Automated mapping and identification of shrub individuals in South Africa's Fynbos biome using drone imagery and deep learning

Nathaniel Allen, Huw Cooksley, Carsten Buchmann, Frank Schurr, Jörn Pagel

Institute of Landscape & Plant Ecology, University of Hohenheim

Abstract

In order to understand population and community dynamics, many ecological studies require comprehensive knowledge of the spatial distribution of individual organisms, but obtaining this data is a time and labor-intensive process. In this study we develop a workflow to automatically determine the species of shrubs of the Proteaceae family in South Africa's Fynbos region from drone-based photogrammetric data. We applied deep learning to segment five species of shrub individuals from the background based on spectral and height information. The spectral-height model achieved an average prediction accuracy of 74.4%, compared to 61.6% when using spectral information alone. Despite the challenge in distinguishing sprawling shrubs from the background, which may be overcome with additional training data, the presented workflow holds promise for the efficient mapping of shrub communities.

1 Introduction

For ecologists, understanding the spatial distribution of individuals in populations and communities within a study area is of critical importance. To obtain the most complete picture, individuals in a given system must be detected, identified, and placed within a spatial context. Traditionally, this mapping is performed by hand via vegetation surveys, a time and labor-intensive process. Improving this process with automation, e.g. with remote sensing techniques such as obtaining data with drones, would greatly increase the scale of future ecological research. Workflows involving drone-obtained data have seen compelling successes such as detecting individual trees[1] and assessing success of wetland restoration[2].

Another technical arena of increasing relevance to ecology is deep learning. Before 2014, there were less than a dozen papers applying deep learning for ecology, but between 2014 and 2019 over 50 were published[3]. Deep learning is a more advanced branch of machine learning, leveraging recent improvements in computational capacity to build classifiers suitable to complex tasks. *Convolutional Neural Networks* (CNNs) are a typical deep learning architecture employed in image-recognition problems. CNNs have been successfully employed at a variety of ecological scales: recognition of wheat spikes[4], describing plankton community composition[5], and counting whale individuals within a region[6]. So far, most species identification applications have focused on using solely image, i.e. spectral, data for classification, yet the incorporation of structural information is suggested to improve identification rates[7].

Proceedings of the 7th bwHPC Symposium (ISBN 978-3-948303-29-7)
Article No. 3; pages 11–16

In this study, we tested the efficacy of a workflow combining drone-based photogrammetric data and deep learning for plant species identification. We did this in the Fynbos biome, which is a key part of a global biodiversity hotspot[8] (the South African Cape Floristic Region) and for which extensive individual-based community maps are available. Specifically, we ask (1) is it possible to reliably identify Proteaceae shrubs using purely spectral information, (2) does inclusion of height information improve species identification?

2 Methods

2.1 Study System & Data

Our study site is located in the Jonaskop mountain range (33° 56' S, 19° 31' E). The site is characteristic of a montane Fynbos biome and has sandy and nutrient poor soils. In this biome, members of the Proteaceae family are key ecosystem members and of considerable interest in research regarding biodiversity and the impacts of climate change[9].

On June 8th 2019, a DJI® Mavic Pro Platinum drone flew an overhead 'grid mission', taking a series of overlapping photos of the site. The drone was flown at 30 meters above the starting point of the mission, and photos were taken with a 70% overlap in both X and Y directions. Photogrammetry, i.e. stitching of photos, was done using AgiSoft's *Metashape* application producing first a 3d point cloud and eventually an orthophoto containing spectral and height information. Guided by hand-mapped ground-truth data obtained in 2011[10], we manually digitized polygons for Proteaceae shrubs on the site, resulting in the following number of individuals: 396 *Leucadendron laureolum*, 342 *L. salignum*, 408 *Protea amplexicaulis*, 88 *P. lorifolia*, and 732 *P. repens* (Fig. 1).

