Digital twinning of all forest and non-forest trees at national level via deep learning

Sizhuo Li, Martin Brandt, Rasmus Fensholt, Ankit Kariryaa, Christian Igel, Fabian Gieseke, Thomas Nord-Larsen, Stefan Oehmke, Ask Holm-Carlsen, Samuli Junttila, et al.

Digital twinning of all forest and non-forest trees at national level via deep learning

Sizhuo Li
Københavns Universitet
Martin Brandt (martin.brandt@mailbox.org)
Københavns Universitet https://orcid.org/0000-0001-9531-1239
Rasmus Fensholt
University of Copenhagen https://orcid.org/0000-0003-3067-4527
Ankit Kariryaa
Københavns Universitet
Christian Igel
University of Copenhagen https://orcid.org/0000-0003-2868-0856
Fabian Gieseke
Københavns Universitet
Thomas Nord-Larsen
Københavns Universitet
Stefan Oehmke
Københavns Universitet
Ask Holm-Carlsen
The Danish Agency for Data Supply and Efficiency
Samuli Junttila
Department of Forest Sciences, University of Eastern Finland
Xiaoye Tong
University of Copenhagen
Alexandre d'Aspremont
Department of Computer Science, École normale supérieure
Philippe Ciais
Laboratoire des Sciences du Climat et de l'Environnement https://orcid.org/0000-0001-8560-4943

Article

Keywords:

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1661442/v1

Digital twinning of all forest and non-forest trees at national level via deep learning

Sizhuo Li*, Martin Brandt*, Rasmus Fensholt, Ankit Kariryaa, Christian Igel, Fabian Gieseke, Thomas Nord-Larsen, Stefan Oehmcke, Ask Holm Carlsen, Samuli Junttila, Xiaoye Tong, Alexandre d'Aspremont, Philippe Ciais

Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark

Sizhuo Li, Martin Brandt, Rasmus Fensholt, Thomas Nord-Larsen & Xiaoye Tong

Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Ankit Kariryaa, Christian Igel, Fabian Gieseke & Stefan Oehmcke

Department of Information Systems, University of Münster, Münster, Germany

Fabian Gieseke

The Danish Agency for Data Supply and Efficiency, Copenhagen, Denmark

Ask Holm Carlsen

Department of Forest Sciences, University of Eastern Finland, Joensuu, Finland

Samuli Junttila

Department of Computer Science, École normale supérieure, Paris, France

Alexandre d'Aspremont

Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

Philippe Ciais

Université Paris-Saclay, Gif-sur-Yvette, France

Sizhuo Li

*Correspondence to sizli@ign.ku.dk or mabr@ign.ku.dk

Abstract

Intelligent forest management is the key to mitigating climate warming, fostering a green economy, and protecting valuable habitats. Detailed knowledge about forests is a prerequisite for such management and is conventionally based on costly plot-scale data, rarely available at resolution of relevance for local forest management strategies. Here, we present a deep learning-based framework that provides location, crown area and height for each individual tree from aerial images at country scale. We quantify and characterize all individual trees in Denmark and show that 26% of the trees grow outside forests, which is typically unrecognized in national inventories. Furthermore, we demonstrate that only marginal effort is needed to transfer our framework to Finland, despite markedly dissimilar landscapes and data sources. Our work lays the foundation for a global database, where every tree has its digital twin and is spatially traceable and manageable.

Introduction

Climate change and rapid losses of forest habitats and biodiversity are the major environmental challenges of the 21st century [1,2]. Intelligent forest management can mitigate these crises by building carbon stocks, providing materials for a green economy, and developing habitats representing the most important reservoir for biodiversity in the world [3,4]. Consequently, policies addressing climate change mitigation and adaptation, sustainable wood production, and biodiversity must rely on timely, detailed, and reliable information on the state and development of tree resources and habitats.

Detailed knowledge of forests at regional and national scales is commonly obtained from inventories such as the national forest inventories (NFI). Here, variables such as tree diameter, height, species, growth, and mortality are recorded during repeated census on a representative sample of widely distributed plots [5–8]. Such inventories provide essential information, but are time-consuming, labor-intensive, and costly. Furthermore, the scarce number of plots is of limited relevance for local forest management strategies. Comprehensive information on forest structure is commonly derived from a more intensive sampling of inventory plots, sometimes complemented by remote sensing and simple machine learning tools [9–11], but only a few countries conduct a detailed sampling on a systematic and frequent basis. Current state-of-the-art approaches for the assessments of individual trees are based on airborne LiDAR data [12–14] and have the potential to meet the requirements for assessing forest carbon stocks and to support management and conservation policies. However, these data are costly, are performed on a snapshot basis during irregular campaigns, and the requirements for obtaining and processing the data are high.

Satellite-based monitoring of forests enables low-cost and wall-to-wall assessments that can be rapidly repeated at a high temporal frequency and a large scale. Yet these assessments are typically limited to the variable “forest cover” [15,16]. This variable has a long tradition in forest monitoring, but has little use from a management perspective, and the misinterpretation of the results can cause far-reaching consequences [17,18]. Moreover, forest biomass estimations based on “cover” information and spatially aggregated height or volume proxies ignore the diversity of trees, i.e., the fact that trees generally have different height and crown sizes, leading to highly uncertain carbon stock estimations [19]. Forest cover also ignores trees outside forests, which can constitute a considerable woody resource and provide a variety of ecosystem services [20,21]. Previous studies have shown that the count of trees outside forests far exceeds expectations [22], but their systematic assessment remains a challenge [23]. Finally, the provisioning of ecosystem services, such as forest resources and habitats as well as climate change

mitigation and adaptation, is closely related to the distribution and size of individual trees, which are difficult to measure by traditional optical satellite systems.

Recently, Brandt *et al.* [22] have shown that advances in computer vision and very high resolution (50 cm) satellite imagery enable the mapping of individual trees and shrubs in the sub-Saharan desert and savanna landscapes. Yet, it has been questioned whether the approach designed for mapping single trees growing in isolation in dry areas could be transferred to the European forest setting, where closed forests prevail. Moreover, Brandt *et al.* [22] did not embark upon assessing the height of individual trees, which is an essential variable for estimating biomass and carbon stocks. Ultra-high-resolution aerial imagery is publicly available for many European countries and is frequently updated. Identifying single tree crowns from these images is not difficult for the human eye, and deep learning methods have achieved great success in solving similar problems, including microscopy cell segmentation [24–26] and scene labeling [27,28]. Studies in the domain of crowd counting have also demonstrated that dense representations of objects could be accurately enumerated via a density estimation approach [29,30]. Though the feasibility of tree crown segmentation and counting in imagery has never been verified for closed forests. Furthermore, inferring height from merely optical images remains an important yet understudied problem, especially at the level of individual trees [31,32].

Here, we present an automatic and scalable airborne tree inventory framework based on state-of-the-art convolutional neural networks [33,34], producing detailed individual tree attributes including location, crown area, and tree height (Fig. 1). The proposed method allows for an unbiased analysis of trees growing in diverse landscapes including dense forests, open fields, and urban areas, and, hence, leads to accurate estimates of tree resources at a large scale. We applied the framework on aerial images covering Denmark (2018) at a spatial resolution of 20 cm and generated country-wide maps on the number, location, crown size, and height of every single tree, both inside and outside forests. We further verified the adaptability of our method by transferring model weights learned from data in Denmark to Finland, the most forested country in Europe with over 56% forest cover (2015) [16]. The Finnish dataset has a different spatial resolution (50 cm) and spectral band composition. We demonstrate that, once a model is trained, it can be easily adapted to different years and countries, and height estimation of each tree can be obtained without further need for LiDAR data. We expect this framework to revolutionize the baseline information from which European forests are managed, becoming an important complement to inventories supporting forest management and policy planning.

Results

The proposed framework involves two separate models addressing three localization and characterization tasks of individual trees (Fig. 1). The first model solves the tree counting and crown segmentation tasks jointly from multi-band aerial images and, if available, a canopy height map derived from LiDAR data. The second model uses LiDAR data as training data and predicts canopy heights from multi-band aerial images. The predicted canopy heights are further combined with the crown segmentation results by zonal statistics to obtain the height per tree.

