International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Real-Time Object Detection System using Caffe Model

Vaishali[1], Shilpi Singh[2]

[1]PG Scholar, CSE Department, Lingaya's Vidyapeeth, Faridabad, Haryana, India
[2]Assistant Professor, CSE Department, Lingaya's Vidyapeeth, Faridabad, Haryana, India

---

Abstract - *An object detection system recognises and searches the objects of the real world out of a digital image or a video, where the object can belong to any class or category, for example humans, cars, vehicles and so on. Authors have used OpenCV packages, Caffe model, Python and numpy in order to complete this task of detecting an object in an image or a video. This review paper discusses how deep learning technique is used to detect a live object, localise an object, categorise an object, extract features, appearance information, and many more, in images and videos using OpenCV and how caffee model is used to implement and also why authors have chosen caffe model over other frameworks. To build our deep learning-based real-time object detector with OpenCV we need to access webcam in an efficient manner and to apply object detection to each frame.*

KeyWords: Object Detection, OpenCV, Images, Videos, Caffe model.

1. INTRODUCTION

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 [16].

It is usually utilized in applications like image retrieval, security, surveillance, and advanced driver assistance systems (ADAS).Object Detection is done through many ways:

- Feature Based Object Detection
- Viola Jones Object Detection
- SVM Classifications with HOG Features
- Deep Learning Object Detection

Object detection from a video in video surveillance applications is the major task these days. Object detection technique is used to identify required objects in video sequences and to cluster pixels of these objects .
The detection of an object in video sequence plays a major role in several applications specifically as video surveillance applications.

Object detection in a video stream can be done by processes like pre-processing, segmentation, foreground and background extraction, feature extraction.

2. RELATED TECHNOLOGY

2.1 R-CNN

R-CNN is a progressive visual object detection system that combines bottom-up region proposals with rich options computed by a convolution neural network [10].

R-CNN uses region proposal ways to initial generate potential bounding boxes in a picture and then run a classifier on these proposed boxes.

2.2 Single Size Multi Box Detector

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 the time of prediction the network generates scores for the presence of each object category in each default box and generates adjustments to the box to better match the object shape [9].

Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

2.3AlexNet

AlexNet is a convolutional neural Network used for classification which has 5 Convolutional layers, 3 fullyconnected layers and 1 softmax layer with 1000 outputs for classification as his architecture.

2.4 YOLO

YOLO is real-time object detection. It applies one neural network to the complete image dividing the image into regions and predicts bounding boxes and possibilities for every region.

Predicted probabilities are the basis on which these bounding boxes are weighted [8]. A single neural network predicts bounding boxes and class possibilities directly from full pictures in one evaluation. Since the full detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

2.5 VGG

VGG network is another convolution neural network architecture used for image classification.

2.6 MobileNets

To build lightweight deep neural networks MobileNets are used. It is baesd on a streamlined architecture that uses depth-wise separable convolutions. MobileNet uses 3×3 depth-wise separable convolutions that uses between 8 times less computation than standard convolution at solely alittle reduction accuracy. Applications and use cases including object detection, fine grain classification, face attributes and large scale-localization [7].

2.7 Tensor flow

Tensor flow is an open source software library for high performance numerical computation. It allows simple deployment of computation across a range of platforms (CPUs, GPUs, TPUs) due to its versatile design also from desktops to clusters of servers to mobile and edge devices. Tensor flow was designed and developed by researchers and engineers from the Google Brain team at intervals Google's AI organization, it comes with robust support for machine learning and deep learning and the versatile numerical computation core is used across several alternative scientific domains.

To construct, train and deploy Object Detection Models TensorFlow is used that makes it easy and also it provides a collection of Detection Models pre-trained on the COCO dataset, the Kitti dataset, and the Open Images dataset [10]. One among the numerous Detection Models is that the combination of Single Shot Detector (SSDs) and Mobile Nets architecture that is quick, efficient and doesn't need huge computational capability to accomplish the object Detection task, an example of which can be seen on the image below. This document is template. We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace(copy-paste) the content with your own material.Number the reference items consecutively in square brackets (e.g. [1]).However the authors name can be used along with the reference number in the running text. The order of reference in the running text should match with the list of references at the end of the paper.

3. APPLICATION OF OBJECT DETECTION

The major applications of Object Detection:-

3.1 Facial Recognition

"Deep Face" is a deep learning facial recognition system developed to identify human faces in a digital image. Designed and developed by a group of researchers in Facebook. Google also has its own facial recognition system in Google Photos, which automatically seperates all the photos according to the person in the image.

There are various components involved in Facial Recognition or authors could say it focuses on various aspects like the eyes, nose, mouth and the eyebrows for recognizing a faces.

