ISPRS Journal of Photogrammetry and Remote Sensing 170 (2020) 205–215

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks

Felix Schiefer [a,*], Teja Kattenborn [a,b], Annett Frick [c], Julian Frey [d,e], Peter Schall [f], Barbara Koch [e], Sebastian Schmidtlein [a]

[a] Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
[b] Remote Sensing Centre for Earth System Research, Leipzig University, 04103 Leipzig, Germany
[c] Luftbild Umwelt Planung GmbH (LUP), Große Weinmeisterstraße 3a, 14469 Potsdam, Germany
[d] Chair of Forest Growth and Dendroecology, University of Freiburg, 79106 Freiburg, Germany
[e] Chair of Remote Sensing and Landscape Information Systems, University of Freiburg, 79106 Freiburg, Germany
[f] Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, 37077 Göttingen, Germany

ARTICLE INFO

Keywords:
Deep learning
Forest inventory
Convolutional neural networks
Tree species classification
Unmanned aerial systems
Temperate forests

ABSTRACT

The use of unmanned aerial vehicles (UAVs) in vegetation remote sensing allows a time-flexible and cost-effective acquisition of very high-resolution imagery. Still, current methods for the mapping of forest tree species do not exploit the respective, rich spatial information. Here, we assessed the potential of convolutional neural networks (CNNs) and very high-resolution RGB imagery from UAVs for the mapping of tree species in temperate forests. We used multicopter UAVs to obtain very high-resolution (<2 cm) RGB imagery over 51 ha of temperate forests in the Southern Black Forest region, and the Hainich National Park in Germany. To fully harness the end-to-end learning capabilities of CNNs, we used a semantic segmentation approach (U-net) that concurrently segments and classifies tree species from imagery. With a diverse dataset in terms of study areas, site conditions, illumination properties, and phenology, we accurately mapped nine tree species, three genus-level classes, deadwood, and forest floor (mean F1-score 0.73). A larger tile size during CNN training negatively affected the model accuracies for underrepresented classes. Additional height information from normalized digital surface models slightly increased the model accuracy but increased computational complexity and data requirements. A coarser spatial resolution substantially reduced the model accuracy (mean F1-score of 0.26 at 32 cm resolution). Our results highlight the key role that UAVs can play in the mapping of forest tree species, given that air- and spaceborne remote sensing currently does not provide comparable spatial resolutions. The end-to-end learning capability of CNNs makes extensive preprocessing partly obsolete. The use of large and diverse datasets facilitate a high degree of generalization of the CNN, thus fostering transferability. The synergy of high-resolution UAV imagery and CNN provide a fast and flexible yet accurate means of mapping forest tree species.

1. Introduction

Forest ecosystems cover about one-third of the Earth's land area (FAO, 2020) providing countless and substantial ecosystem services. There is, therefore, great interest in obtaining information on the state of forest ecosystems. Many problems in this context require the acquisition of tree species composition at a high spatial resolution—a goal to which remote sensing can ultimately contribute significantly (Fassnacht et al., 2016). A combination of two technological and methodological advances offers great potential for accurately mapping forest tree species: the use of unmanned aerial vehicles (UAVs) and deep learning. Whereas the use of very high-resolution UAV-data is no novelty in this regard (Franklin and Ahmed, 2018; Gini et al., 2014; Michez et al., 2016; Nevalainen et al., 2017), deep learning is only recently being introduced into vegetation remote sensing (Audebert et al., 2019; Brodrick et al., 2019; Ma et al., 2019; Zhang et al., 2016; Zhu et al., 2017). The most effective deep learning algorithms in analyzing high spatial resolution remote sensing data are convolutional neural networks (CNNs) since these are specifically designed to analyze spatial patterns. CNNs autonomously extract low-, mid-, and high-level feature representations (e.g., corners, edges, abstract shapes) that best describe targets, such as classes or continuous values, through a series of convolutions and

* Corresponding author.
E-mail address: felix.schiefer@kit.edu (F. Schiefer).

https://doi.org/10.1016/j.isprsjprs.2020.10.015
Received 16 July 2020; Received in revised form 16 October 2020; Accepted 22 October 2020

pooling operations.

Several studies have already used CNNs and very high-resolution remote sensing data for the mapping of tree species. To detect tree individuals outside forests, good results have been reported from urban environments (dos Santos et al., 2019; Hartling et al., 2019; Lobo Torres et al., 2020) and plantations (Csillik et al., 2018; Freudenberg et al., 2019; Li et al., 2017; Osco et al., 2020) but these results are hardly transferable to heterogeneous forest. Specifically targeting forest environments, Fricker et al. (2019) used a CNN for classifying and mapping seven tree species in a mixed-conifer forest from airborne data with very accurate results for hyperspectral and moderately accurate results for pseudo-RGB data. Trier et al. (2018) also used airborne hyperspectral data to classify pine, spruce, and birch trees in a boreal forest using a CNN. Nezami et al. (2020) showed very accurate results for classifying the same tree species testing CNNs with different combinations of hyperspectral and RGB imagery and canopy height models. Thus far, mapping tree species in forests often requires high spectral resolution data, which is cumbersome to access for non-specialist users.

Solely relying on RGB information, individual tree species have been accurately mapped against a background of other species using CNNs (Kattenborn et al., 2020, 2019a; López-Jiménez et al., 2019; Morales et al., 2018; Wagner et al., 2020). Natesan et al. (2019) used a CNN to classify previously extracted tree crowns from RGB data into white pine, red pine, and non-pine. Spectral resolution notwithstanding, many studies used additional preprocessing steps prior to classification (e.g., tree segmentation or tree localization from ancillary remote sensing data, background removal, feature engineering), which limits the transferability and increases the computational load of such applications.

With consumer-grade UAVs on the rise, which enable easy and low-cost acquisition of very high-resolution RGB data, the mapping of tree species in heterogeneous forests using solely RGB imagery is of high interest, as it does not rely on sophisticated sensors, does not require extensive calibration and preprocessing and, therefore, enables the application by a wide audience (Komárek, 2020). The above-mentioned studies demonstrated that, regardless of the spectral resolution, high spatial resolution remote sensing data can be sufficient for mapping tree species when small samples of species or relatively homogeneous environments with little site variability are considered. To further assess the potential of very high-resolution imagery for mapping forest tree species it would be desirable to test CNNs on a large and heterogeneous sample of species with a wide gradient of forest types, site conditions, and stand structures. Moreover, such an assessment based on RGB imagery alone would be valuable since the use of RGB data ensures access to such applications for a wide audience. Recent CNN architectures for semantic segmentation (e.g., U-Net (Ronneberger et al., 2015) or DenseNet (Jegou et al., 2017)) facilitate end-to-end learning that can be directly applied on the raw remote sensing data and enable mapping at the original image resolution and overcome the need for prior segmentation and feature engineering steps.

Here, we would like to assess the potential of very high-resolution RGB imagery from UAVs to map forest tree species with a large and heterogeneous sample on mixed stands of forest trees. We used CNNs to map tree species from UAV-based very high-resolution RGB imagery in temperate deciduous and mixed-coniferous forests in Germany. We used a multiclass semantic segmentation approach (U-net) to simultaneously segment and classify 14 classes (i.e., nine tree species, three genus-level classes, deadwood, and forest floor). Our main research question is as follows: Is RGB imagery sufficient to accurately map tree species in heterogeneous forests? Moreover, given the very recent introduction of CNNs into vegetation remote sensing, little is known about the requirements regarding the remote sensing data. We, therefore, tested several spatial resolutions, the additional value of photogrammetric 3D-information, and different tile sizes of the input images.

2. Material and methods

2.1. Study area

The study area is in the Southern Black Forest region and the Hainich National Park (NP), in the German states of Baden-Württemberg and Thuringia, respectively (Fig. 1). The Southern Black Forest study site is situated in a mountain range between 120 and 1492 m a.s.l. between the Rhine valley and the highest peak at Feldberg. The area is mostly covered by mixed and coniferous forests, largely managed for timber production (Kändler and Cullmann, 2015) and covers a wide range of forest types and age classes (Frey et al., 2018). The main tree species are *Picea abies* L. (40% cover), *Fagus sylvatica* L. (18%), and *Abies alba* Mill. (13%). Less common tree species are *Quercus robur* L. (5%), *Pinus sylvestris* L. (4%), and *Pseudotsuga menziesii* Mirbel (4%). Parent rock mainly consists of granite and gneiss with some admixture of sandstone (Storch et al., 2020).

