Comparing Deep Learning and MCWST Approaches for Individual Tree Crown Segmentation

Wen Fan [1], Jiaojiao Tian [1, 2], Jonas Troles [3], Martin Döllerer [1], Mengistie Kindu [1], Thomas Knoke [1, *]

[1] Institute of Forest Management, TUM School of Life Sciences Weihenstephan,
Technical University of Munich, 85354 Freising, Germany, (wendy.fan, jiaojiao.tian, doellerer, mengistie, knoke) @tum.de
[2] German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), 82234 Wessling, Germany, Jiaojiao.Tian@dlr.de
[3] University of Bamberg, Cognitive Systems Group, 96049 Bamberg, Germany, jonas.troles@uni-bamberg.de

Keywords: UAV imagery, Mask R-CNN, Levelset-Watershed, Individual tree crown segmentation, Instance segmentation.

Abstract

Accurate segmentation of individual tree crowns (ITC) segmentation is essential for investigating tree-level based growth trends and assessing tree vitality. ITC segmentation using remote sensing data faces challenges due to crown heterogeneity, overlapping crowns and data quality. Currently, both classical and deep learning methods have been employed for crown detection and segmentation. However, the effectiveness of deep learning based approaches is limited by the need for high-quality annotated datasets. Benefiting from the BaKIM project, a high-quality annotated dataset can be provided and tested with a Mask Region-based Convolutional Neural Network (Mask R-CNN). In addition, we have used the deep learning based approach to detect the tree locations thus refining the previous Marker controlled Watershed Transformation (MCWST) segmentation approach. The experimental results show that the Mask R-CNN model exhibits better model performance and less time cost compared to the MCWST algorithm for ITC segmentation. In summary, the proposed framework can achieve robust and fast ITC segmentation, which has the potential to support various forest applications such as tree vitality estimation.

1. Introduction

The detection of individual tree crowns (ITC) contributes to assessing the vitality of trees, thereby plays an important role in forest management (Kempf et al., 2021). High-vitality trees will respond and recover from drought or other physiological stressors, which decreases climate-induced tree mortality (Gonzalez et al., 2010). Over the past decade, the utilization of very high-resolution imagery has emerged as a viable method for the identification of individual trees. This boosts various applications, such as assessing tree vitality, monitoring forest disturbances, and conducting forest inventories (Pearse et al., 2020). Although traditional methods of surveying forests are highly accurate, their results may not fully capture the current condition of large areas of forest. Remote sensing is less labor-intensive and makes it possible to collect large amounts of data while reducing the time and resources required for fieldwork. However, there are still challenges in the automated segmentation of ITCs using remote sensing data. Image resolution, overlapping crowns and training datasets are the three main challenges. Insufficient image resolution hinders segmentation at the level of individual trees. Even with commercial satellite imagery with a decimeter-level resolution, it is difficult to detect smaller tree crowns. In forests, the density of trees makes it difficult to delineate individual crowns within overlapping crowns, making even manual annotation difficult. Therefore, obtaining training datasets not only incurs data collection costs but also requires significant time and manual effort for annotation. This scarcity is reflected in the limited availability of publicly available forest datasets.

Different types of remote sensing data are used for ITC segmentation. Airborne Laser Scanning (ALS) uses an active sensor capable of collecting high-resolution 3D information on trees to identify ITC (Holmgren et al., 2022). Dong et al. (2020a) utilized the local maximum filtering method and the MCWST method to extract accurate individual tree crown parameters in orchards with complex backgrounds. However, this method faces challenges in extracting small trees and distinguishing between tree crowns and weeds in the image. Guo et al. (2021) proposed a method that combined multi-radius filtering and fusion with random forest to locate tree crowns. However, it also resulted in false extractions due to similar texture and color features between weeds and tree crowns. In recent years, UAV has been introduced to forest monitoring, as it can produce high-resolution data products quickly and conveniently, covering areas of several square kilometers (Ma et al., 2022). UAVs not only have the flexibility to acquire high spatial resolution data but are also inexpensive. Compared to 30cm panoptic resolution of the commercial satellite Airbus Pleiades Neo, UAV imagery offers a higher 1cm spatial resolution, such as 0.6cm resolution FORTRESS (Schiefer et al., 2020), allowing the detection of smaller tree crowns. In comparison to manned aircraft systems, UAVs can be deployed more faster and at lower costs (Huang et al., 2018).

