

INDIVIDUAL TREE CROWN DELINEATION FROM HIGH SPATIAL RESOLUTION IMAGERY USING U-NET

B. Hu *, W. Jung

Dept. of Earth and Space Science and Engineering, York University, Toronto, ON, Canada - (baoxin@yorku.ca;
wjung1008@gmail.com)

Commission III, WG I/1

KEY WORDS: Individual tree crown delineation, High spatial resolution imagery, Deep learning, U-net, Weighted, Segmentation

ABSTRACT:

The objective of this study was to explore the utilization of deep learning networks in individual tree crown (ITC) delineation, a very important step in individual tree analysis. Even though many traditional machine learning methods have been developed for ITC delineation, the accuracy remains low, especially for dense forests where branches, crowns, and clusters of trees usually have similar characteristics and boundaries of tree crowns are not distinct. Advance in deep learning provides a good opportunity to improve ITC delineation. In this study, U-net, Residual U-net, and attention U-net were implemented for the first time in ITC delineation. In order to ensure that the boundaries of tree crowns were classified correctly, a weight map was generated to give more weights to boundary pixels between two close crowns in the loss function. These three networks were trained and tested using optical imagery obtained over a study site within the Great Lakes-St. Lawrence forest region, Ontario Canada. Based on two test sites dominated by open mixed forest and closed deciduous forests, respectively, the overall accuracies were 0.94 and 0.90, respectively for U-net, 0.89 and 0.62 for Residual U-net, and 0.96 and 0.83 for attention U-net.

1. INTRODUCTION

Information on individual trees is required in a variety of forest-related activities, such as silviculture treatments, selective cuts, and biodiversity assessments. Advances in high spatial resolution remote sensing technologies make individual tree-based analysis feasible. Individual tree crown (ITC) serves as a basic unit for many useful activities such as species identification, gap analysis, and volume or biomass estimation. ITC delineation has thus attracted the attention and research activities of remote sensing communities, which has driven the development of various methods of ITC delineation from remote sensing data (Ke et al., 2011). However, it remains challenging to delineate tree crowns with complicated structures found in natural and mixed wood forests. Over-segmentation may occur due to that the branches and sub-crowns of a deciduous tree may resemble small trees; and the fact that deciduous tree crowns are often touching or close to each other, making between-crown valleys so invisible that a tree clump (a group of trees growing closed together) can be falsely detected as one crown, leading to under-segmentation.

With these existing methods, mostly based on traditional machine learning, hand-crafted features, either related to intensity continuity and/or discontinuity are employed. Deep learning, in contrast, involves automatic learning from examples, allowing features to be extracted directly from data. The potential of deep learning is therefore attracting a lot of attention in the field of remote sensing (Zhu et al., 2019). Specifically, in ITC delineation-related applications, a few studies have shown the promise of deep learning for detecting objects in remote sensing images (Zhu et al., 2019), but only one of these deals with the localization of tree crowns (Weinstein et al., 2019), and none with ITC delineation to be best of our knowledge. The objective of this study was to exploit the use of deep learning networks, specifically U-net (Ronneberger et al., 2015), Residual U-net

(Diakogiannis et al., 2019) and attention U-net (Oktay et al., 2018) in ITC delineation. Different implementations and configurations were attempted and compared based on optical imagery collected over a natural forest site within the Great Lakes-St. Lawrence forest region, Ontario Canada.

2. STUDY AREA AND DATA USED

A forest site near Sault Ste. Marie, Ontario, Canada, within the Great Lakes-St. Lawrence forest region was used. The study area consists of various-sized trees, bushes, grasses, and forbs. Located in the Great Lakes-St. Lawrence forest region near Sault Ste. Marie, Ontario, Canada, the study area ($46^{\circ}33'43'' - 46^{\circ}34'03''N, 83^{\circ}25'13'' - 83^{\circ}25'20''W$) consists of various-sized trees, bushes, grasses, and forbs. Aspen (*Populus tremuloides* Michx.), white birch (*Betula papyrifera* Marsh.), and sugar maple (*Acer Saccharum*) are the most common deciduous species. Jack pine (*Pinus banksiana* Lamb.) and black spruce (*Picea mariana* Mill. BSP) are the most common coniferous species. Stands have closed and multi-layered canopy structures and range from 30 to 80 years old.

