Vegetation Density Estimation in the Wild

Radu P. Mihail[1], Wesley I. Cook[1], Brandi M. Griffin[2], Theodore A. Uyeno[2] and Corey D. Anderson[2]
Valdosta State University
[1]Department of Computer Science, [2]Department of Biology
{rpmihail, wicook, bmgriffin, tauyeno, coreanderson}@valdosta.edu

ABSTRACT

Remote sensing has revolutionized the efficiency of vegetation mapping, but such techniques remain impractical for mapping some types of flora over relatively limited spatial extents. We propose a deep-learning based framework for automated detection and planar mapping of an epiphytic plant in a forest from geotagged static imagery using inexpensive cameras. Our pipeline consists of two steps: segmentation and spatial distribution estimation. We evaluate several segmentation methods on a novel dataset of roughly 375 outdoor images with per-pixel labels indicating the presence of Spanish moss. We also evaluate the accuracy of the spatial distribution estimates with respect to field measurements by ecologists for Spanish moss.

CCS CONCEPTS

• **Computing methodologies** → *Visual inspection*; *Image segmentation*; *Object detection*;

KEYWORDS

Segmentation, In the Wild, Machine Vision, Deep Learning, Plants

ACM Reference Format:
Radu P. Mihail[1], Wesley I. Cook[1], Brandi M. Griffin[2], Theodore A. Uyeno[2] and Corey D. Anderson[2] . 2018. Vegetation Density Estimation in the Wild. In *ACM SE '18: ACM SE '18: Southeast Conference, March 29–31, 2018, Richmond, KY, USA.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3190645.3190690

1 INTRODUCTION

Vegetation mapping has become an essential tool for resource managers and scientists [26]. The development of remote sensing techniques (based on multispectral imagery derived from spacecraft and manned aerial vehicles) has substantially improved the resolution and accuracy of existing vegetation maps [39]. Multispectral optical measurements provide discriminative cues for determining vegetation characterisitics and other remote sensing technologies such as LiDAR (Light Detection and Ranging) have been shown to be effective in mapping biomass density and distribution [21].

Although remote sensing has become the favored practice for vegetation mapping, processing remote sensing data can be an esoteric endeavor, usually requiring specialized training and software. Moreover, remote sensing data obtained by aerial and space cameras may not always be readily available for all field sites at the appropriate spatial resolution. Where they are obtainable, they must often be custom ordered at a cost that is prohibitive for many researchers. While the use of lightweight unmanned aerial vehicles (UAVs) or drones has made aerial sensing more feasible and affordable for local-scale studies [2], such technology may still be cost prohibitive, or even illegal in some areas (e.g., near airports).

As compared to multispectral imagery and LiDAR, visible light photography is inexpensive and widely available, but lacks additional information about material composition and 3D structure. Visible light images are also subject to environmental lighting variations (e.g., haze, shadows, specular reflections) which affect the appearance of the objects of interest. Despite these shortcomings, new methods in image processing that enable processing of natural imagery could provide a cheaper and easier way to classify and map vegetation.

Segmentation of natural imagery into semantically meaningful classes is a difficult problem and has received relatively little attention in the research community. Santos et al. [33] use structure from motion (SfM) to create a point cloud from videos of plants in controlled environments for the purpose of phenotyping. Campos et al. [11] describe a method to segment vegetation in corn fields, presumably as a pre-processing step for agricultural robots. Yanikoglu et al. [40] focus on plant identification from leaves using shape and texture features. These authors emphasize that intra-class variability in appearance is a big problem with automatic plant detection.

In the present study, we use and evaluated deep-learning frameworks to automatically detect and estimate the probability surface for the Spanish moss plant (*Tillandsia usneoides*) from geotagged static imagery taken with a common smartphone. Because the focal species is a pendant epiphyte (i.e., most often hangs from trees), its color and texture can easily be confounded with background vegetation. This makes segmenting outdoor imagery a challenging task. The proposed technique has widespread applications for resource managers and scientists by providing a way to conduct local scale vegetation mapping through the use of an inexpensive camera.

2 RELATED WORK

Background on surveyance of epiphytic organisms. An epiphyte is a plant that uses another plant, usually a tree, for support [8, 45]. Because the spatial distribution of an epiphyte is dependent on the location of suitable hosts, epiphytes represent a biologically interesting opportunity to examine direct and indirect interactions between plant species. As such, ecologists have examined the specificity of many different epiphytes to various hosts [37].

