International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Object Detection and Recognition in Satellite map images using Deep Learning

Shubham Pal[1], Prof. Pramila M. Chawan[2]

[1]M.Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *An Object Detection is related to Computer Vision. Object detection enables detecting instances of objects in images and videos. It identifies the feature of Images rather than traditional object detection methods and generates an intelligent understanding of images just like human vision works. In this paper, We revived begins the brief introduction of deep learning and object detection framework like Convolutional Neural Network(CNN), Recurrent neural network (RNN), faster RNN, You only look once (YOLO). Then we focus on our proposed object detection architectures along with some modifications. The traditional model detects a small object in images. We have some modifications to the model. Our proposed method gives the correct result with accuracy.*

Key Words: k means, OpenCV, CNN, R-CNN, Faster-RNN, and YOLO.

1. INTRODUCTION

Deep learning is part of machine learning. Too many methods have been proposed for object detection. Methods of object detection fall under deep learning. Object detection is a computer technology and widely used in Computer vision. Deep learning has been becoming popular since 2006.

1.1 A brief Overview of object detection

Object Detection is a Computer Vision technique. Object detection is a significant research area in Computer Vision. Which can be applied to many applications such as Driverless cars, security, surveillance, machine inspection, etc. Object Detection is used to identify the location of the object in an image, Face detection, medical imaging, etc. Invention and Evolution of Deep learning have changed the traditional ways of object detection and reorganization system.

Computer Vision identifies features present in images, Classifying Objects in the image, Classifying images along with localization, drawing a bounding box around object present in the image, Object segmentation or semantic segmentation, Neural style Transfer. Deep learning methods are the strongest method for object detection. To understanding images, we not only concentrate on classifying images but also try to estimate the concepts and locations of each object in images.

2. LITERATURE SURVEY

Three are different approaches has presented by many researchers. An algorithm for the first face detector was invented by Paul Viola and Michael Jones 2001. The face had detected in real-time on Webcam feed. It was implemented by Opencv and Face Detection. This was not able to detect some orientation like upside down, titled, wearing a mask, etc. Due to the massive development of Object detection in Deep leaning, object detection classified model into (1) Model-based on region proposal; (2) Model-based on regression/Classification.

2.1 Model-based on Region

2.1.1. CNN: This network was introduced by Authors: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton in 2012. The network consists of five convolutional layers. It takes input as an image which is a 2D array of a pixel with RGB channel. Then Filters or features detector apply to the input image and get output features maps. Multiple convolutional are performed in parallel by applying the ReLU function. CNN works for only one object at a time so it does not work effectively in multiple objects images. CNN became a good standard for image classification after Kriszhevsky's CNN's performance We cannot detect objects which are overlapping and different backgrounds and do not classify these different objects but also do not identify boundaries, differences and relations in other.

2.1.2. RCNN: This network is introduced by Authors: Ross Girshick, Jeff Donahue, Trevor in 2013this network inspired by overfeat. This network includes three main parts, first is region extractor, second is feature extractor and final is classifier. It uses a selective search algorithm for object detection to generate region proposals. Extract 2000 regions for every image. Here 2000 convolutional networks used for each regions of the images. So have one Convolutional network required to process RCNN Region with CNN features divides the image into several regions. Run images through pre-trained AlexNet and finally apply the SVM algorithm.

2.1.3. Fast R-CNN: This network is an improved version of R-CNN which is introduced by Ross Girshick. The article claims that Fast R-CNN 9 times faster than previous R-CNN. Network select sets of bounding boxes then use feature extractor by CNN network then use classifier or regression for output the class of each boxes.

2.1.4. Faster R-CNN: This is an improved version of Fast R-CNN which introduced by Shaoqing Ren, Kaiming He, Ross

Girshick, and Jian Sun in 2015. Image is provided input to a convolutional network that provides convolutional map. To identify the regions here the separate network is used to predict the region proposals.

2.1 Model based on regression/Classification.

2.2.1. YOLO: YOLO (You only look once) at an image to predict what are those objects and where objects are present. A single convolutional network simultaneously predicts multiple bounding boxes and class and probabilities for those boxes. Treats detection as a regression problem. Extremely fast and accurate YOLO takes an image and split it into grids. Each grid cell predicts only one object. YOLO is extremely fast at test time and it requires single network evaluation and performs feature extraction, bounding box prediction, non max suppression, and contextual reasoning all concurrently. YOLO is not applicable for small objects that appears in groups such as flocks of birds. YOLO has several variant like fast YOLO. YOLO is a completely different approach. It looks just once but in clear ways. If a simple image gives through the convolutional network in a single pass and comes out the other end as a 13×13×125 tensor describing the bounding boxes for the grid cells. All you need to do then is compute the final scores for the bounding boxes and throw away the ones scoring lower than 30%.

