The ISPRS International Contest on Individual Tree Crown Segmentation using High-Resolution Images and the Initial Findings

Xinlian Liang[1], Yinrui Wang [1], Jun Pan [1], Mi Wang [1], Jiansi Yang [1], Jianya Gong [2]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China
(xinlian.liang, wangyinrui, panjun1215, wangmi, jsyang, gongjy) @whu.edu.cn

Keywords: Individual Tree Crown, ITC, Instance Segmentation, Deep learning, Forests.

Abstract

Tree canopy plays an essential role in the biophysical activities in forest environment. During the past two decades, individual tree delineation using high-resolution imagery data has become a hot topic in forest sensing research. Individual Tree Crown (ITC) segmentation methods aim to generate masks that delineating the boundary of each ITC, which supports various tree parameter extractions. Thanks to the rapid development of deep learning, the ITC segmentation methods achieved remarkable improvement. However, existing research suffers from the limited availability of the datasets, and the lack of evaluation standards as well as task-orientated neural networks. In 2024, the International Society of Photogrammetry and Remote Sensing (ISPRS) launched the first International Contest of ITC Segmentation. The contest aims to reveal the state-of-the-art of the ITC method development using high-resolution images, to clarify the remaining barriers and challenges, and to guide further explorations in the field. This paper overviews the contest and reports the initial finding regarding the impact factors of the method performance.

1. Introduction

Tree canopy plays an essential role in the biophysical activities in forest environment. At an area scale, the horizontal and vertical structures of canopies significantly influence the radiation regime, vapor concentration, and temperature in forests. At an individual tree level, the position, size, and geometry of a canopy are important structural characteristics that indicate the tree's growth, function, and value.

Individual tree crown (ITC) from remote sensing data, which provides boundaries and structural characteristics of each detected tree, has become a prominent research topic in forest remote sensing over the past two decades. ITC segmentation is specifically targeted to the extraction and segmentation of the tree crowns from earth observation products such as the aerial-images from cameras and the canopy height models (CHM) from LiDAR (light detection and ranging) sensors. Outcomes of ITC segmentation are typically the outer boundaries of trees in a forest, which can be used to generate masks for further analysis to estimate various tree parameters.

Prompted by more and more affordable and convenient earth observation solutions, the interests in ITC using high-resolution images has grown rapidly (Dersch et al., 2024; Hao et al., 2021; Li et al., 2024; Mohamed Barakat A. Gibril and Sachit, 2022; Sani-Mohammed et al., 2022; Sun et al., 2022; Troles et al., 2024; Ventura et al., 2024, 2024; Xie et al., 2024; Yang and Li, 2023; Ye Zhang and Li, 2024). Specifically aided by continuous advancements in deep learning, ITC segmentation methods have achieved remarkable improvements through instance segmentation networks. such as Mask RCNN (He et al., 2017), Cascade Mask R-CNN (Cai and Vasconcelos, 2021), etc.

Due to its well-documented effectiveness, Mask R-CNN, which includes a feature extraction backbone, a detection head to generate the bounding box, and a mask head to delineate the target, is one of the most commonly used instance segmentation networks for ITC segmentation and has become a baseline for benchmarking other networks in computer vision.

The Mask R-CNN model is being actively developed. Some modified its backbone to improve accuracy. (Sani-Mohammed et al., 2022; Ye Zhang and Li, 2024) added the Feature Pyramid Network (FPN) to the backbone to build multi-scale feature maps to improve the segmentation of multi-scale ITCs. (Troles et al., 2024) add a transformer-based context enhancement module for finer context information extraction. Cascade R-CNN generalized the cascade architecture to Mask R-CNN to improve both the detection and mask heads, where the higher accuracy is at the expense of lower efficiency.

Nevertheless, common barriers in the this field include limited reference datasets, a lack of task-oriented neural networks, and the absence of evaluation standards.

In 2024, the International Society of Photogrammetry and Remote Sensing (ISPRS) launched the first International Contest of ITC Segmentation. The contest aims to reveal the state of the art of methods for ITC using high-resolution earth observation images, identify remaining barriers and challenges, and guide further exploration in the field.

