The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W6-2025
3rd International Workshop on Evaluation and BENCHmarking of Sensors, Systems and
GEOspatial Data in Photogrammetry and Remote Sensing (GEOBENCH), 20–21 November 2025, Wroclaw, Poland

Comparison of manual and semi-automated synthetic training data creation for individual tree crown delineation

Janik Steier, Dorota Iwaszczuk

Dept. of Remote Sensing and Image Analysis, Technical University of Darmstadt, 64287 Darmstadt, Germany – (janik.steier, dorota.iwaszczuk)@tu-darmstadt.de

Keywords: Manual Labeling, Synthetic Training Data, Individual Tree Crown Delineation

Abstract

Deep learning models in the field of individual tree detection and crown delineation (ITDCD) rely on large and high-quality annotation datasets to produce accurate predictions. Training data or annotations for most ITDCD studies are collected through manual labeling. Manual labeling, especially for complex structures like tree crowns, is a time-consuming process that often results in error-prone annotations. Error-prone annotations, in turn, can lead to significant errors in the predictions of deep learning models. Semi- or fully-automated training data creation shows the potential to make the creation process more efficient and ensure high quality of the training dataset. In this work, we present a methodology for generating semi-automated synthetic training data for deep learning-based ITDCD applications. Furthermore, a systematic criteria-based - validity, efficiency, variety and scalability - comparison is conducted between the manual and synthetic training data creation methods to structurally and practically illustrate the advantages and disadvantages of the two approaches. Overall, the semi-automated synthetic data approach outperforms manual labeling in terms of validity, efficiency, and scalability; once the algorithm is implemented, it rapidly generates arbitrarily large, high-quality, reproducible tree crown annotation datasets. In contrast, a manual creation approach shows its advantages as an efficient way to create small, low-quality datasets (e.g., for fine-tuning a pre-trained model) compared to developing a semi-automated method from scratch.

1. Introduction

In recent years, deep learning, particularly convolutional neural networks (CNNs), has become established in the field of individual tree detection and crown delineation (ITDCD) based on very high-resolution remote sensing images. Thereby, object detection networks like RetinaNet and instance segmentation models like Mask-RCNN or YOLO have shown a significant improvement in terms of accuracy over traditional ITDCD approaches. Training of these deep learning models requires a large amount of labeled samples, also called annotations. In the case of ITDCD, near-pixel-accurate labels of single tree crowns' contours are necessary. Most studies in the field of ITDCD collect training data or annotations via hand labeling, in which individual tree crowns or canopy cover are manually labeled using bounding boxes or polygons. (Zhao et al., 2023)

In general, the manual creation of training data through visual interpretation in remote sensing imagery as well as individual tree crown annotations specifically are inherently error-prone and are fraught with several unavoidable challenges (Zimmermann et al., 2023). Possible subjective and objective error sources from manual annotations of tree crown are for instance the irregular shape of the tree crown, the overlapping canopies, the natural forests' density as well as the subjective recognition of the crown shape by the individual annotator.

Consequently, manually delineated tree crowns are not considered to represent the true delineations of individual tree crown on site. In a previous study, we critically validated quantitatively manual tree crown annotations on two study sides against tree reference data in form of an official tree register and tree segments extracted from UAV laser scanning data. Our validation results show a low quality of the manual annotations in capturing individual tree crowns. The annotators significantly underestimate the true number of reference trees in the images. In addition, the annotators often summarize multiple tree crowns into one annotation. Based on our research, we conclude that manual annotations of individual tree crowns in forests and areas with a forest-like plantation on remote sensing images are very likely to have significant deficits in capturing the actual conditions on the ground. (Steier et al., 2024)

Deep learning models for ITDCD rely on large and carefully annotated datasets to produce accurate predictions (Fan et al., 2024; Troles et al., 2024). As the annotations are used as training data, error-prone annotations can cause substantial errors in the prediction of deep learning models (Elmes et al., 2020; Karimi et al., 2020).

