelled in the YOLO format. Total of 200 images were used for validation. All the images have a specified .txt associated to them after annotation were done in the format of YOLO. Images can be labelled in Pascal VOC version as well.

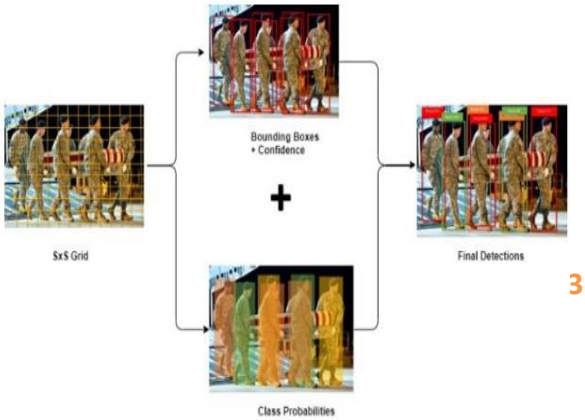

Fig.3 YOLOv3 on custom database

Training

Once labelling is completed, images and annotations/ labels are placed in a directory and all this information is either passed as parameters or code inside the main file and with the help of PyTorch library, YOLOv3 is trained for our custom dataset with the number of epochs decided depending on the size of dataset and trying to achieve maximum accuracy. A weights file is the final output after training which will be used for object detection in our model. An Input video is passed through the system and then at first total number of frames are extracted and forwarded to object detector which is YOLO in this case. Being an object detector YOLO generated bounding boxes with class ID and confidence for each bounding box.

Testing

On several videos, the proposed system is tested. The experiment is divided into two sections, the identification and tracking of objects. The project design is python-based and evaluated on five different video sequences and runs with strong FPS.

Once training is completed the weights files in used for object detection in videos. Input video file is broken down into total number of frames and passes each image to our trained object detector and once detection is done bounding box

ISSN: 2366-1313

information is passed onto algorithm and object tracking performed.

Fig.4 Labelling images using Labelling tool

V. 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.

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

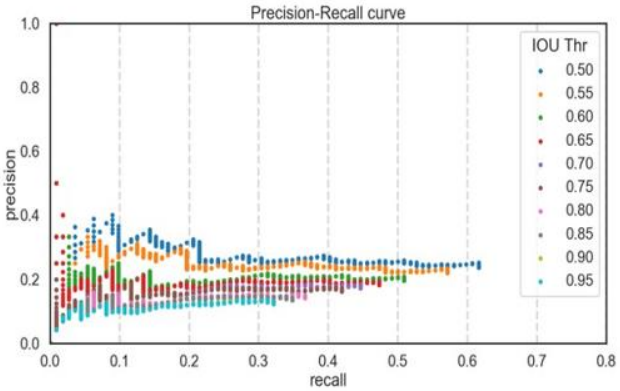

Fig.5 Precision Recall graph for custom dataset.

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.

$$Accuracy = \frac{a+b}{total}$$

$$Precision = \frac{a}{a+c}$$

$$Recall = \frac{a}{a+b}$$

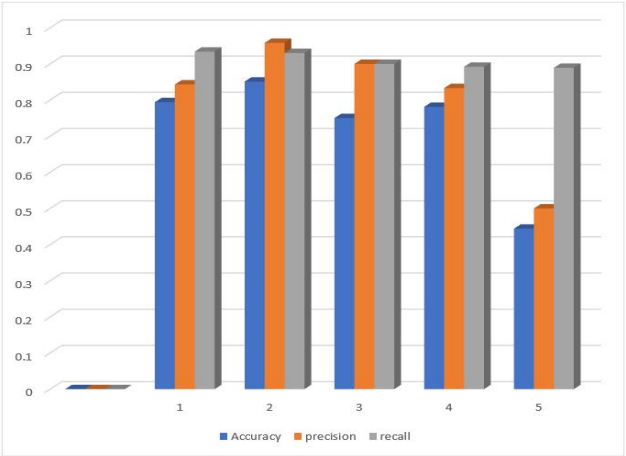

Fig.6 Final performance result of proposed work

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

ISSN: 2366-1313

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.

.

Fig 7 multiple objects are detected as persons

Fig.8 Bottle and glass objects are detected

ISSN: 2366-1313

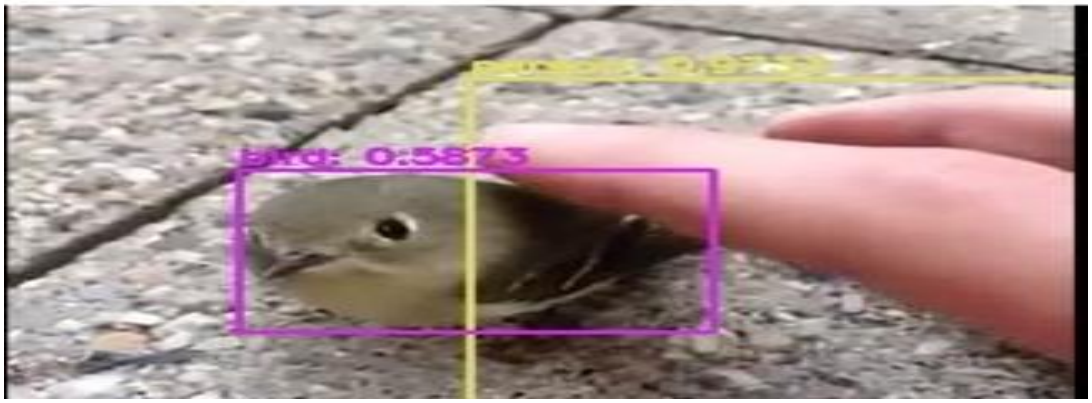

Fig 9 Bird object detected

In figure 7,8,9 different objects are detected.

From fig 7 multiple objects are detected as a person with confidence score.

From fig 8 Objects are detected as bottle and glass and from fig 9 objects are detected as bird and person with confidence score.

VI. CONCLUSION

Objects are detected using YOLO algorithm in real time scenarios. Additionally, YOLO have shown results with considerable confidence level, main Objective of YOLO algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained. The Mobile Nets model is more appropriate for portable and embedded vision-based applications where there is absence of process control. The main objective of Mobile Nets is to optimize the latency while building small neural nets at the same time. It can be Further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance.

REFERENCES

[1] Wei Liu and Alexander C. Berg, “SSD: Single Shot Multibox Detector”, Google Inc., Dec 2016.

[2] Andrew G. Howard, and Hartwig Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, Google Inc., 17 Apr 2017.

[3] Justin Lai, Sydney Maples, “Ammunition Detection: Developing a Real Time Gun Detection Classifier”, Stanford University, Feb 2017

[4] Shreyamsh Kamate, “UAV: Application of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles”, Texas A&M University, 2015.

ISSN: 2366-1313

[5] Adrian Rosebrock, “Object detection with deep learning and OpenCV”, pyimagesearch.

[6] Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, 2016, pp. 1-7.

[7] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, “Object Tracking Algorithms for video surveillance applications” International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680.

[8] Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya, “Object Detection Algorithms for video surveillance applications” International conference on communication and signal processing (ICCSP), India, 2018, pp. 0570-0575.

[9] Manjunath Jogin, Mohana, “Feature extraction using Convolution Neural Networks (CNN) and Deep Learning” 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.

[10] Sheheen Noor, Maria Waqa,“Automatic object tracking and segmentation using unsupervised siamask”2018 IEEE International Conference Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.