In the contest, the participants develop their own methods to perform the ITC delineation, test developed methods over the standard datasets, submit their results, and document the method. The results are evaluated using the standardized method to benchmark the performances of different methods.

This paper overviews the contest, e.g., the datasets, participants, and evaluation method, and reports the initial finding regarding the impact factors of the method performance.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-637-2024

2. Material and Methodology

This section summarizes the study area, the released materials, the contest setup, and the evaluation method.

2.1 Study Area

In the contest, the test data were collected from 11 study areas from multiple climate zones around the world. The forests have diverse types and conditions. The study areas located in 9 countries, i.e., America, Australia, Canada, China, German, Kenya, Malaysia, Norway, and Panama. The test data cover diverse forest types, including the tropical, sub-tropical, temperate, and boreal forest. Besides, both urban and forested areas are considered.

Figure 1 illustrates the distribution of the study areas and example data.

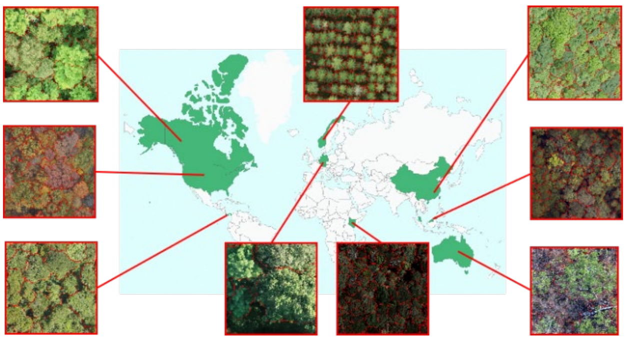


Figure 2. The distribution of the study areas and example data

2.2 The Released Material

The released materials include two parts, i.e., the datasets and sample code.

The high-resolution images were collected from airborne and drone platforms. The orthophotos were randomly cropped into patches of 1024×1024 pixels. All visible individual trees in the images were manually labeled by visual interpretation. In total, the dataset includes more than 11,000 images and 600K ITC masks.

Among the data, two datasets exist only in the training set, i.e., the dataset 1 and 2, and two datasets exist only in the testing set, i.e., the dataset 10 and 11. This setup helps to evaluate the transferability of the developed segmentation methods.

Table 1 summarized the detailed information of datasets.

No.	Area	Resolution (cm)	Dataset		
			Training	Validation	Testing
1	Canada	2.0	1691	—	—
2	Malaysia	10.0	331	—	—
3	Panama	4.5	1200	275	600
4	China	10.0	400	100	200
5	China	2.0	1721	441	786
6	China	3.0	1234	—	346
7	America	5.0	184	—	100
8	Kenya	10.0	300	—	200
9	Norway	10.0	206	—	100
10	German	2.0	—	—	468
11	Australia	2.0	—	—	200
Total	—	—	7267	816	3000

Table 1. The details of the datasets in the contest

A sample code for ITC segmentation network training, validation, testing, and inference was released to the participants as a reference to develop their own methods and to understand the contest setups. The sample code utilizes the Mask R-CNN (He et al., 2017) for ITC segmentation, one the most popular instance segmentation network. The method takes the PNG images and the annotation stored in MS COCO Format as inputs. The predicted ITC masks are stored in MS COCO Format. The sample code was modified based on the detecron2 (Wu et al., 2019), which provide a library for instance segmentation algorithm development based on the Mask R-CNN.

Figure 2 presents the example of the released materials.

