Semantic segmentation on 3D tomato seedling point cloud using deep learning

Weinan Shi Student[1]
weinan-shi@126.com
GW (Gert) Kootstra Assistant Prof[2]
http://gert.kootstra@wur.nl
Huanyu Jiang Prof[1]
hyjiang@zju.edu.cn

[1] Department of Biosystems Engineering and Food Technology
Zhejiang University
Hangzhou, China

[2] Department of Plant Sciences
Wageningen University and Research
Wageningen, The Netherlands

Abstract

Segmentation of 3D plant models is an important task in plant phenotyping and has been a great challenge in recent years. Segmenting plants into organs, such as leaves, stems and nodes can help us get plant-phenotypical data. Recently, deep learning has shown unparalleled results in semantic segmentation. In this work, we address the problem of using fully convolutional neural network (FCN) for semantic segmentation of 3D tomato-seedling models separating leaves, stems, and nodes. We approach the problem by performing semantic segmentation on multiple 2D images from different viewpoints and combining them into the 3D seedling models using the camera-projection matrices. We evaluate the precision and recall of our methods with ground truth labels obtained by hand. Results show that our method is very promising.

1 Introduction

Recently, applying deep learning in plant segmentation is gaining attention among researchers. In agriculture, deep learning is used in areas of weed identification, plant recognition, and fruit counting[3]. Fully convolutional neural network (FCN) is a pioneering field of research which can address the semantic image segmentation problem in pixel-level[8]. Segmentation of 3D models, which provides more objects information than 2D images, has been shown to be effective for vision applications. Algorithms of 3D segmentation using deeping learning fall into three groups. First, voxel-based methods[2, 4, 5, 7, 10, 13, 16] use voxelization of the 3D objects to create 3D tensors to feed a 3D convolutional neural network. Second, point-based methods[11, 12] take unordered point cloud as input and use network with fully-connected and pooling layers. Last, multi-view methods[1, 14, 15] apply neural networks to multiple generated 2D tensors of the point clouds and use CNN to back-project the label predictions to the 3D space. In this work, we present a semantic segmentation on multiple 2D images from different viewpoints and combine results into the 3D seedling model using the camera-projection matrices.

2 Materials and Methods

The main goal of our work is to segment a 3D point cloud seedling into leaves, stems and nodes in pixel-wise level and in an end-to-end way. Our segmentation pipeline is separated in three main steps: First, grey-scale 1280x960 images from MARVIN system[6] were manually annotated into four classes: background, leaf, stem and nodes. All images were cropped and resized into 640x400 pixel resolution. For the training and validation dataset, 270 images (87%) and 40 images (13%) were used, respectively. Second, a pretrained (PASCAL VOC) FCN-VGG16-8s model[9] was employed to provide a per-pixel semantic labeling of the input data. Last, based on the labeling, we used camera-projection matrices to project 2D pixel labels to 3D models. To evaluate the performance of the network, IoU, precision and recall for each class and average IoU (mIoU) for all classes were calculated.

3 Results and Conclusion

Based on the evaluation metrics, the performance of segmenting background, leaf and stem in 2D images was better than nodes (Table 1 and Figure 1). We will show the results of projecting the segmentation of multiple camera viewpoints to a 3D model in poster section.

Table 1: Results of semantic segmentation on 2D images

mIoU(%)	IoU(%)				Precision (%)				Recall(%)			
	Background	Leaf	Stem	Node	Background	Leaf	Stem	Node	Background	Leaf	Stem	Node
60.48	99.91	68.75	49.29	23.97	100.00	69.99	51.57	37.64	99.92	97.48	91.80	39.75

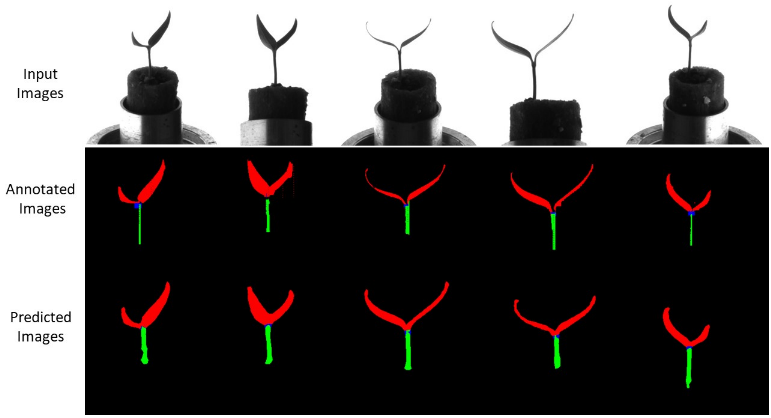

Figure 1: Input, annotated images, and predicted images of a seedling in ten pointviews.Red represents leaf, green represents stem and blue represents node

References

[1] Alexandre Boulch, Joris Guerry, Bertrand Le Saux, and Nicolas Audebert. Snapnet: 3d point cloud semantic labeling with 2d deep segmentation networks. *Computers and Graphics*, 2017. ISSN 00978493. doi: 10.1016/j.cag.2017.11.010.

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias NieÃ§ner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *arXiv preprint arXiv:1702.04405*, 2017.

[3] X Francesc. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 2018. ISSN 0168-1699.

[4] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D Wegner, Konrad Schindler, and Marc Pollefeys. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, 2017.

[5] Jing Huang and Suya You. Point cloud labeling using 3d convolutional neural network. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 2670–2675. IEEE. ISBN 1509048472.

[6] NJJP Koenderink, M Wigham, FBTF Golbach, G Otten, R Gerlich, and HJ van de Zedde. Marvin: high speed 3d imaging for seedling classification. In *Seventh European Conference on Precision Agriculture*, pages 279–286.

[7] Kevin Lai, Liefeng Bo, and Dieter Fox. Unsupervised feature learning for 3d scene labeling. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 3050–3057. IEEE. ISBN 1479936855.

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[9] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[10] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 922–928. IEEE. ISBN 1479999946.

[11] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593*, 2016.

[12] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.

[13] A Quadros, James Patrick Underwood, and Bertrand Douillard. An occlusion-aware feature for range images. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4428–4435. IEEE. ISBN 1467314056.

[14] Konstantinos Sfikas, Theoharis Theoharis, and Ioannis Pratikakis. Exploiting the panorama representation for convolutional neural network classification and retrieval. In *Eurographics Workshop on 3D Object Retrieval*.

[15] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953.

[16] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1912–1920.