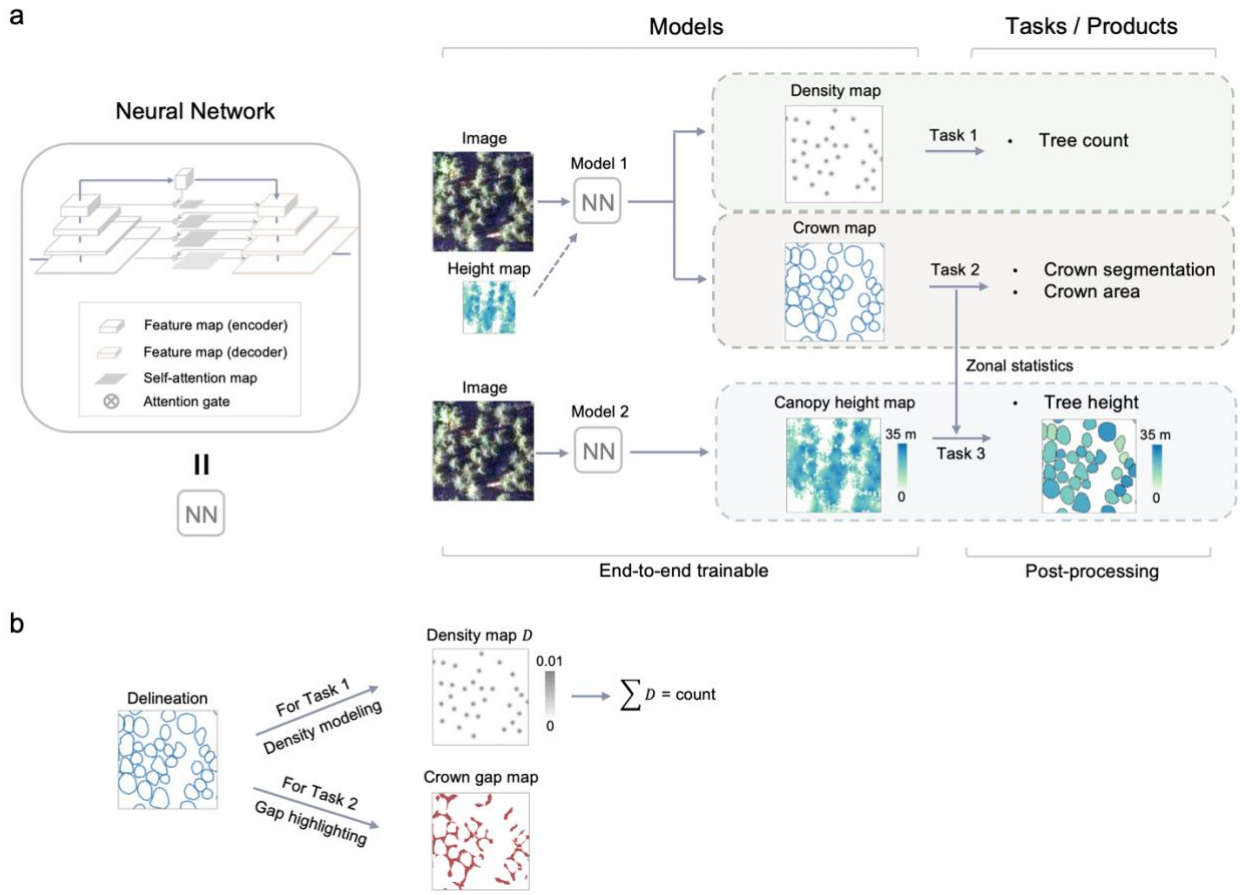

Fig. 1 | **Overview of the framework used to count individual trees and predict their crown area and height. a**, Deep learning-based framework for individual tree counting, crown segmentation, and height prediction. Spatial locations of individual trees are incorporated in the tree density maps and the crown segmentation maps. The canopy height map (CHM) derived from LiDAR data provides pixel-wise height information, which, when available for a specific study area, can optionally be used as an additional input band for the individual tree counting and crown segmentation tasks. **b,** Data preparation and modeling for tree counting and crown segmentation. The manually delineated individual tree crowns are modeled as density maps for the counting task by extracting the polygon centroids. The gaps between adjacent crowns are highlighted for the separation of individual tree crowns during the training phase.

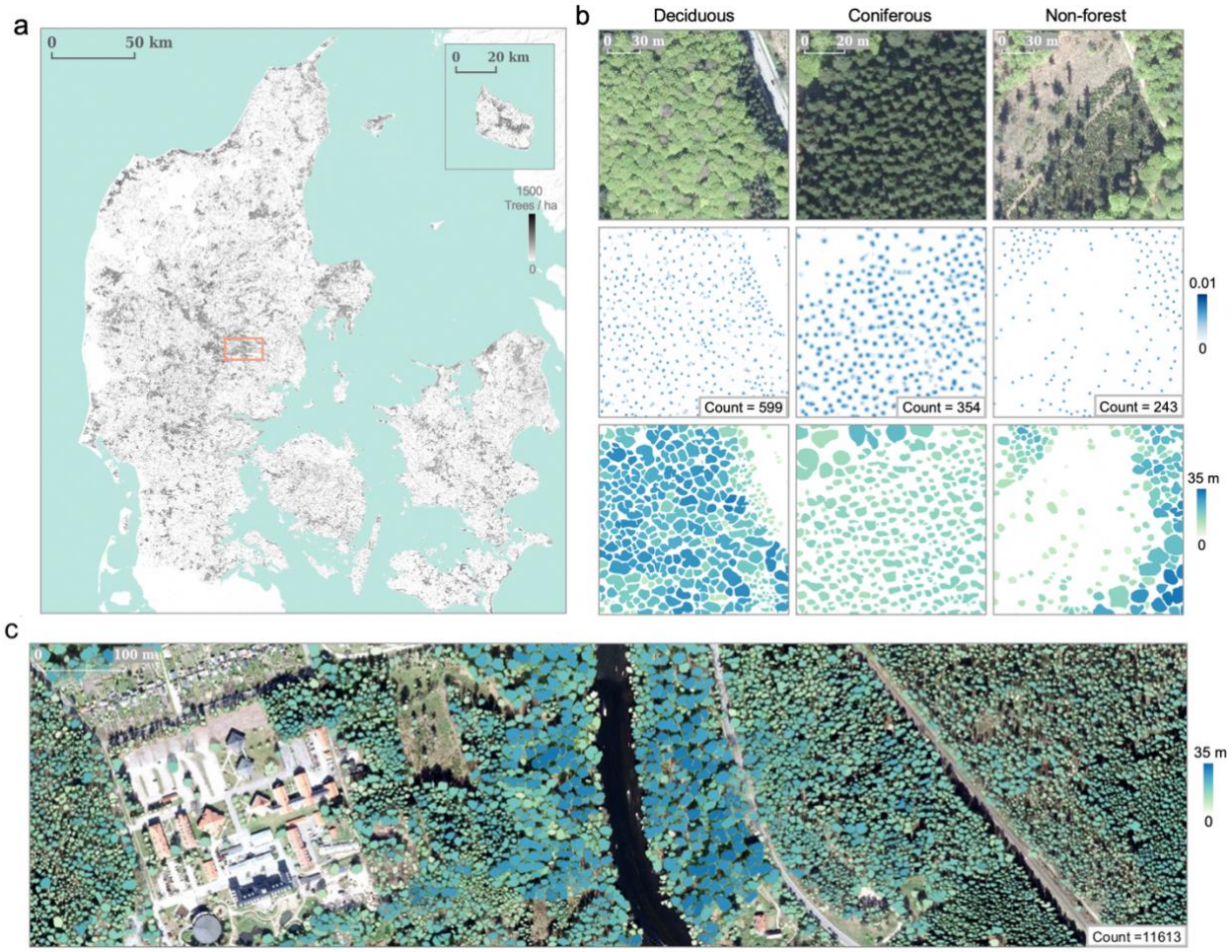

Fig. 2 | **Example products from the proposed framework. a,** Wall-to-wall tree count prediction for Denmark. **b,** Detailed examples showing the individual tree counting (second row), crown segmentation (third row), and height prediction (third row) from three major types of landscapes (deciduous forest, coniferous forest, and non-forest). **c,** Large-scale individual tree crown segmentation results colored by height predictions. Examples in **b** and **c** were sampled from the region indicated by the orange box in **a**.

