Ted: Tree delineation with reduced dimensions using entropy and deep learning

RN Anjani ANJANI_RN@NRSC.GOV.IN
CH Sarvani SARVANI_CH@NRSC.GOV.IN
K Kalyan Deep KALYANDEEP_K@NRSC.GOV.IN
P Aravinda Kumar ARAVINDAKUMAR_P@NRSC.GOV.IN
Sitiraju Srinivasa Rao SSRAO@NRSC.GOV.IN
National Remote Sensing Center, India

Abstract

Trees play a very important role in maintaining the ecosystem. To automate the process of extracting trees from Remote Sensing data and counting thereby, we have used Cartosat-2S merged products @ 0.6m as input. Vegetation indices such as NDVI, and EVI can retrieve the Vegetation class which contains trees and their look-alikes (such as grass, fields, and shrubs) having the same spectral signatures *i.e.*, high reflectance in NIR band, and high absorption in Red band. Extraction of only Trees is not possible with these indices. In this paper we present a novel method for tree delineation from its look-alikes using AIML, concentrating more on preparation of input datasets efficiently. Based on the Spectral Separability analysis, only Red and NIR bands from the satellite imagery are utilized in the proposed method to separate Vegetation from the background classes (such as water, bare soil, terrain, and built-up). In addition to the two bands from satellite imagery entropy layer is computed from the NIR band and utilized as the third band to delineate trees from their look-alikes. Deep Neural Networks have the capability of learning complex patterns that can separate trees and their look-alikes. However, the performance is boosted when the entropy layer is added to the input image. The proposed method showed better performance when utilizing 3 bands Red, NIR, and Entropy bands compared to 4 bands i.e. Red, Green, Blue, and NIR bands. The proposed method obtains a precision of 96%, a recall of 90%, and an F1-score of 93% even with a relatively smaller training dataset. The study is performed on the data collected from various locations of the Indian state Rajasthan.

Keywords: Entropy, Merged product, Tree delineation, look-alikes, Spectral Signatures, Spectral Separability

1. Introduction

Trees benefit the environment and humans in many ways such as reducing pollution and noise because of the vehicles. Trees also protect the roads from erosion. Due to urbanization many trees are cut down every year. Hence many governments across the globe are taking active measures in order to increase the tree count. As a part of the Green Highways policy in 2015, Government of India has taken up the plantation of trees beside highways. As a part of this, there will be a monitoring agency to monitor the progress of planting and the status of plantations on a continuous basis. Such manual assessment of trees on a regular

basis involves a lot of human effort. To overcome this, the automatic segmentation of trees using remote sensing data is being widely studied.

Using high-resolution optical data individual tree delineation has been performed since the early 90s using various techniques such as valley following Gougeon (1995) local maxima Dralle and Rudemo (1997), watershed segmentation Lamar et al. (2005), template matching Brandtberg (1999), region growing Erikson (2003). Recently convolutional neural networks (CNNs), which have shown state-of-the-art results for various tasks such as semantic segmentation Luc et al. (2016), image classification He et al. (2016), and facial recognition Sun et al. (2015), are being widely applied for tree delineation. CNNs are Deep learning models which can extract useful features for performing a task from the raw data. Here, the input data given to the architecture plays a major role in the performance of a deep learning model.

For tree delineation, in terms of input data, many recent works utilize 3D LiDAR (Light Detection And Ranging) Coomes et al. (2017). Though trees and their crowns can be precisely detected using LiDAR data Hao et al. (2021), Windrim and Bryson (2020), due to the costs involved in the data collection, satellite data is preferred for analysis when the data needs to be collected and monitored on a regular basis.

In the proposed method using CNN based architecture UNet Ronneberger et al. (2015) on the high-resolution satellite data from the Cartosat-2 series tree delineation is performed using pixel-wise segmentation. Further to reduce the dimensionality of the input data, band selection is performed and two bands out of four are selected from the satellite imagery for tree delineation. A new band that utilizes information theory metric entropy Cover and Thomas (2006), is utilized as an additional band. This entropy band is derived by applying local entropy on the NIR band where Vegetation has high reflectance and shadow has low reflectance. The experimental results with the final 3 band input shows the efficacy of the proposed method.

Our contribution can be summarized as

1. A novel method for tree extraction based on two bands i.e. Red and NIR bands of High-resolution satellite data, is proposed

2. An additional band based on entropy is utilized for the first time for delineating trees. This is observed to minimize the commission errors where the tree look-alikes are classified as a tree.