3.2 People Counting

People counting is also a part of object detection which can be used for various purposes like finding person or a criminal; it is used for analysing store performance or statistics of crowd during festivals. This process is considered a difficult one as people move out of the frame quickly.

3.3 Industrial Quality Check

Object detection also plays an important role in industrial processes to identify or recognize products. Finding a particular object through visual examination could be a basic task that's involved in multiple industrial processes like sorting, inventory management, machining, quality management, packaging and so on. Inventory management can be terribly tough as things are hard to trace in real time. Automatic object counting and localization permits improving inventory accuracy.

3.4 Self Driving Cars

Self-driving is the future most promising technology to be used, but the working behind can be very complex as it combines a variety of techniques to perceive their surroundings, including radar, laser light, GPS, odometer, and computer vision. Advanced control systems interpret sensory info to allow navigation methods to work, as well as obstacles and it. This is a big step towards Driverless cars as it happens at very fast speed.

3.5 Security

Object Detection plays a vital role in the field of Security; it takes part in major fields such as face ID of Apple or the retina scan used in all the sci-fi movies. Government also widely use this application to access the security feed and match it with their existing database to find any criminals or to detecting objects like car number involved in criminal activities. The applications are limitless.

3.6 Object Detection Workflow

Every Object Detection Algorithm works on the same principle and it's just the working that differs from others.

3.7 Feature Extraction

They focus on extracting features from the images that are given as the input at hands and then it uses these features to determine the class of the image.

4. TECHNIQUES USED

4.1 Deep learning

The field of artificial intelligence is essential when machines can do tasks that typically require human intelligence. It comes under the layer of machine learning, where machines can acquire skills and learn from past

experience without any involvement of human. Deep learning comes under machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data. The concept of deep learning is based on humans' experiences; the deep learning algorithm would perform a task continuously so that it can improve the outcome. Neural networks have various (deep) layers that enable learning. Any drawback that needs "thought" to work out could be a drawback deep learning can learn to unravel.

4.2 OpenCV

OpenCV stands for Open supply pc Vision Library is associate open supply pc vision and machine learning software system library. The purpose of creation of OpenCV was to produce a standard infrastructure for computer vision applications and to accelerate the utilization of machine perception within the business product [6]. It becomes very easy for businesses to utilize and modify the code with OpenCV as it is a BSD-licensed product. It is a rich wholesome libraby as it contains 2500 optimized algorithms, which also includes a comprehensive set of both classic and progressive computer vision and machine learning algorithms. These algorithms is used for various functions such as discover and acknowledging faces. Identify objects classify human actions. In videos, track camera movements, track moving objects. Extract 3D models of objects, manufacture 3D purpose clouds from stereo cameras, sew pictures along to provide a high-resolution image of a complete scene, find similar pictures from a picture information, remove red eyes from images that are clicked with the flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality.

4.3 Caffe Model

Caffe is a framework of Deep Learning and it was made used for the implementation and to access the following things in an object detection system.

•Expression: Models and optimizations are defined as plaintext schemas in the caffe model unlike others which use codes for this purpose.

•Speed: for research and industry alike speed is crucial for state-of-the-art models and massive data [11].

•Modularity: Flexibility and extension is majorly required for the new tasks and different settings.

•Openness: Common code, reference models, and reproducibility are the basic requirements of scientific and applied progress.

Types of Caffe Models

i) Open Pose
The first real-time multi-person system is portrayed by OpenPose which can collectively sight human body, hand, and facial keypoints (in total 130 keypoints) on single pictures.

ii) Fully Convolutional Networks for Semantic Segmentation
In the absolutely convolutional networks (FCNs) Fully Convolutional Networks are the reference implementation of the models and code for the within the PAMI FCN and CVPR FCN papers.

iii) Cnn-vis
Cnn-vis is an open-source tool that lets you use convolutional neural networks to generate images. It has taken inspiration from the Google's recent Inceptionism blog post.

iv) Speech Recognition
Speech Recognition with the caffe deep learning framework.

v) DeconvNet
Learning Deconvolution Network for Semantic Segmentation.

vi) Coupled Face Generation
This is the open source repository for the Coupled Generative Adversarial Network (CoupledGAN or CoGAN) work.These models are compatible with Caffe master, unlike earlier FCNs that required a pre-release branch (note: this reference edition of the models remains ongoing and not all of the models have yet been ported to master).

vii) Codes for Fast Image Retrieval
To create the hash-like binary codes it provides effective framework for fast image retrieval.

viii) GoogleNet_cars on car model classification
On 431 car models in CompCars dataset, GoogleNet model are pre-trained on ImageNet classification task and are then fine-tuned.

ix) SegNet and Bayesian SegNet
SegNet is real-time semantic segmentation architecture for scene understanding.

x) Emotion Recognition in the Wild via Convolutional Neural
This provides models for facial emotion classification for different image representation obtained using mapped binary patterns.

xi) Deep Hand
It gives pre-trained CNN models.

xii) DeepYeast
Deep Yeast may be an 11-layer convolutional neural network trained on biaural research pictures of yeast cells carrying fluorescent proteins with totally different subcellular localizations.

