Automated High-Resolution Field Microscopy for Stomata Detection in Corn Using Deep Learning

Sainath Reddy Gummi[1], James Kemeshi[1], Young Chang[1, *]

[1]Department of Agricultural and Biosystems Engineering, South Dakota State University, Brookings, SD 57007, USA

*Corresponding Author

A paper from the Proceedings of the

16th International Conference on Precision Agriculture

21-24 July 2024

Manhattan, Kansas, United States

Abstract

Stomata, known as guard cells, are crucial for regulating plant water loss and are key targets for breeding drought-tolerant crops. Advanced digital microscopy and Deep Learning can enhance stomata analysis, significantly reducing the manual workload involved in breeding trials. This study aims to develop a precise, field-deployable stomata detection model to facilitate this process. Utilizing the ProScope HR Bodelin digital microscope, we acquired high-resolution images of corn stomata in outdoor settings. The images, captured under various outdoor lighting conditions, were enhanced using a blind deconvolution deblurring algorithm to improve clarity and detail. Following enhancement, the images were accurately annotated to create a robust training dataset for our model. We developed the FieldStomav1 model using Detectron2's Mask R-CNN framework with a ResNet-101 backbone, chosen for its strong feature extraction capabilities. The model underwent extensive training and hyperparameter tuning, optimizing its performance for field conditions. Evaluated using COCO evaluation metrics, the FieldStomav1 model achieved high precision and recall, with an Average Precision (AP) of 0.944 and an Average Recall (AR) of 0.713 at an Intersection over Union (IoU) threshold of 0.50. This high performance indicates the model's robustness in accurately detecting and segmenting stomata. FieldStomav1 demonstrated exceptional efficiency, analyzing 120 high-resolution images in just 1 minute and 11 seconds, a process that would traditionally take around 36 hours manually. Our results validate the model's efficacy in high-throughput stomata analysis, offering a reliable and automated solution for breeders. This model can enhance phenotypic decision-making in drought tolerance breeding programs, providing timely and accurate data to support the development of resilient crop varieties. The full implementation of the model, along with the code and data, can be accessed at our GitHub repository (https://github.com/gummi001/FieldStomav1-Detection).

Keywords

stomata, deep learning, field microscopy, corn, plant breeding, mask R-CNN, drought tolerance, detectron2, plant phenomics

1 Introduction

Under drought conditions, kidney-shaped organs on the epidermal surface of plants, called stomata, play a crucial role in regulating plant health. Resembling pores, stomata open and close during transpiration to regulate water loss (Xie et al., 2022). Recently, the use of advanced digital microscopes to observe the diurnal patterns of stomata has increased. Plant geneticists often study stomata dynamics to breed drought-tolerant varieties. Gaining insights into stomata behavior directly from the field can significantly enhance the breeding of varieties suited to drought conditions (Pathoumthong et al., 2023). Usually, in a breeding trial, there would be thousands of plants, and therefore, millions of stomata images to analyze to achieve precise breeding (Reynolds et al., 2020). Acquiring images from the breeding trials and then using Computer Vision (CV) and Deep Learning (DL) can make this task less time-consuming and advance breeding stages. Convolutional Neural Networks (CNNs) can intelligently analyze metrics in high-resolution images (Wu et al., 2019). Microscopists, biologists, and computer scientists are collaborating to develop advanced DL models. In laboratory settings, lighting can be manipulated to produce desired images. Plant functional genomics and plant signaling labs have already developed models for various crops using CNNs under such conditions. However, when screening stomata for breeding on a large scale, there is usually no time to adjust minor camera settings like brightness and contrast etc., or the timing of image capture (Reynolds et al., 2020; Pathoumthong et al., 2023).

The microscopic imaging of corn leaves in the field can be complicated by factors such as wind, lighting, and precipitation. Despite these challenges, the image resolution of handheld devices might not be sufficient to accurately measure the micrometers of the stomata. Using deblurring algorithms can help observe and extract detailed stomata parameters (Gibbs & Burgess, 2024). Field microscopy requires precision, and as the microscope zooms in on the object of interest, slight movements might cause the target to be missed. Hence, automated capturing techniques are essential for real-time imaging. Crops can be either monocots or dicots. Usually, monocot crops have larger stomata that can be easily observed through commercially available digital microscopes. However, within a single crop, stomata sizes can vary significantly depending on the breeding stage and cultivar (Xie et al., 2022). Therefore, breeders should consider developing their custom DL models for their specific cultivar if open-source models prove inadequate. Most existing stomata detection models have not made their data or user interfaces publicly available, highlighting a critical need for open-sourcing stomata data to advance research (Gibbs & Burgess, 2024).

Stomata detection typically utilizes CV techniques such as object detection, semantic segmentation, instance segmentation, and key point detection. While object detection might not yield precise measurements, segmentation techniques can be more effective. For example, a segmentation model tailored for outdoor corn breeders could significantly advance the breeding pipeline by selecting varieties with optimal diurnal patterns, which are more likely to survive adverse climatic conditions (Pathoumthong et al., 2023; Gibbs & Burgess, 2024). Plant scientists observe stomata traits such as stomata area, pore aperture area, and stomata density. Gaining insights early in the breeding program from these traits will ensure the maintenance of plant diversity. Developing a deep learning (DL) model that can detect all three parameters—stomata area, pore aperture area, and stomata density—in field conditions using rapid handheld digital microscopy is challenging. Hierarchical object detection is required because the pore aperture area is a minute detail within the larger stomata area. From an imaging perspective in the field, it is also challenging to develop a DL model capable of accurately measuring the pore area (Xie et al., 2022; Gibbs & Burgess, 2024).