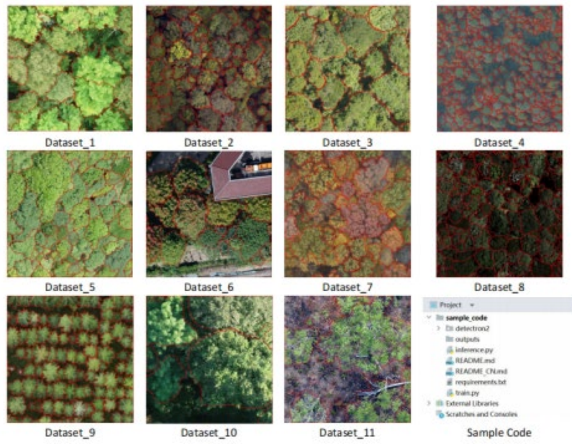


Figure 2. The examples of the released materials

2.3 Contest Setup

The contest is carried out in 2024 and is comprised of two phases. Participants develops and improves their own models during the evaluation phase and test the model performance in the testing phase. The submitted results are ranked in real time in the web-based system. The top six teams submitted all required materials on time are rewarded as the contest winner.

2.3.1 Phase: The contest is comprised of two phases, i.e., the evaluation and testing phase. During the evaluation phase, the participants develop their methods using the training dataset. The inference results are automatically evaluated using the validation set. The developed method can be improved from the evaluation results and the real-time ranking.

During the testing phase, the participants submit the inference results based on the testing set. The feedback of testing results and real-time ranking are provided similarly as what is in the previous evaluation phase, for individual test data sets and for all test datasets as a whole. Thus, the participants are able to fine tune their models during the testing phase.

Among the two phases, the evaluation phase is for the method development that is not compulsory to participate. The testing phase is compulsory for all participants who want to have a solid score/rank.

2.3.2 Schedule: The Contest was launched in Jan. 2024 and the testing phase ended in Jul. 2024. The awarding ceremony is in the ISPRS Technical Commission III Mid-term Symposium on Nov. 4th, 2024.

https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-637-2024

Table 2 reports the schedule of the contest.

Time	Events
Jan. 29th, 2024	The ISPRS ITC Segmentation Contest launched, the training phase started.
Jun. 17th, 2024	The testing phase started.
Jun. 22nd, 2024	The score/results were checked.
Jul. 15th, 2024	Outcome Announcement
Nov. 4th, 2024	The awarding ceremony in the ISPRS Technical Commission III Mid-term Symposium.

Table 2. The schedule of the ISPRS ITC Segmentation Contest

2.3.3 Ranking and Awarding: The submitted results are ranked in real time in the web-based system, which can be followed by all participants at the same time.

The models of the top six teams are validated and their outcomes are replicated before the announcement of the final results. The top six teams are rewarded as the contest winners. The teams win the first-, second-, and third-place receive the prize $3k, $2k, and $1k, respectively. One team that significantly contributes to the success of the contest is rewarded the $1k contribution prize.

It is worth of noting that the ranking aims to encourage continuous studies, instead of a rigid judgment on model performances.

2.4 Participants

More than 40 teams from 13 countries participated the contest, e.g. Australia, America, Canada, China, Czech, Denmark, France, German, Italy, Netherlands, Norway, Switzerland, and UK. The participants are from universities and companies, and also those who are independent researchers.

Figure 3 illustrates the geographical distribution of the participants of this contest.

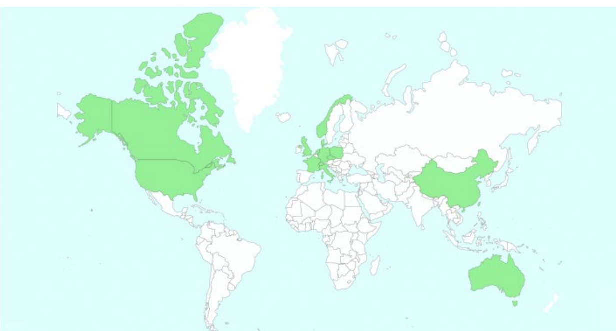


Figure 3. The geographical distribution of the contest participants

2.5 Method of Accuracy Evaluation

The performances of the ITC segmentation methods are evaluated using the Average Precision (AP) in the validation and testing stage, as show in (1) - (3).

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 PdR \quad (3)$$

where TP, FP, FN indicate True Positive, False Positive, and False Negative, respectively; AP represents the areas of the P-R curve, which indicates the correspondence between Precision (P) and Recall (R). The prediction whose IoU is greater than certain threshold is considered as TP, where the IoU represents the ratio of the intersection area and the union area between the prediction and ground truth. The IoU threshold 50% was used in the contest.

