A STUDY OF USING FULLY CONVOLUTIONAL NETWORK FOR TREETOP DETECTION ON REMOTE SENSING DATA

Changlin Xiao[1, 2], Rongjun Qin[2, 3] *, Xu Huang[2], Jiaqiang Li[1]

[1] Future Cities Laboratory, Singapore-ETH Centre, ETH Zurich, 1 Create Way, CREATE Tower, #06-01, 138602, Singapore.
[2] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, 218B Bolz Hall, 2036 Neil Avenue, Columbus, OH 43210, USA. qin.324@osu.edu
[3] Department of Electrical and Computer Engineering, The Ohio State University, 205 Dreese Labs, 2015 Neil Avenue, Columbus, OH, 43210, USA. qin.324@osu.edu

Commission I, WG I/8

KEY WORDS: Treetop Detection, Fully Connected Network, Pseudo Label, Remote Sensing, Satellite Image, DSM

ABSTRACT:

Individual tree detection and counting are critical for the forest inventory management. In almost all of these methods that based on remote sensing data, the treetop detection is the most important and essential part. However, due to the diversities of the tree attributes, such as crown size and branch distribution, it is hard to find a universal treetop detector and most of the current detectors need to be carefully designed based on the heuristic or prior knowledge. Hence, to find an efficient and versatile detector, we apply deep neural network to extract and learn the high-level semantic treetop features. In contrast to using manually labelled training data, we innovatively train the network with the pseudo ones that come from the result of the conventional non-supervised treetop detectors which may be not robust in different scenarios. In this study, we use multi-view high-resolution satellite imagery derived DSM (Digital Surface Model) and multispectral orthophoto as data and apply the top-hat by reconstruction (THR) operation to find treetops as the pseudo labels. The FCN (fully convolutional network) is adopted as a pixel-level classification network to segment the input image into treetops and non-treetops pixels. Our experiments show that the FCN based treetop detector is able to achieve a detection accuracy of *99.7*% at the prairie area and *66.3*% at the complicated town area which shows better performance than THR in the various scenarios. This study demonstrates that without manual labels, the FCN treetop detector can be trained by the pseudo labels that generated using the non-supervised detector and achieve better and robust results in different scenarios.

1. INTRODUCTION

Forest is one of the most important land surfaces of the earth and plays an important role in the global ecosystem. Detailed tree-level attributes such as tree counts, tree heights, and canopy size are essential for monitoring forest regeneration, quantitative analysis of forest structure and dynamics, large-scale ecological simulations and evaluation of deforestations (Mohan et al., 2017; Weng et al., 2015; Zhao et al., 2014). Many works have been proposed to perform tree detection and crown delineation with remote sensing data and have shown great potential in accurately detect in individual level (Hill et al., 2017; Kathuria et al., 2016; Latifi et al., 2015). In most of these methods, the treetop detection is an essential and critical step, of which the detection accuracy is decisive for the final results. And many of these methods are using a window-based filter to find a local maximal point as potential treetop which is feasible but heavily affected by the type of trees. In practice, it is normally hard to find a suitable window size for the treetop detection due to the high variation of crown size, even with methods that adaptively adjust filter size (Ke and Quackenbush, 2011; Özcan et al., 2017; Santoro et al., 2013; Skurikhin et al., 2013; Song et al., 2010; Wulder et al., 2000).

Recently, Convolutional Neural Networks (CNN) with deep learning technology have demonstrated promising performance in tasks such as image classification (Krizhevsky et al., 2012), object detection (Redmon et al., 2016; Ren et al., 2015) and segmentation (Tsogkas et al., 2015). Different from the low-level hand-crafted features, the convolutional neural networks have shown good performance in learning high-level semantic information from the training samples. After training on large-scale datasets like ImageNet (Deng et al., 2009), it has been shown that the CNN is able to learn distinctive information for different object categories and feature points. Considering the treetops have strong geometrical and spectral characteristics, we apply the CNN to the treetop detection in order to remove the time-consuming design of local maximal filter which may need repeating experimentation at different scenarios. To overcome the shortness of training sample and make the network practicable, we innovatively train the network without manual labels but the pseudo ones that come from the results of the non-supervised local maximum detector and get a better detection accuracy.

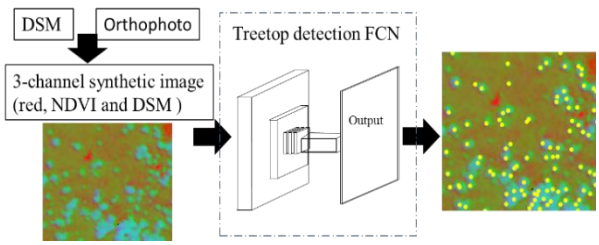

Figure 1. The framework of the FCN based treetop detection. Firstly, the fused image is generated from DSM and multi-band orthophoto. Then the image is feed to the treetop detection FCN to get a segmental result and finally, the result is converted to treetop points.

