Remote Sensing in Ecology and Conservation
Open Access
ZSL LET'S WORK FOR WILDLIFE

ORIGINAL RESEARCH

Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery

Teja Kattenborn[1], Jana Eichel[2], Susan Wiser[3], Larry Burrows[3], Fabian E. Fassnacht[1] & Sebastian Schmidtlein[1]

[1]Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131, Karlsruhe, Germany
[2]Department of Physical Geography, Utrecht University, Princetonlaan 8a, 3584 CB, Utrecht, The Netherlands
[3]Manaaki Whenua – Landcare Research, PO Box 69040, Lincoln 7640, New Zealand

Keywords
Convolutional Neural Networks (CNN), deep learning, invasive species, orthophoto, remote sensing, tree species, Unmanned Aerial Vehicles (UAV), vegetation succession

Correspondece
Teja Kattenborn, Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131, Karlsruhe, Germany. Tel: +49 721 608-43829; Fax: +49 721 608-43738; E-mail: teja.kattenborn@kit.edu

Editor: Ned Horning
Associate Editor: Nicola Clerici

Received: 19 July 2019; Revised: 30 October 2019; Accepted: 24 December 2019

doi: 10.1002/rse2.146

Remote Sensing in Ecology and Conservation 2020;**6** (4):472–486

Abstract

Unmanned Aerial Vehicles (UAV) greatly extended our possibilities to acquire high resolution remote sensing data for assessing the spatial distribution of species composition and vegetation characteristics. Yet, current pixel- or texture-based mapping approaches do not fully exploit the information content provided by the high spatial resolution. Here, to fully harness this spatial detail, we apply deep learning techniques, that is, Convolutional Neural Networks (CNNs), on regular tiles of UAV-orthoimagery (here 2–5 m) to identify the cover of target plant species and plant communities. The approach was tested with UAV-based orthomosaics and photogrammetric 3D information in three case studies, that is, (1) mapping tree species cover in primary forests, (2) mapping plant invasions by woody species into forests and open land and (3) mapping vegetation succession in a glacier foreland. All three case studies resulted in high predictive accuracies. The accuracy increased with increasing tile size (2–5 m) reflecting the increased spatial context captured by a tile. The inclusion of 3D information derived from the photogrammetric workflow did not significantly improve the models. We conclude that CNN are powerful in harnessing high resolution data acquired from UAV to map vegetation patterns. The study was based on low cost red, green, blue (RGB) sensors making the method accessible to a wide range of users. Combining UAV and CNN will provide tremendous opportunities for ecological applications.

Introduction

Unmanned Aerial Vehicles (UAVs) or Remotely Piloted Aircraft Systems (RPAS) have evolved to become an invaluable remote sensing tool for mapping and monitoring vegetation. Specifically, UAV-based photogrammetry based on structure from motion (SfM) algorithms expanded the ability to obtain high resolution orthomosaics and 3D information. Various studies have demonstrated the value of UAV image capture and photogrammetry for vegetation assessments, including the mapping of species at the individual and stand level (Fritz et al. 2013; Kattenborn et al. 2014; Sankey et al. 2017; Cao et al. 2018; Lopatin et al. 2018), plant communities (Husson et al. 2014; Malenovský et al. 2017), canopy structure (Getzin et al. 2012; Sankey et al. 2017) and plant traits (Fritz et al. 2013; Zarco-Tejada et al. 2013, Tian et al. 2017). UAV-based mapping is often conducted at a local scale, but can also be extended to larger scales through a combination with satellite-based remote sensing (Kattenborn et al. 2019).

Despite this demonstrated potential of UAV data, there is still space for improvement. For example, pixel-based remote sensing approaches do not fully exploit the high spatial resolution of UAV data (Zhang et al. 2006; Lopatin et al. 2017; Müllerová et al. 2017). Higher spatial detail results in higher spectral variance per unit area, making it more challenging to assign a pixel-based observation to a class or to fit a regression (Hsieh et al. 2001; Aplin 2006). At the same time it is evident that higher

spatial resolution can reveal characteristic patterns, such as branching architectures or canopy shapes that can be of high value to discriminate both individual species and vegetation types (Kattenborn et al. 2019). Thus, instead of examining single pixels alone, incorporation of the spatial context will allow the spatial detail of high resolution imagery to be fully exploited. A common approach is to derive 2D texture metrics from orthoimagery, such as grey level co-occurrence matrices (Haralick 1979), which have been frequently applied in the context of UAV mapping (e.g. Michez et al. 2016; Lu and He 2017). Another option is to derive metrics describing the 3D structure of the canopy derived from the UAV-based photogrammetric point clouds (Brodu and Lague 2012; Getzin et al. 2012; Kattenborn et al. 2014, 2019; Lopatin et al. 2019). Yet, these 2D texture and 3D structure approaches cannot automatically and sufficiently exploit the available information to detect differences in vegetation characteristics that are readily perceived by human eyes (Lopatin et al. 2018; Kattenborn et al. 2019).

To address this, we test Convolutional Neural Networks (CNNs) to map both individual species and vegetation types, as this deep learning technique has been shown to be accurate and functional in a range of image recognition tasks and contests (Krizhevsky et al. 2012; Shin et al. 2016; Huang et al. 2018). The principle of CNN was inspired by the functioning of the visual cortex, where neurons are sensitive to visual stimuli at varying scales in different and partly overlapping regions of the visual space, also known as receptive fields (Hubel and Wiesel 1962; Cadieu et al. 2014; Angermueller et al. 2016). The combination of multiple firing neurons, excited from multiple visual stimuli, allows the brain to perceive spatial textures and context within the field of vision. CNN behave in an analogous way and, in contrast to previously mentioned approaches using 2D texture or 3D structure metrics, the spatial metrics do not have to be hand-engineered because CNN automatically learn the relevant patterns. As a result, CNN require minimal preprocessing to capture spatial properties.

Convolutional Neural Networks have already been successfully applied in vegetation related applications, for example, the image-based detection of plant diseases (Sladojevic et al. 2016), plant phenotyping (Ubbens and Stavness 2017) and image-based identification of plant species (see e.g. *Pl@ntnet*, *Flora Incognita*, Joly et al. 2016; Wäldchen and Mäder 2018). However, the application of CNN to UAV imagery for mapping vegetation properties remains rare due to various challenges. These include (1) the complexity of natural vegetation canopies, (2) the need for spatially explicit and extensive reference data for training and validation and (3) that mapping approaches are not dedicated to characterize single images, but to locate and characterize specific features within images.

1) Complexity of natural vegetation canopies: CNN are most often used in classification tasks (Krizhevsky et al. 2012; Hu et al. 2015; Wäldchen and Mäder 2018, Wagner et al. 2019). However, as a response to gradual changes of environmental factors, vegetation canopies often feature corresponding gradual changes in species cover, community composition or canopy properties (Foody et al. 1992; Schmidtlein and Sassin 2004; Rocchini et al. 2013). Moreover, pixels may contain more than one vegetation type, even in very high resolution data. Therefore, vegetation often tends to be more appropriately described by continuous metrics (e.g. the coverage of a species [%]) and a robust and flexible mapping approach should ideally characterize the target variable using a continuous scale rather than discrete classes.
2) Reference data availability: The predictive accuracy of CNN commonly benefits from large quantities of training data (also known as labels). Yet, in most remote sensing applications reference data are generally a scarce commodity due to the cost of ground-based sampling and difficulties in accessing sites. Moreover, the use of field data may be impaired by the inability to accurately align the geolocation of field-based observations with remote sensing imagery and sampling bias resulting from ground-based cover estimates (Lunetta et al. 1991; Lepš and Hadincová 1992; Valbuena et al. 2010; Kaartinen et al. 2015; Leitão et al. 2018). One alternative is to use spatially explicit observations from UAV imagery. This is feasible if preexisting ground-based samples are available to aid the visual delineation of the target canopies or if the target variable is the cover of an easily identified species or vegetation type (Vanha-Majamaa et al. 2000; Luscier et al. 2006; Lisein et al. 2015; Kattenborn et al. 2018, 2019).
3) Location and characterization of features within images: Originally, CNN approaches were developed to analyze images where the objects of interest cover a substantial part of the image and the entire image is assigned to a class (Krizhevsky et al. 2012). In contrast, an application of CNN in vegetation remote sensing must enable to locate vegetation features within the orthoimagery and display corresponding spatial gradients. A solution to this problem is to apply CNN to equally spaced tiles extracted from the orthoimagery.