In this study, we acquired images of corn stomata in outdoor settings, applied a deblurring algorithm, and developed a custom stomata model using weights from the Detectron2 model zoo. This model utilizes a Mask R-CNN with a Feature Pyramid Network (FPN) ResNet-101 backbone (Wu et al., 2019). The resulting FieldStomav1 can precisely segment the stomata area. With further post-processing, breeders can export the measurements in micrometers for making phenotypic decisions.

2 Materials and Methods

2.1 Microscopic Image Acquisition Annotation and Deblurring

2.1.1 Image acquisition of corn plants in outdoor lighting

Microscopic images of corn were non-invasively obtained using the ProScope HR Bodelin digital microscope. This imaging was conducted outdoors from May 16 to May 19, 2024, with the plants temporarily moved outside for this purpose (Fig. 1A). Before imaging, the plants were subjected to water stress for 30 days. They were cultivated in 15 cm pots at the Raven Precision Agriculture Center, South Dakota State University (SDSU), within a controlled environment where an artificial lighting system simulated 12 hours of natural daylight. A 400x magnification lens was used, capturing images with a field of view of 0.75 × 0.57 mm, a resolving power of 1 micron, and a pixel density of 652 pixels per millimeter, ensuring high-resolution imagery. The outdoor microscopic imaging was performed using specific settings in the ProScope HR software, including brightness set to 2, contrast to 1, saturation to 20, sharpness to 32, white balance to 5, WB red and blue both set to 140, exposure to -3, with auto exposure disabled, image orientation set to flip, and power frequency set to 60Hz (Table 1).

Table 1 Microscopic Camera Settings for Image Acquisition

Setting	Value
Brightness	2
Contrast	1
Saturation	20
Sharpness	32
White Balance	5
WB Red	140
WB Blue	140
Exposure	-3
Auto Exposure	Disabled
Image Orientation	Flip
Power Frequency	60Hz

2.1.2 Deblurring using blind deconvolution algorithm

In this study, 300 microscopic images of corn stomata were enhanced using a blind deconvolution algorithm implemented through a Python script utilizing the Richardson-Lucy deconvolution method (Guo et al., 2020; Gibbs & Burgess, 2024). The process involved setting up specified input and output directories, applying the deconvolution separately to each RGB channel of the images, and 25 iterations with an initial point spread function (PSF) represented by a normalized 5×5 matrix. After processing, images were recombined and saved, demonstrating significant improvements in image clarity and detail. The effectiveness of the deblurring was visually verified by displaying zoomed-in comparisons of original and processed images, showcasing the enhanced visibility of stomata features (Fig. 1 C & D).

2.1.3 Stomata annotation and image restoration

Following the deblurring of 250 microscopic images of corn stomata, each image was manually annotated on the RoboFlow platform, and the annotation process was completed in two days (Dwyer et al., 2024). For the annotations, the stomata areas were marked using the color #C7FC00 (Cyan Green), utilizing polygon tools to precisely delineate the complex shapes (Fig. 1B). After annotation, the following pre-processing was applied to each image: auto-orientation of pixel data with Exchangeable image file -orientation. No image augmentation and image resizing techniques were applied. This process ensured accurate representation and measurement of stomata traits in the high-resolution images, facilitating subsequent feature extraction.

