Deep Learning Based Tree Detection and Counting for Remote Sensing Images

Nisarg Vora[1], Devam Dave[2], and Kathan Shah[2]

[1] Birla Institute of Technology and Science, Pilani, India
[2] Pandit Deendayal Energy University, Gandhinagar, India

Abstract. Trees are an essential part of the environment as they improve the quality of air around us by providing oxygen, regulate climate and preserve the ecosystem, land and water etc. Some trees are important economic crops and their count in a specific region is an important factor in the prediction of the yield of the product which they give. The number of trees in a particular region helps to monitor the growing situation of trees. In this project, tree detection has been done using a deep learning based framework and the counting of these trees has been done using remote sensing [12] high-resolution images for two regions in the state of Uttarakhand, India. The trees in our areas of study are congested, often leading to an overlap of crowns. Two multi-spectral images have been provided for the project. The first image has four channels namely Red, Green, Blue (RGB) and Near-Infrared (NIR). For the first image provided, we have used a variety of manually interpreted samples for the training as well as the optimization of the convolutional neural network (CNN). Thereafter, using the sliding window technique, we have carried out the prediction of the labels of the samples in the image dataset. The proposed model, which shows a weighted accuracy of over 98% during training and validation. Additionally, the text analyzes the results obtained in case the near-infrared band is removed from this image with four channels (i.e in second image).

Keywords: Tree Detection · Convolutional Neural Networks · Remote Sensing · U-Net Architecture · Deep Learning

1 Introduction

1.1 Remote Sensing

Operational management and control of the green assets of nature, such as trees and green cover entails timely, well-run and reliable information about its current status and developments, especially development of irrigation processes for maximizing productivity. Tree count management is essential for sustenance of ecological biodiversity and conservation stability. A well-organized tree inventory of the forested areas can enable us to investigate the possible causes of the decline of forests and/or vegetation in that particular area and assist in adequate decision-making.

Traditional methods for counting trees are labour-intensive, time consuming, pricey and not pertinent to large areas. Remote sensing plays an important part in a variety of tree-detection studies. Various methods of obtaining remotely sensed data include LiDAR [7], Satellite Images [1], UAV (Unmanned Aerial Vehicles) [4] and Terrestrial Photogrammetry [16], [5].

The tree crown detection research done till date includes traditional as well as modern methods related to computer vision. The traditional methods of tree-detection include the following:

(a) Tree detection and delineation algorithm based on the "local maximum filter and the analysis of local transects extending outward from a potential tree apex."[15]
(b) Tree extraction and counting from high-resolution imagery data, including spectral and texture analysis, edge enhancement, segmentation and morphological analysis.[17]

Machine learning (ML) based methods include the following:

(a) Use of a SIFT (scale-invariant feature transform) and a supervised extreme learning machine classifier for palm tree detection from UAV images. [13]
(b) Use of a local maximum detection algorithm on the spatial distribution of standardized features and circular autocorrelation of the polar shape matrix representation of an image as the shape feature to detect palm trees. [14]

Convolutional neural network (CNN) methods include the following:

(a) Chen et al. [2] applied the stacked auto-encoder method to hyper spectral remote sensing image classification and introduced the concept of deep learning in this field for the first time.
(b) Li et al. [11] built a classification framework for "large scale remote sensing image processing and African land cover mapping" based on the stacked auto-encoder.
(c) Use of a U-Net type of convolutional neural network for the detection of large scale palm trees from high resolution satellite images. [6]
(d) Li et al. [10] introduced the deep learning based method to oil palm tree detection for the first time.

In spite of the higher accuracy, the research behind these methods has been inconclusive and it has been tough to replicate these results. We believe that the performance of a few of these methods is bound to decrease when detecting trees in a few regions of our study area. For example, template matching fails where the crowns of trees often overlap. The resolution, too, plays an important role here, and owing to the resolution of the images provided, we cannot use methods specifically designed for very high resolution satellite images.

In this study, we have proposed another such deep learning based algorithm for carrying out tree detection, similar to the method used by Li et al. [10] for oil palm tree detection. The difference here lies in the fact that the research involved the use of three out of the four bands, while our method involves all the four bands for training the model in the first case. Tree detection has been done using a deep learning based framework (CNN) and the counting of these trees has been done using remote sensing high-resolution images from the state of Uttarakhand, India. The areas we have used have trees that are densely packed and thus, often lap over each other. The proposed method has been discussed further in Section 5.