In the current paper, we propose a workflow for mapping species and vegetation types where CNN are trained using tiles of the orthoimage together with cover values of species and vegetation types determined from the orthoimage itself (see Fig. 1 for an illustration). The fitted CNN is applied to gridded tiles of the entire orthoimage (or further orthoimages) to generate spatially continuous maps of the variable of interest.

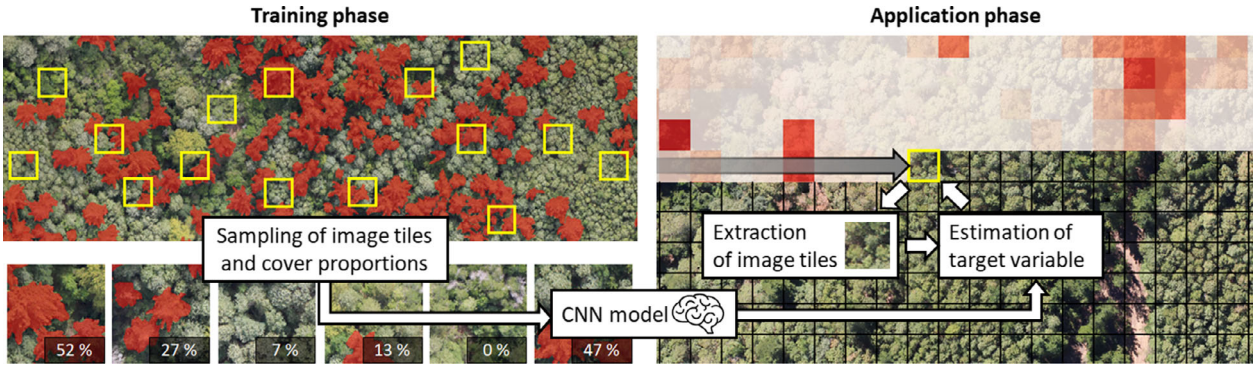

Figure 1. Scheme of the proposed CNN-based procedure for vegetation mapping. The training phase (left) is based on sampling random image tiles and cover proportions of the target class (red). In the application phase (right), the cover [%] of the target class is predicted in form of continuous maps (indicated by the grey arrow and colored grid) using the trained CNN model and regularly extracted image tiles.

This procedure addresses the challenges stated above: Smooth vegetation gradients and fine-scale variation in vegetation cover are accounted for by mapping the target variable on a continuous scale (here, per cent cover instead of assignment to discrete classes). The problem of limited, spatially explicit field reference data are addressed by visual interpretation of the imagery. The problem of locating the vegetation features within the orthoimages is addressed using the tiles approach.

We test the proposed approach in three case studies representing three common remote sensing applications: (1) Mapping plant successions in the foreland of the Mueller Glacier, New Zealand, (2) mapping of two woody invasive species in Central Chile; and (3) tree species mapping in a structurally complex primary forest in Waitutu, New Zealand. We address the following research questions:

- How accurate are CNN models combined with UAV RGB imagery for mapping the spatial extent of different species and different types of plant communities?
- What is the influence of tile size on mapping accuracy?
- Does the addition of photogrammetric 3D information increase mapping accuracy?

Materials and Methods

Study sites description and data acquisition

Case study on herb and shrub communities in a New Zealand glacier foreland ('Vegetation Succession')

We tested the proposed CNN approach for mapping a spatially complex vegetation succession in the Mueller glacier foreland. Located in Mount Cook National Park (New Zealand), it comprises about 450 ha and is characterized by a sequence of lateral and latero-frontal moraines formed 125–3370 ± 290 years ago (Winkler and Lambiel 2018). Previous studies on vegetation succession in the adjacent East Hooker Valley found distinct plant communities on different aged terrain with pioneer and early successional stages characterized by the herb *Epilobium melanocaulon* and the moss *Racomitrium lanuginosum*, intermediate successional stages with *Festuca* and *Chionochloa* grassland and later successional shrubland with woody *Dracophyllum* spp. (Gellally, 1982). RGB data were acquired in seven individual flights with a DJI Phantom 4 Pro+ in February 2018. A flight height of 100 m ensured an image resolution of 5 cm per pixel. The area covered in each flight ranged from 20–50 ha. To define vegetation successional stages, a vegetation survey was conducted using 55 plots (2 × 2 m) distributed across the area following a stratified-random approach based on Normalized Difference Vegetation Index maps calculated from Sentinel-2 data. For each plot, the species cover was quantified based on the vertical projection of the perimeter of the crown of each component plant of that species. Using the species cover data, four vegetation stages were classified using the Isopam algorithm (Schmidtlein and Sassin 2004; implemented in R, distance Bray-Curtis, expert mode). Isopam is a cluster algorithm that is based on a brute force approach to find the optimum separation of all descriptors – here species cover. The four classified vegetation successional stages included a pioneer community, two intermediate and one late successional class. The CNN training data for these classes were delineated in the orthomosaic using visual image interpretation guided by the classified plot data (see supplementary information 2 for sample photographs and orthoimagery of each successional class). Due to the large spatial extent of the study area, the analysis was restricted to seven areas (150 × 150 m). These areas were manually positioned so

that all four successional stages were sufficiently covered (see supplementary information 5 for map overviews).

Case study on invasive woody species in Central Chile ('Plant Invasion')

We tested the potential of CNN for mapping woody plant species invading native vegetation using the examples of *Pinus radiata* and *Ulex europaeus* in Central Chile. *Pinus radiata* is a coniferous tree, which was introduced to Central Chile for timber production (Clapp 1995). It frequently invades from managed plantations into natural *Nothofagus* and sclerophyllous forest stands (Bustamante and Simonetti 2005; Guerrero and Bustamante 2007). *Ulex europaeus* was introduced to Central Chile as natural hedge for livestock and for its ornamental value. *Ulex europaeus* has spread vastly in Chile, causing negative economic impacts to agriculture and silviculture (Norambuena et al. 2000).

For the two invasive species, we used RGB data from four independent octocopter flights (Okto-XL, HiSystems GmbH, Germany). The areas were selected with input from local experts to cover representative situations of the invasion. The octocopter was equipped with a Canon 100D with an 18 mm lens. The UAV flights were performed in Chilean summer and spring (March, November) for *Pinus radiata* and *Ulex europaeus*, respectively. We performed the image flights at an average height of 150 m above ground ensuring a spatial resolution of at least 3 cm for the RGB imagery (depending on the terrain). The area covered in each flight ranged from 21 to 37 ha. Further details on the study sites and the UAV data acquisition are described in Kattenborn et al. (2019) and Lopatin et al. (2019).

Case study on tree species in New Zealand primary forests ('Tree Species')

We tested the mapping of tree species in a primary and structurally complex forest in Waitutu, Southland and New Zealand. The target species *Metrosideros umbellata*, an angiosperm of the Myrtaceae and *Dacrydium cupressinum,* a gymnosperm of the Podocarpaceae, have been selected because they were widespread in the areas of investigation. Other important tree species in these forest canopies include the angiosperms *Weinmannia racemosa, Lophozonia menziesii, Fuscospora cliffortiodes* and the gymnosperms *Podocarpus laetus* and *Prumnopitys ferruginea*. We used UAV data acquired for three plots (edge size of 100–150 m) along with a full inventory of individual trees in each of these plots, with information about position in space, species and diameter at breast height. In-situ data were used to assist the visual delineation of the target species and served as starting point of digitizing the individual crowns. Further details on the individual tree data acquisition, vegetation and site conditions are for example given in Coomes et al. (2005) and Parfitt et al. (2005). The UAV-based RGB imagery was acquired in November 2017 with an average ground sampling distance of 3 cm data using an octocopter (Okto-XL, HiSystems GmbH, Germany) carrying a Canon 100D with an 18 mm lens.

UAV data processing and reference data acquisition

We produced orthomosaics using the SfM-based photogrammetric processing chain in Agisoft Photoscan (Agisoft, Russia, vers. 1.4.2). The applied processing chain included image matching through bundle adjustment and dense point cloud creation. Prior to image matching, we removed blurry images. Based on the dense point cloud, Digital Elevation Models (DEM) were produced. The orthomosaics were created by projecting the single image frames on the DEM. Georeferencing of the dataset was performed automatically in Agisoft Photoscan based on the GNSS trajectories logged during the UAV image flight. The estimated total error of the georeferencing did not exceed 3 m. The orthomosaics were exported at spatial resolutions corresponding to the average pixel size of the single image frames (see Tab. 1 for a summary of the orthoimagery). For compatibility the DEM were resampled to the same spatial resolution. A tabular summary of the UAV data is given in supplementary information 1.