Multi-task deep learning enables simultaneous tree counting and crown segmentation

We established a multi-task deep learning-based network for jointly solving the individual tree counting (Task 1 in Fig. 1) and crown segmentation tasks (Task 2 in Fig. 1) from 2-dimensional imagery for both forest and non-forest trees [16]. The network could be retrained with different spectral band compositions and with or without a height map, leading to separate models with similar performances (Extended Data Fig. 1, Extended Data Table 1, and examples in the last section). Here, we used RGB and near-infrared (NIR) aerial images at 20 cm resolution from summer 2018 and a canopy height map projected from airborne LiDAR data at 40 cm resolution. The network was primarily adapted from the U-Net architecture [33], with two output branches for the counting and crown segmentation tasks (see "Methods"). As target references, 19,771 individual tree crowns from different forest and non-forest landscapes were manually delineated by visually inspecting the aerial images (Supplementary Fig. 1a). We labeled all trees with identifiable shadows, and adjoining crowns were delineated as separate individual segments. For model evaluation as described below, we created an independent test dataset with 2,679 annotated tree crowns in stratified sampling spots distributed all over Denmark (Supplementary Fig. 1b).

Forest trees tend to naturally exhibit a clustered spatial pattern. Thus, we proposed two techniques to ensure both accurate total counts and crown segmentations at the level of individual trees (Fig. 1b). Firstly, for the segmentation branch, the gaps in between neighboring crown boundaries were used during training for computing a loss that fostered the separation of adjoining crowns [22]. We adjusted the penalty weight assigned to the between-crown gaps to optimize both the mapping of crown coverage and separability of the crowns. With an increasing gap penalty from low (1) to high (10), individual crowns could be separated more clearly (counting error - 9.4%), while the predicted crown area declined (error + 10.7%, Extended Data Fig. 2). We found a gap penalty of 5 to adequately balance both the individual tree separation and crown area accuracy (Extended Data Fig. 2). Secondly, instead of counting the segmented tree crowns, we solved the counting task through a partially independent branch, which was not impacted by the crown clustering effects from the segmentation branch. Here, counting was conducted by a density estimation approach [29], where each sample point on the density map represented a tree at the corresponding position and the integral of the density map was equivalent to the total tree count in an image of arbitrary size. The density estimation-based approach improved the counting performance by 8.9% as compared to the segmentation-based approach (Extended Data Fig. 4).

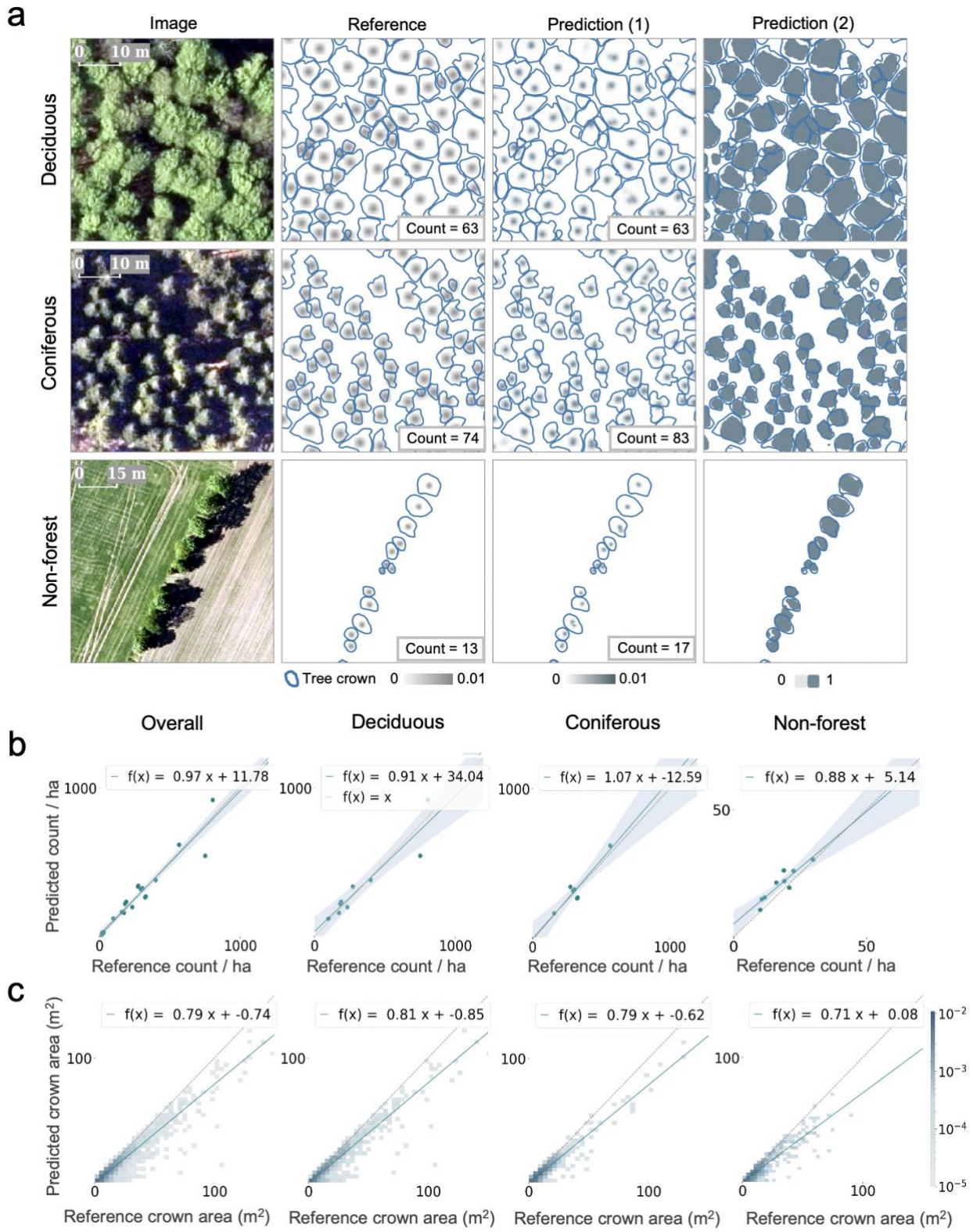

Fig. 3 | **Individual tree counting and crown segmentation performance on the test dataset. a**, Examples from three different forest/landscape types: deciduous forest, coniferous forest, and non-forest areas. Reference shows the target labels, including the manual crown delineations (thin blue lines) and the Gaussian-blurred crown centroids (gray dots). Prediction (1) shows the counting by density estimation results and Prediction (2) shows the crown segmentation results, both overlaid with the manual delineations (thin blue lines). **b**, Evaluation of the tree counts from Prediction (1), grouped in respective landscapes and re-scaled to tree counts per hectare. Here, each scatter point in the plots represents a sampling image of varying size. The regression lines are shown in blue and the identity lines are shown in gray. **c**, Evaluation of the tree crown area predictions from Prediction (2) at individual-tree-level accuracy.

We evaluated the robustness of the multi-task network for three major types of landscapes [16], including high-density deciduous forests (9 images of average size around 0.4 hectares containing 1,279 trees in total), high-density coniferous forests (7 images of average size around 0.6 hectares containing 853 trees in total), and open fields (non-forest) involving trees outside forest in hedgerows and small patches (9 images of average size around 3.4 hectares containing 547 trees in total). The model achieved an overall low bias and high accuracy, as reflected by the close to unity slope (0.97) for the regression of predicted against reference tree numbers and a high coefficient of determination R^2 of 0.93 (see Eq. (6) in "Methods"). However, we noticed a slight overcount in dense coniferous forests (slope = 1.07), yet a marginal undercount in dense deciduous forests (slope = 0.91) and open areas (slope = 0.88) (Fig. 3). We also noted that the model underestimated the crown cover by approximately 20% (Fig. 3c) regardless of the tree density or types, which was likely due to the choice of the gap penalty for improving the separability of individual trees. For large-scale applications, we corrected for this underestimation by rescaling the predicted crown area according to the fitting curves (see "Methods"; Fig. 3c).

Individual tree height prediction from aerial images

The height prediction model received multi-band aerial images as inputs and learned the mapping of the reference height obtained from the LiDAR data through a similar U-Net architecture as used above. To account for the height differences among various landscapes, we constructed a training dataset by sampling aerial images from regions dominated by deciduous, coniferous, and non-forest trees with a ratio of 18:18:1 [16] (Supplementary Fig. 2a). The dataset contained in total 74 images (7.4 k hectares) captured in 2018, with the corresponding LiDAR height data collected primarily from 2018 and partially from 2019 (due to lack of coniferous instances). The pixel-level height prediction, combined with the individual tree crown segmentation, yielded the individual tree height, which we defined as the maximum height within each predicted tree crown.