ACM SE '18, March 29–31, 2018, Richmond, KY, USA

ACM ISBN 978-1-4503-5696-1/18/03…$15.00
https://doi.org/10.1145/3190645.3190690

Although there is widespread interest in epiphyte-host specificity, ground-based censuses are notoriously challenging because many epiphytes occur high in the forest canopy and, if detectable, grow in clumps or festoons that make it difficult to count discrete individuals [23]. While tools, such as canopy cranes [38], and technologies such as unmanned aerial vehicles have helped researchers to access and locate epiphytes in the forest canopy, such techniques (as applied to date) still require manual identification and mapping of epiphytic species to different trees in a forest. Airborne multispectral imaging and LIDAR technologies could potentially provide a new means to locate and measure epiphyte populations, yet there has been virtually no discussion of their applicability beyond epiphytes being a nuisance variable when trying to interpret the spectral signature of other types of vegetation [9, 16]. Because they are difficult to sample, new technologies that aid in detecting and measuring epiphytes could have a transformative effect on the study of their ecology and species interactions.

Background on focal species. In the southeastern United States, Spanish moss is the dominant epiphyte in coastal regions [17] and seems to be especially abundant in hardwood hammock and cypress swamp, and less common in pine forest [17]. Spanish moss can reproduce both by seed and vegetatively [20] and has been found on a wide range of host species, but seems to thrive on certain hardwood species, such as *Quercus virginiana* (live oak) [10, 31]. Recent works have demonstrated that the presence of Spanish moss can have a major impact on local species composition [3, 4], and more efficient methods for detecting and mapping Spanish moss in a forest would facilitate better modeling of the factors that affect its distribution in different types of forest communities.

Semantic Segmentation. Semantic segmentation of images consists of assigning each pixel one of several labels that are semantically meaningful (e.g., car, building, tree). Semantic segmentation has many applications as input to high-level algorithms such as decision-making in autonomous driving [5, 13, 42], handwriting recognition [25, 27], inferring functional relationships between different parts of an image [29], object detection and tracking [19], medical image interpretation [14, 15, 28, 32, 34, 44] and many others. The semantic segmentation literature is vast, and a thorough review is beyond the scope of this work, however Zhao et al. [43] provides an excellent summary. In the past few years, deep neural networks have proven better than specialized, "shallow" algorithms at many tasks in the machine learning and computer vision communities. A special type of deep neural network designed for imagery, the convolutional neural network (CNN) has been applied to segmentation problems.

The successes of CNNs drove the development of more generic models for image segmentation, called fully convolutional networks (FCNNs). In this work, we thoroughly evaluate two state-of-the-art FCNN models: SegNet [6] and PixelNet [7]. In SegNet, the indices of pixels max-pooling layers of a typical CNN are kept to produce sparse maps in the upper layers of the network. The sparse maps are then convolved with filters to perform non-linear upsampling that result in dense segmentation maps. The authors of SegNet show excellent results on outdoor roadway scenes.

PixelNet is a FCNN that makes simultaneous use of information from all convolutional layers with hypercolumns. These hypercolumns are then used independently to compute per-pixel losses that can be sparse through the image due to the many redundancies naturally present, thus speeding the training process significantly. Pixelnet [7] has achieved excellent results in performing semantic segmentation on the Pascal-Context dataset compared to other recent architectures such as in [12] and [30].

Hypercolumn feature extraction occurs for individual pixels across multiple convolutional layers as described in [22]. The hypercolumn feature extraction serves to incorporate low-level features learned from all convolutional layers, as seen in Figure 1. Subsampling is performed during training, where only a small subset of the total pixels is chosen randomly from each image in a batch. The hypercolumn features extracted from a full image induces a high memory footprint, thus the purpose of random subsampling is to reduce memory usage during training. During testing, the full hypercolumn set is used, and the memory required is manageable due to reduction of batch size to one. A softmax cross-entropy loss function is used for training the model and back-propagation through stochastic gradient descent is used to update model parameters.

Contributions. Our primary contribution is the introduction and definition of of image-based, automatic vegetation planar mapping as the computation of a probability density function over a geographic area from a set of labeled, geotagged images. We propose a novel method to estimate geospatial density of Spanish moss from geotagged imagery in two sequential steps: 1) automatic image segmentation and 2) integration of the semantically meaningful labels from geotagged imagery to estimate a spatial probability density for a target of interest.

Our secondary contribution is a novel dataset of labeled outdoor natural imagery. Manually labeling vegetation in cluttered environments is a time-consuming process that requires time and patience. The complex appearance of plants in a natural environment is a result of the relative sparsity of branches and leafs, thus per-pixel labeling is an almost impossible task. We approach this problem using a custom interface, discussed later in this paper.

In the remainder of this paper, we compare the performance of several CNN architectures for the semantic segmentation task on our dataset. We also compare the results of our novel vegetation mapping approach with binary indicator data collected by ecologists in the field. We hope that our work will lead to further developments on semantic segmentation in this difficult domain of unstructured and natural imagery in the wild, where appearance variation plays a critical role.