The reference data required to train and to validate the CNN-based models were derived using GIS-based visual image interpretation and delineation of the target classes. This approach ensured that the reference data and orthoimagery were in direct spatial correspondence (no geolocation mismatches), as they emerged from the same data and perspective (nadir imagery). The visual interpretation was based on knowledge gained through ground-based sampling. For the case studies on plant invasions in Chile and on vegetation succession in the Mueller Glacier forelands, geotagged photographs, which were acquired during the field campaign, were used to aid the image interpretation. The delineation of tree crowns in the Waitutu Tree Species case study was aided and checked using full georeferenced inventory data (Coomes et al. 2005). The polygons created by visual interpretation were cross-checked by at least one other interpreter (Table 1).

Training and applying the CNN-based Deep Learning model

In each case study, we derived spatially continuous estimates of the target classes using CNN models applied to

Table 1. Summary of the case studies, vegetation classes and species of interest.

Case study	Target area [ha]	No of flights	Orthoimage resolution [cm]	Target variable/species
Vegetation succession	327.2	7	5	1. Pioneer class *Racomitrium* spp., *Stellaria gracilenta* 2. Early intermediate class *Coriaria angustissima, Gaultheria crassa, Holcus lanatus* 3. Later intermediate class *Acaena fissistipula, Leucopogon fraseri, Muehlenbeckia axillaris* 4. Late succession *Podocarpus* spp., *Phyllocladus alpinus*
Plant invasion	58.2	4	3	*1. Ulex europaeus*
	93.5	4	3	*2. Pinus radiata*
Tree species	4.3	3	3	1. *Metrosideros umbellata* 2. *Dacrydium cupressinum*

gridded tiles of the imagery (compare Fig. 1). The CNN models were trained and validated with image tiles and visually delineated target canopies. The size of the image tiles was varied from 2 to 5 m to test the trade-off between spatial detail and accuracy of the mapping output. In the second test, we compared the model performance of CNN models incorporating the 3D information (DEM) with models excluding 3D information.

The image tiles used for training and validation of the CNN models were extracted from a regular grid of x- and y-coordinates within the extent of the orthoimages. We set the grid spacing to 5 m to ensure that the image tiles (maximum size of 5 m) did not overlap. At each selected grid position, rectangular tiles of the orthoimagery and DEM were extracted, with edges measuring 2, 3, 4 and 5 m, respectively (Fig. 2). From each tile we extracted the cover [%] of the target class using the digitized polygons of that class (Fig. 2). The DEM values were normalized (0–255 grey values) to remove effects of different ground elevations between the study areas. Using this procedure we randomly sampled 4000 image tiles for each target canopy (together with the reference covers [%]). For the Waitutu Tree Species case study, we used 2500 image tiles due to the smaller spatial extent of UAV data available. For each case study, the image tiles were split into training (66.6%) and validation sets (33.3%).

The CNN processing chain was implemented through an R interface (version 2.2.4, R Core Team 2018) using the Keras deep learning API and the TensorFlow backend (Chollet and Allaire 2019). We trained the models on a local workstation using the CUDA environment (GPU-based processing) and a NVIDIA graphics card (GeForce GTX 980 Ti). We setup the CNN using Depthwise Seperable Convolutions (also known as Xception, Chollet 2017). Depthwise Seperable Convolutions perform convolutions for each channel separately (e.g. the three RGB channels), and in the final layer (the 1×1 convolution) the output is merged. This procedure results in a less complex network with fewer parameters than traditional CNN and is thus more robust for smaller datasets and also computationally more efficient (Chollet 2017). In view of the comparably small size of the reference dataset, we used a relatively streamlined network architecture to avoid overfitting (Fig. 3). The architecture involved eight layers, including the application of six subsequent 3×3 convolutions, each of them including a rectified linear unit (ReLu). After the first two subsequent 3×3 convolutions (layer 2 and 4), a maximum pooling with stride 2 reduced the feature map size by 50%. The number of feature channels was doubled with each maximum pooling iteration. After the last two 3×3 convolutions (layer 6) a global average pooling was used to derive a 1 dimensional layer (layer 7). The last layer of the neural network was based on a sigmoid activation function limiting the range of possible predictions between 0 and 1 (corresponding to 0 and 100%, layer 8). To facilitate a direct comparison, we applied the same CNN architecture for the different case studies (target canopies) and tests (different tile sizes, including and excluding 3D information).

We trained the models using the RMSprob optimizer and the mean squared error (MSE) loss function. To compensate for the limited size of training samples and to increase model robustness, we applied data augmentation during model training. Data augmentation inflated the number of frames by randomly rotating (in 40 degree steps), shearing (0-0.2 radians), shifting (0–15%) and flipping the image frames horizontally. The CNN models were trained in 20 epochs, using a batch size of 32 and 1563 steps (~50 000 steps in total). 20 epochs were

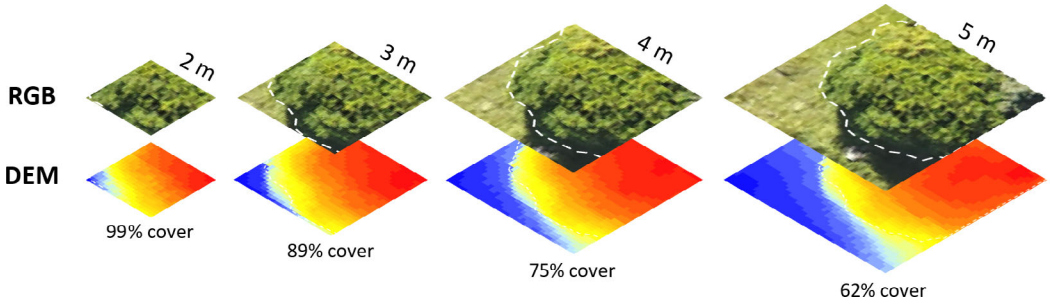

Figure 2. Schematic representation of the input tiles at different sizes (2–5 m) used for training and applying the CNN models. Each tile consists of RGB data, DEM information scaled from 0 (blue, low) to 1 (red, high) and reference data quantifying the per cent cover of the target canopy (polygon boundary indicated by white dashed line).

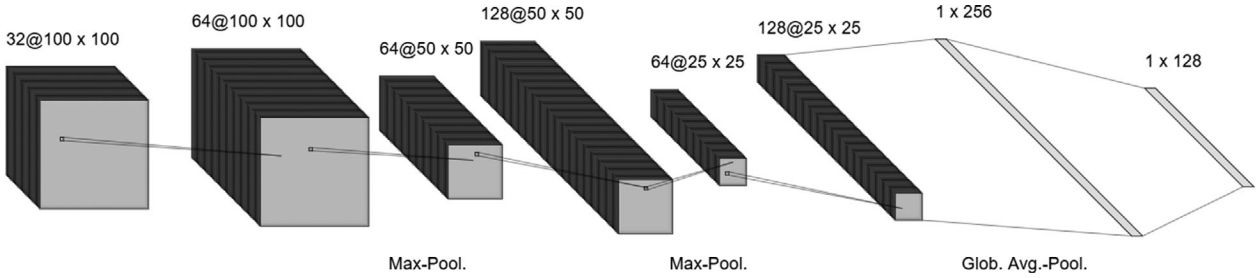

Figure 3. Scheme showing the analysis of tiles in the Convolutional Neural Network (CNN) models used in this study. In the first step, the input tiles (here with a size of 100 × 100 pixels) are analyzed using 32 filters (convolutions). In the subsequent steps, the spatial dimensions of the 32 resulting feature maps are reduced (maximum pooling operations) and further convolutions are applied with increasing numbers of filters with an output of up to 256 feature maps after global average pooling. The actual result, that is % cover, is determined in the last layer (see main text for details).

Table 2. Hyper-parameters used for training the CNN models.

Hyper-parameter	Value
No. of input channels	3 (RGB) or 4 (RGB + 3D)
Filter size	1*32, 3*64, 2*128 (see Fig. 3)
Pooling size	2*2
Convolution kernel size	3
Number of epochs	20
Batch size	32
Steps per epoch	1563 (50 000 in total)
Optimizer	RMSprob (lr = 0.001, rho = 0.9, decay = 0.9)
Activation function	ReLU/sigmoid (last layer)

selected as all models converged before the 20th epoch (usually between the 8th and 14th epoch). The model weights for each epoch were only adjusted if the loss (MSE) was lower than that of the prior epoch. The accuracy of the final models was reported by calculating the R^2 and the RMSE based on the independent validation data. A summary of the hyper-parameters used for the CNN training is given in Table 2.