The Hainich NP lies on a ridge between 225 and 494 m a.s.l. and covers an area of 7600 ha. It is characterized by unmanaged mixed deciduous forests on limestone and dominated by *F. sylvatica*. Subordinate species include *Fraxinus excelsior* L., *Acer pseudoplatanus* L., *Acer platanoides* L., *Q. robur, Quercus petraea* (Matt.) Liebl., *Tilia cordata* Mill., *Tilia platyphyllos* Scop., *Carpinus betulus* L., and others. The heterogeneity of both study areas is exemplified by the forest inventory plots (details see Section 2.2), with species numbers per plot between two and ten and tree densities ranging from 179 and 851 trees per hectare.

2.2. Data acquisition

The ConFoBi-Project (Conservation of Forest Biodiversity in Multiple-Use Landscapes of Central Europe) has implemented 135 research plots (100 × 100 m) within state-owned forests in the Southern Black Forest region (Storch et al., 2020). A full forest inventory was conducted between October 2016 and February 2018. In each plot we recorded tree species, diameter at breast height (DBH), and height of all trees with a DBH ≥ 7 cm. In addition, each plot was inventoried with an octocopter UAV (OktoXL 6S12, Mikrokopter GmbH, Moormerland, Germany) carrying a consumer-grade full-frame RGB camera (Alpha 7R, Sony Europe Limited, Weybridge, Surrey, UK) with a 35 mm prime lens. Flights were carried out in snowless conditions between March 2017 and April 2018. For each flight, the UAV maintained an altitude of 80 m above ground at a flight speed of 3.5 m/s and followed a crisscross pattern using the onboard GNSS (see Frey et al., 2018 for details). The camera was aligned nadir and perpendicular to the flight direction, and triggered automatically every 3–4 m of the flight track. This resulted in forward overlaps of > 95% and ground sampling distances of about 1.1 cm. Because we adopted an area-wide digitization of the reference data to gain a full picture of the model performance across sites, the digitization of all plots would have been too labor-intensive and we randomly selected 47 plots. From all 135 plots, plots with leaf-off conditions, plantation-like forest structures as overly easy targets, or cloud shadows in parts of a scene were excluded.

Within the Biodiversity Exploratories framework (Fischer et al., 2010), 13 research plots (100 × 100 m) were implemented in the Hainich NP. In the off-season from 2014 to 2015, all plots were surveyed and trees with a DBH ≥ 7 cm were recorded with species information, DBH, tree height, and geographic location of the stem (Schall et al., 2018). For four of these plots, UAV-based RGB imagery was acquired in September 2019 with a DJI Phantom 4 Pro+ (DJI Technology Co., Ltd., Shenzhen, China) quadcopter with a ground sampling distance of < 1.35 cm, at a flight speed of 2.8 m/s, and forward overlap of 90%.

We derived a total of 51 orthomosaics using a Structure from Motion-based photogrammetric processing chain in Agisoft Metashape v.1.5.4 (Agisoft LLC, St. Petersburg, Russia). This included filtering of blurry images, image matching, and dense point-cloud creation. Digital elevation models were derived from the dense point cloud.

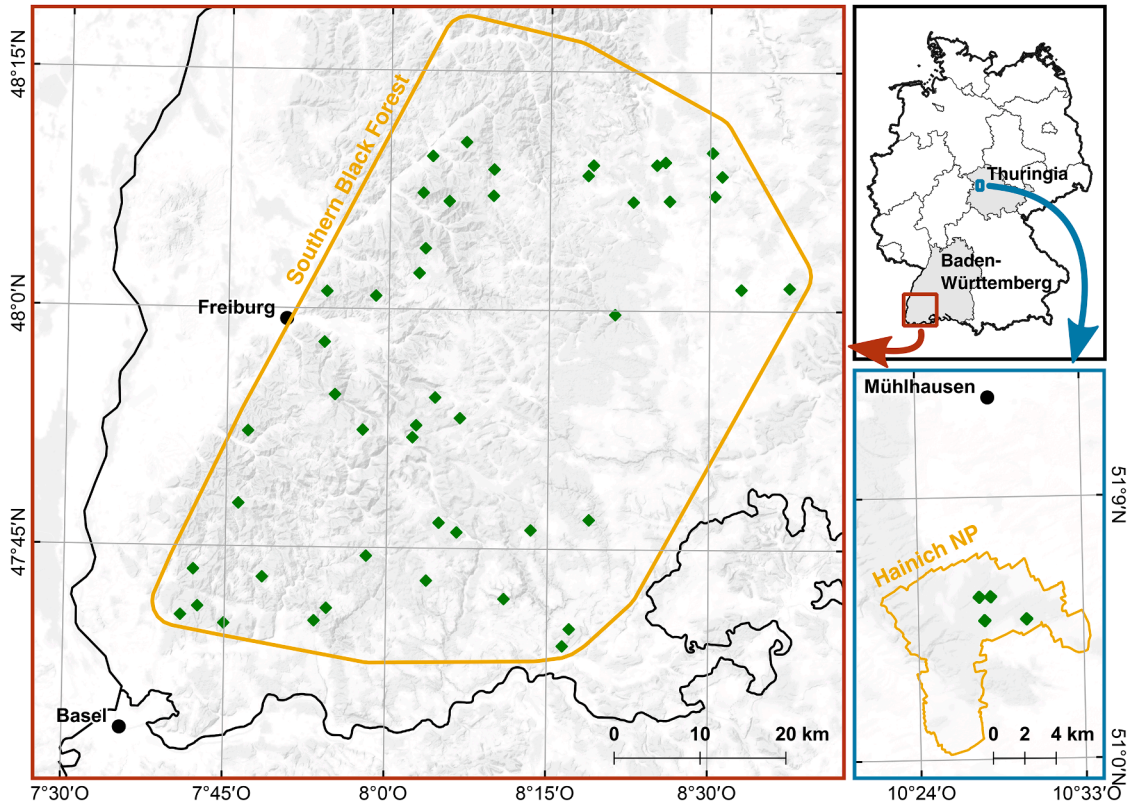

Fig. 1. Map oft the two study areas Southern Black Forest and Hainich NP in Germany. Green markers indicate the locations of the research plots. Projection: WGS84 UTM Zone 32 N. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

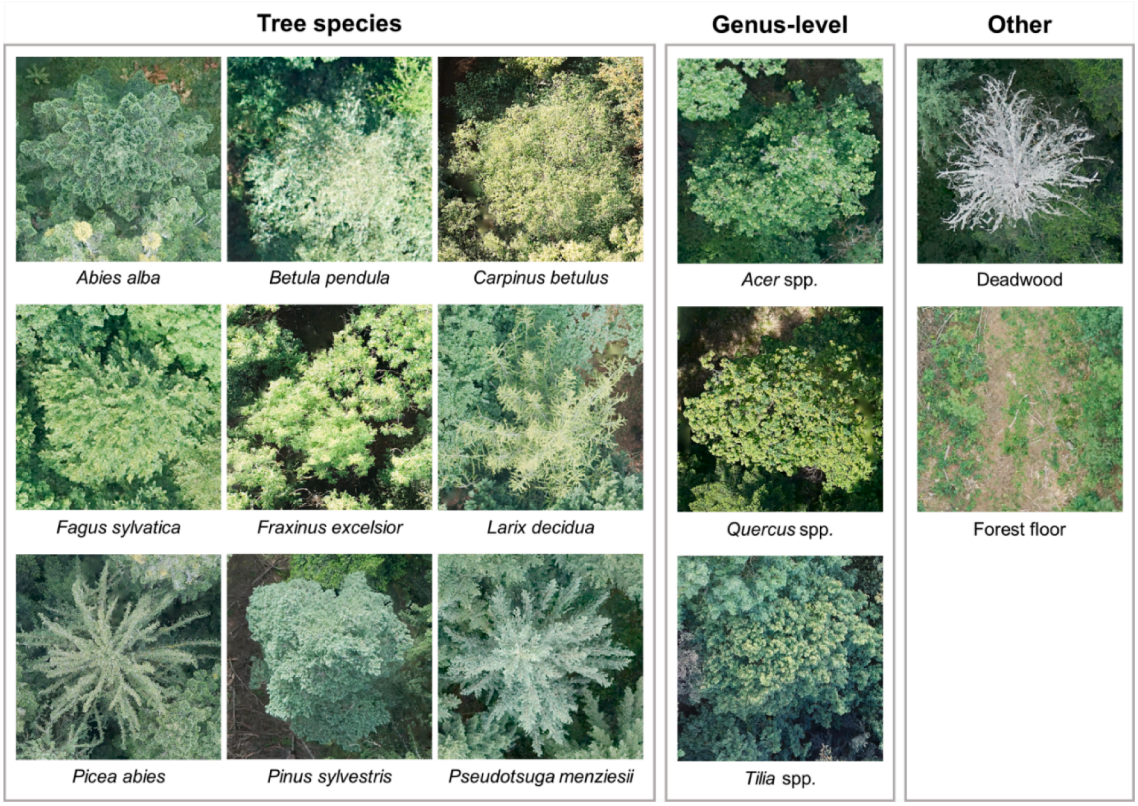

Fig. 2. Detailed overview of the occurring tree species and classes.