The multispectral airborne imagery of the study area was acquired using an Illunis XMV-4021C camera in August 2009 at about 200 m above ground. Each acquired image has three broad spectral bands: blue (with centre wavelength of 450 nm), green (550 nm), and red (625 nm), and has a spatial resolution of 0.15 m. The optical images were geo-referenced using on-board GPS and inertial system. Figure 1 shows the true color composite of the optical imagery over the study area. For the training and validation purpose, the optical imagery was manually segmented by an independent and experienced researcher. Two plots representing typical mixed forest and deciduous forest were selected as test sites. The rest was used for training and validation.

* Corresponding author

3. METHOD

3.1 Preparation of training and test images

To overcome the limited training samples, the original image was randomly cropped into an image size of 128 by 128 pixels. These images were also randomly rotated and added to the training sets. In total, there were 102,400 training images. For each training image, there was a corresponding label image. Figure 2 shows the example of a pair of training image and its label.



Figure 1. The location of the study site (left) and the true colour composite of the optical imagery over this site (right). The two plots were used for independent testing.

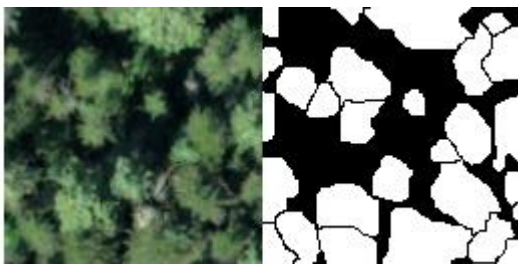


Figure 2. A sample of a pair of a training image and corresponding label with white as tree crowns.

3.2 The implementation of U-net, Residual U-net, and attention U-net

The U-net architecture used in this study is shown in Figure 3. The input was an image with the size of 128 by 128 and with 3 spectral bands. Each pixel in the image was classified into two categories: tree crowns and non-tree crowns (background). The weight map was generated using the equation proposed in Ronneberger et al. (2015) and shown in Equation (1) to force the network to learn features identifying tree crowns close to each other.

$$w(i, j) = w_0 \exp \left(-\frac{(d_1 + d_2)^2}{2\sigma^2} \right), \quad (1)$$

where $w(i, j)$ = the weight for pixel (i, j)
 d_1 = the distance to the border of the nearest crown
 d_2 = the distance to the border of the second nearest crown
 w_0, σ = two user-defined parameters for border weight and width, respectively.

w_0, σ were experimentally determined as 8 and 4, respectively. The effect of these parameters is further discussed in Section 4.

With the same configuration, Residual U-net (Diakogiannis et al., 2019) and attention U-net (Oktay et al., 2018) was also implemented.

In addition, a U-net without weight map was also implemented as comparison. A different implementation to weight the boundary pixels more was also attempted. Instead of generating the weight map, a labelled image with crown boundaries (vs. non-tree boundaries) was derived from the labelled tree crown image. The same network shown in Figure 3 (except for the output layer) was trained to classify tree crowns vs. non-tree crowns and crown boundaries vs. non crown boundaries. The loss functions corresponding to these two classifications could be weighted differently. The final tree crowns were determined by the differences between the detected tree crowns (assigned with a digital number 1) and crown boundaries (assigned with a digital number of 1). The results for this implemented were discussed in Section 4.