We derived spatially continuous estimates for each case study by applying the models to tiles of a continuous grid covering the respective orthoimage (compare Fig. 1). The tile sizes corresponded to the tile sizes used for training.

Results

The CNN regression models for the different case studies had R^2 values between 0.57 and 0.85. We found no clear differences in the predictive accuracy of the CNN approach among the different case studies (Figs. 4–6). The resulting maps (Figs. 4–6) show similar patterns for the predictions at different tile sizes (2–5 m). For all case studies, decreased noise at larger tile sizes corresponded with increased prediction accuracy with increasing tile size (larger R^2 and smaller RMSE, respectively). R^2 values increased by 0.12–0.21 from 2 m to 5 m predictions and RMSE values decreased by 4–10% cover.

Table 3. Comparison between CNN model accuracies (RMSE [% *0.01]) with and without photogrammetric 3D information.

	2 m		3 m		4 m		5 m	
3D information included	No	Yes	No	Yes	No	Yes	No	Yes
Vegetation succession								
Pioneer	0.066	0.064	0.051	0.049	0.040	0.042	0.045	0.043
Early Intermediate	0.134	0.137	0.134	0.135	0.106	0.108	0.106	0.096
Late Intermediate	0.144	0.151	0.123	0.121	0.114	0.109	0.103	0.104
Late successional	0.226	0.220	0.197	0.178	0.157	0.149	0.139	0.138
Plant invasion								
Pinus radiata	0.231	0.239	0.191	0.182	0.155	0.158	0.141	0.139
Ulex europaeus	0.182	0.215	0.181	0.167	0.122	0.121	0.116	0.114
Tree species								
Metrosideros umbellata	0.181	0.155	0.153	0.157	0.129	0.148	0.100	0.099
Dacrydium cupressinum	0.170	0.182	0.136	0.146	0.121	0.132	0.111	0.112

We assessed the contribution of including photogrammetric 3D information by comparing the predictive accuracy (RMSE) of models that were trained with and without elevation data (DEM). Results of the three case studies indicate only minor differences in R^2 and RMSE between models with and without 3D information (Tab1e 3). There is no clear trend in the contribution of photogrammetric 3D information to model accuracy with changing tile size.

Discussion

A primary strength of the approach presented is that it exploits inexpensive, consumer-grade hardware to map the cover of species or vegetation types on continuous scales using ordinary RGB imagery. Mapping continuous values of cover instead of discrete classes (see e.g. Hu et al. 2015; Wagner et al. 2019) accounts for situations where classes (species or plant communities) are mixed within pixels or tiles, respectively. This approach is particularly applicable for landscapes featuring smooth transitions among species or vegetation types and enables co-occurrences of several vegetation classes to be revealed through separate prediction layers. Moreover, the presented CNN-based regression approach is not only applicable for cover fractions of classes, such as species or vegetation types, but also directly applicable for purely continuous metrics.

Performance of the CNN approach in the three case studies

Are CNN regression models trained and applied on image tiles accurate toward common mapping tasks? Overall, the results for all case studies were very accurate. The accuracies presented for mapping the target canopies are comparable or surpass those of previous studies combining UAV-based hyperspectral data and texture metrics with pixel-based approaches on similar datasets (Kattenborn et al. 2019; Lopatin et al. 2019) and results obtained by other authors for different datasets (Lisein et al. 2015; Sankey et al. 2017; Cao et al. 2018). This confirms that CNN algorithms combined with UAV-based RGB imagery are of high value for vegetation mapping.

According to our observations, the accuracy of the CNN-based mapping is determined by characteristic textural and structural differences of the target objects compared to the surrounding vegetation. In the case of the Waitutu Trees Species case study, *Metrosideros umbellata* featured a very heterogeneous reflectance, with flowering, tree age, nutrient supply and various stressors (e.g. soil drainage) as likely reasons. Its canopy structure tended to resemble the surrounding vegetation and the mapping accuracy was accordingly lower. *Dacrydium cupressinum*, in contrast, had a comparably distinct and homogeneous appearance resulting in accurate cover predictions. Similar high accuracies were found for mapping invasions of *Ulex europaeus* in Chile. Here, the combination of yellow blossom and distinct branching patterns formed a clear contrast to the native vegetation matrix at the site.

The high predictive accuracies we obtained using CNN algorithms demonstrate that UAV-based RGB imagery, featuring a low spectral resolution and high spatial resolution, can be very useful to map vegetation types and properties if spatial context is explicitly considered. Where spectral discernibility is low, spatial pattern may be the only key to successful classification. This suggests great potential for low-cost and off-the-shelf UAV platforms, which have become increasingly available and user-friendly during recent years (Colomina and Molina 2014). At the same time, it can be assumed that combining high spatial resolution sensors with high spectral resolution (multi- or hyperspectral data) will provide even more accurate results (Sankey et al. 2017 or Kattenborn

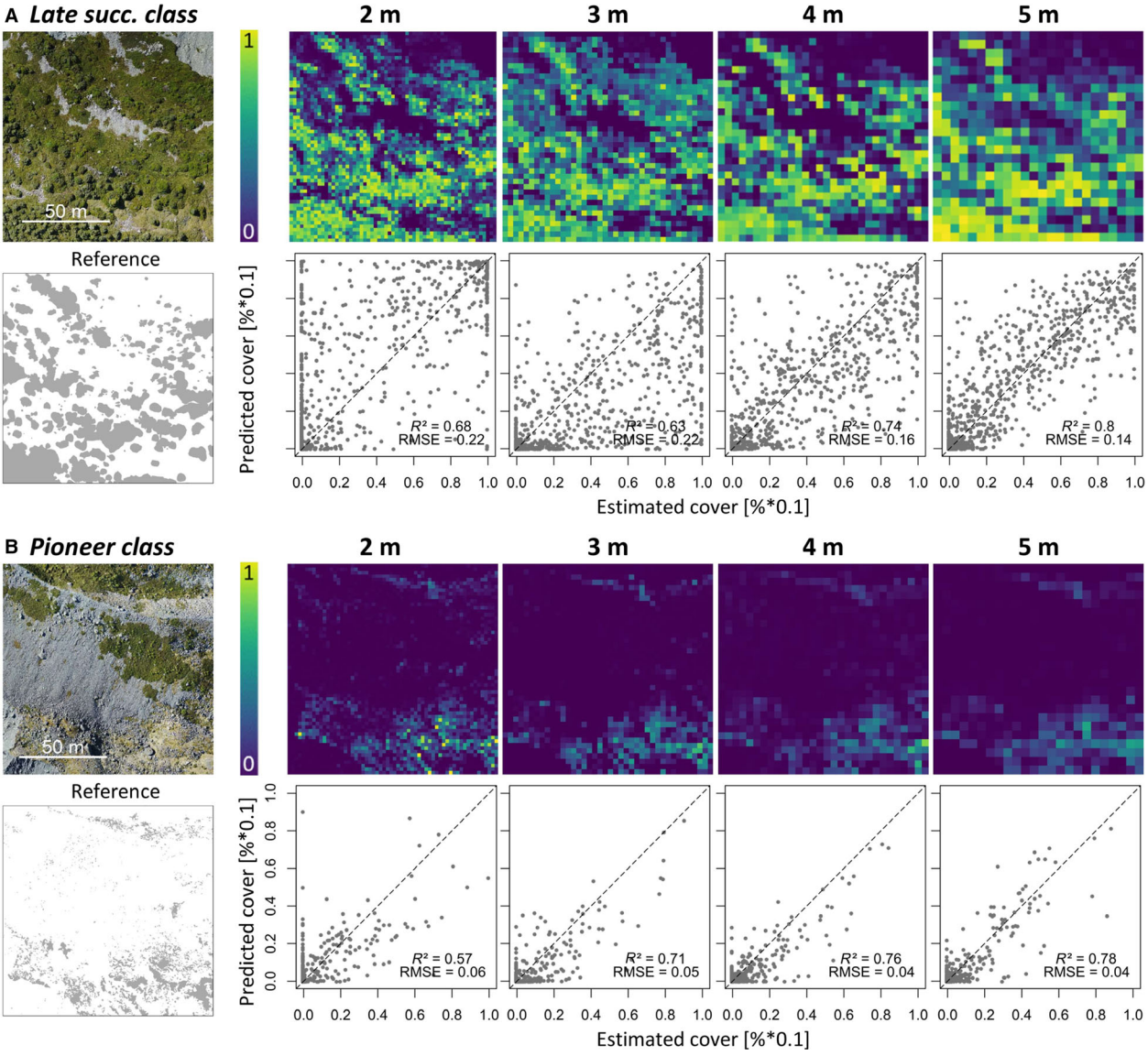

Figure 4. Case study on vegetation succession in the foreland of the Mueller Glacier, New Zealand: RGB imagery, reference data (left) and results (right) for subsets of the (A) late successional class and (B) the pioneer class. Scatterplots depict predictions versus validation data for each tile size. Subsets and predictive accuracy are shown separately for each tile size considered (2–5 m).

et al. 2019), whereas the most accurate results are likely to be obtained by applying such an approach using time-series data (Lisein et al. 2015).

