INDIVIDUAL TREE CROWN BASED TREE SPECIES CLASSIFICATION FROM VERY HIGH-RESOLUTION UAV IMAGES

Mareike Weishaupt[1], *Wen Fan*[2], *Jonas Troles*[3], *Jiaojiao Tian*[1,2]

[1] Remote Sensing Technology Institute (IMF), DLR, 82234, Weßling, Germany
[2] Institute of Forest Management, Technical University of Munich, 85354 Freising, Germany
[3] Cognitive Systems Group at University of Bamberg, 96047 Bamberg, Germany

***Abstract*—Very high-resolution optical imagery from unmanned aerial vehicles (UAVs) enables detailed tree species (TS) classification, surpassing the limitations of satellite imagery. In this paper, we propose an individual tree crown (ITC)-based loss and postprocessing approach, which can optimize classifying single tree species, designed to minimize the impact of overlapping canopies and dense mixed-species regions. Our methodology uses U-Net as the baseline model to integrate the proposed ITC-loss by treating each tree crown as a single unit to refine the predictions through a voting process. Furthermore, we apply a postprocessing method, merging the results with an ITC segmentation to refine the results. This reshapes the ITC and decides on a final species with a voting mechanism. Our results show that both approaches improve the mean intersection over union (mIoU) by 1.60 % for the best test, with a mIoU score for the baseline training of 42.03 % and 44.39 % after including our proposed methods.**

***Index Terms*—tree species classification, ITC loss, deep learning, tree crown segmentation.**

I. INTRODUCTION

Tree species (TS) classification provides essential information on forest biodiversity, which is crucial for forest management. It can be used for various purposes, such as forest inventory [1], biodiversity assessment [2], and monitoring changes in forests, such as disease spread or species distribution [1]. Traditional forest survey and inventory approaches are insufficient for collecting detailed TS information in large areas [2, 3]. In contrast, Remote Sensing (RS) technology allows TS to be distinguished more efficiently and frequently.

Over the past decades, deep learning (DL) and RS technologies have been proposed based on airborne and spaceborne imagery. Research has shown that high spatial resolution data improve the classification of TS, especially in environments with simple TS categories [3–7]. Other studies successfully classified 9 TS [4], while others even achieved classification for 13 species [3]. Currently, various approaches to TS classification have been introduced applying DL algorithms, including CNN and transformer models [3–5, 8]. In addition, approaches have been proposed to improve the classification of densely mixed forests. Using multitemporal data [3] or a digital surface model (DSM) [4] shows improvement.

However, DL-based TS classification remains challenging in dense forests. Overlapping branches result in multiple pixel-wise classifications of an individual tree crown (ITC). Furthermore, differences in individual TS regarding traits, texture, height, and vitality lead to inhomogeneous classification results. This results in unclear boundaries and possible misclassification of individual TS [3, 4, 9]. In addition, a precise ITC segmentation has the potential to improve the TS classification results, since each single tree crown inherently represents only one species. In this paper, we propose an ITC-based TS classification deep learning neural network architecture by introducing an ITC-based loss function. In addition, a postprocessing technique is adapted to achieve further refined classification results. We enhance semantic segmentation by incorporating knowledge of the tree crown shape, improving model accuracy and scalability for reliable TS classification systems.

II. MATERIALS AND STUDY AREA

The BAMFOREST benchmark dataset is adopted to assess the performance of our proposed approach. Captured using UAVs with a resolution of approximately 2 cm, it spans 105 hectares in three forest areas (Stadtwald, Tretzendorf 1 & 2) and one forest-like city park (Hain). The dataset includes 27,160 labeled tree crowns acquired in the summer of 2022 [10, 11]. The Hain area serves exclusively as a test region because of its distinct tree distribution, primarily deciduous species, compared to the predominantly coniferous Stadtwald and Tretzendorf. The data is pre-split into training, validation, and two test sets: one within the training region (Test-2) and one from Hain (Test-1), representing unseen conditions; see Table I. The BAMFOREST dataset features diverse TS with significant size, texture, and color variability. Deciduous species such as Beech and Oak exhibit dense canopies, while conifers such as Spruce and Fir have sparser, structured crowns. Differences in size and vitality further challenge the classification, as trees of the same species can vary significantly due to environmental factors. The dataset contains very high-resolution images and their corresponding ground truth annotations. These provide the input image of the trees