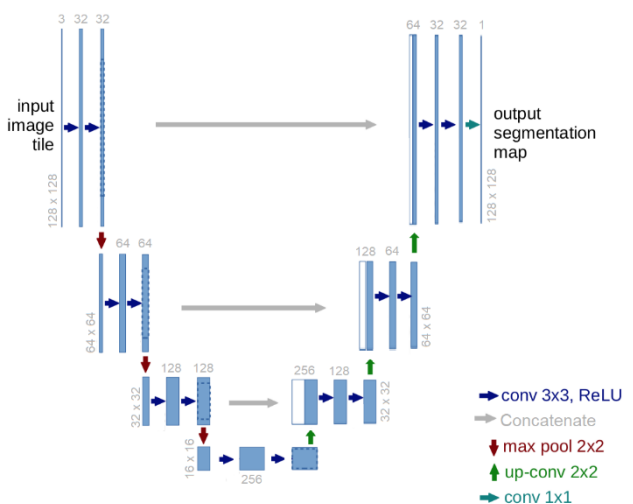


Figure 3. The U-net architecture used in this study.

3.3 Accuracy assessment

In addition to validation during the training processing, we also carry out independent test based on the two plots shown in Figure 1. Hereafter, the manually interpreted and automatically delineated segments are referred to as reference crowns and target segments, respectively. The overall accuracy and omission and commission errors were calculated based on the method proposed by Leckie et al (2003) and Jing et al, (2012). Each reference crown was assigned to each of the following categories based on its relationship with target segments.

- (1) Matched – for a reference crown and a target segment, if their respective overlaps exceeded 50%, the reference crown was considered as a crown matched by the target segment.
- (2) Nearly matched – for a reference crown and a target segment, if their overlaps exceeded 50% of only one segment, the reference crown was counted as a crown nearly matched by the target segment.
- (3) Missed – if a reference crown covered more than half the area of no target segment, the reference crown was considered

as a crown missed in the automatic delineation.

(4) Merged – if there were multiple reference crowns with more than half the area covered by a target segment, the multiple reference crowns were taken as crowns merged in the automatic delineation.

(5) Split – if there were multiple target segments with more than half the area covered by a reference crown, the reference crown was considered as a crown split in the automatic delineation.

Both the matched and the nearly matched reference crowns can be taken as the crowns correctly delineated by the method being tested; the missed and the merged reference crowns jointly respond to the omission errors of the target map; and the split reference crowns, together with the target segments covering more than half the area of no reference crown, represent the commission errors of the target map.

4. RESULTS

Figures 5 and 6 show the ITC delineation results using U-net test plot 1 and plot 2, respectively. The quantitative statistics is shown in Table 1.

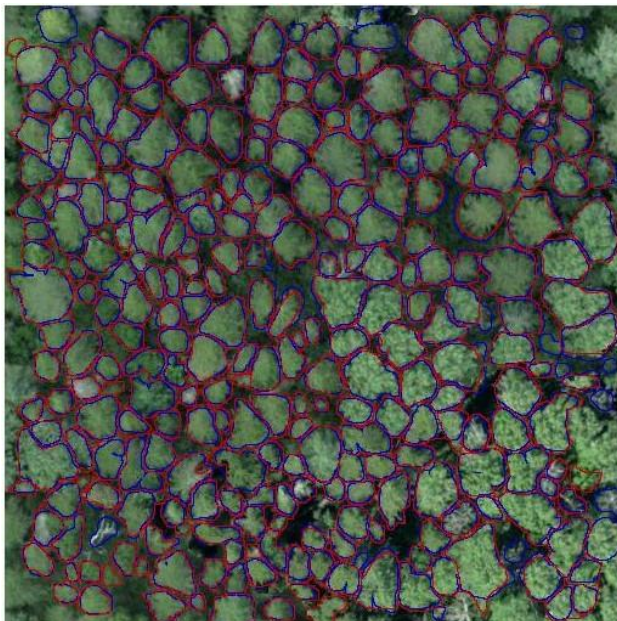


Figure 5. ITC delineation result (blue circle) generated by U-net, together with the tree crowns manually labelled (red circles) for the test size #1.