The trade-off of between spatial detail and predictive accuracy

What is the 'right' tile size for mapping the target variable? Our results indicate that the overall spatial patterns of the target canopies are preserved at all spatial resolutions considered (2–5 m), but the amount of noise increases with finer spatial resolution (Figs. 4–6). Accordingly, we found that predictive accuracy increases with increasing tile size. This can be explained, as larger tiles include more spatial information making it more likely that characteristic diagnostic features, such as branching patterns or the canopy shape, are detected. An additional influence may be inaccuracies in the delineated reference data used for validation, which would have a lower impact at coarser spatial resolutions. Overall, the spatial resolutions tested here are high in comparison to other studies, demonstrating the potential for vegetation

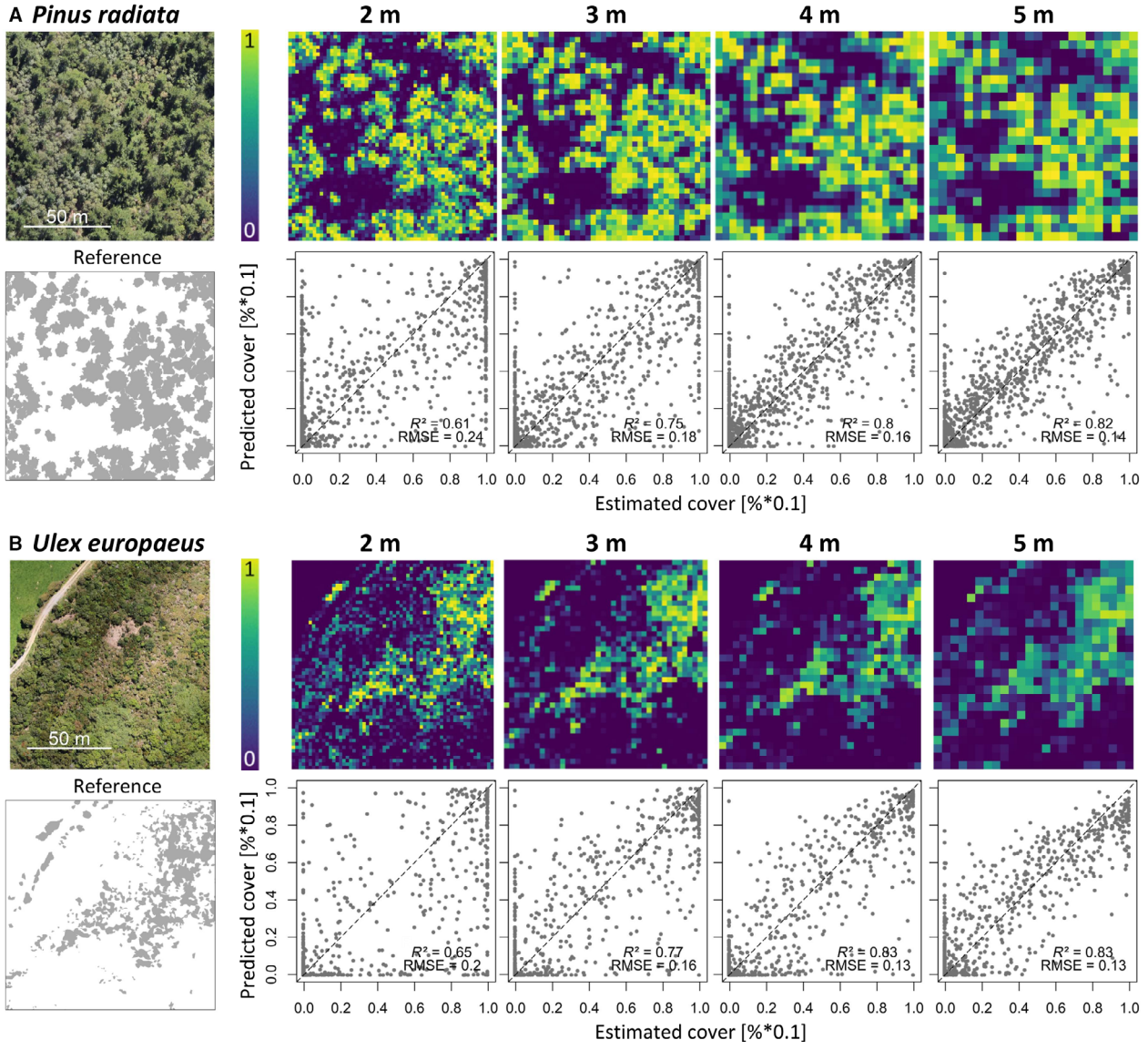

Figure 5. Case study on the mapping of two woody invasive species in Central Chile: RGB imagery, reference data (left) and results (right) for subsets of (A) *Pinus radiata* and (B) *Ulex europaeus*. Scatterplots depict predictions versus validation data for each tile size. Subsets and predictive accuracy are shown separately for each tile size considered (2–5 m).

mapping tasks where high spatial detail matters, for example, for tracking early plant invasions (Cao et al. 2018; Kattenborn et al. 2019), mapping small scale vegetation patterns (Husson et al. 2014; Malenovský et al. 2017; Eichel 2019) or mapping occurrences of individuals within complex vegetation.

The contribution of photogrammetric 3D information

How large is the added value of including 3D information on the canopy structure derived from the SfM processing chain? Our analysis revealed that 3D information did not clearly improve the mapping accuracy. The 3D information on the canopy structure might be redundant as it is already indirectly visible in the orthoimagery through cast shadows and illumination differences within the canopy. Yet, previous studies demonstrated that photogrammetric 3D information can be of high value for species classification or trait retrieval (Kattenborn et al. 2014; Fraser et al. 2016; Alonzo et al. 2018; Lopatin et al. 2019). Hence we suggest that the value of 3D information should be further tested, since it is already available after the photogrammetric processing at no additional cost. It might