3. Experimental results using Red, NIR and entropy bands, showed better performance than utilizing all four bands from satellite data.

2. Literature

This section reviews the works on tree delineation focusing on the ones that use satellite imagery.

The majority of the works on tree delineation utilize LiDAR data Hao et al. (2021) Windrim and Bryson (2020) or multispectral aerial data Safonova et al. (2021). LiDAR data can provide the height information of trees. Deep learning models such as ResNet He et al. (2016) and Mask-RCNN He et al. (2017) are used in recent LiDAR-based tree delineation

works Fromm et al. (2019), Hao et al. (2021). In addition to the cost of equipment and data processing, LiDAR systems cover a relatively small area in each measurement, which can make it challenging and time-consuming to acquire data for large regions. For multi-spectral aerial data, Open source software package Weinstein et al. (2020) is also released for delineating tree crowns, however, it is applied on the minimal resolution of 0.3m.

On the other hand, satellite imagery provides wide coverage and cost-effective data for the study. But the resolution of the satellite imagery is comparatively coarser than the LiDAR or UAV multispectral data. Raja (2016) used traditional watershed segmentation on PCA Wold et al. (1987) data and also the NDVI band. NDVI band resulted in better results than the PCA band. As CNNs became a powerful tool in the image processing field due to their high capabilities, many deep learning-based segmentation methods are used for tree delineation Freudenberg et al. (2022); G. Braga et al. (2020); Yang et al. (2022).

Freudenberg et al. (2022) applied a two step approach that contains two UNet models on 9 band data at 0.3m resolution. The complete network is trained using two additional auxiliary losses. G. Braga et al. (2020) utilized Mask-RCNN on PAN sharpened RGB images of 0.5m resolution. As Mask-RCNN requires more amount of training data simulated forest images are also utilized in G. Braga et al. (2020). Yang et al. (2022) also utilized Mask-RCNN for delineating trees at a much higher resolution of 0.27m.

Compared to the existing methods the proposed method utilizes less number of bands from the satellite imagery, also at a comparatively lower resolution of 0.6m for tree delineation. Simple UNet architecture without any additional losses during training is utilized. Further without any augmentation of the input data the proposed method shows better performance than the existing methods. Advanced models such as Unet++ Zhou et al. (2019), Attention U-net Oktay et al. (2018), and DeepLab Chen et al. (2017) can further enhance the results, at the cost of more learnable parameters.

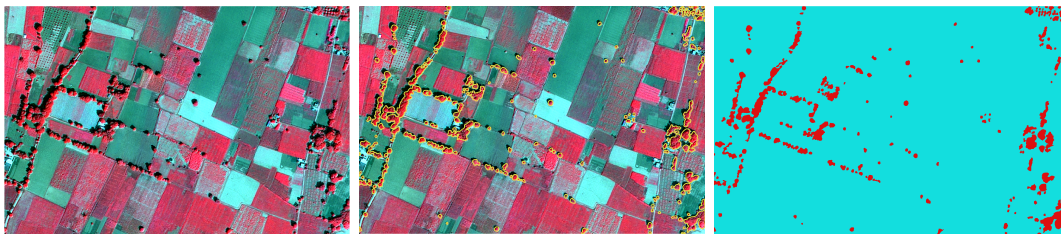
3. Proposed method

This section describes the steps involved in tree delineation as well as the analysis of the bands utilized in the proposed method, *Ted*.

3.1. Study Area and Data

All the images are taken from the Cartosat-2 series merged products from the areas of Rajasthan at a resolution of 0.6m. The satellite images in FCC (False Colour Composite) as shown in Fig. 1(a) containing considerable area with tree look-alikes (that have similar spectral signatures as trees) are extracted. A total area of 108 hectares approximately is utilized in preparing the training dataset. All the trees present in the satellite image are manually labeled by drawing polygons around tree canopies covering few additional pixels of shadow, using the QGIS tool. A common polygon is drawn around trees whose boundary is indistinguishable from the neighboring trees. Finally, the vector file (polygons drawn around tree canopies on training data) is rasterized to match the resolution of the respective satellite images.

The rasterized vector file is classified into 3 different values in the proposed method. Where no data is represented by 0, trees are represented by 1, and background data is represented by 2. The satellite image usually contains no data at the edges to fit its shape



(a) (b) (c)

Figure 1: (a) FCC of a scene from training data (b) with trees segmented by polygons and (c) the rasterized vector file

into a rectangle or square. When a subset of the satellite image is extracted, from the edge portion it can contain no data values, hence an additional no-data class is considered. The vector file overlayed on the satellite imagery and the vector converted to a raster file is represented in Fig. 1(b) and Fig. 1(c). 20% of the dataset is used for validation during training.