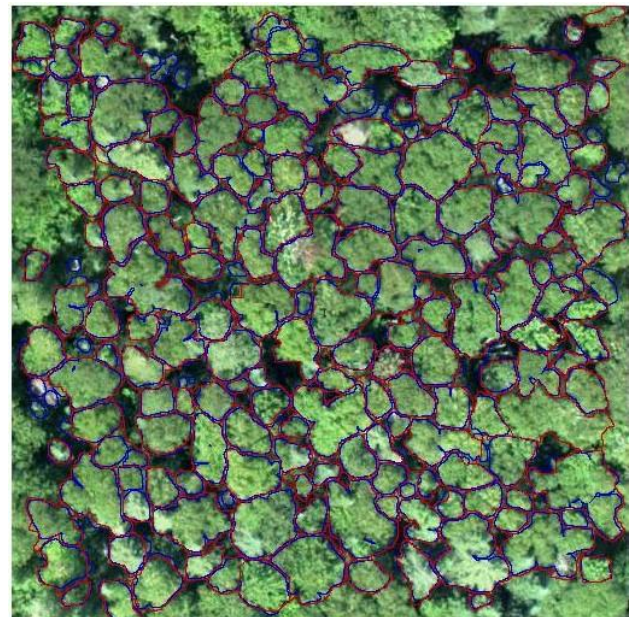


Figure 6. ITC delineation result (blue circle) generated by U-net, together with the tree crowns manually labelled (red circles) for the test size #2.

Most of reference tree crowns were delineated correctly, which was especially true for the test site 1. Based on the definitions in section 3.3, both the “Matched” and “Marginally matched” reference crowns were considered as correctly delineated. As a result, 256 tree crowns (94% of the total 273 crowns) in test site 1 and 151 crowns (90%) in test site 2 were correctly delineated, as shown in Table 1. Both visual observations and quantitative analysis revealed that the U-net implemented could delineate various-sized individual tree crowns in mixed wood and deciduous forests with accuracy comparable to manual interpretation. Further visual examination showed that most of the omitted crowns were low and small; most of the merged crowns belong to tree clusters containing no distinguishable between-crown valleys; and as for the split crowns, their sub-crowns were falsely taken as individual tree crowns. In addition, some of trees (14 and 19 for sites 1 and 2, respectively) were delineated by U-net, not manually. Visual examination indicated that some of them were indeed omitted by the reference.

Site	Matched	Marginally matched	Omitted	Merged	Split
1	256	0	13	4	0
2	151	1	5	3	8

Table 1: The accuracy statistics of the delineation tree crowns using U-net for test sites 1 and 2. The total number of reference crowns are 273 and 168 for these two sites, respectively.

The delineated tree crowns obtained by Residual U-net for these two plots are shown in Figures 7 and 8, respectively and the quantitative statistics is shown in Table 2.

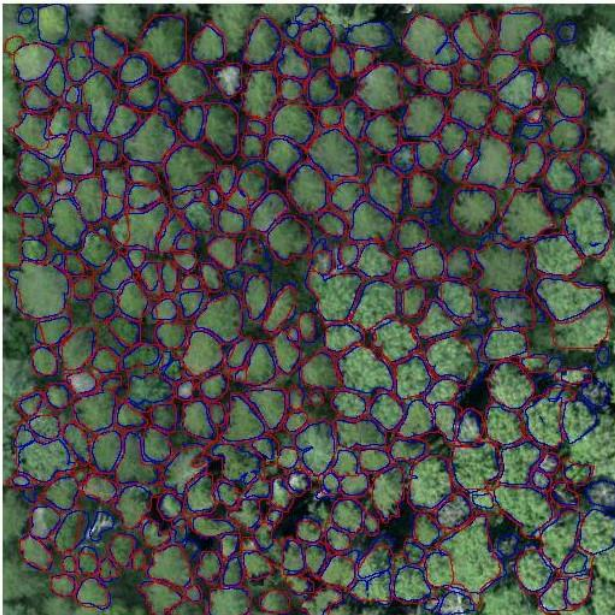


Figure 7. ITC delineation result (blue circle) generated by Residual U-net, together with the tree crowns manually labelled (red circles) for the test size #1.

