Umer Zeb Khan

14031191

Final Year Mid-Project Report

13th January, 2019

Cataloguing trees from Mass plantation aerial imagery

Dr. Junaid Akhtar

Dear fellows in science, this digitally synthesized piece of document shall take you through my quest to train a custom machine learning model to detect trees from aerial images captured by an unmanned aerial vehicle. Before you read any further, please remember this verbatim is not definite and shall be subject to constant change until the final report due in April, 2019 including content, style and format.

A simple organizational structure has been adopted to enhance ease of writing and the flow of creative juices. This happens to be the cover page, followed by a formal introduction and pretext to the matter under scrutiny. Three notions have been adopted in mind while writing this report, specifically: law of contradiction, law of excluded middle and the principle of identity. Critique and suggestions are highly requested. Personal reflections are inclusive of this writing until they are segmented as portfolio. Progress made so far and what remains (tenatatively) is highlighted.

1. Introduction

Pakistan is one of the most prone countries to climate change and global warming, as assessed by several environmental agencies throughout the world. Leaders and people in position of authority usually announce large scale plantation drives. The move certainly earns them the goodwill of many people but independent verification is difficult due to many factors, mostly due to ineffective plantation techniques such as spraying seeds over large tracts of land from aerial vehicles i.e. helicopters and lack of post plantation care of trees for their fragile years of initial growth.

Prime Minister Imran Khan has been elected in 2018 whereby one of his electoral promises has been to plant 10 billion trees. In a recent development, the premier's advisor of climate change, Amin Aslam has invited third parties to enhance transparency in the Clean Green Pakistan initiative by evaluating and monitoring the mass plantation (Anon, 2019).

2. What is happening & why?

The problem statement is: detection of trees in order to determine their count using a drone equipped with camera. The objective of this project is to facilitate the tracking of trees over a period of time and confirm the veracity of the claims made by philanthropists. After all the motive is not to verify the claims of mortals, but rather to devise a mechanism to seamlessly monitor tree plantations for the valuable part they play in our ecosystem and in order to fight the very real and potent threat of global warming.

The aim is to use computer vision techniques and machine learning to teach a neural network to detect trees in a network. Python, Matlab, Weka with several dependencies such as pytorch and advanced computer vision libraries are utilized to process the images and tweak our model along with gaining additional insight into existing methods and techniques regarding machine learning. As training and testing is a computationally expensive process, our hardware requirements are a GPU equipped computer and a drone, in our case, DJI's Tello fitted with a 5 megapixels HD camera to collect our customized data.

3. Document's Flow

The next segment of the report sheds light on various similar technologies and techniques being employed by scientists and engineers to solve applied computer vision problems and how they converge with the defined scope of my project.

Literature review is closely followed by design and analysis section which describes in depth what has been achieved so far pertaining to the project and what needs to be done. Tools and experimentation conducted are discussed.

The last section is the conclusion of this mid year report and my attempt to make sense of the (partially) understood abundance of information absorbed. The end consists of a bibliography in Harvard referencing style.

4. The Deal so far

This chapter deals with the literature consulted in the pursuit of my own agenda for counting trees. An overview is given regarding the imaging techniques available in the market to obtain different types of parameters and how they can be processed. Moreover the crucial impact of the data used for model training and the varying categories of prepared datasets influence of detection and accuracy is discussed.

The current state of the art work done within the domain of computer vision from aerial imagery can be divided primarily on the basis of their input: LIDAR, NIR and/or RGB. The former two are remote sensing technologies that provide radiometric and geometric information regarding the Earth’s surface (Yang, L. et al., 2009) while the latter provides limited information depending on the algorithm used and the manipulation of data set.

Recent advances in computer vision are attributed to the availability of large and small datasets alike, allowing researchers and scientists to not only tweak their algorithms and refine their models but to make innovations possible by the availability of high quality datasets.. Multiple years of hard work have resulted in exhaustive data sets for training and validation, seven years to be precise in the case of Pascal VOC (Everingham, M. et al., 2009) consisting of 20 object classes of more than 11, 000 images with a little more than 27, 000 region of interest annotated images and about 7000 segmentations. Due to the versatility of the data set, it has been used commonly to solve classical computer vision problems such as mentioned in (Mottaghi, R.

et al., 2014) that the context is usually an important cue to weigh in before making a prediction regarding the object i.e. a kid of 4 knows the deformed entity on the sofa is not a tiger but rather a cat, similarly the small object on the table is a fork, not an elephant's tail.

Another comprehensive dataset available is ImageNet (Deng, J. et al., 2009), by far the biggest and largest dataset, consisting of 2200 objects within 15 million images, with average images in the range of 500-1000 belonging to each class of available objects, based on the hierarchical structure of WordNet inclusive of only the nouns (Fellbaum, C., 1998). The dataset has been used extensively in the field of image processing and computer vision due to its scalability, abundance of labelled images and hierarchical structure.