Beyond the inherent susceptibility to errors, manual labeling of individual tree crowns is very labor-intensive and time-consuming, which hinders the efficient creation of a sufficiently large training dataset.

A potential way to address this problem and improve the efficiency and the generalizability of model training is the use of synthetic datasets. To generate large numbers of training data, some studies developed semi- or fully-automated methods in addition to hand-annotated samples (Zhao et al., 2023). For example, Weinstein et al. (2019) proposed an unsupervised method that automatically segments tree crowns from lidar point clouds. The RGB data corresponding to these segmented shapes was then used to train a CNN model. In another work, by employing a morphological algorithm to simulate tree crown shapes in a digital elevation canopy model, Pulido et al. (2020) generated 12,500 synthetic images (with 2 to 7 trees per image) without hand-annotated data. Nevertheless, studies that pursue a semi- or fully-automated approach to training data creation in the field of ITDCD are still in the minority.

This work presents a methodology for generating semi-automated synthetic training data for deep learning-based ITDCD applications. The method is based on the work by Braga et al., 2020.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-227-2025

Furthermore, a systematic criteria-based comparison between the manual and synthetic training data creation methods is conducted. This comparison is designed to structurally and practically illustrate the advantages and disadvantages of the two methods, thereby assisting decision-makers in selecting their preferred training data creation methodology.

2. Method and Materials

2.1 Training Data Creation Approaches

For the criteria-based comparison between the creation methods, both manual and semi-automated synthetic training data are generated.

The manual annotations were created using digital orthophotos (DOPs) with a 20 cm spatial resolution, an 8-bit radiometric resolution and RGB bands of the municipal cemetery of Frankfurt am Main, Germany. The study site was specifically chosen because of the presence of a forest-like plantation and a publicly accessible tree register, which facilitated the validation of the annotations. The image dataset was acquired by an airplane flight in June 2021. The resulting digital orthophotos are published by the federal state of Hesse, Germany. They are freely available and updated in a two-year cycle.

Before the annotation creation, the original image of the study site was split into tiles of 512 × 512 pixels. The larger image size was chosen to avoid excessive tree crown fragmentation when splitting the image into tiles. Furthermore, the chosen tile size proved to be a favorable compromise in the annotation process between effective digitalization of the tree crowns, due to fewer interruptions when creating the annotations, and having a manageable number of trees within a single tile (Steier et al., 2024).

The annotations were created manually by six annotators, each of whom had different levels of experience in creating single tree crown annotations. The annotators were instructed to capture the outline of the individual tree crowns as precisely as possible and to annotate every possible tree crown on the image, which comes into consideration for the annotator. Furthermore, no time limit was set for the annotation process, and the time taken was recorded. The browser-based, open-source platform CVAT (Computer Vision Annotation Tool) was used to manage the annotation process and generate the manual annotations.

Figure 1 provides a representative example of a completely hand-annotated DOP within the study area.

The publicly available algorithm by Braga et al. (2020), written in the Python programming language, was implemented and modified to generate semi-automated synthetic forest images for the synthetic training dataset. The algorithm produces the synthetic forest images using a set of well-delineated tree crowns. The algorithm randomly takes annotations from the set and places them at random in a background image until the desired number of trees for the image is reached. Additionally, the background information, the tree density and the number of trees per image can be defined based on the user's specifications. Each synthetic forest image is therefore created purely by chance and is unique.