Common traditional methods based on high-resolution images include template matching and region-growing methods (Yu et al., 2022). Generally, template matching assumes the shape and size of the crown. Region-growing searches for local maxima points as the centers of tree crowns, and the model performance is highly dependent on the selection of seed points. Currently, the watershed algorithm is the widely used method for individual tree segmentation (Dong et al., 2020b). The watershed algorithm flexibly adapts to the shapes and structures of trees in different images but it is sensitive to the edge information of the target. Therefore, it performs well in segmenting trees and background information. However, due to the limitations of local image information, it has the disadvantage of being prone to over-segmentation (Wallace et al., 2021). Furthermore, due to the dense distribution of trees in forests, the segmentation of overlapping tree crowns is a critical challenge. Recently, re-

* Corresponding author

searchers have proposed a combination of normalized cut and watershed algorithms to improve the accuracy of tree segmentation (Qin et al., 2022). Kempf et al. (2021) used a level-set watershed segmentation transformation to obtain crown contours from the DSM, which corrects or avoids contours resulting from watershed segmentation being in the gaps between crowns.

In recent years, there has been significant progress in object detection tasks on remote sensing data with the advent of deep learning algorithms (Kotaridis and Lazaridou, 2021). Convolutional Neural Networks (CNNs) have the ability to capture spatial features and textures related to target categories or quantities. Hence, CNNs have been widely applied in forestry and remote sensing, such as crown detection, individual tree segmentation, and tree species classification (Kattenborn et al., 2021). Weinstein et al. (2019) utilizes a ResNet-50-based CNN classifier to detect trees in airborne RGB forest imageries. The output of the detection covers bounding boxes which indicates the extent of individual trees. However, it does not provide the specific contours of the trees. Only pixel-level tree contours allow further exploration of applications such as the performance of different bands in reflecting variations in individual tree vitality. Conversely, Mask R-CNN has the ability to recognize objects at a pixel level and provide a comprehensive outline of the object. Therefore, some researchers have explored ITC segmentation based on Mask R-CNN with remote sensing data, such as UAV images (G. Braga et al., 2020; Hao et al., 2021; Ball et al., 2023). Hao et al. (2021) annotate 1605 trees in a 4 ha plot, with 16% of the data reserved for testing and the remaining data used for training and validation. While the model performed well in the experiment, the presence of a single species (spruce) and a large training set could lead to the overfitting of the model. G. Braga et al. (2020) used synthetic data for training and tested in a relatively smaller region.

Without a sufficiently annotated ITC dataset, it is quite challenging to evaluate the deep learning based approaches. Some researchers claim that the traditional algorithms or deep learning methods that they use have advantages in terms of model performance and time cost for individual tree segmentation in UAV imagery. However, it is difficult to assess which method is superior is difficult due to significant differences in the environments and sizes of the datasets used in these studies. Hence, it is essential to evaluate the overall performance of both approaches in the same forests. Fortunately, more research institutes or companies are willing to share their datasets. Troles et al. (2023) created a tree inventory based on RGB and multispectral UAV data, including individual crown contours and tree vitality assessments.

To address this challenge, this paper proposes a three-step method that (1) uses Faster R-CNN to obtain initial detection boxes from a custom Common Objects in Context (COCO) dataset, (2) performs Individual Tree Crown (ITC) segmentation based on MCWST, and (3) accomplishes ITC segmentation using Mask R-CNN. The remaining sections of this paper are organized as follows: Section 2 introduces the UAV data tested in the experiment and the corresponding dataset generation. Section 3 describes the application of these two methods to ITC segmentation. The fourth part contains experimental results and discussions. Finally, section 5 concludes the paper. The main contributions of this study are as follows:

(1) The refinement of the previous Marker Controlled Watershed Transformation (MCWST) segmentation approach by introducing the deep learning model for tree location detection.

(2) The comparison of model performance between traditional methods and deep learning approaches under the same input conditions, providing individual tree information for subsequent analysis of forest disturbance at the individual tree level.

2. Method

In this study, the MCWST algorithm and Mask R-CNN are used for ITC. In terms of data processing, the dataset used utilizes the pre-processed Canopy Height Model (CHM) images and RGB images captured by UAV as inputs. For both methods, we first use Faster R-CNN to obtain initial detection boxes for trees in the images. This approach helps to mitigate the over-segmentation problems caused by traditional methods using blob detection for seed point acquisition, and it additionally also ensures that MCWST and Mask R-CNN share the same initial input. For the MCWST method, detection is performed with the center of the detection box as the starting point on the DSM image. In the case of Mask R-CNN, the instance segmentation is conducted based on the initial detection boxes on RGB images. The key steps of the proposed method are described as follows.

