Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018)
IEEE Xplore Compliant Part Number:CFP18N67-ART; ISBN:978-1-5386-2456-2

Real Time Object Detection and Tracking Using Deep Learning and OpenCV

Chandan G, Ayush Jain, Harsh Jain, Mohana
Telecommunication Engineering
R. V. College of Engineering
Bengaluru, India

Abstract — **Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since past few years. Some of the popular object detection algorithms are Region-based Convolutional Neural Networks (RCNN), Faster-RCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy. 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.**

Keywords — Mobile Nets, Single Shot Detector, COCO.

I. INTRODUCTION

Since AlexNet has stormed the research world in 2012 ImageNet on a large scale visual recognition challenge, for detection in-depth learning, far exceeding the most traditional methods of artificial vision used in literature. In artificial vision, the neural convolution networks are distinguished in the classification of images.

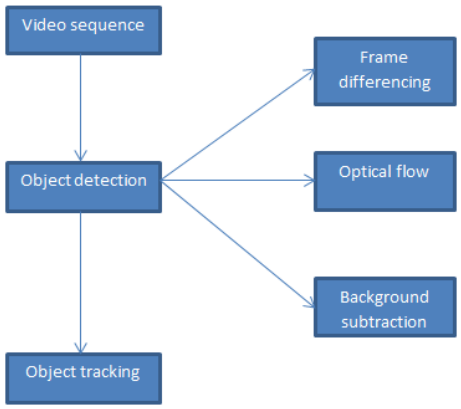

Fig. 1. Basic block diagram of detection and Tracking

Fig. 1 shows the basic block diagram of detection and tracking. In this paper, an SSD and MobileNets 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 [3] [7] [8]. Features are extracted using CNN and deep learning [9]. Classifiers are used for image classification and counting [6]. YOLO based algorithm with GMM model by using the concepts of deep learning will give good accuracy for feature extraction and classification [10]. Section II describes SSD and MobileNets algorithm, section III explains method of implementation, and section IV describes simulation results and analysis.

II. OBJECT DETECTION AND TRACKING ALGORITHMS

A. *Single Shot Detector (SSD) algorithm*

SSD is a popular object detection algorithm that was developed in Google Inc. [1]. It is based on the VGG-16 architecture. Hence SSD is simple and easier to implement.

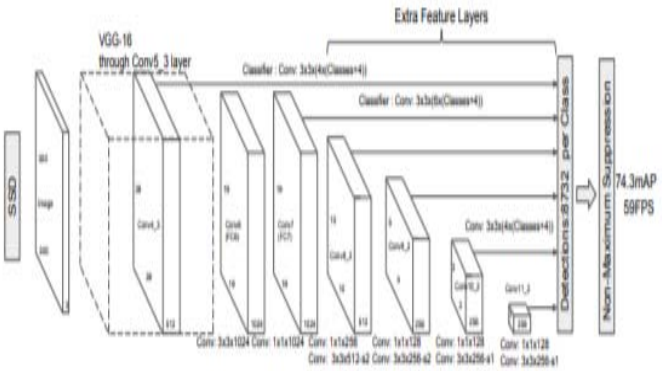

Fig. 2. VGG-16 SSD Model.

Fig. 2 shows VGG 16 SSD model. A set of default boxes is made to pass over several feature maps in a convolutional manner. 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. During training, default boxes are matched to the ground truth boxes. The fully connected layers are discarded by SSD architecture. The model loss is computed as a weighted sum of confidence loss and localization loss. Measure of the deviation of the predicted box from the ground truth box is localization loss.

Confidence is a measure of in which manner confidence the system is that a predicted object is the actual object.

Elimination of feature resampling and encapsulation of all computation in a single network by SSD makes it simple to train with MobileNets. Compared to YOLO, SSD is faster and a method it performs explicit region proposals and pooling (including Faster R-CNN).

B. MobileNets algorithm

MobileNets uses depthwise separable convolutions that helps in building deep neural networks. The MobileNets model is more appropriate for portable and embedded vision based applications where there is absence of process control. The main objective of MobileNets is to optimize the latency while building small neural nets at the same time. It concentrates just on size without much focus on speed. MobileNets are constructed from depthwise separable convolutions. In the normal convolution, the input feature map is fragmented into multiple feature maps after the convolution [2].

