International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Issue **02** | **2021**

Object Detection using Deep Learning - Python

Khushboo Singh
Department of MCA, Computer Science, Jain (Deemed-to-be) University, Bangalore, Karnataka, India

ABSTRACT

Object Detection is a well-known computer technology connected with computer vision and image processing that focuses on detecting objects or its instances of a certain class like humans, animals etc, it also describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization is refers to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. In this project, we are using highly accurate object detection-algorithms and methods such as R-CNN, Fast-RCNN, and SSD. Using these methods and algorithms, based on deep learning which is also based on machine learning require lots of mathematical and deep learning frameworks understanding by using dependencies such as TensorFlow, OpenCV etc.

Keywords: ***Object Detection, OpenCV, R-CNN, Fast-RCNN, SSD.***

1. INTRODUCTION

Object Detection (OD) and location in digital images has become one of the most important applications for industries to ease user, save time and to achieve parallelism. This is not a new technique but improvement in object detection is still required in order to achieve the targeted objective more efficiently and accurately. Given the real time webcam data, this application will use OpenCV library to track an object and allows the user to draw by moving the object in the air. A variety of problems of current interest in computer vision require the ability to track moving objects in real time for purposes such as surveillance, video conferencing, robot navigation, etc.

OD is one of the basic spaces of examination because of routine change moving of article and variety in scene size, impediments, appearance varieties, and sense of self movement and enlightenment changes. In particular, highlight determination is the essential job in object following. It is identified with numerous constant applications like vehicle discernment, video observation, etc. To defeat the issue of discovery, following identified with object development and appearance. A large portion of the calculation centers around the following calculation to smoothen the video succession. Then again, scarcely any techniques utilize the earlier accessible data about object shape, shading, surface, etc. Shading has been generally utilized progressively global positioning frameworks. It offers a few critical benefits over mathematical signals like computational straightforwardness, heartiness under halfway impediment, pivot, scale and goal changes. Despite the fact that shading techniques end up being proficient in an assortment of vision applications, there are a few issues related with these strategies for which shading consistency is quite possibly the most significant. In the global positioning framework carried out, the shading masses are being followed. The idea of masses as a portrayal for picture highlights has a long history in Computer vision and has had a wide range of numerical definitions.

2. LITERATURE SURVEY

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.

Basically, an OD system can be described easily, by seeing Figure 1 which shows the basic stages that are involved in the process of OD. The basic input to the OD system can be an image or a scene in case of videos. The basic aim of this system is to detect objects that are present in the image or scene or simply in other words the system needs to categorize the various objects into respective object classes.

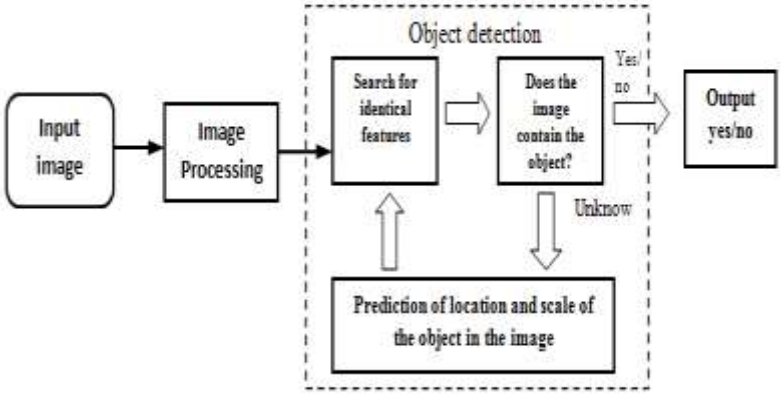

Fig: 1 Object Detection Model

3. OBJECT DETECTION AND TRACKING ALGORITHMS

- **R-CNN**

The Region-based Convolutional Network method (RCNN) is a combination of region proposals with Convolution Neural Networks (CNNs). R-CNN helps in localising objects with a deep network and training a high-capacity model with only a small quantity of annotated detection data. It achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. R-CNN has the capability to scale to thousands of object classes without resorting to approximate techniques, including hashing.