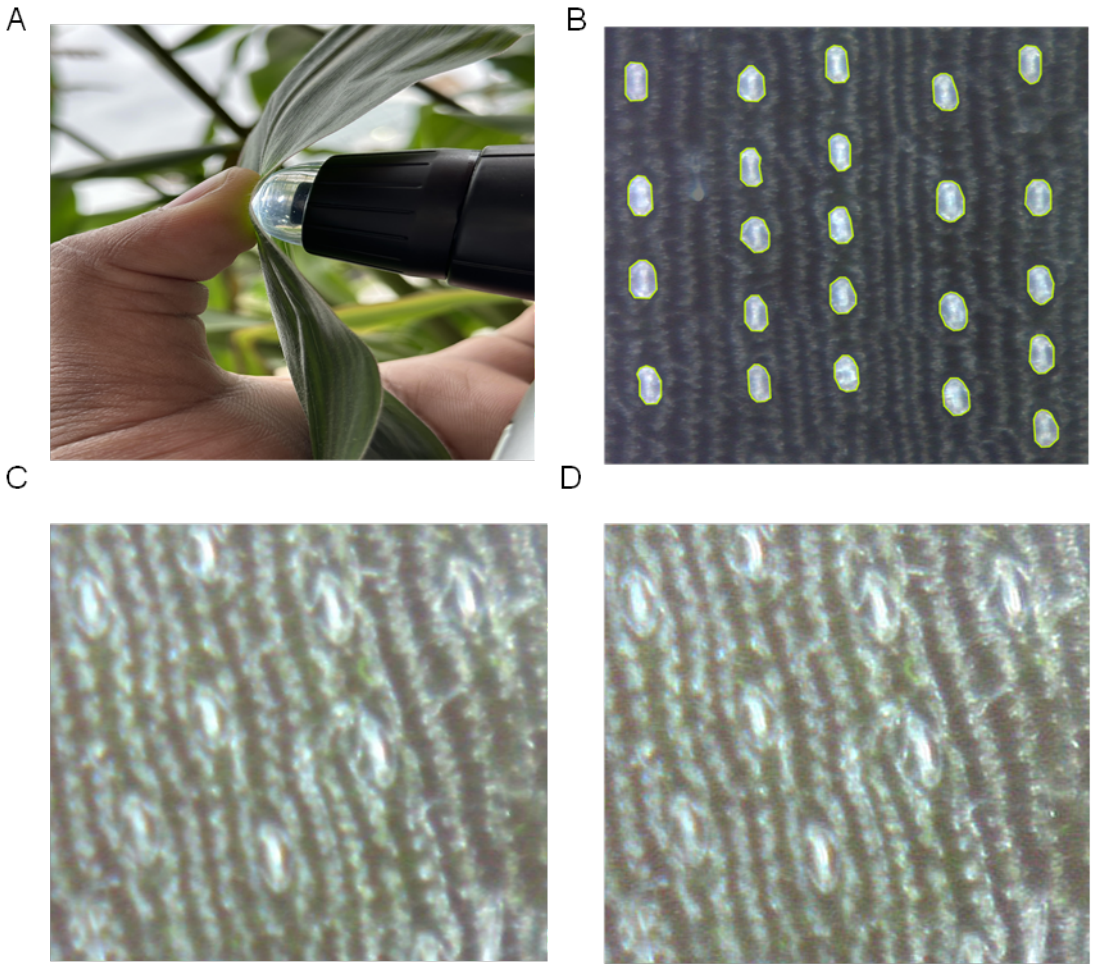

Fig. 1 Overview of outdoor stomata imaging acquisition process (A), annotation in Cyan Green for stomata area extraction (B) and the effectiveness of deblurring algorithms in enhancing image resolution and clarity [C (Original Image) and D (Deblurred Image)]

2.2 Segmentation model training using Detecteron2

2.2.1 Dataset split

The dataset was divided into three subsets: the Train Set, with 164 images and 5772 annotated areas, was used to train the model; the Valid Set, containing 48 images and 1583 areas, was for model validation during training to ensure effective learning; and the Test Set, with 23 images and 794 areas, evaluated the model's performance on new, unseen data to assess its predictive accuracy.

Table 2 Distribution of images and annotated stomata areas across different dataset splits used for training, validation, and testing

Dataset Split	Number of Images	Number of Annotated Stomata Areas
Train Set	164	5772
Valid Set	48	1583
Test Set	23	794

2.2.2 Hyper parameter tuning for the FieldStomav1

For FieldStomav1, hyperparameter tuning was carefully configured with the key parameters as outlined in Table 3. The model utilizes a pre-trained Mask R-CNN architecture with a ResNet-101 backbone and FPN from the Common Objects in Context (COCO) dataset's instance segmentation configurations, ensuring robust feature extraction across scales (Wu et al., 2019).

The Adam optimizer was selected for its efficiency in managing sparse gradients on large datasets. Utilizing the Optuna hyperparameter optimization framework, the base learning rate was set at 5.9217691338383614e-05, and the images per batch at 9 (Akiba et al., 2019). Key adjustments included setting a low base learning rate to fine-tune the pre-trained weights, minimizing perturbations in the learned representations. The model underwent training for 30,000 iterations, with learning rate adjustments planned at the 15,000th and 20,000th iterations to refine learning as the model neared optimal performance. Furthermore, the introduction of horizontal flips as a form of data augmentation and the enabling of cropping by relative range were incorporated. This training was performed in the SDSU "Innovator" HPC cluster, which was equipped with Intel Xeon Gold 6342 CPUs and NVIDIA A100 GPUs. The HPC cluster provided computational power necessary to efficiently process large high-resolution images and perform lengthy training tasks which requires ample amount of GPU memory.

Table 3 Hyperparameter Configuration for FieldStomav1

Configuration Item	Value
Pre-trained Model Weights	COCO-InstanceSegmentation/mask_rcnn_R_101_FPN_3x.yaml
Optimizer	Adam
Images Per Batch	9
Base Learning Rate	5.9217691338383614e-05
Max Iterations	30000
Number of Classes	1 (Single class for stomata_area)
Data loader Workers	8
Crop Enabled	True
Crop Type	relative_range
Random Flip	horizontal
Solver Steps	(15000, 20000)

2.3 Testing model weights on High resolution images

The segmentation process within the FieldStomav1 framework employed the Detectron2 'DefaultPredictor' to generate precise segmentation masks. This tool utilized the FieldStomav1 model weights to accurately delineate stomata, facilitating exact calculations of stomata areas (Wu et al., 2019). Each pixel identified as part of a stomata within these masks was counted, and the total number of true pixels was converted into square micrometers, leveraging the known pixel density of 652 pixels per millimeter. This analysis was conducted on a dataset comprising 120 images captured in outdoor situations between May 23 and May 26, 2024.

3 Results and Discussion

3. Comprehensive Performance Metrics on Validation Dataset

3.1.1 Average Precision and Average Recall from COCO evaluation metrics

The FieldStomav1 model was evaluated on the validation dataset using COCO evaluation metrics, focusing on both bounding box and segmentation performance across varying Intersection over Union (IoU) thresholds (Fig. 2).

For bounding boxes, the model achieved an overall AP of 0.600 across IoU thresholds ranging from 0.50 to 0.95, demonstrating moderate precision. Notably, the highest AP was 0.931 at an IoU threshold of 0.50, indicating excellent detection capability when a more lenient overlap criterion is applied. At a more stringent IoU threshold of 0.75, the AP was 0.687, reflecting the model's ability to maintain reasonable precision even with tighter localization requirements. In terms of recall, the model's AR was 0.700 at the highest IoU threshold of 0.75, highlighting its effectiveness in identifying a high proportion of relevant instances when more overlap is required. Lower recall values at thresholds of 0.50 and 0.50:0.95 (0.228 and 0.022, respectively) suggest that while the model detects fewer instances at lower thresholds, it excels in higher precision settings.

For segmentations, the model exhibited an overall AP of 0.619 across IoU thresholds from 0.50 to 0.95. The peak AP of 0.944 at an IoU threshold of 0.50 underscores the model's robust segmentation capabilities under lenient conditions. As the IoU threshold increased to 0.75, the AP decreased to 0.699, indicating a slight reduction in precision under stricter localization criteria. The AR for segmentations mirrored this trend, achieving a high of 0.713 at an IoU threshold of 0.75, demonstrating the model's strong recall capabilities in identifying relevant segmentation instances accurately under stringent overlap conditions.