3. Results

Figure 4 illustrates the average results in the testing phase from top 20 teams in the ranking list. Note that the rank in the contest is solely based on the AP50 metric. The AP75 results are for comparison and research purposes.

As illustrated in the Figure 3, the state-of-the-art ITC segmentation methods achieve a 50% AP50 and 30% AP75 accuracy.

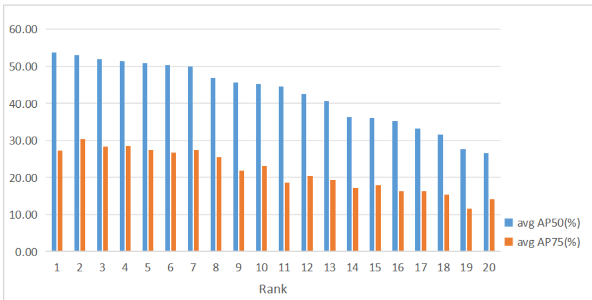
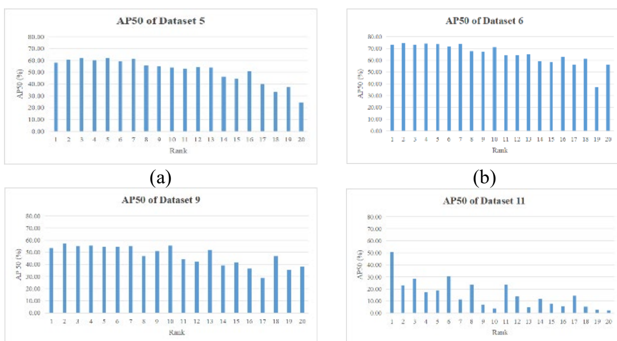


Figure 4. The average segmentation results from the top 20 teams in the testing phase

Figure 5 illustrates the results, i.e., AP50, of four datasets of different forest conditions, i.e., the dataset 5, 6, 9, and 11 from the sub-tropical, urban, boreal, and tropical forest, respectively.



(a) (b) (c) (d)

Figure 5. The average AP50 of 4 datasets of different forest conditions, i.e., dataset 5, 6, 9, and 11 in (a-d), respectively

The dataset 6, i.e., the urban forests, reported the highest accuracy at the 70% level. And the dataset 11, a sparse forest where the forest condition is not represented in the evaluation phase, reported the lowest accuracy.

Figure 6 illustrates the ground truth and the segmentation results of the dataset 5, 6, 9, and 11 in (a-d), respectively.

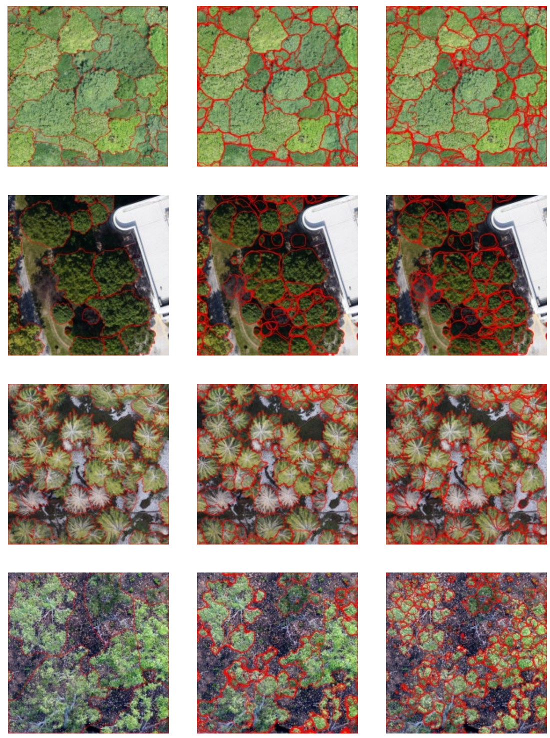


Figure 6. Example of the ground truth and the segmentation results of dataset 5, 6, 9, and 11 in (a-d), respectively. The ground truth and the segmentation results from two teams are illustrated in columns 1-3, respectively.