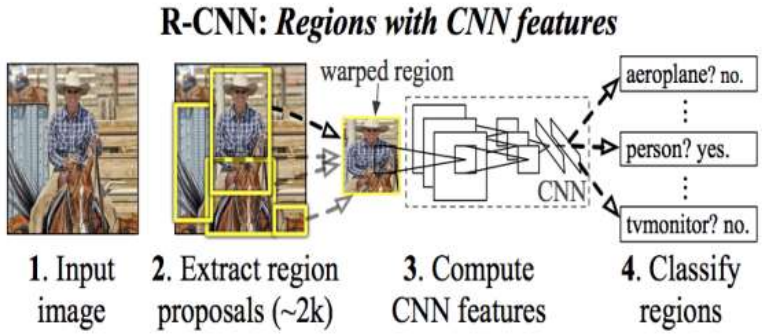

Fig: 2 R-CNN Model

- **Fast R-CNN**

Fast Region-Based Convolutional Network method or Fast R-CNN is a training algorithm for object detection. This algorithm mainly fixes the disadvantages of R-CNN and SPPnet, while improving on their speed and accuracy.

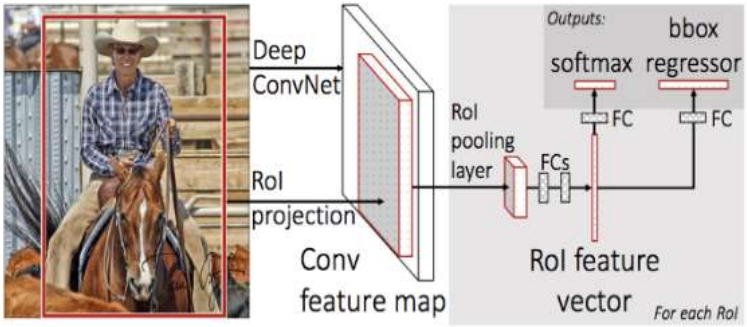

Fig: 3 Fast R-CNN Model

- **SSD**

Single Shot Detector (SSD) is a method for detecting objects in images using a single deep neural network. The SSD approach discretises the output space of bounding boxes into a set of default boxes over different aspect ratios. After discretising, the method scales per feature map location. The Single Shot Detector

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

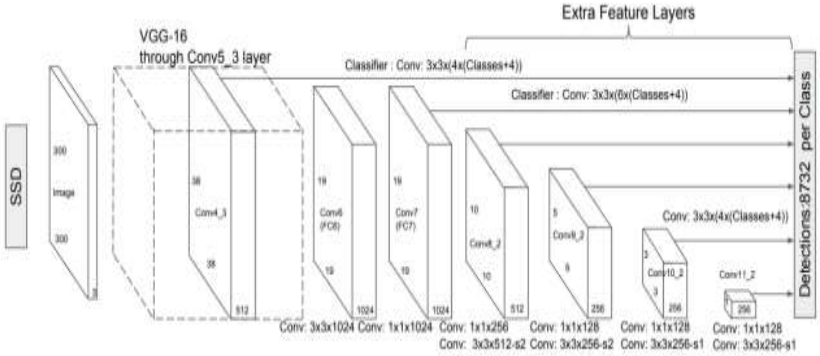

Fig: 4 SSD Model

4. PROPOSED MODEL

We aim to accomplishing an object detection by predicting a set of bounding boxes, which represent potential objects. More precisely, we use a Deep Neural Network (DNN), which outputs a fixed number of bounding boxes. In addition, it outputs a score for each box expressing the network confidence of this box containing an object.

- **Model** To formalize the above idea, we encode the i-th object box and its associated confidence as node values of the last net layer:
- **Bounding Box:** In object detection, we usually use a bounding box to describe the target location. The bounding box is a rectangular box that can be determined by the xx and yy axis coordinates in the upper-left corner and the xx and yy axis coordinates in the lower-right corner of the rectangle. Another commonly used bounding box representation is the xx and yy axis coordinates of the bounding box center, and its width and height.
- **Confidence:** The confidence score reflects how likely the box contains an object and how accurate is the bounding box. If no object exists in that cell, the confidence score should be zero. The confidence score is an output of the underlying neural network, so it is the result of training across a dataset where correct predictions are trained towards 1 and incorrect predictions are trained towards 0.
- **Training Objective:** We train a Deep Neural Network (DNN) to predict bounding boxes and their confidence scores for each training image such that the highest scoring boxes match well the ground truth object boxes for the image. Suppose that for a particular training example, M objects were labeled by bounding boxes gj , j ∈ {1, . . . , M}. In practice, the number of predictions K is much larger than the number of ground truth boxes M. Therefore, we try to optimize only the subset of predicted boxes which match best the ground truth ones. We optimize their locations to improve their match and maximize their confidences. At the same time we minimize the confidences of the remaining predictions, which are considered not to localize the true objects well.

