Deep Learning for Segmentation of Point Clouds with Synthetic Vegetation

Reed Anderson[1,2,3], Jan Bumberger[2], Peter Dietrich[2,4],
W. Stanley Harpole[2,3,5], Hans-Gerd Maas[1]

[1] TU Dresden
[2] Helmholtz Center for Environmental Research – UFZ
[3] German Centre for Integrative Biodiversity Research (iDiv)
[4] University of Tübingen
[5] Martin Luther University Halle-Wittenberg

Keywords: Deep Learning, Synthetic Data, Semantic Segmentation, Point Clouds, Domain Shift, Vegetation

Abstract:

Point clouds provide valuable qualitative representations of captured objects, however automatically summarizing these 3D representations quantitatively into simple metrics is challenging. This holds especially for complex objects such as plants. Recent advances in deep learning, specifically for processing point clouds, might be a solution for improving analysis of point clouds through semantic segmentation, that is, labeling points in a point cloud as particular categories of objects. Despite this promise, a key disadvantage of existing deep learning implementations is dependence on annotated data for supervised deep learning.

In this study we explore creating annotated data from grassland plots rendered with computer graphics as training data for the PointNet++ deep learning model [Qi 2017], which preforms semantic segmentation on point clouds. We then test this model on real-world point clouds generated from structure-from-motion photogrammetry on one square meter grassland plots. The result is a point cloud labeled as either vegetation or ground; this predicted value is compared against real-world biomass harvest.

Supervised machine learning is data intensive. While point clouds representing cityscapes can be hand labeled, doing the same with overlapping vegetation might prove too time consuming. Inspired by Virtual Kitti [Gaidon 2016], a database of synthetic video data used for developing autonomous driving and other tasks, we explore the use of computer generated scenes of vegetation to train deep learning models for real-world application. Some grass rendering modeling exists [Taubert 2012], but we were not able to find any for generating physiologically accuracy renderings of individual, and especially not grouped, grasses.

Instead, we use Blender for rendering a grassland scene using the “Grass Essentials - Grass Models” package [v1.2]. This package is used to create aesthetic grasses for entertainment purposes. Two grass and two forb plant render models were selected for the dataset. We used a uniform distribution for randomly: selecting if a plant be included in the scene, creating perturbations in the planar surface representing the ground, placing plants on this plane, and varying the heights of plants. To create the point cloud from the rendered scene, we captured a single nadir depth image above the synthetic grassland scene. This depth image includes information on whether an object was labeled as vegetation or ground. The dataset includes 10,000 grassland point clouds that qualitatively look like real-world grassland plots.

The synthetic point clouds are used as the input training data for the PointNet++ model [Qi 2017]. On synthetic data only, the model achieves 91 percent accuracy on vegetation points and 89 percent on ground. Without modifying the model trained on synthetic data, we then input a real-world point cloud into the trained model. No color information is input to the trained model. The output is a point cloud labeled as either vegetation or ground [Figure 2].

Next, the accuracy of the models predictions is compared against biomass harvest of 12 grassland plots subplots of size 0.2x0.5 meters. We only use counts of points labeled as vegetation for correlation to biomass harvest, although we plan to explore more complex representations in the future. Values for canopy surface height created with a binned, minimum value digital elevation model were also created from the unlabeled point clouds for comparison. [Figure 1].

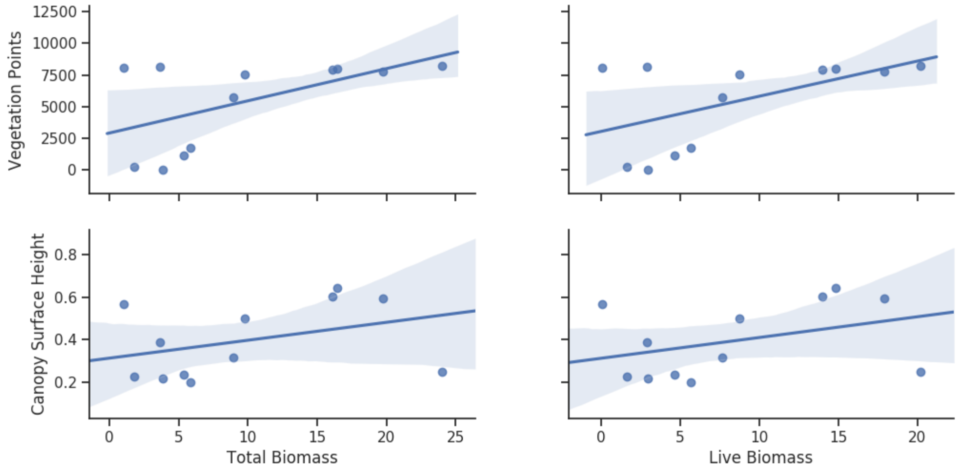

Figure 1: Regressions of harvested aboveground biomass against canopy surface height (live biomass r=0.36, total biomass r=0.38) and counts of vegetation points predicted by the deep learning model (live biomass r=0.53, total biomass r=0.55)

Notably, we observed the accuracy of the deep learning model on real-world data decrease as the model continued optimizing for the synthetic data. The predictions also appeared to have many false positive labels for ground points that were low growing vegetation, possibly because the model excludes color information.