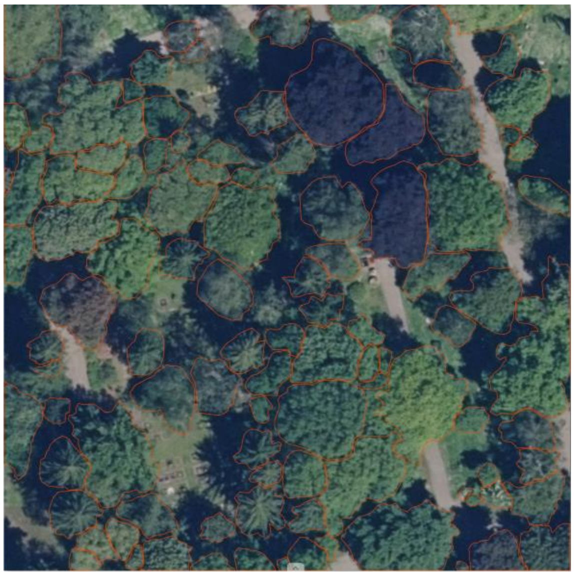

Figure 1. Completely hand-annotated DOP with 103 annotations (orange outline).

We have extended the algorithm by randomizing the orientation of tree crowns, adjusting their radiometric properties, and implementing a blending technique for the crown boarders during the process of synthetic image creation to increase the internal diversity of the synthetic training dataset. We also modified the algorithm to process tree crown annotations based on DOPs with a spatial resolution of 20 cm, instead of the annotations with 50 cm spatial resolution based on WorldView-2 satellite imagery that were used in the study by Braga et al. (2020).

The modified algorithmic steps, based on the work by Braga et al. (2020) are described in Algorithm 1:

Algorithm 1: Generating a randomized synthetic forest image based on a set of single annotations

input: *treecrown_annotation, treecrown_image, background_image*
initialization: *total_forest, total_tree, distance, flip_rotate, brightness_change*
    count = 1;
    create a *matrix* with the same dimensions as the *background_image*
    while (*count* < *total_tree*) **do**
        select one *place* within in the *background_image*
        select one tree crown within the set of *treecrown_image* and corresponding *treecrown_annotation*
        apply the *flip_rotate* and *brightness_change* functions, if selected
        if (*place* is free) **then**
            set the tree crown in the *background_image*
            fill *matrix* with the tree crown values selected from *treecrown_image* using *place*
            count += 1
        end
    end
    polygonize the *matrix* into shapefile
    blending black boarders of tree crowns
    return forest image, forest shapefile

In the Algorithm 1:

- *treecrown_annotation* is the ESRI shapefile containing the manually delineated tree crowns
- *treecrown_image* contains the raster information in RGB channels of the annotated individual trees corresponding to the *treecrown_annotation*
- *background_image* is a background raster with the dimension of *treecrown_image* where the synthetic forest will be created
- *total_forest* - total number of images to be created, *total_tree* - total number trees per image, *distance* - distance between trees, *flip_rotate* - random flip and rotation of tree crown between the angle of 0, 90, 180, 270 degree, *brightness_change* - random change in brightness of the tree crown
- *count* variable to control the numbers of crowns in the *background_image*
- *matrix* is an array where every tree crown in the created forest image is polygonized to get a correspondent shapefile of every tree crown within the forest; this will be the tree crown annotations of the synthetic forest image
- the algorithm checks, if the *place* selected to copy the tree crown into *background_image* is free
- the geometries of each tree crown within in the *matrix* are polygonized and converted into a shapefile
- black boarders are generated at certain points between the tree crowns and the background image during the setting process; these are then blended using a bilateral filter
- the algorithm returns the forest image generated on the *background_image* and the corresponding shapefile

We created our own dataset of high-quality annotations also based on the DOPs with a 20 cm resolution and RGB bands of the municipal cemetery of Frankfurt. During the dataset annotation process, we ensured that we primarily annotated freestanding trees to accurately capture the tree crown outline and to include different tree species (Figure 2). The annotations were created using QGIS (QGIS Development Team, 2024) in the ESRI Shapefile format.