2.2.2. SSD: SSD (Single Shot MultiBox Detector) Objective of localization and classifications are done in a single forward pass of the SSD network. The first advantage of the network is fast with good accuracy. It runs a convolutional network on input images only one time and computes a features map. Histograms of Oriented Gradients are invented by Navneet Dalal and Bill Triggs invented in 2005. We want to look at each pixel that directly surrounding it. Here compare current pixel to every surrounding pixel. It failed in more generalized object detection with noise and distractions in the background.

3. PROPOSED SYSTEM

3.1PROBLEM STATEMENT

Object detection and Recognition in Satellite map images using Deep learning.

Sample paragraph Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

3.2 PROBLEM ELABORATION

The main objective is to detection and recognition Objects in Real-time. We require rich information in real life. We have to observe the objects which are moving respect to the camera. It will help to recognize objects interaction. We focus on accuracy in this paper.

3.3 PROPOSED METHODALOGY

3.3.1 Darknet 53:

YOLO v2 used a custom deep architecture darknet-19, an originally 19-layer network supplemented with 11 more layers for object detection. With a 30-layer architecture, YOLO v2 often struggled with small object detections. This was attributed to loss of fine-grained features as the layers down sampled the input. To remedy this, YOLO v2 used an identity mapping, concatenating feature maps from a previous layer to capture low level features.

However, YOLO v2's architecture was still lacking some of the most important elements that are now stable in most of state-of-the art algorithms. No residual blocks, no skip connections and no upsampling. YOLO v3 incorporates all of these.

First, YOLO v3 uses a variant of Darknet, which originally has 53 layer network trained on Imagenet. For the task of detection, 53 more layers are stacked onto it, giving us a 106 layer fully convolutional underlying architecture for YOLO v3. This is the reason behind the slowness of YOLO v3 compared to YOLO v2. Here is how the architecture of YOLO now looks like.

3.3.2 Detection of three scales

The newer architecture boasts of residual skip connections, and upsampling. The most salient feature of v3 is that it makes detections at three different scales. YOLO is a fully convolutional network and its eventual output is generated by applying a 1 x 1 kernel on a feature map. In YOLO v3, the detection is done by applying 1 x 1 detection kernels on feature maps of three different sizes at three different places in the network.

The shape of the detection kernel is 1 x 1 x (B x (5 + C)). Here B is the number of bounding boxes a cell on the feature map can predict, "5" is for the 4 bounding box attributes and one object confidence, and C is the number of classes. In YOLO v3 trained on COCO, B = 3 and C = 80, so the kernel size is 1 x 1 x 255. The feature map produced by this kernel has identical height and width of the previous feature map, and has detection attributes along the depth as described above. Before we go further, I'd like to point out that stride of the network, or a layer is defined as the ratio by which it downsamples the input. In the following examples, I will assume we have an input image of size 416 x 416.

YOLO v3 makes prediction at three scales, which are precisely given by down sampling the dimensions of the input image by 32, 16 and 8 respectively.

The first detection is made by the 82nd layer. For the first 81 layers, the image is down sampled by the network,

such that the 81st layer has a stride of 32. If we have an image of 416 x 416, the resultant feature map would be of size 13 x 13. One detection is made here using the 1 x 1 detection kernel, giving us a detection feature map of 13 x 13 x 255.

Then, the feature map from layer 79 is subjected to a few convolutional layers before being up sampled by 2x to dimensions of 26 x 26. This feature map is then depth concatenated with the feature map from layer 61. Then the combined feature maps is again subjected a few 1 x 1 convolutional layers to fuse the features from the earlier layer (61). Then, the second detection is made by the 94th layer, yielding a detection feature map of 26 x 26 x 255.

A similar procedure is followed again, where the feature map from layer 91 is subjected to few convolutional layers before being depth concatenated with a feature map from layer 36. Like before, a few 1 x 1 convolutional layers follow to fuse the information from the previous layer (36). We make the final of the 3 at 106th layer, yielding feature map of size 52 x 52 x 255.

3.3.3 Detecting smaller objects

Detections at different layers helps address the issue of detecting small objects, a frequent complaint with YOLO v2. The upsampled layers concatenated with the previous layers help preserve the fine grained features which help in detecting small objects.

The 13 x 13 layer is responsible for detecting large objects, whereas the 52 x 52 layer detects the smaller objects, with the 26 x 26 layer detecting medium objects. Here is a comparative analysis of different objects picked in the same object by different layers.

3.3.4 Choice of anchor boxes

This model total uses 9 anchor boxes for the detection of an object. We are using k-means clustering to generate 9 anchors. For clustering arrange all anchors in descending order according to the dimensions and assign large anchors for the first scales after three anchors for the second scale, and the last three anchors for the third scale.