OverFeat is an intelligent feature extractor developed by (Sermanet, P., 2013) using their best performing model that provides an integrated framework using convolutional neural networks to classify, detect and localize images for an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). A subset of ImageNet with 1000 object and roughly the same number of images in each class is used for the competition. All the three tasks are fundamental in computer vision and consequently recognition of interest regions that are marked using bounding boxes, all of these processes taking place simultaneously using just a single convolutional network and a common feature learning base.

Any mention of datasets shall remain inconclusive without the name of MS COCO, short for Microsoft's Common Objects in Context that has 91 categories of object with about 328, 000 images with 2.5 million instances (Lin, T.-Y. et al., 2014). The state of the art dataset put the agenda of object recognition in the broader context of scene comprehension, something a four year does intuitively but still poses a huge challenge for computers.

Applications of image procession in the field of agriculture are many including fruit grading, plant health, plant count, weed detection and imaging techniques (Vibhute, A. & Bodhe, S.K., 2012). Imaging techniques denote the type of images obtained for analysis, using the change in the wavelength of light as a primary parameter, such as IR imaging, Radio imaging, X-ray imaging, UV imaging, imaging in the visual band ect (Gonzalez, R.C. & Woods, R.E., 1993). Electromagnetic radiations are employed to infer geo-biophysical characteristics and surface features of the earth for identification, the process is known as Remote Sensing. Further applications are determination of water stress that helps in irrigation, management of floods, plant disease monitoring, environment assessment, mapping of manmade and natural plantation etc (Hänsch, R., Schulz, K. & Sörgel, U., 2018).

Palm trees are an important economic crop of Malaysia. A deep learning based solution is proposed for detecting palm trees, their count, estimating their yield and enhancing their productivity (Li, W. et al., 2016). Traditional methods for crown detection of trees, namely, template matching, local maximum filter and artificial neural network have also been implemented to draw a comparison between pre-existing models and the custom deep convolutional neural network. It's a no brainer, a deep learning model specifically tailored to detect and count palm trees fared the best.

Most relevant to my area of interest is the work of Yang, L. et al. that proposes a mechanism for the detection of trees from aerial imagery. A pixel level classifier to applied to an image to distinguish between tree and non-tree regions, being further refined by a partitioning algorithm. Once refining is done, a clean image segmentation is obtained for trees and non-trees followed by template matching. In the end, greedy selection is done for the location of individual tree crowns (Yang, L. et al., 2009). The proposed solution works on pure images only while delivering satisfiable performance on large scale tree plantations.

5. What have I done & what remain yet

I have been reading from various sources and working with introductory machine learning software, algorithms and techniques. One of the first tool recommended for amateur machine learning enthusiasts is a GUI based application called WEKA. Although it’s a data mining software, proficient in numerical analysis and text recognition, yet it provides a comprehensive introduction to various machine learning techniques and algorithms in a graphical user interface that makes the fiddling with parameters intuitive and strong visualization make it possible to see the cause-effect relationship between two or more variables/setting.

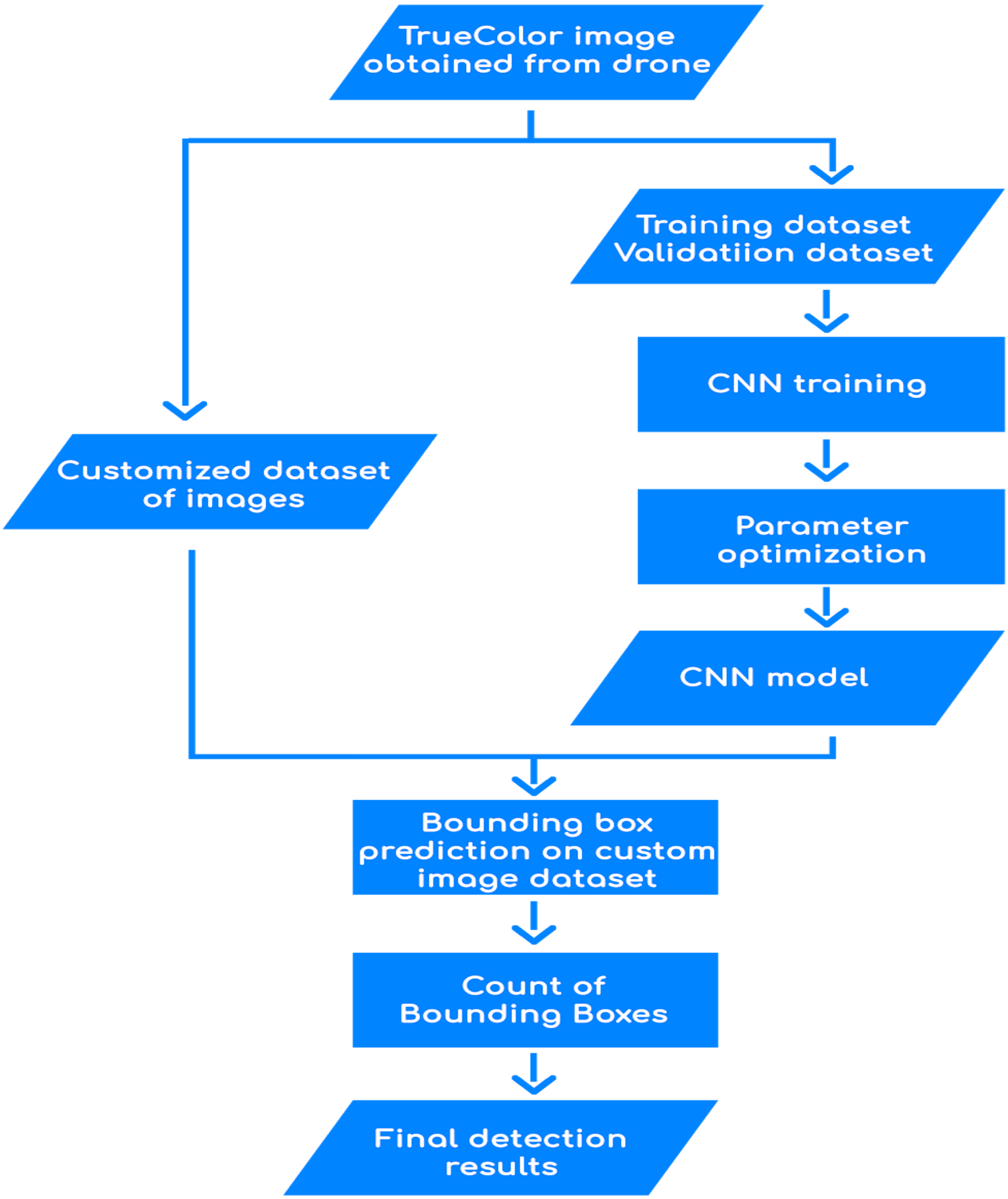
Another very interesting project I undertook to enhance my understanding of neural networks and the technology I intend to use for my custom problem is the implementation of YOLO (You Only Look Once) on my webcam. The process started with downloading and installing Darknet with CUDA and OpenCV (Redmon, J., 2017). With one simple command, the webcam turned into a real time object detecting device, which was quite fascinating to watch. Throughout the webcam feed, the frames per second are displayed along with bounding boxes over the types of objects detected. In my case, it mostly was my face which resulted in the prediction of person class.