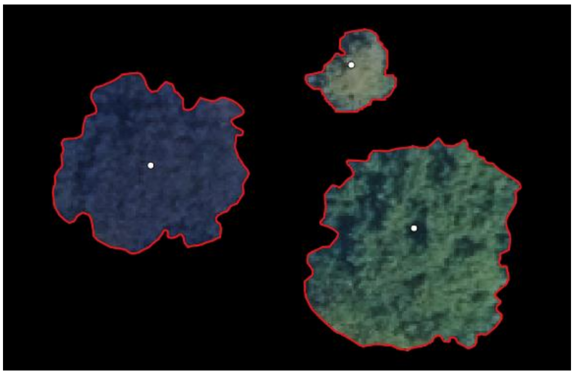

Figure 2. Excerpt of the high-quality annotations set. The individual tree raster is shown with the hand-made annotation (red outline) and the tree register point within each crown (white point).

Furthermore, we validated the annotations in our set against the tree register to ensure that every annotation in the set represents only one individual tree crown on the ground and to be able to assign the metadata from the tree register to the individual annotations. In total, our high-quality annotation dataset comprises 60 validated annotations with 12 different tree genera.

2.2 Evaluation criteria

The creation approaches are evaluated qualitatively based on the following criteria: validity, efficiency, variety and scalability.

Validity evaluates the correctness of an individual annotation in terms of whether the annotator was able to identify an individual as well as the annotation made represents the true outline of the individual tree crown on the ground. Manual annotations can be validated with tree reference data (e.g. tree register, laser scanning data or field-based labeling from expert botanist). In this work, the manual annotations are validated against the public accessible tree register of the city of Frankfurt am Main from 2021. This register information was converted to vector point data for each tree.

Therefore, it can only be assessed whether the annotator captured the individual tree register point with their annotation**.** A comparison of the annotation outline with the true outline of the individual tree crown on the ground is not possible.

To validate the manually generated annotations, we use the following case distinctions based on Steier et al. (2024): “one annotation captures exactly one tree register point”, “a tree register point is not captured by a single annotation” and “multiple tree register points are summarized by one annotation”. The number in the case is divided by the total number of tree register points to obtain the validation metric.

For synthetic training data, a high validity is ensured as the computer-generated imagery is based on high-quality annotations. In the proposed semi-automated approach, this means that the annotations of individual tree crown, which are reproduced on the synthetic images, are well-delineated tree crowns of mostly freestanding trees created by hand and have been validated using tree reference data.

In both datasets the efficiency can be evaluated by the time it takes for the dataset creation. When creating the manual dataset, the time it takes for annotators to label the individual tree crowns is measured. For the semi-automated dataset, in addition to the computer processing time for synthetic image creation, it is also necessary to account for the time spent on implementing the creation tool (e.g. setting up or adapting an algorithm) and the manual creation of a set of well-delineated tree crowns.

The variety of a dataset is judged by its ability to represent various scenarios of the training object. This characteristic is especially important to avoid overfitting and to increase the generalizability of the neural network. A training dataset for an ITDCD model should ideally exhibit high variability. This includes a diverse range of tree species, variations in tree crown shape and size, different tree stand densities, varying degrees of crown overlap, diverse shadow patterns, and a variety of background scenarios.

Scalability of the training data creation approach is defined by the ease with which a sufficiently large amount of training data can be generated and can be applied to larger projects.

3. Results and Discussion

3.1 Synthetic Forest Images

For creating our semi-automated synthetic training datasets, we chose various background images with the size of 512 x 512 pixels with a 20 cm pixel resolution and different configurations regarding the number of trees and their stand density.

Figures 3 and 4 show example synthetic forest images with different creation configurations. In Figure 3, ten trees with a low stand density are positioned on a black background image. In this setup, the individual trees are spaced at different distances, ensuring that crown overlap does not occur.

For Figure 4, an agricultural landscape with paths was used to create a more realistic background for an urban forest. The number of tree crowns was increased to 90 compared to Figure 3, and a high tree stand density was chosen. It is observed that the tree crowns are closely spaced and that neighboring tree crowns overlap to varying degrees, while clear areas also remain.

The algorithm was unable to process more than 90 to 100 tree crowns given our settings in dense distance between trees, spatial resolution, shape of the tree crown annotations in the high-quality annotation dataset and background image size. If a higher number of trees was selected, the algorithm would enter an infinite loop while searching for an available spot to place the tree crown.