4. Discussion

ITC segment results differentiate on the basis of data quality, forest condition, method, and evaluation metric.

4.1 The Overall Accuracy and Transferability

Different datasets have clear different accuracy according to the results in Figure 5. The overall AP50 metrics of Dataset 5, 6, and 9 surpass the Dataset 11.

The AP50 in different test areas is between 50-70%. Among the test sites, the urban forests achieve the highest accuracy among all conditions. The AP50 of the top results are approximately 70%. In general, regular maintenance and improvement take place in urban forests, e.g., for fire prevention, landscape planning, and many other management purposes. Consequently, the density and structure complexity of urban forests are typically lower in comparison with the forested areas. Thus, urban forests present a favorable condition for the ITC segmentation.

The dataset 11 reported a low accuracy, i.e., lower than 30% AP50 for most methods. This test site has a sparse forest. However, this forest condition is not represented in the evaluation phase, as described in section 2.2. The methods developed based on other forest conditions appear to be less applicable in this scenario that they have never met.

Another dataset that is not included in the evaluation phase is the dataset 10. Even though methods have not been trained based on data from this forest, the segmentation results were relatively good, e.g., approximately 50% AP for the top teams. This is most probably because this forest has a similar condition as a few other forests involved in the evaluation phase, e.g., dataset 3.

These results indicate that the current methods can be directly applied to datasets that have not been included in the training data, while the method transferability is not guaranteed in the most current models, just as in many other types of instance segmentation tasks in machine learning.

4.2 The Impacts of Forest Conditions

The performances of the ITC segmentation methods from high-resolution images are significantly impacted by forest conditions. This impact may or may not be significant in an individual test area. However, it becomes clear in diverse forest conditions.

The result differences between the top teams were not significant, according to the average AP50 results from all test areas shown in Figure 4. Similarly, it is also not significant in some individual areas, e.g., in dataset 5, 6, and 9. However, the differences become clear in the dataset 11. The method ranked in the first place in the final list reported a clear better result in comparison with other methods in this dataset, which made it led in the overall performance.

These results indicate the importance and contribution of diverse forest conditions, or in general the testing data, to the evaluation of the method performance. Given limited test data or conditions, it is hard to rationally reveal the algorithm performance.

4.3 The Evaluation Metric

The evaluation in this contest uses the same evaluation metric as what is in most deep learning research, i.e., AP50, as it has become a standard evaluation metric. Thus, the results reported in this contest are in the same contest as what is in any other instance segmentation research.

On the other hand, it is still an open question whether this particular evaluation metric is ideal for forest research. In forest-related disciplines, such as forest science, silviculture, and ecology, the tree position and crown size information have their own values in both practice and research. In addition, they are directly linked to tree growth stage and forest structure, and are further employed in the forest operation planning and research, e.g., to reveal the tree- and stand-level attributes, conditions, and functions.

As AP50 depends on IoU 50 metric, where a true positive refers to that the ratio between the intersection and union area of the predicted and ground-truth tree crown is larger than 50%. Consequently, the question still needs to be answered whether the instance segmentation results, evaluated by the AP50 metric, can answer important forest related questions.

4.4 The Influence of Image Data Quality

The image quality also influences the segmentation results. In the contest, the results from the datasets with lower resolution, e.g., dataset 4 and 8, reported lower accuracy in comparison

https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-637-2024

with those datasets with higher spatial resolution. The low-resolution lead to deficient texture information, which in turn influences the method performances.

5. Conclusion

This paper overviews the first ISPRS International Contest of Individual Tree Crown (ITC) Segmentation using high-resolution images, and reports the initial findings from the contest.

The forest conditions significantly impact ITC segmentation results. For the urban forest, 70% AP50 can be achieved. For the forested area, the state-of-the-art methods can reach 50% AP50 and 30% AP75 accuracy. The ITC segmentation results are also influenced by the image quality. The favorable image resolution is above 10 cm. More detailed analyses regarding the interaction between the neural networks design, image quality, and forest condition will be carried out and reported in future.