3 DATASET

Our dataset contains images with per-pixels labels for Spanish moss. Since labeling each pixel individually is very time consuming and difficult to perform, we have created a labeling tool to annotate images. The tool segments the image into superpixels using SLIC [1] superpixels implemented in VLFeat [36]. The user then swipes over superpixels that are then highlighted with a specific tint to mark them as labeled. To make the labeling more accurate, the user is allowed to change the region size and regularization parameters of the SLIC algorithm. The images were selected manually (due to

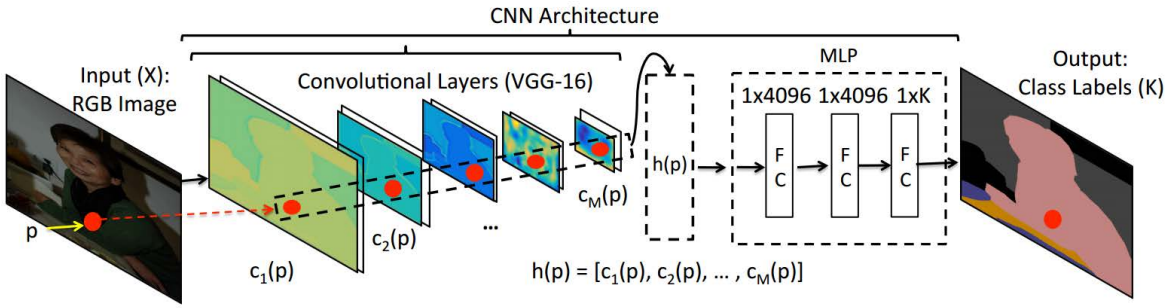

Figure 1: A visualization of the Pixelnet architecture from Bansal et. al [7]

Figure 2: Sample image-label pairs from Spanish moss dataset.

many inaccurate tags) and downloaded from Flickr. The datasets are accessible at http://mypages.valdosta.edu/rpmihail/data.

4 IMAGE-BASED GEOSPATIAL DENSITY ESTIMATION

Given a set of geotagged images $I = \{i_1, i_2, ..., i_n\}$ of images from a calibrated camera, and geotags $G_i = (longitude, latitude, azimuth, FOV)$ we want to estimate $P_{target}(p|I_n, G_n)$, where p is a point in a world. This function is conceptually easy to estimate if we had aerial imagery (e.g., from drones or satellite) where the camera's image plane is parallel to the ground plane and we have a perfect per-pixel classifier and no occlusions. In this ideal scenario, it would suffice to map each pixel from image-space to world-space and compute the probability directly from the image. Such ideal datasets are used to estimate canopy coverage and make-up [18, 35, 41], but are limited to targets directly visible from above. Therefore, these data and methods do not work for targets that are occluded by the canopy.

Here, we consider a more practical image acquisition scenario where 1) images are sparse (several meters apart), 2) the image plane is perpendicular to the ground plane, and the center of projection is roughly eye-level for an average-height adult. These criteria describe an extremely common scenario from from consumer-level devices. Since smartphones are ubiquitous, most ecologists will find obtaining such data both affordable and easily available.

Each calibrated image i_n is run through a segmentation algorithm, which results in a probability $P_{i_{x,y}}$ for a pixel location (x, y). At this stage of the process, we assume each pixel is independent of its neighbors.[1] We then marginalize the probabilities over the columns of the image:

$$P_{i_x} = \frac{1}{\lambda} \sum_y T(x, y) \tag{1}$$

where T is the target segmentation, λ is a normalization constant set to the height of the image in pixels, such that if an entire image column x is positive for the target, P_{i_x} is one. Each image column is then mapped to a global camera azimuth angle, given the camera's horizontal field of view (FOV) and azimuth angle of the optical axis (assumed parallel to the ground plane) at the time of capture, resulting in $P_i(a)$, where $a \in [G_i(azimuth) - \frac{G_i(FOV)}{2}, G_i(azimuth) + \frac{G_i(FOV)}{2}]$.

[1] While the smoothness assumption is ignored here, it is fully captured by the deep learning segmentation method.