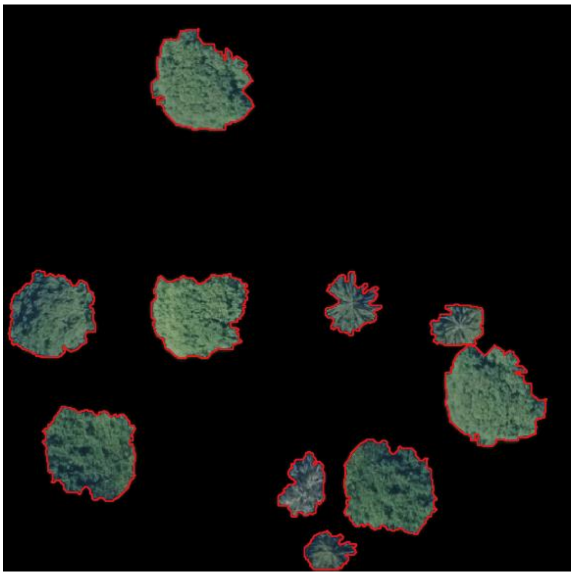

Figure 3. Ten trees with a low tree stand density; the red outline defines the annotation shape.

Furthermore, when a high density and a high number of trees were chosen (Figure 4), while the result was a realistic approximation of an urban forest capable of partially forming a continuous canopy, two issues arose. First, free spaces still remained between the trees. Second, the overlap between adjacent crowns was occasionally so large that the crowns were heavily obscured, giving them an unrealistic shape. This issue may be caused by our specific chosen settings, the tree crown shape within our high-quality annotation dataset, or the general architecture of the algorithm.

Further optimization steps would be necessary here to fully utilize the available space in the synthetic images. This could allow us to generate synthetic forest images that feature a fully closed canopy, similar to natural deciduous forests.

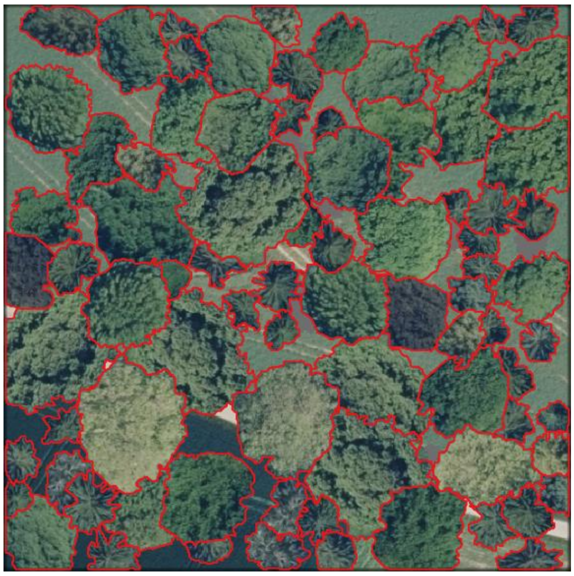

Figure 4. Approx. 90 trees with a high tree stand density placed on a natural background scene.

3.2 Analysis of the training data creation approaches

The training data creation approaches are analyzed and discussed qualitatively using the evaluation criteria - validity, efficiency, variety and scalability - and are presented quantitatively in a radar diagram.

3.2.1 Validity

The validation results of the manual annotations show that 50 % of the annotations capture exactly on tree register point. 21% of the tree register points went unidentified by the annotator and in 29% of cases, the annotator summarized multiple tree register points into a single annotation. These results show that the annotators systematically underestimate the true number of trees in the in the images as well as they often summarize multiple tree crowns into one annotation. Additionally, the annotations occasionally exhibit unrealistic tree crown shapes (Figure 1). This result once again highlights the problem with manual annotations in the area of individual tree crown delineation, as we investigated in Steier et al., 2024: the lack of the ability to capture the true tree crown conditions, thereby generating error-prone training data. Error-prone annotations can significantly impact the performance of deep learning models in many machine learning and computer vision applications, especially if the predicted object is irregularly shaped and has low visibility within the image (Elmes et al., 2020; Karimi et al., 2020).

