e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

REAL TIME OBJECT DETECTION WITH DEEP LEARNING

Tilak Singh Kuntal*1, Dr. Mrinal Paliwal*2

*1Research Scholar, Sanskriti University, Mathura, India.

*2Assistant Professor, Sanskriti University, Mathura, India.

ABSTRACT

Real time object detection is a vast, vibrant and sophisticated area of computer vision aimed towards object identification and recognition. Object detection detects the semantic objects of a class objects using OpenCV (Open source ComputerVision), which is a library of programming functions mainly trained towards real time computer vision in digital images and videos. Visually challenged people cannot distinguish the objects around them. The main aim behind this real time object detection is to helpthe blind to overcome their difficulty. Real time object detection finds its uses in the areas like tracking objects, video surveillance, pedestrian detection, people counting, self-driving cars, face detection, ball tracking in sports and many more. This is achieved using Convolution Neural Networks, which is a representative tool of Deep learning. This project acts as an aiding tool for visually challenged people.

Keywords: Convolutional Neural Network, OpenCV, Deep Learning.

I. INTRODUCTION

Object detection is a technology to detect various objects in digital images and videos too. It is mainly helpful within the self- driving cars, face detection, etc., where the objects are to be continuously monitored. The algorithm or the technique involved for object detection during this project is Convolutional Neural Networks which is a class of Deep learning. This uses MobileNet SSD technique during which MobileNet is a neural network used for image classification and recognition whereas SSD is a framework that is used to realize the multibox detector. The mixture of both MobileNet and SSD can do object detection. The main advantage or purpose of choosing Deep learning is that we do not need to do feature extraction from data as compared to machine learning.

The Haar-like trait play a crucial role in detecting the objects in a picture. They scan the entire picture starting from the top left and compares every small box with the trained data. In this way, even small-detailed objects present within the images are identified.

II. METHODOLOGY

Deep learning, a subset of machine learning which in turn is a subset of artificial intelligence (AI) has networks capable of learning things from the data that is unstructured or unlabeled. The approach utilized in this project is Convolutional Neural Networks (CNN). It uses the Haar-cascade classifiers which help us in the detection of objects.

1. CNN:

The convolutional neural network, or CNN for brief, could also be a specialized kind of neural network model designed for working with two-dimensional image data, although they're going to be used with one-dimensional and three-dimensional data.

Central the convolutional neural network is the convolutional layer that gives the network its name. This layer performs an operation known as “convolution”.

In the context of a convolutional neural network, a convolution may be a linear operation that involves the multiplication ofa group of weights with the input, very similar to a standard neural network. as long as the technique was designed for two- dimensional input, the multiplication is performed between an array of input file and a two-dimensional array of weights, called a filter or a kernel. The filter is smaller than the input file and therefore the before the sort of multiplication applied between a filter-sized patch of the input and the filter may be a scalar product. A scalar product is that the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed, always leading to one value. Because it leads to 1 value, the operation is conventionally represented and mentioned because the “scalar product”. Using a filter smaller than the input is intentional because it allows an equivalent filter (set of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

weights) to be multiplied by the input array multiple times at distinct points on the input. Specifically, the filter is applied systematically to every overlapping part or filter-sized patch of the input file, left to right, top to bottom.

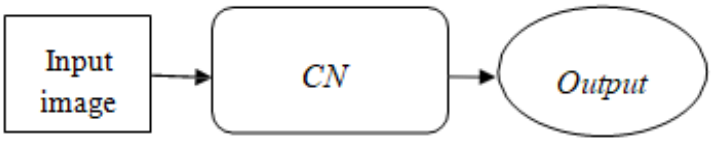

Fig: Sample block diagram indicating the flow of image processing using CNN

This systematic application of an equivalent filter across a picture may be a powerful idea. If the filter is meant to detect aselected sort of feature within the input, then the appliance of that filter systematically across the whole input image allows the filter a chance to get that feature anywhere within the image.

This capability is usually represented and mentioned as translation invariance, e.g. the total altogether concern in whetherthe feature is present instead of where it should had been present.