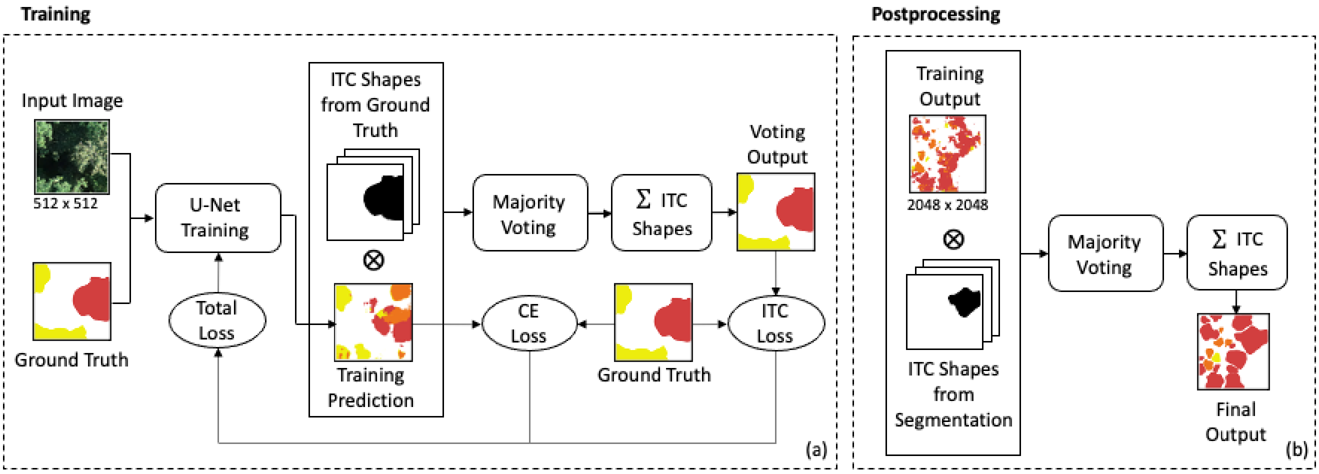

Fig. 1. Schematic overview of the two proposed methods for ITC-based semantic segmentation: (a) the ITC Loss method, which includes U-Net training with total loss combining cross-entropy (CE) loss and ITC loss for segmentation, and (b) the postprocessing method, involving majority voting for aggregation of segmentation maps and summation of ITC shapes.

and the labels that contain the TS information. Additionally, single-shape JSON files focus on individual tree structures, offering a binary representation that is used in training and postprocessing to refine segmentation results. Preprocessing involves categorizing TS, restructuring dead trees into a separate class, and modifying shapefiles to include only relevant data. These files are converted into TIFF format for ground truth usability and further adjusted to extract RGB channels from the original four-channel data. Patches of 512×512 pixels with 50 % overlap were created to ensure that tree traits can be meaningfully learned [10].

TABLE I
SUMMARY OF TREE DISTRIBUTION IN THE DATASET ACCORDING TO THE CLASSES. THE PERCENTAGES ARE ACCORDING TO THE NUMBER OF TREES AFTER TROLES ET AL. [10].

	Train	Val	Test-2	Test-1
N of trees	17,212	4,390	3,580	1,978
Pinus (Pine)	36.23 %	31.94 %	27.54 %	1.11 %
Fagus (Beech)	23.11 %	20.71 %	23.10 %	23.96 %
Quercus (Oak)	23.30 %	20.27 %	22.43 %	19.01 %
Picea (Spruce)	5.53 %	7.22 %	9.11 %	1.06 %
Larix (Larch)	2.70 %	1.75 %	2.21 %	1.26 %
Pseudotsuga (Douglas Fir)	1.12 %	1.80 %	1.20 %	0.15 %
Abies (Fir)	1.19 %	1.12 %	1.01 %	0.00 %
Other	6.83 %	15.19 %	13.41 %	52.88 %
Vital	86.34 %	84.99 %	83.97 %	91.20 %
Degrading	11.88 %	12.35 %	13.18 %	8.49 %
Dead	1.78 %	2.67 %	2.85 %	0.30 %

III. METHODOLOGY

Our approach enhances semantic segmentation for TS classification through innovative approaches that integrate pixel- and ITC-based segmentation techniques. Building on a baseline architecture, U-Net, two significant enhancements are proposed: ITC-based loss and postprocessing strategies. These approaches address the challenges of classifying ITCs from different species within very high-resolution datasets.

A. *ITC Loss Function*

Our first method incorporates an ITC-based loss function that improves end-to-end model training by integrating TIC as single units into the training process and introduces an accuracy loss, named ITC loss (see Figure 1(a)). This will improve the model's prediction of consistent species classifications for entire tree crowns by multiplying the predictions with the ITC shapes and majority voting. The ITC loss complements the standard categorical Cross Entropy (CE) loss by evaluating predictions at the tree crown level rather than solely on individual pixels.