In the study, we use multi-view very-high-resolution (VHR) satellite imagery derived DSM (Digital Surface Model) and multi-band orthophoto as research data. We fuse the DSM and orthophoto into a 3-channel fused multi-clue image which includes the red band, the NDVI values, and the DSM height. For the training samples, instead of manually labeling them, we use the top-hat by reconstruction (THR) operation on the DSM to

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-163-2018 | © Authors 2018. CC BY 4.0 License.

detect the local maximum as treetops and mark them as pseudo labels. For the deep learning network, we adopt the FCN (fully convolutional network) (Long et al., 2015) to segment the input image into two categories: treetops and non-treetops. We train the FCN with small and fixed-size patches. However, due to the character of the fully convolutional layer, an image with any size can be used as the input of this network. The framework of the proposed FCN based treetop detection is shown in figure 1.

Through this trail of using FCN for treetop detection for the first time, we found an efficient and effective way to detect treetops that more robustly and efficiently works at different scenarios. Also, the experiment demonstrated the possibility of using pseudo labels to train the treetop detection neural network without any manually labeled samples which make the FCN detector more practical. The rest of the paper is organized as follows: The related work is in section 2, which followed by the methodology of this study; The experiments are given in section 4; Finally, we conclude in section 5.

2. RELATED WORK

2.1 Treetop Detection with Local Maximal Filter

Extensive works based on remote sensing images have been performed to contribute to forest measurement at the individual tree level. Typically, the first step of these methods is to detect the treetops. An often used assumption is that treetops reflect the light shedding on them in a decreasing manner from top to bottom (Culvenor, 2002; Özcan et al., 2017). Hence, treetops can be detected as bright spots in the satellite or aerial images, and based on this, window-based local maximal filters were proposed to find the brightest points as treetops (Pouliot and King, 2005; Wang et al., 2004; Wulder et al., 2000). However, it is difficult to find an appropriate window size for the maximal filter which heavily depends on the spatial resolution of the image and the size of the trees that usually have huge variations (Ke and Quackenbush, 2011; Özcan et al., 2017). Even there are several local maximal filters which can adaptively change their sizes based on the estimated dimension of the trees (Santoro et al., 2013; Skurikhin et al., 2013; Song et al., 2010), the filter-based methods still do not have enough improvement in the areas that contain a variety of trees. Besides local maximal filtering, tree templates were also used to detect treetops through template matching (Quackenbush et al., 2000; Tarp-Johansen, 2002). Representative and complete templates normally lead to good matching results with high accuracy, while such methods may need a large amount of training data and their poor transferability to other different datasets can be an issue (Mallinis et al., 2008).

On the other hand, 3D presentation of the surface of objects that provided by 3D points is becoming popular. Right now, much attention is given to the use of 3D point cloud data on individual tree detection (Ferraz et al., 2016; Gini et al., 2014; Jakubowski et al., 2013; Kathuria et al., 2016; Saarinen et al., 2017; Strîmbu and Strîmbu, 2015; Turner et al., 2012). Most of the methods used the canopy height model (CHM) generated from the 3D point clouds. The CHM can naturally highlight the treetops and directly derive the tree heights. Similar to the 2D image-based method, the CHM-based methods also use procedures such as image smoothing, local maxima localization, and template matching to detect individual trees and their boundaries (Chen et al., 2006; Koch et al., 2006; Popescu et al., 2002). like the window based method, the size of local maximal filter needs to be carefully considered (Mohan et al., 2017).

2.2 Image Segmentation with Fully Convolutional Network

Recently, deep convolutional neural networks (CNN) have dominated the computer vision community and outperformed many competing methods in various object recognition challenges, such as Pascal VOC Semantic Labeling Challenges (Everingham et al., 2010) and the ImageNet classification competition (Deng et al., 2009). Many CNNs have been applied to pixel-level classification of overhead imagery (Sherrah, 2016; Sun et al., 2017), in which the FCN being one of the most basic networks. By converting the fully connected layer into the fully convolutional layer, it can efficiently perform classification for input image with any size at pixel-level for semantic segmentation. Given its superior performance, in this paper, we adopt FCN to perform a practical task of treetop detection, to validate its feasibility for single class object detection.

