JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Object Detection using SSD

Abhishek Kodannur, Sai Shendge, Jyoti Wahule, Saurabh Kumar
Computer Engineering

***Abstract* - Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. As one of the mainstream detection algorithms, it greatly increases the detection speed and accuracy as detecting objects in a matter of milliseconds. Well researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. This paper includes the OpenCV's Deep Neural Network which is used to train the pre-trained framework Caffe.**

Keywords - SSD algorithm, OpenCV's DNN's, Object Detection, MobileNetSSD.

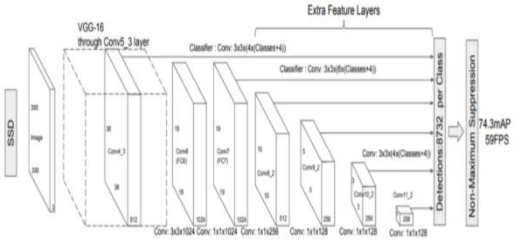
I. INTRODUCTION

Object detection is so important in the world right now as it is used in many fields like Health-care, Agriculture, Autonomous Driving, and more. It provides an efficient way of handling image classification by detecting the object in the image and letting us know where it is in the image using localization, That is, it creates a bounding box around the object. This may sound like just another image classification algorithm but it is super powerful in the current world. Self Driving Cars use object detection to detect what is there in front of them. They are used in health care to understand and classify different types of tumors and diseases in the human body.

The Applications of Object Detection are endless. But what makes it more interesting is to be able to achieve such technology in real-time. This has been very challenging so far. By using a simple technique we can boost the performance of object detection in real-time drastically. This can be observed by an increase in FPS (Frames Per Second) and its faster processing of each frame. We will exactly discuss this methodology in this article by using OpenCV's Deep Neural Network (or) simply called DNNs.

II. SSD ALGORTIHM

SSD (Single Shot Multi-Box Detector) is a Deep Learning model used to detect objects in an image or from a video source. SSD is a simple approach to solve the problem but is still very effective. SSD has 2 components and they are the **Backbone Model** and **SSD Head.** Backbone model is a pre-trained image classification network as a feature extractor. SSD Head is another set of convolutional layers added to this backbone and the outputs are interpreted as the bounding boxes and classes of objects in the spatial location of the final layer's activation.



SSD divides the image as Grids, and each grid cell responsible for detecting objects in that region of the image. if there are no images in the frame then we output as "0" precisely.if there are more than one objects of the same instance in the single image. This is where Anchor box comes in to play. Anchor boxes are simple boxes that are assigned with multiple anchors/prior boxes, which are predefined and have fixed size and shape within the grid cells.

III. OpenCV's DNN's

OpenCV's Deep Neural Network (DNNs) is a module that is used to train and test deep learning models. There is a framework that is used to train models that is **Caffe**. We can train the DNNs using just our CPUs or GPUs. Even using just our CPUs gives us a pretty decent performance.

The OpenCV DNN module supports deep learning interface on images and videos. It is highly optimizable for intel processors. We can use many different frameworks that OpenCV supports and provides. What we have used in the Caffe framework.

Caffe :

We need two things to use a pre-trained Caffe model. One is the model. Caffemodel file that contains the pre-trained weights. The other one is the model architecture file which has a .prototxt extension. It is like a plain text file with a JSON like structure containing all the neural network layers' definitions.

IV. TRAINING OF SSD

Input to SSD is an input image with ground truth bounding boxes for each object. Conv layers evaluate boxes of different sizes at each location in several feature maps with different scales.

Multiboxes are like anchors. We have multiple default boxes of different sizes, aspect ratio across the entire image. SSD used 8732 Boxes. This helps with finding the default box that most overlaps with the grounding for containing objects.

V. MATCHING STRATEGY

During training time the default boxes are matched over aspect ratio, location and scale to the ground truth boxes. We select the boxes with the highest overlap with the ground truth bounding boxes. IoU (Intersection over union) between predicted boxes and ground truth should be greater than 0.5. we finally pick up the predicted box with the maximum overlap with the ground truth.

In any image, we match default boxes corresponding to the objects in the image. They are treated as positive boxes and the rest of the boxes are treated as negatives

Each prediction is composed of :

- Bounding box with shape offset. Δcx, Δcy, h and w, representing the offsets from the center of the default box and its height and width.
- Confidences for all the object categories or all the classes. Class 0 is reserved to indicate absence of object.

Loss function used in SSD and MultiBox loss, which consists of two terms : the confidence loss and the localization loss.

VI. DATA AUGMENTAION

Data Augmentation technique is used ot handle variants of object sizes and shapes using shearing, zoom in, zoom out, flipping, cropping, etc. Application of data augmentation makes the model more robust to various input object sizes and shapes. This help improve the accuracy of the model.

Each training image is randomly sampled by one of the following options:

- Use the entire original input image.
- Sample a patch of the object so that the minimum overlap with the objects is 0.1, 0.3, 0.5, 0.7, or 0.9.
- Randomly sample a batch.

VII. INFERENCE using SSD

SSD uses the default boxes of different scales, shapes and aspect ratio on different output layers.

It uses 8732 boxes for a better coverage of location, scale and aspect ratios. Most of the prediction will not contain any object. SSD drops predictions that have confidence score that is lower than 0.01. we then apply Non Max Suppression (NMS) overlap of 0.45 per class and keep the top 200 detection per image.

VIII. CONCLUSION

The paper presents the SSD model, which is modified by adding OpenCV's Deep Neural Network (DNNs). The given model can detect objects in images, videos (both online and local). the detection process is done without a drop of detection speed. The object counting function is added to the model. The usable object detection system was built using Caffe framework.

REFERENCES

1. IEEE transactions of Qianjun Shuai, Xingwen Wu on Object detection using SSD.
2. JEONG J, PARK H, KWAK N. Enhancement of SSD by concatenating feature maps for object detection.
3. Huieun Kim, Youngwan Lee, Byeounghak Yim, Eunsoo Park, Hakil Kim On-road Object detection using deep neural network 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE -Asia).
4. Ayesha Younis, Li Shixin, Shelembi Jn, Zhang Hai, Real time object detection using pre-trained deep learning models using MobileNet-SSD. Proceedings of 2020 the 6th international conference on computing and data engineering.
5. An enhanced SSD with feature fusion and visual reasoning for object detection, by Jiaxu Leng, Ying Liu at Neural Computing and Applications 31, 65496558, 2019.
6. Ssd: Single shot multi-box detector, by Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, Alexander C Berg. Computer vision ECCS 2016. 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14,21-37, 2016.