3.2. Notation

All the satellite images used have 4 bands of data. Hence each 3-dimensional image is represented by $I \in \mathbb{R}^{h \times w \times b}$. Where h, w, and b denote height, width, and number of bands respectively. For convenience, we denote the four bands of Cartosat-2S data *i.e.*, Blue, Green, Red and NIR as B_1, B_2, B_3, and B_4 respectively. Hence, the original image can be specified as

$$I = B_1 \oplus B_2 \oplus B_3 \oplus B_4 \tag{1}$$

where $\oplus$ denotes depth-wise concatenation.

3.3. Band selection from High resolution satellite images

Before preparing the dataset an analysis is performed for band separability to detect specific class using a spectral separable process on satellite images. In the case of UAV multispectral images Xi et al. (2021) analyzed various band selections and merging, using various techniques and reported, utilizing Red, NIR, and Green bands produced the most accurate result.

In the proposed method utilizing the satellite images, the mean radiance or pixel intensity values of each feature of classes Tree, Look-alikes, Waterbody, Built-up, and Soil, have been calculated and plotted for all four bands as shown in Fig. 2. Look-alikes in Fig. 2. mean the tree look-alikes. It can be observed in Fig. 2 that the Tree and Look-alikes classes that have the same spectral signatures are overlaid one on another and they are highly spectrally separable from the other classes. However, as the tree and its look-alikes have the same spectral signature, delineating trees from their look-alikes is a challenging task. Further to select the minimal bands to extract vegetation classes (Trees and Look-alikes) from background classes we have calculated the band ratios as shown in Table 1.

In Table 1 the ratios of mean radiance of various band combinations are depicted. Band ratio B_3/B_4 for vegetation classes has minimum value of 0.3 with remarkable difference with respect to background classes (*i.e.*, 1.135,0.844, 0.873, 1.958), compared to the other ratios B_1/B_4, B_2/B_3, B_2/B_4, B_1/B_3. B_3 and B_4 bands are therefore sufficient and hence selected to extract the vegetation classes.

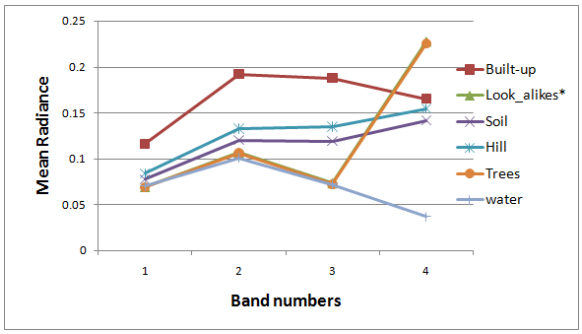


Figure 2: Mean radiance for various classes across four bands. (* Spectral signature of Tree and Look-alikes are same, hence Look-alikes trend is overlaid by Tree trend)

Table 1: Band ratios of mean radiance for various classes.

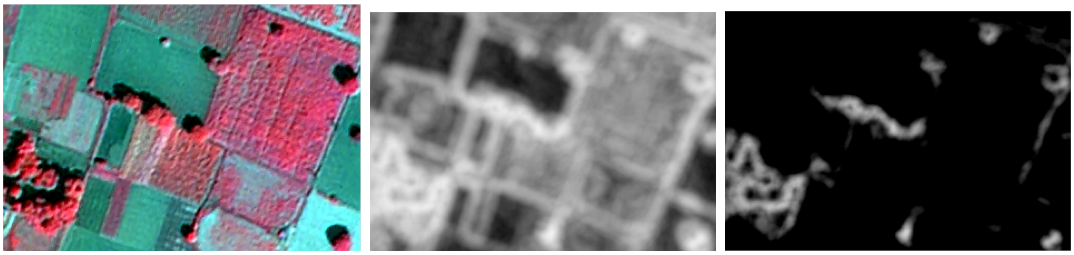
Band ratios / Class	B_1/B_4	B_2/B_3	B_3/B_4	B_2/B_4	B_1/B_3
Built-up	0.706	1.023	1.135	1.162	0.621
Look-alikes	**0.304**	**1.456**	**0.322**	**0.469**	**0.942**
Soil	0.551	1.008	0.844	0.850	0.654
Hill	0.546	0.983	0.873	0.859	0.626
Trees	**0.306**	**1.475**	**0.319**	**0.470**	**0.961**
Water	1.915	1.400	1.958	2.740	0.978

3.4. Entropy band

Shannon entropy identifies the amount of information or the randomness in the data. The entropy of an image M is given by

$$\text{Entropy(M)} = -\sum_{\text{i=1}}^{\text{N}} p_i \log\ p_i \tag{2}$$

where N denotes the total number of pixels and p_i is the probability of the i^{th} pixel value. In the context of an image if all the pixels are of the same value there is no randomness and hence no information can be captured from it. If an image has more number of different-valued pixels it is said to have more information, which results in a higher value of entropy. The local entropy can be used for texture classification and object detection Park et al. (2013).