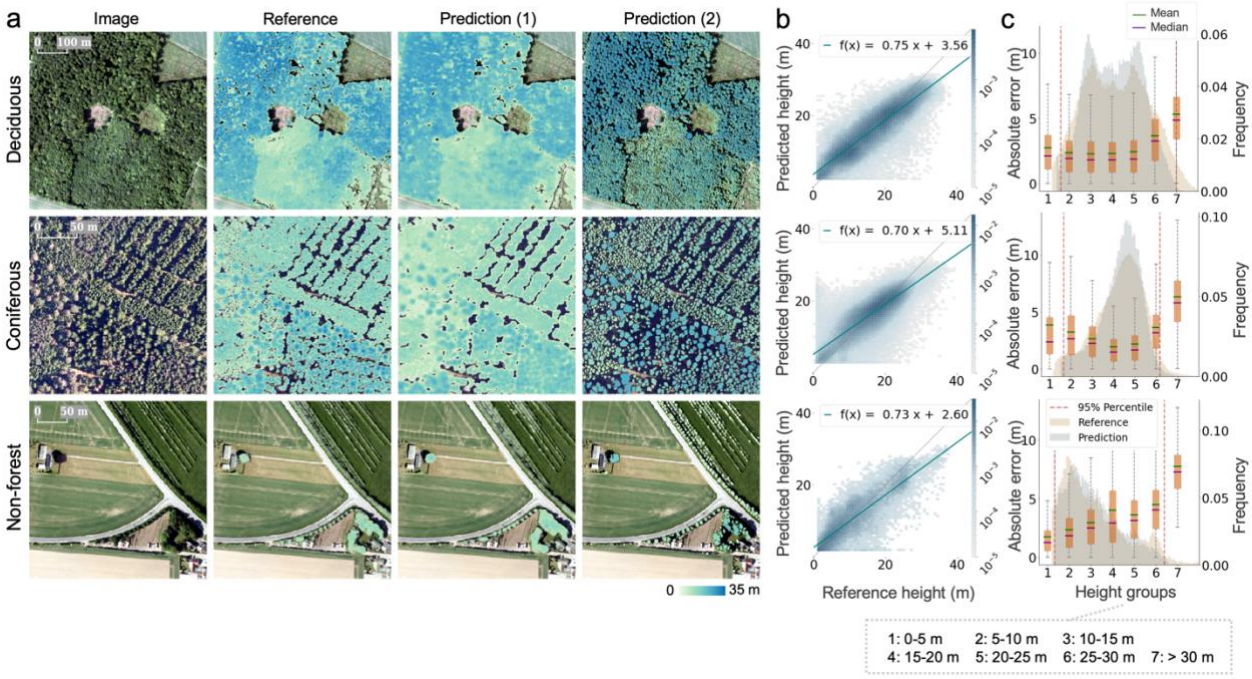

Fig. 4 | **Individual tree height prediction from aerial images.** The evaluation was done on regions randomly sampled from approximately one-third of Denmark (Supplementary Fig. 2b). **a,** Height predictions for three different forest/landscape types:

deciduous, coniferous, and non-forest. Prediction (1) shows the height prediction per pixel and Prediction (2) shows the height prediction per tree obtained by using the maximum value from Prediction (1) within each segmented tree crown. **b,** A comparison between individual tree heights derived from Prediction (1) and from Reference, with the regression lines shown in blue and the identity lines shown in gray. **c,** Absolute errors of the individual tree height prediction, grouped in 5 m height intervals, with the predicted and reference height distributions in the background.

We conducted large-scale individual tree height evaluation on randomly sampled aerial images (in total 3 k hectares containing 479 k predicted tree crowns) captured in 2018 and 2019 (in total 1.8 million hectares, approximately one-third of Denmark, Supplementary Fig. 2b), respectively for each forest/landscape type (Fig. 4). We noticed a high agreement between the predicted and the reference tree heights across all forest/landscape types, with an R^2 of 0.79, 0.66, and 0.73 (Fig. 4a, 4b). The model achieved a global median absolute error of 1.9 m with only rare extreme errors over 5 m. The errors gradually increased from less than 2 m to more than 5 m as the reference tree heights increased from 0-5 m to above 30 m, showing a tendency to underpredict the height of the taller trees (Fig. 4c). The relative error (see "Method") for all trees in the evaluation dataset was 0.68%, implying no systematic errors on large-scale aggregated performances (see also Fig. 4c). When decomposing the errors into bias and variance [35], we found the bias increased with tree height, while the variation, which reflected the model's capability to capture height variability, maintained constant across diverse height ranges and forest/landscape types (Extended Data Fig. 5c).

Nation-wide implementation in Denmark

We evaluated the scalability of the framework by generating an individual tree count map and a crown segmentation map featured with individual tree heights for Denmark (Fig. 2 and Supplementary Fig. 3). A total of 289 million trees were detected based on density estimation and a total crown area of 0.47 million hectares was predicted (Table 1). The results revealed a surprisingly large number of non-forest [16] trees (76 million), which represents around 30% of the national tree crown coverage and 26% of the national tree count. Individual tree heights were obtained by combining the refined crown segmentations with the per-pixel height predictions, which were corrected based on the predicted against reference height regressions (see "Methods"). Compared with the Danish NFI forest tree count from 2018 [36], which upscaled field measured plot information to nation-wide forest areas, our predictions showed an underestimation of tree counts of 15.2% for the medium and tall tree classes (20 - 40 m). For small trees (<= 6 m), the underestimation was severe (94.3%, Extended Data Fig. 6), which can be explained by the fact that understory in dense forests cannot be seen by aerial or satellite images [37]. Note that NFI data only estimated tree counts for forest areas, and we excluded non-forest trees for this comparison.

Our tree cover maps produced from the detection of individual trees leave little space for uncertainty and can be used to evaluate existing products. Here, we compared the tree cover map aggregated from our individual tree crown segmentation with two existing state-of-the-art forest cover maps estimated from satellite imagery at 30 m (Landsat) [15] and 10 m (Sentinel-2) [16] resolutions, and noticed a much higher tree canopy area in dense forests from these existing products (Extended Data Fig. 7a). In particular, the 10 m resolution Copernicus tree cover map (2018) [16] showed 32.9% higher values for deciduous forests and 50.7% for coniferous forests, and conversely an underestimation of 50.3% for non-forest areas (Extended Data Fig. 7b and Extended Data Table 2).

Table 1: Tree count, total crown area and tree height products for Denmark, grouped in three major forest/landscape types [16].

Forest/Landscape type [16]	Tree count	Total tree crown area (ha)	Tree height (median [10% ~ 90%], m)
Deciduous Forest	127,442,201 (44.1%)	233,720 (50.0%)	12.5 [3.6 ~ 24.8]
Coniferous Forest	85,398,027 (29.5%)	92,352 (19.8%)	12.9 [5.1 ~ 23.0]
Non-forest	76,320,387 (26.4%)	141,186 (30.2%)	6.7 [0.5 ~ 17.2]
Total	289,160,615 (100%)	467,257 (100%)	11.6 [2.7 ~ 23.6]

Transfer learning enables cross-national applications

The proposed framework showed high transferability regardless of the input data source, spatial resolution, composition of the spectral bands, and differences in major forest/landscape types. The models pre-trained with the Danish dataset at 20 cm resolution were easily adapted to 50 cm aerial images from Finland by fine-tuning using a small additional training set of the target distribution (Fig. 5). Specifically, the counting and crown segmentation model established for Denmark was further trained with the original data from Denmark and additional data from Finland including up-sampled coarser resolution images and 4,773 tree crown delineations (Supplementary Fig. 4a). Likewise, the pre-trained height prediction model using data from Denmark was adapted to the Finnish setting by fine-tuning with 10.8 k hectares of images and 1 m resolution LiDAR height data collected from three locations in Finland (Supplementary Fig. 4b). Notably, when the models trained for Denmark were directly applied to Finland, reduced performances were observed, yet the main patterns including the general tree crown shapes as well as overall height differences were captured (Extended Data Fig. 8).

