

RESEARCH ARTICLE

LeafNet: A Tool for Segmenting and Quantifying Stomata and Pavement Cells

Shaopeng Li^{a,1}, Linmao Li^{a,1}, Weiliang Fan^a, Suping Ma^a, Cheng Zhang^{c,d}, Jang Chol Kim^a, Kun Wang^a, Eugenia Russinova^{c,d}, Yuxian Zhu^{a,b,2}, and Yu Zhou^{a,b,e,f,2}

^a College of Life Sciences, Wuhan University, Wuhan, P. R. China

^b Institute for Advanced Studies, Wuhan University, Wuhan, P. R. China

^c Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

^d Center for Plant Systems Biology, VIB, Ghent, Belgium

^e State Key Laboratory of Virology, Wuhan University, Wuhan, P. R. China

^f Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, P. R. China

¹ These authors contributed equally to this work

² Corresponding Authors: zhuyx@whu.edu.cn (Yuxian Zhu), yu.zhou@whu.edu.cn (Yu Zhou)

Short title: LeafNet for Segmenting Stomata and Pavement Cells

One-sentence summary: LeafNet is an accurate, robust, automatic, high-throughput analytical tool for identifying and quantifying morphological features of both stomata and pavement cells in leaf epidermal images.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) are: Yuxian Zhu (zhuyx@whu.edu.cn) and Yu Zhou (yu.zhou@whu.edu.cn).

ABSTRACT

Stomata play important roles in gas and water exchange in leaves. The morphological features of stomata and pavement cells are highly plastic and are regulated during development. However, it is very laborious and time-consuming to collect accurate quantitative data from the leaf surface by manual phenotyping. Here, we introduce LeafNet, a tool that automatically localizes stomata, segments pavement cells (to prepare them for quantification), and reports multiple morphological parameters for a variety of leaf epidermal images, especially bright-field microscopy images. LeafNet employs a hierarchical strategy to identify stomata using a deep convolutional network and then segments pavement cells on stomata-masked images using a region merging method. LeafNet achieved promising performance on test images for quantifying different phenotypes of individual stomata and pavement cells compared with six currently available tools, including StomataCounter, Cellpose, PlantSeg, and PaCeQuant. LeafNet shows great flexibility, and we improved its ability to analyze bright-field images from a broad range of species as well as confocal images using transfer learning. Large-scale images of leaves can be efficiently processed in batch mode and interactively inspected with a graphic user interface or a web server (<https://leafnet.whu.edu.cn/>). The functionalities of LeafNet could easily be extended and will enhance the efficiency and productivity of leaf phenotyping for many plant biologists.

INTRODUCTION

1 Stomata are microscopic openings in the epidermis of leaves, stems, and other plant organs that
2 allow for oxygen and carbon dioxide exchange between a plant and the atmosphere as well as for
3 water loss by transpiration (Zoulias et al., 2018). In general, each stoma contains a pair of
4 specialized guard cells. In many plants, two or more subsidiary or accessory cells that are adjacent
5 to guard cells cooperatively regulate stomatal aperture. Stomatal function is essential for
6 photosynthesis and respiration, which are critical for plant survival in the terrestrial environment
7 and for plant productivity (Qi and Torii, 2018). Thus, stomata biology has attracted the interest of
8 many plant researchers over the years.

9 Recent changes in climate, including elevated CO₂ levels, high temperatures, and drought,
10 have significantly influenced the ecosystem structure and the productivity of global agriculture
11 (Xu et al., 2016; Engineer et al., 2016). High-throughput leaf thermal imaging has identified
12 multiple mutants in the CO₂ response (Hashimoto et al., 2006). To optimize the regulatory
13 functions of stomata in response to the changing environment, the generation, development, and
14 patterning of stomata and pavement cells are regulated by the complex interplay between internal
15 developmental programs and various environmental cues (Casson and Hetherington, 2010).
16 However, the irregularity of plant epidermal cells makes quantitative analysis difficult and
17 inefficient. Hence, there is an urgent need for high-throughput technologies for screening large
18 populations of genetic materials to identify regulators of stomatal development.

19 Due to the irregularity of epidermal cells surrounding stomata, traditional phenotyping of
20 stomata and pavement cells in images generally depends on laborious, time-consuming manual
21 work by specialists in plant biology. Fetter et al. developed StomataCounter to identify and count
22 stomata in scanning electron microscopy (SEM) images and differential interference contrast
23 (DIC) images using deep convolutional neural networks (Fetter et al., 2019), but this technique
24 has reduced accuracy for bright-field images. Several tools have recently been developed for the
25 segmentation of pavement cells (to prepare them for quantification), including PaCeQuant (Möller
26 et al., 2017), PlantSeg (Wolny et al., 2020), and Cellpose (Stringer et al., 2021). PlantSeg and
27 PaCeQuant can generate accurate segmentation for confocal and light sheet images, but their
28 accuracy is limited using bright-field images taken under a light microscope. Cellpose performs
29 well with convex polygon-like cells, but fails with puzzle-shaped cells. MorphoGraphX is a 3D

30 image analysis tool that can be used to collect accurate leaf epidermal information from confocal
31 or light sheet images (Barbier de Reuille et al., 2015); however, collecting large numbers of 3D
32 images is expensive and time-consuming compared with 2D bright-field images. Therefore, an
33 automatic tool is needed to accurately identify stomata and pavement cells simultaneously in
34 bright-field images taken under a light microscope.

35 Here, we present an accurate, robust, automatic, high-throughput analytical tool called
36 LeafNet for identifying and quantifying different features of both stomata and pavement cells in
37 light microscopy images for plant biology studies.

38

39 **RESULTS**

40 **Hierarchical strategy for segmenting stomata and pavement cells**

41 In representative bright-field images of *Arabidopsis thaliana* leaf epidermis obtained under a light
42 microscope (Figure 1A and Supplemental Figure S1A), the stomata appear ellipse-like, whereas
43 the pavement cells are extremely irregular in shape. Furthermore, the experimental process of
44 creating the images can generate various types of noise, which makes it difficult to perform
45 accurate segmentation and to quantify their numbers and other features. Manual segmentation can
46 be accurate (Figure 1B), but it is very laborious and time-consuming to label the boundaries of
47 individual stomata and the puzzle-shaped contours of pavement cells. Besides segmentation, it is
48 important to characterize the general features of stomata and pavement cells, such as count and
49 size (Figure 1C), for a large number of leaves from the same or different genotypes.

50 To solve these problems more efficiently, we built an automatic tool for identifying stomata
51 using a deep-learning approach, segmenting pavement cells with a region merging algorithm, and
52 quantifying their features, including count, size, and length-to-width ratio for stomata and 28
53 different morphological parameters (e.g., size, perimeter, circularity, lobe count, and so on) for
54 pavement cells. We initially manually annotated 140 images with fine segmentation of stomata
55 and pavement cells by labeling pavement cell walls in green, stomata in blue, and the background
56 in black (Figure 1D). All of these images with manual annotations are available at the LeafNet
57 website (<https://leafnet.whu.edu.cn/suppdata/>).

58 We developed the LeafNet program, which employs a hierarchical strategy to sequentially
59 identify stomata and pavement cells (Figure 1E). We first trained a stoma detector, the StomaNet
60 (Supplemental Figure S1B) module, based on a deep residual network (Figure 1F and
61 Supplemental Figure S2) to identify reliable stomata in the input image. We then masked the
62 stomata out of the original image by coloring them black. We built the LeafSeg module to reliably
63 identify pavement cell borders using a region merging algorithm (Supplemental Figure S1C, see
64 methods for details). Finally, we merged the stomata and pavement cell borders. Combining the
65 two modules, LeafNet can generate pixel-wise segmentation of the input image (Supplemental
66 Figure S1D) and then collect different morphological features of stomata and pavement cells
67 (Figure 1E).

68

69 LeafNet shows good performance for the segmentation of stomata and pavement cells

70 To evaluate the performance of LeafNet for detecting stomata, we manually labeled 30 images
71 with 1,086 complete stomata as a test set for stomata detection. The stomata detecting module
72 StomaNet achieved an average precision of 98.1%, while StomataCounter (Fetter et al., 2019)
73 reached 89.5% precision as our baseline (Figure 2A). As a result, of the 1,086 ground truth stomata
74 (using manual labeling as the gold standard), StomaNet successfully detected 1,062 stomata
75 (missing 24) and falsely detected 35 stomata (Figure 2B).

76 To compare LeafSeg with a baseline method, the morphological watershed algorithm (which
77 operates on the topographic surface of an image gradient), we performed baseline analysis using
78 ITK, a general image processing library widely used for biological images (McCormick et al.,
79 2014). We added another optional module, ITK morphological watershed, into LeafNet by calling
80 the ITK morphological watershed algorithm for pavement cell segmentation (see Methods for
81 details). The ITK morphological watershed module is interchangeable with LeafSeg in LeafNet.
82 For our 140 manually labeled images, the LeafSeg module of LeafNet achieved an average
83 precision of 89.1% for segmenting ~14,900 pavement cells, while ITK morphological watershed
84 achieved an average precision of 70.7% (Figure 2C). In detail, 79.6% pavement cells were
85 correctly predicted, 6.3% cells resulted from under-segmentation (multiple ground truth cells were
86 merged into one cell), and 14.2% cells resulted from over-segmentation (one ground truth cell was
87 split into multiple cells) (Figure 2D).

88 To further evaluate the ability of LeafNet to quantify different epidermal cell characteristics
89 through pixel-wise segmentation, we compared the predicted stoma sizes, length/width ratios, and
90 pavement cell sizes with the ground truth. The average deviations of stoma sizes were in the range
91 of -13.3% to 13.9% by image, while the average deviations of length/width ratios ranged from -
92 0.1% to 13.2% (Figure 2E). The MAE metrics (mean absolute error) of stoma size and length/width
93 ratio reached 12.6% and 12.4%, respectively. The cumulative error distribution showed that most
94 stomata were correctly predicted, with ~10% of stomata having large errors in size or length/width
95 ratio (Figure 2F). Although under-segmented and over-segmented pavement cells existed, the
96 difference in pavement cell sizes between the predictions and ground truth was not significant (P
97 value as 0.48 from two-tailed t -test on ~11,000 complete pavement cells in images) (Figure 2G).

98 Moreover, we analyzed the success and failure of LeafNet for individual cases. The LeafSeg
99 module of pavement cell segmentation is noise-tolerant (Figure 2H, top row), but occasionally
100 failed when using images with strong noise (Figure 2H, middle row) or with multiple breaks in
101 short borders (Figure 2H, bottom row). As shown in the three representative examples (Figure 2I,
102 left column), the StomaNet module faithfully captured stomata in most cases. The presence of
103 fuzzy contours (Figure 2I, middle column) or ellipse-like cell walls and noise (Figure 2I, right
104 column) seldom prevented StomaNet from correctly identifying stomata.

105 Taken together, LeafNet performed satisfactorily in identifying stomata and pavement cells
106 and quantifying their biological features.

107

108 **Comparison of LeafNet with StomataCounter and PaCeQuant**

109 For stoma detection, StomataCounter was developed to automatically count stomata from
110 micrographs of the leaf epidermis (Fetter et al., 2019). This program performed well for SEM and
111 DIC images, but it had limited capacity to detect stomata in our dataset with bright-field images.
112 In our testing dataset used in Figure 2, StomataCounter only reached an average precision of 89.5%
113 (Figure 2A); it successfully detected 918 of 1,086 true stomata (missed 168 stomata) but falsely
114 detected 157 stomata with the best threshold. In addition, this tool failed to predict the contours of
115 the stomata and as a result could not quantify their sizes.