NumPy: The full form of NumPY is "Numeric Python" or "Numerical Python". It is referred as extension module for Python, and written in C language most of the times. This guarantees a great execution speed of the precompiled mathematical and numerical functionalities of Numpy. NumPy provides the Python many powerful data structures which lets its implement multi-dimensional arrays and matrices which make it more strong language. With matrices and arrays the data structures used give efficient and prominent calculations. Better know under the heading of "big data", the implementation is even aiming at vast matrices and arrays. To operate on these matrices and arrays this module also provides a large library of high level mathematical functions [13].

Python VS other languages for Object Detection: Object detection may be a domain-specific variation of the machine learning prediction drawback. Intel's OpenCV library that is implemented in C/C++ has its interfaces offered during a} very vary of programming environments like C#, Matlab, Octave, R, Python and then on. Why Python codes are much better option than other language codes for object detection are more compact and readable code [5].

- Python uses zero-based indexing.
- Dictionary (hashes) support provided.
- Simple and elegant Object-oriented programming
- Free and open
- Multiple functions can be package in one module
- More choices in graphics packages and toolsets Supervised learning also plays an important role.

The utility of unsupervised pre-training is usually evaluated on the premise of what performance is achieved when supervised fine-tuning. This paper reviews and discusses the fundamentals of learning as well as supervised learning for classification models, and also talks about the mini batch stochastic gradient descent algorithm that is used to fine-tune many of the models.

Object Classification in Moving Object Detection Object classification works on the shape, motion, color and texture. The classification can be done under various categories like plants, objects, animals, humans etc. The key concept of object classification is tracking objects and analysing their features.

i) Shape-Based
A mixture of image-based and scene based object parameters such as image blob (binary large object) area, the as pectration of blob bounding box and camera zoom is given as input to this detection system. Classification is performed on the basis of the blob at each and every frame. The results are kept in the histogram.

ii) Motion-Based
When an easy image is given as an input with no objects in motion, this classification isn't required. In general, non-rigid articulated human motion shows a periodic property; therefore this has been used as a powerful clue for classification of moving objects. based on this useful clue, human motion is distinguished from different objects motion. ColorBased- though color isn't an applicable live alone for police investigation and following objects, but the low process value of the colour primarily based algorithms makes the coloura awfully smart feature to be exploited. As an example, the color-histogram based technique is employed for detection of vehicles in period. Color bar chart describes the colour distribution in a very given region that is powerful against partial occlusions.

iii) Texture-Based
The texture-based approaches with the assistance of texture pattern recognition work just like motion-based approaches. It provides higher accuracy, by exploitation overlapping native distinction social control however might need longer, which may be improved exploitation some quick techniques. I. proposed WORK Authors have applied period object detection exploitation deep learning and OpenCV to figure to work with video streams and video files. This will be accomplished using the highly efficient open computer vision. Implementation of proposed strategy includes caffe-model based on Google Image Scenery; Caffe offers the model definitions, optimization settings, pre-trained weights[4]. Prerequisite includes Python 3.7, OpenCV 4 packages and numpy to complete this task of object detection. NumPy is the elementary package for scientific computing with Python. It contains among other things: a strong N-dimensional array object, subtle (broadcasting) functions tools for integrating C/C++ and fortran code, helpful linear algebra, Fourier transform, and random number capabilities. Numpy works in backend to provide statistical information of resemblance of object with the image scenery caffemodel database. Object clusters can be created according to fuzzy value provided by NumPy. This project can detect live objects from the videos and images.

5. LEARNING FEATURE HIERARCHY

Learn hierarchy all the way from pixels classifier One layer extracts features from output of previous layer, train all layers jointly

Zero-One Loss
The models given in these deep learning tutorials are largely used for classification. The major aim of training a classifier is to reduce the amount of errors (zero-one loss) on unseen examples

Negative Log-Likelihood Loss
Optimizing it for large models (thousands or millions of parameters) is prohibitively expensive (computationally) because the zero-one loss isn't differentiable. In order to achieve this maximization of the log-likelihood is done on the classifier given all the labels in a training set [14].The likelihood of the correct class and number of right predictions is not the equal, but they are pretty similar from the point of view of a randomly initialized classifier. As the likelihood and zero-one loss are different objectives but we should always see that they are co-related on the validation set but sometimes one will rise while the other falls, or vice-versa.