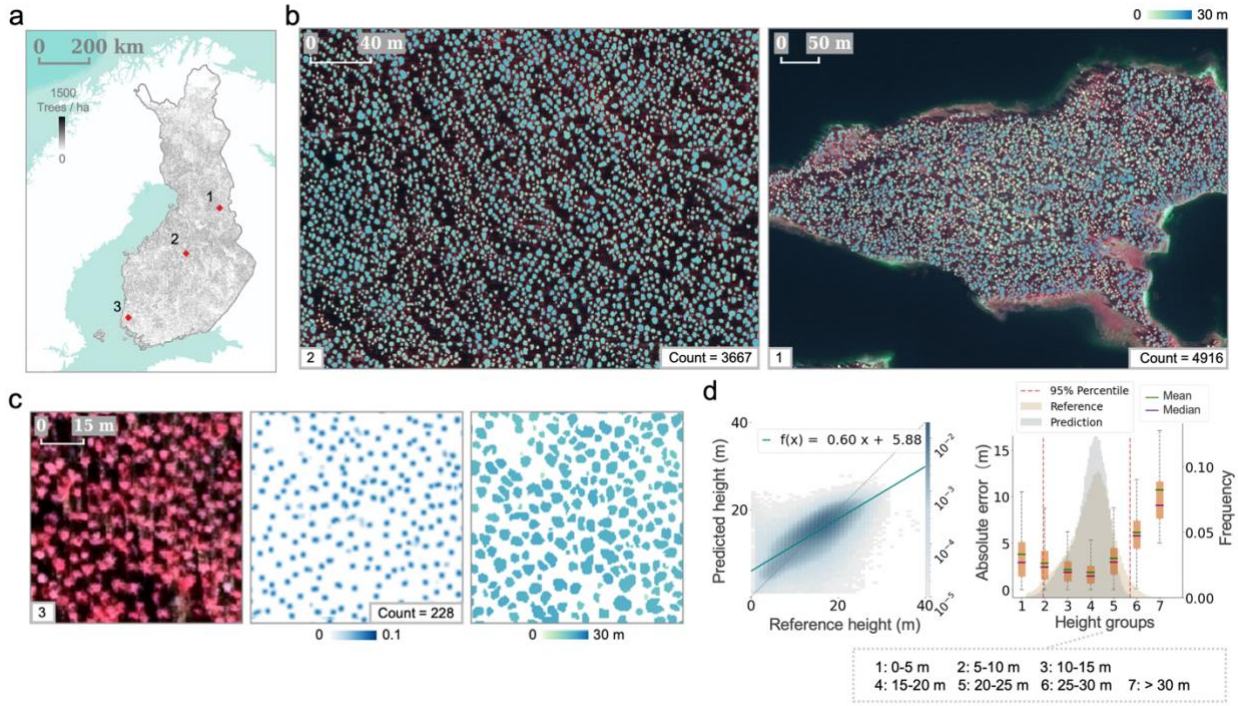

Fig. 5 | **Transferring the proposed framework from Denmark to Finland. a,** Tree count prediction in Finland derived from aerial images using the transferred model from Denmark. The evaluation locations for **b**, **c**, and **d** (red dots). **b,** Large-scale individual tree crown segmentation colored by height prediction. **c,** Detailed examples for individual tree counting, crown segmentation, and height prediction. **d,** A comparison of individual tree heights from predictions and references, with the regression line in blue and the identity line in gray. Absolute errors for the evaluation of individual tree height prediction.

We evaluated the performance of individual tree height prediction on randomly sampled images (3 k hectares containing 833 k predicted tree crowns) from three regions in Finland (21.6 k hectares), where images and LiDAR height data were collected during the same period (2019) (Fig. 5). The selected regions were dominated by either coniferous forests or non-forest areas [16]. The individual tree height predictions showed a reasonably high agreement with the reference heights for most trees ($R^2 = 0.57$), and the global median absolute error was 1.9 m. The absolute errors increased slightly for smaller (< 5 m) and taller trees (> 30 m), while only a tiny proportion (<< 1%) yielded errors up to 10 m (Fig. 5d). The relative error for all trees in the evaluation dataset was 1.0%, indicating low systematic errors at a large scale.

Discussion

We established a novel end-to-end deep learning-based framework for individual tree mapping and height prediction in forest and non-forest areas from high-resolution aerial images and applied it to two European countries with dissimilar datasets and landscapes. Our approach enabled the derivation of the height information, normally only available from high-cost LiDAR data, from less expensive aerial imagery, with an unprecedented accuracy down to the level of single trees [32,38,39]. While aerial images cannot be considered "low cost", the availability of sub-meter resolution images from nano-satellites, such as Skysat, promises a comparable quality for a reasonable price. We propose such an individual tree localization and characterization approach as the means to produce a comprehensive tree database that concerns not only forest but also non-forest trees, which are essential yet often neglected by the conventional forest inventories [23]. Such a database would serve as the new standard for supporting local or national forest management. The highly detailed database consisting of tree counts, crown delineations and heights would also allow for downstream analyses such as monitoring self-thinning phenomena [40] for dense forests (Supplementary Fig. 5) or investigating the impact of global warming and drought on tree mortality and resilience. This information will improve the way how forest resources are estimated and how their sustainable utilization is optimized. Furthermore, we publish a readily available framework pre-trained on data samples from Denmark and Finland, which can be easily adapted to other domains by fine-tuning with a little extra data from specific landscapes. This will enable countries to make use of their aerial images in order to derive annual wall-to-wall airborne tree inventory with a marginal effort.

The resulting individual tree crown and count maps reflect the vegetation structure more accurately than existing conventional forest cover maps, and the products can be intuitively interpreted and validated by the human eye. Our method unambiguously determined trees as objects, whilst other well-known methods estimate the percentage canopy cover from spectral colors at relatively coarse resolution (10 - 30 m) [15,16]. Notably, canopy cover maps obtained by simply aggregating the individual tree crown maps might suffer from an underestimation of the true canopy coverage, since only the topmost tree crowns are clearly recognizable from above and we intentionally emphasized the between-crown gaps. Yet, we argue that existing forest cover maps [15,16] possibly overestimate the canopy coverage by not excluding gaps between individual trees, which are not seen in mixed pixels at coarse resolution [41,42].

Results such as presented here could make the upscaling of plot-based information to a national scale more accurate. The localization of individual trees is particularly important for the monitoring of tree mortality, which would not be based on plot-scale estimations, but on actual counts with wall-to-wall coverage. Moreover, carbon stocks could be reported at the level of individual trees in the frame of climate agreements. Using local or global allometric equations [43], crown diameter and height can be

directly converted to carbon stocks, so the upscaling from inventory plots would remain at the level of single trees. Going beyond areal variables and towards single tree assessments also reduces uncertainties related to forest area definitions. We found that following existing forest maps and definitions, about 26% of the trees in Denmark were missed by the NFI statistics [36], as they were growing outside of forests. Note however that short trees growing under the top canopies are underestimated and NFI data needs to be used to correct for the undercount bias.

Uncertainties could arise when transferring the established models to markedly different regions, considering the divergences in forest type distributions, traits of trees affected by local climate, image acquisition times, viewing angles, and spatial resolutions. Yet, our study revealed the feasibility of pre-trained models for multi-country scale automatic individual tree localization and characterization, which could be further extended to a continental or global scale with the availability of high-quality aerial or satellite imagery. A database on individual trees would serve as a digital twin of global forests, where each tree could be traced, for example, from the forest to the factory. Such detailed information would allow more sophisticated and attentive utilization of the wood material as wood properties are influenced by the local growing conditions, leading towards resource efficiency and sustainable utilization of forests. Our proposed framework may assist researchers, stakeholders, and policymakers in monitoring tree resources in an unprecedented way, ultimately promoting more intelligent and digitalized environmental management in support of the green transition.

References

1. Lindner, M. *et al.* Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Forest Ecology and Management* **259**, (2010).
2. Brondizio, E. S., Settele, J., Díaz, S., Ngo, H. & Guèze, M. *IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.* (2019).
3. Canadell, J. G. & Raupach, M. R. Managing Forests for Climate Change Mitigation. *Science (1979)* **320**, (2008).
4. Forster, E. J., Healey, J. R., Dymond, C. & Styles, D. Commercial afforestation can deliver effective climate change mitigation under multiple decarbonisation pathways. *Nature Communications* **12**, (2021).
5. *Forest Europe (2020). State of Europe's forests*. (2020).
6. Tomppo, E., Gschwantner, T., Lawrence, M. & McRoberts, R. E. *National Forest Inventories*. (Springer Netherlands, 2010). doi:10.1007/978-90-481-3233-1.
7. Tomppo, E. *et al. Designing and Conducting a Forest Inventory - case: 9th National Forest Inventory of Finland*. (2011).
8. Fischer, C. & Traub, B. *Swiss National Forest Inventory – Methods and Models of the Fourth Assessment*.
9. Tomppo, E. *et al.* Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment* **112**, 1982–1999 (2008).
10. Gjertsen, A. K. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment* **110**, 420–430 (2007).
11. Nord-Larsen, T. & Schumacher, J. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. *Remote Sensing of Environment* **119**, 148–157 (2012).
12. Zhang, C., Zhou, Y. & Qiu, F. Individual Tree Segmentation from LiDAR Point Clouds for Urban Forest Inventory. *Remote Sensing* **7**, 7892–7913 (2015).

