International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

Yield Estimation using faster R-CNN

[1]Vidhya Sagar, [2]Sailesh J.Jain and [2]Arjun P.
[1]Assistant Professor, [2]UG Scholar , Department of Computer Engineering and Science
SRM Institute of Science and Technology,Chennai, India
Department of Computer Engineering and Science
SRM Institute of Science and Technology Chennai, India.
Email: saileshlunkad127@gmail.com, arjunp111009@gmail.com

Abstract:
In today's world, computers exist in every domain thinkable. Improvements in computational technology has brought to life many wonderful solutions to problems that were once considered too unreal. One such fields is Computer Vision. The Human eye is "the most" delicate organ in the human body, which plays a major role in providing input to the brain. Understanding and replicating the way in which eyes work is the objective of Computer Vision. It has applications in various fields. Computer Vision applied to Agriculture sector can be a significant advantage to the farmers; it can help speed up the process of farming. In this paper, we talk about applying Object Detection to estimate the yield of tomatoes in a particular farm, using Convolutional Neural Networks. Computer vision is one of the main challenges in DeepLearning domain. Convolutional neural network currently dominate the computer vision landscape. Recently a CNN based model achieved State-of-the-art object detection These networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPnet and Fast R-CNN have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. A Region Proposal Network (RPN) shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultan eously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection. Further after we merge RPN and Fast R-CNN into a single network by sharing their convolutional features—using the recently popular terminol ogy of neural networks with “attention” mechanisms, the RPN component tells the unified network where to look while achieving state-of-the-art object detection accuracy. The whole Faster-RCNN model consists of Fast-RCNN model along with RPN to detect objects andbound the objects. The model is trained using our own dataset for “Object detection for Yield estimation using Faster-RCNN”.

Index terms: Yield estimation; machine learning technology; Deep Learning; Faster-RCNN;computervision convolutional net

1. INTRODUCTION

1.1. A. General:

The goal of this project is meet by combining various different technologies and approaches like Deep Lear- ning, Convolutional Neural Networks, Computer vision and different Regressors. Deep Learning is used for training the model at a deeper level than

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

compared to a standard neural approach. Convolutional neural net is used to process the input image and convert it to a feature map which makes the further processing effici- ent. Soft-max regressor makes classification of the objects possible and finally the Bounding-Box regressor predicts the boundary boxes around the detected objects. Combining all these modules the output obtained after this is output image having bounding boxes drawn around all the objects (Tomato here) with the yield estimation.

1.1.1.B. Objective:

With increase in global population, increasing agricultural producction is the key global food security goals. Here is why yield estimation comes in the picture. Estimating yield is a critical input in crop management to increase productivity. The main objective of this project is to utilise certain technology such as computer vision and machine learning to detect crops such as tomato from the given input image. To make this possible Faster R-CNN is used to allow the user to process an image and estimate the yield. The Faster R-CNN model Contains various modules which performs various tasks to make image processing possible.

2.II. RELATED WORK:

Object Proposals. There is a large literature on object proposal methods. Comprehensive surveys and com- parisons of object proposal methods can be found in. Widely used object proposal metho. ds include those based on grouping super-pixels (e.g.,Selective Search) and those based on sliding windows (e.g., objectness in wind- ows, EdgeBoxes). Object proposal methods were adopted as external modules independent of the detectors (e.g., Selective Search object detectors, R-CNN, and Fast R-CNN).

Deep Networks for Object Detection. The R-CNN method trains CNNs end-to-end to classify the proposal regions into object categories or background. R-CNN mainly plays as a classifier, and it does not predict object bounds (except for refining by bounding box regression). Its accuracy depends on the performance of the region proposal module (see comparisons in). Several papers have proposed ways of using deep networks for predicting object bounding boxes. In the OverFeat method, a fully-connected layer is trained to predict the box coordinates for the localization task that assumes a single object. The fully-connected layer is then turned into a convolutional layer for detecting multiple class-specific objects. The MultiBox methods generate region proposals from a network whose last fully-connected layer simultaneously predicts multiple class-agnostic boxes, generalizing the “single-box” fashion of OverFeat. These class-agnostic boxes are used as proposals for R-CNN. The Multi Box proposal network is applied on a single image crop or multiple large image crops (e.g., 224×224), in contrast to our fully convolutional scheme. MultiBox does not share features between the proposal and detection networks. We discuss OverFeat and MultiBox.