5. FLOWCHAT

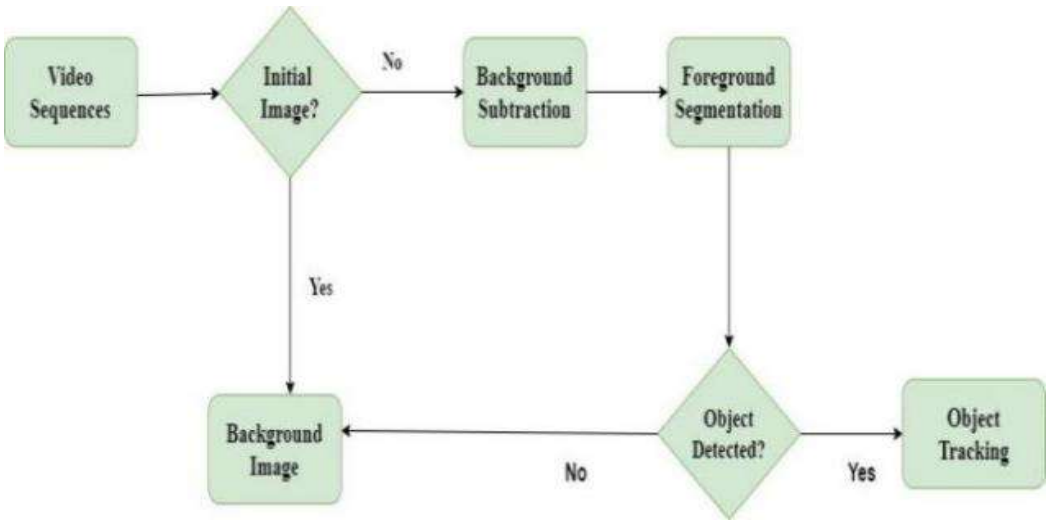

Fig: 5 Flowchat of Object Detection

6. BACKGROUND WORK

The aim of object detection is to recognize all cases of items from a referred to class, like individuals, vehicles or countenances in a picture. For the most part, just few examples of the item are available in the picture, yet there is an enormous number of potential areas and scales at which they can happen and that need to some way or another be investigated. Every identification of the picture is accounted for with some type of posture data. This is pretty much as basic as the area of the article, an area and scale, or the degree of the item characterized as far as a bouncing box. In some other situations, the posture data is more definite and contains the boundaries of a direct or non-straight change. For instance for face recognition in a face finder may figure the areas of the eyes, nose and mouth, as well as the bounding box of the face. An illustration of bike recognition in a picture that indicates the areas of specific parts is appeared in Figure 6. The posture can likewise be characterized by a three-dimensional change indicating the area of the item comparative with the camera.

Fig: 6 Bicycle Detection in an image

7. METHODOLOGY

7.1 Object Detection

- **Frame Differencing:** This technique, we use to detect motion and find out the moving objects and stationary objects in the image scene. Frame Differencing is an algorithm to observe the motion in the image scene and detect the moving objects using a fixed surveillance camera. In this technique, the model captured the image from the static camera and sequence of image from the camera stream. In the second phase, the absolute difference is calculated between the consecutive frames and the record the difference. Finally, the image processing techniques are applied to remove the noise.
- **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.

7.2 Object Tracking

Object Detection 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.

8. RESULTS AND ANALYSIS

Based on SSD algorithm, a python program was developed for the algorithm and implemented in OpenCV. OpenCV is run in PYcharm 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: 7 Results Image frame 1

Fig: 8 Results Image frame 2

Fig: 9 Real-Time Camera view Detection

Fig.7 to 9 shows the real time detection of horse, car, person and dog with confidence levels 99.90%, 98.68%, 99.9% and 97.46% respectively. The model was trained to detect 21 objects class like dog, motorbike etc. with accuracy of 99%.

Average Precision Results table

Model	mAP(%)	MAX RECALL(%)	pd(%)	Speed(ms)
FAST-RCNN	88.44	99.91	61.78	37.63
RCNN	97.92	99.90	56.85	19.26
SSD	97.92	99.93	94.25	17.42

Table 1. Average precision results for the baseline Object Detector, we proposed RCNN, FAST-RCNN and SSD approach and state-of-the-art results on all 21 classes of the COCO and ImageNet dataset.