Stochastic Gradient Descent
Ordinary gradient descent is an easy rule within which we repeatedly create tiny steps downward on an error surface defined by a loss function of some parameters. For the aim of normal gradient descent we take into account that the training data is rolled into the loss function. Then the pseudo code of this algorithm can be represented as
Stochastic gradient descent (SGD) works according to similar principles as random gradient descent (SGD) operates on the basis of similar principles as normal gradient descent. It quickly proceeds by estimating the gradient from simply a few examples at a time instead of complete training set. In its purest kind, we use simply one example at a time to estimate the gradient.

Caffe is a deep learning framework or else we can say a library it's made with expression speed and modularity in mind they will put by Berkeley artificial intelligence research and created by young King Gia there are many deep learning or machine learning frameworks for computer vision like tensorflow ,Tiano, Charis and SVM[2]. But why exactly we implement edition cafe there as on is its expressive architecture we can easily switch between CPU and GPU while training on GPU machine modules and optimization for Our problem is defined by configuration without hard coding. It supports extensible code since cafes are open source library. It is four foot by over twenty thous and developers and github since its birth it offers coding platform in extensible languages like Python and C++. The next reason is speed for training the neural networks speed is the primary constraint. Caffe can process over million images in a single day with the standard media GPU that is milliseconds per image. Whereas the same dataset of million images can take weeks for Tiana and Kara's Caffe is the fastest convolution neural network present community as mentioned earlier since its open source library huge number of research are powered by cafe and every single day something new is coming out of it.

6. CONCLUSION

Deep learning based object detection has been a research hotspot in recent years. This project starts on generic object detection pipelines which provide base architectures for other related tasks. With the help of this the three other common tasks, namely object detection, face detection and pedestrian detection, can be accomplished[1]. Authors accomplished this by combing two 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 as it can create problem in recognizing the object. The end result is a deep learning-based object detector that can process around 6-8 FPS.

ACKNOWLEDGEMENT

I would like to express my gratitude to Ms. Shilpi Singh and Ms. Latha Banda, my research supervisors, for their guidance in this research work. I would also like to thank my teachers for their help in making the implementation of my work successful.
Finally, I would like to thank my parents and my friends in the University for Support and encouragement throughout my study.

REFERENCES

[1] Bruckner, Daniel. Ml-o-scope: a diagnostic visualization system for deep machine learning pipelines. No. UCB/EECS-2014-99. CALIFORNIA UNIV BERKELEY DEPT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES, 2014.

[2] K Saleh, Imad, Mehdi Ammi, and Samuel Szoniecky, eds. Challenges of the Internet of Things: Technique, Use, Ethics. John Wiley & Sons, 2018.

[3] Petrov, Yordan. Improving object detection by exploiting semantic relations between objects. MS thesis. UniversitatPolitècnica de Catalunya, 2017.

[4] Nikouei, Seyed Yahya, et al. "Intelligent Surveillance as an Edge Network Service: from Harr-Cascade, SVM to a Lightweight CNN." arXiv preprint arXiv:1805.00331 (2018).

[5] Thakar, Kartikey, et al. "Implementation and analysis of template matching for image registration on DevKit-8500D." Optik-International Journal for Light and Electron Optics 130 (2017): 935-944..

[6]Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.", 2008.

[7] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[8] Kong, Tao, et al. "Ron: Reverse connection with objectness prior networks for object detection." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[9] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[10] Veiga, Francisco José Lopes. "Image Processing for Detection of Vehicles In Motion." (2018).

[11]Huaizheng Zhang, Han Hu, Guanyu Gao, Yonggang Wen, Kyle Guan, "Deepqoe: A Unified Framework for Learning to Predict Video QoE", Multimedia and Expo (ICME) 2018 IEEE International Conference on, pp. 1-6, 2018.

[12] Shijian Tang and Ye Yuan,"Object Detection based on Conventional Neural Network".

[13] R. P. S. Manikandan, A. M. Kalpana, "A study on feature selection in big data", Computer Communication and Informatics (ICCCI) 2017 International Conference on, pp. 1-5, 2017

[14] Warde-Farley, David. "Feedforward deep architectures for classification and synthesis." (2018).

[15] Shilpi singh et al" An Analytic approach for 3D Shape descriptor for face recognition", International Journal of Electrical, Electronics, Computer Science & Engineering (IJEECSE), Special Issue - ICSCAAIT-2018 | E-ISSN: 2348-2273 | P-ISSN: 2454-1222,pp-138-140. Available Online at www.ijeecse.com

[16] Streitz, Norbert A., and Shin'ichi Konomi, eds. Distributed, Ambient and Pervasive Interactions: Technologies and Contexts: 6th International Conference, DAPI 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings. Vol. 10922. Springer, 2018.