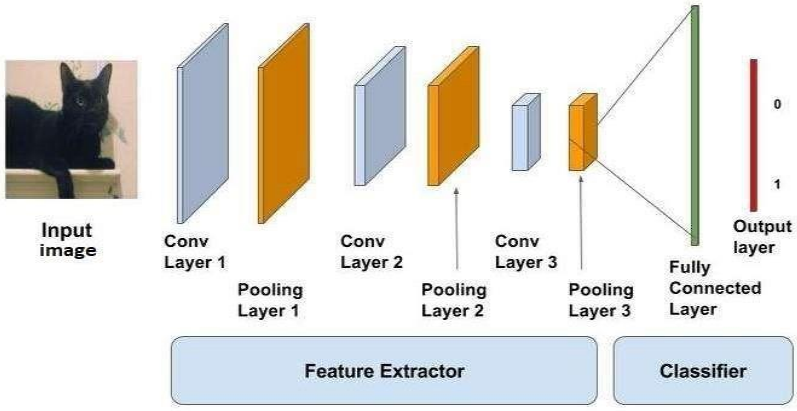

Fig: Image classification using CNN

2. OpenCV:

Open CV stands for open source computer vision. it's a group of libraries in Python. it's a tool by which we will be able to manipulate the pictures , like image scaling, etc. This supports and helps us in developing real time computing applications. It mainly concentrates and targets on image processing, video capture and analysis. It includes several features like face detection and also object detection. Currently OpenCV supports differing types of programming languages like C++, Python, Java etc., and it's available on various platforms including Windows, Linux, OS X, Android etc.

3. Training the data set:

The data set is typically the gathering of knowledge . the info set could also be collection of images or alphabets or numbers or documents and files too. the info set we used for the thing detection is that the collection of images of all the objects that are to be identified. Several different images of every and each object is typically present within the data set. Ifthere are more number of images like each object within the datasets then the accuracy are often improved. The important thing that's to be remembered is that the info within the data set must be labelled. there'll be actually 3 data set. they're the training data set, the validation dataset and therefore the other one is testing data set. The training data set will usually contains around 85-90% of the entire labelled data. This training dataset are going to be training our machine and therefore the model is obtained by training the info set. The validation data set consists of around 5-10% of the entire labelled data. this is often used for the validation purpose. the opposite data set is that the testing dataset and it's wont to test the performance of our machine.

4. Developing a real time object detector:

For developing a true time object detector using deep learning and open cv we'd like to access our web cam during a really effective way then the thing detection is to be applied to each and every frame. we should always install open cv in our systems.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

The deep neural network module should be installed.

Firstly, we should always always import all the specified packages:

1. From imutils.video we'll import VideoStream
2. From imutils.video we'll import FPS
3. we'll import numpy as np
4. we'll import argparse
5. we'll import imutils
6. we'll import time
7. we'll import cv2

The next step is to construct the argument parse then we should always parse the arguments.

--prototxt: provide path to the Caffe prototxt file.

--model: provide path to the pre-trained model.

--confidence: The minimum probability threshold to filter weak detections. The default value is given as 20%. The next step is to initialize CLASS labels and corresponding random COLORS.

Each object when it's detected, it's surrounded by a box with some predefined colour. Thus, we assign each object aspecific color.

After that we'll load our model and that we will provide the regard to our prototxt and also to our model files. With the assistance of imutils we'll read the video and that we will set the amount of frames per second. Now with this some predefined number of frames are going to be loaded per second. Each frame is analogous to the image. Now these images are going to be given because the inputs to the model.

The model will process the input image and produces the output image which consists of labels. in additional practical sense the input raw image is given to the model. Now the model process the input image. within the output image all the thing s are identified and every object is surrounded by an oblong box and therefore the name of the object is additionally displayed. we'll be only observing the output video stream but not the input video stream.

III. RESULT

Here, in this project we've considered around 15 to 20 objects to be detected during the training. Some of those include 'person', 'car', 'train', 'bird', 'sofa', 'dog', ''plant', 'aero plane', 'bicycle', 'bus', 'motorbike', etc.

The output of this project displays the objects detected with a rectangular box around the object with a label indicating it's name and therefore the exactness with which the object has been detected on the top of it. It can dig out any number of objects existing during a single image with certainty.