The refined predictions are then compared to the ground truth to evaluate the consistency of the shapes and species classification. The number of wrong-classified pixels is counted. The overall pixels and the sum of the wrongly predicted classes are summarized. The final accuracy is calculated, resulting in a value between zero and one, described in Equation 1 as the ITC loss (L_{ITC}).

$$L_{\text{ITC}} = \frac{\text{Number of False Predictions}}{\text{Total Number of Predictions}} \tag{1}$$

This ITC-based loss is then summarized with the first pixel-based CE loss to a total loss. This total loss (Equation 2) combines the CE loss and the ITC loss. After the loss is calculated, it is fed into backpropagation.

$$\text{Loss} = \alpha \times L_{\text{CE}} + \beta \times L_{\text{ITC}} \tag{2}$$

The parameters can be weighted differently with α and β, allowing us to optimize the impact of the ITC loss.

B. *ITC-Based Postprocessing*

The second method introduces ITC-based postprocessing to refine initial predictions generated by the semantic segmentation model, see Figure 1(b). This method utilizes ITC seg-

TABLE II
TS CLASSIFICATION ACCURACIES SHOWING THE MIOU OF BOTH TEST-1 AND TEST-2 AND THE PIXEL-RELATED DATA SHARE IN PERCENT. HIGHLIGHTED VALUES SHOW THE BEST-PERFORMING VALUE ON EACH TEST, COMPARING THE BASELINE TESTING RESULTS TO THE TWO PROPOSED METHODS.

Class	Test-1			Test-2		
Name (Latin Name)	CE Loss [%]	CE + ITC Loss [%]	Postprocessing [%]	CE Loss[%]	CE + ITC Loss[%]	Postprocessing[%]
Background	59.56	58.21	**64.44**	62.47	64.96	**65.28**
Pine (Pinus)	**00.04**	00.02	00.00	50.22	50.78	**50.96**
Beech (Fagus)	11.25	21.24	**29.05**	32.83	35.24	**38.58**
Oak (Quercus)	23.99	19.52	**29.68**	33.24	32.40	**37.62**
Spruce (Picea)	00.04	**00.06**	00.00	23.73	23.31	**24.72**
Larch (Larix)	00.55	**00.59**	00.52	10.38	**11.03**	10.06
Douglas Fir (Pseudotsuga)	00.00	00.00	00.00	**00.97**	00.21	00.27
Fir (Abies)	00.00	00.00	00.00	00.00	00.00	00.00
Other Trees	**23.52**	14.53	16.25	09.72	09.63	**10.18**
Dead Trees	01.25	01.01	**01.32**	**17.34**	13.86	16.72
Mean Score	15.46	17.33	**20.93**	42.03	42.79	**44.39**

mentation results that identify individual tree crowns, ensuring spatial and classification precision. The ITC segmentation was obtained by Tian et al. [11]. Combining these shapes with our semantic segmentation output, we isolate each tree crown and apply a voting mechanism to determine its most dominant species classification. The majority voting mechanism selects the species based on the most pixels within a tree crown.

For semantic segmentation, 512×512 sized patches are best to enhance the traits of the trees, while ITC segmentation performed best using 2048×2048 patches because of more whole tree crowns per image. This concluded that the tinier patches of the semnatic segmentation are concatenated to the same size for postprocessing. This approach has no downside after finishing the training since it is a pixel-based result.

These two proposed methods, ITC loss and postprocessing, work synergistically to refine the baseline U-Net segmentation architecture. Postprocessing enhances spatial and classification accuracy after predictions, while ITC loss improves model training by integrating structural and contextual information. Together, these methods significantly advance the precision of TS classification in very high-resolution imagery, making them valuable for ecological mapping and forestry applications.

C. Implementation Strategy

Data were trained with a batch size of 10, using cross-entropy loss. AdamW was used with a weight decay of 0.011. 30 epochs were trained using an early stopping mechanism to avoid overfitting. Weighted loss and augmentations were applied but not finally chosen, as they did not take advantage of the training. A total of 10 classes were trained.

The metrics for evaluating the classification results are measured with intersection over union (IoU). As described by Kattenborn et al. [12], the IoU measures the relative spatial agreement between the reference and predicted surfaces. It is measured with true positives (TP), false positives (FP), and false negatives (FN).

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \quad (3)$$

IV. RESULTS

The ITC loss function, implemented in the U-Net baseline model along with postprocessing, demonstrated measurable performance improvements. For Test-Set-1, the mean intersection over union (mIoU) increased by 5.47%, while Test-Set-2 showed a minor improvement of 2.36% (Table II). The ITC loss showed improvements for some classes. Postprocessing, performed on the output generated with the ITC loss, further enhanced results, improving the IoU across most classes.