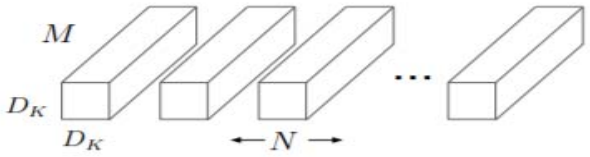

Fig. 3. Normal Convolution [2]

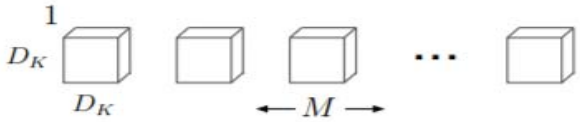

Fig. 4. Depthwise Convolution Filters [2]

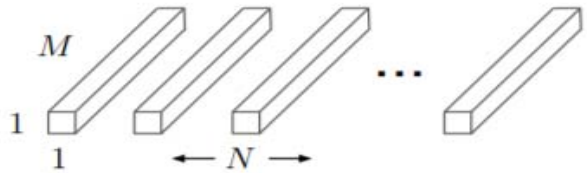

Fig. 5. 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution [2].

The number of parameters is reduced significantly by this model through the use of depthwise separable convolutions, when compared to that done by the network with normal convolutions having the same depth in the networks. The reduction of parameters results in the formation of light weight neural network as shown in fig 3 to 5.

III. METHODS OF IMPLEMENTATION

A. Object detection

Frame differencing

Frames are captured from camera at regular intervals of time. Difference is estimated from the consecutive frames.

Optical Flow

This technique estimates and calculates the optical flow field with algorithm used for optical flow. A local mean algorithm is used then to enhance it. To filter noise a self-adaptive algorithm takes place. It contains a wide adaptation to the number and size of the objects and helpful in avoiding time consuming and complicated preprocessing methods.

Background Subtraction

Background subtraction (BS) method is a rapid method of localizing objects in motion from a video captured by a stationary camera. This forms the primary step of a multi-stage vision system. This type of process separates out background from the foreground object in sequence in images.

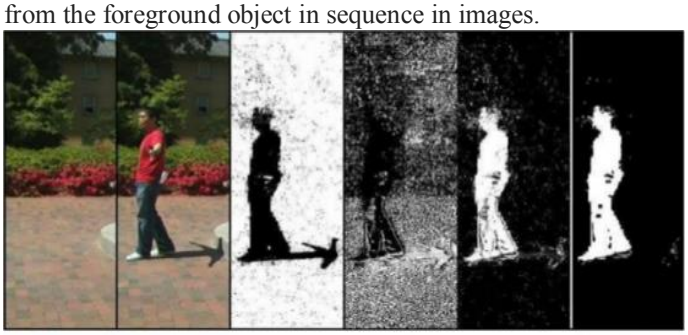

Fig.6. Detection of human from background subtraction

Fig. 6 depicts Detection of human from background subtraction. Foreground or person is detected and separated from the background of the image for further preprocessing. The separation effect is shown step wise, after which localization of region of interest takes place.

B. Object tracking

It is done in video sequences like security cameras and CCTV surveillance feed; the objective is to track the path followed, speed of an object. The rate of real time detection can be increased by employing object tracking and running classification in few frames captured in a fixed interval of time. Object detection can run on a slow frame rates looking for objects to lock onto and once those objects are detected and locked, then object tracking, can run in faster frame speed.

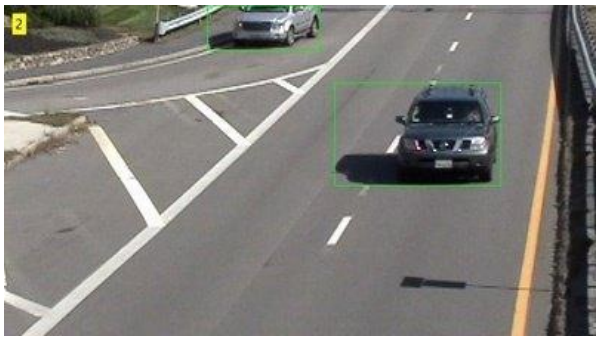

Fig.7. Tracking of car

Fig. 7 shows the tracking of car. Two ways in which the object can be tracked in the above example are: (1)-Tracking in a sequence of detection. In this method a CCTV video sequence of a traffic which is in motion takes place. Suppose someone wants to track a car or person's movement here, he will take different images or frames at different interval of time. With the help of these images one can target the object like a car or person. Then, by checking how my object has moved in different frames of the video, I can track it. Velocity of the

