Stewart T. Schultz
E-mail: sschultz@unizd.hr
Claudia Kruschel
E-mail: ckrusche@unizd.hr
University of Zadar, Department of Ecology, Agronomy and Aquaculture
Trg Kneza Višeslava 9, 2300 Zader, Croatia
Viviane Wolff
E-mail: viviane.wolff@et.fs-fulda.de
Klaus Fricke-Neuderth
E-mail: Klaus.Friscke-Neuderth@et.fs-fulda.de
Fulda University of Applied Sciences
Leipziger Strasse 123, 36037 Fulda, Germany
Dubravko Pejdo
E-mail: dubravkopejdo@yahoo.com
University of Zadar, Department of Ecology, Agronomy and Aquaculture
Trg Kneza Višeslava 9, 2300 Zader, Croatia
Jonas Jaeger
E-mail: jonas.jaeger@et.hs-fulda.de
Fulda University of Applied Sciences
Leipziger Strasse 123, 36037 Fulda, Germany

Semantic Segmentation for Posidonia Oceanica Coverage Estimation

Abstract

One method of assessing the ecological status of seagrass is the analysis of videographic images for variables such as total aerial cover. Georeferenced images can be collected and matched by location over time, and any changes in coverage can be compared statistically to the expected null hypothesis. Since the manual analysis of large datasets approaching over a million images is not feasible, automated methods are necessary. Because of the wide variation in underwater conditions affecting light transmission and reflection, including biological conditions, deep learning methods are necessary to distinguish seagrass from non-seagrass portions of images. Using deep semantic segmentation, we evaluated several deep neural network architectures, and found that the best performer is the DeepLabv3Plus network, at close to 88% (intersection over union). We conclude that the deep learning method is more accurate and many times faster than human annotation. This method can now be used for scoring of large image datasets for seagrass discrimination and cover estimates. Our code is available on GitHub: https://enviewfulda.github.io/LookingForSeagrassSemanticSegmentation

Keywords: Posidonia, machine learning, neural network

1. Introduction

This research used machine learning to identify seagrass in underwater images and calculate the proportion of the image covered by seagrass. We found that the deep neural network DeepLabv3Plus was able to identify seagrass in test images to 88% intersection over union. This method is practicable and immediately useful for videographic monitoring of *Posidonia oceanica* seagrass in the Mediterranean Sea.

Seagrass communities provide a variety of valuable ecological services to coastal communities, including water purification, stabilization of the substrate, erosion control, habitat for native species, nursery grounds for juvenile fish and invertebrates, and primary productivity at the base of the coastal food web [1].

Seagrass populations are in decline worldwide due to a variety of anthropogenic causes, including physical damage (coastal works, bottom trawling, erosion) and eutrophication (mainly sewage and aquaculture). In the Mediterranean, these have caused overall losses of approximately 15 to 50% in *Posidonia* cover from 1842 to 2009, and possibly 50% loss in shoot density in just the last 20 to 25 years [2].

Development of accurate and precise monitoring methods is necessary to prevent or mitigate ongoing loss. Such methods should be able to reliably demonstrate any losses equal to or greater than 10% between two adjacent monitoring events. Two classes of methods are capable of satisfying this criterion: 1) fixed-plot methods such as balisage, in which seagrass descriptors are measured at precise marked underwater locations, and 2) remote methods with very low classification error, such as remote underwater videography (RUV) [3]. The RUV method is the most cost-effective for fieldwork, providing tens of thousands of images during a single field work day [3]. However, RUV requires automated analysis of these images to be cost-effective in the office/laboratory.

Deep learning methods can be used for this task, as for example convolutional neuronal networks (CNNs) [4]. Advantages include the absence of any need for image preprocessing; instead the CNN learns about the structure of the object images and from that structure generates recognized definable features. CNNs historically have been used to classify entire images. In the present study, we use Semantic Segmentation to classify each pixel of an image, in this case into seagrass or non-seagrass. After each pixel is classified, the total proportion of pixels that are classified as seagrass is calculated and reported for each image, giving the estimate of seagrass cover.

Past research on seagrass identification using CNNs has taken the approach of identifying uniform patches of interest within each image, which are then classified and analyzed, with extracted features [5, 6, 7]. Patches can be recognized using superpixel algorithms. Our approach in contrast processes the entire image, using Semantic Segmentation to classify every pixel in the image. Fully CNNs allow this by generating

convolutional layers that can be transposed to return to the original size and total pixel number of the image.

Previous work on *Posidonia oceanica* machine learning has used the VGG network encoder coupled with the FCN-8 decoder architecture for the goal of segmentation of seagrass images, using a training set of 460 images at Mallorca, Spain taken from an AUV [8]. The trained NN achieved a pixel accuracy of 96.1%, 94.0%, and 87.6% as compared to the U-Net result of 93.1%, 82.3%, and 81.2%.

2. Material and methods

Our dataset consisted of 6037 images of *Posidonia oceanica* in its natural habitat along the coast of the island of Murter (Croatia). These images were taken by an AUV using a GoPro Hero 2 camera providing a resolution of 1920 x 1080 pixels in September 2012. The images were taken between 1 and 6 meters distance above the ground, 80% of which were closer than 4 meters. The images contain a mixture of seagrass, algae, sand, and invertebrates. Annotation was performed by two people trained by a marine ecologist (SS) in 2017.