The two test datasets displayed different results due to their distinct characteristics. Test-Set-2, which closely resembles the training data, aligned better with the model improvements. Classes well represented and trained in the baseline U-Net model, such as Beech and Oak, showed significant performance gains. In contrast, poorly performing classes with limited training data below 5%, such as Larch, Douglas Fir, and Dead Trees, showed minimal or no improvement.

Test-Set-1, on the other hand, exhibited a different species distribution than the training data. This led to lower baseline performance, although species with abundant data, such as Beech, Oak, and Dead Trees, still showed notable improvements. Classes combining multiple TS or representing dead trees were less responsive to the ITC loss, with the classes ”Other Trees” and ”Dead Trees” showing limited performance gains due to insufficient training and data variability.

For species with more than 20% data share, the proposed methods consistently improved the results over the baseline CE loss model under the condition that the baseline is well trained. These findings underscore the effectiveness of the ITC loss function and postprocessing in improving semantic segmentation, particularly for well-trained and well-represented classes.

Figure 2 displays selected results showing the improvement of both methods. The ITC loss improved the identification of wrong-classified classes, refined shapes, and reduced noise. After postprocessing, the ITCs can be distinguished clearly in the results, and one species is voted per tree. This eliminates noise and further refines the classification into full tree shapes.

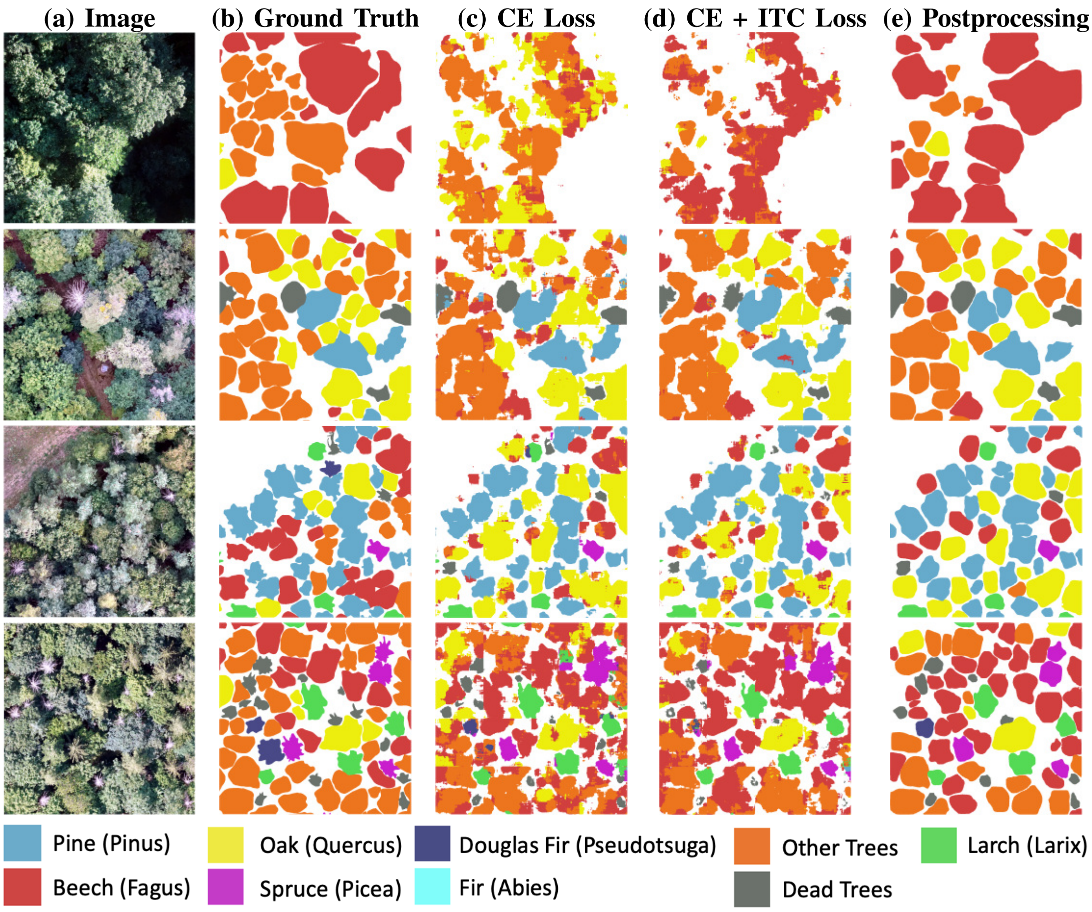

Fig. 2. Result comparison of 2048 x 2048 images (a), ground truth (b), predictions (CE loss) (c), predictions (CE and ITC loss) (d), and postprocessing (e).