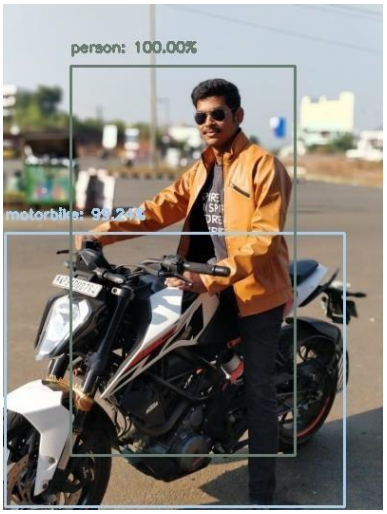

Fig 1

Fig 2

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:06/June-2022 Impact Factor- 6.752 www.irjmets.com

Fig 3

Fig 4

Fig 5

IV. APPLICATIONS

Here are a some of the future implementation of object detection.

1. Face detections and recognition:

Face detection perhaps be a separate class of object detection. We wonder how some applications like Facebook, Faceapp, etc., detect and recognize our faces. this is often a sample example of object detection in our day to day life. Face detection is already in use in our lifestyle to unlock our mobile phones and for other security systems to scale back rate .

2. Object tracking:

Object detection is additionally utilized in tracking objects like tracking an individual and his actions, continuously monitoring a ball within the game of Football or Cricket. As there's an enormous interest for people in these games, these tracking techniques enables them to know it during a better way and obtain some additional information. Tracking of the ball is of maximal importance in any ball-based games to automatically record the movement of the ball and adjust the video frame accordingly.

3. Self-driving cars:

This is often one among the main evolutions of the planet and is that the best example why we'd like

object detection. so as for a car to travel to the specified destination automatically with none human interference or to form decisions whether to accelerate or to use brakes and to spot the objects around it. this needs object detection.

4. Emotions detection:

This permits the system to spot the type of emotion the person puts on his face. the corporate Apple has already tried to use this by detecting the emotion of the user and converting it into a respective emoji within the smart phone.

5. Biometric identification through retina scan:

Retina scan through iris code is one among the techniques utilized in high security systems because it is one among the foremost accurate and unique biometric.

6. Smart text search and text selection (Google lens)

In recent times, we've encountered an application in smart phones called google lens. this will recognize the text and also images and search the relevant information within the browser without much effort.

V. CONCLUSION

Deep-learning based object detection has been a search hotspot in recent years. This project starts on generic object detection pipelines which give base architectures for other related tasks. With the assistance of this the 3 other common tasks, namely object detection, face detection and pedestrian detection, are often accomplished. Authors accomplished this by combing 2 things: Object detection with deep learning and OpenCV and Efficient, threaded video streams with OpenCV. The camera sensor noise and lightening condition can change the result because it can create problem in recognizing the objects. generally, this whole process requires GPU's rather than CPU's. But we've done using CPU's and executes in much less time, making it efficient. Object Detection algorithms act as a mixture of both image classification and object localization. It takes the given image as input and produces the output having the bounding boxes adequate to the amount of objects present within the image with the category label attached to every bounding box at the highest. It projects the scenario of the bounding box up the shape of position, height and width.

VI. REFERENCES

[1] Geethapriya S, N. Duraimurugan, S.P. Chokkalingam, "Real-Time Object Detection with Yolo", InternationalJournal of Engineering and Advanced Technology (IJEAT)

[2] Abdul Vahab, Maruti S Naik, Prasanna G Raikar an Prasad S R4, "Applications of Object Detection System", International Research Journal of Engineering and Technology (IRJET)

[3] Hammad Naeem, Jawad Ahmad and Muhammad Tayyab, "Real-Time Object Detection and Tracking", IEEE

[4] Meera M K, & Shajee Mohan B S. 2016, "Object recognition in images", International Conference on Information Science (ICIS).

[5] Astha Gautam, Anjana Kumari, Pankaj Singh: "The Concept of Object Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016,pp. 779-788

[7] V. Gajjar, A. Gurnani and Y. Khandhediya, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach," in 2017 IEEE International Conference on Computer Vision Workshops, 2017.