Overcoming Small Minirhizotron Datasets Using Transfer Learning

Weihuang Xu, Guohao Yu, Alina Zare, Brendan Zurweller, Diane Rowland, Joel Reyes-Cabrera, Felix B Fritschi, Roser Matamala, Thomas E. Juenger

1Department of Electrical and Computer Engineering, University of Florida, FL, USA
2Department of Plant and Soil Science, Mississippi State University, MS, USA
3Department of Agronomy, University of Florida, FL, USA
4Division of Plant Sciences, University of Missouri, MO, USA
5Argonne National Laboratory, IL, USA
6Department of Integrative Biology, University of Texas at Austin, TX, USA

weihuang.xu@ufl.edu, azare@ufl.edu

Abstract

Minirhizotron technology is widely used for studying the development of roots. Such systems collect visible-wavelength color imagery of plant roots in-situ by scanning an imaging system within a clear tube driven into the soil. Automated analysis of root systems could facilitate new scientific discoveries that would be critical to address the world’s pressing food, resource, and climate issues. A key component of automated analysis of plant roots from imagery is the automated pixel-level segmentation of roots from their surrounding soil. Supervised learning techniques appear to be an appropriate tool for the challenge due to varying local soil and root conditions, however, lack of enough annotated training data is a major limitation due to the error-prone and time-consuming manually labeling process. In this paper, we investigate the use of deep neural networks based on the U-net architecture for automated, precise pixel-wise root segmentation in minirhizotron imagery. We compiled two minirhizotron image datasets to accomplish this study: one with 17,550 peanut root images and another with 28 switchgrass root images. Both datasets were paired with manually labeled ground truth masks. We trained three neural networks with different architectures on the larger peanut root dataset to explore the effect of the neural network depth on segmentation performance. To tackle the more limited switchgrass root dataset, we showed that models initialized with features pre-trained on the peanut dataset and then fine-tuned on the switchgrass dataset can improve segmentation performance significantly. We obtained 99% segmentation accuracy in switchgrass imagery using only 21 training images. We also observed that features pre-trained on a closely related but relatively moderate size dataset like our peanut dataset are more effective than features pre-trained on the large but unrelated ImageNet dataset.

1. Introduction

Minirhizotron camera systems are a minimally-invasive imaging technology for monitoring and understanding the development of plant root systems[14]. A variety of root phenotypes can be determined from minirhizotron RGB root imagery, such as lengths, diameters, patterns and distributions at different depths. However, manual tracing roots in minirhizotron imagery is tedious and extremely time-consuming, which limits the number and size of experiments. Thus, techniques that can automatically and accurately segment roots from minirhizotron imagery are crucial to improve the efficiency of data collection and post-processing.

Semantic image segmentation is one of the most challenging tasks in computer vision. Instead of assigning labels at the whole-image level for image classification problems, semantic image segmentation requires a model to predict a label for each pixel. Many methods based on deep convolutional neural networks (DCNN) have been proposed to address semantic segmentation tasks such as fully convolutional networks[13], SegNet[1], U-net[18], and deeplab[6]. Models based on the above methods have achieved success in segmentation of medical images[18, 15, 4, 23], satellite images[7, 16, 2], and plant images[5, 24]. Such segmentation models are based on supervised learning of large networks with a very large number of parameters, requiring a huge amount of data with ground truth to achieve a satisfactory performance. Deep neural networks trained on small datasets can quickly overfit for the small sets and perform poorly in larger applications. Thus, a fundamental issue of
using those models for many applications, including plant science, is limited availability of training data.