object can be calculated by verifying the object's displacement with the help of different frames taken at different interval of time. This method is actually a flaw where one is not tracking but detecting the object at different intervals of time. Improved method is "detection with dynamics". In this method estimation of car's trajectory or movement takes place. By checking it's position at a particular time 't' and estimating its position at another time interval let's say 't+10'.From this actual image of car at 't+10' time can be proposed with the help of estimation.

IV. SIMULATION RESULTS AND ANALYSIS

Based on SSD algorithm, a python program was developed for the algorithm and implemented in OpenCV [5]. OpenCV is run in Ubuntu IDE. Total 21 objects were trained in this model. The following results are obtained after successful scanning, detection and tracking of video sequence provided by camera.

Fig.8. Detection of Bicycle with confidence level of 99.49%

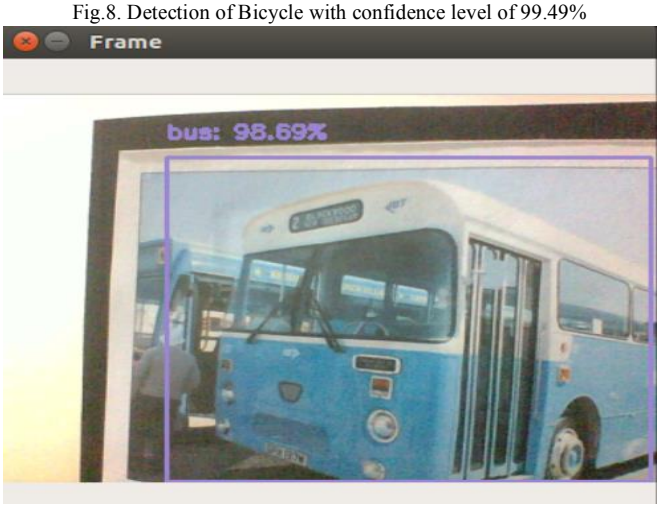

Fig.9.Detection of Bus with confidence level of 98.68%

Fig. 10. Detection of Train with confidence level of 99.99%

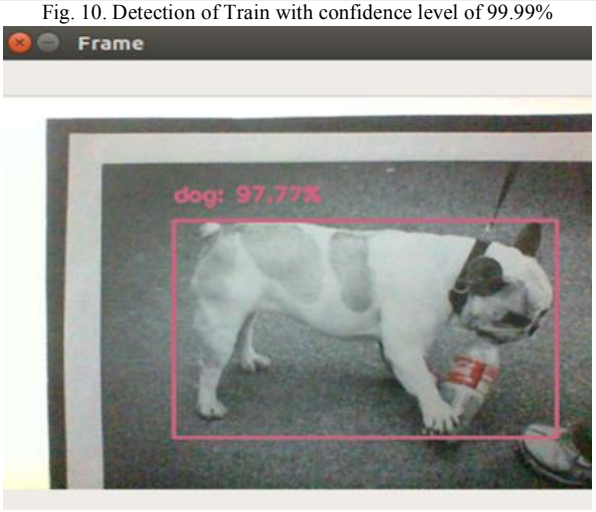

Fig.11. Detection of Dog with confidence level of 97.77%

Fig. 8 to 11 shows the real time detection of bicycle, bus, train, and dog with confidence levels 99.49%, 98.68%, 99.99% and 97.77 % respectively. The model was trained to detect 21 objects class like dog, motorbike, person, potted plant, bird, car, cat, sofa, sheep, bottle, chair, aero plane, train, bicycle etc. with accurately of 99%.

V. CONCLUSION

Objects are detected using SSD algorithm in real time scenarios. 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. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance. This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise. Further extending the work to detect ammunition and guns in order to trigger alarm in case of terrorist attacks. The model can be deployed in CCTVs, drones and other surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.

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.

[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] Arka Prava Jana, Abhiraj Biswas, Mohana, "YOLO based Detection and Classification of Objects in video records" *2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT)* 2018, India.