116 For pavement cell segmentation, we first tried PaCeQuant, a recently developed tool for
117 pavement cell segmentation and morphological analysis of fluorescence microscopy images
118 (Möller et al., 2017). This tool also performed well for 2D images converted from confocal images
119 by maximum intensity z-projection (Supplemental Figure S3A). When our dataset (representative
120 image in Figure 3A) was examined with PaCeQuant, almost all pavement cells were over-
121 segmented into tiny areas, and no reasonable results were generated (Figure 3B) in contrast to
122 LeafNet (Figure 3C), suggesting that PaCeQuant is not suitable for analyzing bright-field images.
123 We further examined the results from the PaCeQuant output in detail and concluded that
124 PaCeQuant is very sensitive to various types of noise in bright-field images, such as dots and lines
125 (Figure 3D, top row) or a dirty background (Figure 3D, middle row). The missing feature of
126 PaCeQuant to detect stomata also affected its performance to segment cells adjacent to those
127 stomata (Figure 3D, bottom row), whereas the accuracy of LeafNet in segmenting pavement cells
128 is substantially enhanced by masking stomata before performing pavement cell segmentation.

129 To further verify that the poor results of PaCeQuant were due to its low tolerance to noise,
130 we used the segmentation from LeafNet to generate input images without these types of noise for
131 PaCeQuant. In this case, PaCeQuant successfully segmented the pavement cells (Figure 3E, left),
132 as shown in the three areas in Figure 3E (right), in contrast to when PaCeQuant was directly applied
133 to the bright-field images (Figure 3D, middle column). These results indicate that LeafNet
134 performs well in tolerating various types of noise in bright-field images and that its hierarchical
135 strategy is effective for avoiding the interference from stomata during pavement cell segmentation.

136 To perform further morphological analysis, we implemented a script to parse the annotation
137 image generated by LeafNet and to directly feed the pavement cell segmentation results into
138 PaCeQuant without calling its own segmentation function. The combination of LeafNet's cell
139 segmentation and PaCeQuant's feature extraction enabled us to obtain 28 morphological
140 parameters such as perimeter, circularity, lobe counts, and so on in bright-field images. These
141 quantification results were then visualized within the segmentation (Supplemental Figure S4). In
142 addition, the annotation image from LeafNet can be manually corrected using GIMP or Photoshop
143 before extracting the morphological features (see Methods for details).

144

145 **Quantitative evaluation of pavement cell segmentation using LeafNet and other programs**

146 We quantitatively evaluated LeafNet using three different metrics: recognition quality (RQ,
147 evaluating LeafNet's ability to correctly report pavement cells from an image), segmentation
148 quality (SQ, evaluating LeafNet's ability to closely match the predicted borders with cell walls),
149 and panoptic quality (PQ, the product of SQ and RQ; see Methods for details). Meanwhile, we
150 used the same dataset to investigate the performance of ITK morphological watershed and several
151 recently developed programs, including PaCeQuant (Möller et al., 2017), PlantSeg (Wolny et al.,
152 2020), Cellpose (Stringer et al., 2021), and CSU-CN from Cell Segmentation Benchmark. As
153 expected, LeafNet performed much better than PaCeQuant, especially for recognition quality and
154 panoptic quality (Figure 4A), with a representative example shown in Figure 4B. Consistent with
155 the results of our comparison based on average precision (Figure 2C), the performance of ITK
156 morphological watershed was worse than that of LeafNet with the LeafSeg module for these three
157 metrics (Figure 4A, B).

158 PlantSeg, which contains a pretrained model to segment 2D images, performed well on 2D
159 images converted from confocal images by maximum intensity z-projection (Supplemental Figure
160 S3B). However PlantSeg's performance on our dataset was worse than LeafNet's, probably
161 because the pretrained model was not trained for bright-field images (Figure 4A, B). Please note
162 that PlantSeg has multiple algorithm options for pavement cell segmentation, and the results are
163 based on the default GASP algorithm.

164 Cellpose accurately segmented convex polygon-like cells in bright-field images
165 (Supplemental Figure S3C). On our dataset of bright-field images from Arabidopsis leaves,
166 Cellpose did not perform well with its pretrained cyto model. We retrained a new model called
167 Cellpose-retrained with the images from our dataset. Cellpose-retrained showed improvement over
168 Cellpose; however, its performance was still worse than LeafNet's (Figure 4A, B), probably
169 because the Cellpose algorithm does not support non-convex-shaped cells.

170 We chose the CSU-CN method as a representative tool from the Cell Segmentation
171 Benchmark in Cell Tracking Challenge (<http://celltrackingchallenge.net/>) because it achieved the
172 highest score on the Fluo-N2DH-GOWT1 dataset, which contains high-contrast bright cells
173 separated by a dark background, similar to our images. CSU-CN accurately segmented the image
174 from the testing set Fluo-N2DH-GOWT1 (Supplemental Figure S3D). However, CSU-CN failed

175 to segment our bright-field images of Arabidopsis leaves (Figure 4A, B), likely because all of its
176 training data were morphologically different from our images.

177 These results indicate that LeafNet well tolerates various types of noise when bright-field
178 images are used, and its hierarchical strategy is effective for avoiding the interference from stomata
179 during pavement cell segmentation.

180

181 **Extension of LeafNet to confocal images**

182 To explore LeafNet's flexibility and broad utility, we first examined its performance using bright-
183 field images from a different plant species, tobacco (*Nicotiana tabacum*). As expected, as shown
184 by a representative example in Figure 5A, LeafNet can faithfully identify the two stomata and
185 segment pavement cells, as in the ground truth image (manual labeling as the gold standard), while
186 StomataCounter can only count stomata without giving exact borders. For our dataset of 14 images
187 from *N. tabacum*, there were 79 stomata and 285 pavement cells based on ground truth labeling.
188 LeafNet successfully detected 75 and falsely detected 8 stomata, while StomataCounter
189 successfully detected only 36 and falsely detected 30 stomata. The overall performance of LeafNet
190 (F1 score 92.6%) was much better than that of StomataCounter (F1 score 49.7%) (Figure 5B). The
191 performance of StomataCounter for *N. tabacum* was much worse than that for *A. thaliana*, while
192 LeafNet's performance showed a small loss (Figure 5B, less than 5% in F1 score). On tobacco
193 pavement cells, LeafNet achieved a good panoptic score that appeared to be slightly better than
194 that for *A. thaliana* (Figure 5C), pointing to LeafNet's good adaptability to species with similar
195 cell morphology.

196 Next, we extended LeafNet to analyze confocal images. Confocal imaging can generate a 3D
197 stack of images of leaf tissue, as shown in Figure 5D (top panel) (Erguvan et al., 2019). As LeafNet
198 is designed to analyze 2D bright-field images, we established a pipeline to convert 3D image stacks
199 from confocal or light sheet microscopes to a maximum intensity z-projection 2D image using
200 MorphoGraphX (Barbier de Reuille et al., 2015) and ImageJ (Schindelin et al., 2012) (Figure 5D
201 bottom, Supplemental Figure S5A, and see Methods for details). We noticed that LeafNet's default
202 mode accurately segmented pavement cells but had difficulty in precisely identifying stomata,
203 probably because new confocal patterns were not observed in bright-field images. Thus, we trained
204 a StomaNet confocal model using transfer learning on a confocal dataset of Arabidopsis leaves

205 (Erguvan et al., 2019) after pre-processing them into 2D images and manually labeling all stomata
206 and pavement cells (Figure 5E).

207 We compared LeafNet with four other state-of-the-art programs for analyzing confocal
208 images, including StomataCounter for stoma detection and PlantSeg, PaCeQuant, and
209 MorphoGraphX for pavement cell segmentation (Figure 5F-H and Supplemental Figure S5B).
210 PlantSeg and PaCeQuant perform segmentation on maximum intensity z-projections of image
211 stacks in a similar manner to LeafNet, while MorphoGraphX performs 2.5D surface segmentation
212 on the original image stacks, and we converted the results to 2D segmentation with z-projection,
213 allowing us to compare the results with other tools (see Methods for details).

214 For the 38 ground truth stomata, LeafNet successfully detected 37 and falsely detected 1
215 stoma, while StomataCounter successfully detected 29 and falsely detected 5 stomata: the
216 performance scores are summarized in Figure 5F. For the 92 ground truth pavement cells, LeafNet
217 achieved a panoptic score of 85.0%, while MorphoGraphX obtained a score of 66.1% and PlantSeg
218 obtained a score of 81.0%. PaCeQuant accurately identified most pavement cells but falsely
219 recognized stomata as 74 pavement cells. As PaCeQuant rejects pavement cells that are adjacent
220 to image edges, we only quantified the performance of MorphoGraphX and PlantSeg to compared
221 with LeafNet (Figure 5G). LeafNet achieved the highest accuracy in terms of both stomata
222 (LeafNet F1 score of 0.97 vs. 0.80 for StomataCounter) and pavement cells (LeafNet F1 score of
223 0.85 vs. 0.66 for MorphoGraphX and 0.81 for PlantSeg). In the representative patch in Figure 5H
224 (see complete image in Supplemental Figure S5B), StomataCounter missed true stoma,
225 MorphoGraphX and PlantSeg mis-segmented pavement cells, and PaCeQuant frequently
226 miscalled one or more pavement cells in one stoma.

227 In summary, LeafNet outperforms the state-of-the-art programs in handling confocal images
228 and has good adaptivity to different species, pointing to its flexibility and potential broad utility.

229

230 **Extension of LeafNet to a wide range of species**

231 To further examine the capability of LeafNet to analyze images from a broad range of species and
232 images obtained using different micrographic methods, we performed systematic comparisons of
233 datasets with a variety of images.

234 For stoma detection, we trained another universal model in StomaNet using the same deep
235 learning network structure with the training data used by StomataCounter (Fetter et al., 2019). The
236 training data contain more than 4000 leaf epidermal images from more than 600 species taken by
237 DIC microscopy, SEM, and bright-field microscopy; however, no label information is available
238 for this dataset (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.kh2gv5f>). We used
239 StomataCounter's prediction on 960 images for rough training and then manually labeled 140
240 randomly selected images for further training. The StomaNetUniversal model has a similar
241 performance to StomataCounter on 47 manually labeled testing images randomly selected from
242 the testing set (Figure 6A). StomaNet achieved an average precision of 86.6%, similar to that of
243 StomataCounter (86.7% average precision) (Figure 6B).

244 For pavement cell segmentation, we used a dataset containing leaf epidermal images from
245 different species with stained cell walls (Vöfely et al., 2019). The authors indicated that it is very
246 difficult to perform automatic segmentations on this dataset due to various image defects, which
247 prompted them to manually track the boundaries of pavement cells. The boundaries for each
248 pavement cell were individually saved as coordinates relative to the cell center, but their positions
249 in the original image are not available. We successfully mapped 4188 pavement cells to 223 leaf
250 epidermal images from 86 different species as our testing dataset (Figure 6D, see Supplemental
251 Figure S6A, B for three representative examples), which are provided in LeafNet's website
252 (<https://leafnet.whu.edu.cn/suppdata/>).