3. METHODOLOGY

3.1 Study Area and The Data

The study area is located in Don Torcuato, a small city on the west side of Buenos Aires, Argentina (figure 2). The size of this area is 6.740 km by 6.914 km (22469 pixels × 23048 pixels) and the scene contains both forested prairies and urban areas. The satellite images in this work are from the multi-view benchmark dataset provided by John’s Hopkins University Applied Physics Lab (JHUAPL) (Bosch et al., 2016; Bosch et al., 2017), containing multiple worldview2/3 images over this area across two years with a total of approximately 50 images. They were taken under various conditions containing on-track and off-track stereos with the ground resolution around 0.3 meters, a complete set of meta information can be found on their hosting website. To derive an accurate DSM, we selected five pairs of the on-track stereo image from the year of 2015 in December, with the maximal off-nadir angle between 7-19 degrees and the average intersection angle between 15-21 degrees. We applied a fully automated pipeline proposed by (Qin, 2014, 2017; Qin et al., 2016) that consists of 1) pansharpening, 2) automatic feature matching, 3) pair-wise bundle adjustment, 4) dense matching and 5) a bilateral-filter based depth-fusion, to generate a high-quality DSM and subsequently true orthophoto. The core method is a hierarchical semi-global matching (Qin, 2016). In (Qin, 2017), he reported an absolute accuracy on this particular dataset varying between 2.5-4 meters (inclusion of blunders of all types of objects). The readers may find more details about the method on this data in (Qin, 2017). Figure 2 shows the orthophoto and the DSM for the experimental sites.

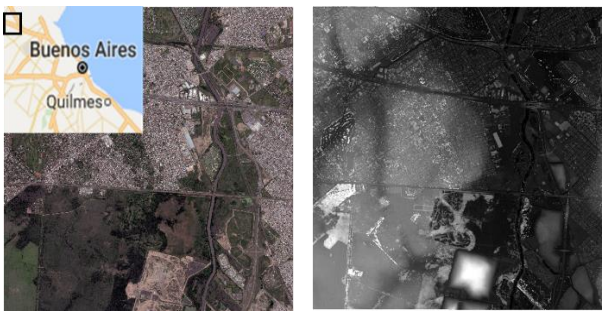

Figure 2. The orthophoto (RGB band) and the DSM generated from satellite imagery with 0.3 m spatial resolution.

3.2 The Top-Hat Local Maximal Treetop Detector

Instead of manually labeling the treetops for the training samples, we use the top-hat by reconstruction operation to detect local maximal points and further refine them by height check and non-maxima suppression to generate treetop masks. These treetop masks are used to train the network as well as a comparison.

For trees, we naturally assume that the local maximal points in the DSM are the treetops. However, since the filter-based method requires a careful tuning of the window size, we adopt the grey-level morphological top-hat by reconstruction operator (THR) to find the local maximal points, as it is an effective method to detection blob-like shapes and is less sensitive to window sizes (Qin and Fang, 2014; Vincent, 1993). Morphological top-hat (MTH) is defined as the peaks of an image grid computed by morphological operations. In the detection, we first use a disk-shaped structuring element (SE) to do the grey-level morphology erosion on the DSM to generate a marker image $\varepsilon(\mathrm{DSM}, \mathrm{e})$. Then the morphological reconstruction mask $\mathrm{B}_{\varepsilon(\mathrm{DSM,e})}$ is generated from maker image with an iterative operation. Finally, by subtracting morphological reconstruction mask from the DSM, the peaks on DSM can be extracted. Unlike the filter-based methods that only offer one maxima in each filter region and the number of total local maxima is heavily dependent on the filter size, in most cases, the THR produces local maxima as a blob-like region and the size of SE specifies the maximally-detectable region, thus still keeping all possible local maximal points even with overestimation. To locate single local maxima, we use opening operation in morphology to reduce (one local maximal point in one region) or separate (several sub-local-maximal points in one region) large regions and the find maximal points in these small regions are the treetops.

In the detection, we use Normalized Difference Vegetation Index (NDVI) to remove the points that in the non-vegetation area and check the height of the points by subtracting the height of nearby terrain area on the DSM. Finally, we use a non-maxima suppression to refine the treetops that are too close to each other. Figure 3 shows an example of the local maximal point detection and the final treetops. The left images (left-top) are the false-color satellite image of the test area and the NDVI image (left-bottom) helps to remove the local maximal points in the non-vegetated area. The final treetops are shown as blue dots in the right image where red dots are the local maxima that filtered out by the height check and the green stars without blue dots are the ones that filtered out by the non-maxima suppression. The size of non-maxima suppression window (red rectangles in the image) is decided by the height of the local maxima.