To address such problems, so-called transfer learning techniques have been developed that apply model-weights pre-trained on large-scale data as initial parameters, and then they fine-tune the models on target problems that usually have more limited training data. This process will work based on the assumption that those pre-trained features are fairly general and applicable to many visual image applications, and can be re-used for a different specific problem. When target dataset is small, pre-trained features can significantly improve the performance and help with faster convergence. Leveraging this idea, features pre-trained on massive scale data such as ImageNet are widely used as initial weights in recent work, which achieved state-of-the-art results on a variety of different tasks, such as image classification, object detection and image segmentation. However, more and more work is questioning the effects of ImageNet pre-training. Yosinski et al. showed that features in shallow layers are more general and effective when transferred to other specific problem. On the contrary, features from higher layers are more problem-specific. Huh et al. illustrated that transfer learning performance is similar with features pre-trained only on half of ImageNet dataset as opposed to the full dataset. Thus, how to balance the application of available large-scale but less problem-specific data with very limited but more relevant data for pre-training is a topic needing exploration.

In this work, we collected a moderately sized peanut root minirhizotron imagery dataset and a small sized switchgrass root minirhizotron imagery dataset and manually traced root segments in both sets. We trained U-net based models with different depths on the peanut root dataset to achieve automated, precise pixel-wise root segmentation. We also investigated and compared the effect on segmentation performance of model depth. Furthermore, we used a transfer learning approach to apply pre-trained features from the peanut root data and the ImageNet dataset on the small-scale switchgrass root dataset to achieve high segmentation accuracy. We also found that features pre-trained on a moderate-sized dataset that was highly related to the target dataset were more effective than the large-scale but less relevant data.

In the following sections we describe our datasets, the semantic segmentation methods employed, our experiments using those methods with our datasets and other visual imagery datasets, and finally draw some conclusions based upon that work.

2. Dataset

We have compiled two minirhizotron root image datasets. The first dataset contains 17,550 peanut root RGB images and the second dataset has 28 switchgrass root RGB images. All the images in both datasets were acquired using minirhizotron systems in the field, and paired with manually labeled ground truth masks indicating the location of roots in each image. The details of data collection and labelling process are as follows.

**Peanut root dataset** was collected in a field trial at the Plant Science Research and Education Unit (PSREU) during the 2016 growing season. Minirhizotron tubes 2 m in length were installed directly under and parallel to the row at a 45° angle to the soil surface after crop emergence using a hydraulic powered coring machine (Giddings Machine Company, Windsor, CO). After installation the portion of the minirhizotron tube was covered with reflectance insula-
tion (Reflectix Inc., Markleville, IN) to avoid root UV exposure and precipitation from entering the tube. At each measurement date, images were captured at 13.5 mm increments (resulting typically in 112 image frames) along the minirhizotron tubes using a BTC 100X video camera and BTC I-CAP image capture software (Bartz Technology Corporation, Carpinteria, CA). Root parameter analysis was conducted using WinRHIZO Tron software (Rgent Instruments Inc., Quebec, Canada) by hand tracing root segments within each image frame. The binary ground truth masks were generated by hand using the WinRHIZO Tron software package. The process consists of manually drawing different sizes of rectangles to highlight the area of roots while attempting to leave the soil blank. Examples of collected peanut root images and corresponding labeled ground truth masks are shown in Figure 1. Labelling by WinRHIZO is faster than labelling images pixel by pixel, since a large area of root pixels can be labeled at once. However, two shortcomings of this method are that: 1) the width of each rectangle is constant indicating that the roots in the labeled area have the same diameter, which is not true in practice; and 2) when the roots are not straight, there are gaps between labeled regions and the labeled edges of roots are not smooth.

Switchgrass root dataset was collected using a CI-602 in-situ root imager (CID Bio- Science, Camas, WA, USA) in minirhizotron tubes in a 2-year old switchgrass field at the U.S. Department of Energy National Environmental Research Park at Fermilab in Batavia, IL, USA. Minirhizotron tubes were installed with an angle of $60^\circ$ to an approximate maximum vertical depth of 120 cm using an angled guided soil core sampler. Foam caps were installed over the top end to protect the tubes from UV damage. Root images were taken at 300 dots per inch (11.8 pixels per millimeter) from eight depth intervals along minirhizotron tubes. All the binary ground truth masks were manually labeled pixel by pixel. Manual labeling of the imagery took the annotator approximately 2 hours per image. Examples of raw switchgrass images and corresponding ground truth masks are shown in Figure 2.