(a) (b) (c)

Figure 3: (a) FCC of scene from training data (b) Result after applying local entropy on NIR band (c) Result after applying threshold on entropy.

Local Entropy is utilized in the proposed method to minimize the commission error of look-alikes as trees. In B_4 though look-alikes and trees both have high reflectance value, a tree is always accompanied by shadow whereas look-alikes are not. Hence look-alike texture is more uniform than a tree with a shadow, which results in a higher entropy value for trees and shadow combination as shown in Fig. 3.

To delineate the urban trees and grass Qian et al. (2020) proposed a new index using a panchromatic band at 0.5m resolution. However, in this method, the threshold is manually selected by visual inspection of the results at various values. In the proposed method local entropy $B_{entropy}$ is calculated using a sliding window size of 9×9 from the NIR band without using any threshold. Local entropy with window size $m \times n$ at location (x, y) is calculated as below.

$$B_{entropy}(x,y) = \sum_{i=m-m'}^{m+m'} \sum_{j=n-n'}^{n+n'} p_{x+i,y+j} \log\ p_{x+i,y+j} \tag{3}$$

Where $p_{a,b}$ denotes probability of pixel value at location (a, b) from B_4, $m' = \lfloor m/2 \rfloor$ and $n' = \lfloor n/2 \rfloor$.

The obtained entropy band (without applying any threshold) is stacked to the input whose information can be utilized by the DNN models for learning textural parameters. Fig. 3(a) and 3(b) show FCC and results after applying entropy on the NIR band of the scene respectively. From Fig. 3(b) it can be observed that the trees and look-alike portions are brighter compared to other portions. For better visualization of higher entropy values a threshold of 0.4 is manually applied, which resulted in Fig. 3(c) highlighting the tree portions and suppressing the look-alikes. This shows that the entropy band benefits in differentiating trees and their look-alikes.

Barbieri et al. (2011) segmented an image capturing the entropy distribution of each band for each of the distinct classes of water, rural and urban. This method cannot be generalized to classes that have similar spectral signatures. Hržić et al. (2019) utilized local entropy in medical imaging. Katiyar and Arun (2014) and Aalan Babu and Mary Anita Rajam (2020) studied local entropy in satellite imagery for various scenarios. However, as per our best knowledge Local entropy is utilized for the first time in satellite imagery for additional advantage of separation between tree and it's look-alikes. Though works exist for tree and grass delineation they are applicable for only aerial imagery Qian et al. (2020)].

Input size	256 X 256
Epochs	200
Learning rate	0.001
Weight decay	1e-4
Batch-size	2
Input Bands	Red, NIR and Entropy

Table 2: Training details.

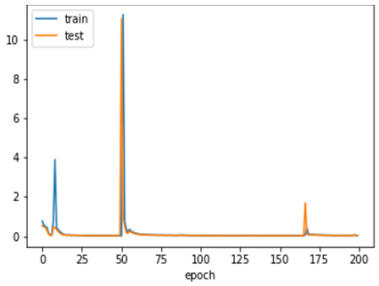


Figure 4: Loss curve of the Unet model trained using the proposed method.

3.5. Training

The final dataset has 2 bands i.e. Red and NIR from satellite data and an additional entropy band derived from the NIR band, which is denoted as

$$I_{train} = B_3 \oplus B_4 \oplus B_{entropy} \tag{4}$$

The CNN model UNet is trained using I_{train} along with the rasterized vector file as shown in Fig. 1(c) as labels for pixel-wise segmentation. Cross-entropy loss and L2 regularizer are utilized during training. After training the model is used for the prediction of the unseen data. For convenience, we refer to 3 bands utilized in the proposed method as *3-band input* and conventional four bands as *4 bands*.

3.6. Performance metrics and evaluation

To compare with the existing methods the commonly used metrics precision, recall, and F1-score for the tree class are utilized. Pixel-wise IoU metric is also utilized to compare the results of 3-band input with 4 bands. All the metrics are evaluated on the test data and the results are presented in the Section 4.