2.1 Faster R-CNN Based Tree Location Detection

The Faster R-CNN (Ren et al., 2015) process is illustrated in Figure 1. The main steps are (1) feature extraction and fusion through ResNet and RPN networks, resulting in the generation of five feature maps, (2) the feature maps are fed into the RPN, generating anchors of different sizes to produce a given number of region proposals, (3) extract Region of Interest (RoI) feature maps from the CNN feature map based on the coordinates of the region proposals, (4) utilize the RoI Pooling layer for spatial pooling to ensure that the output size of all feature maps is the same. (5) input all feature maps into subsequent Fully Connected (FC) layers for classification and regression.

Faster R-CNN is employed to obtain initial bounding boxes on the dataset. Faster R-CNN utilizes a deep neural network to generate region proposals for object detection, followed by classification and bounding box regression on these proposed regions. Both ITC detection methods are based on the results of Fast R-CNN. Specifically, the design of the MCWST algorithm considers the centers of the Faster R-CNN detection boxes as initial seed points for segmentation. Mask R-CNN extends Faster R-CNN by adding additional branches for instance segmentation. Then we intend to evaluate the performance of these two algorithms on the same initial detection boxes.

2.2 MCWST Segmentation Algorithm

In this paper, we adopt the method proposed by Kempf et al. (2021), which primarily involves the modified DSM-level set watershed segmentation. The Chan-Vese (CV) model is first employed to segment the CHM into foreground and background. The blob detection is used to acquire the initial seed points, and the MCWST segmentation method is used for individual tree segmentation. The CV model achieves image segmentation based on energy minimization of level set functions. In this process, the level set function evolves and approximates the boundaries of the foreground and background in the image. Thus, the image is segmented into foreground and background. MCWST is an improvement based on the watershed segmentation method. The Watershed segmentation simulates

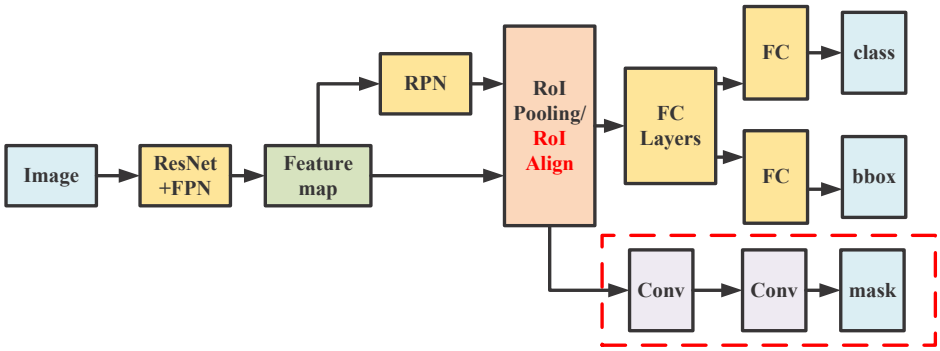

Figure 1. The Workflow of Faster R-CNN and Mask R-CNN.

the flow of water starting from local minima in the image. By constructing dams that represent high-value areas in the image, the algorithm prevents water flows from merging. Therefore, the different regions of the image are segmented. Due to noise or other factors, images often contain many local minima, leading to potential over-segmentation. In order to mitigate over-segmentation, the MCWST algorithm is enhanced on the watershed algorithm by incorporating foreground and background labels generated from the CV model results.

As an improvement, we employ Faster R-CNN to obtain initial seed points, which is able to decrease over-segmentation issues in blob detection. High-resolution CHMs often contain multiple local minima, which may cause severe over-segmentation problems. Faster R-CNN, which is the basis for Mask R-CNN, not only provides high-quality seed points but also ensures approximate consistency in the initial regions for both algorithms. The improved MCWST algorithm used in the experiment is shown in Figure 2.