### 3. Methods

A U-Net[18] based encoder-decoder neural network was used for root segmentation. The network architecture is shown in Figure 3. The left half of the architecture works as an encoder where each block consists of two 3x3 convolution layers followed by one 2x2 max-pooling layer to down sample feature maps. The right half of the architecture works as a decoder where each block consists of one transpose convolution layer and two 3x3 convolution layers. The transpose convolution layer up-samples the size of feature maps by two. The encoder blocks are trained to extract dense feature maps from minirhizotron RGB imagery. Via skip connections, those feature maps will be concatenated with higher-level ones in corresponding decoders to offer more spatial information in output mask. The last layer is a 1x1 convolution layer (i.e., a weighted sum across all feature layers) to convert feature maps to a heat map. Then, the softmax function is used to assign class labels to each pixel. As it uses fully convolutional network architecture [13], a U-Net can be trained end-to-end with input images of any size. In order to keep the dimension of the output segmentation mask to be the same as input images, zero padding is used in every convolution layer. The model was implemented using the Pytorch library and trained on a GTX 1080TI GPU with 12GB of RAM. The receiver operating characteristic (ROC) curve is plotted along with the value of area under the curve (AUC) measure to evaluate segmentation performance. True positive rate (TPR) and false positive rate (FPR) are calculated at the pixel-level by comparing the output mask to the manually labeled ground truth of the test data.

### 4. Experiments and Results

To investigate the segmentation performance of our model on minirhizotron imagery and the influence of network depth on root segmentation performance, we implemented three models with depth 4, depth 5 and depth 6, where model depth refers to the number of encoders in the down-sampling path. The peanut root dataset was used for training due to its larger size. We used 90% of the images for training and the remaining 10% for testing. The inputs
to the model were entire images (instead of stacking small randomly selected patches). We set the batch size to be two and used binary cross-entropy as loss function. All models were trained for 100 epochs with randomly initialized parameters.

We designed experiments on the small-scale switchgrass root dataset to explore the effect of pre-trained features from popular massive-scale ImageNet dataset and our own peanut root dataset. Compared with the ImageNet dataset that has 14 million images, our peanut root dataset is quite small, but much more relevant to the switchgrass root dataset. As the goal in this experiment is to figure out the role of pre-trained features on a general model instead of finding the highest performing network architecture, we implemented the U-net based model with down path architecture the same as the VGG13 network[21]. Besides the features in the encoder, decoder blocks also extract higher-level feature maps for up-sampling. We believe those feature maps are also crucial for improving segmentation performance as in the limited switchgrass root dataset. Thus, we studied pre-trained features not only in encoder, but also in the combination of encoder and decoder. To make a comprehensive comparison of different pre-trained features, we implemented four models named: 1) S-model, whose weights are randomly initialized; 2) I-model whose encoder is initialized with pre-trained weights on ImageNet dataset; 3) P-model whose encoder is initialized with pre-trained weights on our peanut dataset; and 4) P-modelV2 whose encoder and decoder are initialized with pre-trained weights on our peanut dataset. These four models have exactly the same architecture but different weight initialization. All models were fine tuned on the switchgrass dataset for 100 epochs using the same learning rate. Since randomly initialized weights can cause variance in segmentation results, each model was trained five times to compare the performance consistency. We used 75% of the switchgrass minirhizotron images for training and the remaining 25% for evaluating the performance of each model. Due to the limitation of GPU memory, each switchgrass root image is evenly cropped into 15 small images with size 720x510 pixels with a batch size of two.

### 4.1. Segmentation Performance on Peanut Root Dataset

The segmentation results of the three models with different depths are shown in Figure 4. Column (a) shows the raw peanut root images taken from the minirhizotron system. These images were taken across differing depths,
dates, and local environments. The corresponding manually labeled ground truth masks are shown in column (b). Column (c)-(e), show segmentation masks of our models with depth 4, depth 5 and depth 6, respectively. Qualitatively, all three models provided good segmentation performance. Most of the roots can be segmented from complicated soil backgrounds. Our method solved two major issues caused by the manual labelling process using WinRHIZO software, those of fixed label width along rectangles, and gaps between neighboring rectangles. Our model can capture the real thickness and diameter variation along each root. The segmented roots are consistent and smooth instead of having a boxy shape with gaps in manually labeled ground truth masks. As shown in column (c)-(e), our segmentation masks have a better representation of roots, which can with accurate determination of root traits in subsequent measurements.