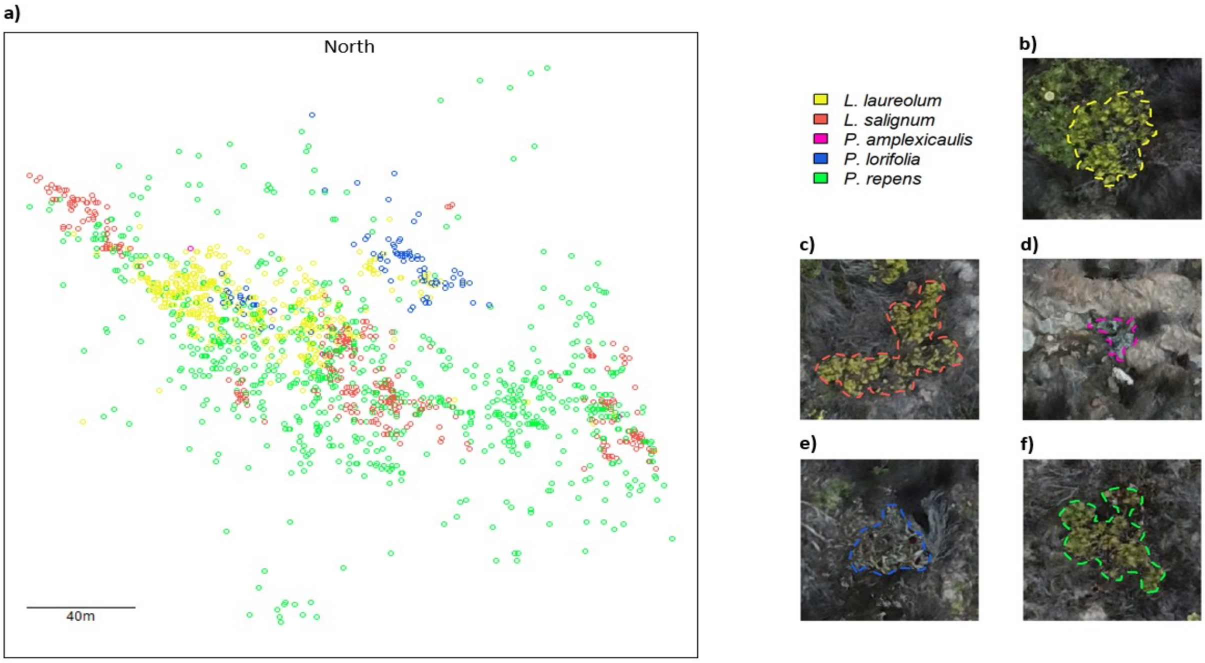

Figure 1: -a) Field-recorded map of the study site produced in 2011 with each circle representing a shrub. b-f) 2x2m subsets of the orthophoto centered around an individual shrub. In both plots color indicates the species identity of an individual.

2.2 Deep Learning

In order to obtain spatially representative information on individual plants, we require a model capable of "semantic segmentation". Semantic segmentation performs pixel-level predictions for classes (i.e. species) and remains one of the more complex topics in computer vision. This complexity arises from the requirement to simultaneously extract class-specific features while also retaining spatial relationships between pixels (normally lost in CNN convolution-pooling layers). Here, we employ a modified version of U-Net, a CNN which uses an encoder-decoder architecture to segment classes from background in images[11]. The model was written in *R* and *python* and uses the *keras* library[12].

CNN model input consists of 3d arrays containing different input channels for cells of a 2d grid. These grids (or tiles) are usually randomly split into train/validation/test data-sets. We selected a tile size of 256 x 256 pixels (2.7 x 2.7 m^2). As we are interested in identification of individuals rather than pixels, we require predictions for a data-set be contiguous, i.e. tiles within a data-set should be adjacent. This gives us tile sets which contain a higher proportion of complete shrubs, rather than cropping shrubs at the tile edges. To meet this criterion, we first divide the site into subsites, each comprised of 5x5 tiles, and semi-randomly sort them into train/validation/test. We then manually filtered subsites out based on data distortions and/or insufficient (< 5%) shrub presence. Using a 80/10/10 ratio we end up with 1044 train, 141 validation, and 141 test tiles.