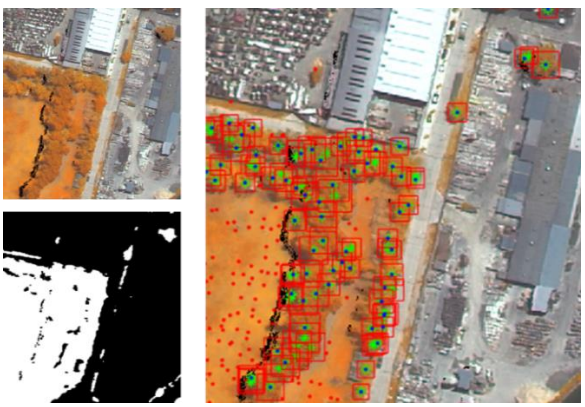

Figure 3: The illustration of the treetop detection. The details are in the text above.

3.3 The Sample Generation

In the study area, we randomly select 10 patches with 1000 × 1000 pixels (300m × 300m) as training (8 patches) and test (2 patches) areas. In each training patch, 3000 sub-patches are generated as training samples. The size of the training sample is designed as 48 × 48 pixels corresponding to 14.4 × 14.4m^2 which is normally large enough to cover the crown of a tree. The treetop mask (0 for the non-treetops area, 1 for treetop area) is generated by finding treetops through top-hat by reconstruction introduced in above section. Instead of using single pixel as treetop, we marked the area that around the treetop pixel as treetop area. The treetop area size in the experiment is set as 3×3 pixels.

In this study, we convert the 8 bands input orthophoto into a 3-channel fused multi-cue image. The tree channels are red, NDVI and DSM values that normalized into 0-1. Figure 4 gives an illustration of the training sample generation processing.

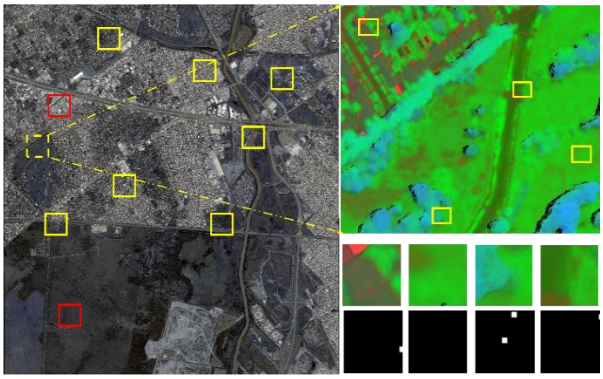

Figure 4. The generation of training samples for the FCN. In the left image, the rectangles represent patches used for training (yellow) and test (red) samples. The right-top image is the 3-channel fused image and the rectangles show four examples of the training samples and their labels (right-bottom).

3.4 The Architecture of The FCN and The Training

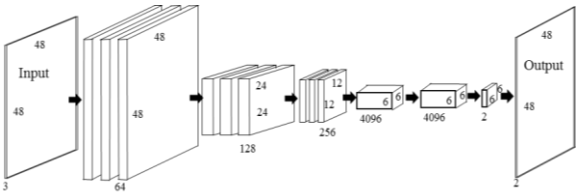

Figure 5. The architecture of the adopted FCN which is shorter than the original one.

In this study, we adopt the basic FCN architecture as our network. We cropped the FCN that instead of using 5 blocks of convolution and max-pooling layers, we only use 3 due to the input size is 48 × 48. After 3 max-pooling layers, the feature map's size will reduce to 6 × 6 and the two fully convolutional layers will produce a coarse prediction for the 2 classes at the downsampled resolution. The last layer is an up-sampling layer used to resize the output image as big as the input one. The readers can find more details about the architecture of the original FCN in (Long et al., 2015).

The output of the FCN is a 2-channel classification probability distribution map. The values in the first and second channel represent the possibility of being treetops and non-treetops separately. From these two channels, we can generate a

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-163-2018 | © Authors 2018. CC BY 4.0 License.

segmentation map that includes the two categories. Figure 6 shows an example of the segmentation output of the FCN. Since the segments are regions, we need to further convert them into single point treetops. We use the highest point in each region to represent the treetops in it.