Orthomosaics were created by projecting single images on digital elevation models. Georeferencing was performed automatically based on the GNSS trajectory logs of the respective UAV.

We calculated normalized digital surface models (nDSM) via subtraction of digital terrain models. The digital terrain models were derived from airborne laserscans with 1 m resolution and were provided by the states Baden-Württemberg (State Agency for Spatial Information and Rural Development Baden-Wuerttemberg, LGL, Stuttgart, www.lgl-bw.de) and Thuringia (State Agency for Land Management and Geoinformation, TLBG, Erfurt, www.geoportal-th.de). Orthomosaics were resampled to a spatial resolution of 2 cm. To compensate for differences in the illumination properties of the individual UAV scenes, we applied a histogram stretch to the 0.01 and 99.99% percentiles to all orthomosaics.

2.3. Reference data extraction

Training of the U-net segmentation algorithm requires regular tiles of the RGB imagery. Besides, classified areas in the form of masks need to be provided for training. We derived these masks by visual interpretation and manual delineation of classes in the orthomosaics and nDSM using ArcGIS v.10.6.1 (ESRI, Redlands, CA, USA). A total of nine tree species, three genus-level classes, deadwood, and forest floor were classified in this study (Fig. 2). Tree species composition, tree height, DBH, and relative position of trees from the forest inventory data aided the visual classification. For each plot, we digitized the classes. We did not explicitly delineate tree individuals because this was beyond the scope of the study. Delineation and the class assignment were cross-checked by at least one other interpreter. The visual, area-wide classification is not a necessity of the CNN approach but it was, as already mentioned above, a requirement for gaining a comprehensive picture of the model performance across sites and with different tile sizes. Parts of the canopy that could not be assigned to classes with certainty (0.07% of the area, i.e., due to blurry image areas) were excluded from further analysis. The area-related share of a species in the dataset and the number of sites in which the species occurred is shown in Table 1 (two right columns).

We tested squared tiles with three different edge sizes: 128 pixel, 256 pixel, and 512 pixel corresponding to 2.56 m, 5.12 m, and 10.24 m, respectively. We seamlessly cropped orthomosaics and class delineations into non-overlapping tiles, resulting in a maximum of 36^2, 18^2, and 9^2 tiles for the respective tile sizes per scene. Tiles containing empty raster cells (artifacts from the SfM-workflow caused by too little image overlap) in the orthomosaics or unidentified species in the mask were excluded from further analysis. In total, we extracted 62826, 15094, and 3112 tiles for the respective tile sizes.

2.4. Data splitting

Training of a CNN is performed in epochs, which are defined as one complete pass through a training dataset. To assess whether a CNN is starting to over-optimize on training data, the CNN is evaluated using a validation dataset after each epoch. To get an independent assessment of the model accuracy, a model has to be evaluated with independent test data. Prior to model training, we randomly sampled 10% of the dataset (based on the 512-pixel tiles) as independent test data. Additionally, for visual inspection of the results, the UAV-scene of an entire 100 × 100 m plot was set aside. The area covered by the 512-pixel test tiles was also used for the test datasets of the smaller tiles, with an accordingly higher resulting number of tiles. With the same procedure as for the test dataset, we randomly split the remaining dataset into 75% for model training and 25% for model validation.

2.5. CNN-based tree species mapping

For tree species mapping, we adapted the U-net CNN-architecture (Ronneberger et al., 2015, Fig. 3). The U-net consists of a contracting path (Fig. 3, left side) to capture context and a symmetric expanding path (Fig. 3, right side) to map the contextual information to the original image resolution. In our implementation, the contracting path featured four blocks. Each block consisted of two 3 × 3 convolutions, both followed by batch normalization and rectifier linear unit (ReLU) activation. A 2 × 2 max pooling operation with a striding of two concluded each block, reducing the spatial dimensions of the feature maps by half. After each max pooling operation, we doubled the number of feature maps. Each block of the expanding path consisted of up-sampling of the feature maps and subsequent 2 × 2 convolution ("up-convolution"), reducing the number of feature maps by half. The resulting feature maps were concatenated with the feature maps of the corresponding blocks from the contracting path. This was followed by repeated 3 × 3 convolutions, batch normalization, and ReLU activation. With each block of the expanding path, we halved the number of feature maps and doubled the

Table 1
Tree species mapping model accuracies. Classes are sorted descending by their area-related share. For each class, the highest class-specific F1-score is highlighted in bold.

	Tile size [pixel]						Spatial resolution [cm]					Area-related share[a]	Occurences[b]
Input data	RGB			RGB + nDSM			RGB + nDSM						
Tile size/resolution	128	256	512	128	256	512	2	4	8	16	32		
F1-Score													
Picea abies	0.89	0.93	0.91	0.93	**0.93**	0.93	0.93	0.91	0.86	0.81	0.70	32.97	45
Fagus sylvatica	0.89	0.90	0.87	**0.90**	0.90	0.86	0.90	0.86	0.79	0.75	0.66	29.80	46
Abies alba	0.79	0.85	0.86	0.86	**0.87**	0.86	0.87	0.83	0.60	0.60	0.34	10.91	37
Pseudotsuga menziesii	0.84	0.89	0.74	0.89	**0.91**	0.88	0.91	0.86	0.79	0.77	0.36	3.89	12
Pinus sylvestris	0.89	0.90	0.89	**0.91**	0.91	0.87	0.91	0.81	0.78	0.60	0.24	3.59	19
Acer spp.	0.70	0.72	0.53	**0.80**	0.73	0.40	0.73	0.60	0.40	0.37	0.12	2.33	23
Fraxinus excelsior	0.75	0.79	0.16	**0.87**	0.82	0.52	0.82	0.59	0.28	0.15	–	1.01	14
Larix decidua	0.80	0.82	0.80	0.83	**0.89**	0.82	0.89	0.65	0.21	0.17	–	0.98	19
Quercus spp.	**0.64**	0.49	0.28	0.58	0.39	0.02	0.39	0.38	0.00	–	–	0.88	10
Carpinus betulus	**0.45**	0.33	–	0.38	0.36	0.00	0.36	0.24	0.08	0.06	–	0.39	4
Tilia spp.	0.26	0.20	–	**0.50**	0.02	–	0.02	0.01	–	–	–	0.24	4
Betula pendula	0.07	**0.33**	–	0.27	–	–	–	–	–	–	–	0.20	8
Forest floor	0.78	0.83	0.82	0.83	**0.84**	0.84	0.84	0.82	0.80	0.77	0.72	11.79	50
Deadwood	0.71	0.73	0.68	**0.72**	0.75	0.69	0.75	0.70	0.53	0.57	0.44	0.95	44
Mean F1-Score	0.68	0.69	0.54	**0.73**	0.67	0.55	0.67	0.59	0.44	0.40	0.26		
Overall Accuracy	0.86	0.88	0.86	**0.89**	0.89	0.87	0.89	0.85	0.78	0.73	0.62		

[a] Area-related share of the class in the dataset [%].
[b] Occurrence of class in number of sites.

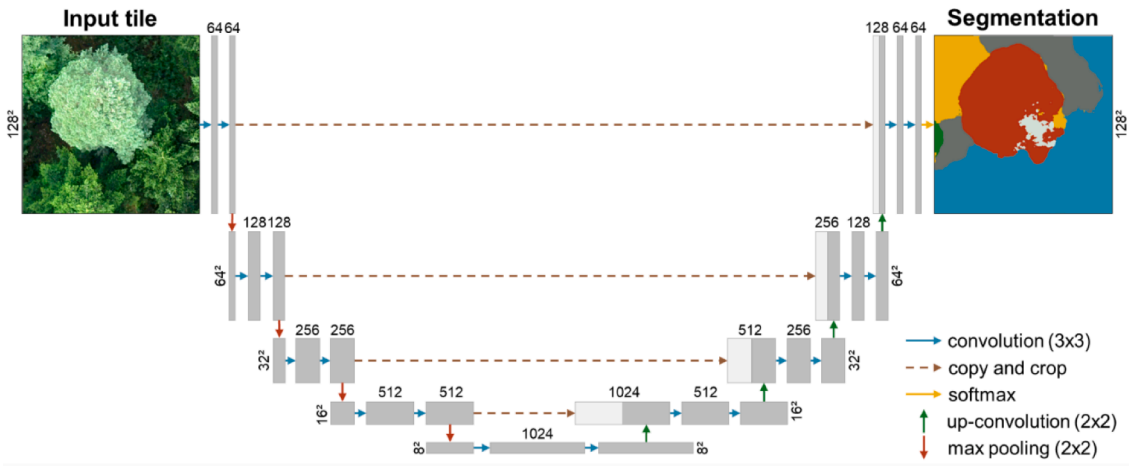

Fig. 3. Adapted U-net CNN-architecture for the tree species segmentation (Ronneberger et al., 2015). This scheme illustrates how 128 × 128 pixel tiles were analyzed. Values on top of the boxes depict the number of calculated feature maps with the respective x-y-dimensions as vertically oriented labels.

spatial dimensions. The pixel-wise classification was performed at a subsequent 1 × 1 convolutional layer with a softmax activation. This softmax activation mapped the learned features to the final class probabilities. The maximum class probability of a pixel represented the final class of the respective pixel.