253 Based on this large testing dataset, we evaluated the LeafSeg module of LeafNet compared
254 to other existing methods including ITK morphological watershed, PlantSeg, PaCeQuant, and
255 Cellpose. LeafSeg and Cellpose were more robust to various types of noise in regularly shaped
256 cells comparing to the other programs (Figure 6E top and Supplemental Figure S6C, D left), and
257 they both achieved F1 scores >0.95 and panoptic quality scores >0.75 for more than 30% of the
258 images (Figure 6F, G). However, Cellpose showed reduced performance for images with puzzle-
259 shaped pavement cells or with uneven lighting, which had little impact on LeafSeg's performance
260 (Figure 6E-G and Supplemental Figure S6C, D middle and right). PlantSeg and ITK morphological
261 watershed showed similar performance for images with both regularly shaped and puzzle-shaped
262 cells (Figure 6E and Supplemental Figure S6E, F), but their overall performance scores were worse
263 than that of LeafSeg (Figure 6F-G). We also tested CSU-CN in this dataset, but the results were
264 unusable (Supplemental Figure S6G).

265 Overall, for this complex dataset, LeafSeg had the best performance, achieving an average
266 F1 score of 0.74 and a panoptic quality score of 0.64, while ITK morphological watershed obtained
267 scores of 0.64 and 0.52, Cellpose obtained scores of 0.58 and 0.50, and PlantSeg obtained scores
268 of 0.40 and 0.35, respectively. We then examined the predicted pavement cells compared with the
269 ground truth for LeafSeg and Cellpose, which generated more acceptable results (F1 score >0.95
270 and panoptic quality >0.75) than PlantSeg and ITK morphological watershed. Of the 4188
271 manually labeled pavement cells, LeafSeg correctly segmented 3050 cells (72.8%), under-
272 segmented 697 cells (16.6%) into 304 cells, and over-segmented 439 cells (10.4%) into 1194 cells
273 (Figure 6H). Cellpose correctly segmented 2535 pavement cells (60.5%), under-segmented 94
274 cells (3.5%) into 39 cells, over-segmented 504 cells (12.0%) into 2485 cells, and reported 1055
275 cells as background (Figure 6I). These results are consistent with the finding that Cellpose
276 performed worse than LeafSeg based on F1 score and panoptic quality.

277 In summary, these results suggest that the performance of StomaNet for stoma counting is
278 similar to that of StomataCounter based on the 140 manually multi-species training data and show
279 that StomaNet can obtain accurate boundaries for stomata with clear signals. For pavement cell
280 segmentation, LeafSeg showed the best performance among the tools examined on a large set of
281 images with various defects from different species. By combining StomaNet and LeafSeg, LeafNet
282 can be extended to a wide range of species and to images taken using different methods.

283

284 **LeafNet detects significant biological differences**

285 Next, we applied LeafNet to evaluate its ability to automatically analyze large-scale microscopy
286 image datasets and to assess its difference from manual labeling using statistics. In total, we
287 analyzed 460 images using LeafNet, manually inspected the segmentations of stomata and
288 pavement cells, and recorded the correct counts as ground truth. We compared and evaluated the
289 differences between the predicted and manual results. The predicted counts of stomata and
290 pavement cells had good linear relationships with the manual counts (Figure 7A-B), and the
291 deviation of counts by image showed a tight distribution centered around 0 (Figure 7C-D), with a
292 mean absolute error of 5.80% and 5.45% for stomata and pavement cells, respectively.

293 Furthermore, we tested whether LeafNet can detect statistically significant differences in the
294 densities of stomata and pavement cells between two different *Arabidopsis* genotypes in the

295 *Columbia* (*Col-0*) genetic background (M1 and M2). M1 is *Pro35S:PIF4* (expressing
296 *PHYTOCHROME INTERACTING FACTOR4* under the control of the 35S promoter), and M2 is
297 the wild-type control (*Col-0*). We compared the counts of stomata and pavement cells generated
298 by LeafNet with those obtained from a manual annotation of 40 images. As shown in the
299 representative examples for stomata (Figure 7E) and for pavement cells (Figure 7F), we observed
300 consistent results between the predicted (blue) and manual (orange) results for non-significant
301 differences in M1 or in M2 (not significant *P*-value), and significant differences between M1 and
302 M2, from paired *t*-tests.

303 These results suggest that LeafNet is a useful tool for plant biologists to process large numbers
304 of images and quantify the biological differences in a reliable manner.

305

306 Using LeafNet in the CLI, GUI, and web server

307 To make LeafNet widely accessible to different users, we designed a standalone program that can
308 be run on most computer systems. We also developed a graphic user interface (GUI, Supplemental
309 Figure S7A) and a web server (Supplemental Figure S7B) for users without Linux experience, and
310 a command-line interface (CLI) for experienced users with servers (Supplemental Figure S7C).
311 The web server is hosted at <https://leafnet.whu.edu.cn/>. We have also created two Conda packages
312 for both CPU and GPU environments, and thus users can easily install LeafNet with one command
313 “conda install -c anaconda -c conda-forge -c zhoyulab leafnet(-gpu)” in Linux, Mac OS, or
314 Windows systems.

315 LeafNet can use images generated from bright-field, confocal z-projection, and other imaging
316 methods as input by using different modules. During preprocessing, Peeled denoiser works with
317 bright-field images from peeled leaf epidermal (e.g. Figure 2A) and with confocal images (e.g.
318 Figure 5D), while Stained denoiser works with leaf epidermal images with stained cell walls (e.g.
319 Supplemental Figure S6A) and is recommended for other types of images. For stoma detection,
320 StomaNet is trained for bright-field images from Arabidopsis by peeling off leaf epidermis,
321 StomaNetConfocal is trained for confocal images from Arabidopsis, and StomaNetUniversal is
322 trained to detect stomata in a wide range of species.

323 LeafNet can generate three types of output results. The first type is a preview image provided
324 for visualizing the segmentation, which is shown on the original image with cells of different colors
325 and stomata marked in blue. The second type is an annotation image provided for further analysis,
326 which uses green lines to label pavement cell walls and labels stomata with blue ellipses. The third
327 type is a statistical text with quantified data, including morphological features of the leaf epidermal
328 image, such as the counts and sizes of stomata and pavement cells ([Supplemental Figure S8A-C](#)).
329 All three output results are available in the CLI, GUI, and web server.

330 In addition, users can use LeafNet in annotation mode and manually correct the output. Users
331 can load the annotation image from LeafNet together with the input image into GIMP or Photoshop
332 and then correct the annotation image for further analysis (see Methods for details). Existing
333 epidermal image processing pipelines could also benefit from LeafNet by simplifying the image
334 annotation procedure, as the annotation image generated by LeafNet could easily be handled by
335 other tools. [Supplemental Figure S4](#) shows a feasible scenario in which manual correction can be
336 performed on the annotation image from LeafNet and the corrected image can be fed into
337 PaCeQuant to extract morphological information.

338 We also provide training utilities to extend LeafNet for other types of images. Advanced users
339 can perform transfer learning to improve the stoma detection model, and our investigations and
340 codes provide an exemplar workflow. New types of stomata for other species can be analyzed
341 using newly labeled images with transfer learning in a similar manner ([Figure 5E](#)). Meanwhile,
342 LeafNet's segmentation can be used as a good starting point for further manual correction to
343 efficiently construct training data.

344

345 **DISCUSSION**

346 Here we introduce LeafNet, a fully automatic program capable of precisely detecting stomata and
347 segmenting pavement cells. We devised a hierarchical strategy to accurately identify stomata first
348 and then segment pavement cells in stomata-masked images. By incorporating two modules,
349 StomaNet and LeafSeg, LeafNet sequentially conquers the two challenges of precisely detecting
350 stomata and segmenting pavement cells while avoiding the interference between these two types
351 of objects with different characteristics. The StomaNet module accurately segments stomata using
352 a deep neural network. The LeafSeg module tolerates various types of noise and puzzle-shaped

353 cell shapes, exhibiting acceptable performance on bright-field images using a region merging
354 algorithm. LeafNet adapts to images from a broad range of species and outperforms several state-
355 of-the-art programs, enabling biologists to perform fast and simple experiments using easily
356 accessible bright-field microscopy, SEM, or confocal microscopy.

357 Object detection and cell segmentation are classic tasks in biological image processing, and
358 many tools have been built for different scenarios. At the level of stomata detection,
359 StomataCounter (Fetter et al., 2019) performed well on different types of images. StomaNet
360 achieved higher accuracy than StomataCounter for bright-field images from Arabidopsis or species
361 with similar morphology using 140 training images. StomaNetUniversal achieved similar accuracy
362 to StomataCounter on images from a wide range of species using 960 roughly labeled and 140
363 manually labeled training images, in contrast to StomataCounter's much larger training data of
364 4618 images, suggesting that StomaNet requires less training data than StomataCounter and that
365 StomaNetUniversal has the potential to evolve once more training data have been introduced.
366 Moreover, StomaNet's ability to accurately segment the borders of stomata enables LeafNet to
367 mask stomata, which prevents the stomata from interfering with subsequent pavement cell
368 segmentation.

369 For pavement cell segmentation, we tested three non-deep learning-based tools including
370 MorphoGraphX (Barbier de Reuille et al., 2015), PaCeQuant (Möller et al., 2017), and ITK
371 morphological watershed (McCormick et al., 2014), and three recently developed deep learning-
372 based tools including PlantSeg (Wolny et al., 2020), Cellpose (Stringer et al., 2021), and CSU-CN
373 (from Cell Tracking Challenge), based on a systematic evaluation on our own and published
374 datasets. MorphoGraphX has been used to perform surface segmentation on confocal image stacks
375 (Sapala et al., 2018). However, its performance was slightly worse than that of LeafNet and
376 PlantSeg (Figure 5G-H), and it could not process bright-field images. PaCeQuant could not tolerate
377 the noise, uneven lighting, and inconsistent border signals in bright-field images. ITK
378 morphological watershed showed better tolerance to image defects, but it did not perform as well
379 as LeafSeg (Figure 4 and Figure 6). PlantSeg, Cellpose, and CSU-CN achieved state-of-the-art
380 performance on their own preferred input images (Supplemental Figure S3B-D), but their
381 performance on bright-field leaf epidermal images was worse than that of LeafSeg, especially on
382 puzzle-shaped pavement cells (Figure 4 and Figure 6). Based on the results for these representative
383 tools, we conclude that the segmentation of pavement cells in bright-field images is challenging

384 and that LeafSeg represents a significant advancement: it is well-adapted to this task and performs
385 even better than the three deep learning-based methods examined.

386 To explore the possibility of using a convolutional neural network (CNN) to enhance cell
387 wall signals as in PlantSeg, we retrained a PlantSeg CNN model with 100 training images from
388 our dataset in Figure 4. We applied our Stained Denoiser before training and prediction to improve
389 its generalization ability. We evaluated the performance of this retrained network (named
390 CNNwall) on 40 other images from the same dataset. With CNNwall enhancement, PlantSeg's
391 panoptic quality increased from 13.5% to 70.4% on testing images, outperforming ITK
392 morphological watershed (59.1%), and LeafSeg's panoptic quality increased from 77.4% to 81.4%
393 ([Supplemental Figure S9A](#)).