The obvious, and perhaps unsurmountable, limitation of synthetic data is that it exists in a separate domain from the target object, in this case point clouds of real grassland plots. One approach to minimizing differences between the synthetic and real domain is improving the accuracy of the synthetic model, for instance, by creating physiologically accurate models of plants and grouping these synthetic models in representative scenes that will not strongly bias the deep learning model. Another approach is to align domains with an adversarial framework by minimizing a shared feature between model inputs [Sankaranarayanan 2018]. Such unsupervised domain adaptation could take the form of adversarial point perturbation [Xiang 2018]. Because of the cost and time associated with creating physiologically accurate models of plants, our next focus will be on unsupervised domain adaptation of the synthetic data using the real-world point clouds. Additionally, PointNet++ was developed for multiclass segmentation with approximately 20 classes of objects; we intend to explore how far this method can be applied in beyond binary classification.

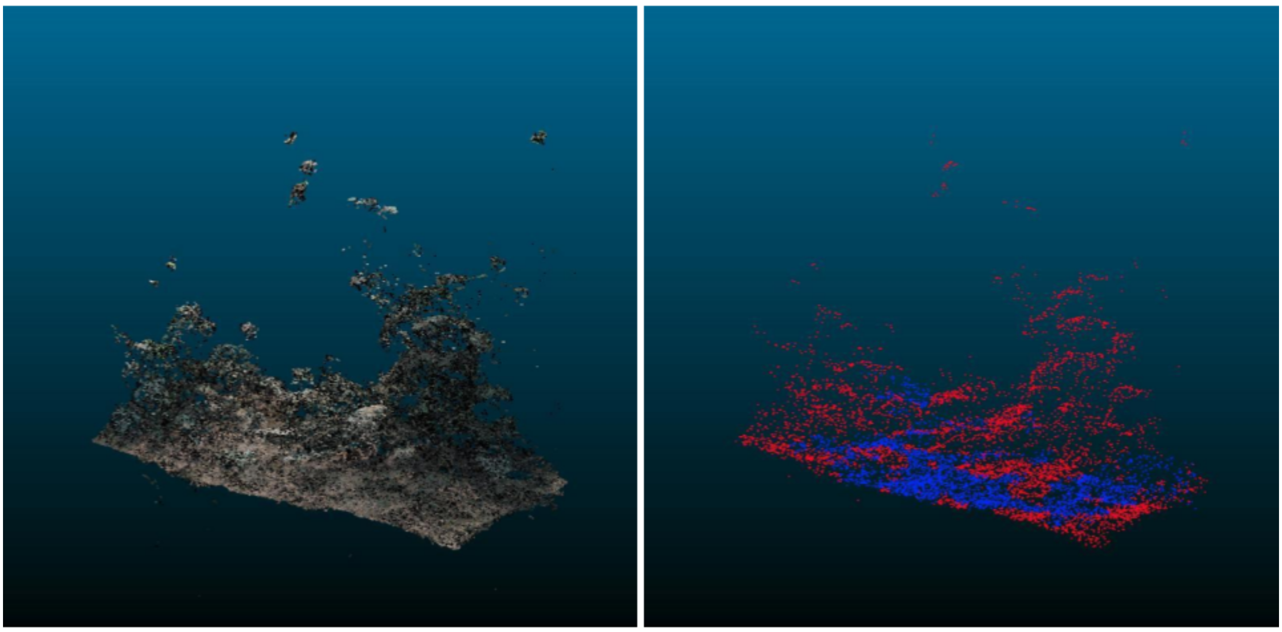

Figure 2: An example 0.2 x 0.5 meter point cloud captured on a grassland, which was input to the model without color values (left); An exceptional example of predicted point labels on the same point cloud; red is vegetation, blue is ground (right)

References:

Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE* 1.2 (2017): 4.

Gaidon, Adrien, et al. "Virtual worlds as proxy for multi-object tracking analysis." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Taubert, Franziska, Karin Frank, and Andreas Huth. "A review of grassland models in the biofuel context." *Ecological modelling* 245 (2012): 84-93.

Sankaranarayanan, Swami, et al. "Generate to adapt: Aligning domains using generative adversarial networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

Xiang, Chong, Charles R. Qi, and Bo Li. "Generating 3D Adversarial Point Clouds." *arXiv preprint arXiv:1809.07016* (2018).