Overall, the FieldStomav1 model demonstrates a balanced performance with high precision and recall at lower IoU thresholds, while maintaining desirable performance at higher thresholds. These results reflect the model's effectiveness in both detecting and segmenting stomata with high accuracy when overlap criteria is applied. The high AP and AR values at lower IoU thresholds suggest the model is well-suited for applications requiring less localization, whereas its sustained performance at higher thresholds highlights its flexibility and strength in harder settings.

The F_1 score achieved by the FieldStomav1 model was high at 0.977, reflecting the model's excellent precision and recall. This score, detailed in the appendix (Fig. A1, A2, A3), underscores the model's efficiency in identifying true positives while minimizing false positives and negatives, thus confirming its capability to deliver reliable and accurate stomata detection. The high F_1 score near the decision threshold of 0.5, as discussed in the appendix, illustrates the optimal balance the model achieves between precision and recall, making it particularly effective for field applications.

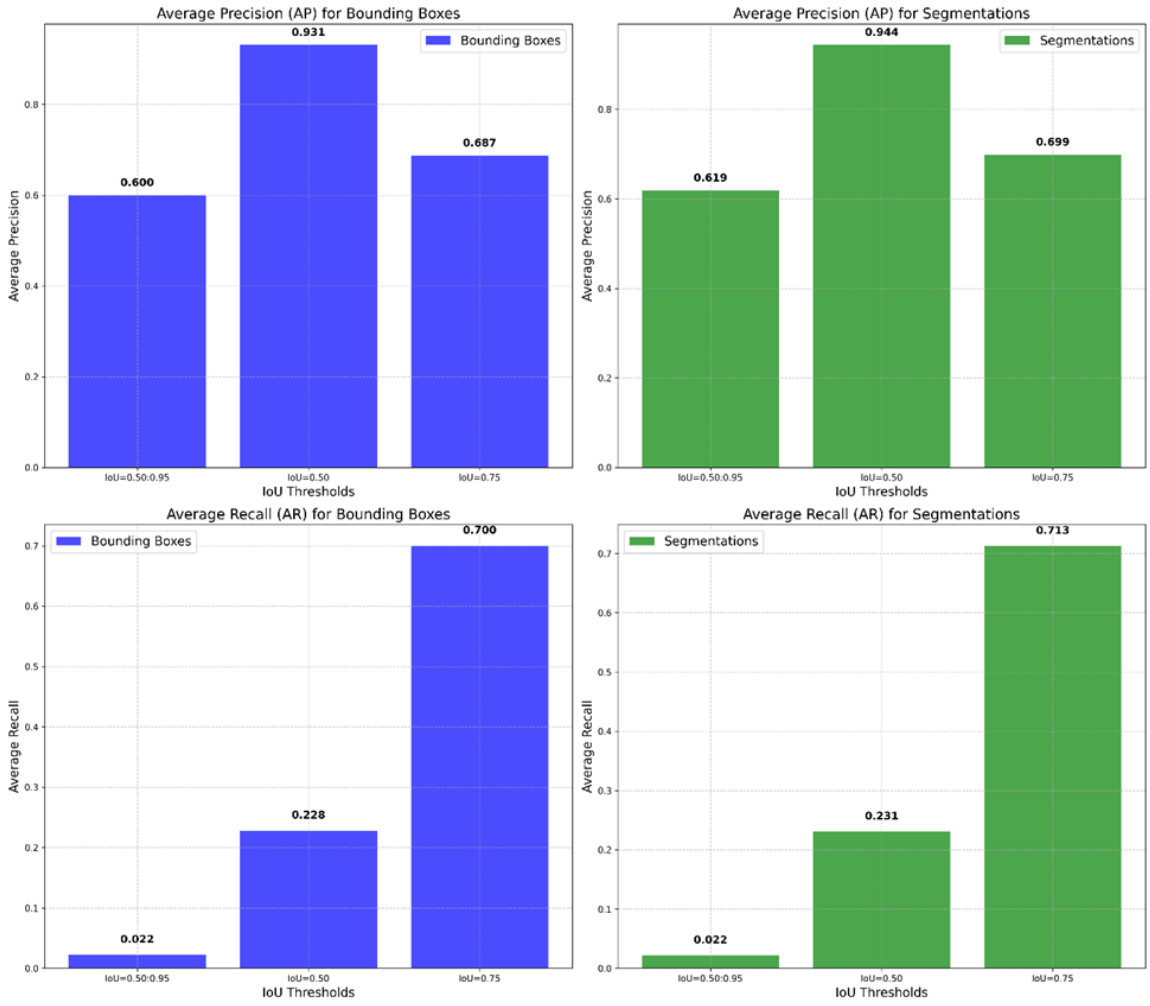

Fig. 2 The FieldStomav1 model performance by Average Precision (AP) and Average Recall (AR) metrics

3.1.2 Iterative Loss Reduction and Model Convergence

The training process of FieldStomav1 as shown in Fig. 3, indicates a total loss over iterations. Initially, the model undergoes a rapid learning phase characterized by a steep decline in total loss from above 5 to below 1 within the first few thousand iterations. This phase demonstrates the model's ability to quickly grasp fundamental patterns in the data. As training progresses, the loss reduction rate decelerates, indicating the model's transition into a fine-tuning phase where it makes more nuanced adjustments to its weights. Beyond approximately 15,000 iterations, the total loss stabilizes around a low value, signifying the model's convergence to optimal weights. This stabilization suggests that the model has effectively minimized the error between its predictions and the ground truth, reaching an equilibrium where further significant improvements are unlikely. The overall trendline underscores the strong correlation between iterations and model accuracy, affirming the model's learning efficacy. The training process took 8 hours, 15 minutes, and 28 seconds, utilizing a maximum GPU memory of 13,000 to 15,000 MB.