394 We further tested CNNwall's broad utility using the Vofely dataset (Vófély et al., 2019) used
395 in Figure 6D-I. PlantSeg's panoptic quality increased from 34.5% to 55.7%, outperforming ITK
396 morphological watershed (52.9%), suggesting that CNNwall is better adapted to various input
397 images than PlantSeg's original model ([Supplemental Figure S9B-C](#)). However, LeafSeg's
398 panoptic quality dropped from 64.5% to 56.5%, a value similar to PlantSeg's (55.7%)
399 ([Supplemental Figure S9B-C](#)). As illustrated in the exemplar images, we reasoned that although
400 CNNwall could enhance the signals of cell walls ([Supplemental Figure S9D](#)), it could also
401 introduce extra artifacts when input images are different from those in the training dataset, which
402 would impair cell segmentation ([Supplemental Figure S9E](#)). Consistently, the differences in
403 pavement cells in a wide range of species and using different imaging methods had a huge impact
404 to the performance of previously reported deep learning-based cell segmentation methods
405 ([Supplemental Figure S3B-D](#) and [Supplemental Figure S6D, E, G](#)). Therefore, a much larger
406 dataset may be required to train a universal deep learning-based pavement cell segmentation model
407 than to train a good stoma detection model. Considering that manually labeling pavement cell
408 boundaries takes more time than labeling stomata, we currently provide CNNwall-enhanced
409 LeafSeg as an optional method and provide the original LeafSeg as our default universal method
410 for pavement cell segmentation.

411 Nevertheless, we believe that pavement cell segmentation and stoma detection could be better
412 solved using a single joint deep learning model in the future. Many articles reported to date by the
413 plant community only contain morphological information but do not provide manually corrected

414 segmentation. Platforms such as Cell Tracking Challenge (Ulman et al., 2017) provide different
415 types of images from cultivated cell lines with pixel-wise labels for researchers and programmers
416 to test and compare their methods, but there no such platform is currently available for pavement
417 cell segmentation and stoma detection. We believe that the plant community needs more well-
418 labeled datasets, and thus we shared all our training datasets, testing datasets, and the results from
419 LeafNet and existing tools in the Download page of the LeafNet web server. In the future, as more
420 datasets are released by researchers, a deep learning-based universal model could be created to
421 segment pavement cells and stomata simultaneously with better accuracy and to further enhance
422 the performance of automatic morphological analysis tools on bright-field images.

423 The LeafNet program, the associated web server, strategy, and codes are provided for the
424 plant community with the potential to replace manual work, enhance productivity, and increase
425 reproducibility. We have shown that LeafNet is flexible and can be extended to different species
426 or confocal images, and we anticipate that it will be useful for a broad range of researchers
427 interested in quantifying stomata and pavement cells.

428

429 **METHODS**

430 **Plant culture**

431 All *Arabidopsis thaliana* plants used in this study were of the *Columbia (Col-0)* genetic
432 background. The seeds were sterilized with 75% ethanol for 2 min and 2% sodium hypochlorite
433 solution for 15 min. The seeds were sown in Petri dishes containing 1/2 strength Murashige and
434 Skoog medium solidified with agar and placed at 4°C for 3 days in complete darkness, followed
435 by growth under short days (10 h light/14 h dark) or long days (14 h light/10 h dark) at 22°C.
436 Seedlings were grown under white fluorescent light at a light intensity of 100 $\mu\text{mol photons/m}^2/\text{s}$.

437 **Image collection**

438 The light microscope images in this study were taken using a modified method as described
439 (Engineer et al., 2014). Briefly, plant tissues (leaves) were sampled throughout plant growth. To
440 obtain epidermal peels, glass slides (CITOGLAS 9821) were sprayed with Hollister Medical
441 Adhesive (3.8 oz. Spray HH7730), and the abaxial epidermal surfaces of leaves from independent
442 seedlings were gently pressed onto the slides. The mesophyll tissues were removed from the slides

443 with a blade, and the epidermal peels were imaged under a Leica DMi8 microscope at 200 x
444 magnification.

445

446 **Manual labeling to create the training dataset**

447 We used the GIMP program to annotate stomata and pavement cells in input images. Briefly,
448 we created a new annotation layer on top of the sample layer and set its opacity to 50%. We
449 manually labeled the boundaries of pavement cells with a green line with 100 hardness and filled
450 the stomata with blue coloring on the annotation layer. We then set the opacity of the annotation
451 layer back to 100% and set its background to black. Finally, we removed the sample layer, flattened
452 the image, and saved it as an annotation figure. The manual annotations were validated by one or
453 more other annotators independently.

454

455 **LeafNet workflow**

456 The LeafNet workflow consists of image preprocessing, the StomaNet module, and the
457 LeafSeg module.

458 **Preprocessing.** Images taken under a light microscope can be noisy and must be denoised
459 before training or prediction. LeafNet has two different preprocessing modules. For our images of
460 the epidermal surface peeled from leaves, we used the Peeled Denoiser based on the noise
461 reduction function from the generic graphics library GEGL. Preprocessing involves the following
462 steps: 1) resize the image to the resolution of the trained model in PIL with Image.ANTIALIAS;
463 2) invert the image only if it is a fluorescence image; 3) separate the image into dark and bright
464 parts with Otsu threshold; 4) perform an adaptive linear transformation on the bright part to scale
465 the mean of pixel gray scale to 200; 5) merge the dark part and transform the bright part; 6) denoise
466 the image with the noise reduction function from GEGL.

467 For other types of images, we used the Stained Denoiser. This denoiser is better adapted to
468 different types of images and is recommended for most scenarios. This denoiser involves the
469 following steps: 1-2) the same steps as for the Peeled Denoiser; 3) apply a median filter to the
470 image; 4) apply a high-pass filter to the image; 5) perform adaptive area normalization on the
471 image; 6) apply mean curvature blur on the image.

472

473 **The StomaNet module for detecting stomata**

474 **StomaNet input.** As we used valid convolution and transpose convolution without padding,
475 a fixed network size is needed, as the layer output size must be a positive integer. Therefore, the
476 sliding window method is used for network input, including the following steps: 1) As the network
477 output is smaller than the input, denoised images are copy padded with OpenCV using the border
478 reflect method; 2) the sliding window method is used to split the input images into smaller images
479 as network input. The input images are broken into patches whose sizes match the network input
480 (186 for StomaNet) with a stride of network output (104 for StomaNet).

481 **StomaNet network architecture.** StomaNet is a deep residual network inspired by ISL
482 (Christiansen et al., 2018) built with TensorFlow and Keras. The network comprises three sub-
483 networks (called towers) of different input sizes to collect information at different scales. The
484 towers are composed of residual blocks. Residual blocks are sub-networks of the structure shown
485 in [Supplemental Figure S2A](#). Three types of residual blocks are used in the network: down-scale,
486 in-scale, and up-scale, the parameters of which are shown in [Supplemental Figure S2B](#).

487 Residual blocks consist of two parts. The first part directly copies the data to output, forming
488 a residual connection. Up sampling, cropping, and average pooling are used in different types of
489 residual blocks to keep the data in the same shape with convolution filters. The second part
490 consists of two convolution layers, which form an encoder-decoder structure. Input is batch
491 normalized and activated by ReLU and Tanh (concatenated), encoded by a convolution layer
492 called Conv2D expand. The output of the encoder is concatenated with the max pooling result of
493 block input, batch normalized, and activated by ReLU and Tanh (concatenated) again. The 1*1
494 convolution is used as a decoder to reduce the count of filters to make it match the block input.
495 The results of the two steps are added to generate the output of a residual block.

496 **StomaNet output.** The output of the network are images that represent the pixel-wise
497 plausibility of stomata in the area (104 for StomaNet) of the center of the input (186 for StomaNet).
498 As the input is generated by the sliding window approach, output images are stacked, generating
499 a full-sized probability heatmap of stomata.

500 **Training samples.** For StomaNet, we manually labeled stomata in blue (0,0,255) in raw
501 images from Arabidopsis using GIMP software, and saved the annotated images with the same
502 name and resolution as the original image in a label folder. A 4 px Gaussian Blur was applied to
503 stomata labels. The labeled images were broken into patches using the sliding-window method

504 (the step length equals the size of the network output in generating non-overlapping results) and
505 split into a training set and validation set. An additional 15 negative images (including 5
506 photographs, 5 instances of random noise, and 5 different cells) without any stomata were added
507 to the training set to prevent overfitting. For StomaNet confocal mode, we labeled six additional
508 2D images transformed from 3D confocal images for transfer learning. For StomaNetUniversal
509 model, we manually labeled 140 images of diverse species using the same method.

510 **Model training.** StomaNet was built to be trainable with devices accessible to most
511 researchers in an acceptable timeframe. We trained StomaNet with 4 Tesla K40m in approximately
512 2 hours, and it could be trained with a GTX 1060 6G in ~8 hours. It is possible to train StomaNet
513 with cheaper video gaming cards, but the batch size should be decreased according to the VRAM
514 of the card used. We used Nadam optimizer with a batch size of 25 for StomaNet.

515 **Transfer learning.** Transfer learning is implemented by using another pre-trained model's
516 weights as initial weights and then applying the above training procedure to fine-tune the model.
517 We did not freeze the weights of any layers in the initial model, as low-level features also vary
518 using different imaging methods. The StomaNetConfocal model was trained with six manually
519 labeled images based on StomaNet's model. The StomaNetUniversal model was trained with 140
520 manually labeled images based on an initial model trained on 960 images with rough labels from
521 StomataCounter's prediction.

522 **Prediction.** The prediction of stomata involves three steps: 1) stack results from the network
523 to produce a heatmap with the same size as the input image; 2) assign the score as weight for the
524 pixels in the heatmap with scores > 0.5 and perform DBSCAN (eps = 10, minimum samples = 40)
525 on these pixels to generate clusters; 3) perform PCA on each cluster and use the two principal
526 components to describe the stoma as an ellipse. The stoma center is the weighted average of all
527 pixels, and stoma size is the count of all pixels in the cluster multiplied by the size correction ratio
528 (0.85 for StomaNet). The stoma length/width ratio is computed as $(PC1/PC2)^{0.5}$, and stoma angle
529 is the angle of PC1.

530

531 **The LeafSeg module for segmenting pavement cells**

532 After detecting stomata in the input image, the LeafSeg module segments pavement cells
533 using a region merging algorithm as follows: 1) mask stomata in black (grayscale=0) with the

534 length and width multiplied by 1.5 and copy the masked image; 2) perform a median blur on the
 535 masked image; 3) binarize the blurred image with Otsu threshold and copy it; 4) skeletonize the
 536 image, remove isolated skeletons <64 px in size, and dilate the remaining skeletons by 4 px to
 537 obtain smooth borders; 5) perform a Euclidean distance transform on the border image; 6) perform
 538 the watershed algorithm on the masked image from step 1, with the peaks in the distance image
 539 from step 5 as different labels; 7) obtain border dots from watershed areas, and define the border
 540 score as the average value of the binarized image from step 3 under the border; 8) merge the areas
 541 based on border score. An area can only merge with one other area (with the lowest score) at a
 542 time. Border score is recalculated after each merge. The final borders are used as the boundaries
 543 of pavement cells for counting and statistical analysis.

544 The borders of an image (50 px wide after resizing) are considered invalid areas, which are
 545 marked with a red line in the segmentation results. If the stoma center is located in a valid area,
 546 the stoma is counted as 1 stoma; otherwise, it is not counted in the final statistical result. Pavement
 547 cells covering the edge of a valid area are counted as 0.5 pavement cell and are not included in
 548 other statistical results such as cell size.