Due to the imbalanced distribution of the tree species (see Table 1), we used weighted categorical cross entropy as loss function during model training. Thereby, the categorical cross entropy between masks and model output was weighted by the area-related share of a species; in this case inversely proportional. As optimizer, we chose RMSprop with a learning rate of 1e-4. For better model generalization, we performed a random data augmentation during model training. This augmentation included inflating the training dataset to four times its size, applying random horizontal and vertical flips, and randomly changing brightness (90–110%) and contrast (80–120%) values of input tiles. Models were trained for 40 epochs with batch sizes of 3, 12, and 46 for 128 × 128, 256 × 256, and 512 × 512 pixel tiles, respectively. The epoch with the lowest loss value from the validation dataset was kept as the final model.

All code was written in R v.3.6.3 (R Core Team, 2020), using the packages 'tensorflow' (Allaire and Tang, 2019), 'keras' (Allaire and Chollet, 2019), 'tfdatasets' (Allaire et al., 2019), and 'tibble' (Müller and Wickham, 2019), and is available at https://github.com/FelixSchiefer/TreeSeg. We used the R interface to Keras (Chollet and Allaire, 2017) with the TensorFlow backend v.2.0.0 (Abadi et al., 2016). Training of a CNN model on a CUDA-compatible NVIDIA GPU (GeForce RTX 2080 Ti, 11 GB RAM) and the cuDNN library (Chetlur et al., 2014) took between 7 and 14 h. Upon request, the data used in this study can also be made available.

2.6. Accuracy assessment

To analyze the effects of the tile size, height information, and spatial resolution on CNN accuracy, we compared the results of several models. Three CNNs were trained with RGB data; each with a different tile size. Another three CNNs were trained with RGB + nDSM data; each with a different tile size. To analyze the influence of spatial resolution, we trained four CNNs with RGB + nDSM data and a fixed tile size of 256 × 256 pixel; each with a different spatial resolution (4, 8, 16, and 32 cm).

We compared manually delineated tree crowns from the test dataset with CNN predictions to evaluate CNN models based on overall accuracy (OA), precision, recall, and F1-score (harmonic mean of precision and recall). The reported accuracies are based on the pixel-level. For visual inspection, we applied the best model to an entire UAV-scene that was not used during model training. We used a moving window approach with a half tile size overlap in x- and y-direction. From the resulting nine predictions per pixel, final predictions were derived through majority vote.

3. Results

3.1. Model training

For each model, the validation loss reached a minimum during the 40 epochs (Fig. 4). After reaching its minimum, the training loss for all models converged towards zero (not depicted) whereas the validation loss stagnated or increased again. Models that were trained with smaller tiles, displayed a faster decrease in validation loss.

3.2. Model results

The model that performed best was trained with RGB + nDSM data and a tile size of 128 × 128 pixel (OA = 89%, mean F1-Score = 73%), albeit only marginally better than models trained only with RGB data or with a larger tile size (Table 1). A coarser spatial resolution resulted in overall accuracy reduction from 89% at 2 cm to 62% at 32 cm resolution, and mean F1-scores from 67% to 26%. Class-specific F1-scores were highest for *P. abies* (93%). Moreover, these scores did not differ much between models with different tile sizes, especially not for abundant species (i.e., *P. abies, F. sylvatica, A. alba, P. menziesii,* and *P. sylvestris*). For underrepresented species (i.e., *Acer* spp., *F. excelsior, L. decidua, Quercus* spp., *C. betulus, Tilia* spp., and *B. pendula*), however, larger tile sizes resulted in lower F1-scores, with rare classes no longer being classified. The use of weighted categorical cross entropy did not compensate for the imbalanced dataset. Setting the weights higher even worsened the results (see Appendix A for details). The same applied for models with a decreasing spatial resolution; 13 out of 14 classes were recognized at a spatial resolution of 4 cm, but only 8 classes at 32 cm resolution. Such decrease in model accuracy was even more evident for classes with a lower share. For example, *Larix decidua* had a high F1-score at 2 cm spatial resolution (F1 = 89%), but was not classified at 32 cm spatial resolution. This variation was species dependent. For example, for *P. abies* the F1-score decreased far less, from 93% at a spatial resolution of 4 cm to 70% at 32 cm resolution. Site-specific F1-scores did not show large fluctuations over the research plots from different study areas and years (see Appendix B for details).

3.3. Prediction on an independent scene

We applied the best model (i.e. CNN trained with RGB + nDSM on 128 × 128 pixel tiles) to a UAV-scene that had not been used for training

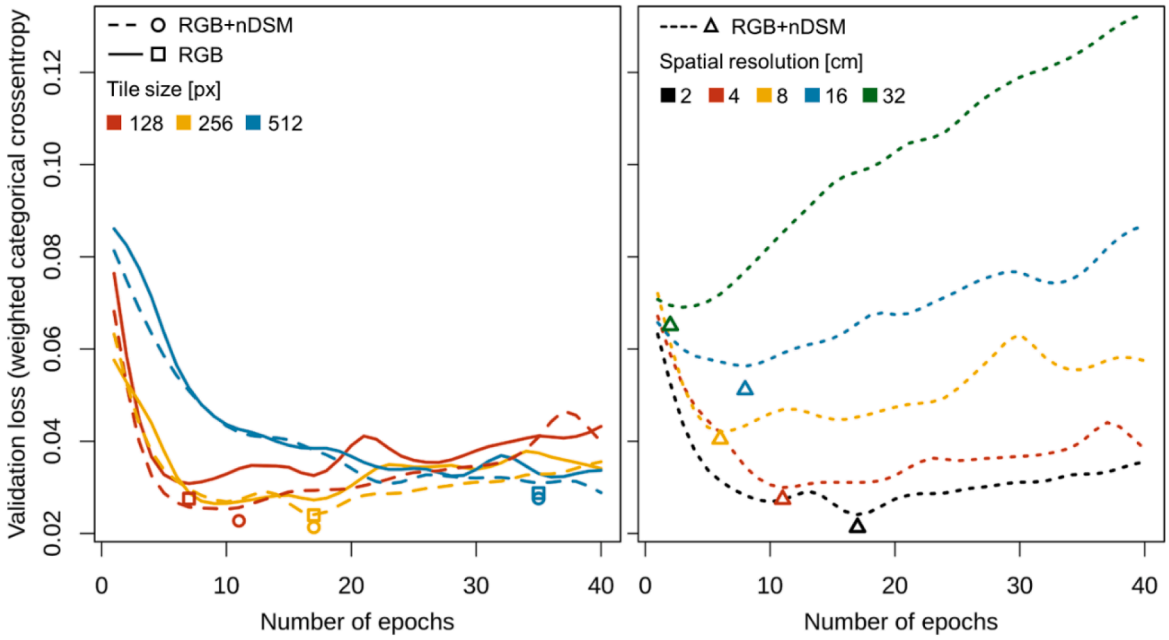

Fig. 4. Validation loss during CNN model training. Curves were smoothed for better visualization. The symbols represent the unsmoothed validation loss of the best epoch.

(Fig. 5). Model inference took about 3 min for the entire 100 × 100 m UAV-scene. Abundant classes were almost perfectly predicted, but the model struggled with underrepresented classes. The CNN predictions on larger tiles resulted in similar patterns, but edge effects of the tiles were less pronounced (not shown).

4. Discussion

4.1. Model performance

The model accuracies achieved in our study were relatively high, especially when considering the high number of 14 classes (i.e., nine tree species, three genus-level classes, deadwood, and forest floor) and the fact that we only used RGB imagery. Moreover, our data are characterized by a high degree of heterogeneity, as they include different forest types (i.e. mixed, deciduous and coniferous), different types of use (i.e. unmanaged forests in Hainich NP and commercial forest in the Southern Black Forest), and feature a diverse age structure. By using a semantic segmentation approach, no tree segmentation or localization steps prior to model inference were required, allowing us to fully exploit the end-to-end learning capabilities of CNNs.