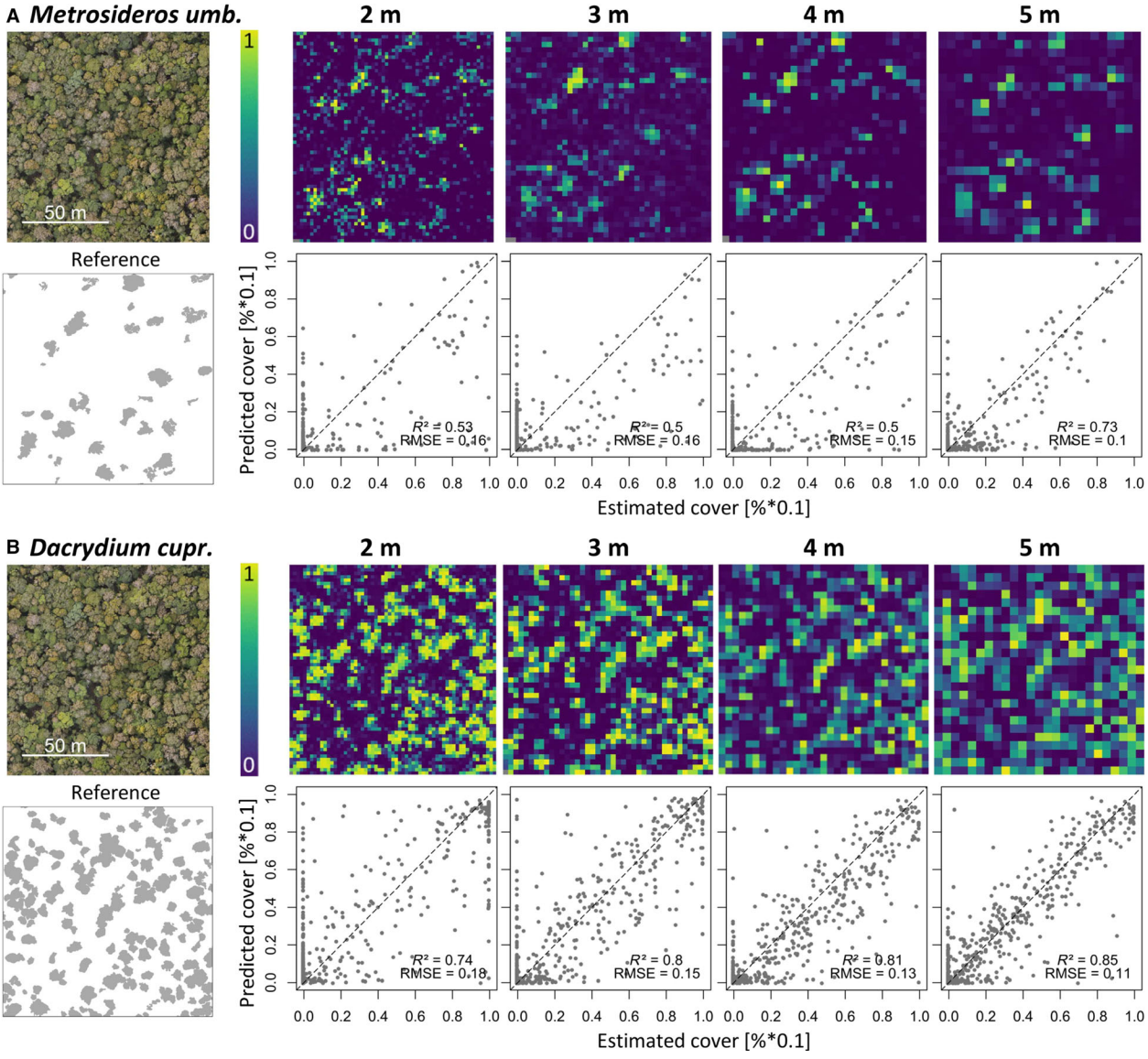

Figure 6. Case study on tree species mapping in a structurally complex primary forest in Waitutu, New Zealand: RGB imagery, reference data (left) and results (right) for subsets of (A) *Metrosideros umbellata* and (B) *Dacrydium cupressinum*. Scatterplots depict predictions versus validation data for each tile size. Subsets and predictive accuracy are shown separately for each tile size considered (2–5 m).

be especially useful in situations with little presence of cast shadows and illumination effects, as where sun angles are high or light conditions are diffuse. Furthermore, the value of 3D information might be in particular useful if the target class or quantity is directly related to canopy height. The value of 3D information may be further increased when normalizing the relative heights to absolute canopy heights by incorporating a Digital Terrain Model (not available for all datasets tested here). A promising way to include photogrammetric information are algorithms that use point clouds instead of elevation models in raster format (e.g. *PointNet*, Garcia-Garcia et al. 2016). Another alternative may be the combined use of orthoimagery and LiDAR data, as LiDAR data have been proven to be a powerful tool to assess geometric vegetation characteristics (Wallace et al. 2012, Sankey et al. 2017). Combining multiple data sources, however, can also be challenging as this requires accurate geometric alignment.

Limitations and practical considerations

We used a straight forward CNN architecture (Fig. 3) to avoid the overfitting that can result from small sample

sizes. With increased availability of high quality data, more sophisticated CNN architectures with more convolutional layers and filters might perform better. Additional strategies such as the use of pretrained libraries (Schwarz et al. 2015) or freezing (Nogueira et al. 2017) can be used to boost the performance of such models. Depending on the target feature and the data, a different type of CNN might be applicable. Here, we implemented a CNN-based regression model to predict a continuous metric on tiles. Given that the target variable is a discrete class (e.g. binary absences/presences of a species) and the available data are of high quality and resolution, a semantic segmentation approach may be applicable (Kattenborn et al. 2019). Semantic segmentation architectures, such as the *U-net* (Ronneberger et al. 2016) or the *DenseNet* (Jégou et al. 2017) can be advantageous, since they enable the extent of the target class to be predicted at the original resolution of the input imagery (Kattenborn et al. 2019). If the objective is to identify single occurrences of a class (e.g. individual trees) instance segmentation algorithms, such as *Mask R-CNN* (He et al. 2017), can be used. However, the latter requires that the individual are clearly definable in the orthoimagery.

Further, we found the regression approach advantageous when the image quality or resolution does not allow for explicit segmentation of the objects. Such a regression approach is also applicable for mapping continuous metrics, such as plant diseases (Sladojevic et al. 2016) and structure-related plant and community traits such as crown width, canopy gap fraction or Leaf Area Index (Getzin et al. 2014; Chianucci et al. 2016; Panagiotidis et al. 2017; Tian et al. 2017).

For training and validation, the workflow presented here included the visual delineation of the target class from the UAV-based RGB imagery. The delineation and image interpretation was carefully checked using photographs and geocoded field data. Still, inaccuracies during the delineation process are inevitable and therefore might have affected the training and thus predictive accuracy of the CNN models. We noted cases where target canopies were overlooked during the delineation resulting in 'false' prediction that were in reality true (see Fig. 7 for an example). Validation based on holdouts of the visually delineated reference data can be affected by such inaccuracies, implying that a model might perform better than indicated by the reference data (Kattenborn et al. 2019). Despite this, we believe that, if the target class can be identified in the imagery, training CNN models using visual delineation has the advantage that (1) the visual delineation in orthomosaics is spatially explicit, (2) it is not affected by geolocation alignment issues, (3) it is derived from the same perspective (nadir) facilitating the generation of statistical links, (4) it allows many observations to be acquired and (5) overcomes barriers and biases resulting from inaccessible sites (Lunetta et al. 1991; Lepš and Hadincová 1992; Valbuena et al. 2010; Kaartinen et al. 2015; Leitão et al. 2018).

The resolution of the datasets acquired within the case studies presented here ranged from 3 cm to 5 cm. However, important features of plant canopies might only be identifiable at higher spatial resolutions, for example, leaf or needle shape and arrangement, branching architecture, presence and form of flowers or seeds (Müllerová et al. 2013; Ghazi et al. 2017; Wäldchen et al. 2018). We believe the predictive accuracy of the approach could potentially be enhanced by increasing the spatial resolution. However, traditionally there has been a trade-off between spatial resolution and area coverage of remote sensing imagery. With time, we expect technological advances will continue to increase spatial detail without compromising spatial extent.

The spatial extents at which UAV data are commonly acquired for vegetation mapping is generally limited to not more than a few hectares or square kilometers. This could be considered a substantial limitation of UAV-applications for vegetation mapping, as many research

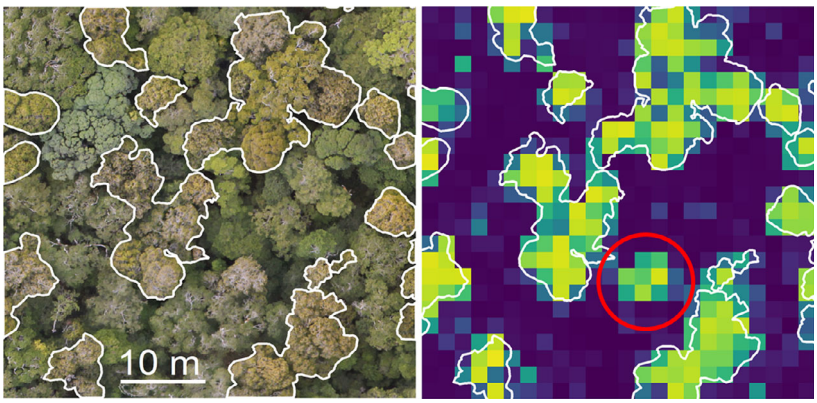

Figure 7. Map extracts of the RGB imagery (left) and predictions with 2 m tile size for *Dacrydium cupressinum* (right). The red circle marks plausible cover values for an overlooked Dacrydium cupressinum tree, suggesting that the CNN models may be even more accurate than indicated by the delineated reference data (white polygons).

questions are centered at regional to global and not local scales. However, UAV mapping products can be of high value for large-scale analyses, as they can contribute large training data for satellite-based mapping procedures (Kattenborn et al. 2019; Riihimäki et al. 2019).

Some authors consider that pattern recognition and neural networks are limited by the computational demands to train the models (Siéler et al. 2010; Culbert et al. 2012; Nogueira et al. 2017). Yet, recent developments in the computer industry have created multiple possibilities for an efficient application of image analysis (in particular for CNN algorithms) and neural networks such as graphic-processing unit (GPU)-based computations using the parallel computing platform CUDA (Nvidia). Here we used CUDA with a medium priced GPU (GeForce GTX 980 Ti), on which training of the CNN architecture with 5 m window size (27889 pixels) and 1562 steps and 20 epochs lasted approximately 2.5 h. Once the model is trained, its application on further images using the tile approach is computationally very efficient and takes only seconds.