Our label data consists of segmentation masks, generated from our digitized polygons. Any non-Proteaceae pixel is treated as "background", which ultimately describes over 80% of the pixels in the training data. To mitigate biased learning in an unbalanced data-set, we implemented a "double focal" loss function which usually outperforms the standard "cross-entropy" loss used for multiclassification[13]. We tested two models, one using purely spectral (Hue-Saturation-Value) input channels, and one which included height as an additional channel (HSV+Z). We chose the HSV color space as it is considered more robust when handling variation in light conditions. We then tuned the model by testing multiple combinations of dropout rate and number of filters in the network to determine the appropriate model size for our data.

With the best model selected, our ultimate performance metric is prediction of individual shrubs. As a first step, we ran the predictions produced by our model through a conditional random field (CRF) function[14]. CRF helps clean up prediction noise ("stray" pixels) and sharpen edges of the shrubs. We then dissolved contiguous pixel segments into a polygon of the predicted species. Finally, for each ground-truth polygon in the validation set we evaluate if 33% or more of its area was covered by a predicted polygon of the same class (Fig. 2). In the scope of this study, we did not worry about separating monospecific clusters into individual shrubs.

Tuning of the U-Net models was performed on the BwUniCluster 2.0. Our workload required a JupyterHub instance with a single GPU (1 NVIDIA Tesla V100 32GB), 4-8 CPUs, and 32-64GB of RAM. When training our spectral-height model, our Institute's workstation required 70s/epoch, whereas the bw-hpc instance required only 10s, for a total training time of 10 minutes. This improvement in efficiency greatly increased our capacity to test additional model configurations.

3 Results & Discussion

The spectral-height model performed better than a model using only spectral input. For the spectral-height model, identification rates of individual shrubs ranged from 64% to 95% per species with an average prediction accuracy of 74.4%, whereas the spectral model achieved an average of 61.6% (Table 1). The spectral-height model also (mis-)predicted 52 additional shrubs at locations that were classified as background in the ground-truthing (one *Leucadendron laureolum*, two *L. salignum*, 21 *Protea amplexicaulis*, 21 *P. lorifolia*, and seven *P. repens*).

Investigating the (overhead) perspective of species growth form may hint as to why the accuracy varies so dramatically across species. Of the species studied, *L. laureolum, L. salignum & P. repens* (Fig. 1bcf) could be broadly classified as spectrally vibrant. Our results show that *L. laureolum* is rarely confused with background, but rather with the most common shrub species (*P. repens*). In the case of *L. salignum* there is some confusion with background despite apparently distinct foliage. However, when we compare *L. salignum* to the growth forms of the two species most commonly confused with background, *P. amplexicaulis* and *P. lorifolia* (Fig. 1de), we see that they have something in common. All three species commonly have a sprawling and a rather open canopy. So, it may be that the spectral signals for certain *L. laureolum* and *L. salignum* individuals are strong enough for U-net to recognize them as shrubs, but due to the unbalanced data-set there is a bias for predicting *P. repens*. To test this theory, we plan further missions across additional sites, greatly increasing the amount of training data for all species (as well as including more species).