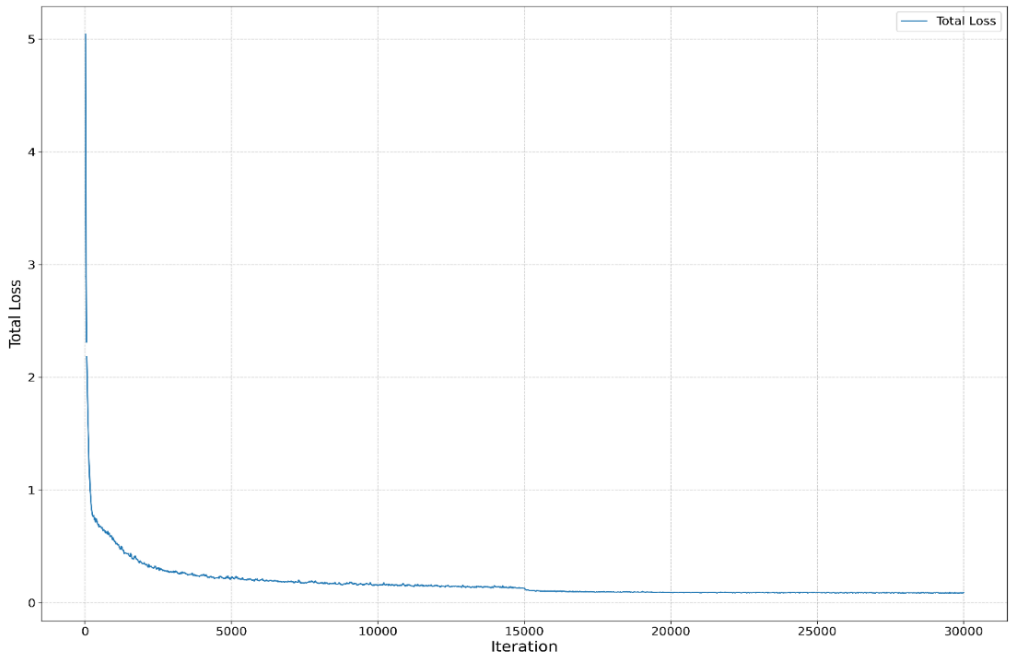

Fig. 3 Total Loss over Iterations for FieldStomav1

3.2 Performance Evaluation of FieldStomav1 Model

The FieldStomav1 model was tested on a dataset of 120 high-resolution images, focusing on the detection and segmentation of stomata. The evaluation process, conducted in an HPC Cluster, took only 1 minute and 11 seconds, which is significantly faster compared to the estimated 36 hours required for manual annotation. The results are presented in a combined figure (Fig. 4 and Fig. 5) that highlights the model detection capabilities and the variability in stomata areas.

In Fig. 4, the number of stomata detected per image is shown as blue bars on the left y-axis, while the distribution of stomata areas in square micrometers (um^2) is illustrated with an orange violin plot on the right y-axis. The consistent detection of stomata across images, with counts ranging from 20 to 50, emphasizes the model's reliability. The violin plot provides a detailed view of the stomata area distribution, revealing variations across different images.

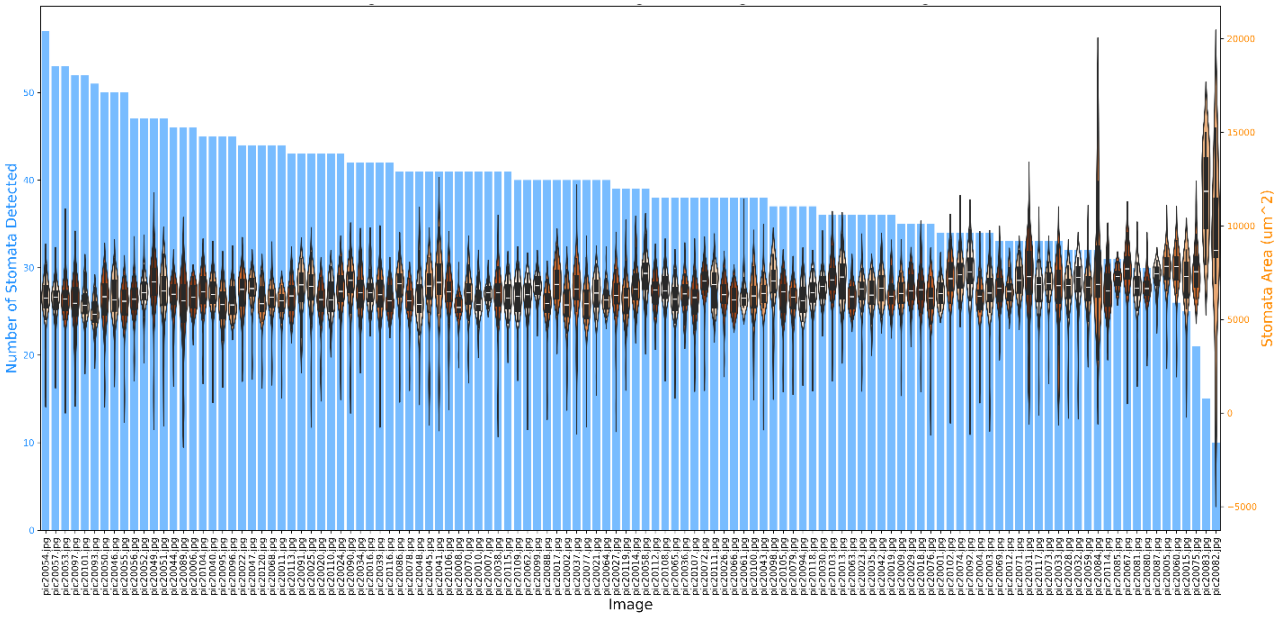

Fig. 4 Testing FieldStomav1 Model Weights on High Resolution Images

Further analysis was conducted on a specific high-resolution image, pic20038.jpg, to provide detailed insights into the model's segmentation performance. As depicted in Fig. 5A shows the original image alongside the detected image with segmented stomata. The detected image includes annotations indicating the confidence score for each stomata detection, demonstrating the model's precision in identifying stomata.