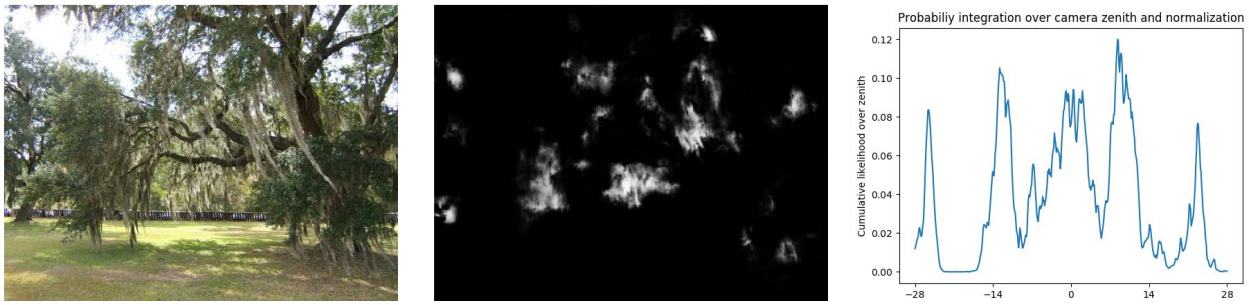

Figure 3: Given an image (left), we use the per-pixel probability estimate of a target (middle) to compute a density over camera azimuth by marginalizing the image columns (right).

Looking at the world from a top-down (bird's eye) view, for some point $p = (lat, lon)$, we can determine the images that may potentially include the target using simple triangulation. If two or more images "see" a given point, we assign $P(target|p)$ the weighted sum of $\sum_{i,a} P_i(a)w_{i,a}$, where w is a decay variable that decreases exponentially with distance, thus minimizing the effect of distant points[2]. We illustrate one case with two geotagged images pointed at a target (a tree) in Figure 4. The flattened segmentation results are projected to the world space and if the camera's view frustums intersect, the detector results are summed, weighted and normalized.

Our method is most accurate when a target is imaged from three or more positions with azimuth angles that minimize the intersected area within a detectable potential target. In Figure 5 we show two synthetic examples with density estimates for two and three cameras. With only two cameras, there is an ambiguity in the position of a target. Three or more cameras are sufficient to determine the position of an object in the world.

5 EXPERIMENTS

In this section, we evaluate the results of segmentation and spatial distribution estimation. We use the Flickr dataset to evaluate FCNNs on the segmentation task. We randomly select 80% of images for training, and 20% for testing. We report several performance metrics for the FCNNs on the testing set. The evaluation of our spatial distribution estimation is qualitative and results are compared with data from field measurements.

5.1 Moss Segmentation

Similar to Long et. al [30], we evaluate the performance of the four CNNs using the following metrics: the pixel accuracy, the mean pixel accuracy between classes (referred to as mean accuracy), the mean intersection over union (IoU) scores for each class, and the frequency weighted IoU. The best results with respect to pixel accuracy was achieved by PixelNet. The greatest mean accuracy between classes was achieved by SegNet. A few sample segmentation results are shown in Figure 6. Overall performance metrics are shown in Table 1. All models were trained from scratch.

[2]The distance at which targets are detectable varies greatly for different settings, thus w is set empirically. In our experiments, $w = z \times e^{-d}$, where d is the distance from the image to the point in the world, and z is a normalization constant.

Table 1: Results of segmentation. The pixel accuracy is defined as the ratio of correctly labeled pixels and total number of pixels. Mean accuracy is defined as the average of the pixel accuracy of each class. FW IU is a frequency weighted intersection over union. We evaluated two flavors of PixelNet: the base model as released by its authors, and PixelNet with median frequency balancing.

	Pixel Acc.	Mean Acc.	Mean IoU	FW IoU
PixelNet Base	**87.4**	57.6	51.0	**78.7**
PixelNet MFB	82.7	66.3	**52.4**	75.0
SegNet	73.3	**68.9**	46.4	65.5

Class Balancing. Due to a relatively low number of target labels compared to background, CNNs often converge to the trivial solution of labeling everything background. To prevent this, we test our FCNN configurations with an implementation of median frequency balancing (MFB), which addresses inter-class frequency by assigning a higher weight in the loss function to classes which occur less frequently. Kampffmeyer et al. [24] demonstrate the efficacy of MFB in cases of high class imbalance with remote sensing of urban landscapes. The effect of adding MFB, is an increase in the mean IoU and mean accuracy, but leads to a decrease in pixel accuracy with respect to the original implementation.

5.2 Spatial Distribution Estimation

The method of our spatial distribution estimation produces a discrete probability map over the imaged region, where each cell represents the probability of finding Spanish moss in that area. In our experiment we used 206 images.