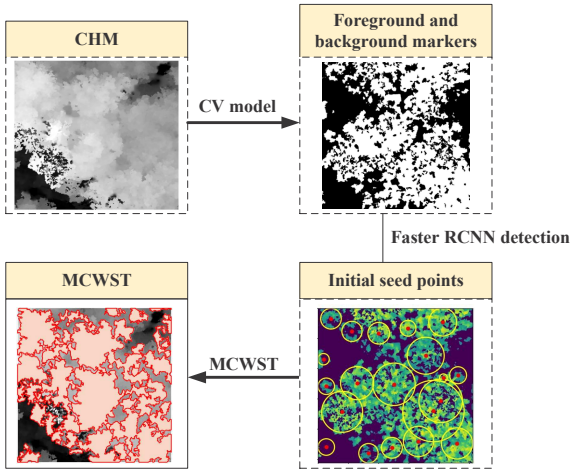

Figure 2. The workflow of MCWST method.

2.3 Mask R-CNN Model

Mask R-CNN is a deep learning model for object detection and instance segmentation (He et al., 2017), the changes compared to Faster R-CNN are highlighted in red in Figure 2. The mentioned models have two main differences, including (1) RoI Pooling is replaced by RoI Align, and (2) an additional branch in Mask R-CNN is added to output predictions about whether each pixel belongs to the object mask. RoI Align is introduced to address quantization issues present in RoI Pooling and enhance sampling precision for features. In addition, a branch dedicated to generating target masks is added, making it possible to obtain pixel-level boundaries of the targets, not limited to bounding boxes.

For each candidate region, RoI Align is introduced to address quantization issues present in RoI Pooling and enhance sampling precision for features. The model is trained and validated using a dataset of annotated tree crowns, and its performance is evaluated based on test results. The method is implemented using MMdetection 3.0 (Chen et al., 2019) and trained on an Nvidia GeForce RTX 3080 GPU. Mask R-CNN adopts ResNet50 as the backbone network. The overall loss function L for Mask R-CNN is represented by Equation (1).

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

L_{cls} is the target classification loss, L_{box} is the bounding box regression loss, and L_mask is the mask segmentation loss. Both L_{cls} and L_{mask} utilize the Cross-Entropy Loss function, while L_{box} employs the Smooth L1 Loss.

The loss allows the model to simultaneously perform target classification, bounding box regression, and instance segmentation at the pixel level. The training process includes twelve epochs. According to the experiments, increasing the number of training iterations did not lead to a better model. The minimum IoU threshold for bounding box predictions and ground truth boxes is set to 0.5.

2.4 Evaluation Metrics

Precision, recall, and F1 scores are used to evaluate model performance, which is assessed by comparing image annotations and algorithm predictions. The performance of models is assessed on the comparison between image annotations and algorithm predictions. If the center of each instance predicted by the algorithm falls within the corresponding annotated crown, and there is only one instance center within the crown, we calculate the Intersection over Union (IoU) between the annotated crown and the instance. If the IoU is greater than 50%, it is defined as true positive (TP). Otherwise, if the instance center is not contained within any crown, the IoU between a single

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-X-1-2024-67-2024 | © Author(s) 2024. CC BY 4.0 License.

instance and its corresponding annotation is less than 0.5, or if there are multiple instances within a patch, it is defined as false positives (FP). The absence of any instance center within a crown is labeled as a false negative (FN). Precision is defined as the ratio of TP to the total detected trees. Recall is defined as the ratio of TP to the annotated trees.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$IoU = \frac{area(B_{predict} \cap B_{crown})}{area(B_{predict} \cup B_{crown})} \tag{4}$$

$B_{predict}$ is indicative of the predicted scope for an individual instance, while B_{crown} delineates the range of the annotated crown corresponding to this instance.

3. Study Area and Data

3.1 Study Area

The study area is located in and around Bamberg, a city in northern Bavaria, Germany. It consists of the following four different forest and forestlike areas: (1) the citypark, called Hain, with a forestlike structure of deciduous trees and an area of 50 hectares, (2) the Stadtwald, a mostly coniferous managed forest with an area of 190 hectares in the southeast of Bamberg, and (3) two areas of 60 and 45 hectares of mixed managed forest called Tretzendorf_1 and Tretzendorf_2 about 20 km east of Bamberg. The dataset includes 19 tree species such as Picea abies, Fagus sylvatica and Abies alba. Table 1 gives a detailed description of these four areas of the dataset. The study area was divided into sample plots with a length and width of 100 m. The distribution of the sample plots is shown in Figure 3, where the red dots represent the centers of the grid cells in which the plots are located, not the centers of the plots themselves.