Fig. 5B represents a bar plot of the detected stomata areas for pic20038.jpg. The stomata areas are color-coded to reflect their relative sizes, with the x-axis representing the stomata index and the y-axis showing the stomata area in square micrometers (um²). The bar plot indicates a range of stomata areas, showcasing the model's ability to accurately measure stomata sizes. These results validate the FieldStomav1 model's efficiency and accuracy in high-throughput stomata analysis, offering a reliable and automated alternative to manual annotation.

Fig. 6 illustrates a linear regression analysis comparing ground truth annotations with predictions from the FieldStomav1 model for pic20038.jpg, a sample image was manually annotated to provide a high-resolution benchmark for stomata analysis. Despite the inherent biological and environmental variability, the model achieved an R^2 value of 0.50, signifying a moderate to strong correlation between the predicted and actual stomata areas. This suggests that the model's predictive accuracy is robust enough for practical applications in field settings. The plot shows the distribution of predicted versus actual stomata areas, with deviations around the regression line highlighting areas for potential refinement. Regions where predictions diverge significantly from the ground truth are prime candidates for targeted model training adjustments or for the integration of additional training data. These enhancements could help to minimize prediction errors and improve the model’s accuracy. This visual and statistical analysis serves as vital feedback for continuous improvement of the FieldStomav1 model, guiding iterative developments that enhance its precision and reliability in stomata measurement.

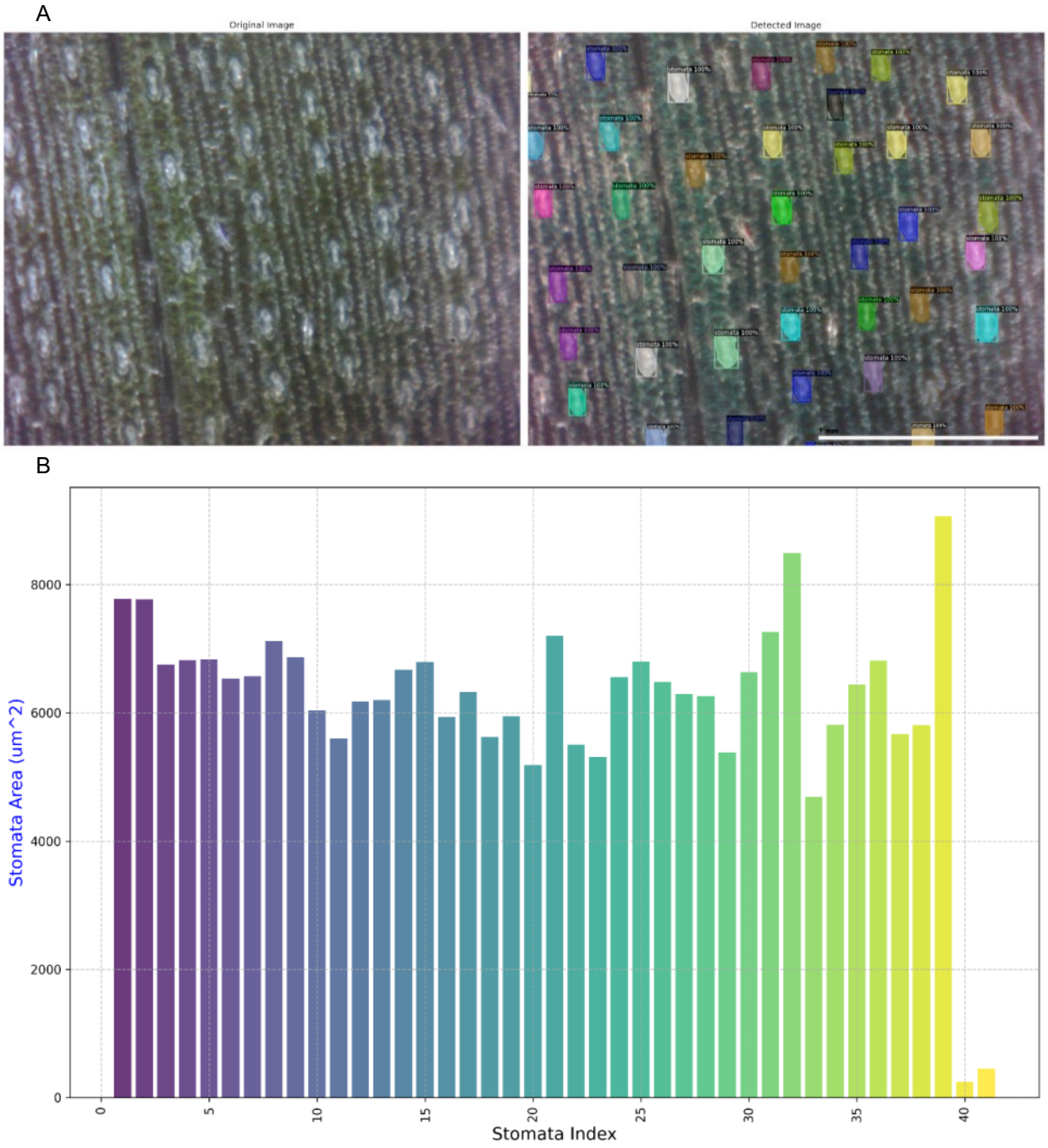

Fig. 5A) The original and detected image versions of stomata in pic20038.jpg which illustrates the precision of the FieldStomav1 model. Fig. 5B) The bar chart represents the area measurements of detected stomata, indicating the variability and distribution of stomata sizes across the sample image pic20038.jpg