That the application of a trained CNN model is very fast and a large share of the community has access to UAV with comparable RGB systems opens up new possibilities and advantages of sharing pretrained CNN models. A potential future direction is the establishment of databases that offer CNN models for various classification and regression tasks. For common tasks (e.g. fractional vegetation cover mapping of common invasive species) such a framework would be highly valuable for the community. Databases would also increase the quantity of training data, as CNN and neural networks require comprehensiveness of training data. More specifically such databases could either consist of labelled UAV image frames or orthoimagery together with delineated polygons of the target species.

Conclusion

Convolutional Neural Networks regression models are a powerful tool to harness high resolution data acquired from UAV to predict vegetation patterns. In many cases, where spectral information is scarce or does not help in identifying the given vegetation or species, spatial patterns can be essential. This cutting-edge technique, in concert with hyperspectral remote sensing in a multi-temporal setting will pave the way toward unprecedented accuracy in future vegetation mapping. At the same time, CNN alone will revolutionize the way we use high resolution spatial imagery. The high predictive accuracies obtained in our case studies using low cost RGB sensors highlights the potential application for a wide range of users. We conclude that combining UAV and CNN will provide ground-breaking opportunities for applied vegetation mapping. Moreover, satellite images are already approaching the high spatial resolutions relevant for the methods tested in this contribution, opening up a wealth of further applications.

Acknowledgements

Parts of this study were conducted within the SaMovar project (Satellite-based Monitoring of invasive species in central Chile), a collaborative project of Karlsruhe Institute of Technology (KIT) and Technical University of Berlin. It was funded by the German National Space Agency DLR (Deutsches Zentrum für Luft- und Raumfahrt e.V.) on behalf of the German Federal Ministry of Economy and Technology based on the Bundestag resolution 50EE1535 and 50EE1536. J. Eichel appreciates funding by the Hanna Bremer Foundation for fieldwork in Mueller glacier forelands. The Waitutu case study was part of a project funded by the Catalyst: Leaders program financed by the New Zealand Ministry of Business, Innovation and Employment (MBIE) and administered by the Royal Society of New Zealand and the MBIE Strategic Science Investment Fund to Manaaki Whenua-Landcare Research. Many thanks to Timo Schmid, Felix Schiefer, Denis Debroize and Jorge Petri for assistance in image interpretation. Furthermore, we thank David Hedding, Stefan Winkler and Daniel Draebing for assistance during UAV and vegetation surveys and the Department of Conservation for granting permission for research and drone use at the New Zealand test sites.

Conflict of Interest

The authors declare no competing interest.

References

Alonzo, M., H. E. Andersen, D. Morton, and B. Cook. 2018. Quantifying boreal forest structure and composition using UAV structure from motion. *Forests* **9**, 119.

Angermueller, C., T. Pärnamaa, L. Parts, and O. Stegle. 2016. Deep learning for computational biology. *Mol. Syst. Biol.* **12**, 878.

Aplin, P. 2006. On scales and dynamics in observing the environment. *Int. J. Remote Sens.* **27**, 2123–2140.

Brodu, N., and D. Lague. 2012. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. *ISPRS J. Photogramm. Remote Sens.* **68**, 121–134.

Bustamante, R. O., and J. A. Simonetti. 2005. Is *Pinus radiata* invading the native vegetation in central Chile? Demographic responses in a fragmented forest. *Biol. Invasions* **7**, 243–249.

Cadieu, C. F., H. Hong, D. L. Yamins, N. Pinto, D. Ardila, E. A. Solomon, et al. 2014. Deep neural networks rival the

representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* **10**, e1003963.

Cao, J., W. Leng, K. Liu, L. Liu, Z. He, and Y. Zhu. 2018. Object-Based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. *Remote Sens.* **10**, 89. https://doi.org/10.3390/rs10010089.

Chianucci, F., L. Disperati, D. Guzzi, D. Bianchini, V. Nardino, C. Lastri, et al. 2016. Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing UAV. *Int. J. Appl. Earth Obs. Geoinf.* **47**, 60–68.

Chollet, F. 2017. Xception: deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1251–1258.

Chollet, F., and J. J. Allaire. 2019. Keras: R interface to Keras. R package version 2.2.0. https://CRAN.R-project.org/package=keras

Clapp, R. A. 1995. The unnatural history of the Monterey pine. *Geogr. Rev.* 1–19.

Colomina, I., and P. Molina. 2014. Unmanned aerial systems for photogrammetry and remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* **92**, 79–97.

Coomes, D. A., R. B. Allen, W. A. Bentley, L. E. Burrows, C. D. Canham, L. Fagan, et al. 2005. The hare, the tortoise and the crocodile: the ecology of angiosperm dominance, conifer persistence and fern filtering. *J. Ecol.* **93**, 918–935.

Culbert, P. D., V. C. Radeloff, V. St-Louis, C. H. Flather, C. D. Rittenhouse, T. P. Albright, et al. 2012. Modeling broad-scale patterns of avian species richness across the Midwestern United States with measures of satellite image texture. *Remote Sens. Environ.* **118**, 140–150.

Eichel, J. 2019. Vegetation succession and biogeomorphic interactions in glacier forelands. Pp. 327–349 *inGeomorphology of proglacial systems*. Springer, Cham.

Foody, G. M., N. A. Campbell, N. M. Trodd, and T. F. Wood. 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogramm. Eng. Remote Sensing* **58**, 1335–1341.

Fraser, R. H., I. Olthof, T. C. Lantz, and C. Schmitt. 2016. UAV photogrammetry for mapping vegetation in the low-Arctic. *Arctic Sci.* **2**, 79–102.

Fritz, A., T. Kattenborn, and B. Koch. 2013. UAV-based photogrammetric point clouds—Tree stem mapping in open stands in comparison to terrestrial laser scanner point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **40**, 141–146.

Garcia-Garcia, A., F. Gomez-Donoso, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla, and J. Azorin-Lopez. 2016. Pointnet: a 3D convolutional neural network for real-time object class recognition. Pp. 1578–1584*in* 2016 International Joint Conference on Neural Networks (IJCNN), IEEE.

Gellally, A. F. 1982. Lichenometry as a relative-age dating method in Mount Cook National Park, New Zealand. *New Zeal. J. Bot.* **20**, 343–353.

Getzin, S., K. Wiegand, and I. Schöning. 2012. Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods Ecol. Evol.* **3**, 397–404.

Getzin, S., R. Nuske, and K. Wiegand. 2014. Using unmanned aerial vehicles (UAV) to quantify spatial gap patterns in forests. *Remote Sens.* **6**, 6988–7004.

Ghazi, M. M., B. Yanikoglu, and E. Aptoula. 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing.* **235**, 228–235.

Guerrero, P. C., and R. O. Bustamante. 2007. Can native tree species regenerate in Pinus radiata plantations in Chile?: evidence from field and laboratory experiments. *For. Ecol. Manage.* **253**, 97–102.

Haralick, R. M. 1979. Statistical and structural approaches to texture. *Proc. IEEE* **67**, 786–804.

He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask r-cnnIn Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Hsieh, P. F., L. C. Lee, and N. Y. Chen. 2001. Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *IEEE Trans. Geosci. Remote Sens.* **39**, 2657–2663.

Hu, F., G. S. Xia, J. Hu, and L. Zhang. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **7**, 14680–14707.

Huang, B., K. Lu, N. Audeberr, A. Khalel, Y. Tarabalka, J. Malof, et al. (2018). Large-scale semantic classification: outcome of the first year of inria aerial image labeling benchmark. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 6947-6950). IEEE.

Hubel, D. H., and T. N. Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154.

Husson, E., O. Hagner, and F. Ecke. 2014. Unmanned aircraft systems help to map aquatic vegetation. *Appl. Veg. Sci.* **17**, 567–577.

Jégou, S., M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. 2017. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. Pp. 11–19.

Joly, A., P. Bonnet, H. Goëau, J. Barbe, S. Selmi, J. Champ, et al. 2016. A look inside the Pl@ ntNet experience. *Multimedia Syst.* **22**, 751–766.

Kaartinen, H., J. Hyyppä, M. Vastaranta, A. Kukko, A. Jaakkola, X. Yu, et al. 2015. Accuracy of kinematic positioning using global satellite navigation systems under forest canopies. *Forests* **6**, 3218–3236.