Table 1: Confusion matrix of model predictions for the validation set. Bold face indicates the spectral-height model, and the normal face the spectral model. The rows represent ground-truth individuals, the columns are the predictions. Each cell is the proportion of observations categorized as that class, those highlighted in blue indicate correct identification.

	predicted					
observed (n)	*Leucadendron laureolum*	*L. salignum*	*Protea amplexicaulis*	*P. lorifolia*	*P. repens*	background
Leucadendron laureolum (48)	**0.73** 0.90	**0.02** -	- -	- -	**0.19** 0.06	**0.06** 0.04
L. salignum (36)	- 0.22	**0.64** 0.36	- -	- -	**0.11** 0.06	**0.25** 0.36
Protea amplexicaulis (74)	- -	- -	**0.70** 0.68	- 0.01	- -	**0.30** 0.31
P. lorifolia (10)	- -	- -	- -	**0.70** 0.20	- -	**0.30** 0.80
P. ***repens***	-	-	-	-	**0.95** 0.94	**0.05** 0.06

Current research recommends the use of geometric information to improve species identification and here we incorporated pixel-level height as model input. Although the results are slightly nuanced (increased confusion of *L. laureolum* with *P. repens),* we see a meaningful improvement in overall accuracy. This indicates that height data, as in the case with spectral data, will be most useful when the features of species (e.g. foliage or height) are significantly distinct from each other and the background. In this study system, that would imply older sites where the shrubs are mature (i.e. tall). Nevertheless, distinguishing shorter shrubs (*P. amplexicaulis* and *P. lorifolia*) from the ground remains a challenge, which could be due to the following convoluting factors. First, a recent study showed that drone-based data collection likely misrepresent the true heights of small to medium sized shrubs[15]. Second, our study site is not only rugged, but full of non-shrub vegetation which can reach heights similar to short Proteaceae, potentially leading to "background" pixels with relatively large heights. Some methods to improve accuracy of height information would be to explore near-ground photogrammetry (e.g. with a stick-mounted camera) or to use a fine-resolution digital elevation model to assist in normalizing the site height values.

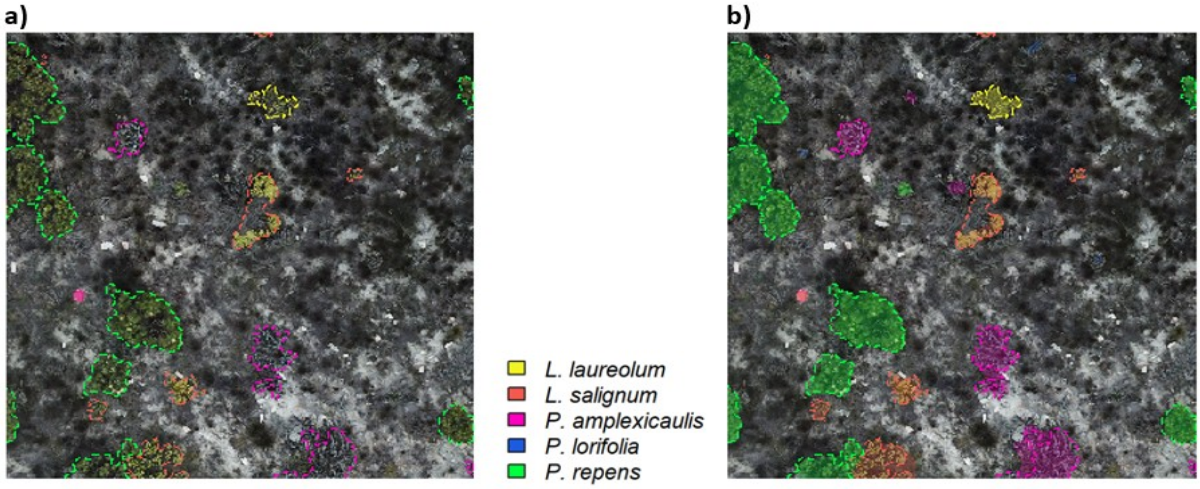

Figure 2: -a) Orthophoto with ground-truth data for one subsite (13.5 x 13.5 m²) in the validation data-set; -b) Predictions of the spectral-height model for the same subsite. In both plots ground-truth shrubs are delineated by a dashed line. In b) model predictions are represented as shading. Both lines and shading are colored by species.