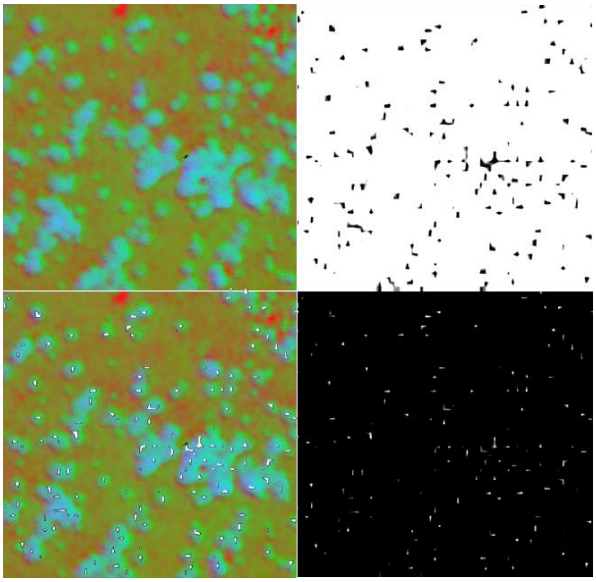

Figure 6. The treetop detection with FCN output. The top-left image is the input fused image with arbitrary size. The left-bottom image is the segmentation result (white for the treetops) overlapping the input image. Right images are the two output channels that represent the possibility of been non-treetops (right-top) and treetops (right-bottom). And the brighter the pixel, the higher possibility it belongs to the current category.

To train the network, we generate total 24000 image patches and labels as described in section 3.3. We set the training epoch as 200 and the size of training batch is set as 256 and the learning rate is set to 0.0001 with the Adam optimizer. Since it is a classic classification problem, we use the cross-entropy loss as the target function. However, the treetop only takes a small part of the image, we weight the loss as [1, 10] for the non-treetops area and treetops separately.

4. EXPERIMENT RESULTS

4.1 Reference Data

Besides the pseudo labels that generated by the THR operation, we still need true labels as reference data to assess the performance of the treetop detection. In this study, due to our limitation to collect the field samples, we generate the reference data by labeling the individual trees with visual inspection as some previous research did in their works (Brandtberg et al., 2003; Ke and Quackenbush, 2011). The reference samples are collected by visual inspection through 3D visualization of the orthophoto and DSM as shown in figure 7. In this study, we labeled the treetops in two test areas which including a town and a sparsely forested prairie like the red rectangle marked in figure 4. In the experiment, the FCN and top-hat based treetop detectors are separately used to find the treetops in these two test areas and compared to the reference data.

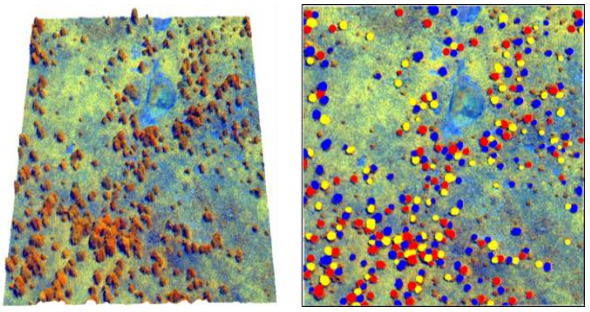

Figure 7. The 3D visualization of the data (left) and the reference masks (color dots) for the prairie area.

4.2 Accuracy Assessment Measure

To quantitatively validate the individual treetop detection accuracy, we use true positives (TP), false positives (FP) and false negatives (FN) to compute the correct detection, wrong detections, and the missing detections, respectively denoted as detection accuracy (DA) or recall (r), commission error (e_{com}) and the omission error (e_{om}):

$$DA/r = \frac{n_{TP}}{N}, \quad (1)$$

$$e_{com} = \frac{n_{FP}}{n_{TP}+n_{FP}}, \quad (2)$$

$$e_{om} = \frac{n_{FN}}{n_{TP}+n_{FN}}, \quad (3)$$

where n_{TP}, n_{FN} and n_{FP} are the number of treetops in TP, FN and FP category. N is the total number of the reference treetops. Other metrics like precision (P) and F-score(F) can be derived as:

$$p = \frac{n_{TP}}{n_{TP}+n_{FP}}, \quad (4)$$

$$F = \frac{2rp}{r+p}. \quad (5)$$

In the experiment, if the detected treetops have an overlap with the reference mask, we make it as a correct detection, otherwise a false detection.

4.3 Experimental Results and Discussions

The two experimental sites variably include densely and sparsely distributed trees, buildings, cars, shrubs and glasses which make the test site cover various surfaces like the urban area. For each site, we calculate the results of DA, e_{com} and e_{om}, as well as the precision P and F-scores and the final results can be found in table 1 and the visualized results can be found in figure 8 and figure 9.

Site /N_R.	Detector	N_D.	DA/r	e_{com}	e_{om}	P	F
Town /187	FCN	178	**0.529**	0.444	**0.471**	0.556	**0.543**
	THR	134	0.444	**0.381**	0.556	**0.619**	0.517
Prairie /307	FCN	272	0.746	**0.158**	0.254	**0.842**	0.791
	THR	287	**0.782**	0.164	**0.218**	0.836	**0.808**

Table 1. The experiment results in the two test areas. N_R. and N_D. refer to the number of the reference and detected treetops. DA/r is detection accuracy or recall ratio. e_{com} and e_{om} are the commission error and the omission error while precision and F-score are represented as P and F. The best numbers are red and bolded.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-163-2018 | © Authors 2018. CC BY 4.0 License.