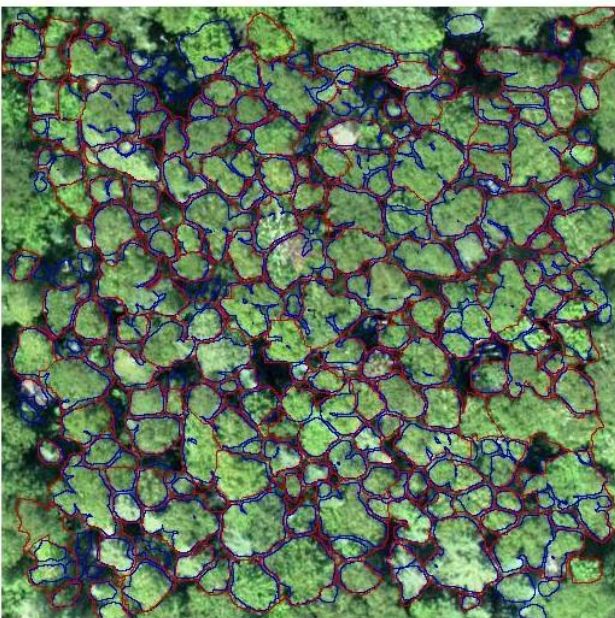


Figure 8. ITC delineation result (blue circle) generated by Residual U-net, together with the tree crowns manually labelled (red circles) for the test size #2.

Site	Matched	Marginally matched	Omitted	Merged	Split
1	235	8	17	10	3
2	92	13	33	7	23

Table 2: The accuracy statistics of the delineation tree crowns using Residual U-net for test sites 1 and 2. The total number of reference crowns are 273 and 168 for these two sites, respectively.

The results obtained by using Residual U-net were worse than those by using U-net, compared with the reference crowns. The decreasing in the accuracy was larger for site 2 than site 1. For the scene with dominantly open canopies (site 1), both networks

worked fine. However, for the scene with closed canopies (site 2), U-net worked better than Residual U-net. In addition, 22 and 44 crowns were delineated that were not part of reference crowns for test sites 1 and 2, respectively.

The delineated tree crowns obtained by attention U-net for these two plots are shown in Figures 9 and 10, respectively and the quantitative statistics is shown in Table 3. In addition, 17 and 14 crowns were delineated that were not part of reference crowns for test sites 1 and 2, respectively. Compared with the results obtained by U-net, the overall accuracy was slightly increased by using attention U-net for the test site 1 but decreased for the test site 2.

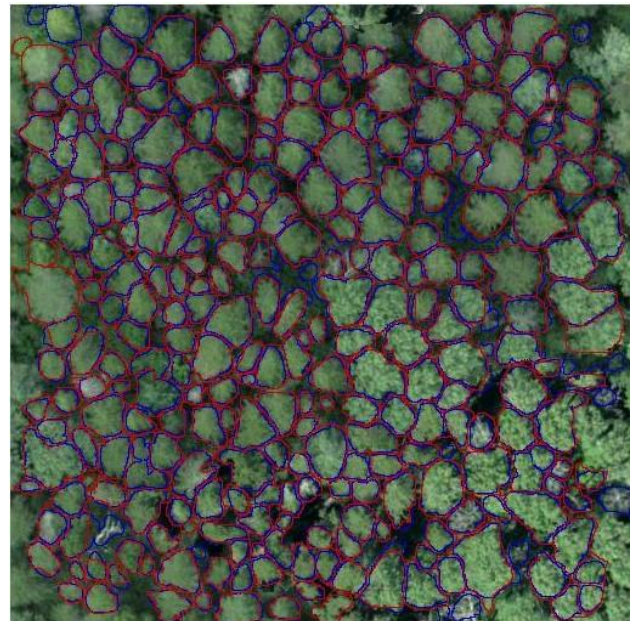


Figure 9. ITC delineation result (blue circle) generated by attention U-net, together with the tree crowns manually labelled (red circles) for the test size #1.