549

550 **Metrics for evaluating pavement cell segmentation**

551 To evaluate pavement cell segmentation, three metrics were used: recognition quality (RQ),
 552 segmentation quality (SQ), and panoptic quality (PQ). The PQ is the product of SQ and RQ, which
 553 are defined as follows:

$$554 \quad PQ = \underbrace{\frac{\sum_{(p,g) \in TP} \text{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + 0.5|FP| + 0.5|FN|}}_{\text{recognition quality (RQ)}},$$

555 where TP represents true positive; FP represents false positives; FN represents false negatives; p
 556 and g represents prediction and ground truth; IoU stands for Intersection over Union; and $|x|$
 557 represents the number of x .

558

559 **2D segmentation with ITK morphological watershed, PlantSeg, PaCeQuant, and Cellpose**

560 For ITK morphological watershed, we applied the following steps for a fair comparison with
 561 LeafNet: 1) preprocess the image with Stained Denoiser, as described in the LeafNet workflow; 2)
 562 perform ITK morphological watershed on the preprocessed image, try different segmentation

563 levels, and choose the one with the best panoptic quality; 3) remove the areas darker than the Otsu
564 threshold of the full image to remove cell walls; 4) perform watershed with the remaining areas as
565 labels to fill in the removed areas.

566 For PlantSeg, we used the `confocal_2D_unet_bce_dice_ds3x` model for boundary detection,
567 applied rescaling based on the image resolution, and used the GASP algorithm for segmentation.
568 We tried different Under-/Over-segmentation factors and CNN prediction thresholds and chose
569 the values with the best panoptic quality, and used the default values for the other parameters.

570 For PaCeQuant and Cellpose, we only set one parameter based on the input image resolution
571 and used default values for other parameters.

572

573 **Pre-processing of 3D images to 2D images**

574 MorphoGraphX is used to generate 2D images from 3D image stacks via the following
575 processes: 1) use Stack->Canvas->ReverseAxes to reverse z-axis when the image stack is upside
576 down; 2) use Stack->Filters->Gaussian Blur Stack to denoise the image by 1 μm ; 3) use
577 Stack->Multi-stack->Copy Work to Main Stack to save the denoised stack; 4) use
578 Stack->Morphology->Edge Detect to create a solid shape; 5) use Mesh->Creation->Marching
579 Cube Surface to create a mesh surface; 6) use Mesh->Structure->Subdivide and
580 Mesh->Structure->Smooth Mesh to smooth the mesh; 7) use Stack->Multi-stack->Copy Main to
581 Work Stack to load the input signal; 8) use Stack->Mesh Interaction->Annihilate($\text{minDist}=6\mu\text{m}$,
582 $\text{maxDist}=8\mu\text{m}$) to remove the surface; 9) save the work stack, and use ImageJ to generate a
583 maximum intensity z-projection image.

584

585 **Surface segmentation with MorphoGraphX**

586 To segment 3D image stacks with MorphoGraphX, we first used the above steps 1-6 to create
587 a mesh surface and then performed the following operations: 1) use Stack->Multi-stack->Copy
588 Main to Work stack to load input signal; 2) use Mesh->Signal->Project Signal ($\text{minDist}=6\mu\text{m}$,
589 $\text{maxDist}=8\mu\text{m}$) to project the signal onto the mesh surface; 3) use Mesh->Segmentation->Auto-
590 Segmentation (blur for seeding = $5\mu\text{m}$, radius for auto seeding = $5\mu\text{m}$, blur for cell outlines = $1\mu\text{m}$,
591 normalize radius = $20\mu\text{m}$, border distance = $0.5\mu\text{m}$, merge threshold = 1.5 for input signal) to
592 segment the mesh surface.

593

594 **Prediction of stomata using StomataCounter**

595 To predict stomata with StomataCounter, we used the model named `sc_feb2019.caffemodel`,
596 set the scale parameter as 2 for our regular (*A. thaliana*, bright-field) and confocal (*A. thaliana*,
597 confocal) dataset, and set the scale as 1 for the *N. tabacum*, bright-field dataset. Confocal
598 maximum intensity z-projection images should be inverted for better performance. We tried
599 different values for the threshold parameter in console mode to call `stoma` and used the one with
600 the best F1 score (1.625 for the regular dataset, 2.25 for the *N. tabacum* dataset, and 0.3 for the
601 confocal dataset).

602

603 **Manual correction procedure with LeafNet**

604 All results of LeafNet reported in this article are original output without any correction. We
605 added this section for users to improve the segmentation results. The manual correction procedure
606 involves the following steps: 1) use LeafNet to generate a segmentation image; 2) load the sample
607 image and annotation image into GIMP or Photoshop; 3) place the annotation image in a layer
608 above the sample image; 4) set the opacity of the annotation layer to 50%; 5) correct pavement
609 cell boundaries with 100 hardness and in green (0,255,0) on the annotation layer, and correct
610 stomata with 100 hardness and in blue (0,0,255) on the annotation layer; 6) set the opacity of the
611 annotation layer back to 100%; 7) set the background of the image to black (0,0,0); 8) remove the
612 sample image layer; 9) flatten the image, and save the corrected annotation.

613

614 **Quantification and statistical analysis**

615 The quantifications of stomata and pavement cells are from LeafNet version 1.0. To test for
616 the differences between manual and LeafNet predictions across different genotypes (Figure 6E-F),
617 we used the paired two-tailed *t*-test.

618

619 **Data and Code Availability**

620 Source code and released LeafNet packages are available in the GitHub repository:
621 <https://github.com/zhouyulab/leafnet>, and the web application is available at

622 <https://leafnet.whu.edu.cn/>. The training data and the results from all the analysis, as well as
623 detailed configurations to run these tools, are available at <https://leafnet.whu.edu.cn/suppdata/>.

624

625

626 **Supplemental Data**

627 **Supplemental Figure S1.** Graphical representation of the LeafNet workflow.

628 **Supplemental Figure S2.** Detailed structure of StomaNet.

629 **Supplemental Figure S3.** Representative examples of segmentation results from four programs
630 using their preferred images.

631 **Supplemental Figure S4.** Integration of LeafNet with PaCeQuant.

632 **Supplemental Figure S5.** Pre-processing of a 3D image and representative results from five
633 programs.

634 **Supplemental Figure S6.** Representative examples of segmentation results from five tools using
635 complex datasets.

636 **Supplemental Figure S7.** User interfaces for LeafNet.

637 **Supplemental Figure S8.** Sample output for LeafNet.

638 **Supplemental Figure S9.** Performance of CNNwall enhancement for pavement cell
639 segmentation.

640

641

642 **ACKNOWLEDGEMENTS**

643 This study was supported by grants from the China NSFC projects (31922039 to Y.Z. and
644 31830057 to Y.X.Z.), the National Key R&D Program of China (2017YFA0504400 to Y.Z.), the
645 Foundation of Hubei Grants (2020CFA057 to Y.Z.), the Chinese Scholarship Council (predoctoral
646 fellowships to C.Z.) and the Ghent University “Bijzondere Onderzoekfonds” (BOF/CHN/010 to
647 C.Z.). Part of computation in this work was done on the supercomputing system in the
648 Supercomputing Center of Wuhan University. The authors thank the undergraduates in Class 4 of
649 Biological Science in Grade 2019 from College of Life Science at Wuhan University, especially
650 Keyu Lu, Dongzi Liu, Yuxuan Yang, Liran Mao, Yinghao Sun, Miao Lou, Zhiming He, and
651 Shilong Zhao for their help in manual annotation of data.

652

653 **AUTHOR CONTRIBUTIONS**

654 Y.Z., Y.X.Z., and K.W. designed the research. L.M.L. performed the experiments. L.M.L., S.P.L.,
 655 S.P.M., and J.C.K. generated manually labeled data. S.P.L., Y.Z., and C.Z. developed the
 656 computational method. S.P.L. implemented the standalone tool and W.L.F. implemented the web
 657 server. Y.Z., Y.X.Z., and E.R. wrote the paper with inputs from S.P.L. and L.M.L. All authors
 658 discussed the results and approved the manuscript.

659 **REFERENCES**

- 660 Barbier de Reuille, P., Routier-Kierzkowska, A.-L., Kierzkowski, D., Bassel, G. W., Schüpbach,
 661 T., Tauriello, G., Bajpai, N., Strauss, S., Weber, A., Kiss, A., et al. (2015).
 662 MorphoGraphX: A platform for quantifying morphogenesis in 4D. *eLife* 4:e05864.
- 663 Casson, S. A., and Hetherington, A. M. (2010). Environmental regulation of stomatal
 664 development. *Curr. Opin. Plant Biol.* 13:90–95.
- 665 Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount,
 666 E., O’Neil, A., Shah, K., Lee, A. K., et al. (2018). In Silico Labeling: Predicting
 667 fluorescent labels in unlabeled images. *Cell* 173:792-803.e19.
- 668 Engineer, C. B., Ghassemian, M., Anderson, J. C., Peck, S. C., Hu, H., and Schroeder, J. I.
 669 (2014). Carbonic anhydrases, EPF2 and a novel protease mediate CO₂ control of
 670 stomatal development. *Nature* 513:246–250.
- 671 Erguvan, Ö., Louveaux, M., Hamant, O., and Verger, S. (2019). ImageJ SurfCut: a user-friendly
 672 pipeline for high-throughput extraction of cell contours from 3D image stacks. *BMC Biol.*
 673 17:38.
- 674 Fetter, K. C., Eberhardt, S., Barclay, R. S., Wing, S., and Keller, S. R. (2019). StomataCounter: a
 675 neural network for automatic stomata identification and counting. *New Phytol.* 223:1671–
 676 1681.
- 677 Hashimoto, M., Negi, J., Young, J., Israelsson, M., Schroeder, J. I., and Iba, K. (2006).
 678 Arabidopsis HT1 kinase controls stomatal movements in response to CO₂. *Nat. Cell Biol.*
 679 8:391–397.
- 680 McCormick, M., Liu, X., Jomier, J., Marion, C., and Ibanez, L. (2014). ITK: enabling
 681 reproducible research and open science. *Front. Neuroinformatics* 8.
- 682 Möller, B., Poeschl, Y., Plötner, R., and Bürstenbinder, K. (2017). PaCeQuant: A tool for high-
 683 throughput quantification of pavement cell shape characteristics. *Plant Physiol.* 175:998–
 684 1017.