We compared the results of our spatial distribution estimation with hand-collected data in the field. The hand-collected data represent the presence (= 1) or absence (= 0) of Spanish moss in the vicinity of regularly spaced points (n = 195) within a ≈8.3 ha region of a Slash-pine dominated forest stand at the Lake Louise Research Center in Lowndes County, Georgia, USA. At each sample point, all trees within 10 m search radius (with a diameter at breast height ≥ 5 cm) were surveyed for Spanish moss with the naked eye and with the aid of binoculars. A total of 1254 trees were sampled, with 209 of the trees containing Spanish moss (= 17%). Overall, Spanish

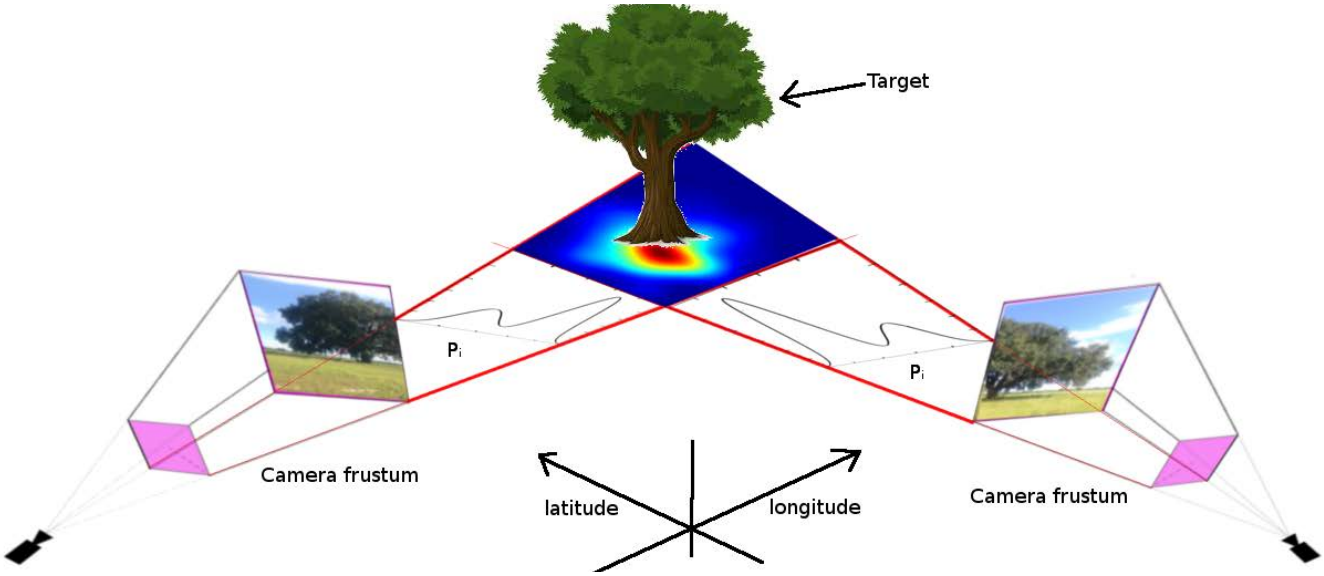

Figure 4: We consider the case of two geotagged images from calibrated cameras pointed at a tree—the target, from two different perspectives. For each image, we have detector results, P_i. The final density is estimated by summing and weighing detector results of intersecting rays in the world space.

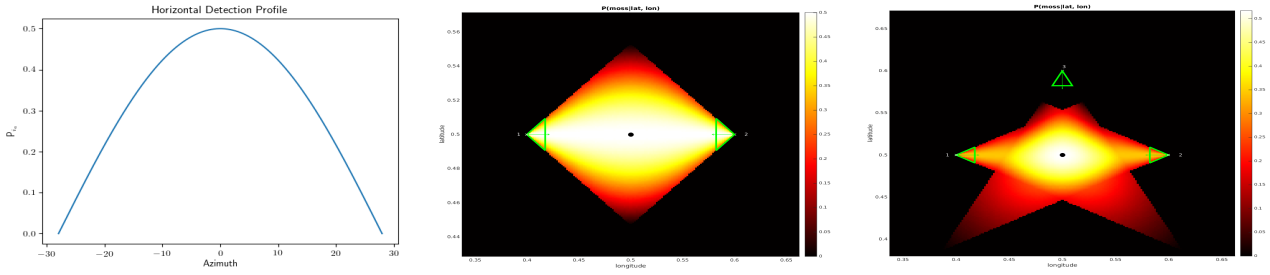

Figure 5: We synthesize a detection profile (left) for all images and compute the spatial density given two images taken at diametrically opposite sides of a target (middle). In the two image configuration, there is an ambiguity along the optical axis. Adding a third image from solves the ambiguity (right).

was found in 83 of the 195 sample units (= 43%), with most of the Spanish moss located around the periphery of the stand, in ecotonal areas near old growth hardwood forest.