For the purpose of prototype, I was able to find a deep CNN model used with the CIFAR-100 dataset, which includes the superclass of trees able to detect six different types. For a start, it’s a great model to learn about how the entire mechanism works. The difference between classification, detection, localization, tracking and template matching was learnt.

Furthermore as my intended work is to detect trees, the model does just that along with many other objects detection. I tweaked with the dataset to include only the tree superclass images, which are roughly 600 in number, but to my dismay the resulting performance although quick, was quite terrible at detecting trees.

In the future, I will have to curate my personalized dataset of images captured via an unmanned aerial vehicle. Along with that, I shall be training my own custom detection model exclusively for trees, mainly following the methods and techniques discussed by Yang, L. et al.,

2009. A flowchart is given below to illustrate the proposed solution.



6. What does all this mean

The entire learning experience has been great. The observation pertaining my report's lack of structure and ambiguity in proposed solution is of merit enough to be disproved. If answers were known and mapping of intricate real world problems onto a binary scale seamless, I would not have tried to teach a machine how to learn to solve a problem. Instead I would have taught it a certain set of rules to abide by, which shall always result in a solution.

Hopefully with the provision of more computational resources, my ability to fiddle with with the neural network architecture will be enhanced. It'd be interesting to see the performance of neural network after introduction of customized dataset of aerial imagery.

Bibliography

Deng, J. et al., 2009. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition.

Everingham, M. et al., 2009. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), pp.303–338.

Fellbaum, C., 1998. WordNet: an electrical lexical database, The MIT Press.

Gonzalez, R.C. & Woods, R.E., 1993. Digital image processing, Addison-Wesley.

Hänsch, R., Schulz, K. & Sörgel, U., 2018. Machine learning methods for remote sensing applications: an overview. Earth Resources and Environmental Remote Sensing/GIS Applications IX.

Li, W. et al., 2016. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. Remote Sensing, 9(1), p.22.

Lin, T.-Y. et al., 2014. Microsoft COCO: Common Objects in Context. Computer Vision – ECCV 2014 Lecture Notes in Computer Science, pp.740–755.

Mottaghi, R. et al., 2014. The Role of Context for Object Detection and Semantic Segmentation in the Wild. 2014 IEEE Conference on Computer Vision and Pattern Recognition.

Anon, 2019. PM Imran targets 142.6 million trees in fresh plantation drive. The Express Tribune. Available at: https://tribune.com.pk/story/1883972/1-pm-imran-targets-142-6-million-trees-fresh-plantation-drive/.

Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R. & LeCun, Y. (2013), 'OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks.', CoRR abs/1312.6229 .

Vibhute, A. & Bodhe, S.K., 2012. Applications of Image Processing in Agriculture: A Survey. International Journal of Computer Applications, 52(2), pp.34–40.

Yang, L. et al., 2009. Tree detection from aerial imagery. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09.

Redmon, J., 2017 , YOLO: Real-Time Object Detection. Available at: https://pjreddie.com/darknet/yolo/.