Further research is encouraged to develop application-orientated neural networks, to explore the method transferability across different forest conditions, and to discover more applicable evaluation matrix.

Acknowledgements

This work was supported by the National Science Fund for Distinguished Young Scholars grant number 62425102, National Key Research and Development Program of China grant number 2023YFF1303901.

References

Cai, Z., Vasconcelos, N., 2021: Cascade R-CNN: High Quality Object Detection and Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 1483–1498. https://doi.org/10.1109/TPAMI.2019.2956516

Dersch, S., Schöttl, A., Krzystek, P., Heurich, M., 2024: Semi-supervised multi-class tree crown delineation using aerial multispectral imagery and lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing* 216, 154–167. https://doi.org/10.1016/j.isprsjprs.2024.07.032

Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021: 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. https://doi.org/10.1016/j.isprsjprs.2021.06.003

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017: Mask R-CNN, in: *2017 IEEE International Conference on Computer Vision (ICCV)*. pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322

Li, Z., Deng, X., Lan, Y., Liu, C., Qing, J., 2024: Fruit tree canopy segmentation from UAV orthophoto maps based on a lightweight improved U-Net. *Computers and Electronics in Agriculture* 217, 108538. https://doi.org/10.1016/j.compag.2023.108538

Mohamed Barakat A. Gibril, S.J. bin H., Helmi Zulhaidi Mohd Shafri, Abdallah Shanableh, Rami Al-Ruzouq, Aimrun Wayayok, Sachit, M.S., 2022: Deep convolutional neural networks and Swin transformer-based frameworks for individual date palm tree detection and mapping from large-scale UAV images. *Geocarto International* 37, 18569–18599. https://doi.org/10.1080/10106049.2022.2142966

Sani-Mohammed, A., Yao, W., Heurich, M., 2022: Instance segmentation of standing dead trees in dense forest from aerial imagery using deep learning. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 6, 100024. https://doi.org/10.1016/j.ophoto.2022.100024

Sun, Y., Li, Z., He, H., Guo, L., Zhang, X., Xin, Q., 2022: Counting trees in a subtropical mega city using the instance segmentation method. *International Journal of Applied Earth Observation and Geoinformation* 106, 102662. https://doi.org/10.1016/j.jag.2021.102662

Troles, J., Schmid, U., Fan, W., Tian, J., 2024: BAMFORESTS: Bamberg Benchmark Forest Dataset of Individual Tree Crowns in Very-High-Resolution UAV Images. *Remote Sensing* 16. https://doi.org/10.3390/rs16111935

Ventura, J., Pawlak, C., Honsberger, M., Gonsalves, C., Rice, J., Love, N.L.R., Han, S., Nguyen, V., Sugano, K., Doremus, J., Fricker, G.A., Yost, J., Ritter, M., 2024: Individual tree detection in large-scale urban environments using high-resolution multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation* 130, 103848. https://doi.org/10.1016/j.jag.2024.103848

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019: Detectron2.

Xie, Y., Wang, Y., Sun, Z., Liang, R., Ding, Z., Wang, B., Huang, S., Sun, Y., 2024: Instance segmentation and stand-scale forest mapping based on UAV images derived RGB and CHM. *Computers and Electronics in Agriculture* 220, 108878. https://doi.org/10.1016/j.compag.2024.108878

Yang, B., Li, Q., 2023: Individual tree crown extraction of natural elm in UAV RGB imagery via an efficient two-stage instance segmentation model. *Journal of Applied Remote Sensing* 17, 044509. https://doi.org/10.1117/1.JRS.17.044509

Ye Zhang, C.M., Moyang Wang, Joseph Mango, Liang Xin, Li, X., 2024: Individual tree detection and counting based on high-resolution imagery and the canopy height model data. *Geo-spatial Information Science* 0, 1–17. https://doi.org/10.1080/10095020.2023.2299146