4. RESULTS AND DISCUSSION

4.1. Results on the training dataset

This section presents the results of the proposed method on the test data, which is 20% of the unutilized training data. The training dataset consists of high-resolution images from the areas of Sumerpur, Sheoganj, Chhibagaon, and Sutharon ka Gurha of the Indian state, Rajasthan. The input and the training details are described in Table 2. The proposed method uses Unet architecture which has 1.9M parameters.

The same hyperparameters as mentioned in Table 2 are used for training the model with different combinations of input bands. The loss curve of training and test data using 3-band input is depicted in Fig. 4. The results obtained on the 20% of the non-utilized data from the training dataset, for various band combinations, are depicted in Table 3, where PCA denotes principal component analysis.

It can be observed from Table 3 that the method M1 with 3-band input attained higher scores of precision, recall, and F1-scores compared to the other methods. The precision,

Table 3: Results of various band combinations.

Method	**Bands**	**Precision(%)**	**Recall(%)**	**F1-score(%)**
M1	B_3,B_4,$B_{entropy}$	89	86	87
M2	B_1,B_2,B_3,B_4	85	83	84
M3	PCA(B_1,B_2,B_3,B_4),$B_{entropy}$	88	82	85

recall, and F1-score values of the tree class are presented in Table 3. Method M3 is observed to fail in capturing many trees present in the scene which is reflected by its recall value. In M3, PCA has retained 97% of the total variance from B_3 and B_4 bands and still has not shown better prediction results. Among the methods from Table 3 M1 resulted in a better performance by identifying more trees and also distinguishing trees from look-alikes.

4.2. Results on the test scene

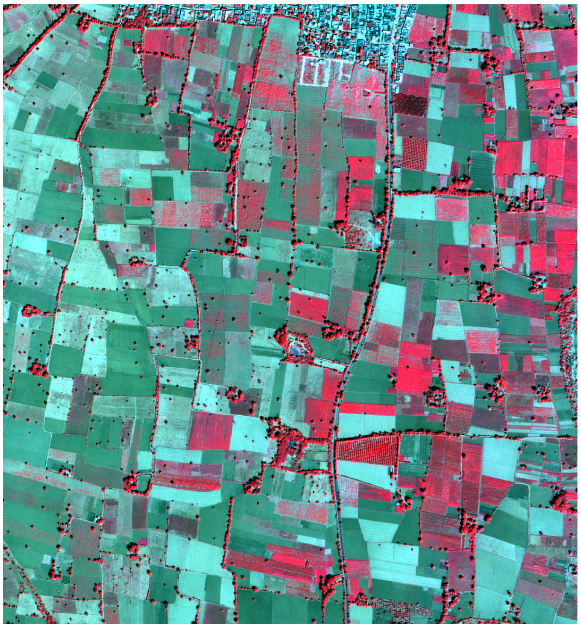
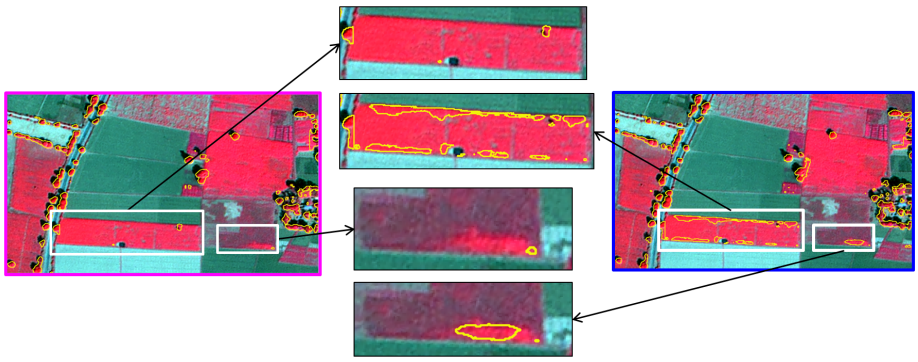
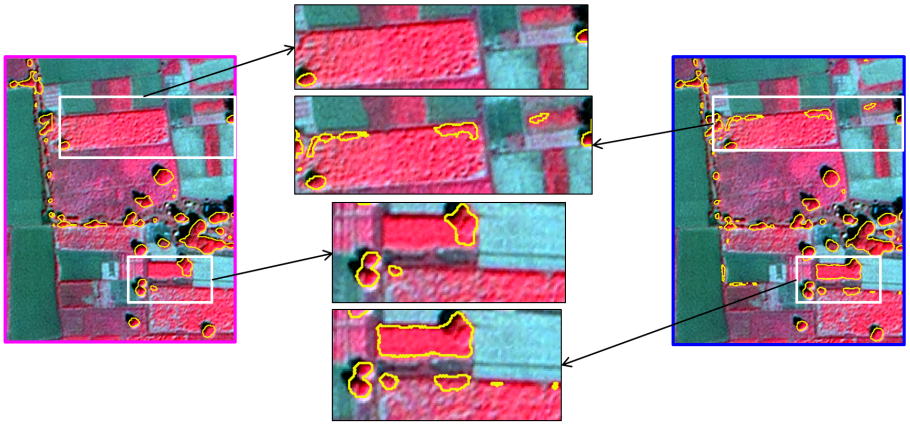