4 Conclusions

We tested the efficacy of using drone photogrammetry and deep learning to automatically identify Proteaceae shrubs in South Africa's Fynbos region. The results are promising and highlight what factors will influence the success of such a workflow. Species with vibrant foliage and/or large size (e.g. *P. repens*) will contrast strongly with the background and should be easily detected, but those of a greyer color or more sprawling structure (e.g. *P. lorifolia*) will likely be missed due to visual and structural similarity with the background. We saw meaningful improvements in species identification by including height values, and plan to run the workflow on additional sites to demonstrate its transferability.

5 Acknowledgements

This work was funded by the German Research Foundation (DFG) grant SCHU 2259/3-3 to F.M.S. We utilized the computational resources of BwUniCluster funded by the Ministry of Science, Research and Arts Baden-Wurttemberg and the Universities of the State of Baden-Wurttemberg, Germany, within the frame-work program bwHPC.

References

1. Chen X, Jiang K, Zhu Y, Wang X, Yun T. Individual Tree Crown Segmentation Directly from UAV-Borne LiDAR Data Using the PointNet of Deep Learning. *Forests*. 2021;12(2):131. doi:10.3390/f12020131
2. Dale J, Burnside NG, Hill-Butler C, Berg MJ, Strong CJ, Burgess HM. The Use of Unmanned Aerial Vehicles to Determine Differences in Vegetation Cover: A Tool for Monitoring Coastal Wetland Restoration Schemes. *Remote Sens*. 2020;12(24):4022. doi:10.3390/rs12244022
3. Christin S, Hervet É, Lecomte N. Applications for deep learning in ecology. Ye H, ed. *Methods Ecol Evol*. 2019;10(10):1632-1644. doi:10.1111/2041-210X.13256
4. Pound MP, Atkinson JA, Townsend AJ, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*. 2018;7(7). doi:10.1093/gigascience/giy042
5. Beijbom O, Hoffman J, Yao E, et al. Quantification in-the-wild: data-sets and baselines. *ArXiv151004811 Cs*. Published online November 28, 2015. Accessed February 20, 2021. http://arxiv.org/abs/1510.04811
6. Guirado E, Tabik S, Rivas ML, Alcaraz-Segura D, Herrera F. Whale counting in satellite and aerial images with deep learning. *Sci Rep*. 2019;9(1):14259. doi:10.1038/s41598-019-50795-9
7. Wäldchen J, Mäder P. Machine learning for image based species identification. Cooper N, ed. *Methods Ecol Evol*. 2018;9(11):2216-2225. doi:10.1111/2041-210X.13075
8. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000;403(6772):853-858. doi:10.1038/35002501
9. Schurr FM, Esler KJ, Slingsby JA, Allsopp N. Fynbos Proteaceae as model organisms for biodiversity research and conservation. *South Afr J Sci*. 2012;108(11/12):4 pages. doi:10.4102/sajs.v108i11/12.1446
10. Schmid B, Nottebrock H, Esler KJ, et al. Responses of nectar-feeding birds to floral resources at multiple spatial scales. *Ecography*. 2016;39(7):619-629. doi:10.1111/ecog.01621
11. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs*. Published online May 18, 2015. Accessed October 6, 2021. http://arxiv.org/abs/1505.04597
12. Allaire JJ, Chollet F. *Keras: R Interface to "Keras". R Package Version 2.3.0.0.*; 2020. https://CRAN.R-project.org/package=keras
13. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. *ArXiv170802002 Cs*. Published online February 7, 2018. Accessed November 18, 2021. http://arxiv.org/abs/1708.02002
14. Krähenbühl P, Koltun V. *Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials*. NIPS; 2011.
15. Swetnam TL, Gillan JK, Sankey TT, et al. Considerations for Achieving Cross-Platform Point Cloud Data Fusion across Different Dryland Ecosystem Structural States. *Front Plant Sci*. 2018;8:2144. doi:10.3389/fpls.2017.02144