In the proposed semi-automated approach, the annotations of individual tree crown, which are reproduced on the synthetic images, have been validated in advance using tree register points. That means each reproduced annotation within the synthetic forest imagery corresponds to a unique, existing tree crown and completely avoids the error of one annotation representing no tree crown or multiple tree crowns.

While generating the high-quality annotations involved careful attention to capturing the tree crown outline, this procedure is nonetheless susceptible to the possible subjective and objective error sources inherent in manual labeling. A validation of the

https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-227-2025 | © Author(s) 2025. CC BY 4.0 License.

high-quality annotations against the tree crown outlines on the ground was not possible.

However, the semi-automated synthetic training data creation method is able to significantly reduce the error within the individual tree crown annotations compared to manual annotation.

3.2.2 Efficiency

In our experiment, the annotators required approximately 38 hours to manually label 140 images, which involved creating 8,230 annotations of tree crowns. This equates to an average of 18 minutes per image. In contrast, the algorithm requires only approx. six seconds per image and 14 minutes in total to generate the same number of images with the same number of annotations.

For both approaches, an additional time investment for preparation must be considered. The manual annotation approach requires the selection of an annotation tool and the instruction of the annotators. For this project, the total time required was approximately 6 hours.

For the semi-automated approach, the time required to create the high-quality annotation set must be considered; this took approximately 15 hours for this project.

Furthermore, an algorithm for the synthetic image creation must be set up or implemented. The independent creation of a new algorithm from scratch was not necessary for this work, as the existing algorithm from Braga et al. (2020) was implemented and modified. The time required for the implementation and modification, in our case, should be accounted for as approximately 24 hours. In addition, an algorithm must be written to convert the synthetic created annotations from the shapefile format into a format readable by deep learning models, such as the COCO JSON format. This feature is already available in the standard annotation tools for manual annotations.

When creating a small dataset without certain quality standards, the manual data creation method is advantageous because it can be implemented more easily and efficiently using a state-of-the-art annotation tool than by employing an entire semi-automated method from scratch. The manual method is particularly useful for tasks such as fine-tuning a pre-trained model using a custom training dataset and adapting the ITDCD model to a specific geographical region or specializing the model to recognize a particular tree species.

3.2.3 Variety

The variety of manual created tree crown annotations, including the range of tree species, variations in tree crown shape and size, different tree stand densities, varying degrees of crown overlap, diverse shadow patterns and a variety of background scenarios, is defined by the area being annotated and the existing tree properties on the ground and is therefore fixed and cannot be changed within an annotation project.

The study area selected for tree crown annotations contains a total of 44 distinct tree genera with crown radius ranging from 1 to 15 meters. In comparison, the German National Forest Inventory (BWI), which assesses large-scale forest conditions in Germany, records 51 distinct tree species or species groups (BMEL, 2024). Given an annotation area limited to 75 hectares, our dataset nonetheless captures a considerable diversity in tree species, tree crown size and shape. The tree density corresponds to approximately 96 trees per hectare. For reference, a German forest consisting of mature trees, the stem density per hectare ranges from approximately 100 trees for oak stands to around 400 trees for spruce stands (Scholz, o.D.)

In comparison to the complete manual dataset, which includes 44 distinct tree genera, our high-quality annotation dataset for the synthetic dataset creation contains only 12 different tree genera with crown radius ranging from 1 to 11.5 meters. If these few high-quality annotations are reproduced on the synthetic forest images, it necessarily leads to a high redundancy of those annotations within the image set. However, it must be highlighted that the high-quality annotation dataset can be structured and expanded according to the user's preferences to increase its variety.