Figure 5: Test area (360 hectares) from Rajasthan.

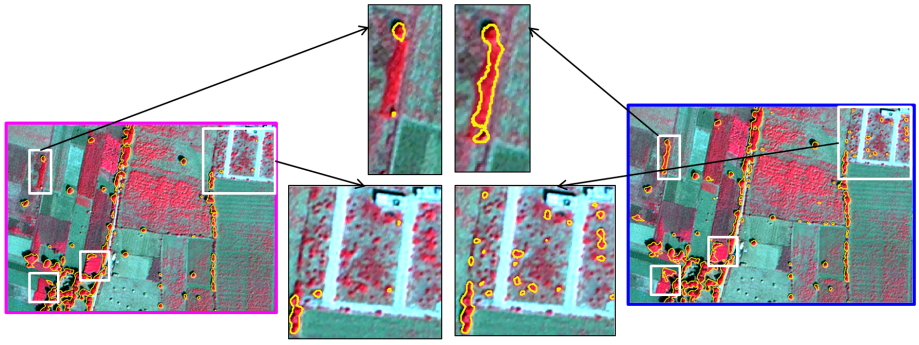
The test scene in Fig. 5 is extracted from the Sheoganj area of Rajasthan centered at 25.115781N 72.982314E covering 360 hectares. The test scene consists of many tree look-alikes that seem to have similar spatial features and spectral signatures to the trees. However, trees are accompanied by shadows having different spectral signatures from that of trees. Due to the differences in the pixel values for tree and shadow combination, a higher entropy value is resulted at trees compared to their look-alikes (as depicted in Fig. 3(c)). The scene also covered a portion of built-up in the top portion with few trees in between. The comparison of the predicted results using 3-band input and all the 4 bands from the satellite imagery are depicted in Fig. 6(a), 6(b), 6(c) and Fig. 6(d) respectively (Images on



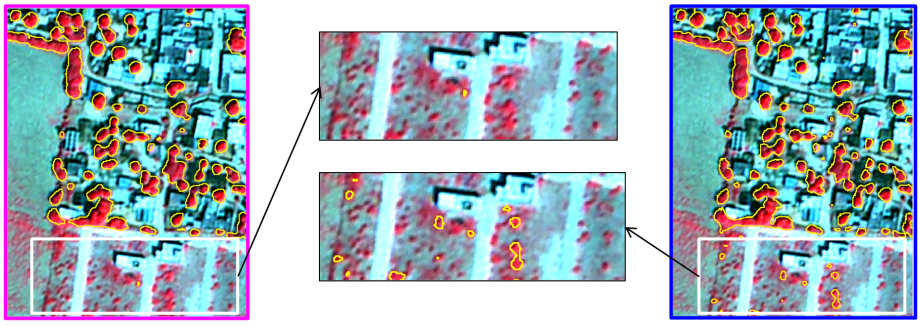
(a)



(b)



(c)



(d)

Figure 6: Predicted results on satellite data using Ted (Left side with magenta background) and Original 4 bands (Right side with blue background) respectively; Portions from the predicted images where the Ted exhibits prominent changes are emphasized.

the left represent results of the scenes with 3-band input and Images on the left represent their respective results using all the bands from satellite data).

From Fig. 6(a) it can be observed that a portion of thick vegetation present at the bottom of the image is misclassified as trees using 4 bands. Also, the vegetation below the cluster of trees at the right-bottom is misclassified as a tree using 4 bands. In Fig. 6(a) both the former cases are correctly classified as background using Ted. In Fig. 6(b) also the multiple vegetation areas are misclassified as trees, and the same are correctly classified as background by Ted. Similar observations where tree look-alikes are delineated as trees are found in the predicted images Fig. 6(c) and Fig. 6(d) using 4 bands. Even though there is no built-up area in the training data Fig. 6(d) shows that the proposed method is able to identify the trees in built-up correctly. Hence Fig. 6. indicates that the proposed method with entropy band is able to extract trees well by distinguishing trees from their look-alikes in a better way.