13. Hyyppä, J. *et al.* Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* **29**, (2008).
14. Budei, B. C., St-Onge, B., Hopkinson, C. & Audet, F.-A. Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. *Remote Sensing of Environment* **204**, 632–647 (2018).
15. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science (1979)* **342**, (2013).
16. European Environment Agency (EEA). Copernicus Land Monitoring Service <2018>.
17. Ceccherini, G. *et al.* Abrupt increase in harvested forest area over Europe after 2015. *Nature* **583**, 72–77 (2020).
18. Palahí, M. *et al.* Concerns about reported harvests in European forests. *Nature* **592**, E15–E17 (2021).
19. Santoro, M. *et al.* The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data* **13**, 3927–3950 (2021).
20. Skole, D. L., Mbow, C., Mugabowindekwe, M., Brandt, M. S. & Samek, J. H. Trees outside of forests as natural climate solutions. *Nature Climate Change* **11**, 1013–1016 (2021).
21. Schnell, S., Altrell, D., Ståhl, G. & Kleinn, C. The contribution of trees outside forests to national tree biomass and carbon stocks—a comparative study across three continents. *Environmental Monitoring and Assessment* **187**, 4197 (2015).
22. Brandt, M. *et al.* An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **587**, (2020).
23. Schnell, S., Kleinn, C. & Ståhl, G. Monitoring trees outside forests: a review. *Environmental Monitoring and Assessment* **187**, 600 (2015).
24. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods* **16**, 67–70 (2019).
25. Lugagne, J.-B., Lin, H. & Dunlop, M. J. DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLOS Computational Biology* **16**, e1007673 (2020).
26. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3431–3440 (IEEE, 2015). doi:10.1109/CVPR.2015.7298965.
27. Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1915–1929 (2013).
28. Pinheiro, P. & Collobert, R. Recurrent Convolutional Neural Networks for Scene Labeling. in *Proceedings of the 31st International Conference on Machine Learning* 82–90 (PMLR, 2014).
29. Zhang, C., Li, H., Wang, X. & Yang, X. Cross-Scene Crowd Counting via Deep Convolutional Neural Networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015).
30. Boominathan, L., Kruthiventi, S. S. S. & Babu, R. V. CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. in *Proceedings of the 24th ACM international conference on Multimedia* (ACM, 2016). doi:10.1145/2964284.2967300.
31. Potapov, P. *et al.* Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment* **253**, 112165 (2021).
32. Lang, N., Schindler, K. & Wegner, J. D. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment* **233**, 111347 (2019).

33. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *International Conference on Medical image computing and computer-assisted intervention* 234–241 (2015). doi:10.1007/978-3-319-24574-4_28.
34. Oktay, O. *et al.* Attention U-Net: Learning Where to Look for the Pancreas. (2018).
35. Kobayashi, K. & Salam, M. U. Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. *Agronomy Journal* **92**, 345–352 (2000).
36. Nord-Larsen, T., Johannsen, V. K., Riis-Nielsen, T., Thomsen, I. M. & Jørgensen, B. B. *Skovstatistik 2018: Forest statistics 2018. (2 udg.)*. (2020).
37. Melin, M., Korhonen, L., Kukkonen, M. & Packalen, P. Assessing the performance of aerial image point cloud and spectral metrics in predicting boreal forest canopy cover. *ISPRS Journal of Photogrammetry and Remote Sensing* **129**, 77–85 (2017).
38. Potapov, P. *et al.* Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment* **253**, 112165 (2021).
39. Li, W. *et al.* High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation* **92**, 102163 (2020).
40. Westoby, M. The Self-Thinning Rule. *Advances in ecological research* 167–225 (1984) doi:10.1016/S0065-2504(08)60171-3.
41. Chen, X., Wang, D., Chen, J., Wang, C. & Shen, M. The mixed pixel effect in land surface phenology: A simulation study. *Remote Sensing of Environment* **211**, 338–344 (2018).
42. Helman, D. Land surface phenology: What do we really 'see' from space? *Science of The Total Environment* **618**, 665–673 (2018).
43. Jucker, T. *et al.* Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biology* **23**, 177–190 (2017).

Methods

Tree counting and crown segmentation

We used a multi-task deep neural network with two partially connected branches for the tree counting and crown segmentation tasks. The crown segmentation branch solved a semantic segmentation problem, where each pixel in a given image was classified as either object or background [1]. The counting branch predicted the tree count by regressing density maps for a given image. The ground truth density maps were generated by applying normalized 2-dimensional Gaussian kernels on the manual crown delineations [2,3]. Following the strategy from Zhang *et al.* [2], given an image with a total of C tree delineations, the density map D is defined as in Eqs. (1-2):

$$G_{\sigma,\mathbf{m}}(\mathbf{p}) = \begin{cases} e^{-\frac{||\mathbf{p}-\mathbf{m}||^2}{2\sigma^2}} \cdot \left(\sum_{r,s=-M}^{M} e^{-\frac{r^2+s^2}{2\sigma^2}} \right)^{-1} & \text{if } \mathrm{m}_d - M \leq \mathrm{p}_d \leq \mathrm{m}_d + M \text{ for } d = 1,2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$D(\mathbf{p}) = \sum_{c=1}^{C} G_{\sigma,\mathbf{m}_c}(\mathbf{p}) \quad (2)$$

Here, $G_{\sigma,\mathbf{m}}(\mathbf{p})$ is a sampled truncated Gaussian kernel evaluated at pixel position $\mathbf{p} = (\mathrm{p}_1, \mathrm{p}_2)^{\mathsf{T}}$ in the image. The kernel is centered around $\mathbf{m} = (\mathrm{m}_1, \mathrm{m}_2)^{\mathsf{T}}$with bandwidth parameter σ and is truncated to

height/width $2M + 1$. $D(\mathbf{p})$ denotes the density map for C tree delineations centered at the position $\mathbf{m}_1, \ldots, \mathbf{m}_C$ evaluated at pixel position $\mathbf{p}$. We used a fixed Gaussian filter with a kernel size of 15×15 (i.e., $M = 7$ in Eq. (1)) and a standard deviation (σ) of 4 (Extended Data Fig. 3). Through the normalization term in Eq. (1), we ensured that each Gaussian kernel was normalized to unity. The total tree count could then be estimated by summing up the density values across the whole image.

The network was primarily based on the U-Net architecture [4], a widely used neural network for computer vision applications. Following the approach from Oktay *et al.* [5], we extended the standard U-Net with self-attention blocks to extract more relevant information from the down-sampling path (details of the model can be found in Supplementary Table 1). Batch normalization was applied after each convolutional layer to stabilize and to speed up the training process [6]. The majority of the model weights were shared across the two branches, while only those for producing the final output predictions were task specific. For the segmentation branch, the sigmoid activation was used in the final output layer to produce probabilities in the range of $[0,1]$, which were then converted into binary labels with a threshold of 0.5. For the counting branch, the linear activation was used to maintain the Gaussian kernel values. In each epoch of training, random patches of size 256×256 pixels from all available training images were extracted to generate training and validation data with a batch size of 8. The generated image patches were standardized (per instance and per channel) to zero mean and unit standard deviation before being fed into the network as inputs. We used the Adam optimizer [7] for training.

The network was trained in a fully supervised manner with 19,771 manual tree crown delineations from sampling plots of varying sizes distributed over Denmark, covering diverse landscapes including dense deciduous forests (49% of tree delineations), dense coniferous forests (30% of tree delineations) and open fields with hedgerows or scattered trees (21% of tree delineations). We generated two types of target outputs from the referential annotations for training the two branches: (1) binary masks with tree pixels denoted as ones and background pixels as zeros for the segmentation branch and (2) density maps with single trees represented by Gaussian kernels for the counting branch. Another independent validation set consisting of 2,016 crown delineations (46% in dense deciduous forests, 38% in dense coniferous forests, and 15% in non-forest areas) was used for model selection and hyper-parameter tuning (including the Gaussian parameters and the gap penalty weight as described below).