To compare the ground truth data with our image-based density estimates, we used indicator kriging in ArcGIS 10.4 (Environmental Research Systems Institute, Redlands, CA) to interpolate a continuous probability surface from the binary Spanish moss presence/absence data. When performed on binary data, indicator kriging amounts to a nonparametric version of ordinary kriging that does not assume normality. We chose indicator kriging because: 1) it is naturally well-suited to binary response data, 2) it provides a probability surface and information about the standard error of the estimates, 3) cross-validation can be used to optimize model parameters, and 4) optimized model parameters are based on information about spatial structure in local search neighborhoods around each interpolated position, thus providing important detail about both local and global patterns of Spanish moss within the study area.

6 CONCLUSION

In this paper, we propose a framework for measuring the spatial density of Spanish moss from geotagged visible-light photographs, and our results show the potential for consumer-grade cameras to assist with vegetation mapping. We compared various performance metrics for segmentation of three FCNNs: two variations of PixelNet and SegNet. We found PixelNet to be the best performing architecture with respect to the pixel accuracy and frequency weighted intersection over union (FW IoU).

Implementing median frequency balancing (MFB) in PixelNet was observed to increase the mean accuracy and the mean IoU over the base PixelNet. Although we found SegNet to perform the best with regard to the mean accuracy, the results from SegNet appeared worse according to our other metrics: pixel accuracy, mean IoU, and FW IoU.

Our spatial density estimation results are encouraging. We compare our method to interpolated estimates made during a more

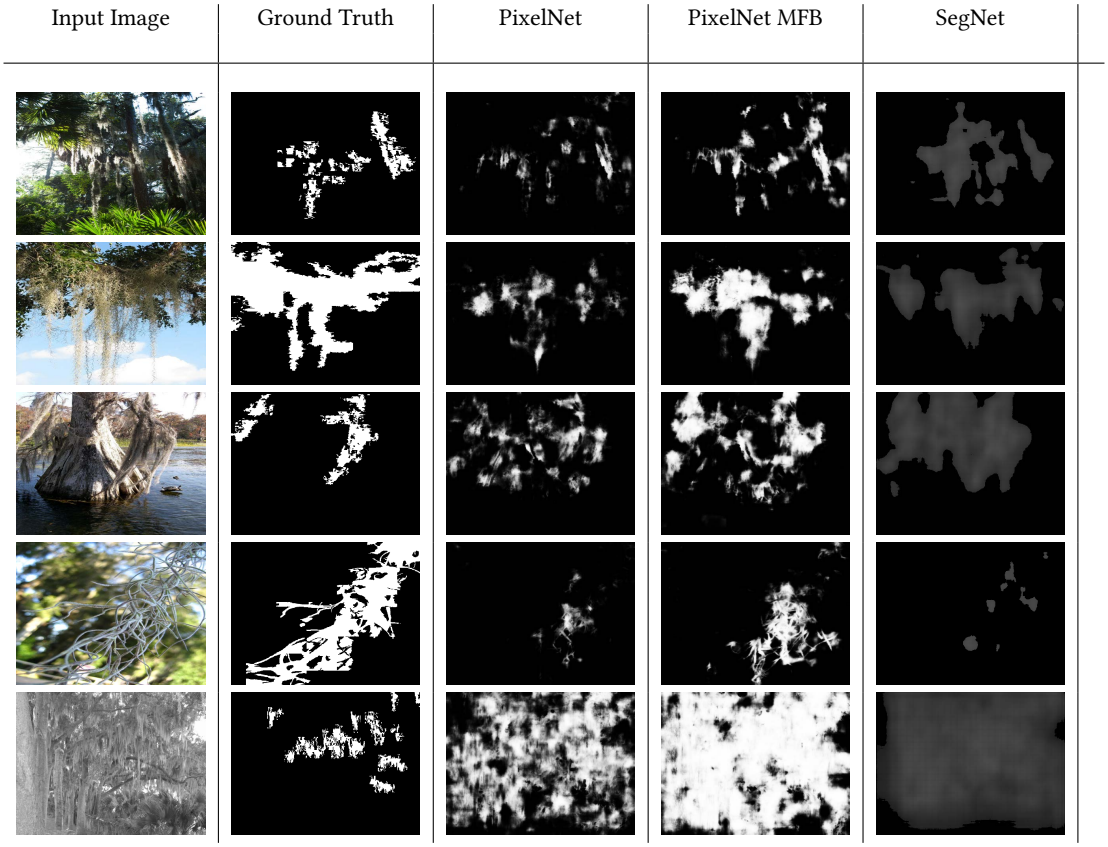

Figure 6: Sample segmentations from the Flickr dataset.