Comparison with the other methods is presented in Table 4. As shown in Table 4 compared to the other methods, even with the basic Unet model the proposed method has achieved better results with considerably less number of bands at lower resolution. The precision, recall, and F1-score values of the tree class are presented in Table 4. The proposed method shows the highest recall and F1-score values and the second highest precision value. It is to be noted that all the methods are evaluated on various areas of interest. The high precision of the proposed method indicates that there are comparatively fewer False positives, which are mainly due to the misclassification of the look-alikes as trees. The high value of Recall indicates that the majority of the trees available in a given scene are identified by the algorithm.

Table 4: Comparision of various tree delineation methods. (n denotes number of bands and Res denotes resolution)

Method	**n**	**Res**	**Precision(%)**	**Recall(%)**	**F1-score(%)**
Freudenberg et al. (2022)	9	0.3m	85	83	84
G. Braga et al. (2020)	3	0.5m	89	68	77
Gomes et al. (2018)	9	-	60	67	63
Yang et al. (2022)	3	0.3m	99	82	90
Ted (Proposed method)	3	0.6m	96	90	93

Table 5: Comparision of IoU for different threshold values on NDVI.

Threshold	**IoU of 4 bands**	**IoU of Ted**
0.20	0.74	0.92
0.25	0.74	0.92
0.30	0.75	0.93

The pixel-wise IoU metric is also calculated for the vegetation pixels only, from the test image. From Fig. 5 it can be observed that the background pixels are more in number compared to trees and their look-alikes. Moreover, the background class is classified correctly as depicted in Fig. 6, as it has a completely different spectral signature as shown in Fig.

2. Therefore including all the pixels (including background class) in IoU metric results in a higher value as majority of the pixels *i.e.*, background class are classified correctly. Hence we considered only the vegetation part from the test image. To extract the vegetation part, NDVI is calculated from the test data, and the locations that have NDVI greater than 0.2 are retrieved, which included all the vegetation. The IoU metric is calculated only at those vegetation locations of the test scene. This resulted in 0.92 IoU for 3-band input and 0.74 when all the satellite bands are used as input. The lesser IoU using four bands is due to the misclassification of look-alikes as trees and smaller trees as background from the test scene. Table 5 shows the IoU values of 4bands and the proposed method for different threshold values on NDVI.

5. Conclusion

In this paper, a novel method for extracting trees is proposed using only Red and NIR bands from the satellite image. A new band derived from the local entropy of the NIR band is also added as the input, totaling 3 bands of input to the UNet architecture. The entropy layer is observed to capture the tree and shadow combination with a higher entropy than the tree look-alikes which have a comparatively uniform texture. The results obtained on the test scene have shown a precision value of 96%, recall value of 90%, and F1-score of 93%. Pixel-wise IoU metric has obtained values of 0.74 and 0.92 utilizing 4 bands and 3-band input showcasing the efficacy of the proposed method. Lower IoU with 4 bands from the satellite imagery reflects that more look-alikes are misclassified as trees. On the whole, the 3-band input is effective in extracting the trees even in the environment surrounded by the tree look-alikes with similar spectral signatures.

Acknowledgments

We thank National Remote Sensing Center- Bhuvan division for the guidance and support

References

A Aalan Babu and V Mary Anita Rajam. Water-body segmentation from satellite images using kapur's entropy-based thresholding method. *Computational Intelligence*, 36(3): 1242–1260, 2020.

Andre L Barbieri, GF De Arruda, Francisco A Rodrigues, Odemir M Bruno, and Luciano da Fontoura Costa. An entropy-based approach to automatic image segmentation of satellite images. *Physica A: Statistical Mechanics and its Applications*, 390(3):512–518, 2011.

Tomas Brandtberg. Automatic individual tree based analysis of high spatial resolution aerial images on naturally regenerated boreal forests. *Canadian Journal of Forest Research*, 29 (10):1464–1478, 1999.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous

convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

David A Coomes, Michele Dalponte, Tommaso Jucker, Gregory P Asner, Lindsay F Banin, David FRP Burslem, Simon L Lewis, Reuben Nilus, Oliver L Phillips, Mui-How Phua, et al. Area-based vs tree-centric approaches to mapping forest carbon in southeast asian forests from airborne laser scanning data. *Remote Sensing of Environment*, 194:77–88, 2017.