Some interesting details were observed in the segmentation results in the last three rows in Figure 4. In the third row, part of the top of the root is covered by soil in the raw peanut root image. An example of the robustness of our method is that a small area of the ground truth mask was mislabeled; however, our method obtained the correct answer in all three models. In the fourth row, the root at the bottom left of the picture is partially covered by soil. We expected that it would be considered as a single piece of root
Figure 5. ROC curves plotted for models with depth 4 (Blue), depth 5 (Orange), and depth 6 (Green), respectively. Each model was trained five times with random weight initialization.

as it was labeled as such in the ground truth mask. The shallow model (depth 4) generated three separated small pieces of root instead of one unbroken piece. In contrast, deeper models (depth 5 and depth 6) were capable of filling the gap and generated an unbroken root. This ability is very important when considering the density or number of roots in a specific area. The last row shows the case of a very complicated background. Because of a wide variety of reflections, it is very difficult to eliminate water bubbles in segmentation results. Our method was able to remove most of the water bubbles, but there still was some residual noise in the segmentation results. The output masks from depth 6 model are much cleaner than depth 5 and depth 4 models, which indicates that the deeper network is more powerful to accommodate complex noise in order to match ground truth masks as close as possible. This is reasonable, because a deeper network can extract higher-level features to further improve the reconstruction step in decoders.

In order to evaluate the consistency of the model, we trained each model 100 epochs for five trials with different random weight initialization. We calculate the TPR and FPR using the entire test dataset containing 0.7 billion pixels. The ROC curves for each model are shown in Figure 5. The average and STD of AUC for each model are shown in Table 1. The depth 6 model had the highest average AUC of 0.9904 indicating the best segmentation accuracy among all the models. Additionally, the method showed good consistency as all three models had small variance in AUC.

### 4.2. Transfer Learning on Limited Switchgrass Root Dataset

Figure 6 shows the ROC curves evaluated on the switchgrass root test set. The average and standard deviation value of the AUC is shown in Table 2. Under the same training condition, pre-trained features improve the segmentation performance substantially. Specifically, features pre-trained on the peanut root dataset were more effective than features pre-trained on ImageNet dataset, even thought the ImageNet dataset is 1000 times larger than peanut dataset. Also, models used with the peanut-pre-trained encoder were more consistent (lower standard deviation of AUC) than models with ImageNet encoder. Furthermore, P-modelV2 had the highest average AUC value indicating that pre-
Figure 7. Heat maps generated by models with different pre-trained features. (a) Raw switchgrass root image. (b) Manually labeled ground truth mask. (c) Heat map of model with encoder pre-trained on ImageNet. (d) Heat map of model with encoder pre-trained on peanut root dataset. (e) Heat map of model with encoder and decoder pre-trained on peanut root dataset.

Figure 8. Learning curve for models with different weight initialization. Blue curve: S-model, weights in both encoder and decoder are randomly initialized. Orange curve: I-model, weights in encoder are initialized with ImageNet pre-trained features and randomly initialized weights in decoder. Green curve: P-model, weights in encoder are initialized with peanut pre-trained features and randomly initialized weights in decoder. Red curve: P-modelV2, weights in both encoder and decoder are initialized with peanut pre-trained features.

Our results show that (perhaps, intuitively) pre-trained features from a massive-scale dataset are not always the best pre-trained features for imagery with different visual appearances. Although the size of a pre-trained dataset is important, it appears that the relevance of the pre-trained and the target datasets is more crucial for segmentation performance. If those two datasets are very different, only the features in shallow layers can help with segmentation results, since low-level features are more general such as edges or textures[22]. The higher-level features are problem-specific, which could mislead the decision of a model on the target dataset. This is even worse when the model is deeper, because the proportion of effective parameters in shallow layers is getting smaller. In contrast, features from a small-scale but highly related to the target dataset are more valuable regardless of how deep the model is, because both low-level and high-level features are useful to the target dataset.

5. Conclusions

In this work, we propose the use of U-net based deep neural networks for automated, precise, pixel-wise segmentation of plant roots in minirhizotron imagery. Our model achieved high quality segmentation masks with 99.4% accuracy at the pixel-level and overcame errors in human-labeled ground truth masks. We also found that deep networks can better resolve more challenging images (more complicated backgrounds) than shallow networks. Furthermore, we improved the segmentation performance on a small-scale switchgrass root dataset by using pre-trained features from the massive-scale ImageNet dataset and a mid-scale peanut root dataset, then fine tuning on a small switchgrass root dataset. We obtained above 99% segmentation accuracy in switchgrass root segmentation with pre-trained encoder and decoder.
trained encoder and decoder from our peanut root dataset. Our results indicate that both pre-trained encoder and decoder can help with segmentation performance when the target dataset is small. Also, features pre-trained from the dataset that is relatively small but highly related to the target dataset are more effective than the massive-scale but less relevant dataset.

Acknowledgement

This work was supported by U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research award number DE-SC0014156 and by the Advanced Research Projects Agency - Energy award number DE-AR0000820.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, 2015. 1
[2] Y. Bai, E. Mas, and S. Koshimura. Towards operational satellite-based damage-mapping using u-net convolutional network: A case study of 2011 tohoku earthquake-tsunami. Remote Sensing, 10(10):1626, 2018. 1
[3] Y. Bengio. Deep learning of representations for unsupervised and transfer learning. In Proceedings of ICML Workshop on Unsupervised and Transfer Learning, pages 17–36, 2012. 2
[4] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng. Dcan: Deep contour-aware networks for object instance segmentation from histology images. Medical image analysis, 36:135–146, 2017. 1
[5] J. Chen, Y. Fan, T. Wang, C. Zhang, Z. Qiu, and Y. He. Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. Agronomy, 8(8):129, 2018. 1, 2
[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4):834–848, 2018. 1
[7] A. Constantin, J.-J. Ding, and Y.-C. Lee. Accurate road detection from satellite images using u-net model. In 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), pages 423–426. IEEE, 2018. 1
[8] J. Dai, K. He, and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3150–3158, 2016. 2
[9] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In International conference on machine learning, pages 647–655, 2014. 2
[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014. 2
[11] M. Huh, P. Agrawal, and A. A. Efros. What makes imagenet good for transfer learning? arXiv preprint arXiv:1608.08614, 2016. 2
[12] V. Iglovikov and A. Shvets. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv preprint arXiv:1801.05746, 2018. 2
[13] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015. 1, 3
[14] H. Majdi. Root sampling methods-applications and limitations of the minirhizotron technique. Plant and Soil, 185(2):255–258, 1996. 1
[15] R. K. Pandey, A. Vasan, and A. Ramakrishnan. Segmentation of liver lesions with reduced complexity deep models. arXiv preprint arXiv:1805.09233, 2018. 1
[16] A. Rakhlin, A. Davydow, and S. Nikolenko. Land cover classification from satellite imagery with u-net and lovász-softmax loss. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 257–2574. IEEE, 2018. 1
[17] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015. 2
[18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015. 1, 3
[19] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013. 2
[20] A. Sharif Razavian, H. Azizzpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 806–813, 2014. 2
[21] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. 4
[22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In Advances in neural information processing systems, pages 3320–3328, 2014. 2, 7
[23] Y. Zhang, J. Wu, W. Chen, Y. Chen, and X. Tang. Prostate segmentation using z-net. arXiv preprint arXiv:1901.06115, 2019. 1
[24] Y. Zhu, M. Aoun, M. Krijn, J. Vanschoren, and H. T. Campus. Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants. Computer Vision Problems in Plant Phenotyping (CVPPP2018), 2018. 1