- 685 Qi, X., and Torii, K. U. (2018). Hormonal and environmental signals guiding stomatal
686 development. *BMC Biol.* 16:21.
- 687 Sapala, A., Runions, A., Routier-Kierzkowska, A.-L., Das Gupta, M., Hong, L., Hofhuis, H.,
688 Verger, S., Mosca, G., Li, C.-B., Hay, A., et al. (2018). Why plants make puzzle cells,
689 and how their shape emerges. *eLife* 7:e32794.
- 690 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
691 S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for
692 biological-image analysis. *Nat. Methods* 9:676–682.
- 693 Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist
694 algorithm for cellular segmentation. *Nat. Methods* 18:100–106.
- 695 Ulman, V., Maška, M., Magnusson, K. E. G., Ronneberger, O., Haubold, C., Harder, N., Matula,
696 P., Matula, P., Svoboda, D., Radojevic, M., et al. (2017). An objective comparison of
697 cell-tracking algorithms. *Nat. Methods* 14:1141–1152.
- 698 Vöföly, R. V., Gallagher, J., Pisano, G. D., Bartlett, M., and Braybrook, S. A. (2019). Of puzzles
699 and pavements: a quantitative exploration of leaf epidermal cell shape. *New Phytol.*
700 221:540–552.
- 701 Wolny, A., Cerrone, L., Vijayan, A., Tofanelli, R., Barro, A. V., Louveaux, M., Wenzl, C.,
702 Strauss, S., Wilson-Sánchez, D., Lymbouridou, R., et al. (2020). Accurate and versatile
703 3D segmentation of plant tissues at cellular resolution. *eLife* 9:e57613.
- 704 Zoulias, N., Harrison, E. L., Casson, S. A., and Gray, J. E. (2018). Molecular control of stomatal
705 development. *Biochem. J.* 475:441–454.
- 706

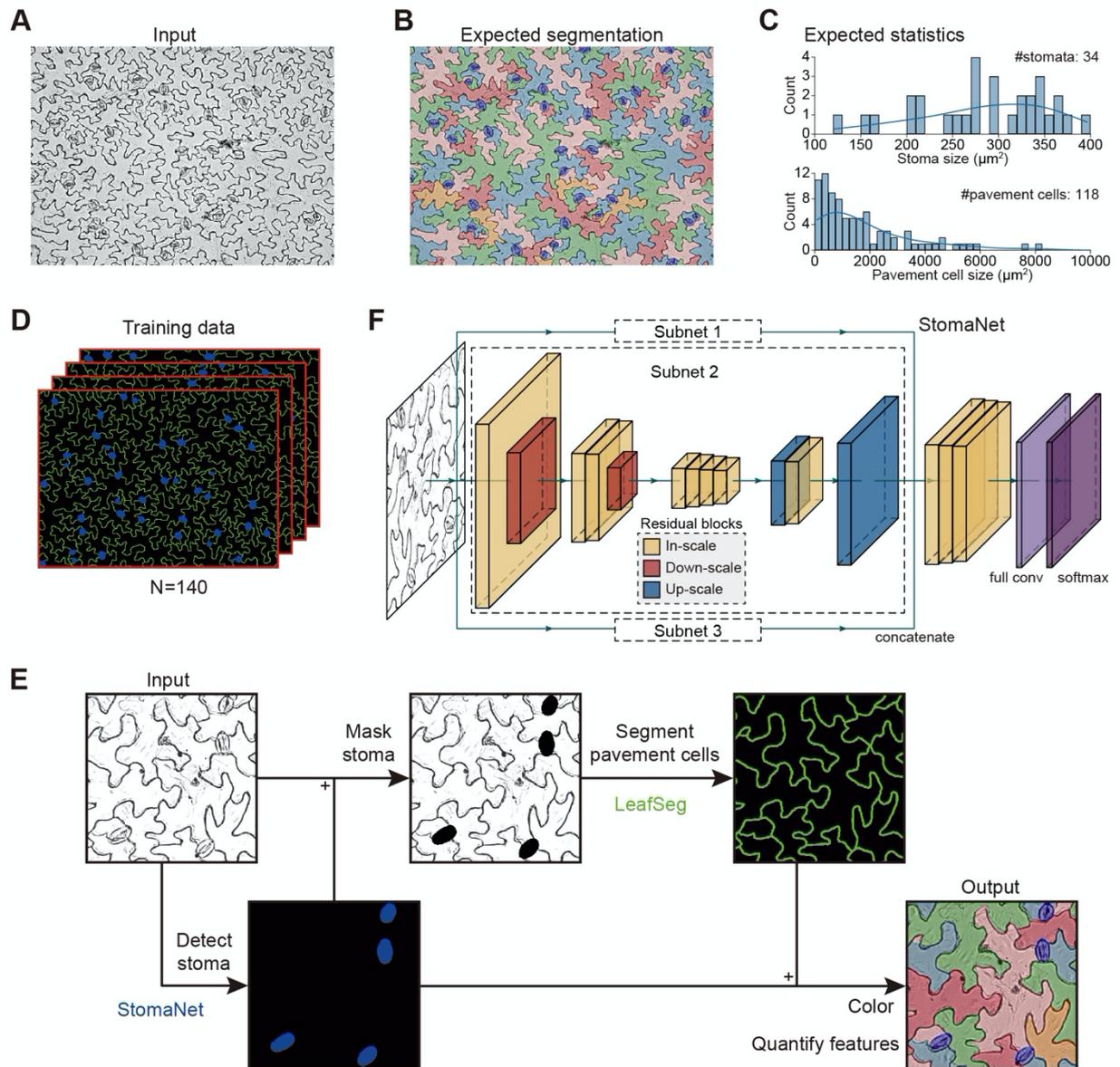


Figure 1. Hierarchical strategy of LeafNet to segment stomata and pavement cells.

(A) Representative bright-field image of stomata and pavement cells. (B) Representative result from manual segmentation of the input image in A. Stomata are labeled in blue, and pavement cells are filled with different colors. (C) Expected statistics from the segmentation in B on the size distribution for stomata (top) and for pavement cells (bottom). (D) Training data prepared from manual segmentation. The stomata are shown in blue, and the borders of pavement cells are labeled in green. (E) Hierarchical strategy and workflow of LeafNet with the StomaNet module for detecting stomata, and the LeafSeg module for segmenting pavement cells on stoma-masked input. A graphical illustration of each step is shown in Supplemental Figure S1. (F) Graphical illustration of the deep residual neural network for the StomaNet module. This module is composed of three subnets with in-scale (orange), down-scale (red), and up-scale (blue) residual blocks.

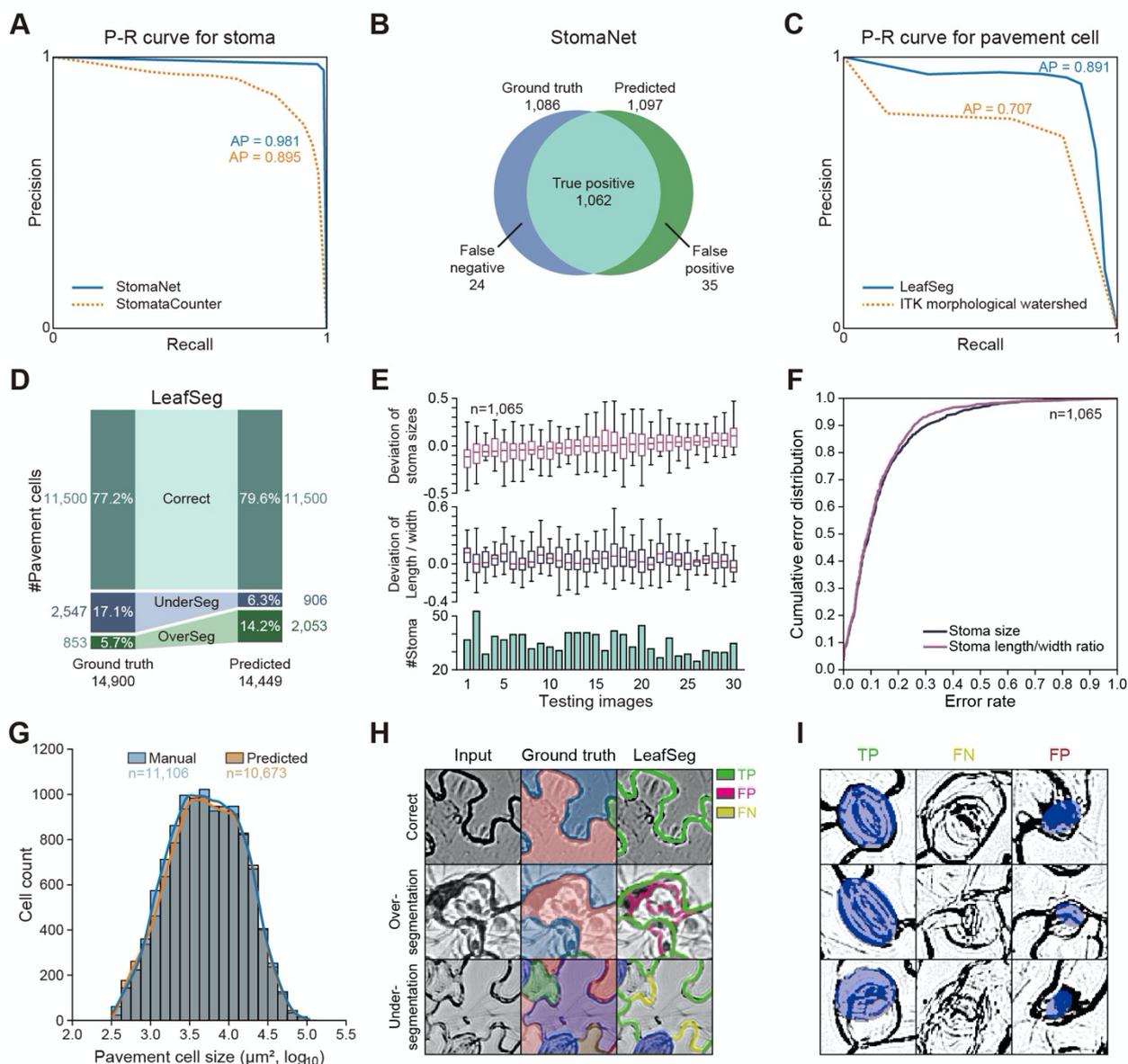


Figure 2. Performance of LeafNet for the recognition and quantification of stomata and pavement cells.

(A) Precision-recall curve of StomaNet (blue) and StomataCounter (orange dotted line) for counting stomata in 30 testing images. The thresholds were evaluated from 0.1 to 0.9 to calculate average precision (AP). (B) Venn diagram showing the performance of StomaNet for detecting stomata in the testing dataset using default settings. (C) Precision-recall curve for segmenting pavement cells with LeafSeg (blue) and the ITK morphological watershed algorithm (orange). (D) Performance of LeafSeg for segmenting pavement cells in the testing dataset using default settings. The numbers and percentages of correct, under-segmented (UnderSeg), and over-segmented (OverSeg) cells are shown. (E) Deviation to the ground truth of the quantification of size (top) and length/width ratio (middle) for stomata in each image. The per-image counts of stomata are shown as a bar graph (bottom). (F) Cumulative distribution of the size deviation (blue) and length/width ratio deviation (red) for all stomata ($n=1,065$) in the testing dataset. (G) Distribution of the pavement cell sizes obtained by manual annotation (blue) and LeafSeg prediction (orange). (H) Representative examples of LeafSeg predictions in three cases: correct (top row), over-segmentation (middle row), and under-segmentation (bottom row). The middle column (Ground truth) shows the correct segmentation of cells filled with different colors. In the right column, the true positive (TP), false positive (FP), and false negative (FN) borders are shown in green, red, and yellow, respectively. (I) Representative stomata correctly identified (TP), missed (FN), and falsely identified (FP) by StomaNet.