The detailed information about the equipment is available in the work of (Troles et al., 2023), which involves the use of two different UAVs to collect data in suburban and urban forests. UAV images, Digital Surface Model (DSM), and Digital Terrain Model (DTM) of the study area were collected in July 2022. The pixel resolution of the RGB orthophotos varies from 1.6 to 1.8 centimeters, DSMs have a resolution of 3.2 to 3.6 centimeters and DTM has a pixel resolution of 1 meter. To further process and obtain the CHM, the DTM is first resampled to the resolution of the DSM, then the DTM is subtracted from the DSM and lastly affixed value is subtracted, so ground areas pixels are as close to zero as possible. The delineation of tree crowns and the creation of the dataset were carried out from March 2023 to December 2023.

3.2 Generation of the Experimental Dataset

In this study, the delineation of individual tree crowns is carried out by a forester with extensive knowledge of forestry and remote sensing. This process is based on field observations and orthophoto imagery. As shown in Table 1, the delineation of a total number of 27,167 tree crowns are provided by the University of Bamberg (Troles et al., 2023). First, the CHM is generated based on the DSM and DTM, which serves as input data for the MCWST. Additionally, we organized the experimental

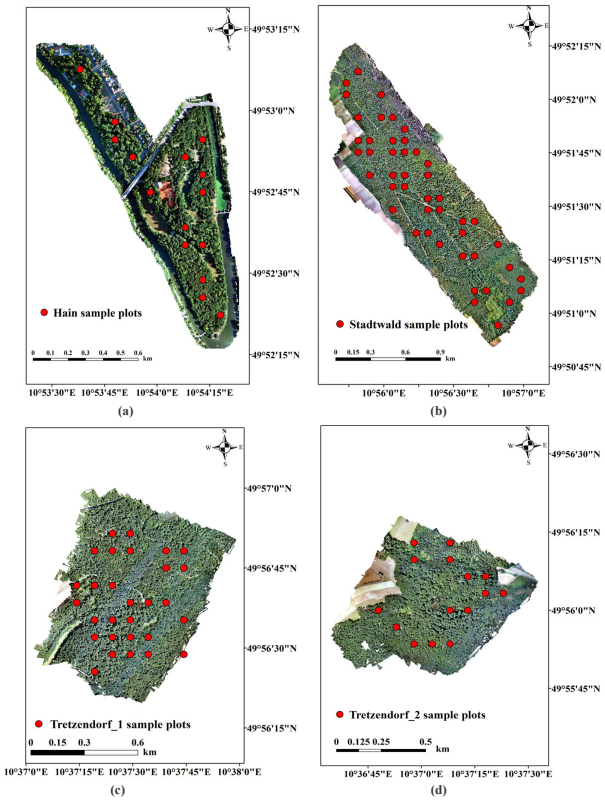

Figure 3. Distribution of sample plots and the study area (a) Hain, (b) Stadtwald, (c) Tretzendorf_1 (d) Tretzendorf_2.

data according to the COCO dataset format, segmenting the image tiles and associated annotations randomly for each original plot. For this purpose, 40% of the data is used for model training, 30% for model validation and the remaining 30% is used as a test set to assess the final performance of the model. The dimensions of the segmented images are 2048 pixels in both width and height. At this stage, the CHM image is cropped to the same size as the images in the COCO training set. It should be noted that the edges of the image can lead to the segmentation of a single crown into multiple crowns. Detailed information about the dataset is presented in Table 2 and Figure 4. To ensure that the training set of the dataset includes all tree species and that the model trained on the dataset shows transferability. The training set includes a total of 42 plots, most of which are from the Stadtwald area. Images from the Hain areas are mainly used for testing and validation. The training and test images are from different regions to ensure that overfitting is avoided in the experiment.

4. Result and Discussion

In the experimental part, we apply the deep learning based approach and the MCWST to our datasets. The results of the two approaches are compared both at the pixel level and the object level.

4.1 Experimental Results

The instance segmentation results for both algorithms used in the experiment are shown in Figure 5. From Figure 5, it can be seen that Faster R-CNN provides good initial detection boxes. Despite modifications to the seed point input in the experiments,

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-X-1-2024-67-2024 | © Author(s) 2024. CC BY 4.0 License.

Region	Size	Sample plots	Number of delineated ITCs	Description
Hain	50ha	15	1978	Region with deciduous forest-like areas
Stadtwald	190ha	46	15477	Mostly coniferous forest
Tretzendorf_1	60ha	29	6898	Mixed forest
Tretzendorf_2	45ha	15	2814	Mixed forest

Table 1. The study areas in and around Bamberg, Germany.