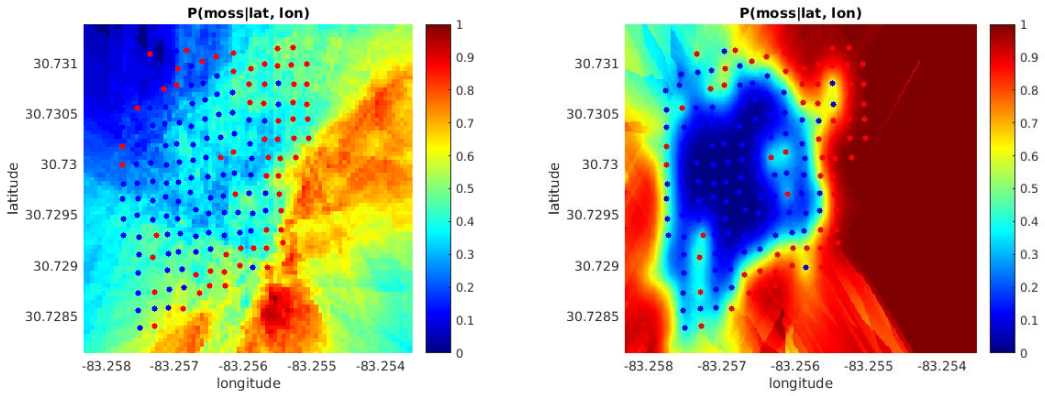

Figure 7: Left: our results superimposed on hand-collected binary indicators. Right: interpolation using indicator kriging from the binary indicators. Our results provide high resolution density estimates of Spanish moss using significantly less resources than standard measurements by ecologists.

traditional survey performed by an ecologist. Our results are highly correlated with the ecologists' data, which suggests this approach is a viable alternative for larger areas. In the future, we will evaluate our method on different species of plants and larger areas.

ACKNOWLEDGMENTS

The authors would like to thank the Valdosta State University Office of Sponsored Programs and Research Administration for funding project "Green Vision".

REFERENCES

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* 34, 11 (2012), 2274–2282.

[2] Karen Anderson and Kevin J Gaston. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment* 11, 3 (2013), 138–146.

[3] Christine Angelini and Kristin Lauren Briggs. 2015. Spillover of secondary foundation species transforms community structure and accelerates decomposition in oak savannas. *Ecosystems* 18, 5 (2015), 780–791.

[4] Christine Angelini and Brian R Silliman. 2014. Secondary foundation species as drivers of trophic and functional diversity: evidence from a tree–epiphyte system. *Ecology* 95, 1 (2014), 185–196.

[5] Anelia Angelova, Alex Krizhevsky, Vincent Vanhoucke, Abhijit S Ogale, and Dave Ferguson. 2015. Real-Time Pedestrian Detection with Deep Network Cascades.. In *BMVC*. 32–1.

[6] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).

[7] Aayush Bansal, Xinlei Chen, Bryan Russell, Abhinav Gupta Ramanan, et al. 2017. Pixelnet: Representation of the pixels, by the pixels, and for the pixels. *arXiv preprint arXiv:1702.06506* (2017).

[8] DH Benzing. 1990. Vascular Epiphytes 354. (1990).

[9] Marcus Borengasser, William S Hungate, and Russell Watkins. 2007. *Hyperspectral remote sensing: principles and applications*. Crc Press.

[10] Ragan M Callaway, Kurt O Reinhart, Georgianne W Moore, Darrin J Moore, and Steven C Pennings. 2002. Epiphyte host preferences and host traits: mechanisms for species-specific interactions. *Oecologia* 132, 2 (2002), 221–230.

[11] Yerania Campos, Erik Rodner, Joachim Denzler, Humberto Sossa, and Gonzalo Pajares. 2016. Vegetation segmentation in cornfield images using Bag of Words. In *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 193–204.

[12] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2016. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915* (2016).

[13] Xiaozhi Chen, Kaustav Kundu, Ziyu Zhang, Huimin Ma, Sanja Fidler, and Raquel Urtasun. 2016. Monocular 3d object detection for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2147–2156.

[14] Zhipeng Cui, Jie Yang, and Yu Qiao. 2016. Brain MRI segmentation with patch-based CNN approach. In *Control Conference (CCC), 2016 35th Chinese*. IEEE, 7026–7031.

[15] El-Sayed A El-Dahshan, Heba M Mohsen, Kenneth Revett, and Abdel-Badeeh M Salem. 2014. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert systems with Applications* 41, 11 (2014), 5526–5545.

[16] SK Fyfe. 2003. Spatial and temporal variation in spectral reflectance: Are seagrass species spectrally distinct? *Limnology and Oceanography* 48, 1part2 (2003), 464–479.

[17] Richard Edwin Garth. 1964. The ecology of Spanish moss (Tillandsia usneoides): its growth and distribution. *Ecology* 45, 3 (1964), 470–481.

[18] Philipp Gärtner, Michael Förster, Alishir Kurban, and Birgit Kleinschmit. 2014. Object based change detection of Central Asian Tugai vegetation with very high spatial resolution satellite imagery. *International Journal of Applied Earth Observation and Geoinformation* 31 (2014), 110–121.