2 Concepts Used

2.1 Convolutional Neural Network (CNN/ConvNet)[?, 9]

A neural network, or a connectionist system, is a set of algorithms, modelled roughly around the human brain. CNN is one such deep learning algorithm which takes inputs in the form of images and assigns certain parameters to various aspects of the image in order to compare one image to another image. ConvNet is one of the main categories in neural network and is used to perform image recognition, image classification, object detection, face recognition, etc.

2.2 Chlorophyll and Near Infrared Radiation

Chlorophyll, a pigment found in plants and algae, absorbs visible light and reflects near infrared. The health of a tree or plant is directly correlated to the amount of near infrared reflected by it. Xiao et al. [20] and S. Jacquemond et al. [8] have provided valuable insight regarding the absorption spectrum of leaves. This near infrared radiation that is reflected by chlorophyll can be detected by satellites, thus enabling the researchers to analyse the flora from space.

2.3 Sliding Window Technique

A rectangular region of fixed height and width that slides across an image is known as a sliding window. The sliding window technique is an image processing technique used in object classification.

3 Study Area and Datasets

In this project, two satellite images are used. The first multi-spectral (MS) image has four bands, with 1.6-m spatial resolution. On the other hand, the second multi-spectral image has three bands, with 0.6-m spatial resolution. The aforementioned image with three bands is used at a much later stage, primarily to figure out one of the limitations of the deep learning based framework built earlier using the first image. Henceforth, the image with four channels is used for the majority of the discussion.

The training and validation datasets for further image processing and analysis have been obtained through manual interpretation, using QGIS software (version

3.6.3, Quantum Geographic Information System, 'Noosa', released on 22" Feb, 2019).

The study areas of this project are located in Uttarakhand, India. The first satellite image corresponds to a region in Dehradun, while the second image corresponds to a region in Rishikesh (see Figure 1). For each multi-spectral image, the manually interpreted samples have been collected from throughout the image and used as an input for the respective model for training and validation. For the first image, an additional dataset, comprising 50 cropped images not included in the training or validation dataset has been utilized to further evaluate the accuracy of the proposed method by carrying out a comparison of the ground truth labels with the predicted labels.

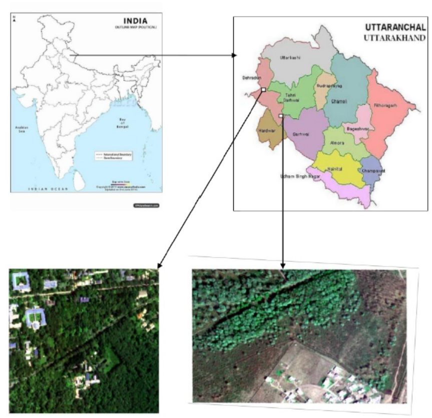

Fig. 1: 4- band multi-spectral image(left), multi-spectral image with three channels(right)

4 Methodology

Overview In Figure 2, the proposed method has been represented in the form of a flowchart. Primarily, using the Tensorflow framework, the convolutional neural network (CNN) was constructed. We then used a variety of training samples to carry out training of the CNN, and estimated the weighted classification accuracy from a few validation and test samples obtained independently from the training samples. The major parameters of the CNN were kept constant. Hyper parameters such as learning rate and batch size were also fixed. However, the number of epochs was adjusted continuously, until we found the value of the number of epochs for which the weighted accuracy was the maximum on the validation and test samples. The validation accuracy was calculated at each epoch to ensure that we weren't overfitting with the model. We achieved the best

CNN model post hyper parameter tuning and conserved it for future use. Next, we obtained the image dataset using the sliding window technique (window size: 14x14, sliding step: 14 pixels). The window size and the step size were decided keeping in mind the size of the cropped images of trees and the resolution of the original image. Among the various CNN models we obtained and constructed earlier, we have used the one with the best accuracy and performance for prediction. In this manner, the number of trees in the image was counted and final tree detection results were obtained.