9. 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 great discovery and following outcomes on the article prepared and can additionally used in explicit situations to identify, track and react to the specific focused on objects in the video observation. This continuous investigation of the biological system can yield incredible outcomes by empowering security, request and utility for any undertaking. 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.

10. REFERENCES

[1] Cadena, C., Dick, A., and Reid, I. (2015). “A fast, modular scene understanding sys-tem using context-aware object detection,” in Robotics and Automation (ICRA),2015 IEEE International Conference on (Seattle, WA).

[2] Correa, M., Hermosilla, G., Verschae, R., and Ruiz-del-Solar, J. (2012). Humandetection and identification by robots using thermal and visual information indomestic environments. J. Intell. Robot Syst. 66, 223–243. doi:10.1007/s10846-011-9612-2

[3] Dalal, N., and Triggs, B. (2005). “Histograms of oriented gradients for humandetection,” in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEEComputer Society Conference on, Vol. 1 (San Diego, CA: IEEE), 886–893. doi:10.1109/CVPR.2005.177

[4] Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2014). "Scalable object detec-tion using deep neural networks," in Computer Vision and Pattern Recognition Frontiers in Robotics and AI www.frontiersin.org November 2015
[5] Bourdev, L. D., and Malik, J. (2009). "Poselets: body part detectors trained using 3dhuman pose annotations," in IEEE 12th International Conference on ComputerVision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE),1365–1372.
[6] Bourdev, L. D., Maji, S., Brox, T., and Malik, J. (2010). "Detecting peopleusing mutually consistent poselet activations," in Computer Vision – ECCV2010 – 11th European Conference on Computer Vision, Heraklion, Crete, Greece,September 5-11, 2010, Proceedings, Part VI, Volume 6316 of Lecture Notes inComputer Science, eds K. Daniilidis, P. Maragos, and N. Paragios (Heraklion:Springer), 168–181.
[7] Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490. doi:10.1109/TPAMI.2004.108
[8] Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. Int. J.Comput. Vis. 1, 333–356. doi:10.1007/BF00133571
[9] Azizpour, H., and Laptev, I. (2012). "Object detection using strongly-supervisedeformable part models," in Computer Vision-ECCV 2012 (Florence: Springer),836–849.
[10] P. Druzhkov and V. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," Pattern Recognition and Image Anal., vol. 26, no. 1, p. 9, 2016.
[11] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. LeCun, "Learning convolutional feature hierarchies for visual recognition," in NIPS, 2010.
[12] D. Tome, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, and ` S. Tubaro, "Deep convolutional neural networks for pedestrian detection," Signal Process.: Image Commun., vol. 47, pp. 482–489, 2016.
[13] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multitask learning framework for face detection, landmark localization, pose estimation, and gender recognition," arXiv:1603.01249, 2016.
[14] L. Huang, Y. Yang, Y. Deng, and Y. Yu, "Densebox: Unifying landmark localization with end to end object detection," arXiv:1509.04874, 2015.
[15] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion," IEEE Trans. Cybern., vol. 48, pp. 103–114, 2018.
[16] Benbouzid, D., Busa-Fekete, R., and Kegl, B. (2012). "Fast classification using sparsedecision dags," in Proceedings of the 29th International Conference on MachineLearning (ICML-12), ICML '12, eds J. Langford and J. Pineau (New York, NY:Omnipress), 951–958.
[17] Azizpour, H., and Laptev, I. (2012). "Object detection using strongly-supervisedeformable part models," in ComputerVision-ECCV 2012 (Florence: Springer),836–849.
[18] S. Hong, B. Roh, K.-H. Kim, Y. Cheon, and M. Park, "Pvanet: Lightweight deep neural networks for real-time object detection," arXiv:1611.08588, 2016.
[19] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Convolutional channel features," in ICCV, 2015.
[20] S. Azadi, J. Feng, and T. Darrell, "Learning detection with diverse proposals," in CVPR, 2017.
[21] J. Dong, X. Fei, and S. Soatto, "Visual-inertial-semantic scene representation for 3d object detection," in CVPR, 2017.
[22] Y. Fang, K. Kuan, J. Lin, C. Tan, and V. Chandrasekhar, "Object detection meets knowledge graphs," in IJCAI, 2017.
[23] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3d object proposals for accurate object class detection," in NIPS, 2015.