[19] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2016. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence* 38, 1 (2016), 142–158.

[20] Arthur T Guard and Marilee Hen. 1968. Reproduction of Spanish moss, Tillandsia usneoides L., by seeds. *Bulletin of the Torrey Botanical Club* (1968), 327–330.

[21] Xuan Guo, Nicholas C Coops, Piotr Tompalski, Scott E Nielsen, Christopher W Bater, and J John Stadt. 2017. Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecological Informatics* 38 (2017), 50–61.

[22] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. 2015. Hypercolumns for object segmentation and fine-grained localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 447–456.

[23] Dick Johansson. 1974. *Ecology of vascular epiphytes in West African rain forest*. Ph.D. Dissertation. Sv. växtgeografiska sällsk.

[24] Michael Kampffmeyer, Arnt-Borre Salberg, and Robert Jenssen. 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–9.

[25] Daniel Keysers, Thomas Deselaers, Henry A Rowley, Li-Lun Wang, and Victor Carbune. 2016. Multi-Language Online Handwriting Recognition. *IEEE transactions on pattern analysis and machine intelligence* (2016).

[26] August Wilhelm Küchler and Isaak Samuel Zonneveld. 2012. *Vegetation mapping*. Vol. 10. Springer Science & Business Media.

[27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[28] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 305–312.

[29] Liang Lin, Guangrun Wang, Rui Zhang, Ruimao Zhang, Xiaodan Liang, and Wangmeng Zuo. 2016. Deep structured scene parsing by learning with image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2276–2284.

[30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3431–3440.

[31] Wm T Penfound and FG Deiler. 1947. On the ecology of Spanish moss. *Ecology* 28, 4 (1947), 455–458.

[32] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *International conference on medical image computing and computer-assisted intervention*. Springer, 246–253.

[33] Thiago Teixeira Santos, Luciano Vieira Koenigkan, Jayme Garcia Arnal Barbedo, and Gustavo Costa Rodrigues. 2014. 3D plant modeling: localization, mapping and segmentation for plant phenotyping using a single hand-held camera. In *European Conference on Computer Vision*. Springer, 247–263.

[34] Dinggang Shen, Guorong Wu, and Heung-Il Suk. 2017. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* 0 (2017).

[35] Alexei N Skurikhin, Steven R Garrity, Nate G McDowell, and Dongming M Cai. 2013. Automated tree crown detection and size estimation using multi-scale analysis of high-resolution satellite imagery. *Remote sensing letters* 4, 5 (2013), 465–474.

[36] A. Vedaldi and B. Fulkerson. 2008. VLFeat: An Open and Portable Library of Computer Vision Algorithms. http://www.vlfeat.org/. (2008).

[37] Katrin Wagner, Glenda Mendieta-Leiva, and Gerhard Zotz. 2015. Host specificity in vascular epiphytes: a review of methodology, empirical evidence and potential mechanisms. *AoB Plants* 7 (2015), plu092.

[38] Y Wu, WY Liu, L Song, X Chen, HZ Lu, S Li, and XM Shi. 2016. Advances in ecological studies of epiphytes using canopy cranes. *Chinese Journal of Plant Ecology* 40 (2016), 508–522.

[39] Yichun Xie, Zongyao Sha, and Mei Yu. 2008. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology* 1, 1 (2008), 9–23.

[40] B Yanikoglu, Erchan Aptoula, and Caglar Tirkaz. 2014. Automatic plant identification from photographs. *Machine vision and applications* 25, 6 (2014), 1369–1383.

[41] Gong Zhang, Sangram Ganguly, Ramakrishna R Nemani, Michael A White, Cristina Milesi, Hirofumi Hashimoto, Weile Wang, Sassan Saatchi, Yifan Yu, and Ranga B Myneni. 2014. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment* 151 (2014), 44–56.

[42] Ziyu Zhang, Sanja Fidler, and Raquel Urtasun. 2016. Instance-level segmentation for autonomous driving with deep densely connected mrfs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 669–677.

[43] Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan. 2017. A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing* (2017), 1–17.

[44] Tian Zhou, Ilknur Icke, Belma Dogdas, Sarayu Parimal, Smita Sampath, Joseph Forbes, Ansuman Bagchi, Chih-Liang Chin, and Antong Chen. 2017. Automatic segmentation of left ventricle in cardiac cine MRI images based on deep learning. In *SPIE Medical Imaging*. International Society for Optics and Photonics, 101331W–101331W.

[45] Gerhard Zotz. 2016. *Plants on Plants: The Biology of Vascular Epiphytes*. Springer.