The network, when randomly initialized, could be retrained from scratch using any composition of input bands, resulting in several final models with slightly different performances (Extended Data Fig. 1 and Extended Data Table 1). The architecture of the network could also be modified slightly to allow for multi-resolution input images (details of the model can be found in Supplementary Table 2). Specifically, the bands with the highest spatial resolution would be fed into the topmost input layer, while the bands with coarser resolutions would be fed into the network after specific down-sampling layers when the spatial resolutions matched. In our experiments, the input data consisted of aerial images (RGB + NIR bands) with a spatial resolution of 20 cm and canopy height maps at 40 cm resolution derived from LiDAR data. The coarser resolution height maps were fed into the network after the first down-sampling layer.

The model was trained by minimizing a combined loss l_{seg_count} from the two branches. The segmentation loss l_{seg} was based on the Tversky index [8], a generalized version of the dice coefficient [9], which penalizes false positives and false negatives differently. To account for the segmentation failures in separating densely connected or overlaid tree crowns, we highlighted the between-crown gaps. Specifically, crown gap maps were generated based on the crown delineations by morphological operations, with the gap pixels being assigned a higher weight than other pixels. The pixel-wise weights were applied in the loss computation so that the misclassified gap pixels were penalized more heavily than others [4,10]. Given a training image I_n, let $p_{0i} \in [0,1]$ denote the predicted probability of pixel $i \in$

I_n being an object. Let $g_{0i} = 1$ if i is an object and $g_{0i} = 0$ otherwise, and let $p_{1i} = 1 - p_{0i}$ and $g_{1i} = 1 - g_{0i}$. Suppose w_i indicates the weight for pixel i, the pixel-wise weighted Tversky loss was then defined as shown in Eq. (3) [11]:

$$l_{seg} = 1 - \frac{\sum_{i \in I_n} w_i p_{0i} g_{0i}}{\sum_{i \in I_n} w_i p_{0i} g_{0i} + \alpha \sum_{i \in I_n} w_i p_{0i} g_{1i} + \beta \sum_{i \in I_n} w_i p_{1i} g_{0i}} \quad (3)$$

For the counting branch, the pixel-wise mean squared error (MSE), as defined in Eq. (4), was employed as the loss function for evaluating the differences between the predicted density map D^{pred} and the ground truth density map D^{gt} :

$$l_{count} = \frac{1}{M} \sum_{i=1}^{M} \left(\mathcal{D}^{pred}(i) - \mathcal{D}^{gt}(i) \right)^2 \quad (4)$$

Here, M denotes the number of pixels in the image. The total loss was a weighted summation of the segmentation loss and the density estimation loss: $l_{seg\_count} = l_{seg} + \lambda_t l_{count}$. The weighting factor λ_t was initially set to 100 and increased steadily during training to ensure that the two losses were re-scaled to a similar magnitude. The final model was determined as the one achieving the lowest error on 800 randomly chosen validation patches.

Individual tree height prediction

Canopy height prediction We formulated the canopy height prediction task as a pixel-wise regression problem. Given a set of multi-band aerial images $\{I_1, I_2, \ldots, I_n\}$ (20 cm resolution) and a set of corresponding LiDAR derived canopy height maps $\{H_1, H_2, \ldots, H_n\}$ (40 cm resolution), a pixel-wise mapping from I to H was established. To balance the differences in different height groups and forest type distributions, we randomly sampled 45 aerial images (4.5 k hectares) from regions dominated by coniferous forest (collected in summer 2018 and 2019) and deciduous forest [12] (collected in summer 2018) with an average LiDAR height over 8 m, respectively. Besides, 3 aerial images covering 300 hectares were randomly sampled from all available non-forest [12] area images taken in summer 2018. The whole dataset was split into a training (74 images) and validation (19 images) set with a ratio of 4:1 using stratified sampling. The training data was used for learning the model parameters and the validation data was used for model selection. The input multi-band aerial images were globally standardized to zero mean and unit standard deviation based on the training dataset before being fed into the network. Data augmentation techniques [13] including random flipping, cropping, Gaussian blurring, and brightness adjustment were applied during training. The network shared a similar U-Net architecture [4,5] as the counting branch of the multi-task network, with the last activation function being a linear transformation (details of the model can be found in Supplementary Table 3). The final decoding block was removed due to the coarser resolution of the LiDAR height maps. A weighted mean absolute error (wMAE) was used as the loss function, where heights over 10 m were given a higher weight w of 5 to penalize particularly the underprediction of the taller trees. Denoting the reference and the predicted tree height at pixel $i \in I$ as y_i and $\hat{y}_i$, the pixel error $l^I(i)$ was formulated as Eq. (5):

$$l^I(i) = \begin{cases} |y_i - \hat{y}_i| & y_i < 10\ \text{m} \\ w * |y_i - \hat{y}_i| & y_i \geq 10\ \text{m} \end{cases} \quad (5)$$

After training, we further adjusted the bias parameter from the last layer to minimize the systematic errors normally caused by non-optimal neural network learning in practice [14]. Specifically, another

validation set was generated by sampling 10 aerial images of size 1 km^2 for each forest/landscape type [12] (deciduous forest, coniferous forest, and non-forest). Given the original bias parameter b, the reference height y, the predicted height $\hat{y}$, and the total number of predictions Z, the optimal bias parameter was calculated by $b^* = \frac{1}{Z}\left(\sum_{i=1}^{Z} y_i - \sum_{i=1}^{Z} \hat{y}_i\right) + b$ [14]. We corrected the bias parameter for each forest/landscape type, respectively (Extended Data Fig. 5).

Individual tree height prediction Individual tree heights were obtained by combining the canopy heights with the individual tree crown segmentation results through the following steps. Firstly, the predicted individual tree crowns were polished by removing tiny segments with an area of less than 2 pixels (0.08 m^2 on site). Secondly, to mitigate the uncertainties induced by slight mismatches between the aerial images and LiDAR data [15] (Supplementary Fig. 8), each predicted tree crown was expanded by a distance ($d = \alpha\sqrt{s/\pi}$), proportional to its area s. We set the expanding factor α to 0.2 in the experiments. Finally, the maximal height within each refined tree crown was determined as the tree height. The individual-tree-level height prediction performance was evaluated by comparing the referential and the predicted tree height derived from the LiDAR height references and the canopy height predictions. Notably, imprecise tree height references (< 0.1%) were removed by filtering out tree crowns with a maximal NIR value higher than 80 yet a maximal height lower than 1 m, as a high NIR value is normally closely associated with trees and is therefore regarded as an anomaly when the height is low.

Transfer learning

The individual tree counting and crown segmentation model was adapted to Finland by fine-tuning the pre-trained weights established on the data from Denmark (using G+B+NIR bands as inputs). The fine-tuning dataset consisted of the original dataset from Denmark (19,771 tree crown delineations in 84 plots) and a small dataset from Finland (4,773 tree crown delineations in 19 plots), which was over-sampled five times to balance the two datasets. The aerial images at 50 cm resolution in Finland were up-sampled to 25 cm using bi-linear interpolation to match with the 20 cm images from Denmark.

The height prediction model was adapted to Finland by fine-tuning with a canopy height dataset from Finland. The Finland dataset consisted of 3 aerial images (10.8 k hectares, G+B+NIR bands) at 50 cm resolution and the corresponding LiDAR height data at 1 m resolution, both collected in 2019. The multi-band aerial images were up-sampled to 25 cm resolution and globally standardized to zero mean and unit standard deviation before being fed into the model. The LiDAR height data was up-sampled to 50 cm resolution accordingly. The wMAE loss (Eq. (5)) was also used for fine-tuning the model.

Production of large-scale country-wide statistics and maps

We applied the established tree counting and crown segmentation model as well as the height prediction model for the whole of Denmark. The final nation-wide products include: (1) Individual tree crown segmentation maps (20 cm resolution), tree count density maps (20 cm resolution), and canopy height maps (40 cm resolution, see examples in Supplementary Fig. 3). (2) An individual tree database featured by tree crown area, tree count evaluated per crown, tree height, and tree type (deciduous, coniferous, or non-forest) determined by Copernicus forest type map (2018) [12]. (3) Tree count and crown area maps re-sampled to spatial resolutions of 10 m and 1 hectare.