The classification of comparably high numbers of tree species using CNNs has been demonstrated in subtropical forests (OA = 84%), but hyperspectral UAV data was used and the targeted tree crowns were previously extracted from the imagery (Sothe et al., 2020). Similar accuracies have been reported for the classification of seven tree species in mixed coniferous forest using airborne hyperspectral data (F1 = 87%) and pseudo-RGB data (F1 = 64%), after previous identification of the trees in LiDAR-derived canopy height models (Fricker et al., 2019). Atfer the removal of shadowed, low-, and non-vegetated pixels prior to CNN-classification, *P. abies, P. sylvestris*, and *B. pendula* have been mapped in boreal forests in airborne hyperspectral data (OA = 87%) and RGB data (OA = 74%) (Trier et al., 2018). The same species have been mapped with different combinations of hyperspectral data, RGB imagery, and canopy height models with highest accuracies (OA = 98%) (Nezami et al., 2020). CNNs have been successfully used to classify two *Pinus* species and non-*Pinus* in previously extracted tree crowns from UAV-based RGB imagery (F1 = 80%) (Natesan et al., 2019). However, a more detailed comparison of our results with the existing literature is hampered by the variety of applied CNN approaches (i.e., object detection, image classification/regression, and semantic segmentation), CNN architectures, forest types, and most of all tree species studied.

4.2. Tile size

For most of the classes, the tile size did not have a prominent effect on model performance. Only for underrepresented classes a larger tile size was disadvantageous. This depended less on the tile size itself but rather on the species coverage within the tiles. With smaller tiles, the area percentage of rare species on the tile was larger and underrepresented species thus contributed more to the model update during training. Whereas with larger tiles, underrepresented species got lost in the surrounding information of more frequent species. This was despite the use of weighted categorical cross entropy as loss function to compensate for such imbalances. The situation was different for classes that feature distinct characteristics (i.e. L. *decidua* and deadwood) as they were modeled equally well, regardless of the tile size. From the prediction map (Fig. 5c) it becomes evident that the CNN, despite the employed moving window approach, sufferd from edge effects, a known problem with CNNs. Obviously, this effect is more problematic with a smaller tile size. Hence, if sufficiently enough reference data for the targeted classes is available a larger tile size should be preferred. This allows a larger spatial context to be considered—which is key information to CNNs—and speeds up model inference over large spatial extents.

The fact that the models with smaller tile sizes reached their minimal validation loss earlier can be explained by the different batch sizes. The batch size is limited by the computational complexity of the CNN-architecture, the available RAM, and the size of the images. To analyze the influence of the batch size on the model performance is beyond the scope of this study. With the different batch sizes for the CNNs of the different tile sizes, we ensured that the models were exposed to the same amount of information in terms of area coverage.

4.3. Canopy height information

Adding height information from nDSM to the CNN slightly increased the model accuracies for most of the classes. This contrasts with Sothe

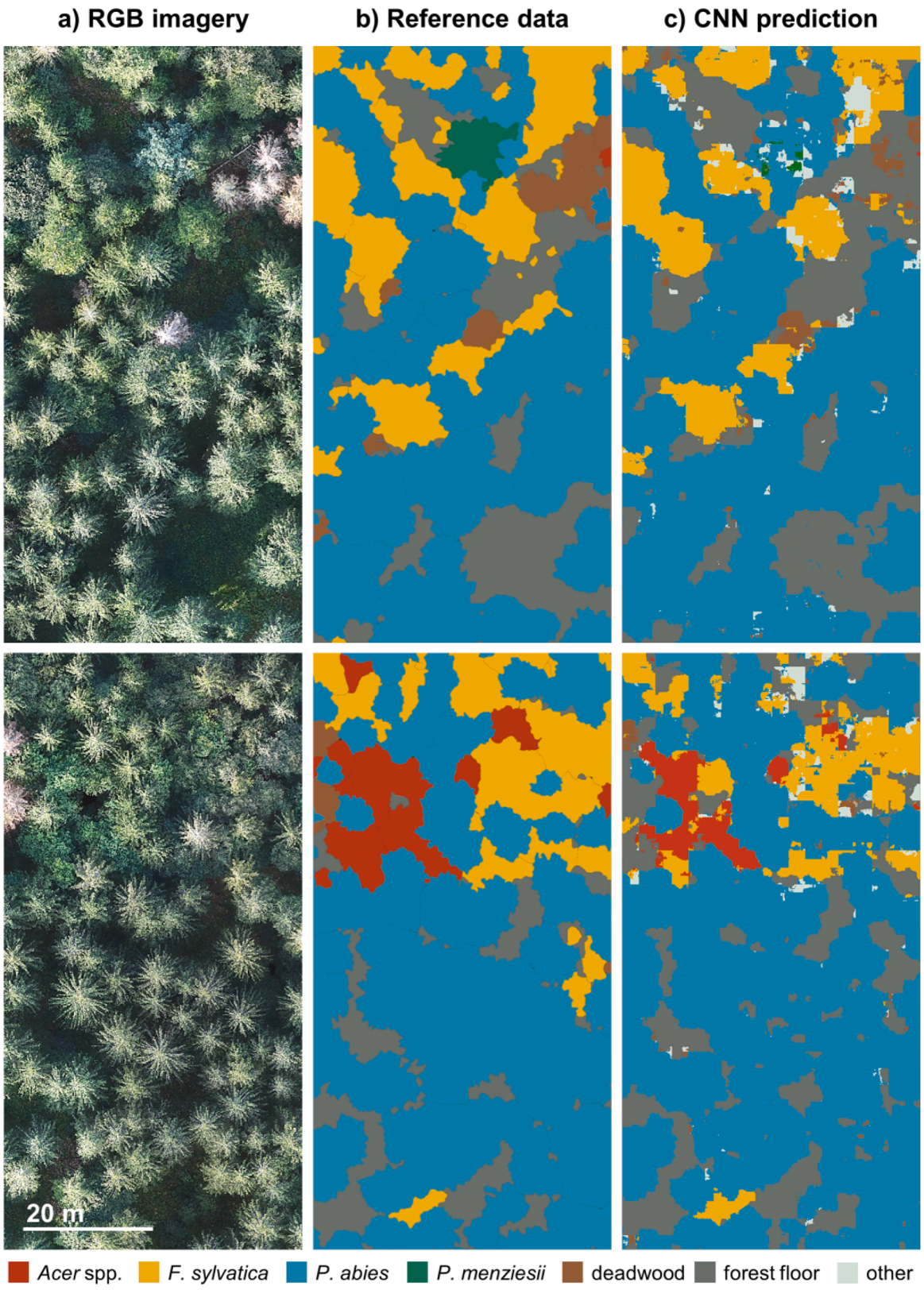

Fig. 5. Predictions of a trained CNN on a 100 × 100 m plot. (a) UAV-based RGB orthomosaic, (b) manually delineated reference data, (c) CNN prediction based on 128 × 128 pixel tiles (RGB + nDSM). For illustrative purposes, the two sides of the plot are shown one above the other. Classes that did not appear in the reference data are grouped in the category "other".

et al. (2020) and Hartling et al. (2019) who found additional height information to decrease the model performance. Kattenborn et al. (2020) found no clear positive effect of combining height information with RGB data and suggested that the structural aspect is redundant in both height and RGB information. Analogous to our visual perception of the tree crowns, we assume that the basic structural information of nDSMs is already inherently included in RGB data through shadows and illumination differences. Whereas the creation of a digital surface model from UAV data is required for the calculation of the orthomosaic anyway, one should keep in mind that for the calculation of nDSMs a digital terrain model is needed (Wallace et al., 2019), which in turn requires additional processing steps. Furthermore, including additional layers to the CNN increases the number of parameters and thus computational complexity and could outweigh the benefit introduced.

4.4. Spatial resolution

Our results showed that very-high spatial resolution was essential for accurate mapping of forest tree species using RGB data. These findings underline the key role that UAVs can play for the remote sensing-based forest assessment, given that airborne and satellite remote sensing data currently do not provide a comparable spatial resolution. While most species with small shares of the dataset could not be identified with coarsening spatial resolution deadwood could still be sufficiently identified, despite its small share of the dataset (0.95%). This is probably because the visual characteristics of deadwood were still represented at coarser spatial resolutions. This indicates that for some classes mapping might be possible even at coarser spatial resolutions if prominent features exist. Accordingly, Safonova et al. (2019) used CNNs on UAV-based RGB imagery with 5–10 cm spatial resolution to detected damaged and dead trees of *Abies sibirica* after bark beetle infections with F1-scores up to 93%.

For a qualitative inspection of the effect of the spatial resolution and to obtain a causal explanation for our results we inspected the learned features of the CNN based on filter visualizations (Fig. 6). The latter are synthetic images that would maximally activate the respective filter of a trained network—in other words, they reflect what the network is looking for (technical details on the filter visualization are given in Appendix C). The filter visualizations of the fourth block and the center block of the CNN revealed fine-scale patterns that resemble typical canopy features, e.g., conifer-like branching structures (Fig. 6a,c), or broad-leaf-like canopy structures (Fig. 6b,d). Such patterns could not be revealed with coarser spatial resolutions, which underlines our findings that a very high resolution is key to identify forest tree species. It, therefore, seems possible that further increasing the spatial resolution (e.g. sub-centimeter) may even improve the capabilities for a CNN-based tree species mapping.