This model predicts more bounding boxes. This model predicts boxes at 3 different scales, for the images of 416 x 416, the number of predicted boxes are total 10647

In Class Prediction, Softmax is not used. Independent logistic classifier and binary cross entropy loss are used

3.3.5 More bounding boxes per image

For an input image of same size, YOLO v3 predicts more bounding boxes than YOLO v2. For instance, at it's native resolution of 416 x 416, YOLO v2 predicted 13 x 13 x 5 = 845 boxes. At each grid cell, 5 boxes were detected using 5 anchors.

On the other hand YOLO v3 predicts boxes at 3 different scales. For the same image of 416 x 416, the number of predicted boxes are 10,647. This means that YOLO v3 predicts 10x the number of boxes predicted by YOLO v2. You could easily imagine why it's slower than YOLO v2. At each scale, every grid can predict 3 boxes using 3 anchors. Since there are three scales, the number of anchor boxes used in total are 9, 3 for each scale.

3.3.6 No more soft maxing the classes

YOLO v3 now performs multilabel classification for objects detected in images.

Earlier in YOLO, authors used to softmax the class scores and take the class with maximum score to be the class of the object contained in the bounding box. This has been modified in YOLO v3.

Softmaxing classes rests on the assumption that classes are mutually exclusive, or in simple words, if an object belongs to one class, then it cannot belong to the other. This works fine in COCO dataset.

However, when we have classes like Person and Women in a dataset, then the above assumption fails. This is the reason why the authors of YOLO have refrained from softmaxing the classes. Instead, each class score is predicted using logistic regression and a threshold is used to predict multiple labels for an object. Classes with scores higher than this threshold are assigned to the box.

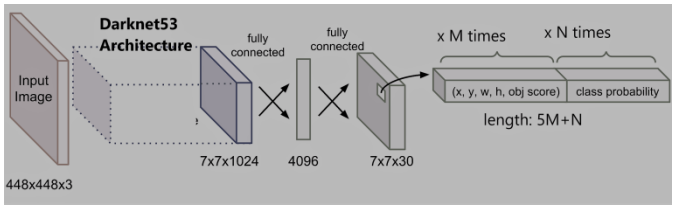

Fig -1: Object Detection model

4. RESULT AND ANALYSIS

The network outputs bounding boxes are each represented by a vector of number of classes + 5 elements.

The first 4 elements represent the center_x, center_y, width and height. The fifth element represents the confidence that the bounding box encloses an object.

The rest of the elements are the confidence associated with each class (i.e. object type). The box is assigned to the class corresponding to the highest score for the box.

The highest score for a box is also called its confidence. If the confidence of a box is less than the given threshold, the bounding box is dropped and not considered for further processing.

The boxes with their confidence equal to or greater than the confidence threshold are then subjected to Non Maximum Suppression. This would reduce the number of overlapping boxes.

Table -1: Compare framework

OS	FRAMEWORK	CPU	TIME PER FRAME
Linux 16.04	Darknet	Intel Core i7-6850K CPU @ 3.60GHz	9370

Linux 16.04	Darknet + OpenMP	Intel Core i7-6850K CPU @ 3.60GHz	1942
Linux 16.04	OpenCV [CPU]	Intel Core i7-6850K CPU @ 3.60GHz	220
Linux 16.04	Darknet	NVIDIA GeForce 1080 Ti GPU	23
macOS	Darknet	2.5 GHz Intel Core i7 CPU	7260
macOS	OpenCV [CPU]	2.5 GHz Intel Core i7 CPU	400

Fig -2: Image after detection

5. CONCLUSION

This paper provides a detailed review of many of the models in this paper related to object detection such as R-CNN, YOLO SSD, etc. Then we have introduced the limitations of each technology. Deep learning-based object detection research has been a hotspot in recent years. This paper provides those models had. This proposed model focus on accuracy than speed. The previous models are not accurate when images have small object Small object in images need to be detected.

REFERENCES

[1] The Object Detection Based on Deep Learning Cong Tang 1,2,3 , Yunsong Feng 1,2,3 , Xing Yang 1,2,3 , Chao Zheng 1,2,3 , Yuanpu Zhou 1,2,3 2017 4th International Conference on Information Science and Control Engineering

[2] Moving object detection and tracking Using Convolutional Neural Networks Shraddha Mane Prof.Supriya Mangale Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3

[3] Pedestrian Detection Based on YOLO Network Model Wenbo Lan ; Jianwu Dang ; Yangping Wang ; Song Wang 2018 IEEE International Conference on Mechatronics and Automation (ICMA)

[4] Bones detection in the pelvic area on the basis of YOLO neural network Zuzanna Krawczyk ; Jacek Starzyński 19th International Conference Computational Problems of Electrical Engineering

[5] Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO . 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)

[6] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun

[7] You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi in 2016

https://arxiv.org/abs/1506.02640