Furthermore, the limited variety inherent in a small limited high-quality annotation set is addressed by randomly reselecting and rearranging the annotations for each newly created synthetic forest image. When a user also chooses a high stand density, the individual tree crowns will naturally overlap (see Figure 4), which alters the original annotation shapes in every generated synthetic image.

Additionally, in the semi-automated creation approach, the background information, the tree density, the number of trees per image, the orientation and the radiometric properties of the individual tree crowns can be defined based on the user's specifications. In the manual annotation approach, these characteristics are predetermined by the acquisition settings and the natural features of the tree.

3.2.4 Scalability

If we wanted to generate high numbers of manually annotated tree crowns for a training dataset in our project, this would require either increased time expenditure or an increase in the number of annotators. Assuming the study area remains the same, the annotation time would not simply increase linearly with a higher number of images. Several factors can lead to a disproportionate increase in the time required for a manual annotation process relative to the need for larger training datasets.

For instance, naturally and legally mandated rest periods for human annotators must be adhered to, and consequently, more annotators may need to be onboarded and trained. Furthermore, fatigue and waning motivation can slow down the annotation process over an extended period and affect the quality of the annotations (Mei et al., 2024). Conversely, an increase in annotator competence can lead to an acceleration of the process.

The computerized semi-automated creation approach is largely exempt from these human-related constraints. In our work, once the existing algorithm has been implemented and a high-quality annotation dataset has been created, the number of synthetic forest images is freely scalable. The time required per image proceeds constant and the creation process can be operated largely without interruption.

However, the scalability of the creation of synthetic forest images is limited by the fact that maximal space utilization in the tree crown placement on the background image was not achieve and a synthetic forest image with a full closed canopy could not be generated yet.

Our results for the four evaluation metrics for both training data creation approaches in the area of individual tree crown delineation are summarized quantitatively on a scale from 0 (not applicable) to 5 (very good applicable) and are graphically presented in Figure 5.

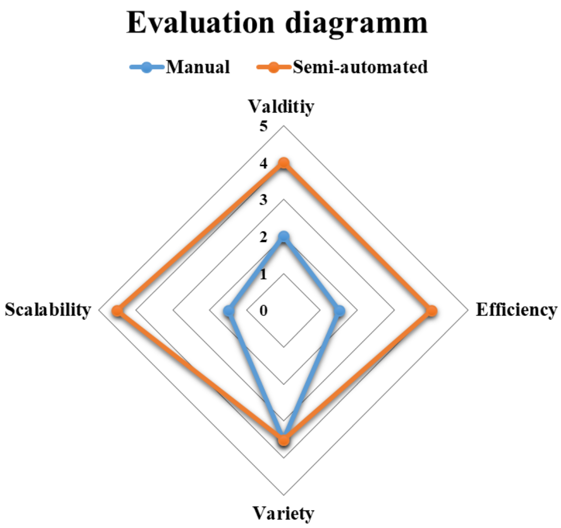

Figure 5. Quantitative evaluation of the manual and semi-automated training data creation approaches.

4. Conclusion and Outlook

In this work, we implemented and modified an algorithm, for generating semi-automated synthetic training data for ITDCD applications. Using our well-delineated annotation set, we were able to create an extensive, high-quality training dataset with this methodology. High quality was ensured by only reproducing annotations, which were validated against a tree inventory, onto the synthetic forest images. Furthermore, user settings such as background information, tree density, number of trees, variations in geometric crown orientation, and changes in brightness allow us to both specify the dataset and increase its generalizability by boosting variety. Only the placement of the tree crown annotations in the synthetic images needs further development to allow us to generate synthetic forest images as close to natural tree density as possible.

Overall, the semi-automated synthetic training data approach offers significant advantages across the evaluation criteria validity, efficiency and scalability compared to manual labeling method. Once the semi-automated algorithm is implemented and a reproducible tree crown annotation dataset is created, it can produce an arbitrarily high quantity and high-quality of synthetic forest images in a relatively short amount of time. In our experiment, the human annotators required approximately 38 hours to manually label 140 images, which involved creating 8,230 annotations of tree crowns. In contrast, the algorithm requires approximately 14 minutes to generate the same number of images with the same number of annotations. Furthermore, the systematic human error proneness of the tree crown annotations was reduced by utilizing validated and well-delineated tree crown annotations. Yet, a state-of-the-art annotation tool makes manual data creation for small, low-quality datasets easier and more efficient than developing a semi-automated method from scratch.