Dataset	Image numbers	Tree numbers
Train data	367	16124
Test data	255	7722
Validation data	226	6391

Table 2. Dataset description.

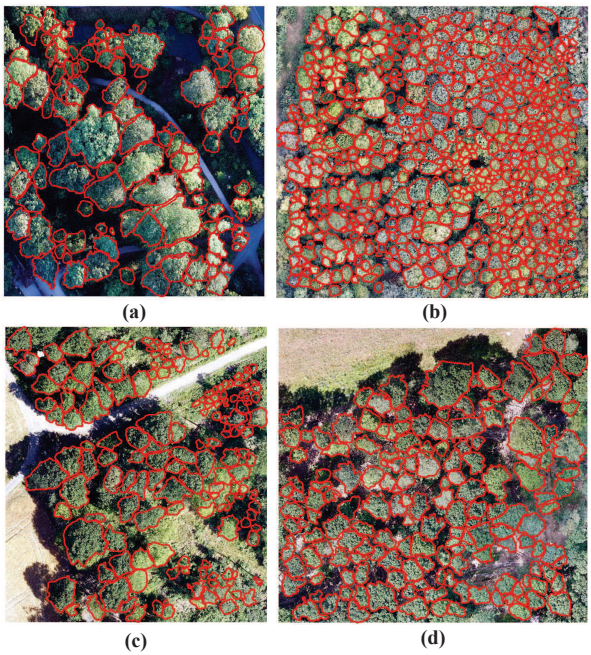

Figure 4. Image annotations for (a) the Hain plot, (b) the Stadtwald plot, (c) the Tretzendorf_1 plot, and (d) the Tretzendorf_2 plot.

the MCWST performance of the model remains suboptimal. Not only are there noticeable segmentation omissions, but for larger tree crowns, the segmentation results are significantly larger than the actual regions. MCWST exhibits significant under segmentation. Regarding Mask R-CNN, the algorithm has a good visual performance. The contour of Mask R-CNN visually segments different tree crowns well. It can be observed that there are still instances of missed segmentation in the results. The quantitative evaluation results for ITC segmentation using two different methods are presented in Table 3. TP is the actual number of crowns extracted, FN represents the number of missed extractions, and FP indicates the number of erroneously extracted crowns. Mask R-CNN shows a better performance, with precision and recall rates of 67.72% and 70.14%, respectively, compared to MCWST.

4.2 Discussion

For MCWST, this paper uses the detection boxes from Faster R-CNN as initial inputs to ensure the same range of inputs. In addition, if traditional blob detection is used to obtain seed points, there would be multiple low-value points in the high-resolution CHM. This leads to severe over-segmentation problems as shown in Figure 6. The number of seed points obtained from traditional blob detection is significantly higher than the number of trees, resulting in overly fragmented image segmentation. Faster R-CNN provides a more reasonable initial number and position of seed points, thus reducing this problem. Moreover, it can be seen from Figure 5 that the edges produced by the MCWST method are not as smooth as those produced by Mask R-CNN. It is partly due to that the ground truth is manually annotated, resulting in smoother edges than the actual ground truth edges.

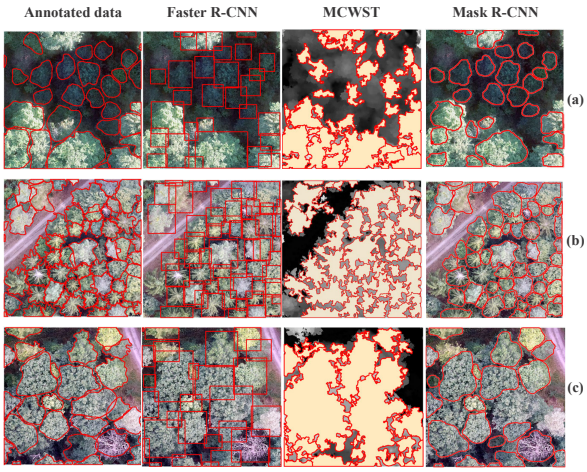

Figure 5. Experiment results. (a), (b), and (c) represent results of different test images.