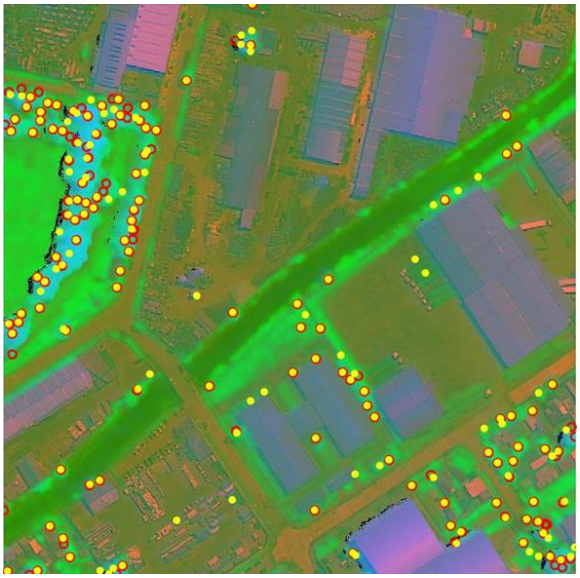

Figure 8. The treetop detection result of the town area. The yellow dots are the treetops detected by FCN and the red circles are the treetops detected by THR operation.

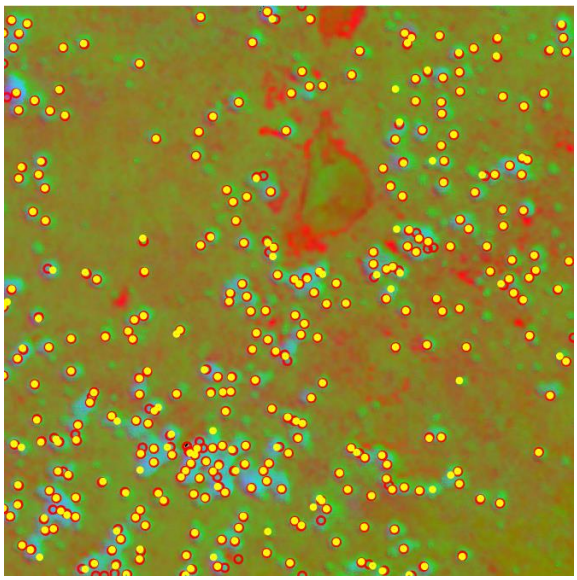

Figure 9. The treetop detection result of the prairie area. The yellow dots are the treetops detected by FCN and the red circles are the treetops detected by THR operation.

As we can observe from the table 1, comparing the two methods, the FCN greatly improved the performance in the town area, there are total *187* reference trees while FCN detected *52.9%* and THR detected *44.4%*. The surface objects of town area are quite complicated: trees at courtyards and on the street vary with crown size and height. Shrubs, glasses, and buildings close to trees can easily affect the tree area extraction. However, as shown in figure 8, the FCN based treetop detector detected most treetops even they have different height and close to buildings. Even it is trained by the THR results, the FCN has much higher DA/r values than THR and better or very close values in other assessments. The deep learning based network is an end-to-end method which does not need to adjust parameters for specified scenarios. In this study, the FCN has learned tens of thousands of various samples including trees and non-tree objects and it has found the best parameters for all the scenarios. On the other hand, the detector using top-hat by reconstruction (THR) operation in this area only has a mediocre performance. We think the complexity in this area impeded the prior-knowledge we have injected into the THR detector and further reduced the detection ability. This is also a limitation of the manually crafted feature-based method that needs prior-knowledge and adjusting the parameter to being adapted to different situations. Another reason we believe is that, unlike FCN, the top-hat based method never learned the non-treetops objects that similar to treetops and it would miss-classify some confused objects.

In the prairie area, as shown in figure 9, both two methods have a good performance. The surface in this area is relatively smooth and the trees are easier to be detected. However, even though this experiment site is deemed in general as an easier tree detection area, the trees are still relatively dense in some places and there is no pattern of their distribution. For this area, there are total *307* reference trees and the FCN algorithm detected *272* treetops while the THR get *287* treetops. Comparing to the town area, the DA/r can reach as high as *0.782* (THR) and *0.746* (FCN). In this area, the two methods have very similar performance that THR based method has a little better DA/r. We think this is due to the fact that the trees in this area have the typical shapes that we have considered in the design of top-hat based treetop detection. Meanwhile, without noise and distractions, the top-hat have a precise detection of the treetops which also taught the FCN to have a fair performance in this area.