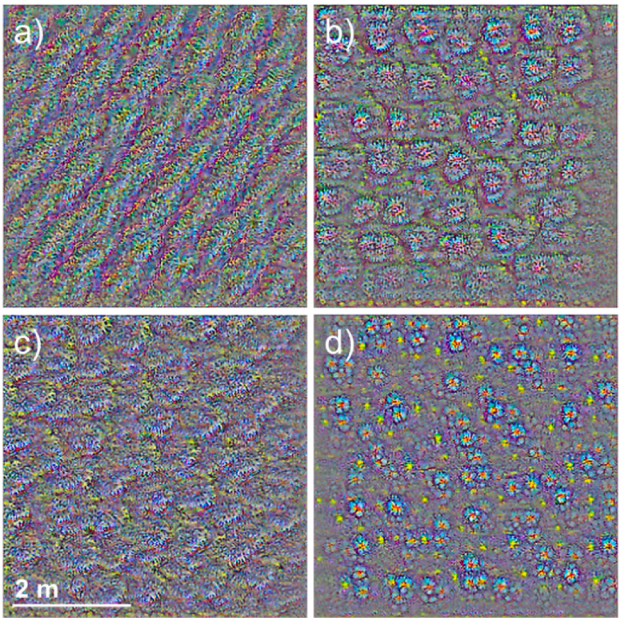

Fig. 6. Selection of synthetic filter visualizations resembling patterns that would most stimulate the CNN (for technical details see Appendix C). The filter visualizations correspond to the dimensions of the tile size (here 256 × 256 pixel) and highlight the importance of fine-scale features.

Nevertheless, a higher resolution also comes in hand with lower spatial coverage of the UAV data and therefore the ideal trade-off between area coverage and spatial resolution should be considered when designing imaging campaigns. Given its good performance at very-high spatial resolution, CNNs applied to small extents can aid in the generation of reference data for remote sensing approaches at large spatial extents with a coarse spatial resolution (Kattenborn et al., 2019b).

4.5. Model generalization

The validation from the test dataset revealed high generalization abilities for the identification of 14 classes with a mean F1-score of 73% (128 × 128 pixel tiles, RGB + nDSM) and evenly distributed site-specific F1-scores across all sites and years. Sothe et al. (2020) reported problems in generalizing the learned features of nine tree species when individual CNNs were trained locally on different sites. For the discrimination of two *Pinus* species from non-*Pinus*, Natesan et al. (2019) reported a higher F1-score (80%) when CNNs were trained with samples from several years than when trained with only one year (50%). Similarly, Weinstein et al. (2020) reported high generalization abilities of CNNs for the detection of individual trees over four different forest types. They found a CNN trained on all available forest types to outperform individual, locally trained CNNs. Their results suggest high model transferability when CNNs are trained over large and heterogeneous data.

The data used in this study were collected in 51 one-hectare plots in two different forest types (temperate deciduous and mixed coniferous forests), different managements (managed and unmanaged), and study areas (Southern Black Forest and NP Hainich), and included a variety of growth stages. UAV data acquisition took place in the years 2017–2019 from June–September (day of the year 110–307) and covered a variety of illumination situations due to the different recording times from 7 am to 6 pm. In addition, data augmentation was used to increase size and variance of the training dataset, and to minimize spatial autocorrelation of adjacent tiles. We, therefore, assume that the high generalization abilities of the CNNs, as indicated by the overall accuracy, as well as the evenly distributed F1-scores across all sites, are the result of including many sites from different areas, different forest structures, different seasons and years, and varying illumination properties. This way it can be ensured that the CNN learns features of tree species that are representative for different growth stages and site conditions. In line with Weinstein et al. (2020) we assume that more training data and increased heterogeneity will further enhance the accuracy and generalization of CNNs. Coupled with the establishment of large databases of remote sensing and reference data (Zhu et al., 2017), this opens the possibilities of transfer learning or even the creation of universal models. In the case of transfer learning, CNNs are pre-trained on large and heterogeneous datasets and the model weights are fine-tuned for the respective use case, while a universal model is trained on all existing data and is therefore transferable across sites. Weinstein et al. (2020) already demonstrated this future perspective for the detection of trees over various landscapes. Our results show a path for widely applicable mapping of tree species in temperate forests using only low-cost UAV-based RGB data and CNNs.

4.6. CNN architecture

Frequently applied approaches for mapping vegetation from remote sensing data using CNN comprise image classification/regression or object detection. Classification and regression approaches either rely on prior segmentation of the target (cf. Hartling et al., 2019; Natesan et al., 2019; Sothe et al., 2020) or predictions are assigned to an entire image tile (cf. Kattenborn et al., 2020; Qian et al., 2020; Rezaee et al., 2018). The output of the object detection task is typically a bounding box drawn around the object of interest (cf. Chen et al., 2019; Csillik et al., 2018; dos Santos et al., 2019; Fromm et al., 2019; Safonova et al., 2019; Weinstein et al., 2020, 2019). The capability of such approaches to derive spatially explicit maps can be limited by a number of reasons: (1) they require additional preprocessing steps (e.g. segmentation, background removal), or (2) classification on the single-pixel level is required to retrieve pixel-based predictions, or (3) the results represent object location and (rectangular) extent rather than spatially explicit objects. In contrast, semantic segmentation is an end-to-end learning approach that combines segmentation and classification in a pixel-based fashion at the original spatial resolution and is thus ideally suited for mapping tree species in forests. No prior segmentation or classification is necessary apart from the creation of training samples.

In this study, we used the U-net architecture given its good performance even with small amounts of labeled data (Ronneberger et al., 2015). Besides its relatively low computational complexity, several studies have successfully demonstrated the suitability of the U-net for mapping single plant species (Kattenborn et al., 2019a; Wagner et al., 2020), individual trees (Freudenberg et al., 2019; Lobo Torres et al., 2020), forest damage and disturbance (Hamdi et al., 2019; Kislov and Korznikov, 2020; Wagner et al., 2019), forest types (Wagner et al., 2019), and plant communities (Kattenborn et al., 2019a). Since we were interested in the general applicability of CNNs for mapping forest tree species, we did not aim for benchmarking multiple architectures. Besides U-net, a variety of more elaborate model architectures for semantic segmentation exist (e.g., FC-DenseNet (Jegou et al., 2017), SegNet (Badrinarayanan et al., 2017), or DeepLabv3+ (Chen et al., 2017)). Lobo Torres et al. (2020) compared five models of varying complexity, namely U-net, FC-DenseNet, SegNet, and two variants of the DeepLabv3+ for semantic segmentation of tree species in urban environments. Their results suggest the model accuracies of the architectures to be comparable, whereas more complex models (i.e. DeepLabv3+) required up to two or four times more time during model training and inference.

Another alternative to semantic segmentation is instance segmentation, i.e. segmenting not only classes but also individuals. Detecting individual trees would truly be of high value for forestry and conservation. However, from our experience from the visual interpretation, many tree crowns of the same species are hard to differentiate because branches may have crown-like characteristics (e.g., *F. sylvatica*, *F. excelsior*). This suggests that generating labels for the segmentation of individuals requires more sophisticated procedures that either require in-situ data with high-quality GNSS data on tree stem locations or a sophisticated link to ancillary remote sensing data (e.g. LiDAR data) to aid visual inspection. However, even if labels were available, we doubt that instance segmentation algorithms would be able to locate individuals in RGB orthomosaics given the above-described difficulties.

4.7. Reference data

Reference data were derived through manual delineation in the orthomosaics after visual interpretation. Given the very high spatial resolution (<1.35 cm) of the imagery tree species were clearly identifiable. To minimize errors in the visual interpretation, we used additional information from forest inventories (i.e., tree height, DBH, and partly tree stem coordinates), cross-checked the delineations by at least one other interpreter, and removed tree crowns that we could not identify with certainty. Several reasons suggest that when using very high-resolution image data, no other method is appropriate for obtaining reference data, especially in the case of deep learning: (1) acquisition of in-situ data of the required amount is costly, time- and labor-intensive and might thus not be feasible; (2) reference data from visual interpretation of the image data is not subject to geolocation errors of GNSS-measurements as for in-situ measurements. Such errors are typically in the range of decimeters to meters when using differential GNSS and might even exceed several meters when using stand-alone GNSS, particularly under dense canopies (Kaartinen et al., 2015; Valbuena et al., 2012). Especially when using very high-resolution imagery, errors might exceed the spatial resolution by far, which makes in-situ measurements difficult to use; (3) in-situ data that can be recorded with the least effort in forests are typically point observations (e.g. tree stem coordinates) that do not necessarily allow for a spatially explicit link with the targeted variable (e.g. tree crowns). However, visual interpretation from RGB imagery is not free of misinterpretation, but due to the high amount of reference data required for CNNs and the need for high precision geolocation within the high-resolution imagery, it seems to be the most effective way of collecting reference data. Furthermore, it has been shown that CNNs can compensate for faulty labels to some extent and that correct classes were predicted despite incorrectly labeled reference data (Hamdi et al., 2019; Kattenborn et al., 2020).