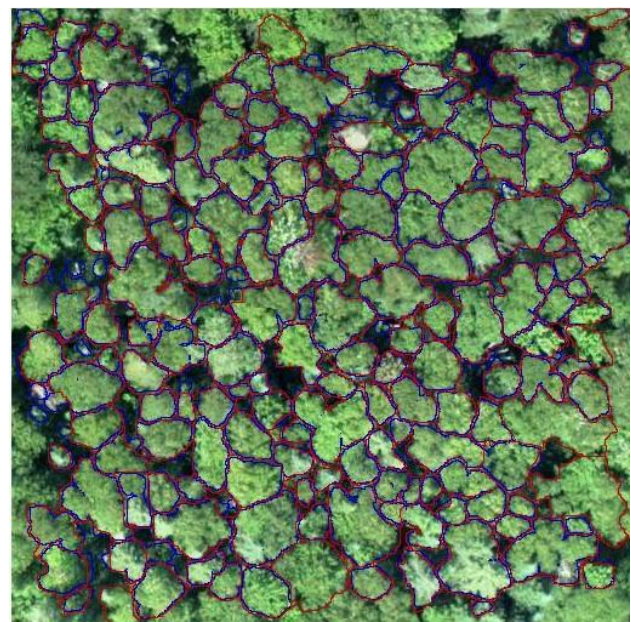


Figure 10. ITC delineation result (blue circle) generated by attention U-net, together with the tree crowns manually labelled (red circles) for the test size #2.

Site	Matched	Marginally matched	Omitted	Merged	Split
1	262	1	9	1	0
2	139	0	19	6	4

Table 3: The accuracy statistics of the delineation tree crowns using attention U-net for test sites 1 and 2. The total number of reference crowns are 273 and 168 for these two sites, respectively.

5. DISCUSSION

Among the three networks implemented in this study, overall accuracies in the ITC delineation obtained by U-net were the highest, and those by ResU-net were the lowest. Caution should be exercised when interpret the different results obtained from U-net and ResU_net. For the fair comparison, same number of epochs were used for all networks. Increasing the number of epochs may improve the performance in Residual U-net as it was originally designed to perform better when trained for a long period of time. Further analysis is warranted to examine the results in detail. The results between U-net and attention U-net were comparable. With the attention gates implemented in the attention U-net, attention U-net automatically suppressed features in background regions. For the test site 1, individual crowns were dominated and visible from the optical imagery and thus attention U-net performed better than U-net. For the test site 2, tree crowns were very close to each other, and individual crowns were, thus, not obvious. The attention gates might not capture the crowns. The delineation result was worse compared with that obtained from U-net. Different features and attention mechanism may be needed, which will be pursued in future work.

The accuracies obtained from the deep learning networks, specifically U-net and attention U-net were higher than those obtained by traditional machine learning methods (Jing et al., 2012 and Qiu et al., 2020) tested using the same data sets. This was especially true to the test site 2 dominated by dense deciduous forest. It is worth mentioning that the machine learning methods compared were unsupervised and thus no training data was required.

To seek the best configuration for the U-net, different window size for the convolution layers were tested and 3 by 3 filters generated the best results. With the increasing of the window size, more and more crowns were merged together, which was as expected. In future work, multi-scale networks will be explored.

Experiments carried out in this study to explore the ways to ensure the crown boundaries were classified correctly. As mentioned earlier, as a comparison, a basic U-net (without weight map) was implemented as well. The results showed that a loss function with more weights on the border pixels between tree crowns was beneficial. The delineation result (not shown here) without the weight map was very poor, and most the tree crowns were connected together. The weight map generated using the method proposed by Ronneberger et al (2019) was more effective than the implementation with dual mode of classification (crowns vs. non-crowns and crown boundaries vs. non crown boundaries). One reason with the relatively poor performance with the dual mode of classification might be the inaccurate localization of the boundary pixels generated from the labelled training image. This will be further investigated in the future study.