From these scores, we can find out that the FCN treetop detector shows comparative performance as THR at the prairie area while has an outperformance at the town area. The FCN was trained from the pseudo labels that generated by THR which may have minor errors. But due to the global optimization of all the training samples including treetops and non-treetops, the minor errors could be corrected. Besides that, the FCN can learn high-level semantic features which are also robust in complicated scenarios. Hence, comparing to the THR, even it is trained from THR results, the FCN based treetop detector can have a fair performance and be more robust to complicated areas.

5. CONCLUSIONS

In the study, we use multi-view high-resolution satellite imagery derived DSM (Digital Surface Model) and orthophoto as research data to analyze the possibility of using a neural network to detect treetops. Since the training of CNN needs a large number of labeled samples, instead of manually labeling them, we generate the labels by finding treetops with the top-hat by reconstruction (THR) operation on DSM. For the deep learning network, we adopt the FCN (fully convolutional network) as a pixel-level classification network to segment the input image into treetops and non-treetops. We train the FCN with small and fixed-size patches, but due to the character of the fully convolutional layer, arbitrary size image can be the input of this network for treetop detection. Through the experiment, we proofed that the FCN based detector has a robust performance at various scenes due to its ability to learn high-level semantic features from various samples. And this study proved that the fully convolutional network can be trained with pseudo labels that from the non-supervised detector and achieve better performance.

There are still some errors in the FCN detection results such as that the treetops are too close to each other and miss-classify some object as treetops. We believe good training sample can lead better results and our next step is working on how to refine the training samples by the initial training result.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-163-2018 | © Authors 2018. CC BY 4.0 License.

ACKNOWLEDGEMENTS

The authors would like to thank John Hopkins University Applied Physics Lab for providing the Multi-view 3D Benchmark dataset used in this study. We would also like to thank Xing Pei and Ruopeng Wang from Lanzhou Jiaotong University for providing the tree labels.

REFERENCES

Bosch, M., Kurtz, Z., Hagstrom, S., Brown, M., 2016. A multiple view stereo benchmark for satellite imagery, *Applied Imagery Pattern Recognition Workshop (AIPR), 2016 IEEE*. IEEE, pp. 1-9.

Bosch, M., Leichtman, A., Chilcott, D., Goldberg, H., Brown, M., 2017. Metric evaluation pipeline for 3d modeling of urban scenes. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, 239.

Brandtberg, T., Warner, T.A., Landenberger, R.E., McGraw, J.B., 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in north america. *Remote sensing of Environment* 85, 290-303.

Chen, Q., Baldocchi, D., Gong, P., Kelly, M., 2006. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering & Remote Sensing* 72, 923-932.

Culvenor, D.S., 2002. Tida: An algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences* 28, 33-44.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, pp. 248-255.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 303-338.

Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016. Lidar detection of individual tree size in tropical forests. *Remote sensing of environment* 183, 318-333.

Gini, R., Passoni, D., Pinto, L., Sona, G., 2014. Use of unmanned aerial systems for multispectral survey and tree classification: A test in a park area of northern italy. *European Journal of Remote Sensing* 47, 251-269.

Hill, S., Latifi, H., Heurich, M., Müller, J., 2017. Individual-tree- and stand-based development following natural disturbance in a heterogeneously structured forest: A lidar-based approach. *Ecological Informatics* 38, 12-25.

Jakubowski, M.K., Li, W., Guo, Q., Kelly, M., 2013. Delineating individual trees from lidar data: A comparison of vector-and raster-based segmentation approaches. *Remote Sensing* 5, 4163-4186.

Kathuria, A., Turner, R., Stone, C., Duque-Lazo, J., West, R., 2016. Development of an automated individual tree detection model using point cloud lidar data for accurate tree counts in a pinus radiata plantation. *Australian Forestry* 79, 126-136.

Ke, Y., Quackenbush, L.J., 2011. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *International Journal of Remote Sensing* 32, 4725-4747.

Koch, B., Heyder, U., Weinacker, H., 2006. Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing* 72, 357-363.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097-1105.

Latifi, H., Fassnacht, F.E., Müller, J., Tharani, A., Dech, S., Heurich, M., 2015. Forest inventories by lidar data: A comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest. *International Journal of Applied Earth Observation and Geoinformation* 42, 162-174.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., Karteris, M., 2008. Object-based classification using quickbird imagery for delineating forest vegetation polygons in a mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing* 63, 237-250.