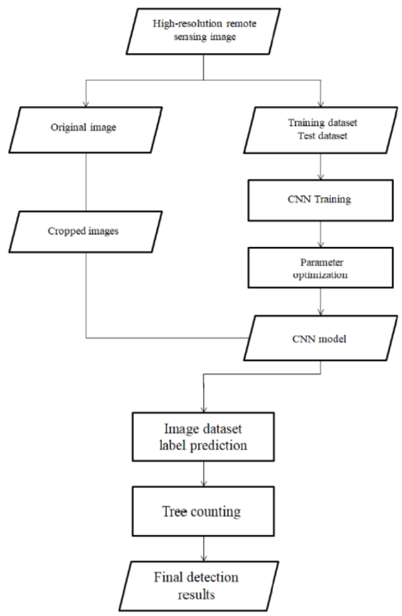

Fig. 2: The flowchart of our proposed method

CNN Training and Parameter Optimization The CNN we have used in our study draws its inspiration from VGG-16[18] network architecture. This "mini" VGGNet is constructed of a fully connected classifier (Figure 3) to go along with four convolutional blocks. Every convolutional block consists of a convolutional layer and a max pooling layer. A flattened layer, a dropout layer and two dense layers together make up the fully connected classifier. The set of all feature maps is used as an input to the fully connected classifier. We have used ReLU as the activation function in our convolutional neural network (CNN) for every layer except the last one. In the last layer, we have used a sigmoid activation function.

In this project, 150 tree samples and 85 background samples were manually interpreted from our study area. Then, we randomly select 148 of the aforementioned samples to be used for training the CNN and another 37 samples as its validation dataset and the remaining 50 samples for testing. An instance of the image dataset with a single tree situated at the center is labelled as a ”tree”. Cropping has been done keeping in mind that every cropped image contained at the most one tree. Each sample has different size, but has four bands (Red, Green, Blue and Near-Infrared).

The cropped images are fed into the model after resizing them to 32 x 32, without changing the channels. The number of epochs is adjusted continuously while keeping all the other parameters fixed, until we find the value of the number of epochs for which the weighted classification accuracy is the highest. The prime CNN model thus obtained is used in the next process of the prediction of the labels of images from the dataset. Table 1 illustrates our model.

Table 1: Results.

Block Name	Layer Name	Number of Filters/Hidden Units	Filter Size
First Convolutional Block	Convolutional Layer Max Pooling Layer	NIL	2x2
Decision Trees	98.18%	82.71%	
Support Vector Machines	89.63%	87.37%	
CNN (U-Net architecture)	95.06%	89.31%	

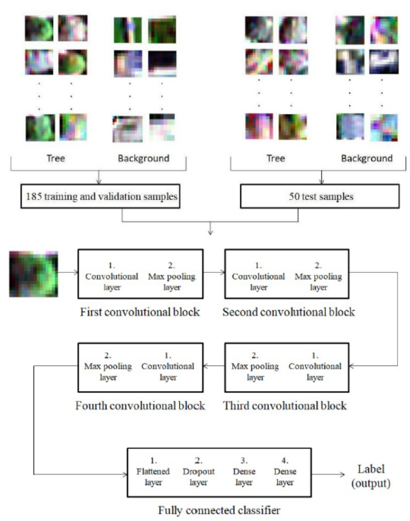

Fig. 3: The structure of the convolutional neural network (CNN)

Label Prediction and Tree Counting The images for prediction of labels are collected, one by one, by using the sliding window technique (window size: 14x14, sliding step: 14 pixels). We have taken into consideration the mean size of the cropped images of trees and the resolution of the original image. The sliding step should be chosen carefully, as it has a major impact on the final results for tree detection. Too large, and we see many tree samples not being spotted. Too small, and a single tree sample may be spotted more than one time. As each image is collected, it is resized to 32 x 32 pixels and used as an input for the CNN model to predict its label, and if it is a tree, a rectangle is drawn around the 14 x 14 image. In this manner, the number of trees in the original multi-spectral image is obtained.