2. III. PROPOSED METHODOLOGY:

Object detection is one of the most fundamental problems in computer vision. Currently top leading results on PASCAL VOC, MS COCO object detection challenges all utilize Faster

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

R-CNN framework. The pioneering R-CNN decomposes object detection into two primary tasks. Firstly thousands of candidate object locations are generated by traditional object proposal methods (e.g., Selective Search). Then these candidates are classified and further refined by deep convolutional neural network (DCNN). Based on R-CNN, Faster R-CNN introduces region proposal network (RPN) to generate high quality proposals and adopts Fast R-CNN to perform RoI-wise classification and refinement. Both shared convolutional layers and end-to-end joint training push object detecti on to real-time and state-of-art accuracy. Faster R-CNN model basically consists of two parts. The first part is a deep fully convolutional network that proposes regions, and the second part is the Fast R-CNN detector that uses the proposed regions. The entire system is a single, unified network for object detection. Using the recently popular terminology of neural netw orks with 'attention' mechanisms, the RPN module tells the Fast R-CNN module where to look. To unify RPNs with Fast R-CNN object detection networks, we propose a training scheme that alternates between fine-tuning for the region proposal task and then fine-tuning for object detection, while keeping the proposals fixed. This scheme converges quickly and produces a unified network with convolutional features that are shared between both tasks.

