

DeepXScope: Segmenting Microscopy Images with a Deep Neural Network

Philip Saponaro^{1,2} Wayne Treible¹ Abhishek Kolagunda¹ Timothy Chaya³
 Jeffrey Caplan³ Chandra Kambhamettu¹ Randall Wisser²

¹Department of Computer and Information Sciences, University of Delaware, Newark, DE

²Department of Plant and Soil Sciences, University of Delaware, Newark, DE

³Delaware Biotechnology Institute, University of Delaware, Newark, DE

{saponaro,wtreible,abhi,tchaya,jcaplan,chandrak,rjw}@udel.edu

Abstract

High-speed confocal microscopy has shown great promise to yield insights into plant-fungal interactions by allowing for large volumes of leaf tissue to be imaged at high magnification. Currently, segmentation is performed either manually, which is infeasible for large amounts of data, or by developing separate algorithms to extract individual features within the image data. In this work, we propose the use of a single deep convolutional neural network architecture dubbed DeepXScope for automatically segmenting hyphal networks of the fungal pathogen and cell boundaries and stomata of the host plant. DeepXScope is trained on manually annotated images created for each of these structures. We describe experiments that show each individual structure can be accurately extracted automatically using DeepXScope. We anticipate that plant scientists will be able to use this network to automatically extract multiple structures of interest, and we plan to release our tool to the community¹.

1. Introduction

1.1. Background of Maize Imaging

Microscopic phenotypes extracted from image data on plant-pathogen interactions provides a basis for studying mechanisms of pathogenesis. Recently, a platform to systematically capture three-dimensional (3D) image data across large areas of maize leaf tissue infected with fungal pathogens was developed [3].

In many microscopy applications, fluorescent stains are used that bind to specific molecules constituting biological structures. Imaging deep into the tissue of plants and

other organisms to obtain 3D image data has been facilitated by recent developments in optical clearing [1, 6, 3] combined optical clearing with fluorescent confocal microscopy to image fungal infected plant tissues where the host and pathogen cells were differentiated by stains that fluoresce at distinct wavelengths. However, gathering enough image data to make statistical inference requires large-scale microscopy imaging. To address this, Minker et al. [3] also developed "Macroscopic Microscopy," an approach where images from partially overlapping fields of view are stitched together into one large, high-resolution image of the original tissue specimen. This methodology was used to capture 3D image stacks of maize leaf tissue with each stack comprised of 150 slices at $1.2\mu\text{m}$ resolution with approximately 36mm^2 surface area per slice.

1.2. Image Processing

Capturing large 3D areas of a numerous tissue samples requires a very large amount of data storage. Traditional image processing methods are not effective on this data due to low amounts of contrast between desired features and the background [5], as well as differing intensities, orientations, and contrasts between samples.

With the goal of segmenting biological structures and features from the data, methods were originally developed using a suite of automatic thresholding schemes, template matching, active contours, and shape fitting. However, many of these methods would only work on small subsets of samples and needed to be consistently fine-tuned when new datasets were captured on the microscope.

We propose a new method of adapting a deep convolutional neural network (CNN) for segmenting the maize leaf image data in order to consistently and accurately segment biological features needed for phenotyping pathogen resistance in maize at a microscopy level. A single deep

¹Code will be made available at <http://www.drmaize.org/>

CNN can be trained to segment multiple structures of interest using annotated training (ground truth) data. This network is applied to segment fungal hyphae networks, host cell boundaries, and host stomatal structures.

1.3. Paper Outline

In the Section 2, we describe the CNN architecture and the training methodology that was used in the segmentation scheme. We then present the experiments and results in Section 3 for each of the segmented biological structures. Finally, we close with a conclusion and future work in Section 4.

2. Method

Traditional image processing methods are insufficient at solving many of the problems that arise in data collected by Minker et al. [3]. Hand-picked features and ad-hoc methods not only perform poorly, but do not generalize to maintain a unified workflow. Therefore, we developed a single workflow that uses the same pipeline of training and applying our CNN (Section 2.1). Some steps are slightly modified depending on the type of training data input and the desired output. For instance, surface features, including stomata, and cell boundaries use an estimated leaf surface image [2], while hyphal networks use the entire leaf image stack. Each training scheme is also slightly different as described in Section 2.2.

2.1. Network Description

We adapted the U-net architecture from Ronneberger et al. [4] because of its success in segmenting features in biomedical images, which often exhibit the same difficulties for segmentation as our microscopy data. The U-Net architecture has an initial contracting stage to capture context and a mirrored expansion stage for localization. This architecture is particularly useful in our application because it requires relatively few training samples to get acceptable segmentations, and the segmentation is robust to close or touching objects in the same class.

As shown in Figure 1, we maintain a similar structure to that of the Unet architecture but with a slightly more shallow network depth and different numbers of filters at each layer. We also chose to implement the architecture in Keras using Tensorflow as opposed to the original Caffe implementation in order to easily fit into our current deep learning workflow.

2.2. Training Schemes

We trained a separate instance of one network architecture for each biological structure of interest. That is, the number of nodes and layers remained the same for each structure, but different hyper-parameters were used when

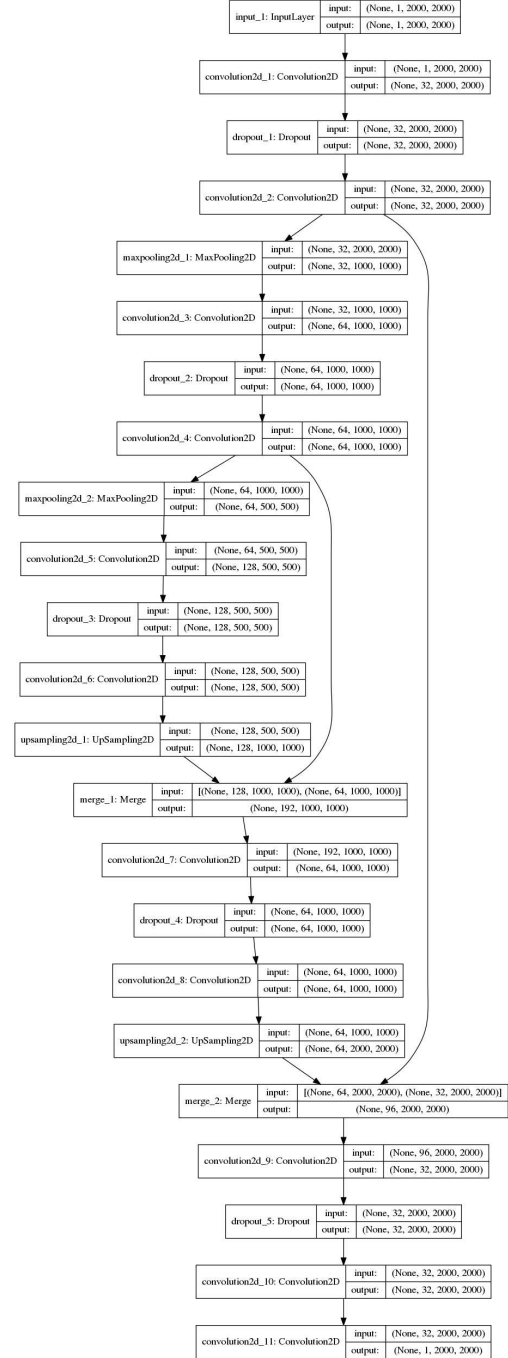


Figure 1. Our adaptation of the U-net architecture with input and output shapes. Each convolution layer is followed by a non-linear activation function (ReLU). In this graph, the input was a grayscale 2000x2000 surface image. For the input shape, the dimensions are as follows: (ImageBatchSize, ImageChannels, ImageY, ImageX) For intermediate convolution shapes, the dimensions are: (ImageBatchSize, NumberOfFilters, ImageY, ImageX). In both instances, batch size is reported as "None" because it is variable to the network.

training. For each of the structures, ground truth data was supplied as patches of 2D data. 3D fungal hyphae structures were segmented by applying 2D segmentation for each Z-slice without a 3D constraint. The batch size varied depending on the amount of training data for each structure, and is given in each corresponding subsection.

2.2.1 Cell Boundaries

For cell boundaries, training datasets were created by manually labeling images and generating masks. Cell boundaries were manually labeled in 60 48x48 image patches from surface images. A batch size of 50 was used. The CNN for cell boundary segmentation did not apply a final softmax layer for normalization, but instead was normalized and thresholded in a post-processing step.

2.2.2 Fungal Hyphae

Due to the three dimensional nature of hyphae in the leaf tissue, as well as blurring between layers, it was very difficult to create manually annotated training data. Because of this limitation, a semi-automatic synthetic fungus generation scheme was developed to create training data. Synthetic hyphal networks were generated in manually chosen areas of the leaf samples that were free of fungus. These synthetic volumes were then blurred to simulate intensity changes across layers in real data. For more details on this process, see [5]. This process generated synthetic data used to train the CNN. 2D patches were extracted from the training volume and segmentation was applied on each Z-slice in the 3D stack. A batch size of 100 was used.

2.2.3 Stomata

Training data for stomata was generated by a semi-automatic method. Approximately 30 stomatal locations were manually identified per sample. A 25x25 image patch was cropped around the identified stomatal locations and then used as templates to identify more stomates on the leaf's surface image. A threshold of 0.75 on the normalized template matching score was used to facilitate identification of additional stomatal locations. A circular disk with a radius of 15 pixels centered around each stomate was used as a mask. The images and the corresponding masks were divided into non overlapping patches of size 48x48. Image patches that had at least 75 pixels set in the corresponding patch from the mask were used to train the network. 1,391 image patches containing 1,107 stomata were used for training. A batch size of 128 was used, which is more than the other structures due to the amount of training samples.

The network used for segmenting stomates included a final softmax layer to normalize the final response.

3. Experiments and Results

In this section we describe our experiments to test each subsystem – fungus, cells, stomata – all using the same CNN architecture for segmentation.

3.1. Cell Boundary Segmentation

The cell boundaries segmented for quantification were taken from 2D surface images from the initial 3D image stack [2]. Due to the enormous number of cells in a single surface image, quantification was done on 36 manually segmented 48x48 image patches from six different surface images. An example of the patches generated and their CNN responses is shown in 4

Cell boundaries for training data were labeled at single pixel resolution, so quantification metrics leveraged a leniency threshold that allowed for up to a 2 pixel radius in any direction to account for boundary thickness. Figure 3 gives the precision recall curve generated using the 36 image patches.

The network was also designed to run on the entire surface image. An example of the CNN applied to a surface image can be found in Figure 2. The CNN segmentation is fast to apply, taking on average 4.69s per 2000x2000 cropped surface image on a NVidia Titan X (Pascal).

3.2. Fungal Hyphae Segmentation

Fungal hyphae networks were present in 3D images with a resolution of 2357x2366x150. Manually annotating 3D ground truth is not feasible to do, as even single slices of the stack can take hours to annotate. To give a quantitative analysis, we manually annotate maximum intensity projections (MIPs), which are 2D projections where the intensity at each X,Y coordinate is the maximum intensity along the Z dimension. We manually annotated 11 samples from low, medium, and high fungal growth image stacks for a total of 351,059 positive pixels and 60,597,609 negative pixels. A summary of these statistics is in Table 1.

Table 1. Number of positive pixels and number of connected fungal networks in samples containing low, medium, and high levels of fungal infection.

Infect Lvl	Pos Px	# Nets
Low	9383	181
Medium	128991	694
High	212685	261
Overall	351059	1136

Using the manually annotated MIPs, we calculated three metrics. The first metric was based on precision and recall. The Equal Error Rate (EER) is the threshold that causes precision and recall to be equal, where a value closer to 1.0 is better. Another way to combine precision and recall is the F1 score, which is the harmonic mean of the precision and

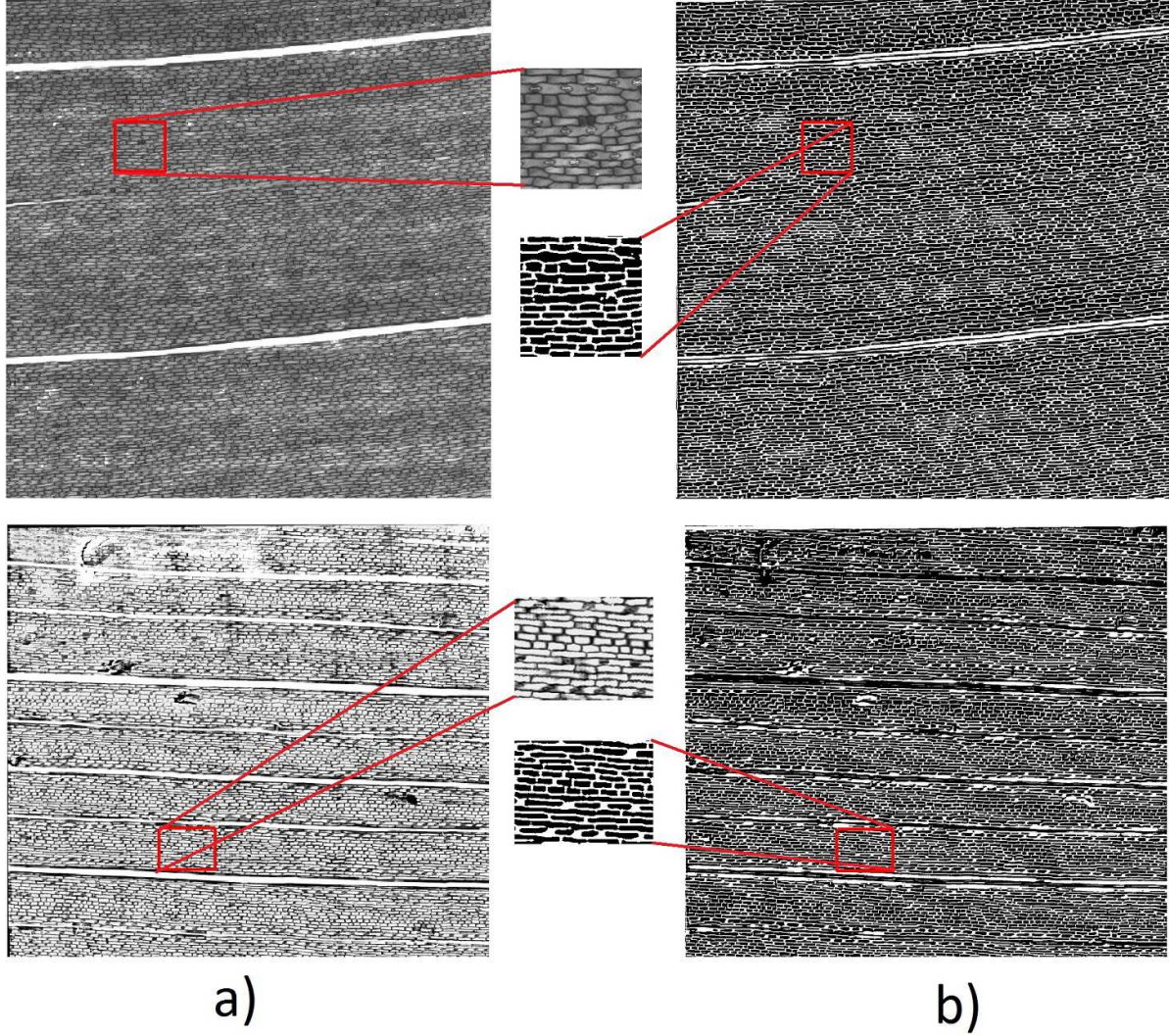


Figure 2. Example of cell boundary segmentation of entire surface images from different samples. **a)** Raw imagery **b)** Thresholded CNN response.

recall defined by $f = 2 * \frac{P * R}{P + R}$. This metric is a single measure of the performance of the CNN for fungal hyphae detection, and can represent its performance even given the sparsity of the positive pixels.

Due to the fact that manually segmented pixels and automatically segmented pixels can have different line thickness, we allow a leniency threshold of 10 pixels. That is, if the proposed positive pixel is within a radius to the ground truth defined by the leniency threshold, the pixel is marked correct. Without this leniency threshold, only pixel perfect segmentations to the manually marked ground truth are marked as correct, which gives a misleading view of the performance. Figure 6 gives a precision-recall curve on our dataset for fungal hyphae segmentation and Table 2 gives tabular results across different levels of fungal infections.

The second metric we calculate is the number of con-

Table 2. Precision, recall, and F1score of the deep CNN for various levels of fungal infection with a leniency threshold of 10px.

Infect Lvl	Prec	Recall	F1
Low	0.62	0.83	0.68
Medium	0.71	0.97	0.81
High	0.89	0.99	0.94
Overall	0.712	0.93	0.79

nected fungal hyphae networks. Since most fungal hyphae networks did not overlap in the Z dimension, the MIPs provide a reasonable estimate of the number of total fungal hyphae networks. Accurately measuring the number of fungal networks is an important part of quantifying fungal infection. Moreover, it is possible to have a high F1-score but a poor estimate of the number of connected fungal hyphae.

To measure performance in terms of the number of con-

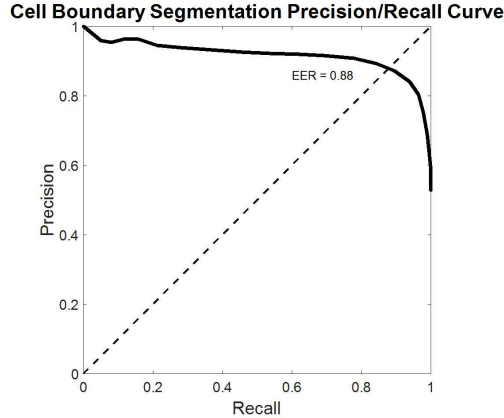


Figure 3. Precision-Recall curve generated by varying the threshold on the response from the deep CNN applied to surface image patches for cell boundary detection.

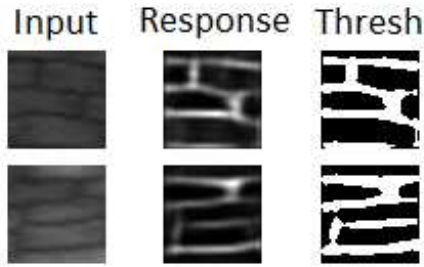


Figure 4. Example of two 48x48 patches from two separate leaf surface images used in the testing set for cell boundary segmentation. Left: Input patch, Center: CNN Response image, Right: Automatic threshold image

nected hyphae, we calculated the percentage of the number of detected networks from the CNN segmentation relative to the MIP ground truth data. We calculated this with and without applying the minimum spanning tree (MST) connection algorithm described by Saponaro et al. [5]. Experimentally, we found 75 pixels to be the highest performing threshold for the MST connection algorithm. We note that the MST algorithm does not significantly affect the recall, and reduces the precision by 3%. The results with this metric are summarized in Table 3.

Table 3. Percent increase of number of fungal networks from automatically segmented to ground truth across low, medium, and high levels of infection. The first row is without using the MST algorithm, and the second row is using it with an experimentally determined optimal gap threshold.

Gap Threshold	Low	Medium	High	Overall
N/A	47.13	341.53	1011.28	456.18
75	10.92	19.04	-18.04	8.85

The final metric we calculated was the distribution of pixel intensity. Figure 7 shows this as a box and whisker

plot for the positive ground truth pixels, positive CNN pixels, and the negative ground truth pixels. Note the whiskers of the plot represent a standard deviation of 2.7 for 99.3% coverage of the data.

Additionally, we provide a qualitative comparison with representative samples taken from the low, medium, and high infection image stacks in Figure 5.

From the intensity distributions in Figure 7, 75% of the negative pixels fall beneath 75% of the positive pixels, and the boxes do not overlap. However, 25% of the negative pixel intensities overlap with the box of the positive samples, which suggests simpler methods such as thresholding would give higher errors than our schema, which was also supported by Saponaro et al. [5]. The intensity distribution for the CNN result overlaps with that of the positive ground truth distribution, with the median of the CNN corresponding to the lower quartile of the ground truth (at 26%). On the other hand, the medians of intensity distributions for the CNN and negative ground truth do not overlap. These results further support the ability of the CNN to detect the targeted feature in the image data.

For connected components, Table 3 shows that the MST connection algorithm from Saponaro et al. [5] is mandatory to achieve a reasonable fungal hyphae network count. Without it, small gaps in the segmentation led to a large over-estimation of the number of hyphal networks. This could be explained by the intensity distributions as above – 25% of the negative pixel intensities overlapped with the positive pixel intensities.

In Table 2, the F1-score is maximal for the high infection levels and is lowest for low infection levels. One explanation of this is that the CNN was trained on data that more closely resembles full fungal networks, whereas the low infection level fungus takes on different shapes/sizes. One way to possibly improve the performance is to classify the infection level and segment each infection level with a differently trained network.

An additional note is that lowering the leniency threshold to 5 pixels lowers the F1-score by 5% overall, and lowering the threshold to 3 pixels lowers the F1-score by 15% overall.

3.3. Stomata Segmentation

Given 2D fluorescence images of the leaf surface, we manually annotated the locations of 1,834 stomata for testing (apart from the stomatal locations used in training). The whole leaf surface image was given as input to the network and the response image from the network was thresholded to compare against the ground truth. A true positive is when the location of the ground truth stomata lies within a connected component of the thresholded response image. With this, we calculated the precision, recall, and F1-score. Note that compared to the other structures, a precision/recall curve was not needed because the response

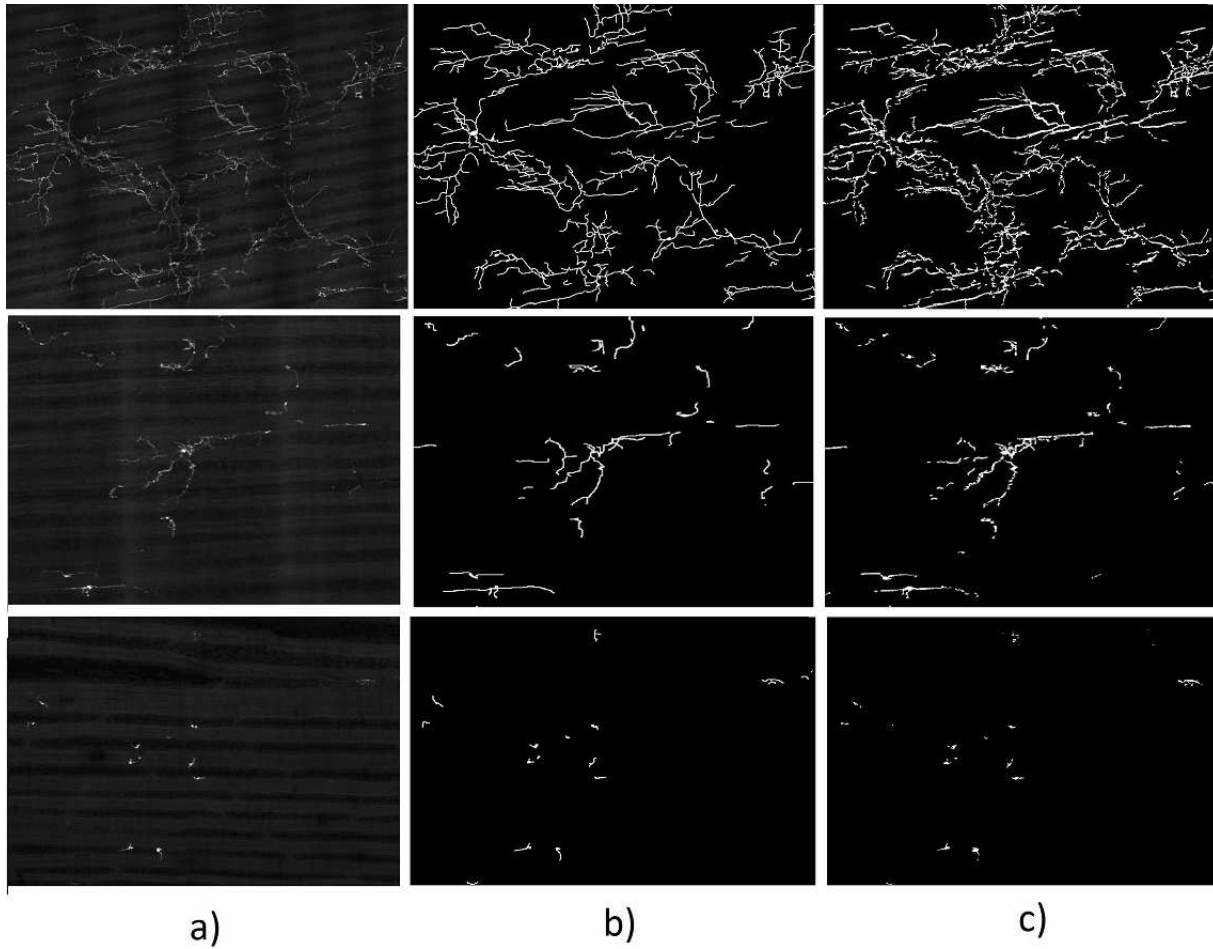


Figure 5. Qualitative comparison between the automatically segmented and ground truth hyphae MIPs from three samples – high, medium, and low fungal infection levels. **a)** MIP Image. **b)** Ground truth. **c)** Thresholded CNN Response.

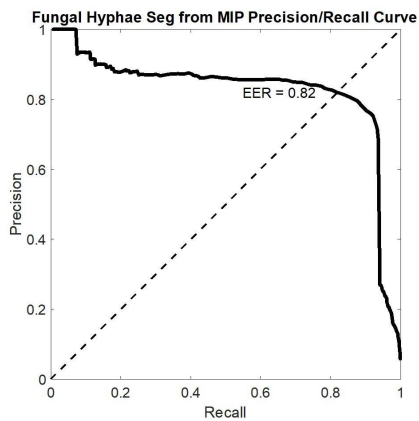


Figure 6. Precision Recall curve generated by varying the threshold on the response from the deep CNN applied to fungal hyphae image stacks and compared to MIP ground truth. A 10px leniency threshold was used for the precision/recall calculation.

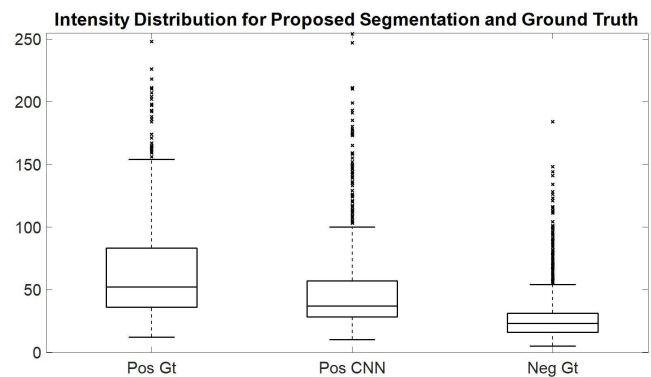


Figure 7. Intensity distribution for fungal hyphae between the positive ground truth pixels, positive pixels according to the CNN, and negative ground truth pixels. The given whiskers represent a standard deviation of 2.7 for 99.3% coverage of the data. Outliers outside this range are displayed as xs.

from the CNN response was nearly binary. A threshold of

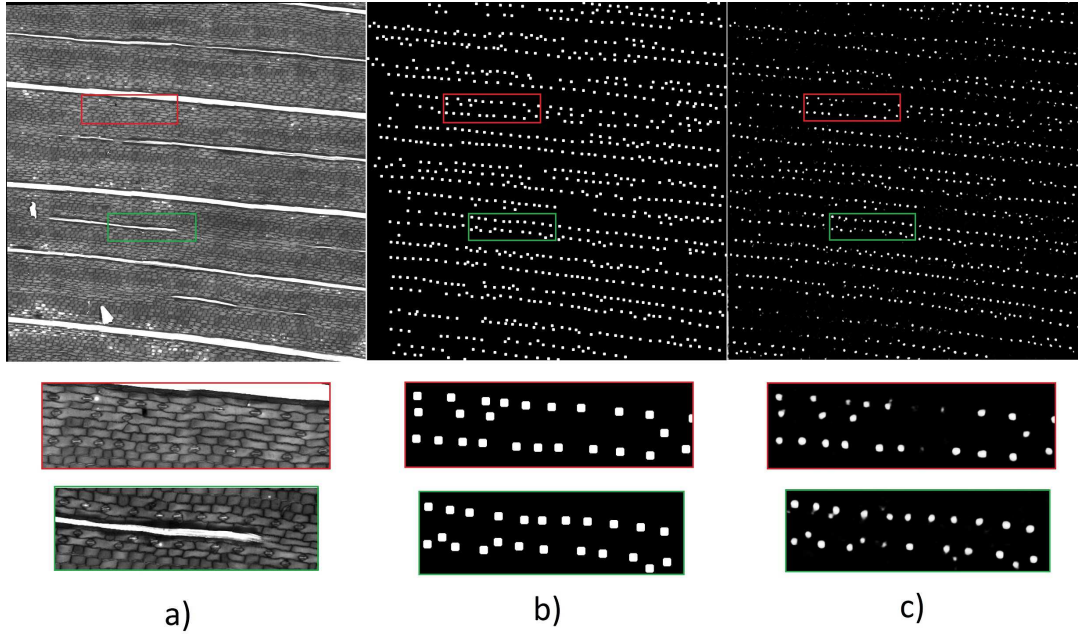


Figure 8. Example result of CNN compared to ground truth for stomata segmentation. **a)** Raw surface image. **b)** Ground truth segmentation. **c)** CNN segmentation.

0.90 for the response from the network was applied to calculate the precision, recall, and F1-score in Table 4 (image regions that were used in training were not included in these metrics). An example full resolution stomata segmentation result is shown in Figure 8.

Table 4. Precision, recall, and F1-score for stomata segmentation using 1834 manually annotated stomata.

Prec	Recall	F1-score
0.78	0.69	0.73

4. Conclusion

We introduce a unified workflow for accurately segmenting specific biological features from maize leaf tissue using a single deep neural network architecture.

Detecting useful information in microscopy data is a difficult task for traditional image processing methods. Hand-crafted methods are usually not robust to changes in intensity and noise, especially across multiple samples. Biological features are also difficult to manually segment and label across a large number of images. Despite these challenges, we have shown that our CNN adaptation and workflow performs well on 2D surface feature segmentation and 3D fungal network segmentation.

In the future, we will implement cross-communication into the workflow such that segmenting one type of feature will assist in the segmentation of all other features. We also hope to expand the workflow to include additional maize subsurface features such as vascular bundle segmentation.

5. Acknowledgements

This work was supported by the U.S. NSF Plant Genome Research Program (IOS-1127076). Microscopy access was supported by the NIH-NIGMS (P20 GM103446), the NSF (IIA-1301765) and the State of Delaware.

References

- [1] H. Hama, H. Kurokawa, H. Kawano, R. Ando, T. Shimogori, H. Noda, K. Fukami, A. Sakaue-Sawano, and A. Miyawaki. Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nat Neurosci*, 14(11):1481–1488, Nov 2011. 1
- [2] A. Kolagunda, R. Wisser, T. Chaya, J. Caplan, and C. Kambhamettu. Detection of fungal spores in 3d microscopy images of macroscopic areas of host tissue. *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 00:479–483, 2016. 2, 3
- [3] K. R. Minker, M. L. Biedrzycki, A. Kolagunda, S. Rhein, F. J. Perina, S. S. Jacobs, M. Moore, T. M. Jamann, Q. Yang, R. Nelson, P. Balint-Kurti, C. Kambhamettu, R. J. Wisser, and J. L. Caplan. Semiautomated confocal imaging of fungal pathogenesis on plants: Microscopic analysis of macroscopic specimens. *Microscopy Research and Technique*, 2016. 1, 2
- [4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]). 2
- [5] P. Saponaro, W. Treible, A. Kolagunda, S. Rhein, J. Caplan, C. Kambhamettu, and R. Wisser. Three-dimensional segmen-

tation of vesicular networks of fungal hyphae in macroscopic microscopy image stacks. *arxiv*, 2017. [1](#), [3](#), [5](#)

- [6] C. A. Warner, M. L. Biedrzycki, S. S. Jacobs, R. J. Wisser, J. L. Caplan, and D. J. Sherrier. An optical clearing technique for plant tissues allowing deep imaging and compatible with fluorescence microscopy. *Plant Physiol.*, 166(4):1684–1687, Dec 2014. [1](#)