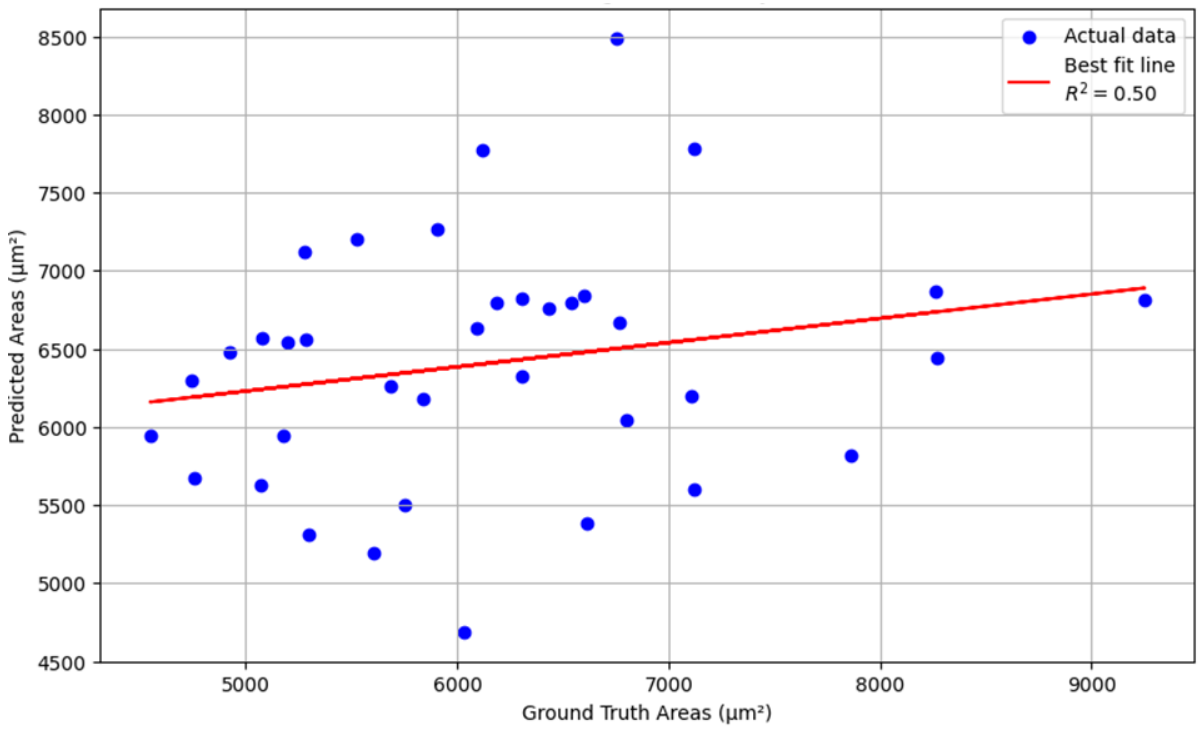

Fig. 6 Linear Regression Analysis of Stomata Areas on pic20038.jpg illustrates the comparison of manually annotated ground truth areas with areas predicted by the FieldStomav1 model. R^2 value of 0.50 indicates that the model explains half of the variance in stomata area measurements, reflecting its predictive accuracy amidst biological and environmental variabilities.

4 Conclusion

The FieldStomav1 model presents a significant advancement in high-resolution field microscopy for stomata detection, offering a rapid and accurate alternative to manual annotation. This study underscores the potential of integrating DL and CV in plant breeding research, particularly in breeding programs aimed at enhancing drought tolerance. The model's high precision and recall metrics, coupled with its efficient processing time, highlight its practical application for large-scale phenotypic analysis in diverse environmental conditions. The full implementation of the model, along with the code and data, can be accessed at our GitHub repository https://github.com/gummi001/FieldStomav1-Detection.

Future work should focus on refining the model for various crops and expanding its capabilities to include other critical stomata traits like pore aperture area. Future research could look towards using Generative Adversarial Networks (GANs) and autoencoder architectures to further enhance the model's accuracy and robustness. GANs could be particularly effective in generating synthetic stomata images for training, thereby enriching the dataset without the need for extensive field image capture. This approach would enable the model to learn from a more comprehensive range of conditions. Additionally, Autoencoders can be utilized for noise reduction and feature extraction in stomata imaging, helping to improve the clarity and detail of input images before segmentation. This would be crucial in minimizing errors related to external variables such as lighting and shadow effects, which often challenge the accuracy of phenotyping tools.

5 Acknowledgements

This research was funded by the Hatch Project (3AH777) and Multi Hatch Project (3AR730) of USDA NIFA through the South Dakota Agricultural Experimental Station at South Dakota State University.

6 References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623-2631. https://doi.org/10.1145/3292500.3330701

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. Information, 11(2), 125. https://doi.org/10.3390/info11020125

Dwyer, B., Nelson, J., Hansen, T., et al. (2024). Roboflow. Available from https://roboflow.com

Gibbs, J. A., & Burgess, A. J. (2024). Application of deep learning for the analysis of stomata: A review of current methods and future directions. Journal of Experimental Botany. Advance online publication. https://doi.org/10.1093/jxb/erae207

Guo, M., Li, Y., Su, Y., Lambert, T., Nogare, D. D., Moyle, M. W., ... & Shroff, H. (2020). Rapid image deconvolution and multiview fusion for optical microscopy. Nature Biotechnology, 38(11), 1337-1346. https://doi.org/10.1038/s41587-020-0505-4

Pathoumthong, P., Zhang, Z., Roy, S. J., et al. (2023). Rapid non-destructive method to phenotype stomatal traits. Plant Methods, 19, 36. https://doi.org/10.1186/s13007-023-01016-y