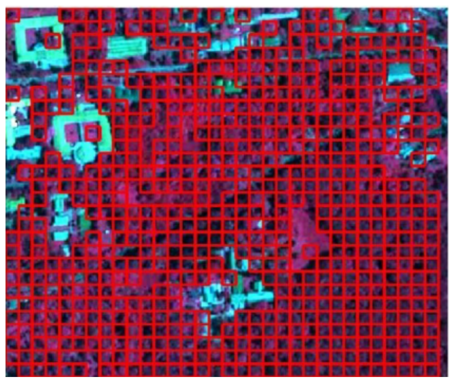

Fig. 4: Tree detection result obtained after applying sliding window technique. Each red rectangle indicates the location of a single detected tree.

5 Results

Weighted Classification Accuracy and Parameter Optimization In this project, the train accuracy, validation accuracy and test accuracy were obtained using 148 training samples, 37 validation samples and 50 test samples. It is important to note here that the three datasets are disjoint. These accuracies were used to obtain the weighted classification accuracy using the following formula:

$$acc_w = \frac{n_{train} * acc_{train} + n_{val} * acc_{val} + n_{test} * acc_{test}}{n_{train} + n_{val} + n_{test}} \tag{1}$$

In this formula, acc_w is the weighted classification accuracy of the model, n_{train} is the number of training samples (=148), n_{val} is the number of validation samples (=37), n_{test} is the number of test samples (=50), acc_{train} is the train accuracy, acc_{val} is the validation accuracy, and acc_{test} is the test accuracy

The performance of the model in classification can potentially be influenced by various parameters, like the size of the max-pooling kernel and the convolutional kernel, the amount of feature detectors in every convolutional layer, the number of hidden units present in the fully connected classifier, etc. In our CNN model, for the first three convolutional blocks, the size of the convolutional kernel is three and the size of the max-pooling kernel is two. For the fourth convolutional block, the sizes are two and one respectively.

The batch size has been chosen to be 16, while the learning rate, , has been kept fixed at 1.0 x 10^{-4}. A dropout layer has been added in the fully connected classifier for regularization. The quantity of filters in each layer and the amount of hidden units in the fully connected classifier are also fixed. Through "hit-and-trial", we adjusted the number of epochs to obtain the maximum weighted classification accuracy of 98.39% after 30 epochs. In this case, acc_{train}=98.12%, acc_{val}=100% and acc_{test}=98%.

5.1 Tree Detection and Counting

The tree detection result for the original multi-spectral image is already shown in Figure 4. With the exception of a few false positives and false negatives, the result obtained looks pretty accurate to the naked eyes, considering the 1.6-m spatial resolution of the satellite image. The model predicts a total of 603 trees in the image.

6 Limitations

There are a few limitations to this project, which are as follows:

(a) The training, validation and test datasets were generated by us; no standard dataset was picked for training the model. Usually, the datasets available online for object detection are large-scale and quite rich, resulting in better subsequent training of the model and improving its reliability. In our case, though, we couldn't find any dataset which satisfied the requirements, resolution being the key here. As a result, we were compelled to generate the dataset on my own by manually cropping the images of trees and background from the original images.

(b) A deep learning model is very specific with regard to the training dataset, as discussed earlier. The small training datasets further added to the specificity of the models created.

(c) The size of the sliding window is fixed. However, in a forest, trees may have, and usually have, different widths. The window size, coupled with the size of the sliding step, decides the final tree detection results. However, the resolution of the images could not allow us to choose the appropriate window size and the step size for both the images; their values were decided considering the mean size of the cropped images of the trees for each of the cases. They do not take into account the following cases:

– Multiple trees in a single window and hence being counted as one tree.
– A single tree lying on the border of two or more squares and hence being detected and counted more than once.

7 Conclusions

In this project, we have administered tree detection using a deep learning based framework and the counting has been done from two satellite images, one with a 1.6- m resolution and four channels, and the other with a 0.6-m resolution and three bands.Two models were designed and trained independently of each other, and best CNN models were chosen for obtaining the tree count in the respective images. The accuracy and performance of the aforementioned method is given by the preliminary results.The text discusses the need to design the second model, and the importance of the near infrared channel for obtaining more accurate results.It must be admitted, though, that the tree counts obtained are approximations, highly governed by the sizes of the manually cropped images of trees in each case considered separately. Furthermore, the improvements are highly restricted due to inadequate resolution of the images, particularly of the four-band image.