A probable reason for the decreasing accuracy with decreasing share of the species might be that less abundant species share similar features with more abundant species and are therefore misclassified. This could be the case especially with *F. sylvatica* and *C. betulus* whose leaves have a similar size and shape. On the other hand, rarely occurring species that show no or less similarities with more abundant species (e.g. small leaves and distinct habitus of *B. pendula*) have also been poorly classified, most likely due to their underrepresentation in the data set. The majority of observations in this study was situated along a gradient of forest connectivity and structure (Storch et al., 2020) and, hence, not optimized for representing all species for a remote sensing application. Thus, designing or updating a database towards sufficient observations for rare taxa, may be key for an accurate species mapping. More technical alternatives for improving the accuracy for underrepresented species include tuning the weights in the loss function and setting them higher for less frequent classes (which in our case however rather worsened the results at some point), weight updating (i.e., updating the weights of an already trained CNN using solely data of less frequent species), or sampling tiles containing less frequent species more often during model training. The latter, however, was not an option due to the large range of occurrences in our dataset (0.2–33% area-related share), as it would have drastically reduced the dataset size or assumedly would have introduced large redundancies.

For the genera *Acer, Tilia*, and *Quercus*, we grouped the respective species into genus-level classes, since they were only present in very small quantities in the plots. While for some of these species a distinction in the RGB data might be easier due to visible differences in tree habitus or leaf shape (e.g. *Acer platanoides* and *Acer pseudoplatanus*), for other species with only subtle differences it might be very difficult or even impossible (e.g. *Quercus petraea* and *Quercus robur*). The mapping of such species using very high-resolution UAV-based RGB data and CNNs could prove to be very difficult and should be subject to further research.

5. Conclusion

We showed that RGB imagery from consumer-grade UAVs in concert with a CNN-based semantic semgnetation enables to map tree species across heterogeneous temperate forests with high accuracies. We tested CNN-based tree species mapping with different tile sizes, incorporation of height information (nDSM), and varying spatial resolutions. The tile size had no prominent influence on the model accuracy if enough reference data was available. By choosing a larger tile size, a larger spatial context was considered by the CNN, thereby minimizing edge

effects, and accelerating the application over a large spatial extent. Additional height information from nDSMs slightly increased the model accuracy. Still, the inclusion of nDSMs should be carefully considered, since the increased computational complexity of the CNN and the need for a digital terrain model are major drawbacks. A high spatial resolution was indeed decisive for the accurate mapping of forest tree species using RGB data. Overall, our results showed that CNN models generalize well over the diverse dataset in terms of site conditions, forest types, stand structure, phenology, and illumination properties.

Our findings underline the synergies between high resolution UAV imagery and CNN-based segmentation procedures. In view of the increasingly easy and affordable way to obtain very high-resolution RGB imagery with consumer-grade UAVs, and given that air- and spaceborne data currently do not provide comparable spatial resolutions, UAVs can play a crucial role in the mapping of forest tree species. CNN are able to learn species-specific features from such high resolution imagery, while their end-to-end learning capabilities make extensive preprocessing of remote sensing data obsolete and simplify a widespread application. Our study demonstrates the potential of a concerted use of UAVs and CNNs and thus provides promising future perspectives for applications in forestry or large-scale and long-term ecological research. Such applications usually require large-scale and accurate maps of forest tree species, for which field-based methods might be too labor-intensive while commonly used machine learning approaches might not be accurate enough.

Given that training data generation for semantic segmentation is a laborious task and generalization across forest types is of primary concern, a flexible, widespread, and operational application of such an approach may be facilitated by incorporating transfer learning (i.e. updating and refining the learned feature representations of an already trained CNN by retraining the model with new image data) or the development of universal models (i.e. one single model that has been trained over a variety of landscapes and many species).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the public forest authorities involved for facilitating this research and the State Agency for Spatial Information and Rural Development of Baden-Württemberg (LGL) for the provisioning of data. We thank the administration of the Hainich National Park for the opportunity for research within the National Park and the Biodiversity Exploratories project (German Research Foundation - DFG Priority Program 1374 “Infrastructure-Biodiversity-Exploratories”) for their cooperation, support, and provision of data. We want to especially thank Johannes Penner from the ConFoBi coordination team and Andrey Lessa for the provision of the forest inventory data for the black forest region. Many thanks to Kathrin Wagner and Benjamin Stöckigt for their aid in UAV image acquisition and Johannes Hoffmann and Timo Schmid for the assistance in visual image interpretation. The study has been funded by the German Aerospace Centre (DLR) on behalf of the Federal Ministry of Economics and Technology (BMWi), FKZ 50EE1909A. The data acquisition within the black forest was founded by the German Research Foundation DFG (GRK 2123).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2020.10.015.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.

Allaire, J.J., Chollet, F., 2019. keras: R Interface to “Keras.” R Packag. version 2.2.5.0. https//CRAN.R-project.org/package=keras.

Allaire, J.J., Tang, Y., 2019. tensorflow: R Interface to “TensorFlow.” R Packag. version 2.0.0. https//CRAN.R-project.org/package=tensorflow.

Allaire, J.J., Tang, Y., Ushey, K., 2019. tfdatasets: Interface to “TensorFlow” Datasets. R Packag. version 2.0.0. https//CRAN.R-project.org/package=tfdatasets.

Audebert, N., Le Saux, B., Lefevre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. IEEE Geosci. Remote Sens. Mag. https://doi.org/10.1109/MGRS.2019.2912563.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615.

Brodrick, P.G., Davies, A.B., Asner, G.P., 2019. Uncovering ecological patterns with convolutional neural networks. Trends Ecol. Evol. 34, 734–745. https://doi.org/10.1016/j.tree.2019.03.006.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv. http://arxiv.org/abs/1706.05587v3.

Chen, Y., Lee, W.S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., He, Y., 2019. Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. Remote Sens. 11, 1584. https://doi.org/10.3390/rs11131584.

Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., Shelhamer, E., 2014. cuDNN: Efficient Primitives for Deep Learning.

Chollet, F., Allaire, J.J., 2017. R Interface to Keras. GitHub. https://github.com/rstudio/keras.

Csillik, O., Cherbini, J., Johnson, R., Lyons, A., Kelly, M., 2018. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. Drones 2, 39. https://doi.org/10.3390/drones2040039.

dos Santos, A.A., Marcato Junior, J., Araújo, M.S., di Martini, D.R., Tetila, E.C., Siqueira, H.L., Aoki, C., Eltner, A., Matsubara, E.T., Pistori, H., Feitosa, R.Q., Liesenberg, V., Gonçalves, W.N., 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. Sensors 19, 1–11. https://doi.org/10.3390/s19163595.

FAO, 2020. Global Forest Resources Assessment 2020 – Key findings, Rome. https://doi.org/10.4060/ca8753en. https://doi.org/10.4060/ca8753en.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2016.08.013.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.D., Weisser, W.W., 2010. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl. Ecol. 11, 473–485. https://doi.org/10.1016/j.baae.2010.07.009.

Franklin, S.E., Ahmed, O.S., 2018. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. Int. J. Remote Sens. 39, 5236–5245. https://doi.org/10.1080/01431161.2017.1363442.

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., Kleinn, C., 2019. Large scale palm tree detection in high resolution satellite images using U-Net. Remote Sens. 11, 1–18. https://doi.org/10.3390/rs11030312.

Frey, J., Kovach, K., Stemmler, S., Koch, B., 2018. UAV photogrammetry of forests as a vulnerable process. A sensitivity analysis for a structure from motion RGB-image pipeline. Remote Sens. 10, 912. https://doi.org/10.3390/rs10060912.

Fricker, G.A., Ventura, J.D., Wolf, J.A., North, M.P., Davis, F.W., Franklin, J., 2019. A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. Remote Sens. 11 https://doi.org/10.3390/rs11192326.

Fromm, M., Schubert, M., Castilla, G., Linke, J., McDermid, G., 2019. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. Remote Sens. 11 https://doi.org/10.3390/rs11212585.

Gini, R., Passoni, D., Pinto, L., Sona, G., 2014. Use of unmanned aerial systems for multispectral survey and tree classification: A test in a park area of northern Italy. Eur. J. Remote Sens. 47, 251–269. https://doi.org/10.5721/EuJRS20144716.