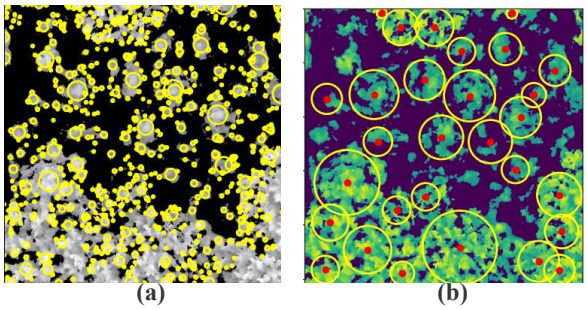

Figure 6. Differences in seed point settings. (a) shows seed points obtained from traditional spot detection, and (b) shows seed points centers of detection boxes from Faster R-CNN.

The Mask R-CNN algorithm has instances of both under-segmentation and over-segmentation, as shown in Figure 7. Under-segmentation occurs due to dense forest areas, making it difficult to accurately segment trees of the same species based on images alone. On the other hand, over-segmentation results from significant differences in the tree distribution between the training and test sets, leading to suboptimal segmentation. Therefore, future work should focus on designing experiments to improve the robustness of the model.

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-X-1-2024-67-2024 | © Author(s) 2024. CC BY 4.0 License.

Method	TP	FP	FN	Precison	Recall
MCWST	2881	2564	5547	0.529	0.342
Mask R-CNN	5453	2599	2322	0.677	0.701

Table 3. Experimental results.

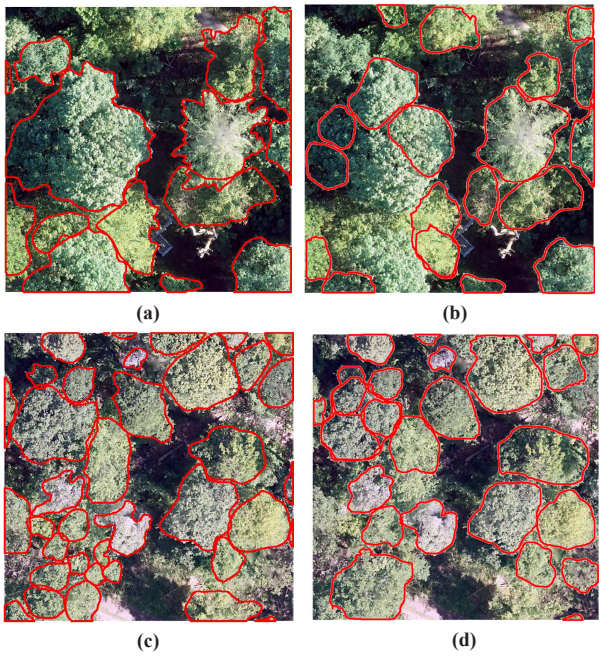

Figure 7. Analysis of Mask R-CNN Results. (a) and (c) represent RGB images and annotations, (b) exhibits instances of under-segmentation, while (d) displays instances of over-segmentation.

Both methods are capable of segmenting individual trees in forests, providing data for tree resource management and forest disturbance analysis. However, there are still some problems. Random proportional sampling has the potential to produce an unbalanced sample. In subsequent datasets, manual selection is added to reduce this problem. The training and test sets are selected from different regions to ensure portability of the model. In addition, the annotated data introduces errors for both algorithms. From a ground truth annotation perspective, the segmentation of overlapping crowns proves to be challenging. Even experienced foresters cannot guarantee the correct segmentation of two overlapping crowns. There is unannotated data in the training images, where some tree crowns are not assigned to any instance, leading to numerous misidentifications. Despite the superior performance of Mask R-CNN, there is still considerable potential for improving the model’s performance on the ITC segmentation task using the given dataset. The dataset used in the experiment will also be released soon, allowing more researchers to explore this topic. Besides, orthophotos suffer from tree occlusion. Only the top canopy is visible in the image. In the future, the vertical structural information of trees will be explored to segment individual trees using 3D data

5. Conclusion

Reliable segmentation of individual trees in the forest is essential for forest management and ecological assessment. On the one hand, the varying contours of tree crowns of different species, coupled with the challenge of segmenting overlapping crowns, add a significant degree of complexity to the task. On the other hand, the acquisition of UAV data and subsequent manual delineation requires time and significant human resources to overcome the lack of annotated datasets. This study investigates the algorithmic performance of the MCWST method and the Mask R-CNN approach for ITC segmentation in UAV imagery. Specifically, Mask R-CNN achieves an accuracy about 15% higher than MCWST coupled with an object detection approach, reaching 67.72%. The recall rate is twice that of MCWST, at 70.14%. Mask R-CNN provides a valid initial contour of individual trees, but it is still affected by the complexity of the tree species and tree density. There are instances of over-segmentation and under-segmentation in the experiment, which needs to further improve the transferability of the model. Based on the results of the individual tree segmentation, our future work will focus on the investigation of forest disturbances and the analysis of tree vitality at the individual tree level. In addition, the aforementioned ITC segmentation dataset will be released as a publicly available dataset for wider use by researchers.