3.1. A. System Architecture

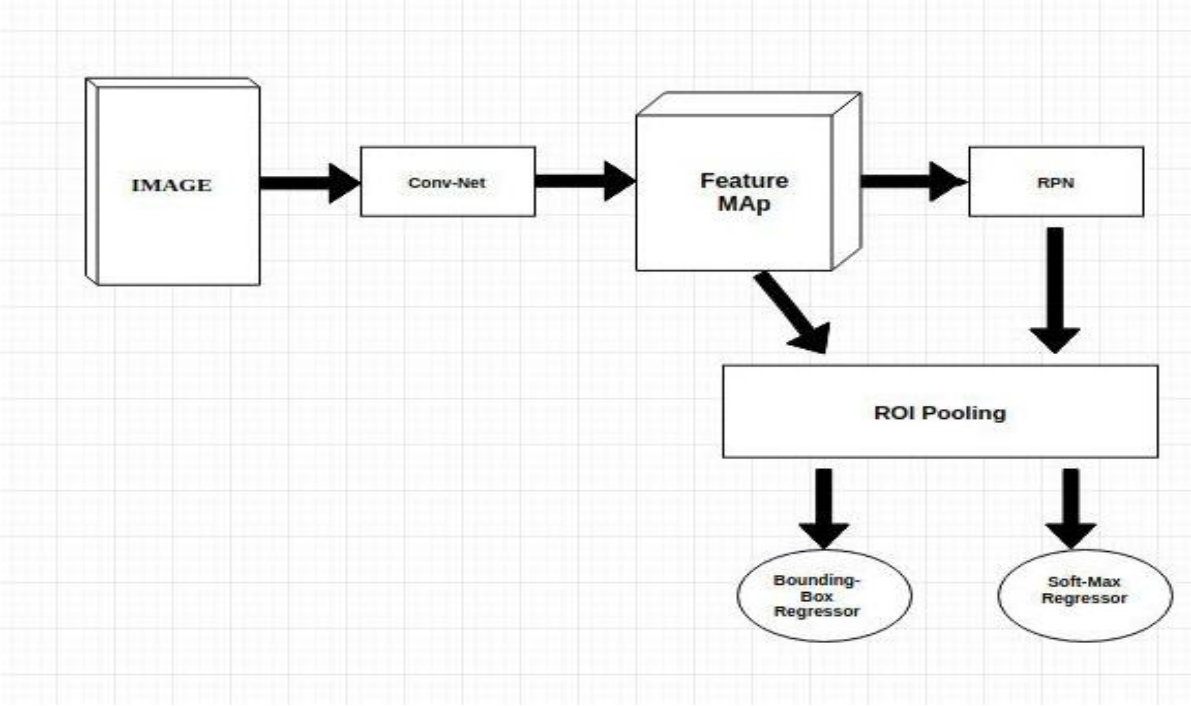


Figure 3. System Architecture

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

In Faster R-CNN, RPN is built on top of high-level convolutional feature layer (e.g., layer “conv5 3” in VGG-16)which is shared with Fast R-CNN. Specifically, a small network is slided over the feature layer. Each sli- ding window takes as input an 3 × 3 spatial window of the convolutional feature maps. Then it is mapped to a lower-dimensional feature (e.g., 512-d for VGG-16). After that, this feature is fed into two sibling fully-conn- ected layers, a box-regression layer (reg) and a box-classification layer (cls) to generate object proposals. Using multiple reg and cls layers, RPN can simultaneously predict multiple object proposals centered at each sliding window with different scales and aspect ratios. Then top scored proposals are selected. Finally, Fast R-CNN, also built on top of high-level convolutional features (e.g., “conv5 3” in VGG16), are followed to further classify and refine the proposals.

3.2B. TECHNICAL MODULE / IMPLEMENTATION:

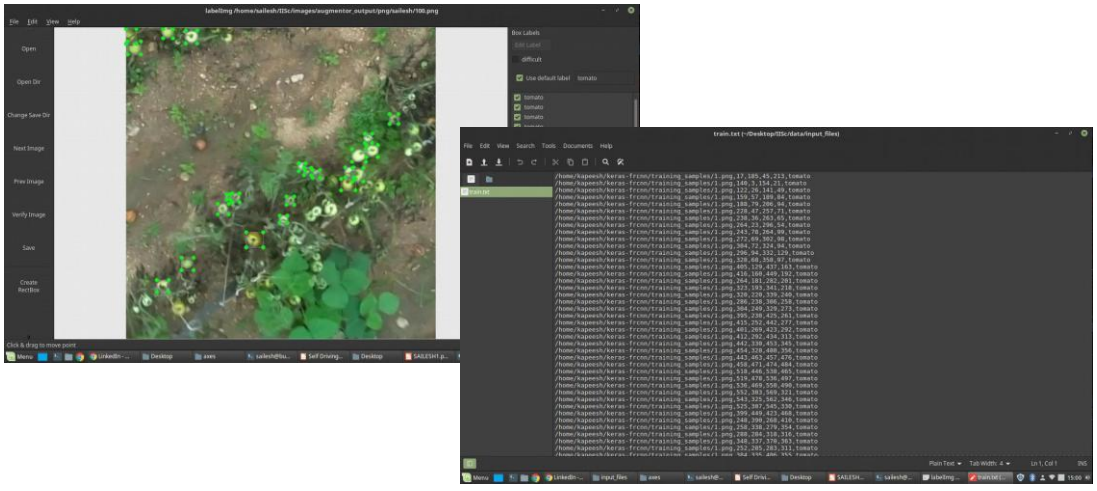
3.2.C. MODULE 1: Creating the Data set:

Here a UAV(Unmanned Aerial Vehicle) is used to capture some snapshots of the tomato field, which is used For obtaining the Data. The Data obtained from the UAV isof resolution 1366*768, which is then Augmented Using a Python library tool Augmentor to create a fixed sized sample images of resolution 600*600. By Aug- mentation we mean to perform certain operation on this images such as right flip, left flip, zoom in and many More ofa respective probability. This images are then Annotated using a LabelImg tool. The annotated data in The LabelImg tool isof pascal-VOC format, so we have to convert this annotated file into respective CSV files. After this we need to put all the data from the CSV files into a single txt file. This txt file acts as a dataset to our mdel, the data in this txt file should be of a particular format as given below.

/home/kapeesh/keras-frcnn/training_samples/1.png,17,185,45,213,tomato

Challenges:

- Annotation has to be done with 100% accuracy
- More Data set is required for training



International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

3.2. B. MODULE 2: CNN Model(VGG-16 / Resnet) :

In this the images along with the annotation is passed through the CNN model. Here the CNN model is also termed as a 'Feature extractor' for our model. The model then trains itself to generate a feature map. We can also trainour CNN-model by loading the predefined weights of the existing dataset say Image-net and pascal VOC. Here feature map is the output of the last conv layer of our CNN-mode, without passing through the last Fully Connected layer. This feature map is then passed to the Region Proposed Network and ROI pooling of our network. For our VGG-16 network the feature map is of size 40*60*512.
Challenges:

- Training the models requires a highly configured system
- Training takes up a lot of time

3.2. C. MODULE 3: Region Proposed Network(RPN):

Here the feature map generated from our last module is passsed through this RPN model. RPN is a small network where a 3*3kernel slides over a feature map. Simultaneously for each kernel position on the feature map there are k no.of anchors generated. Each anchors in all the position are then classified as positive or negative based on the IOU value.This K anchors are of different scales and ratios based on our requirement. As assumed above if the feature map obtained from VGG-16 network is of size 40*60*512,then this gives us a total of 2400 position. Using a 3*3 sliding window in this 2400 positions gives us a total of 2400*K anchors generated. If the no.of anchors are K=9 then there will be 21000 anchors generated, where cross boundary and ovberfitting anchors are eliminated. This in last gives us a total of 2000 approx anchors after which a minibatch of size 512-d is genera ted containing 256 positive anchors and 256 negative anchors. Output of this RPN model is region proposals which is then passed to the ROI pooling part.

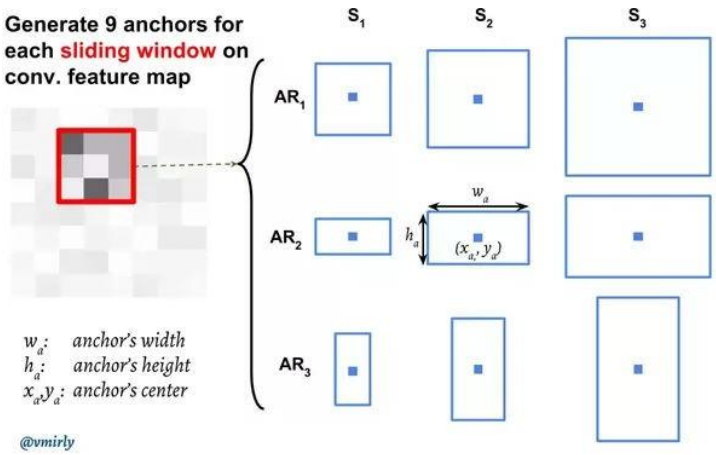


Figure 2: Sliding window technique

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

3.2. D.Module 4: Bounding Box Regression :

The Bounding Box Regressors are essential because the initial region proposals might not fully coincide with the region that is indicated by the learned features of the Convolutional Neural Network. It is a kind of a refinement step.Therefore,based on the weights of the classifier (eg. Neural Networks, SVMs), the region proposals are regressed. Keep in mind that that the features used for the regression are the features obtained at the end of the final pooling layer.This kind of regression gives a better estimate of the object position than our simple Proposal Generators, since it is based on the features generated by the Feature extractor. The output of this regressor deter mines a predicted bounding box, consisting of four values(x,y,h,w). Where x and y are the centre coordinates of the bounding box and h and w are the height and width of the bounding box.

3.2. F.Module 5: Soft-max Regression :

Soft-max Regressor is used in hand to hand with Bounding box Regressor. Soft-max regressor is used to classify between objects. It fills up the gap of question “what the object is” in the output. If the object is positive the classifier value of its becomes 1 and 0 if the object is negative. In our case it states 1 for the object with class tomato and 0 for the object with the class background.

Figure 3.Bounding box and Soft-max Regressor example

IV. CONCLUSION:

We have presented RPNs for efficient and accurate region proposal generation. By sharing convolutional features with the down-stream detection network, the region proposal step is nearly cost-free. Our method enables a unified, deep-learning-based object detection system to

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 3, No. 1, March - May, 2018
Pages 110-116 | Cosmos Impact Factor (Germany): 5.195
Received: 28.03.2018 Published : 23.04.2018

run at near real-time frame rates. The learned RPN also improves region proposal quality and thus the overall object detection accuracy. This thus helps us in agricultural field for yield estimation using object detection thus increasing productivity.

REFERENCES

[1] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders,"Selective search for object recognition," International Journal of Computer Vision (IJCV), 2013.
[2] https://datascience.stackexchange.com/questions/27277/faster-rcnn-how-anchor-work-with-slider-in-rpn- layer
[3] https://github.com/rbgirshick/py-faster-rcnn
[4] http://blog.csdn.net/u014365862/article/details/77887230