Mohan, M., Silva, C.A., Klauberg, C., Jat, P., Catts, G., Cardil, A., Hudak, A.T., Dia, M., 2017. Individual tree detection from unmanned aerial vehicle (uav) derived canopy height model in an open canopy mixed conifer forest. *Forests* 8, 340.

Özcan, A.H., Hisar, D., Sayar, Y., Ünsalan, C., 2017. Tree crown detection and delineation in satellite images using probabilistic voting. *Remote Sensing Letters* 8, 761-770.

Popescu, S.C., Wynne, R.H., Nelson, R.F., 2002. Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size. *Computers and electronics in agriculture* 37, 71-95.

Pouliot, D., King, D., 2005. Approaches for optimal automated individual tree crown detection in regenerating coniferous forests. *Canadian Journal of Remote Sensing* 31, 255-267.

Qin, R., 2014. Change detection on lod 2 building models with very high resolution spaceborne stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, 179-192.

Qin, R., 2016. Rpc stereo processor (rsp)–a software package for digital surface model and orthophoto generation from satellite stereo imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3, 77.

Qin, R., 2017. Automated 3d recovery from very high resolution multi-view satellite images, *ASPRS (IGTF) annual Conference*, Baltimore, Maryland, USA, p. 10.

Qin, R., Fang, W., 2014. A hierarchical building detection method for very high resolution remotely sensed images combined with dsm using graph cut optimization. *Photogrammetric Engineering & Remote Sensing* 80, 873-883.

Qin, R., Tian, J., Reinartz, P., 2016. 3d change detection–approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 41-56.

Quackenbush, L.J., Hopkins, P.F., Kinn, G.J., 2000. Using template correlation to identify individual trees in high resolution imagery, *American Society for Photogrammetry & Remote Sensing (ASPRS) 2000 Annual Conference Proceedings, Washington DC*.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, pp. 91-99.

Saarinen, N., Vastaranta, M., Näsi, R., Rosnell, T., Hakala, T., Honkavaara, E., Wulder, M., Luoma, V., Tommaselli, A., Imai, N., 2017. Uav-based photogrammetric point clouds and hyperspectral imaging for mapping biodiversity indicators in boreal forests. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 42.

Santoro, F., Tarantino, E., Figorito, B., Gualano, S., D'Onghia, A.M., 2013. A tree counting algorithm for precision agriculture tasks. *International Journal of Digital Earth* 6, 94-102.

Sherrah, J., 2016. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv preprint arXiv:1606.02585*.

Skurikhin, A.N., Garrity, S.R., McDowell, N.G., Cai, D.M., 2013. Automated tree crown detection and size estimation using multi-scale analysis of high-resolution satellite imagery. *Remote sensing letters* 4, 465-474.

Song, C., Dickinson, M.B., Su, L., Zhang, S., Yaussey, D., 2010. Estimating average tree crown size using spatial information from ikonos and quickbird images: Across-sensor and across-site comparisons. *Remote sensing of environment* 114, 1099-1107.

Strîmbu, V.F., Strîmbu, B.M., 2015. A graph-based segmentation algorithm for tree crown extraction using airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, 30-43.

Sun, X., Shen, S., Lin, X., Hu, Z., 2017. Semantic labeling of high-resolution aerial images using an ensemble of fully convolutional networks. *Journal of Applied Remote Sensing* 11, 042617.

Tarp-Johansen, M.J., 2002. Automatic stem mapping in three dimensions by template matching from aerial photographs. *Scandinavian journal of forest research* 17, 359-368.

Tsogkas, S., Kokkinos, I., Papandreou, G., Vedaldi, A., 2015. Semantic part segmentation with deep learning. *CoRR, abs/1505.02438*.

Turner, D., Lucieer, A., Watson, C., 2012. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (uav) imagery, based on structure from motion (sfm) point clouds. *Remote Sensing* 4, 1392-1410.

Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE transactions on image processing* 2, 176-201.

Wang, L., Gong, P., Biging, G.S., 2004. Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering & Remote Sensing* 70, 351-357.

Weng, E., Malyshev, S., Lichstein, J., Farrior, C., Dybzinski, R., Zhang, T., Shevliakova, E., Pacala, S., 2015. Scaling from individual trees to forests in an earth system modeling framework using a mathematically tractable model of height-structured competition. *Biogeosciences* 12, 2655-2694.

Wulder, M., Niemann, K.O., Goodenough, D.G., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of environment* 73, 103-114.

Zhao, D., Pang, Y., Li, Z., Liu, L., 2014. Isolating individual trees in a closed coniferous forest using small footprint lidar data. *International journal of remote sensing* 35, 7199-7218.