8 Direction for Future Work

It is evident that background suppression and/or foreground detection may prove handy along with IHS (Intensity Hue Saturation)[19] fusion method for appropriate tree detection, but this would involve an absolute change in the image processing technique to be followed for detection and delineation. Wagner et al. proposed an altogether different methodology [3]. They used a rolling ball method along with morphological mathematical operations for tree crown detection and delineation, though it works on only grayscale images. This requires converting RGB images to HSL (Hue Saturation Lightness) and using only the "L" part. This method, coupled wisely with our proposed method, can yield more accurate tree detection results.

References

1. A. Carleer and E. Wolff. Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering Remote Sensing*, 70(1):135–140, 2004.
2. Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2014.
3. Fabien Hubert Wagner et al. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:362–377, 2018. SI: Latin America Issue.

4. Matheus Pinheiro Ferreira, Danilo Roberti Alves de Almeida, Daniel de Almeida Papa, Juliano Baldez Silva Minervino, Hudson Franklin Pessoa Veras, Arthur Formighieri, Caio Alexandre Nascimento Santos, Marcio Aurélio Dantas Ferreira, Evandro Orfanó Figueiredo, and Evandro José Linhares Ferreira. Individual tree detection and species classification of amazonian palms using uav images and deep learning. *Forest Ecology and Management*, 475:118397, 2020.
5. Mona Forsman, Niclas Börlin, and Johan Holmgren. Estimation of tree stem attributes using terrestrial photogrammetry with a camera rig. *Forests*, 7(3), 2016.
6. Maximilian Freudenberg, Nils Nölke, Alejandro Agostini, Kira Urban, Florentin Wörgötter, and Christoph Kleinn. Large scale palm tree detection in high resolution satellite images using u-net. *Remote Sensing*, 11(3), 2019.
7. Haiyan Guan, Yongtao Yu, Zheng Ji, Jonathan Li, and Qi Zhang. Deep learning-based tree classification using mobile lidar data. *Remote Sensing Letters*, 6(11):864–873, 2015.
8. S. Jacquemoud and F. Baret. Prospect: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(2):75–91, 1990.
9. Fei-Fei Li, Justin Johnson, and Serena Yeung. cs231n - lecture slides, 2017.
10. Weijia Li, Haohuan Fu, Le Yu, and Arthur Cracknell. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing*, 9(1), 2017.
11. Weijia Li, Haohuan Fu, Le Yu, Peng Gong, Duole Feng, Congcong Li, and Nicholas Clinton. Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of african land-cover mapping. *International Journal of Remote Sensing*, 37(23):5632–5646, 2016.
12. Chor Pang Lo. Applied remote sensing. 1986.
13. S. Malek, Y. Bazi, N. Alajlan, H. AlHichri, and F. Melgani. Efficient framework for palm tree detection in uav images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4692–4703, 2014.
14. Abhishek Manandhar, Ludwig Hoegner, and Uwe Stilla. Palm tree detection using circular autocorrelation of polar shape matrix. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3:465–472, 06 2016.
15. D.A Pouliot, D.J King, F.W Bell, and D.G Pitt. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, 82(2):322–334, 2002.
16. John Roberts, Andrew Koeser, Amr Abd-Elrahman, Benjamin Wilkinson, Gail Hansen, Shawn Landry, and Ali Perez. Mobile terrestrial photogrammetry for street tree mapping and measurements. *Forests*, 10(8), 2019.
17. Helmi Z. M. Shafri, Nasrulhapiza Hamdan, and M. Iqbal Saripan. Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery. *International Journal of Remote Sensing*, 32(8):2095–2115, 2011.
18. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
19. Harpreet Singh, Prabhakar Alok Verma, and Sameer Saran. Qualitative assessment of geostatistical and non-geostatistical fusion techniques: a case study on landsat 8 images. *Spatial Information Research*, 27(3):275–283, 2019.
20. Yi Xiao, Danny Tholen, and Xin-Guang Zhu. The influence of leaf anatomy on the internal light environment and photosynthetic electron transport rate: exploration with a new leaf ray tracing model. *Journal of experimental botany*, 67(21):6021–6035, 2016.