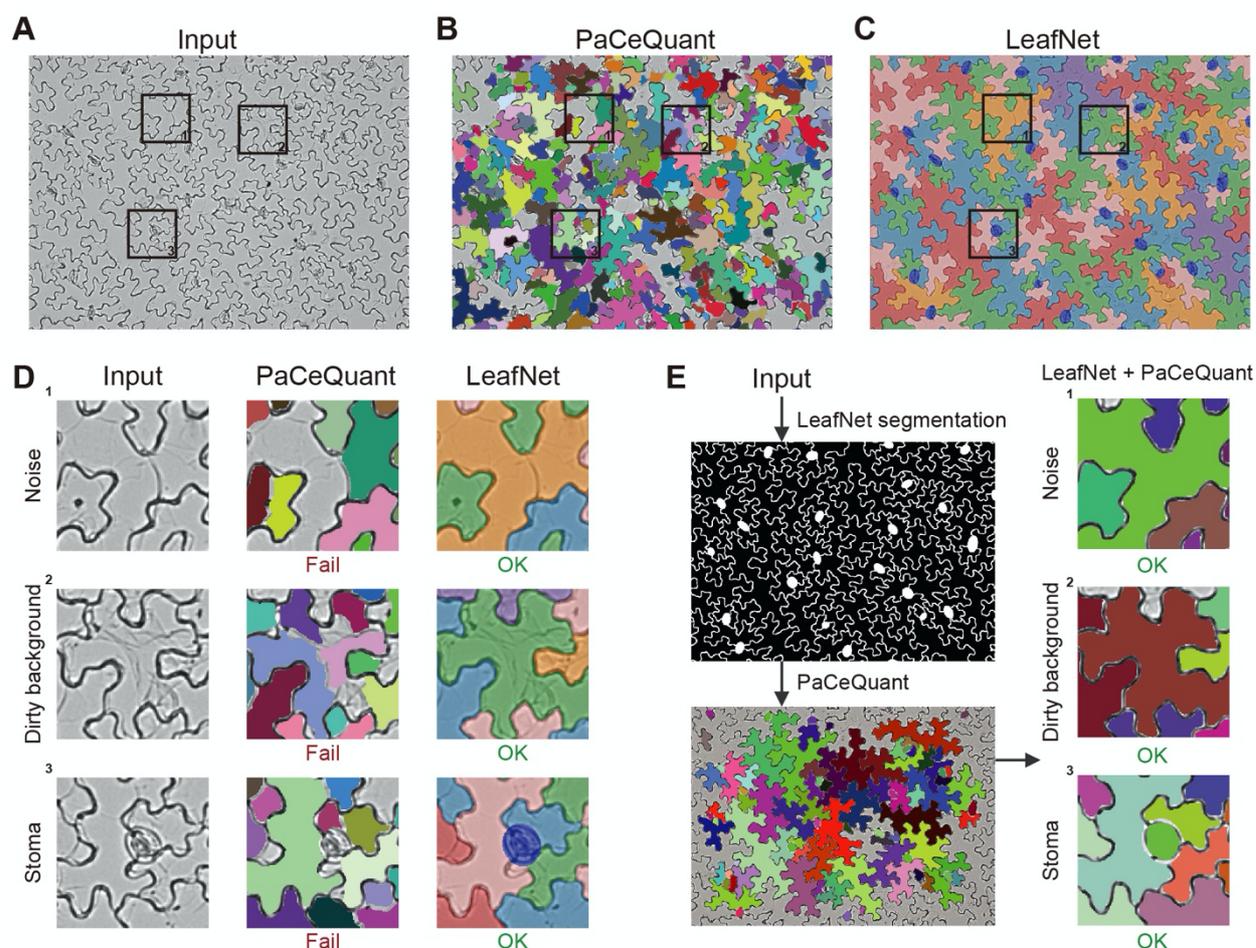


Figure 3. Comparison of LeafNet with PaCeQuant.

(A) Representative raw input image with three areas highlighted in boxes. (B) Results of segmentation of the image in A using the PaCeQuant program with the default configuration. Individual cells are filled with different colors. (C) Results of segmentation of the image in A using LeafNet with the default configuration. Stomata are colored in blue, and neighboring pavement cells are filled with different colors. (D) Zoom-in views of the three representative areas in the image in A showing typical noise and difficulties encountered in light microscope images. The raw input (left), segmented cells from PaCeQuant (middle), and those from LeafNet (right) are filled with different colors as in B and C. (E) Segmentation results of PaCeQuant using LeafNet's segmentation as input (left). The combination of LeafNet and PaCeQuant achieved good results for the three representative areas, highlighting the advantages of LeafNet's tolerance to various types of noise.

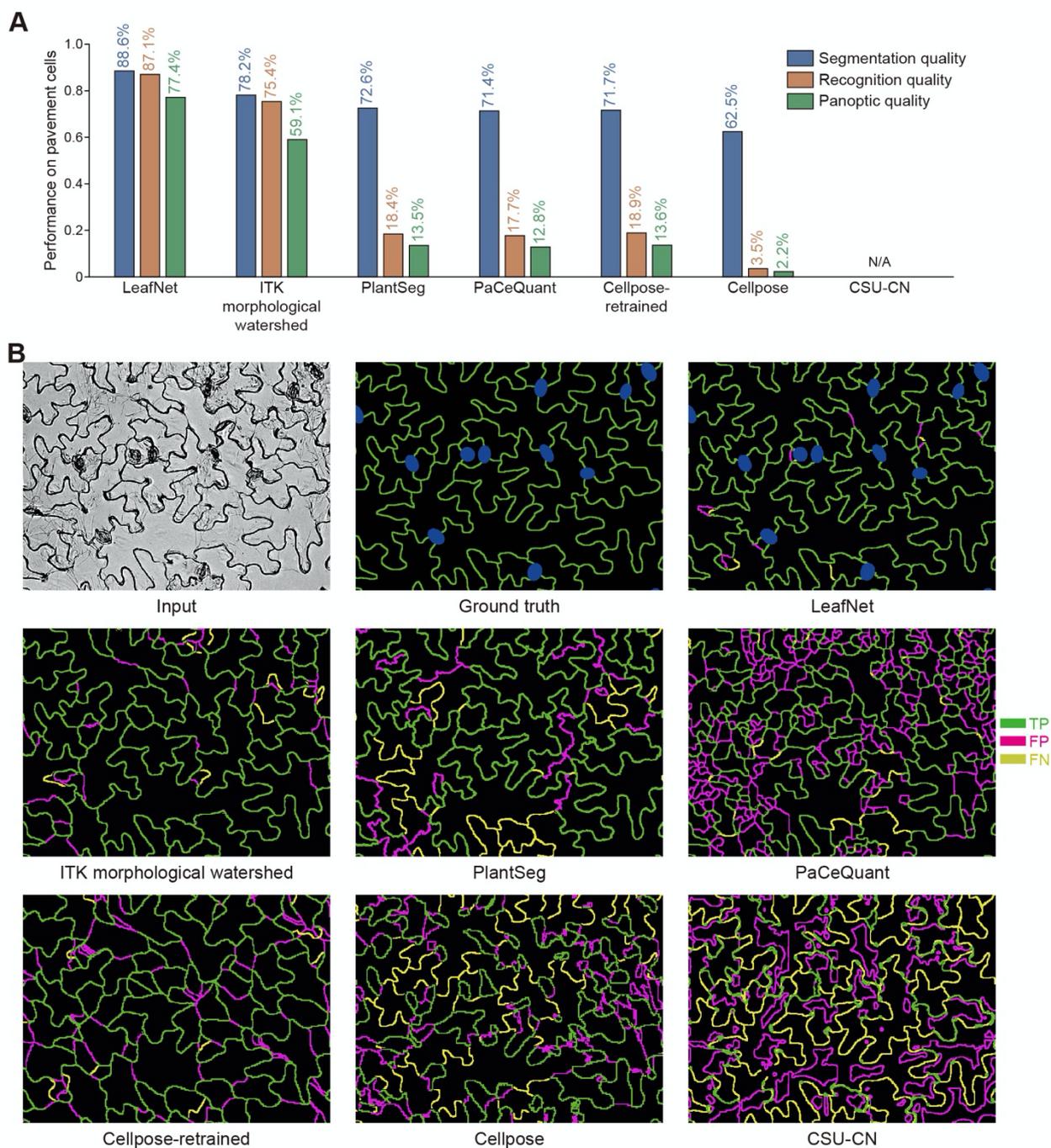


Figure 4. Quantitative evaluation of the performance of LeafNet and related tools for pavement cell segmentation.

(A) Performance of different tools for pavement cell segmentation in bright-field images of *Arabidopsis* leaves using three different metrics. (B) Representative examples of segmentation results using different tools. The true positive (TP), false positive (FP), and false negative (FN) borders are shown in green, red, and yellow, respectively.

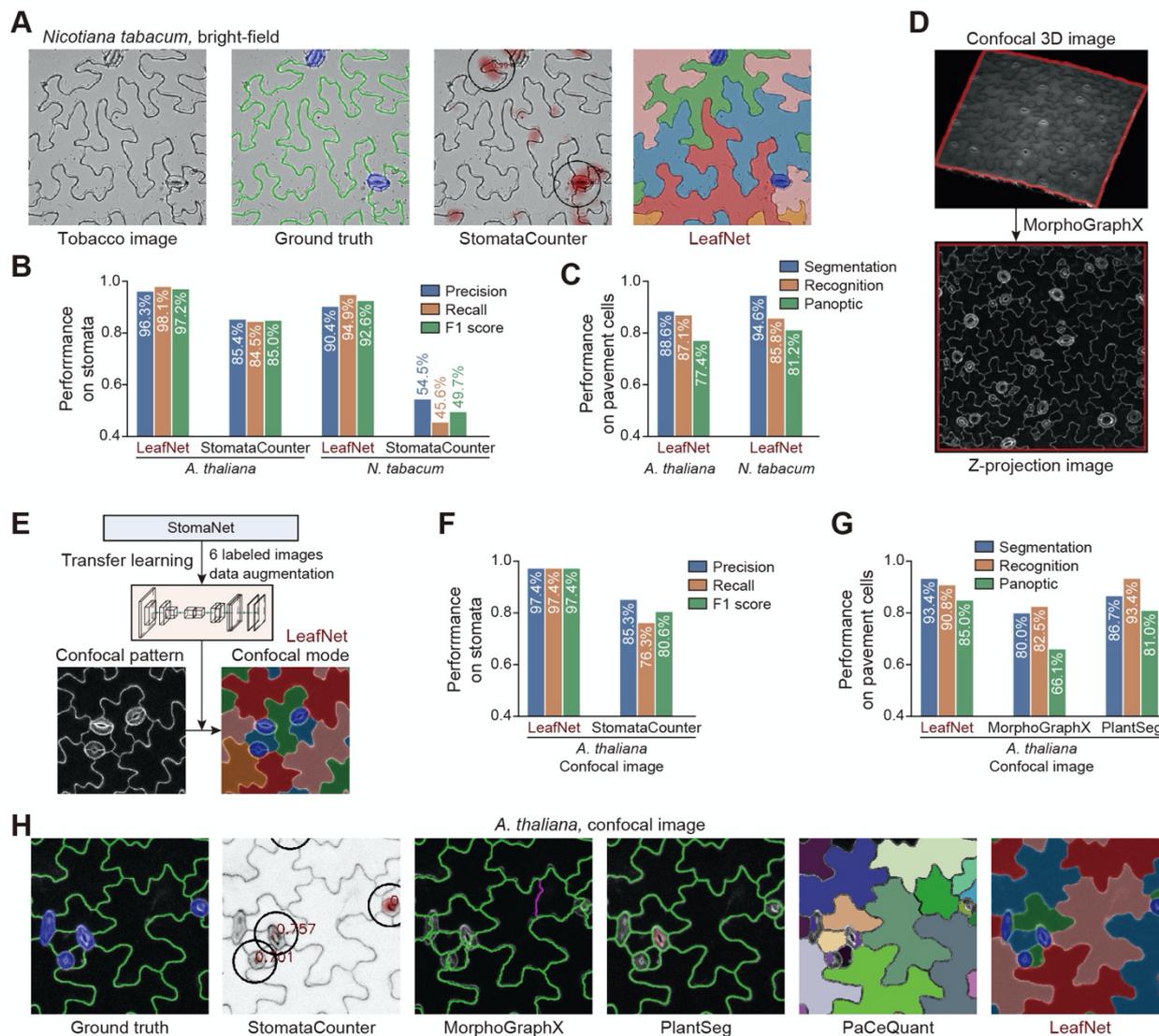


Figure 5. Extension of LeafNet to analyze species with similar morphology as well as confocal images.