6. Acknowledgements

This research is jointly supported by the China Scholarship Council and the Bavarian State Ministry for Digital Affairs.

References

Ball, J. G., Hickman, S. H., Jackson, T. D., Koay, X. J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G., Coomes, D. A., 2023. Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sensing in Ecology and Conservation*, 9(5), 641–655.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C. C., Lin, D., 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155*.

Dong, X., Zhang, Z., Yu, R., Tian, Q., Zhu, X., 2020a. Extraction of information about individual trees from high-spatial-resolution UAV-acquired images of an orchard. *Remote Sensing*, 12(1), 133.

Dong, X., Zhang, Z., Yu, R., Tian, Q., Zhu, X., 2020b. Extraction of Information about Individual Trees from High-Spatial-Resolution UAV-Acquired Images of an Orchard. *Remote Sensing*, 12(1). https://www.mdpi.com/2072-4292/12/1/133.

G. Braga, J. R., Peripato, V., Dalagnol, R., P. Ferreira, M., Tarabalka, Y., OC Aragão, L. E., F. de Campos Velho, H., Shiguemori, E. H., Wagner, F. H., 2020. Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, 12(8), 1288.

Gonzalez, P., Neilson, R. P., Lenihan, J. M., Drapek, R. J., 2010. Global patterns in the vulnerability of ecosystems to vegetation shifts due to climate change. *Global Ecology Biogeography*.

Guo, X., Liu, Q., Sharma, R. P., Chen, Q., Ye, Q., Tang, S., Fu, L., 2021. Tree recognition on the plantation using UAV images with ultrahigh spatial resolution in a complex environment. *Remote Sensing*, 13(20), 4122.

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.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Holmgren, J., Lindberg, E., Olofsson, K., Persson, H. J., 2022. Tree crown segmentation in three dimensions using density models derived from airborne laser scanning. *International Journal of Remote Sensing*, 43(1), 299–329.

Huang, H., Li, X., Chen, C., 2018. Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(7), 2253–2262.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 173, 24–49.

Kempf, C., Tian, J., Kurz, F., D’Angelo, P., Schneider, T., Reinartz, P., 2021. Oblique view individual tree crown delineation. *International Journal of Applied Earth Observation and Geoinformation*, 99, 102314.

Kotaridis, I., Lazaridou, M., 2021. Remote sensing image segmentation advances: A meta-analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 309-322. https://www.sciencedirect.com/science/article/pii/S0924271621000265.

Ma, K., Chen, Z., Fu, L., Tian, W., Jiang, F., Yi, J., Du, Z., Sun, H., 2022. Performance and sensitivity of individual tree segmentation methods for UAV-LiDAR in multiple forest types. *Remote Sensing*, 14(2), 298.

Pearse, G. D., Tan, A. Y., Watt, M. S., Franz, M. O., Dash, J. P., 2020. Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168, 156–169.

Qin, H., Zhou, W., Yao, Y., Wang, W., 2022. Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. *Remote Sensing of Environment*, 280, 113143.

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*, 28.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., Schmidtlein, S., 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170, 205–215.

Troles, J., Nieding, R., Simons, S., Schmid, U., 2023. Task planning support for arborists and foresters: Comparing deep learning approaches for tree inventory and tree vitality assessment based on uav-data. *International Conference on Innovations for Community Services*, Springer, 103–122.

Wallace, L., Sun, Q. C., Hally, B., Hillman, S., Both, A., Hurley, J., Martin Saldias, D. S., 2021. Linking urban tree inventories to remote sensing data for individual tree mapping. *Urban Forestry Urban Greening*, 61, 127106. https://www.sciencedirect.com/science/article/pii/S161886672100131X.

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11), 1309.

Yu, K., Hao, Z., Post, C. J., Mikhailova, E. A., Lin, L., Zhao, G., Tian, S., Liu, J., 2022. Comparison of classical methods and mask R-CNN for automatic tree detection and mapping using UAV imagery. *Remote Sensing*, 14(2), 295.