As mentioned earlier, to calculate the weight map (Ronneberger et al., 2019), two user-defined parameters were required, w_0 and σ (Equation 1). Different values were attempted in this study to investigate their impacts on the results. Table 3 summarizes the results with various w_0 , while σ was set as 5. Due to the limited space, the two categories of “matched” and “marginally matched” were merged. Table 4 summarizes the impact of σ .

site	w_0	Matched	Omitted	Merged	Split
1	6	239	21	10	3
	8	257	11	3	2
	10	253	13	6	1
2	6	120	24	9	15
	8	116	14	7	31
	10	100	40	12	16

Table 3: The accuracy statistics of the delineation tree crowns using U-Net for test sites 1 and 2. σ was set as 5.

site	σ	Matched	Omitted	Merged	Split
1	3	262	7	2	2
	4	256	13	4	0
	5	257	11	3	2
2	3	137	22	3	6
	4	152	5	3	8
	5	116	14	7	31

Table 4: The accuracy statistics of the delineation tree crowns using U-Net for test sites 1 and 2. w_0 was set as 8.

The results were sensitive to both w_0 and σ , especially for the scenes with complicated canopy structure (such as site 2). In addition, the selection of the values for these two parameters needs to be adaptive to local characteristics. For test site 1 where mix-forest was dominated and with open canopies, w_0 of 8 generated the best result (Table 3) while for test site 2 where dense deciduous trees were dominated, w_0 of 10 was preferred. When the value of w_0 was reduced, the width of the boundary became smaller. In other words, the gaps between tree crowns were decreased. A larger w_0 should be selected for sites with trees that are very close together, such as test site 2. Similarly, for σ , the value of 3 and 4 was the best for test site 1 and 2, respectively. A smaller σ value in close canopies, gaps between trees tends to be reduced, leading to that some tree crowns were merged.

6. CONCLUSIONS

In this study, three deep learning networks, U-net, Residual U-net, and attention U-net, were investigated in the ITC delineation. Based on the test results, U-net achieved a high accuracy for both sites (0.94 and 0.90, respectively), significantly higher than those obtained by traditional machine learning methods validated on the same study area (Jing et al 2012 and Qiu et al, 2020). All three networks could delineate tree crowns in the forest site of open canopies (site 1) with high accuracies, such as higher than 0.89; the highest accuracy was obtained by attention U-net (0.96). For the forest site with dense deciduous trees (site 2), the accuracy achieved by Residual U-net was lower (0.62), compared with those obtained by U-net (0.90) and attention U-net (0.83). Additional tests are required to determine the robustness of the trained networks for operational forest inventory purposes.

For the U-net, a weight map significantly improved the delineation accuracy, but the results were sensitive to the parameters employed to generate the weight map. In future work, we will seek better way to generate the weight map.

Even though the deep learning networks implemented in this study outperformed traditional machine learning method. The challenge of providing good quality label data was encountered. Effective strategies are needed to generate ground truth.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

- Diakogiannis, F. I., Waldner, F. et al., 2020. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. of Photogramm. Remote Sens.*, 162, 94-114.
- Jing, L., Hu, B., Li, J., and Noland, T., 2012. An individual tree crown delineation method based on multi-scale segmentation of imagery". *ISPRS Journal of Photogrammetry & Remote Sensing*, 70, 88-98.
- Ke, Y., Quackenbush, L.J. 2011. A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution Imagery. *Int. J. Remote Sens.* 32, 3625–3647.
- Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing* 29 (5), 1–17.
- Oktay, O., schlemper, J. et al., 2018. Attention U-Net: Learning Where to Look for the Pancreas, *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*, Amsterdam, The Netherlands.
- Qiu, L., Jing, L., Hu, B., Li, H. and Tang, Y., 2020. An improved multiscale method to delineate individual tree crowns from high resolution multispectral imagery of rainforest. *Remote Sensing*, 12(3), 585.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- Weinstein, B., S. Marconi, et al., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sens.* 11, 1309.
- Zhu, M., Y. He, et al., 2019. A review of researches on deep learning in remote sensing application. *International Journal of Geosciences*, 10, 1-11. doi: 10.4236/ijg.2019.101001.