Our method consisted of two steps: 1) training of the segmentation networks, and 2) application of the trained network to test images to calculate the accuracy of the method.

We used four deep neural network architectures with multiple hidden layers, based on the Deeplablv3Plus by Tensorflow [9]. These were netFCN, uNet, dilNet, and MobileNetv2 [10, 11, 12, 13]. These networks differ in several respects, relating to the degree of convolution, the presence of dilation, the conservation of features from the downsampling layer, and the presence of depth-wise convolution which conserves on computational resources (Table 1).

Training was performed on images classed in two groups: taken at distances 1) closer than 2 meters to the ground, and 2) between 2 and 6 meters from the ground. For the first three networks training occurred in 60 epochs with a batch size of 4; the fourth used a batch size of 1. To speed the training all images were downsized to 512 x 256 pixels. Hardware used was an Intel i7-8700k with NVIDIA 1080ti, 16GB or RAM, running on Ubuntu 16.04LTS. All code is available at: https://enviewfulda.github.io/LookingForSeagrassSemanticSegmentation/

To evaluate the performances of the CNNs, we evaluated pixel accuracy, mean accuracy, mean intersection over union, and the frequency weighted intersection over union. The final result is the maximum of the last softmax layer, the normalized exponential function: each pixel is assigned the identity of the highest probability according to the exponential function.

Table 1: Architecture of networks: layer/module name, kernel size and feature map size.

netFCN	uNet	dilNet
conv- 3x3, 128	encode1 - 3x3, 64	encode1 - 3x3, 64
maxPool - 2x2, 2	encode2 - 3x3, 128	encode2 - 3x3, 128
conv- 3x3, 256	encode3 - 3x3, 256	encode3 - 3x3, 256
maxPool - 2x2, 2	encode4 - 3x3, 512	dilationBlock - x2, 512
conv- 3x3, 256	sole - 3x3, 512	dilationBlock - x4, 512
maxPool - 2x2, 2	decode1 - 3x3, 512	pixelDeconv - 3x3, 512
conv- 3x3, 256	decode2 - 3x3, 256	pixelDeconv - 3x3, 256
conv- 3x3, 512	decode3 - 3x3, 128	pixelDeconv - 3x3, 128
pixelDeconv - 3x3, 512	decode4 - 3x3, 64	conv - 1x1, 2
pixelDeconv - 3x3, 256	conv - 1x1, 2	
pixelDeconv - 3x3, 128		
conv - 1x1, 2		

3. Results

Overall results on the four accuracy measures, presented in Table 2, indicate that the best performing method was the Deeplabv3+ method, which gave percentages of 88, 94, 96, and 92 for the mean IU, frequency weighted IU, pixel accuracy, and mean accuracy. Images taken at 6 meters distance were classified only slightly less accurately, with a loss of only about 1% or less (Table 2, Figure 1).

Table 2: Results of the three re-implementations with batch size of 4 and the deeplabv 3+ trained on normal and flipped dataset compared to the implementation of reus [14].

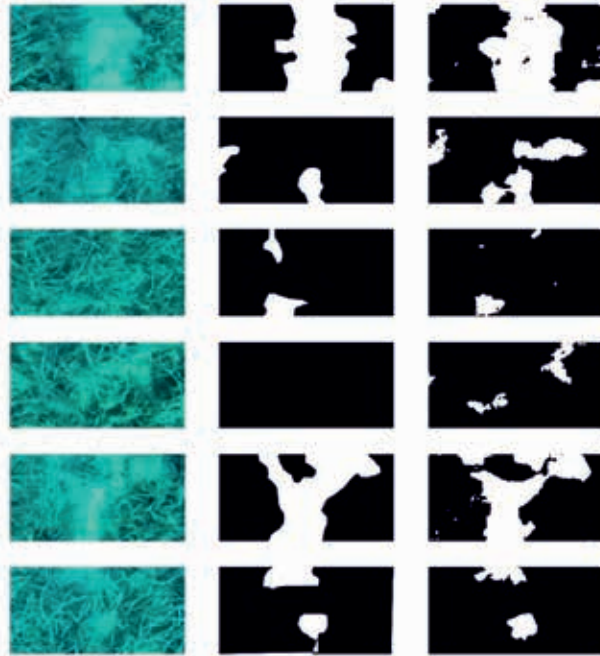| Method	mean IU	f.w. IU	pixel acc.	mean acc.
Reus - 2m	83.25%	90.93%	94.16%	88.48%
FCN - 2m, flipH	85.52%	91.49%	94.21%	92.19%
dilNet - 2m, flipH	85.52%	91.62%	94.37%	92.04%
uNet - 2m, flipH	85.03%	90.96%	93.67%	91.64%
Deeplabv3+ - 2m	86.23%	93.11%	95.85%	90.71%
Deeplabv3+ - 6m	86.97%	92.78%	95.67%	91.70%
Deeplabv3+ - 2m, flipH	**87.78%**	**93.93%**	**96.42%**	**92.37%**
Deeplabv3+ - 6m, flipH	87.33%	93.02%	95.85%	91.70%

Figure 1: Left original color image, midle the ground-truth image and right is the CNN prediction (black: Posidonia).