Thomas M Cover and Joy A Thomas. Wiley series in telecommunications and signal processing. In *Elements of information theory*. Wiley-Interscience USA, 2006.

Kim Dralle and Mats Rudemo. Automatic estimation of individual tree positions from aerial photos. *Canadian Journal of Forest Research*, 27(11):1728–1736, 1997.

Mats Erikson. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, 33(8): 1557–1563, 2003.

Maximilian Freudenberg, Paul Magdon, and Nils Nölke. Individual tree crown delineation in high-resolution remote sensing images based on u-net. *Neural Computing and Applications*, pages 1–11, 2022.

Michael Fromm, Matthias Schubert, Guillermo Castilla, Julia Linke, and Greg McDermid. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sensing*, 11(21):2585, 2019.

José R G. Braga, Vinícius Peripato, Ricardo Dalagnol, Matheus P. Ferreira, Yuliya Tarabalka, Luiz E OC Aragão, Haroldo F. de Campos Velho, Elcio H Shiguemori, and Fabien H Wagner. Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, 12(8):1288, 2020.

Marília Ferreira Gomes, Philippe Maillard, and Huawu Deng. Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model. *Remote Sensing of Environment*, 211:184–195, 2018.

François A Gougeon. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian journal of remote sensing*, 21(3):274–284, 1995.

Zhenbang Hao, Lili Lin, Christopher J Post, Elena A Mikhailova, Minghui Li, Yan Chen, Kunyong Yu, and Jian Liu. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask r-cnn). *ISPRS Journal of Photogrammetry and Remote Sensing*, 178:112–123, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

Franko Hržić, Ivan Štajduhar, Sebastian Tschauner, Erich Sorantin, and Jonatan Lerga. Local-entropy based approach for x-ray image segmentation and fracture detection. *Entropy*, 21(4):338, 2019.

Sunil Kumar Katiyar and PV Arun. A review over the applicability of image entropy in analyses of remote sensing datasets. *arXiv preprint arXiv:1405.6133*, 2014.

W Robert Lamar, James B McGraw, and Timothy A Warner. Multitemporal censusing of a population of eastern hemlock (tsuga canadensis l.) from remotely sensed imagery using an automated segmentation and reconciliation procedure. *Remote Sensing of Environment*, 94(1):133–143, 2005.

Pauline Luc, Camille Couprie, Soumith Chintala, and Jakob Verbeek. Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*, 2016.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

Bo-Young Park, Hyo-Hun Kim, and Byung-Woo Hong. A multilabel texture segmentation based on local entropy signature. *Mathematical Problems in Engineering*, 2013, 2013.

Yuguo Qian, Weiqi Zhou, Christopher J Nytch, Lijian Han, and Zhiqiang Li. A new index to differentiate tree and grass based on high resolution image and object-based methods. *Urban Forestry & Urban Greening*, 53:126661, 2020.

Shaik Karimullaand A Ravi Raja. Tree crown delineation from high resolution satellite images. *Indian Journal of Science and Technology*, 9:S1, 2016.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

Anastasiia Safonova, Yousif Hamad, Egor Dmitriev, Georgi Georgiev, Vladislav Trenkin, Margarita Georgieva, Stelian Dimitrov, and Martin Iliev. Individual tree crown delineation for the species classification and assessment of vital status of forest stands from uav images. *Drones*, 5(3):77, 2021.

Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*, 2015.

Ben G Weinstein, Sergio Marconi, Mélaine Aubry-Kientz, Gregoire Vincent, Henry Senyondo, and Ethan P White. Deepforest: A python package for rgb deep learning tree crown delineation. *Methods in Ecology and Evolution*, 11(12):1743–1751, 2020.

Lloyd Windrim and Mitch Bryson. Detection, segmentation, and model fitting of individual tree stems from airborne laser scanning of forests using deep learning. *Remote Sensing*, 12(9):1469, 2020.

Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

Xiangshu Xi, Kai Xia, Yinhui Yang, Xiaochen Du, and Hailin Feng. Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and uav multispectral imagery in urban forest. *Computers and Electronics in Agriculture*, 191:106506, 2021.

Mingxia Yang, Yuling Mou, Shan Liu, Yanrong Meng, Zelin Liu, Peng Li, Wenhua Xiang, Xiaolu Zhou, and Changhui Peng. Detecting and mapping tree crowns based on convolutional neural network and google earth images. *International Journal of Applied Earth Observation and Geoinformation*, 108:102764, 2022.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.