Recent studies have shown, that the performance of ITDCD models is highly sensitive to training data (Fan et al., 2024; Troles et al., 2024; Khan et al., 2025). Consequently, efforts in training data provisioning for ITDCD models must prioritize ensuring high data quality, quantity, and variance. This is essential to significantly enhance model training efficiency and bolster generalizability.

Our findings this work show that a semi-automated method outperforms a purely manual approach in training data creation for individual tree crown delineation. Consequently, future research should focus on how semi- or fully-automated methods can be made easily accessible and the associated implementation effort can be reduced.

Data Availability Statement

The modified algorithm presented in this word is available on request from the corresponding author Janik Steier.

References

BMEL (Ed.), 2024. Bundeswaldinventur Der Wald in Deutschland. Ausgewählt Ergebnisse der vierten Bundeswaldinventur. Bundesministerium für Ernährung und Landwirtschaft.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J., Fishgold, L., Friedl, M., Jain, M., Kohli, D., Laso Bayas, J., Lunga, D., McCarty, J., Pontius, R., Reinmann, A., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z.-F., Estes, L., 2020. Accounting for Training Data Error in Machine Learning Applied to Earth Observations. *Remote Sensing* 12 (6), 1034.

Braga, J.R., Peripato, V., Dalagnol, R., P. Ferreira, M., Tarabalka, Y., O. C. 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.

Fan, W., Tian, J., Troles, J., Döllerer, M., Kindu, M., Knoke, T., 2024. Comparing Deep Learning and MCWST Approaches for Individual Tree Crown Segmentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* X-1-2024, 67–73.

Karimi, D., Dou, H., Warfield, S.K., Gholipour, A., 2020. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis* 65, 101759.

Khan, T., Krebs, J., Gupta, S., Renkel, J., Arnold, C., Nölke, N., 2025. Validation Challenges in Large-Scale Tree Crown Segmentations from Remote Sensing Imagery Using Deep Learning: A Case Study in Germany. *arXiv*

Mei, Q., Steier, J., Iwaszczuk, D., 2024. Integrating Crowd-sourced Annotations of Tree Crowns using Markov Random Field and Multispectral Information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-2-2024, 257–263.

Pulido, D., Salas, J., Rös, M., Puettmann, K., Karaman, S., 2020. Assessment of Tree Detection Methods in Multispectral Aerial Images. *Remote Sensing* 12 (15), 2379.

QGIS, 2024. QGIS Geographic Information System. QGIS Association, Version 3.28.2. Available online: http://www.qgis.org

Scholz, o.D. Schätzhilfen: Wieviel Festmeter und Bäume stehen in meinem Wald?, https://www.lwk-niedersachsen.de/lwk/news/36164_Schaetzhilfen_Wieviel_Festmeter_und_Baeume_stehen_in_meinem_Wald. (Accessed 13 October, 2025).

Steier, J., Goebel, M., Iwaszczuk, D., 2024. Is Your Training Data Really Ground Truth? A Quality Assessment of Manual Annotation for Individual Tree Crown Delineation. *Remote Sensing* 16 (15), 2786.

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 (11), 1935.

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.

Zhao, H., Morgenroth, J., Pearse, G., Schindler, J., 2023. A Systematic Review of Individual Tree Crown Detection and Delineation with Convolutional Neural Networks (CNN). *Current Forestry Reports* 9 (3), 149–170.

Zimmermann, E., Szeto, J., Ratle, F., 2023. An Empirical Study of Uncertainty in Polygon Annotation and the Impact of Quality Assurance. *arXiv*