4. Discussion

Trainings for the 6-m images can be improved by doubling the number of training images simply by flipping each image horizontally. This method increased the classification accuracy for the 6-m images to almost indistinguishable from the 2-m images.

One difficulty of the present method was the annotating procedure: we annotated by simply drawing threshold regions over the image, with rough and variable boundaries (Figure 1). In reality seagrass images are defined by their leaf blades, which do not occupy sharply thresholded regions within the image. This approximation in the annotation process reduces the accuracy of the machine learning classification, causing false negative classifications where there are lighter colored regions in the image, where seagrass is nevertheless present (Figure 1). Future improvements to the method should consider more accurate methods for defining the boundaries of a seagrass area.

Our method classified images approximately 20 times faster than our previous method for seagrass identification, from 4 seconds to 0.15 second per image. This

speed is 5-10 times faster than manual human classification. The previous method used a logistic regression classifier that used rectangular patches within the image [14].

6. Conclusions

We investigated the success of four deep semantic segmentation neural network methods for estimating seagrass coverage in underwater images containing mixtures of seagrass, sand, algae, and assorted invertebrate animals. The Deeplabv3+ produced the most accurate seagrass covers, as compared to human annotators: a mean IU of 88% for 2-m images and 87% for 6-m images. These results are superior to previous methods using rectangular patch classification, and are 20 times faster than these previous methods and 5-10 times faster than human annotations. The method appears to be more accurate than human annotations, and can be used to guide human annotators.

Acknowledgements

References

1. BOUDOURESQUE C, BERNARD G, BONHOMME P, CARBONNEL E, DIVIACCO G, MEINESZ A, et al. *"Protection and conservation of Posidonia oceanica meadows"*. Monaco: Ramoge; 2012.
2. MARBA N, DIAZ-ALMELA E, DUARTE CM. *"Mediterranean seagrass (Posidonia oceanica) loss between 1842 and 2009"*. Biological Conservation. 2014 Aug; 176:183–190. doi: 10.1016/j.biocon.2014.05.024
3. SCHULTZ, S.T., KRUSCHEL, C., BAKRAN-PETRICIOLI, T. PETRICIOLI, D., 2015. *"Error, power, and blind sentinels: The statistics of seagrass monitoring"*. PloS one, *10*(9), p.e0138378.
4. A. KRIZHEVSKY, I. SUTSKEVER, G. E. HINTON, *"Imagenet classification with deep convolutional neural networks,"* in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/ 4824-imagenet-classification-with-deep-convolutional-neural-networks. Pdf
5. M. MASSOT-CAMPOS, G. OLIVER-CODINA, L. RUANO-AMENGUAL, M. MIR-JULI, *"Texture analysis of seabed images: Quantifying the presence of Posidonia oceanica at Palma Bay,"* in 2013 MTS/IEEE OCEANS -Bergen, June 2013, pp. 1–6.
6. A. BURGUERA, F. BONIN-FONT, J. L. LISANI, A. B. PETRO, G. OLIVER, *"Towards automatic visual sea grass detection in underwater areas of ecological interest,"* in 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Sept 2016, pp. 1–4.
7. F. BONIN-FONT, A. BURGUERA, J. L. LISANI, *"Visual discrimination and large area mapping of posidonia oceanica using a lightweight auv,"* IEEE Access, vol. PP, no. 99, pp. 1–1, 2017.

8.   M. MARTIN-ABADAL, E. GUERRERO-FONT, F. BONIN-FONT, Y. GONZALEZ-CID, *"Deep semantic segmentation in an AUV for online Posidonia oceanica meadows identification,"* IEEE Access, vol. 6, pp. 60 956– 60 967, 2018.
9.   L.-C. CHEN, Y. ZHU, G. PAPANDREOU, F. SCHROFF, H. ADAM, *"Encoder-decoder with atrous separable convolution for semantic image segmentation,"* in ECCV, 2018.
10. E. SHELHAMER, J. LONG, T. DARRELL, *"Fully convolutional networks for semantic segmentation,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651, April 2017.
11. F. YU AND V. KOLTUN, *"Multi-scale context aggregation by dilated convolutions,"* 11 2015.
12. O. RONNEBERGER, P. FISCHER, T. BROX, *"U-net: Convolutional networks for biomedical image segmentation,"* CoRR, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597
13. M. SANDLER, A. HOWARD, M. ZHU, A. ZHMOGINOV, AND L.-C. CHEN, *"Mobilenetv2: Inverted residuals and linear bottlenecks,"* in CVPR, 2018.
14. REUS, G., MÖLLER, T., JÄGER, J., SCHULTZ, S.T., KRUSCHEL, C., HASENAUER, J., WOLFF, V., FRICKE-NEUDERTH, K., 2018, May. *"Looking for Seagrass: Deep Learning for Visual Coverage Estimation"*. In *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)* (pp. 1-6). IEEE.