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

Weinan Shi Student¹
weinan-shi@126.com

GW (Gert) Kootstra Assistant Prof²
http://gert.kootstra@wur.nl

Huanyu Jiang Prof¹
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 deep learning fall into three groups. First, voxel-based methods[4, 9, 10, 11, 12, 13] 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.

© 2018. The copyright of this document resides with its authors. It may be distributed unchanged freely in print or electronic forms.
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 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 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>
<thead>
<tr>
<th></th>
<th>Background</th>
<th>Leaf</th>
<th>Stem</th>
<th>Node</th>
<th>Background</th>
<th>Leaf</th>
<th>Stem</th>
<th>Node</th>
<th>Background</th>
<th>Leaf</th>
<th>Stem</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>mIoU (%)</td>
<td>60.48</td>
<td>99.91</td>
<td>68.75</td>
<td>49.29</td>
<td>23.97</td>
<td>100.00</td>
<td>69.99</td>
<td>51.57</td>
<td>37.64</td>
<td>99.92</td>
<td>97.48</td>
<td>91.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Background</th>
<th>Leaf</th>
<th>Stem</th>
<th>Node</th>
<th>Background</th>
<th>Leaf</th>
<th>Stem</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>IoU (%)</td>
<td>99.91</td>
<td>68.75</td>
<td>49.29</td>
<td>23.97</td>
<td>100.00</td>
<td>69.99</td>
<td>51.57</td>
<td>37.64</td>
</tr>
<tr>
<td>Precision (%)</td>
<td>99.92</td>
<td>97.48</td>
<td>91.80</td>
<td>39.75</td>
<td>99.92</td>
<td>97.48</td>
<td>91.80</td>
<td>39.75</td>
</tr>
</tbody>
</table>

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