Hamdi, Z.M., Brandmeier, M., Straub, C., 2019. Forest damage assessment using deep learning on high resolution remote sensing data. Remote Sens. 11, 1–14. https://doi.org/10.3390/rs11171976.

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., Carron, J., 2019. Urban tree species classification using a worldview-2/3 and liDAR data fusion approach and deep learning. Sensors 19, 1284. https://doi.org/10.3390/s19061284.

Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society, pp. 1175–1183. https://doi.org/10.1109/CVPRW.2017.156.

Kaartinen, H., Hyyppä, J., Vastaranta, M., Kukko, A., Jaakkola, A., Yu, X., Pyörälä, J., Liang, X., Liu, J., Wang, Y., Kaijaluoto, R., Melkas, T., Holopainen, M., Hyyppä, H.,

2015. Accuracy of kinematic positioning using global satellite navigation systems under forest canopies. Forests 6, 3218–3236. https://doi.org/10.3390/f6093218.
Kändler, G., Cullmann, D., 2015. Regionale Auswertung der Bundeswaldinventur 3. Wuchsgebiet Schwarzwald. Freiburg, Germany. Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg (FVA).
Kattenborn, T., Eichel, J., Fassnacht, F.E., 2019a. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. Sci. Rep. 9, 17656. https://doi.org/10.1038/s41598-019-53797-9.
Kattenborn, T., Eichel, J., Wiser, S., Burrows, L., Fassnacht, F.E., Schmidtlein, S., 2020. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. Remote Sens. Ecol. Conserv. 1–15 https://doi.org/10.1002/rse2.146.
Kattenborn, T., Lopatin, J., Förster, M., Braun, A.C., Fassnacht, F.E., 2019b. UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. Remote Sens. Environ. 227, 61–73. https://doi.org/10.1016/j.rse.2019.03.025.
Kislov, D.E., Korznikov, K.A., 2020. Automatic windthrow detection using very-high-resolution satellite imagery and deep learning. Remote Sens. 12, 1145. https://doi.org/10.3390/rs12071145.
Komárek, J., 2020. The perspective of unmanned aerial systems in forest management. Do we really need such details? Appl. Veg. Sci. avsc.12503 https://doi.org/10.1111/avsc.12503.
Li, W., Fu, H., Yu, L., Cracknell, A., 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sens. 9 https://doi.org/10.3390/rs9010022.
Lobo Torres, D., Feitosa, R.Q., Nigri Happ, P., Elena Cué La Rosa, L., Marcato Junior, J., Martins, J., Olã Bressan, P., Gonçalves, W.N., Liesenberg, V., 2020. Applying Fully Convolutional Architectures for Semantic Segmentation of a Single Tree Species in Urban Environment on High Resolution UAV Optical Imagery. Sensors 20, 563. https://doi.org/10.3390/s20020563.
López-Jiménez, E., Vasquez-Gomez, J.I., Sanchez-Acevedo, M.A., Herrera-Lozada, J.C., Uriarte-Arcia, A.V., 2019. Columnar cactus recognition in aerial images using a deep learning approach. Ecol. Inform. 52, 131–138. https://doi.org/10.1016/j.ecoinf.2019.05.005.
Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.
Michez, A., Piégay, H., Lisein, J., Claessens, H., Lejeune, P., 2016. Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. Environ. Monit. Assess. 188, 1–19. https://doi.org/10.1007/s10661-015-4996-2.
Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of Mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning. Forests 9, 736. https://doi.org/10.3390/f9120736.
Müller, K., Wickham, H., 2019. tibble: Simple Data Frames. R Packag. version 2.1.3. https//CRAN.R-project.org/package=tibble.
Natesan, S., Armenakis, C., Vepakomma, U., 2019. Resnet-based tree species classification using UAV images, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 475–481. https://doi.org/10.5194/isprs-archives-XLII-2-W13-475-2019.
Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyyppä, J., Saari, H., Pölönen, I., Imai, N.N., Tommaselli, A.M.G., 2017. Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. Remote Sens. 9, 185. https://doi.org/10.3390/rs9030185.
Nezami, S., Khoramshahi, E., Nevalainen, O., Pölönen, I., Honkavaara, E., 2020. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. Remote Sens. 12, 1–19. https://doi.org/10.20944/preprints202002.0334.v1.
Osco, L.P., de Arruda, M. dos S., Marcato Junior, J., da Silva, N.B., Ramos, A.P.M., Moryia, É.A.S., Imai, N.N., Pereira, D.R., Creste, J.E., Matsubara, E.T., Li, J., Gonçalves, W.N., 2020. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. ISPRS J. Photogramm. Remote Sens. 160, 97–106. https://doi.org/10.1016/j.isprsjprs.2019.12.010.
Qian, W., Huang, Y., Liu, Q., Fan, W., Sun, Z., Dong, H., Wan, F., Qiao, X., 2020. UAV and a deep convolutional neural network for monitoring invasive alien plants in the wild. Comput. Electron. Agric. 174, 105519 https://doi.org/10.1016/j.compag.2020.105519.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput, Vienna, Austria https://www.R-project.org.
Rezaee, M., Mahdianpari, M., Zhang, Y., Salehi, B., 2018. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 3030–3039. https://doi.org/10.1109/JSTARS.2018.2846178.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Munich, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4.
Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., Herrera, F., 2019. Detection of Fir Trees (Abies sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. Remote Sens. 11, 643. https://doi.org/10.3390/rs11060643.
Schall, P., Schulze, E.D., Fischer, M., Ayasse, M., Ammer, C., 2018. Relations between forest management, stand structure and productivity across different types of Central European forests. Basic Appl. Ecol. 32, 39–52. https://doi.org/10.1016/j.baae.2018.02.007.
Sothe, C., de Almeida, C.M., Schimalski, M.B., la Rosa, L.E.C., Castro, J.D.B., Feitosa, R.Q., Dalponte, M., Lima, C.L., Liesenberg, V., Miyoshi, G.T., Tommaselli, A.M.G., 2020. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. GIScience Remote Sens. 57, 369–394. https://doi.org/10.1080/15481603.2020.1712102.
Storch, I., Penner, J., Asbeck, T., Basile, M., Bauhus, J., Braunisch, V., Dormann, C.F., Frey, J., Gärtner, S., Hanewinkel, M., Koch, B., Klein, A., Kuss, T., Pregernig, M., Pyttel, P., Reif, A., Scherer-Lorenzen, M., Segelbacher, G., Schraml, U., Staab, M., Winkel, G., Yousefpour, R., 2020. Evaluating the effectiveness of retention forestry to enhance biodiversity in production forests of Central Europe using an interdisciplinary, multi-scale approach. Ecol. Evol. 10, 1489–1509. https://doi.org/10.1002/ece3.6003.
Trier, Ø.D., Salberg, A.B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., Aarsten, D., 2018. Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. Eur. J. Remote Sens. 51, 336–351. https://doi.org/10.1080/22797254.2018.1434424.
Valbuena, R., Mauro, F., Rodriguez-Solano, R., Manzanera, J.A., 2012. Accuracy and precision of GPS receivers under forest canopies in a mountainous environment. Spanish J. Agric. Res. 8, 1047–1057.
Wagner, F.H., Sanchez, A., Aidar, M.P.M., Rochelle, A.L.C., Tarabalka, Y., Fonseca, M.G., Phillips, O.L., Gloor, E., Aragão, L.E.O.C., 2020. Mapping Atlantic rainforest degradation and regeneration history with indicator species using convolutional network. PLoS One 15, e0229448. https://doi.org/10.1371/journal.pone.0229448.
Wagner, F.H., Sanchez, A., Tarabalka, Y., Lotte, R.G., Ferreira, M.P., Aidar, M.P.M., Gloor, E., Phillips, O.L., Aragão, L.E.O.C., 2019. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. Remote Sens. Ecol. Conserv. 1–16 https://doi.org/10.1002/rse2.111.
Wallace, L., Bellman, C., Hally, B., Hernandez, J., Jones, S., Hillman, S., 2019. Assessing the ability of image based point clouds captured from a UAV to measure the terrain in the presence of canopy cover. Forests 10, 284. https://doi.org/10.3390/f10030284.
Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A., White, E.P., 2020. Cross-site learning in deep learning RGB tree crown detection. Ecol. Inform. 56, 101061 https://doi.org/10.1016/j.ecoinf.2020.101061.
Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A., White, E.P., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sens. 11, 1309. https://doi.org/10.3390/rs11111309.
Zhang, Liangpei, Zhang, Lefei, Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40. https://doi.org/10.1109/MGRS.2016.2540798.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a review. IEEE Geosci. Remote Sens. Mag. https://doi.org/10.1109/MGRS.2017.2762307.