Kattenborn, T., M. Sperlich, K. Bataua, and B. Koch. 2014. Automatic single palm tree detection in plantations using

UAV-based photogrammetric point clouds. *Int. Arch. Photogramm.* **40**, 139–144.

Kattenborn, T., J. Hernandez, J. Lopatin, G. Kattenborn, and F. E. Fassnacht. 2018. Pilot study on the retrieval of DBH and Diameter Distribution of deciduous forest stands using cast shadows in UAV-based orthomosaics. *ISPRS Ann. Photogramm. Rem. Sens. Spatial Inform. Sci.* **1**, 93–96.

Kattenborn, T., J. Eichel, and F. E. Fassnacht.2019. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific reports* **9**(1), 1–9.

Kattenborn, T., J. Lopatin, M. Förster, A. C. Braun, and F. E. Fassnacht. 2019. 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.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097–1105).

Leitão, P. J., M. Schwieder, F. Pötzschner, J. R. Pinto, A. M. Teixeira, F. Pedroni, et al. 2018. From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. *Ecosphere* **9**, e02298.

Lepš, J., and V. Hadincová. 1992. How reliable are our vegetation analyses? *J. Veg. Sci.* **3**, 119–124.

Lisein, J., A. Michez, H. Claessens, and P. Lejeune. 2015. Discrimination of deciduous tree species from time series of unmanned aerial system imagery. *PLoS ONE* **10**, e0141006.

Lopatin, J., F. E. Fassnacht, T. Kattenborn, and S. Schmidtlein. 2017. Mapping plant species in mixed grassland communities using close range imaging spectroscopy. *Remote Sens. Environ.* **201**, 12–23.

Lopatin, J., K. Dolos, T. Kattenborn, and F. E. Fassnacht. 2018. How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. *Rem. Sens. Ecol. Conserv.* 1–16.

Lopatin, J., T. Kattenborn, M. Galleguillos, J. F. Perez-Quezada, and S. Schmidtlein. 2019. Using aboveground vegetation attributes as proxies for mapping peatland belowground carbon stocks. *Remote Sens. Environ.* **231**, 111217.

Lu, B., and Y. He. 2017. Species classification using Unmanned Aerial Vehicle (UAV)-acquired high spatial resolution imagery in a heterogeneous grassland. *ISPRS J. Photogramm. Remote Sens.* **128**, 73–85.

Lunetta, R. S., R. G. Congalton, L. K. Fenstermaker, J. R. Jensen, K. C. Mcgwire, and L. R. Tinney. 1991. Remote-sensing and geographic information-system data integration – Error sources and research issues. *Photogramm. Eng. Remote Sensing* **57**, 677–687.

Luscier, J. D., W. L. Thompson, J. M. Wilson, B. E. Gorham, and L. D. Dragut. 2006. Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots. *Front. Ecol. Environ.* **4**, 408–413.

Malenovský, Z., A. Lucieer, D. H. King, J. D. Turnbull, and S. A. Robinson. 2017. Unmanned aircraft system advances health mapping of fragile polar vegetation. *Methods Ecol. Evol.* **8**, 1842–1857.

Michez, A., H. Piégay, L. Jonathan, H. Claessens, and P. Lejeune. 2016. Mapping of riparian invasive species with supervised classification of Unmanned Aerial System (UAS) imagery. *Int. J. Appl. Earth Obs. Geoinf.* **44**, 88–94.

Müllerová, J., J. Pergl, and P. Pyšek. 2013. Remote sensing as a tool for monitoring plant invasions: testing the effects of data resolution and image classification approach on the detection of a model plant species Heracleum mantegazzianum (giant hogweed). *Int. J. Appl. Earth Obs. Geoinf.* **25**, 55–65.

Müllerová, J., J. Brůna, T. Bartaloš, P. Dvořák, M. Vítková, and P. Pyšek. 2017. Timing is important: unmanned aircraft vs. satellite imagery in plant invasion monitoring. Frontiers. *Plant Sci.* **8**, 887.

Nogueira, K., O. A. B. Penatti, and J. A. dos Santos. 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recogn.* **61**, 539–556.

Norambuena, H., S. Escobar, and F. Rodriguez. 2000. The Biocontrol of Gorse, Ulex europaeus, in Chile: a progress report. In Proc. of the International symposium on biological control of weeds, Montana State University, Bozeman, Montana, USA, 955–961.

Panagiotidis, D., A. Abdollahnejad, P. Surový, and V. Chiteculo. 2017. Determining tree height and crown diameter from high-resolution UAV imagery. *Int. J. Remote Sens.* **38**, 2392–2410.

Parfitt, R. L., D. J. Ross, D. A. Coomes, S. J. Richardson, M. C. Smale, and R. A. Dahlgren. 2005. N and P in New Zealand soil chronosequences and relationships with foliar N and P. *Biogeochemistry* **75**, 305–328.

R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Riihimäki, H., M. Luoto, and J. Heiskanen. 2019. Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data. *Remote Sens. Environ.* **224**, 119–132.

Rocchini, D., G. M. Foody, H. Nagendra, C. Ricotta, M. Anand, K. S. He, et al. 2013. Uncertainty in ecosystem mapping by remote sensing. *Comput. Geosci.* **50**, 128–135.

Ronneberger, O., P. Fischer, and T. Brox. 2016. U-Net: convolutional networks for biomedical image segmentation.

Sankey, T., J. Donager, J. McVay, and J. B. Sankey. 2017. UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sens. Environ.* **195**, 30–43.

Schmidtlein, S., and J. Sassin. 2004. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sens. Environ.* **92**, 126–138.

Schwarz, M., H. Schulz, and S. Behnke. 2015. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features. In 2015 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1329-1335). IEEE.

Shin, H. C., H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, et al. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imag.* **35**, 1285–1298.

Siéler, L., C. Tanougast, and A. Bouridane. 2010. A scalable and embedded FPGA architecture for efficient computation of grey level co-occurrence matrices and Haralick textures features. *Microprocess. Microsyst.* **34**, 14–24.

Sladojevic, S., M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic. 2016. Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 3289801, 1–11.

Tian, J., L. Wang, X. Li, H. Gong, C. Shi, R. Zhong, et al. 2017. Comparison of UAV and WorldView-2 imagery for mapping leaf area index of mangrove forest. *Int. J. Appl. Earth Obs. Geoinf.* **61**, 22–31.

Ubbens, J. R., and I. Stavness. 2017. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* **8**, 1190.

Valbuena, R., F. Mauro, R. Rodriguez-Solano, and J. A. Manzanera. 2010. Accuracy and precision of GPS receivers under forest canopies in a mountainous environment. *Span. J. Agric. Res.* **8**, 1047–1057.

Vanha-Majamaa, I., M. Salemaa, S. Tuominen, and K. Mikkola. 2000. Digitized photographs in vegetation analysis-a comparison of cover estimates. *Appl. Veg. Sci.* **3**, 89–94.

Wagner, F. H., A. Sanchez, Y. Tarabalka, R. G. Lotte, M. P. Ferreira, M. P. Aidar, et al. 2019. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Rem. Sens. Ecol. Conserv.* **5**, 360–375. https://doi.org/10.1002/rse2.111

Wäldchen, J., and P. Mäder. 2018. Plant species identification using computer vision techniques: a systematic literature review. *Archiv. Comput. Methods Eng.* **25**, 507–543.

Wäldchen, J., M. Rzanny, M. Seeland, and P. Mäder. 2018. Automated plant species identification—Trends and future directions. *PLoS Comput. Biol.* **14**, 1–19.

Wallace, L., A. Lucieer, C. Watson, and D. Turner. 2012. Development of a UAV-LiDAR system with application to forest inventory. *Remote Sens.* **4**, 1519–1543.

Winkler, S., and C. Lambiel. 2018. Age constraints of rock glaciers in the Southern Alps/New Zealand-Exploring their palaeoclimatic potential. *Holocene* **28**, 778–790.

Zarco-Tejada, P. J., M. L. Guillén-Climent, R. Hernández-Clemente, A. Catalina, M. R. González, and P. Martín.2013. Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). *Agr. Forest. Meteorol.* **171**, 281–294.

Zhang, L., X. Huang, B. Huang, and P. Li. 2006. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **44**, 2950–2961.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Summary of the UAV data acquired.
Data S2. Overview on successional classes.
Data S3. Results for the early and late intermediate class of the case study 'Vegetation Succession'.
Data S4. Center coordinates of all map extracts.
Data S5. Overviews of the UAV orthoimagery for each case study.