(A) Representative example of LeafNet results for a bright-field image of *Nicotiana tabacum*. Raw image, ground truth labeling, StomataCounter results, and LeafNet results are shown from left to right. (B) Performance of LeafNet for stoma detection in *A. thaliana* and *N. tabacum* bright-field images compared with StomataCounter. (C) Performance of LeafNet for segmenting pavement cells in *A. thaliana* and *N. tabacum* bright-field images using three different metrics. (D) Representative example of pre-processing a confocal 3D image to a z-projection image using MorphoGraphX. (E) Extension of LeafNet to analyze z-projection image using transfer learning based on limited number of newly labeled data to identify stomata with its confocal mode. (F) Performance of LeafNet for detecting stomata in *A. thaliana* confocal images compared with StomataCounter. (G) Performance of LeafNet for pavement cell segmentation in *A. thaliana* confocal images compared with MorphoGraphX and PlantSeg for three different metrics. (H) Representative outputs of five programs using confocal images. StomataCounter, PlantSeg, PaCeQuant, and LeafNet use max intensity z-projection images as input, while MorphoGraphX takes 3D image stacks as input.

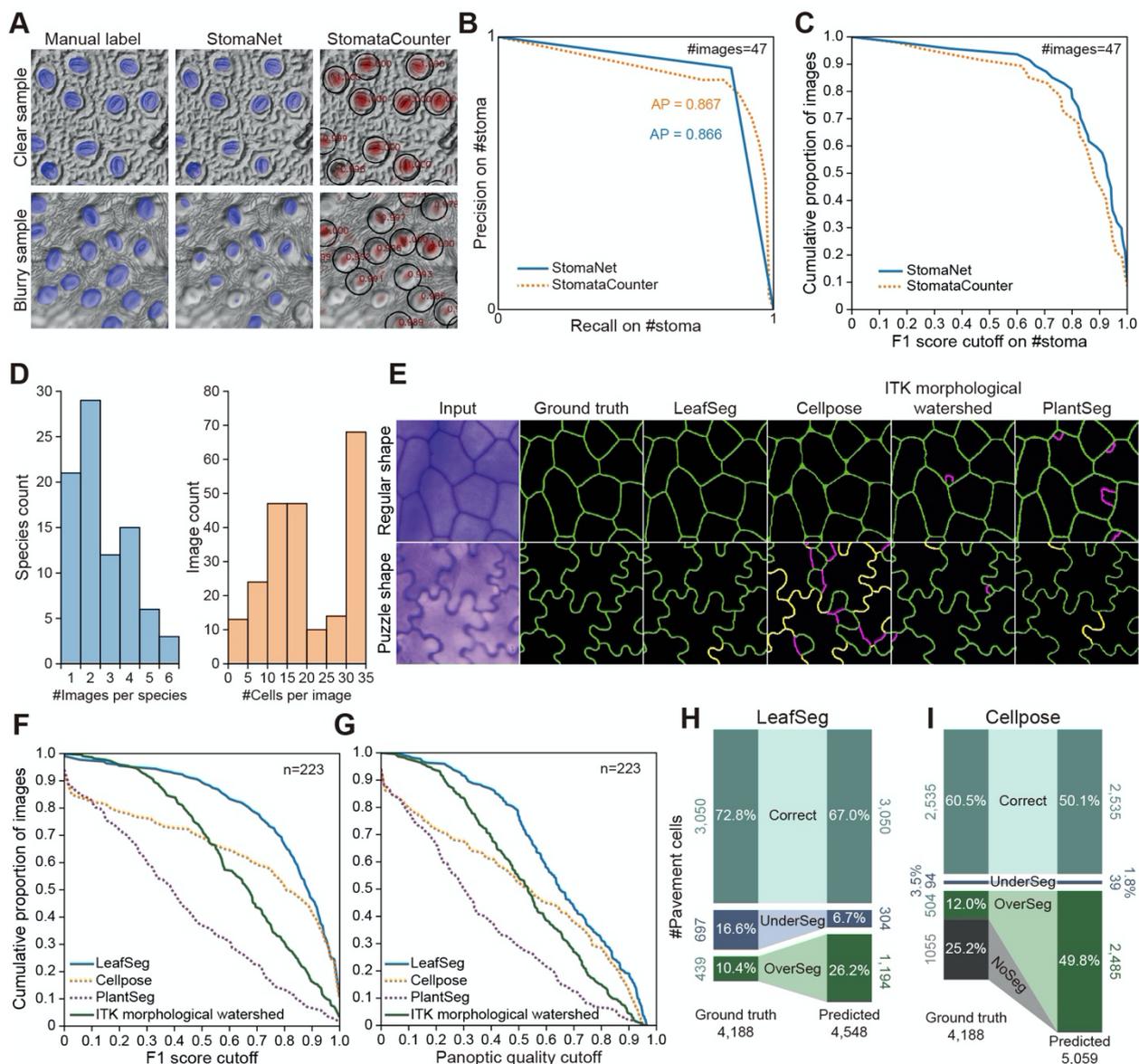


Figure 6. Evaluation of complex datasets using LeafNet and other tools.

(A) Representative examples of the results of StomaNet and StomataCounter for evaluating StomataCounter's dataset. StomaNet and StomataCounter both achieved high quality with stomata with clear boundaries (top) and suffered performance loss with blurry images (bottom). (B) Precision-recall curve of StomaNet (blue) and StomataCounter (orange dotted line) for counting stomata in 30 testing images. The thresholds were evaluated from 0.1 to 0.9 to calculate average precision (AP). (C) Cumulative distribution of F1 scores for stoma detection in all images of the stoma detection testing dataset ($n=47$). (D) The distribution of species counts by the number of images per species (left) and the distribution of image counts by the labeled cell counts per image (right). This pavement cell data set contains 4,188 cells in 223 images from 86 species. (E) Representative examples of the segmentation results from different programs for regularly shaped (top) and puzzle-shaped cells (bottom). (F-G) Cumulative distribution of F1 scores (F) and panoptic quality scores (G) of pavement cell segmentation from different programs using the pavement cell segmentation testing dataset ($n=223$). (H-I) Performance of LeafSeg (H) and Cellpose (I) in segmenting pavement cells for a testing dataset from a wide range of species. The numbers and percentages of correct, under-segmented, and over-segmented cells are shown in the comparison of predictions to ground truth.

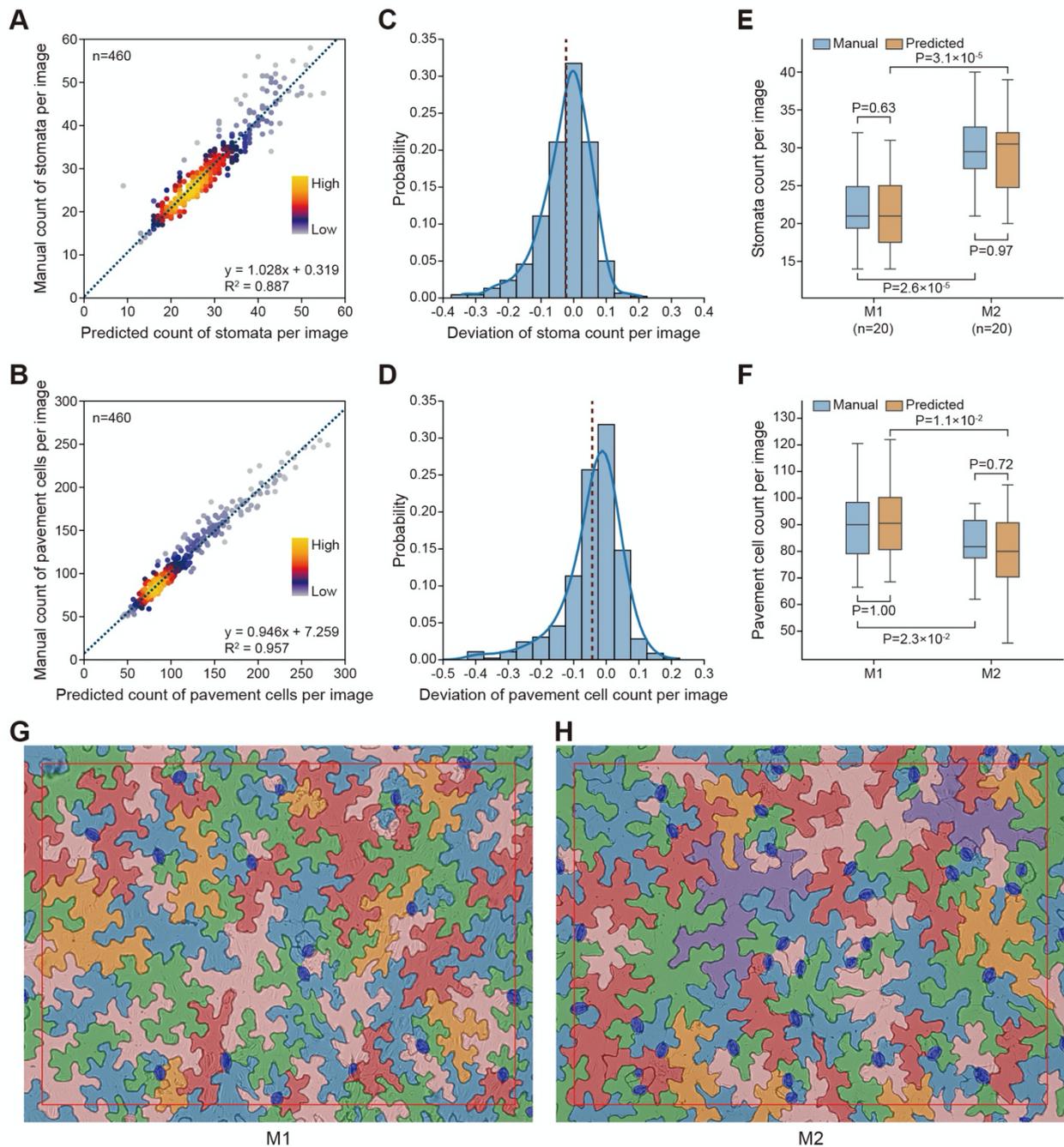


Figure 7. Application of LeafNet for evaluating large-scale microscope images.

(A-B) Linear regression between LeafNet's results and manual counting of stomata (A) and pavement cells (B). The point densities are represented by heatmap with KDE smoothing. (C-D) The deviation of stomata (C) and pavement cell (D) counts per image from LeafNet's results versus manual annotation. (E-F) LeafNet's performance in quantifying the counts of stomata (E) and pavement cells (F) in leaves from two different genotypes: M1 and M2. P-values are based on paired two-tailed *t*-test. (G-H) Representative examples of LeafNet predictions using images from genotype M1 (G) and M2 (H).