V. Discussion and Conclusion

Improving TS classification is crucial for effective forest management and biodiversity monitoring. This research contributes to more accurate and efficient forest inventories and ecological surveys by applying advanced DL models. The findings can support similar efforts in other forest ecosystems, contributing to global biodiversity conservation. By providing information at individual tree level, they enable large-scale mapping and analysis based on detailed, tree specific data.

The proposed ITC-based loss function and postprocessing methods improved significantly for classes with sufficient representation and initial performance. We show that using ITC-based methods to enhance training supports a more precise use of TS classification for forest management and analysis. The performance gap between Test-1 and Test-2 emphasizes the model's reliance on dataset similarity. The poor results of Test-1 indicate limited generalizability. However, since the testing data differed from the training data, this can not be associated with our proposed method. Postprocessing proved effective in correcting pixel-based classification and separating dense and overlapping tree crowns, benefiting all well-trained classes. It has to be mentioned, reliance on predefined shapes from the ITC-segmentation results may restrict generalizability. Developing adaptive postprocessing strategies to accommodate more diverse datasets could enhance its utility. The visual output after postprocessing facilitates individual tree analysis in forest management.

However, the inability to improve minority classes, such as Douglas Fir and Fir, underscores the need for balanced datasets and further baseline training optimization. Future work should focus on strategies such as more diverse data implementation and training, further data augmentation, or weighted methods, to achieve sufficient baseline training for all classes that can guarantee a better baseline for the proposed method that uses natural forests as data with highly varying species occurrence.

References

[1] BMEL, "Ergebnisse der vierten Bundeswaldinventur," 4 2024. [Online]. Available: https://www.bmel.de/DE/themen/wald/wald-in-deutschland/bundeswaldinventur.html

[2] Y. Sun, J. Huang, Z. Ao, D. Lao, and Q. Xin, "Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne LiDAR and high-spatial-resolution remote sensing images," vol. 10, no. 11, p. 1047, number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1999-4907/10/11/1047

[3] V. Ramesh, A. Ouaknine, and D. Rolnick, "Tree semantic segmentation from aerial image time se-

ries," number: arXiv:2407.13102. [Online]. Available: http://arxiv.org/abs/2407.13102

[4] F. Schiefer, T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, and S. Schmidtlein, "Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks," vol. 170, pp. 205–215. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620302938

[5] M. P. Ferreira, D. R. A. d. Almeida, D. d. A. Papa, J. B. S. Minervino, H. F. P. Veras, A. Formighieri, C. A. N. Santos, M. A. D. Ferreira, E. O. Figueiredo, and E. J. L. Ferreira, "Individual tree detection and species classification of amazonian palms using UAV images and deep learning," vol. 475, p. 118397. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037811272031166X

[6] G. D. Pearse, M. S. Watt, J. Soewarto, and A. Y. S. Tan, "Deep learning and phenology enhance large-scale tree species classification in aerial imagery during a biosecurity response," vol. 13, no. 9, p. 1789, number: 9 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/13/9/1789

[7] S. Nezami, E. Khoramshahi, O. Nevalainen, I. Pölönen, and E. Honkavaara, "Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks," vol. 12, no. 7, p. 1070, number: 7 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/12/7/1070

[8] S. Ecke, F. Stehr, J. Frey, D. Tiede, J. Dempewolf, H.-J. Klemmt, E. Endres, and T. Seifert, "Towards operational UAV-based forest health monitoring: Species identification and crown condition assessment by means of deep learning," vol. 219, p. 108785. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924001765

[9] M. Cloutier, M. Germain, and E. Laliberté, "Influence of temperate forest autumn leaf phenology on segmentation of tree species from UAV imagery using deep learning," vol. 311, p. 114283. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425724003018

[10] J. Troles, U. Schmid, W. Fan, and J. Tian, "BAMFORESTS: Bamberg benchmark forest dataset of individual tree crowns in very-high-resolution UAV images," vol. 16, no. 11, p. 1935, number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/16/11/1935

[11] J. Tian, D. Panangian, W. Fan, B. Siegmann, and X. Yuan, "Deep learning based individual tree crown delineation from panchromatic aerial imagery," in *ISPRS Geospatial Week 2025*, 2025, submitted.

[12] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," vol. 173, pp. 24–49. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620303488