To avoid error accumulation for large-scale implementations, we corrected for the underestimation of the crown area of individual trees potentially induced by the gap penalty enforced to separate densely

connected tree crowns. Additionally, we corrected for the underestimation of tree heights, which was especially significant for the tall trees. Both corrections were calculated using the validation data. We corrected the predicted crown area based on a simple linear regression fitted on the predicted values against the reference values, respectively for each forest/landscape type (deciduous, coniferous, and non-forest, see Supplementary Fig. 6). We corrected the predicted tree heights based on a second-degree polynomial regression between the predictions and the references, respectively for each forest type (Supplementary Fig. 7).

Comparison with field data and existing products

We compared our tree count predictions with the Danish NFI forest tree count estimation re-sampled from plot-based field-collected data. The comparison involved merely forest trees (deciduous and coniferous), and the tree counts were grouped into 2 m height intervals (Extended Data Fig. 6).

In addition, we generated tree crown cover (%) maps by aggregating individual crown areas at coarser spatial resolutions (10 m and 100 m). Compared with similar existing forest/tree cover products [12,16], our crown cover maps showed visually lower values (Extended Data Fig. 7a). We numerically compared our refined crown areas with the Copernicus tree cover areas from the same year (2018) [12] for each forest type and observed lower estimates of our crown areas in forest regions yet higher estimates in non-forest regions (Extended Data Fig. 7b). Numerous non-forest trees that were mainly found in hedgerows invisible from low-resolution satellite imagery were clearly visible from high-resolution aerial images. The comparably low estimates of forest cover resulting from our analysis were partially due to the differences in the "tree cover" definitions. We defined the individual tree crown as the part of crown visible from an aerial view, therefore excluding shadows or between-crown gaps. Oppositely, the Copernicus maps used the common "forest/tree cover" definition, which estimated the percentage of tree cover at a coarse resolution (≥10 m) (Extended Data Fig. 7c). We evaluated the bias caused by the definition differences by plotting crown areas obtained from our manual delineations against those from the Copernicus maps. The Copernicus-based crown area was 63% higher than the delineation-based crown area, indicating a systematic bias (Extended Data Fig. 7d).

Evaluation metrics

For evaluating the counting and segmentation performance, we computed the dice coefficient [9] and the regression between the predicted values and the reference values. We also computed the coefficient of determination defined as in Eq. (6):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{6}$$

Here, y, $\hat{y}$, and $\bar{y}$ denote the reference, prediction, and the mean reference value, respectively. Furthermore, we computed the relative error/bias defined by $\delta = \frac{|\sum_i (y_i - \hat{y}_i)|}{|\sum_i y_i|}$ [14].

For evaluating the tree height predictions, we computed the regression lines, the coefficient of determination scores, and the relative errors. We also computed the median/mean absolute errors and the mean squared errors. The mean squared errors were further decomposed into squared bias (the first term) and mean squared variation (the second term) [17], *i.e.*, $l = \left(\bar{y} - \bar{\hat{y}}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}[(y_i - \bar{y}) - \left(\hat{y}_i - \bar{\hat{y}}\right)]^2$.

Uncertainties and limitations

Individual tree counting and crown segmentation could be subject to several uncertainties and limitations, mainly due to the manually delineated tree crown references. Firstly, we excluded small trees, shrubs, and bushes with no visible shadow or with a crown area below 0.08 m^2. Secondly, labeling individual trees was not always obvious, in particular for touching or overlying crowns, heavily shadowed crowns induced by non-vertical shooting angles, and the coexistence of single and multi-branched trees. Thirdly, the individual variation in manual labeling (from two independent data labelers) might also aggravate the model performance [18].

The aerial image-based tree height prediction is particularly challenging since the 2-dimensional spectral features are insufficient to fully reflect the third spatial dimension of heights. We noticed significant height underestimation for the taller trees, which is likely due to the fact that tall trees, despite methodological data balancing, are generally rare and are extremely hard to infer from optical imagery. Secondly, the mismatch between the aerial imagery and the LiDAR height dataset (Supplementary Fig. 8) could cause problems for high-resolution studies. The mismatch occurred since regular orthoimages are generated using a terrain model. True orthoimages could be a potential solution for fixing this issue [15]. Thirdly, a single model generalized on all trees might not be sensitive enough to capture the height differences between different tree species with diverse traits. Lastly, the individual tree height was defined as the maximum height within each predicted tree crown, thereby making the tree height products dependent on the individual tree crown segmentations. If the segmentation model had a systematic bias, then it might be propagated to the error calculation of the height model. Potential future studies could involve evaluations with field-collected data.

References - Methods

1. Wang, P. *et al.* Understanding Convolution for Semantic Segmentation. in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1451–1460 (IEEE, 2018). doi:10.1109/WACV.2018.00163.
2. Zhang, Y., Zhou, D., Chen, S., Gao, S. & Ma, Y. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 589–597 (IEEE, 2016). doi:10.1109/CVPR.2016.70.
3. Zhang, C., Li, H., Wang, X. & Yang, X. Cross-Scene Crowd Counting via Deep Convolutional Neural Networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015).
4. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *International Conference on Medical image computing and computer-assisted intervention* 234–241 (2015). doi:10.1007/978-3-319-24574-4_28.
5. Oktay, O. *et al.* Attention U-Net: Learning Where to Look for the Pancreas. (2018).
6. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. (2015).
7. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014).
8. Tversky, A. Features of similarity. *Psychological Review* **84**, 327–352 (1977).
9. Milletari, F., Navab, N. & Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. in *2016 Fourth International Conference on 3D Vision (3DV)* 565–571 (IEEE, 2016). doi:10.1109/3DV.2016.79.
10. Brandt, M. *et al.* An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **587**, (2020).

11. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. in *International workshop on machine learning in medical imaging* 379–387 (2017). doi:10.1007/978-3-319-67389-9_44.
12. European Environment Agency (EEA). Copernicus Land Monitoring Service <2018>.
13. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* **6**, 60 (2019).
14. Igel, C. & Oehmcke, S. Remember to correct the bias when using deep learning for regression! (2022).
15. Habib, A. F., Kim, E.-M. & Kim, C.-J. New Methodologies for True Orthophoto Generation. *Photogrammetric Engineering & Remote Sensing* **73**, (2007).
16. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science (1979)* **342**, (2013).
17. Kobayashi, K. & Salam, M. U. Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. *Agronomy Journal* **92**, 345–352 (2000).
18. Song, H., Kim, M., Park, D., Shin, Y. & Lee, J.-G. Learning from Noisy Labels with Deep Neural Networks: A Survey. (2020).

Data availability

The aerial imagery from spring and the LiDAR derived canopy height maps are free and publicly available and can be downloaded from https://dataforsyningen.dk/. The user rights of aerial imagery from summer (as used for this study) are generally subject to a fee and the ownership and user rights of these images differ annually. The use of the imagery in research applications can be granted free-of-charge but it is the owners or user rights holders that decide. Contact the Danish Agency for Data Supply and Efficiency for more information. The aerial images and canopy height maps from Finland are publicly available and can be downloaded at https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/. The Copernicus forest cover maps and forest type maps (2018) for Denmark and Finland were downloaded from https://land.copernicus.eu/pan-european/high-resolution-layers/forests.

Code availability

The code will be available upon acceptance of the manuscript.

Author contributions

M.B., S.L., and P.C. designed the study; S.L. implemented the methods, with support from C.I., A.K., F.G., S.O., and A.A.; A.C., T.N., and S.J. provided the airborne data and NFI data; S.L. and M.B. drafted the manuscript and all authors participated in the editing of the manuscript. S.L. and X.T. designed the figures.

Acknowledgements

This study was supported by Institut Europlace de Finance, Kayrros, the DFF Sapere Aude grant (no. 9064–00049B), the Academy of Finland grant (no. 330422), and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no.

947757 TOFDRY). We also acknowledge support by the Villum Foundation through the project 'Deep Learning and Remote Sensing for Unlocking Global Ecosystem Resource Dynamics' (DeReEco), the Pioneer Centre for AI, DNRF grant number P1, and the Independent Research Fund Denmark through the grant 'Monitoring Changes in Big Satellite Data via Massively-Parallel Artificial Intelligence' (9131-00110B).

Ethics declarations

The authors declare no competing interests.

Competing financial interest

The authors declare no competing financial interests.

Materials & Correspondence

Correspondence to S. L. (sizli@ign.ku.dk) or M. B. (mabr@ign.ku.dk).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf