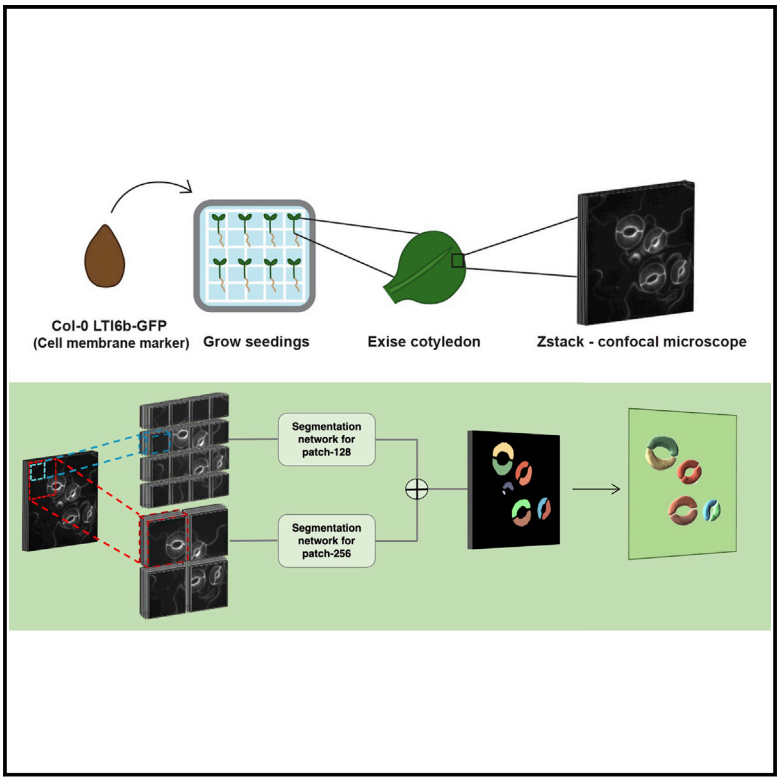
Article

Patterns

Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics

Graphical abstract



Highlights

- Attention-gated, memory-efficient 3D segmentation model is presented
- Model can be trained with a few labeled images while capturing hard-to-localize regions
- Rapid and accurate 3D morphological measurements of stomata from the plant epidermis
- Study demonstrates the model's effectiveness in an analysis of stomatal biomechanics

Authors

Dolzodmaa Davaasuren, Yintong Chen, Leila Jaafar, Rayna Marshall, Angelica L. Dunham, Charles T. Anderson, James Z. Wang

Correspondence

jwang@ist.psu.edu

In brief

Precise, rapid segmentation of volumetric images of stomatal guard cells is crucial for biologists performing analyses of stomatal biomechanics. This is provided by 3DCellNet, introduced in the article and proven practical in the ablation experiments on *Arabidopsis thaliana* images. By studying these underlying mechanics of stomata, we can gain more insights into how they help balance gas exchange at the plant surface to optimize photosynthesis and water transport in response to a multitude of environmental conditions.

Davaasuren et al., 2022, Patterns *3*, 100627
December 9, 2022
https://doi.org/10.1016/j.patter.2022.100627

CellPress

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

Patterns

CellPress
OPEN ACCESS

Article

Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics

Dolzodmaa Davaasuren,[1,4] Yintong Chen,[2,4] Leila Jaafar,[2] Rayna Marshall,[2] Angelica L. Dunham,[2] Charles T. Anderson,[2,3] and James Z. Wang[1,3,5,*]

[1]Data Science and Artificial Intelligence Area, College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA

[2]Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

[3]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence: jwang@ist.psu.edu

https://doi.org/10.1016/j.patter.2022.100627

THE BIGGER PICTURE A major challenge for biologists studying cell dynamics in 3D is labor-intensive manual segmentation and measurements of cells. Plant biologists seek to study the dynamic deformations of stomatal guard cells, which can allow plants to use water more efficiently, a major limiting factor in global agricultural production and an area of increasing concern due to climate change. With this aim, we present a one-stage automated segmentation network for 3D images of stomatal guard cells (3D CellNet) that enables rapid and accurate morphological measurements. When applied to 3D confocal data, it allowed us to quantitatively test how neighboring pavement cells in the epidermis of *Arabidopsis thaliana* plants impose physical constraints on stomatal complexes, a refinement of the "polar stiffening" model of stomatal biomechanics. We anticipate that our model will allow biologists to efficiently test new hypotheses on cell dynamics and biomechanics with a handful of labeled 3D images of their own.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Automating the three-dimensional (3D) segmentation of stomatal guard cells and other confocal microscopy data is extremely challenging due to hardware limitations, hard-to-localize regions, and limited optical resolution. We present a memory-efficient, attention-based, one-stage segmentation neural network for 3D images of stomatal guard cells. Our model is trained end to end and achieved expert-level accuracy while leveraging only eight human-labeled volume images. As a proof of concept, we applied our model to 3D confocal data from a cell ablation experiment that tests the "polar stiffening" model of stomatal biomechanics. The resulting data allow us to refine this polar stiffening model. This work presents a comprehensive, automated, computer-based volumetric analysis of fluorescent guard cell images. We anticipate that our model will allow biologists to rapidly test cell mechanics and dynamics and help them identify plants that more efficiently use water, a major limiting factor in global agricultural production and an area of critical concern during climate change.

INTRODUCTION

Guard cells surround stomatal pores on the leaf surface of plants, and the degree to which stomata are open or closed controls the rate of CO_2 influx and water loss and is thus critical for plant physiology. Stomata help plants maintain a balance between water loss and carbon gain during photosynthesis, water transport, and responses to environmental stresses.[1,2] By studying the underlying mechanics of stomata, we can gain more insights into how they help balance gas exchange at the plant

surface to optimize photosynthesis and water transport in response to a multitude of environmental conditions, including stressors such as drought and excessive heat. To gain a closer view of stomatal guard cells and study their physiology and biomechanics in detail, we need to be able to accurately and quantitatively measure their three-dimensional morphology.[3] For example, despite decades of study, it remains unclear exactly how stomatal aperture is determined by the structures and mechanical properties of paired guard cells and their neighboring pavement cells. Cell ablation has been used to investigate the relationship between mechanical stress and biological responses such as cell polarity and microtubule orientation[4] and is thus a useful tool to study guard cell biomechanics. Employing single-cell ablation combined with confocal microscopy has the potential to help reveal the mechanical interactions and functional interdependencies between guard cells and pavement cells.

One challenge in such efforts is the tremendous amount of time required to manually count and measure a large number of stomata. Software solutions such as 3D Slicer,[5] CellProfiler,[6] MorphoGraphX,[7] and Imaris (Oxford Instruments, Santa Barbara, CA, USA) aim to automate this process, but these software programs require experts to either manually mark certain features of cells or to tune a large set of parameters.

Earlier approaches have relied on traditional image processing and machine-learned features to detect stomata in two-dimensional (2D) images.[8–11] Recently developed methods use more sophisticated features, including histogram of gradients (HOGs), maximum stable external regions (MSERs), wavelet spot detection, and template matching techniques, to identify stomata.[12–15] However, classical image processing techniques are susceptible to noise and perform poorly if the image is not sharply in focus, making analysis of 3D datasets from confocal microscopy challenging since light scattering often blurs cell outlines deeper into the sample. These earlier methods either require the image to contain rich background features, the stomata to be in focus and visible, or the user to pre-select proper shapes and parameters.

Recent advancements in deep learning, especially convolutional neural networks (CNNs), have shown promising performance in object detection, localization, and segmentation in various fields of biomedical imaging. Toda et al.[16] firstly detected stomata with the HOGs feature followed by a CNN to classify cropped pores as open or closed; subsequently, they used binary image segmentation to complete automatic pore measurement. Vialet-Chabrand et al.[17] used a Haar feature-based cascade classifier to determine whether the image contains stomata or not. Bhugra et al.[18] employed a super-resolution CNN (SRCNN) along with a single-shot multi-box to detect stomatal pores in scanning electron microscopy (SEM) images, followed by the use of a fully convolutional network (FCN)[19] to segment the pores. A method for stomatal pore segmentation that used the Chan-Vese (CV) model was presented by Li et al.[20] The authors first used a faster region-based CNN (faster R-CNN) to localize stomata and then extracted the detected stomata, but the method requires manually adjusting the parameters of the CV model. Recently, Jayakody et al.[21] proposed a three-stage pipeline where a feature pyramid network is embedded in the mask R-CNN to identify stomata at different scales. However, all of these methods use a localization step to detect stomata before performing the segmentation, and none of them accomplish volumetric guard cell or stomatal segmentation.

In the medical imaging field, methods for volumetric semantic segmentation work based on 3D augmentation of a U-Net architecture[22] such as 3D U-Net[23] have shown promising results in segmentation tasks for brain tumor images.[24] Recently, the standard 3D U-Net and 2D U-Net models have been used as the main segmentation networks for general cellular segmentation works such as Cellpose[25] and PlantSeg.[26] Series of convolutions are used to reduce the initial image dimension to a set of high-level features. These features are then upsampled back into a scale equivalent to the original image. Annotating volumetric images is very time consuming as it requires consistent labeling in all three orthogonal views. Another challenge associated with 3D images is that they require large amounts of memory when training deep-learning models. Therefore, we aimed to leverage very few annotated patches of the dataset while targeting high accuracy in terms of segmentation.

Here, we developed a patch-wise, attention-gated, fully convolutional 3D framework for guard cell segmentation (Figure 1), which only requires a few hand-annotated volumetric images to train. There are three key parts to our model: (1) two subnetworks for different image resolutions are trained and fused to improve the model's robustness towards different scales. (2) Input images are divided into small patches, and the model is only trained with those patches. This makes the model more memory efficient when analyzing large volumetric images. (3) We improve the model's sensitivity and accuracy with volumetric attention gates by making the model automatically target relevant regions while suppressing irrelevant regions. This eliminates the need to employ an external network for volume of interest (VOI) localization, which has been used frequently in previous approaches. Also, we employ combined dice and focal loss to reduce the inherent class imbalance issue. In addition, we provided an extended pipeline to automatically measure the volume of each guard cell and the width and length of each stomatal complex. This allowed us to perform analyses of these parameters, which are relevant to stomatal biology, in a fully automated fashion. We applied our automated segmentation pipeline to analyze 3D confocal imaging data from an experimental study wherein different subsets of neighboring pavement cells were ablated to determine the effects of these ablations on guard cell morphology. The results of these analyses allow us to refine existing notions of the biomechanical properties of stomatal guard cells and their interactions with neighboring cells into an updated biomechanical model. In the updated model, epidermal cells adjoining the junctional regions between sister guard cells constrain guard cell volume to a greater degree than epidermal cells flanking the sides of the guard cells, potentially due to the heterogeneity of the guard cell wall or higher cell wall permeability at the junction sides than the other sides. This finding highlights the importance of intrinsic and externally imposed mechanical constraints on the poles of stomatal complexes in modulating stomatal opening in response to water influx into the guard cells.

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

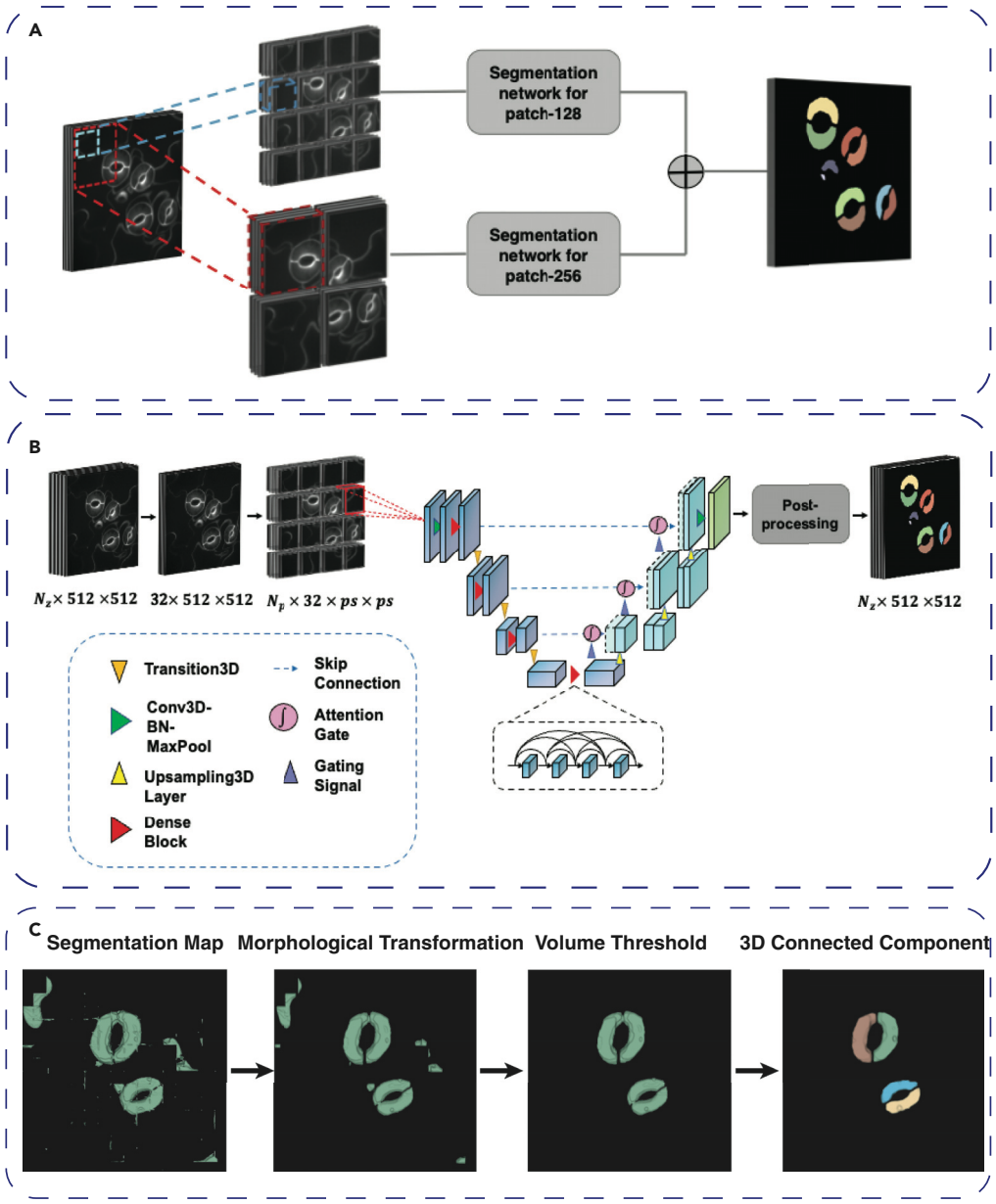


Figure 1. The overall pipeline of the guard cell segmentation network

(A) Final segmentation map results from the fusion of two separate models: one for a patch size of 128 × 128 and the other for a patch size of 256 × 256.

(B) The main segmentation network architecture is shown. The raw image goes through a pre-processing step, which includes automatic noise reduction. Then, the image is resized to a depth of 32 slices and divided into small volume patches. N_z refers to the number of slices in the z stack, and N_p refers to the total number of patches. The encoder part of the network leverages pre-trained, densely connected convolutional networks. The decoder employs an attention gate at the skip connection. Finally, post-processing operations are performed to clean out the raw segmentation map and make it useful for further analysis.

(C) The flow diagram shows the sequence of post-processing steps. Morphological transformations are applied to fill the holes, remove the extrusions, and smooth out the surface. Small artifact blobs are then removed after thresholding on a certain volume dimension. Finally, separate connected 3D objects are classified into different classes, allowing for morphological measurements on individual guard cells.

RESULTS

Model design

Previous segmentation approaches for stomata used localization networks to first localize the region of interest before segmenting, which also helped deal with large images to fit them into the computer's memory. We aimed to build an end-to-end, fully automated network without using any external networks to aid the segmentation pipeline. We collected 3D stacks of images with a 63× objective on a spinning disk confocal microscope. To make the model flexible with an arbitrary number of *z*-slices, we resized the original image with linear interpolation so that the input dimension to the model is 32 × 512 × 512. Since this image size exceeds the GPU memory when training the

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

Table 1. Ablation study on model architecture

Models	mIoU	Precision	F1 score
Baseline (B)	0.6933 ± 0.014	0.7841 ± 0.022	0.8187 ± 0.084
B + combined loss	0.7557 ± 0.027	0.8409 ± 0.028	0.9517 ± 0.014
B with Dense121 (BD)	0.7061 ± 0.011	0.8047 ± 0.018	0.8276 ± 0.024
BD + combined loss	0.7886 ± 0.010	0.9019 ± 0.023	0.8816 ± 0.017
BD + 3D Attention (Attn) + combined loss (ours)	0.8003 ± 0.014	0.9387 ± 0.018	0.9589 ± 0.020
ResNet50 + 3D Attn + combined loss	0.7219 ± 0.032	0.7995 ± 0.012	0.8018 ± 0.019
SeResNet50 + 3D Attn + combined loss	0.7221 ± 0.019	0.8111 ± 0.020	0.8332 ± 0.013
Efficient Net + 3D Attn + combined loss	0.5844 ± 0.024	0.6654 ± 0.025	0.6977 ± 0.044
Ours with 128 × 128	0.8413 ± 0.009	0.9087 ± 0.015	0.9623 ± 0.020
Ours with multi-resolution (final)	0.8200 ± 0.010	0.9034 ± 0.018	0.963 ± 0.026

segmentation model, we divided the original image into two different sizes of patches: 256 × 256 and 128 × 128. Using large image patches provides more abundant contextual information compared with small patches. When limited to a small amount of data for training, the patch-based model provides efficiency and has been shown to improve localization accuracy in the medical imaging domain.[27]

The model architecture for guard cell segmentation from the patches is based on traditional encoder-decoder architecture. We performed an ablation study on different backbone models and modalities to attest to our proposed model's efficacy as demonstrated in Table 1. The baseline model here is the standard and pre-trained 3D U-Net model. The encoder part of the model leverages pre-trained, densely connected convolutional layers (DenseNet)[28] as the model performance was much better compared with the other backbones. By using a pre-trained network, we can avoid training a large number of parameters. We trained two models of the same architecture for each patch size and employed late fusion to regenerate the final segmentation map.

To capture a sufficient receptive field and contextual information, the feature map grid is gradually downsampled in standard CNN architectures. However, it remains difficult to minimize false-positive predictions for small objects that show a large shape variability, such as plant and animal cells. To improve the accuracy, current segmentation frameworks in the biomedical imaging domain[17,20] rely on additional preceding object localization models to simplify the task into separate localization and subsequent segmentation steps. Oktay et al.[29] demonstrated that attention gates (AGs) could replace the localization step in a standard CNN model without introducing a large number of parameters. In contrast to the localization model in multi-stage CNNs, AGs progressively suppress feature responses in irrelevant background regions without the requirement to crop a VOI between networks. We further expanded this AG approach in a 3D setting.

One of the main challenges in volumetric image segmentation is the class imbalance issue. In our confocal datasets, there are at most three to five guard cells in the volume, and neighboring pavement cells occupy large portions of the images. When training small patches, the class imbalance issue is further aggravated. In the training, we avoided using patches that do not include any part of the guard cell and found out the weighted combination of dice and binary focal loss performs the best among the different loss combinations we tried in our ablation experiment.

Training dataset

For a training dataset, we used fluorescent 3D images of the cotyledon epidermis in seedlings of *Arabidopsis thaliana* that were collected via spinning disk confocal microscopy. Figure 2 shows training image samples. Overall, training images varied extensively in terms of voxel intensity even after normalization as shown in Figure 2C. We ran a labeling trial among three cell biologists who are well trained in guard cell labeling. Our results (Figures 3B and 3C) show that there can be large discrepancies in volume measurements even among experts. Toward the beginning and end of the z stack, the human labelers tended to disagree more as it is challenging to distinguish a clear boundary in all three orthogonal views. The average mean intersection over union (IoU) score among the labelers was lower than our model accuracy, indicating that the 3D CellNet model is as good as or better than manual labeling for determining cell volume. A total of eight volumes were fully labeled by expert users in orthogonal *xy*, *yz*, and *xz* slices using Slicer3D[5] into two classes: guard cell and background. Unlike 2D image annotation, labeling 3D images requires continuous adjustment of the labels in all three dimensions. Also, experience in detecting the boundaries between guard cells and neighboring pavement cells is needed. On average, it takes 5 to 8 h to segment one volumetric image for a trained annotator. Therefore, we aimed to leverage as few annotated images as possible for model training. We conducted 5-fold cross-validation on our model and other benchmark models to test their stability with sparse dataset training. The results are shown in the Table 2.

Benchmarks

We compared our model performance with 3D U-Net,[23] 2D U-net[22] pre-trained on ImageNet, and mask RCNN[30] models pre-trained on the COCO dataset. The 2D U-Net and mask RCNN networks were trained with 2D slices and 3D U-Net was trained with 3D image patches identical to the input to our proposed model. Previous methods for automated analysis of stomatal images only segment stomata pores rather than guard cell volumes and require at least a two-step process with some manual interpretation. Thus, we decided it was not relevant to compare our model with existing algorithms for stomatal pore segmentation. Figure 3A shows the performance comparison in mean IoU

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

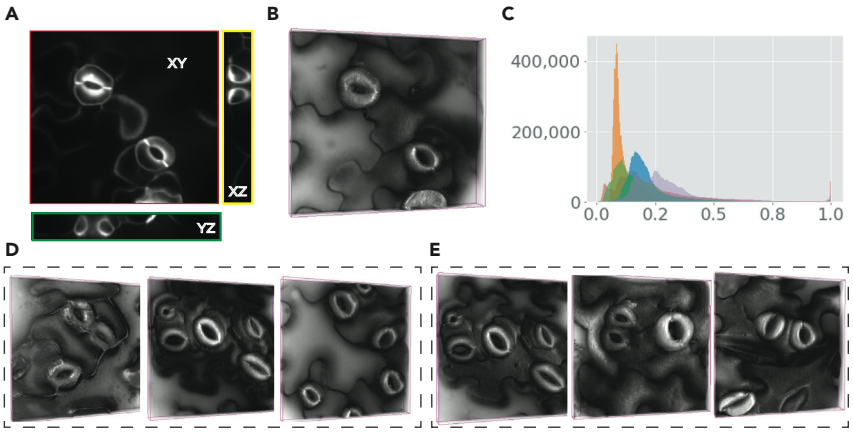


Figure 2. Sample training images

(A) Orthogonal views of a sample image.
(B) 3D rendering of a sample image.
(C) Intensity histogram of the training images.
(D) 3D rendered sample training images (non-ablation).
(E) 3D rendered sample training images (ablation).

(mIoU) score, F1/Dice score, and precision. Quantitative results from 5-fold cross-validation for the models are shown in Table 3. Here, inference time (Inf/T) is reported in seconds. These results show only segmentation performance before any post-processing. Also, we calculated the inference time to segment a single image on both CPU and GPU for all the models. Visual, qualitative results for each segmentation network are shown in Figure 4. Post-processing steps were performed for only qualitative results.

It can be seen that 3D models (3D CellNet and 3D U-Net) outperform 2D models (U-Net and MaskRCNN) by a large margin, further supporting the concept that utilizing volumetric contextual information over slices provides much richer information than separately processing individual slices. When compared with 3D U-Net, 3D CellNet achieves consistently higher scores on all metrics and showed greater accuracy in terms of localization of the VOI.

Stomatal biomechanics study

The "see-saw" hypothesis[31] states that the opening and closing dynamics of grass stomatal complexes are aided by opposite changes in osmotic and turgor pressure between guard cells and neighboring subsidiary cells, which are specialized rounded cells that flank the narrow, dumbbell-shaped guard cells of those species. However, in many eudicot plants, e.g., *A. thaliana*, morphologically distinct subsidiary cells are absent, bringing the see-saw hypothesis into question for those species. Although pavement cells have recently been implicated in guard cell dynamics at the molecular level,[32] how the epidermal pavement cells that surround stomatal complexes on their flanks or across the junctions between the guard cells might differentially influence stomatal biomechanics has not been studied in detail. We applied targeted cell ablation in combination with our 3D CellNet method to test the hypothesis that pavement cells that flank stomatal complexes constrain the widening of the stomatal pore, whereas junctional pavement cells prevent the stomatal complexes from lengthening, potentially synergistically with polar stiffening of the guard cells themselves.[33] We employed confocal microscopy to image stomatal complexes and surrounding cells in the cotyledon epidermis before and after the laser-based ablation and depressurization of some or all neighboring pavement cells, using the plasma membrane marker LTI6b-GFP to both assess protoplast integrity and detect cell boundaries. We collected a total of 50 images for four different ablation types: (1) all neighboring pavement cells are killed; (2) only pavement cells across guard cell junctions are killed; (3) only pavement cells flanking guard cells are killed; and (4) no cells are ablated.

Sample images for the first three ablation types are shown in Figures 6F–6H. After pre-processing followed by segmentation, we calculated stomatal complex width (which scales with pore width) and complex length by computationally fitting a rectangle to each complex and measured the volume of each guard cell by calculating the total voxel numbers in each connected component as demonstrated in Figure 5A. Figures 6A–6D show the before and after width, length, and volume changes for each type of ablation. As shown in Figure 5B, ground-truth width, length, and volume measurements correlate closely with the predicted values.

R-squared values for lines with slope = 1 are shown in each graph. Our data revealed that stomatal complexes responded differently to the ablation of different subsets of neighboring pavement cells. The stomatal complex became wider and shorter after all surrounding pavement cells were ablated (Figure 6A), suggesting the existence of a lateral, but not longitudinal, compression force from pavement cells on the stomatal complex. Ablating junctional cells resulted in no change in complex length but increases in both complex width and guard cell volume (Figure 6B). Consistent with the above model of lateral constraint, ablating pavement cells from the flanking sides of the stomatal complex had a larger effect on complex width than ablating junctional cells (Figures 6B, 6C, and 6E). The finding that neither flanking nor junctional ablation affected complex length, in combination with the fact that ablating all surrounding cells caused the complexes to shorten on average, suggests that surrounding cells might exert longitudinal tension on stomatal complexes (Figures 6B and 6C). Given that the "after" images were collected within seconds of ablation, a significant volume change in guard cells was unexpected, but cell volume in fact increased when all pavement cells or junctional pavement cells were killed. This suggests that constraints at the poles of the stomatal complex limit guard cell inflation. We were also able to compute the cell depth by measuring the height of the stomata in the other orthogonal view (xz). The model-predicted values and the ground-truth measurements are shown in Figure S1.

DISCUSSION

Precise, rapid segmentation of volumetric images of stomatal guard cells is crucial for biologists performing analyses of

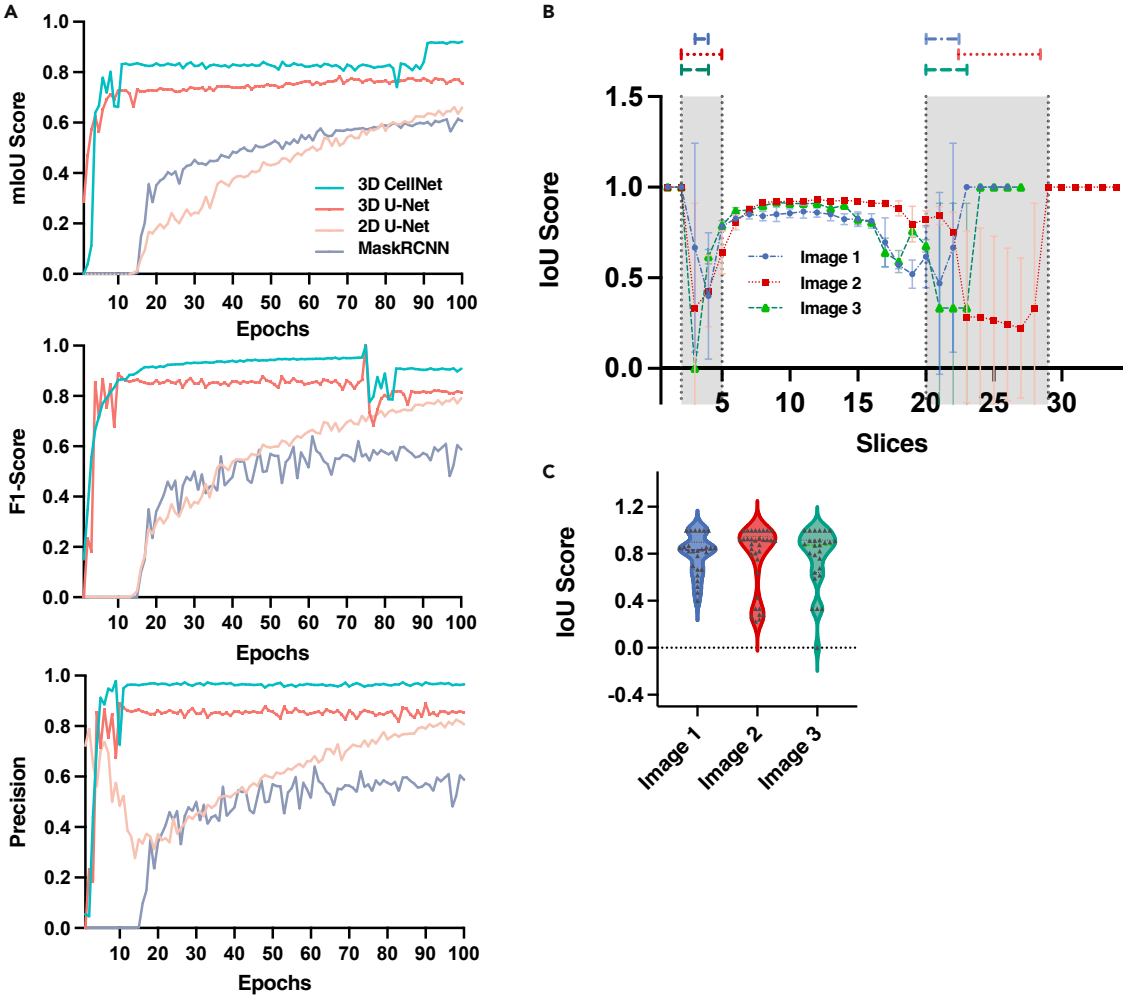


Figure 3. Model performance comparison and annotation analysis between labelers

(A) Comparison between benchmark models on IoU score, F1 score, and precision.

(B) Mean IoU score with standard error per slice for three different volumetric images. The shaded area represents high variance regions.

(C) Violin plot for labelers' IoU score distributions per volumetric image.

stomatal biomechanics. Manual segmentation of such volumetric images is challenging and time consuming as it requires simultaneous labeling in all three orthogonal views. From our experience, these shortcomings can lead to months-long delays in cell mechanics analyses. We provide a single automated solution to the guard cell segmentation that can achieve expert-level accuracy. Our framework leverages only a few volumetric images while tackling inherent class imbalance and memory exhaustion issues with patch-wise, attention-gated mechanisms. Hence, our work can serve as a benchmark framework for many types of volumetric cell segmentation work as it yields accurate segmentation prediction when the inevitable challenges inherent to volumetric images are presented.

Table 2. 5-fold cross-validation result

Test set	mIoU	Precision	F1 score
1	0.8076	0.9497	0.8933
2	0.8020	0.9424	0.8229
3	0.7941	0.8936	0.8947
4	0.8184	0.9001	0.9013
5	0.7877	0.8991	0.8427

Training datasets were divided into 5 sets, and each set was used as a test set.

We further challenged our framework's effectiveness in cell mechanics study by evaluating stomatal morphological changes after ablating neighboring pavement cells. Analysis of the experimental data suggests that differential lateral and longitudinal compression forces generated by pavement cells are critical for maintaining stomatal complex shape and guard cell volume. This idea is consistent with the previous notion that turgor pressure in subsidiary cells has antagonistic effects with turgor pressure in guard cells in determining stomatal aperture[31,34–36] and allows the expansion of the see-saw hypothesis to include eudicot stomatal complexes, which contain kidney-shaped guard cells and lack morphologically distinct subsidiary cells. Instantaneous increases in guard cell volume

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

Table 3. Result comparison of stomata segmentation

Method	mIoU	Precision	F1 score	Inf/T (GPU)	Inf/T (CPU)
3D U-Net	0.7557 ± 0.027	0.8409 ± 0.028	0.9517 ± 0.014	20.7	157.2
2D U-Net	0.4967 ± 0.033	0.6070 ± 0.025	0.7062 ± 0.169	10.1	22.6
MaskRCNN	0.4438 ± 0.037	0.5776 ± 0.021	0.6343 ± 0.011	11.2	25.7
3D CellNet	0.8240 ± 0.017	0.9034 ± 0.010	0.9630 ± 0.016	26.7	177.2

after ablating all or only junctional side pavement cells suggest either a sudden influx of water and/or the release of a restraining force that is exerted mainly at poles of the stomatal complexes by junctional pavement cells. Future challenges in exploring stomatal dynamics in the intact epidermis using a combination of targeted biomechanical interventions and automated, quantitative image analysis include the need for segmenting neighboring pavement cells with as few labels as possible. Our framework sets a new standard for a wide range of future volumetric cell analyses as it provides solutions to the common hurdles encountered in most high-resolution, 3D microscopy images of living cells.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Request for information and resources used in this article should be addressed to Dr. James Wang (jwang@ist.psu.edu).

Materials availability

We used *A. thaliana* images in this study. Please refer to preparation of the materials for more details.

Data and code availability

The dataset and code for the model training and evaluation are available at Mendeley Data: https://data.mendeley.com/datasets/2dvtw8cyhx/draft?a=6673d3a7-fd16-49c0-bad2-f790c8cddb16. The reserved DOI is https://doi.org/10.17632/2dvtw8cyhx.1. We also published our repository at https://github.com/Dolzodmaa/GuardCellSegmentation.

Preparation of the materials

Arabidopsis LTI6b-GFP[37] seeds were sterilized in in 30% bleach + 0.1% SDS for 20 min, then stratified at 4°C for 3 to 10 days before being plated on Murashige and Skoog (MS) plates containing 2.2 g/L MS salts (Caisson Laboratories, Smithfield, UT, USA), 0.6 g/L MES, 1% (w/v) Suc, and 0.8% (w/v) agar (Sigma) (pH 5.6). Seedlings were grown at 22°C under 24 h of approximately 800 PPFD illumination.

Ten-day-old LTI6b-GFP seedlings that were grown under continuous light conditions were ablated using a Micropoint laser (Photonic Instruments, South Windsor, CT, USA). A Zeiss Cell Observer SD microscope with a Yokogawa CSU-X1 spinning disk head and 63× oil 1.4 numerical aperture (NA) objective was used to take z stack images before and after the cell ablated.

We first pre-processed the images by reducing the noise with a non-local means denoising algorithm.[38] The non-local means algorithm replaces the value of a pixel with an average of a selection of other pixels values: small patches centered on the other pixels are compared with the patch centered on the pixel of interest, and the average is calculated only for pixels that have patches close to the current patch. As a result, this algorithm can restore well textures that would be blurred by other denoising algorithms. The denoising operation was performed in a 3D setting. Confocal images are prone to glares in the top and bottom layers of the z stack due to light reflection off the coverslip and/or light scattering deeper into the biological sample. We excluded images with too much glare as they are regarded as rare outliers. This issue affects the model performance on the earlier and latter layers of the slices (Figure S2).

Model architecture

Our segmentation architecture was built on the basis of the U-Net model.[22] U-Net architecture has been shown to perform better with very few labeled biomedical images than sliding-window-based architectures. However, if we input the entire image into the architecture, it will not fit into the GPU memory. To overcome these limitations, we proposed an overlapping patch-wise U-Net architecture with DenseNet as an encoder and AGs in the decoder. The main advantage of our proposed architecture is the patch-wise splitting of a slice obtained from the volumetric image, which helps in better localization because the trained network can focus more on local details in a patch. Moreover, small patches require much less memory for training and testing.

Figures 1A and 1B show the splitting of an input slice into a number of patches. The input slices with 32 × 512 × 512 pixel size are divided into overlapping patches with 128 × 128 and 256 × 256 pixels and a step size of 32. During the training stage, slices of each guard cell volume and their corresponding ground-truth segmentation maps are divided into different patches. These small patches are applied as input to the model for training. We employed the late fusion method to enhance the resulting segmentation maps from the two models. The predictions from the two models are averaged to output the final segmentation.

We designed and evaluated a dual fully CNN (3D CellNet) with a pre-trained 3D DenseNet encoder for each patch size. Pre-trained weights on ImageNet[39] dataset were transferred from 2D to 3D, as demonstrated in Solovyev et al.[40] Using a pre-trained model, we can leverage the learned features from the large dataset and reduce the total number of parameters to

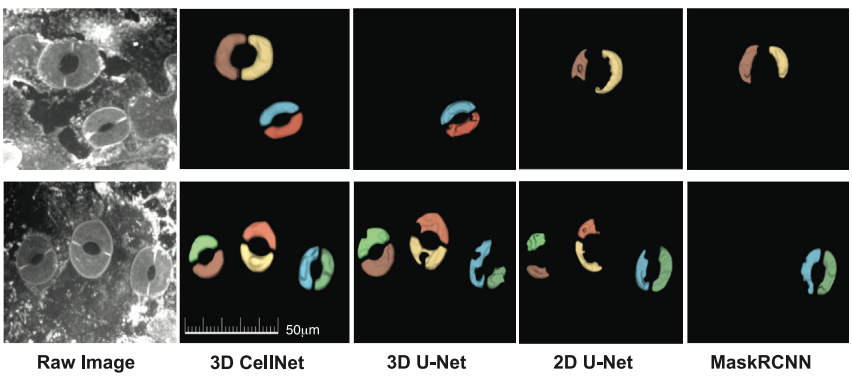


Figure 4. Model comparisons on sample test set

Volumetric segmentation result on sample images from the test set. Post-processing steps are applied on the results from all models. If only tiny portions of the cells are detected, then it is highly likely to be removed in the post-processing step.

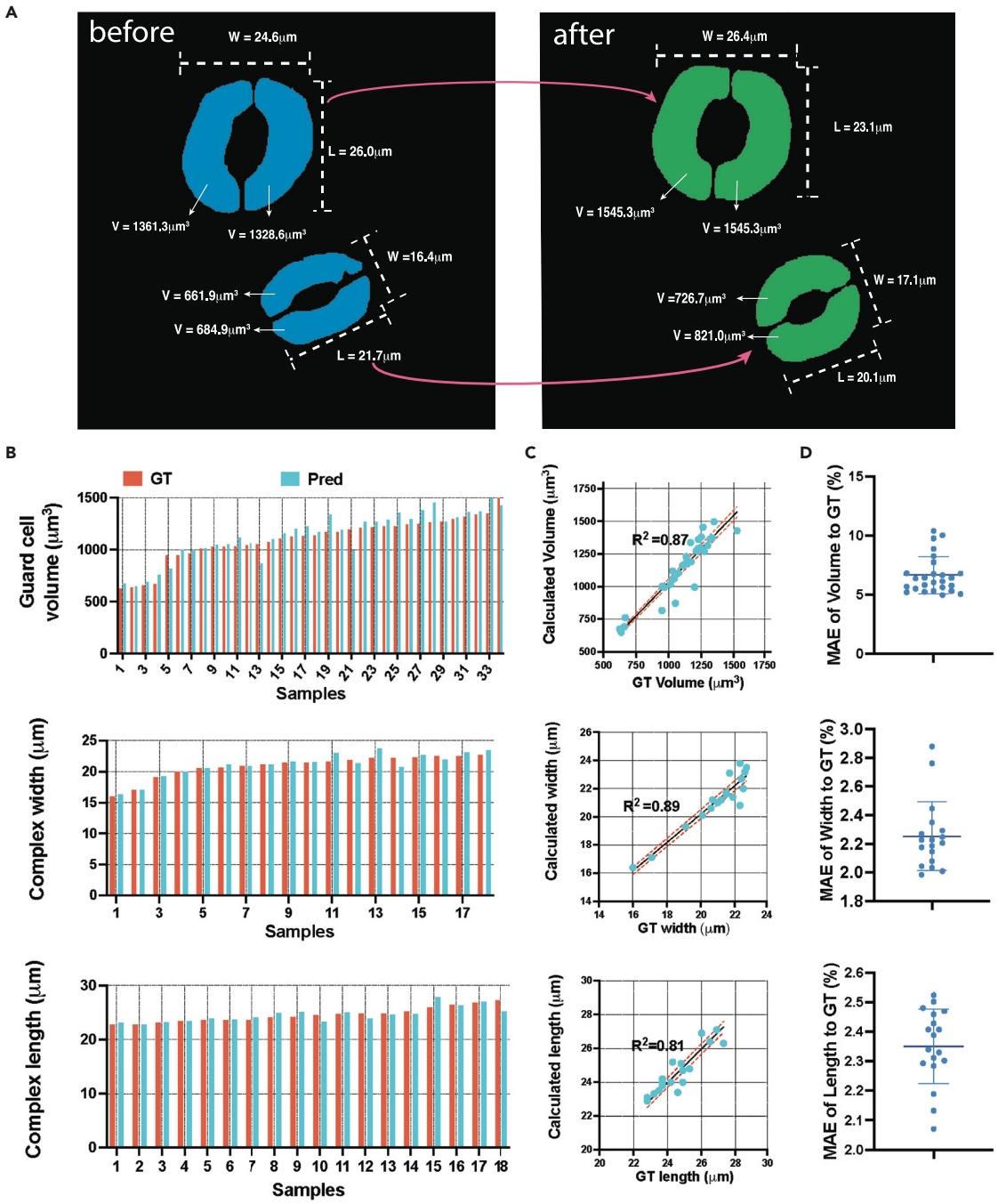


Figure 5. Cell morphological analysis on the predicted segmentation map

(A) Before and after ablation image prediction and corresponding morphological measurement demonstration. Cell width and length are measured from the fitting rectangle on the maximum intensity projection image of the volumetric segmentation map.

(B) Ground-truth (GT) and prediction comparisons on complex width, length, and guard cell volume measurements. Each bar on the bar plots refers to a single cell.

(C) Linear regression between the GT and predicted measurements for complex width and length, as well as guard cell volume for each sample.

(D) Mean absolute error (MAE) of predicted measurements compared with GT values in percentage.

Please cite this article in press as: Davaasuren et al., Automated 3D segmentation of guard cells enables volumetric analysis of stomatal biomechanics, Patterns (2022), https://doi.org/10.1016/j.patter.2022.100627

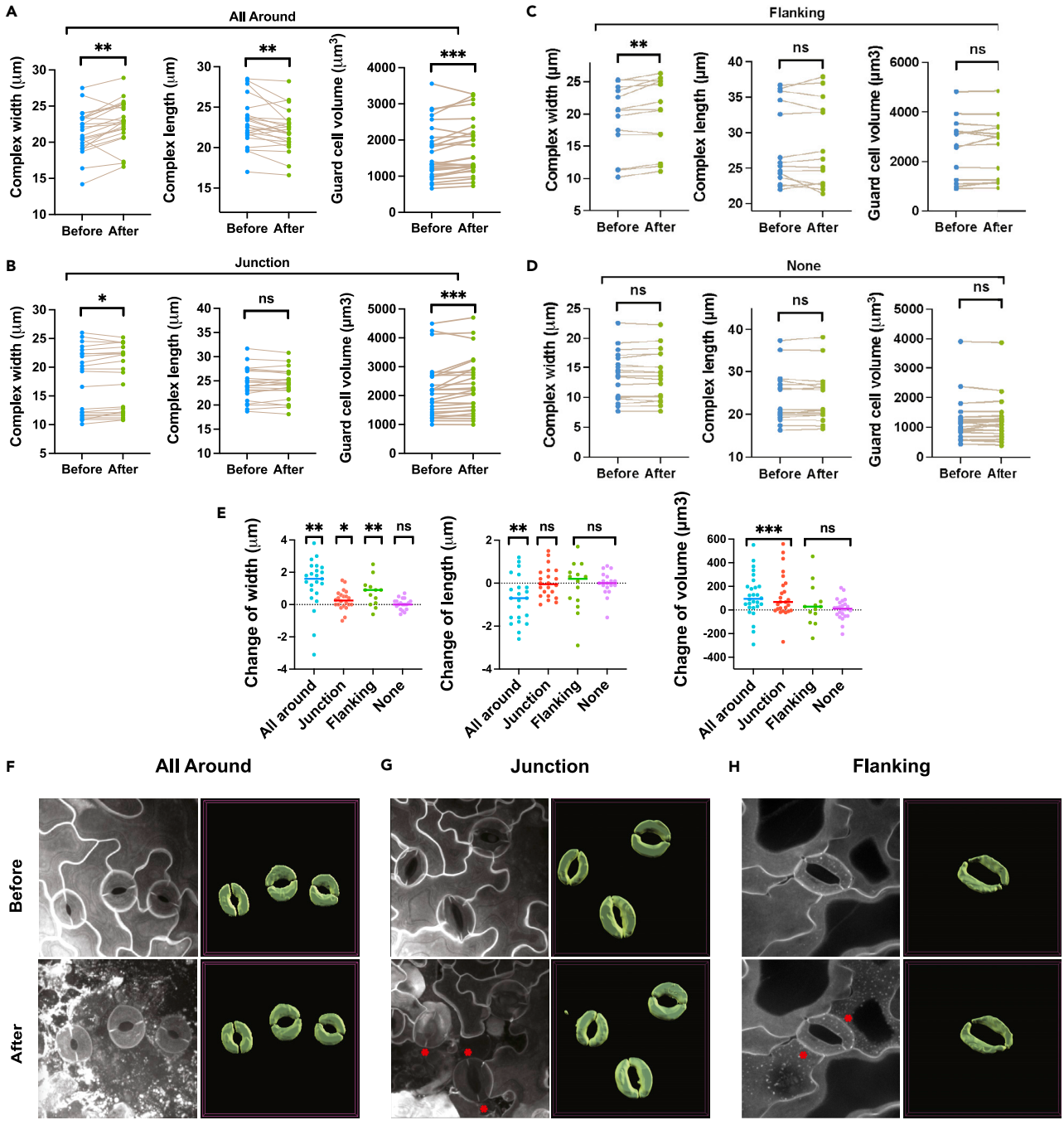


Figure 6. Guard cell volume measurements and comparisons in four different ablation settings

(A–D) Width, length, and volume changes in all-around, junction, flanking, and none-ablation settings, respectively.

(E) Width, length, and volume changes in all four settings. Wilcoxon two-sided signed-rank test is used for the statistical analysis; ** $p < 0.01$; ***$p < 0.001$; ****$p < 0.001$.

(F–H) Comparison between predicted and original images for guard cell complex in before and after ablation for three types of settings: (F) all around, (G) junction, and (H) flanking.

train. DenseNet[28] introduced the concept of shortcut connection, where the input of each layer includes the outputs of all previous layers in a feedforward manner. These connections imply deep supervision and maximum information flow throughout the network, which yielded consistent improvement in accuracy and efficiency. In a CNN, we denote x_l as the output of l^{th} layer. x_l is computed by

$$x_l = H_l(x_{l-1}), \quad \text{(Equation 1)}$$

where H_l is a non-linear transformation, usually composed of convolution, pooling, batch normalization (BN), and non-linear activation rectified linear units (ReLU) layers. As the network becomes deeper, the performance of

the network saturates due to exploding or vanishing gradients calculated in the backpropagation. DenseNet mitigates this degradation issue by concatenating feature maps of all previous layers as

$$x_l = H_l([x_0, x_1, \cdots, x_{l-1}]), \quad \text{(Equation 2)}$$

where $(\cdot)$ refers to concatenation operation. The encoder part of the network performs downsampling to reduce the feature map resolution while increasing the receptive field. In our model, the downsampling is performed by four dense blocks and three transition blocks in between the dense blocks. Dense blocks are composed of BN-ReLU-Conv3D (1 × 1 × 1)-BN-ReLU-Conv3D (3 × 3 × 3) with a growth rate of 16, and the transition blocks include BN-ReLU-Conv3D (1 × 1 × 1) followed by AveragePool3D(2) to reduce the feature map.

In the upsampling path, we used UpSampling3D to recover the resolution and ensure the size of the prediction map is consistent with the corresponding input size, which is important when using skip connections. Skip connections directly concatenate the output of feature maps in encoder layers to the corresponding prediction map in the decoder layers to further strengthen the deep supervision and stabilize the model learning.

3D AGs

It is challenging to reduce false-positive predictions for small objects and cells in microscopic images. With gradually downsampled feature maps in the encoder part of the segmentation, model prediction is conditioned on the collected information from a large receptive field. To improve the segmentation accuracy, existing cell segmentation frameworks usually include a localization step to help with detecting small objects prior to the segmentation. Recent work on introducing the attention mechanism in the CNN[29,41] has shown promising results on segmentation work without needing multi-stage CNNs.

The AGs are incorporated into the upsampling stage of our segmentation architecture to direct the attention to salient features that are passed through the skip connections (see Figure 1B). AGs filter the neuron activations both in the forward pass and the backward pass. Gradients originating from non-relevant regions are downweighted during the backpropagation. This allows model parameters in earlier layers to be updated mostly based on spatial regions that are relevant to the task. In the skip connection, to merge only relevant activations, information from the coarse scale is used in gating to filter out the irrelevant regions in skip connections. The gating signal for each voxel i is denoted as $g_i \in R$. The other input to the AG comes from the skip connection and is denoted as $x_i^l \in R^{F_l}$, where F_l denotes the number of feature maps in layer l. Since it comes from the earlier layers, it contains better spatial information. The output of AGs is the element-wise multiplication of the input feature maps and attention coefficients: $\hat{x}_{i,c}^l = x_{i,c}^l \cdot \alpha_i^l$, where $\alpha_i \in [0, 1]$ are the attention coefficients that identify salient image regions to preserve only activations relevant to the task. Attention is formulated as follows:

$$\alpha_i^l = \sigma_2\left(\psi^T\left(\sigma_1\left(W_x^T x_i^l + W_g^T g_i + b_g\right)\right) + b_\psi\right), \quad \text{(Equation 3)}$$

where σ_1 is ReLU and $\sigma_2(x_{i,c})) = (1 + \exp(-x_{i,c}))^{-1}$ is the sigmoid activation function. Here, the AG is characterized by linear transformations W_x and W_g, which are computed using Conv3D (1 × 1 × 1) and biases b_g and b_ψ. Lastly, the UpSampling3D operation is used to restore the resolution before the concatenation.

Loss function

We used a combined loss function that helps the model to focus more on the hard-to-find class, as the dataset has a high class imbalance. The total loss function equation is the weighted sum of Dice and binary focal losses:[42]

$$L_{\text{Dice}} = 1 - \frac{2\sum p_i g_i}{\sum p_i^2 + \sum g_i^2}, \quad \text{(Equation 4)}$$

$$L_{\text{BinaryFocal}} = -\alpha(1 - p_i)^\gamma g_i \log(p_i) - p_i^\gamma (1 - g_i)\log(1 - p_i), \text{ and} \quad \text{(Equation 5)}$$

$$L_{\text{Total}} = L_{\text{Dice}} + \beta L_{\text{BinaryFocal}} \quad \text{(Equation 6)}$$

where $p_i \in P$ and $g_i \in G$ are the predicted and ground-truth segmentation volumes, respectively. In the binary focal loss, we set α as 0.25, γ as 2, and in the total loss, we chose β as 2. Focal loss introduces a modulation term to the regular cross-entropy loss, and by setting $\gamma > 0$, the focal loss puts more focus on hard and misclassified examples.

Evaluation metrics

The metrics used for evaluating the performance of the segmentation are mIoU, F1 score, and precision. mIoU values are calculated by taking the ratio of the intersection of the segmentation volume and the union of the segmentation volume between ground-truth and predicted segmentation. mIoU is equivalent to TP/(TP + FP + FN), the F1 score, also known as the Dice coefficient,[43] is equivalent to 2TP/2TP + FP + FN), and the precision score is TP/(TP + FN), where TP is true positive, FP is false positive, and FN is false negative.

Experimental setting

We trained the network on five volumetric images and tested it on the remaining three images. To allow the model to analyze images with various depths, we resized the image into constant a depth size of 32 pixels. Each slice size is 512 × 512 pixels. We trained with a batch size of 4 for 500 epochs. To avoid overfitting, we employed an early stopping mechanism. Usually, models tend to converge at around 100 epochs. We implemented the proposed framework in TensorFlow and used NVIDIA Tesla V100.

Post-processing

The resulting segmentation maps still produced some artifacts around the guard cell. With proper post-processing algorithms, we could clean the segmentation results further. First, morphological dilation and erosion operations were performed on each slice to fill small holes in the guard cells (Figure 1C). Then, a morphological opening operation was applied to the whole 3D image to separate the connected guard cells. This step is crucial to calculate the volume for each guard cell in the image. Secondly, we identified 3D connected components[44] in the segmentation map. All guard cells were identified as one class in the segmentation map, but with the classification of 3D connected components, we could further expand the model to 3D instance segmentation. Figure S3 shows false and true positive rates for before and after post-processing. Finally, 3D volumes were rendered from z stack segmentation maps in 3D Slicer with boundary-smoothing filters.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100627.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant MCB-2015943/2015947. The authors thank Hojae Yi, Eoin McEvoy, and Mythili Subbanna for helpful discussion and members of the Anderson lab for technical support.

AUTHOR CONTRIBUTIONS

D.D. developed the pipeline, analyzed the data, wrote the manuscript, and contributed to the data curation. Y.C. collected the majority of the data, labeled a portion of the images, participated in theoretical discussion, and contributed to the writing. L.J. and A.L.D. collected some parts of the images and contributed to the image labeling. R.M. contributed to the image labeling. C.T.A. engaged in theoretical discussion and contributed to the writing. J.Z.W. engaged in technical discussion and commented extensively on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Patterns
Article

CellPress
OPEN ACCESS

Received: June 20, 2022
Revised: August 2, 2022
Accepted: October 12, 2022
Published: November 9, 2022

REFERENCES

1. Berger, D., and Altmann, T. (2000). A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in Arabidopsis thaliana. Genes Dev. *14*, 1119–1131.
2. Hetherington, A.M., and Woodward, F.I. (2003). The role of stomata in sensing and driving environmental change. Nature *424*, 901–908.
3. Meckel, T., Gall, L., Semrau, S., Homann, U., and Thiel, G. (2007). Guard cells elongate: relationship of volume and surface area during stomatal movement. Biophys. J. *92*, 1072–1080.
4. Bringmann, M., and Bergmann, D.C. (2017). Tissue-wide mechanical forces influence the polarity of stomatal stem cells in Arabidopsis. Curr. Biol. *27*, 877–883.
5. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging *30*, 1323–1341.
6. McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). Cellprofiler 3.0: next-generation image processing for biology. PLoS Biol. *16*, e2005970.
7. Strauss, S., Sapala, A., Kierzkowski, D., and Smith, R.S. (2019). Quantifying plant growth and cell proliferation with MorphoGraphX. Methods Mol. Biol. *1992*, 269–290.
8. Karabourniotis, G., Tzobanoglou, D., Nikolopoulos, D., and Liakopoulos, G. (2001). Epicuticular phenolics over guard cells: exploitation for in situ stomatal counting by fluorescence microscopy and combined image analysis. Ann. Bot. *87*, 631–639.
9. Sanyal, P., Bhattacharya, U., and Bandyopadhyay, S.K. (2008). Analysis of SEM images of stomata of different tomato cultivars based on morphological features. In Proceedings of the Second Asia International Conference on Modelling & Simulation (AMS) (IEEE), pp. 890–894.
10. Bourdais, G., McLachlan, D.H., Rickett, L.M., Zhou, J., Siwoszek, A., Häweker, H., Hartley, M., Kuhn, H., Morris, R.J., MacLean, D., and Robatzek, S. (2019). The use of quantitative imaging to investigate regulators of membrane trafficking in Arabidopsis stomatal closure. Traffic *20*, 168–180.
11. Omasa, K., and Onoe, M. (1984). Measurement of stomatal aperture by digital image processing. Plant Cell Physiol. *25*, 1379–1388.
12. Laga, H., Shahinnia, F., and Fleury, D. (2014). Image-based plant stomata phenotyping. In Proceedings of the 13th International Conference on Control Automation Robotics & Vision (IEEE), pp. 217–222.
13. Liu, S., Tang, J., Petrie, P., and Whitty, M. (2016). A fast method to measure stomatal aperture by MSER on smart mobile phone. In Applied Industrial Optics: Spectroscopy, Imaging and Metrology, G. Bennett, ed. (Optica Publishing Group). AIW2B–2.
14. Duarte, K.T., de Carvalho, M.A.G., and Martins, P.S. (2017). Segmenting high-quality digital images of stomata using the wavelet spot detection and the watershed transform. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (Springer), pp. 540–547.
15. Jayakody, H., Liu, S., Whitty, M., and Petrie, P. (2017). Microscope image based fully automated stomata detection and pore measurement method for grapevines. Plant Methods *13*, 1–12.
16. Toda, Y., Toh, S., Bourdais, G., Robatzek, S., Maclean, D., and Kinoshita, T. (2018). DeepStomata: facial recognition technology for automated stomatal aperture measurement. Preprint at bioRxiv. https://doi.org/10.1101/365098.
17. Vialet-Chabrand, S., and Brendel, O. (2014). Automatic measurement of stomatal density from microphotographs. Trees (Berl.) *28*, 1859–1865.
18. Bhugra, S., Mishra, D., Anupama, A., Chaudhury, S., Lall, B., Chugh, A., and Chinnusamy, V. (2018). Deep convolutional neural networks based framework for estimation of stomata density and structure from microscopic images. In Proceedings of the European Conference on Computer Vision Workshops (Springer), pp. 412–423.
19. Shelhamer, E., Long, J., and Darrell, T. (2016). Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. *39*, 640–651.
20. Li, K., Huang, J., Song, W., Wang, J., Lv, S., and Wang, X. (2019). Automatic segmentation and measurement methods of living stomata of plants based on the CV model. Plant Methods *15*, 1–12.
21. Jayakody, H., Petrie, P., Boer, H.J.d., and Whitty, M. (2021). A generalised approach for high-throughput instance segmentation of stomata in microscope images. Plant Methods *17*. 27–13.
22. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 234–241.
23. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 424–432.
24. Feng, X., Tustison, N.J., Patel, S.H., and Meyer, C.H. (2020). Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features. Front. Comput. Neurosci. *14*, 25.
25. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106.
26. Wolny, A., Cerrone, L., Vijayan, A., Tofanelli, R., Barro, A.V., Louveaux, M., Wenzl, C., Strauss, S., Wilson-Sánchez, D., Lymbouridou, R., et al. (2020). Accurate and versatile 3D segmentation of plant tissues at cellular resolution. Elife *9*, e57613.
27. Luna, M., and Park, S.H. (2018). 3D patchwise U-Net with transition layers for MR brain segmentation. In Proceedings of the International MICCAI Brainlesion Workshop (Springer), pp. 394–403.
28. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 4700–4708.
29. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al. (2018). Attention U-Net: learning where to look for the pancreas. Preprint at arXiv. 1804.03999. https://doi.org/10.48550/arXiv.1804.03999.
30. He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In In proceedings of the IEEE International Conference on Computer Vision (IEEE), pp. 2961–2969.
31. Franks, P.J., and Farquhar, G.D. (2007). The mechanical diversity of stomata and its significance in gas-exchange control. Plant Physiol. *143*, 78–87.
32. Nieves-Cordones, M., Azeem, F., Long, Y., Boeglin, M., Duby, G., Mouline, K., Hosy, E., Vavasseur, A., Chérel, I., Simonneau, T., et al. (2022). Non-autonomous stomatal control by pavement cell turgor via the K+ channel subunit AtKC1. Plant Cell *34*, 2019–2037.
33. Carter, R., Woolfenden, H., Baillie, A., Amsbury, S., Carroll, S., Healicon, E., Sovatzoglou, S., Braybrook, S., Gray, J.E., Hobbs, J., et al. (2017). Stomatal opening involves polar, not radial, stiffening of guard cells. Curr. Biol. *27*, 2974–2983.e2.
34. Glinka, Z. (1971). The effect of epidermal cell water potential on stomatal response to illumination of leaf discs of Vicia faba. Physiol. Plant. *24*, 476–479.
35. MacRobbie, E.A.C., and Lettau, J. (1980). Ion content and aperture in "isolated" guard cells of Commelina communis L. J. Membr. Biol. *53*, 199–205.

36. MacRobbie, E.A.C., and Lettau, J. (1980). Potassium content and aperture in "intact" stomatal and epidermal cells of Commelina communis L. J. Membr. Biol. *56*, 249–256.

37. Cutler, S.R., Ehrhardt, D.W., Griffitts, J.S., and Somerville, C.R. (2000). Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. Proc. Natl. Acad. Sci. USA. *97*, 3718–3723.

38. Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition *2*, 60–65.

39. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: a large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 248–255.

40. Solovyev, R., Kalinin, A.A., and Gabruseva, T. (2022). 3D convolutional neural networks for stalled brain capillary detection. Comput. Biol. Med. *141*, 105089.

41. Wang, X., Han, S., Chen, Y., Gao, D., and Vasconcelos, N. (2019). Volumetric attention for 3D medical image segmentation and detection. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 175–184.

42. Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (IEEE), pp. 2980–2988.

43. Dice, L.R. (1945). Measures of the amount of ecologic association between species. Ecology *26*, 297–302.

44. William, S. (2021). connected-components-3d. https://pypi.org/project/connected-components-3d/. Version-3.8.0.