Reynolds, M., Chapman, S., Crespo-Herrera, L., et al. (2020). Breeder friendly phenotyping. Plant Science, 295, 110396. https://doi.org/10.1016/j.plantsci.2020.110396

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. Available from https://github.com/facebookresearch/detectron2

Xie, J., Wang, Z., & Li, Y. (2022). Stomatal opening ratio mediates trait coordinating network adaptation to environmental gradients. New Phytologist, 235(3), 907-922. https://doi.org/10.1111/nph.18189

Appendix

Supplementary FieldStomav1 Performance metrics

In Fig. A1, the PR curve for the FieldStomav1 model shows high precision value maintained across a wide range of recall values, which indicates that the model has a low false positive rate while also capturing most of the true positives. The curve remains close to the upper right corner of the plot, reflecting the model's effectiveness in detecting stomata with high accuracy. The slight drop in precision at higher recall values indicates the trade-off where achieving higher recall might slightly reduce precision. In Fig. A2, the plot shows how the F1 score varies with different threshold values. The FieldStomav1 model achieves high F1 scores across a wide range of thresholds, indicating that the model maintains a good balance between precision and recall at various decision boundaries. The peak F1 score near a threshold of 0.5 suggests that this threshold provides the best trade-off between precision and recall for the model. In Fig. A3, the confusion matrix for the FieldStomav1 model shows a high number of true positives and true negatives, with relatively few false positives and false negatives.

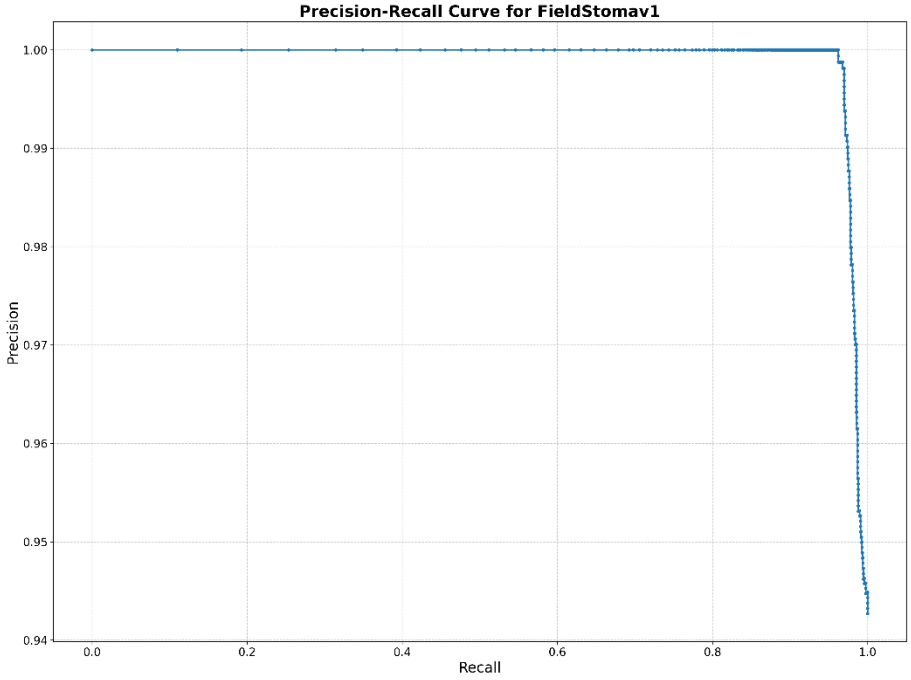

Fig. A1 Precision-Recall Curve for FieldStomav1

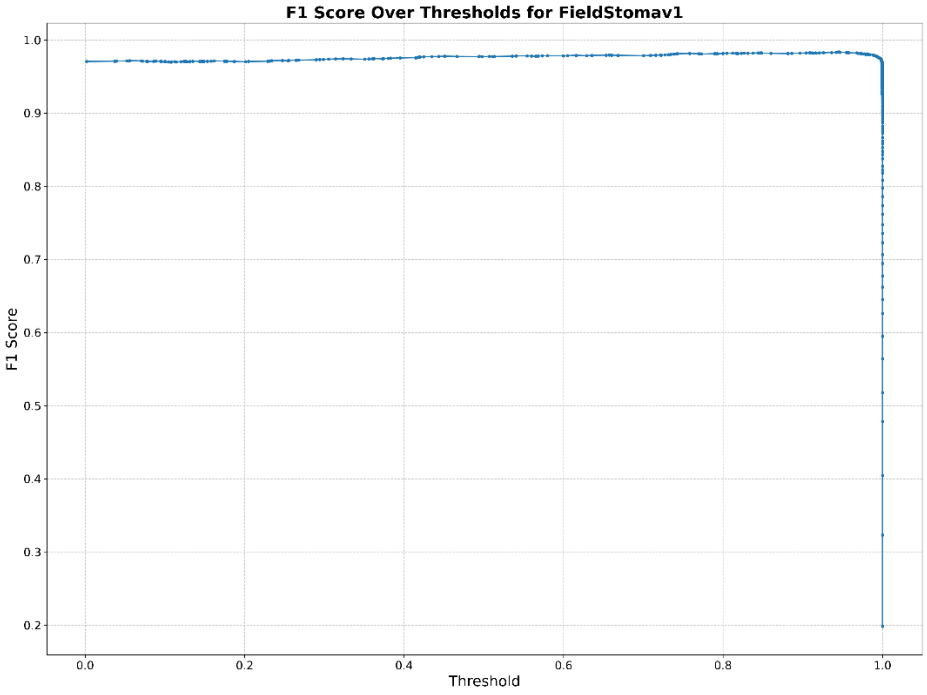

Fig. A2 F1 Score Over Thresholds for FieldStomav1

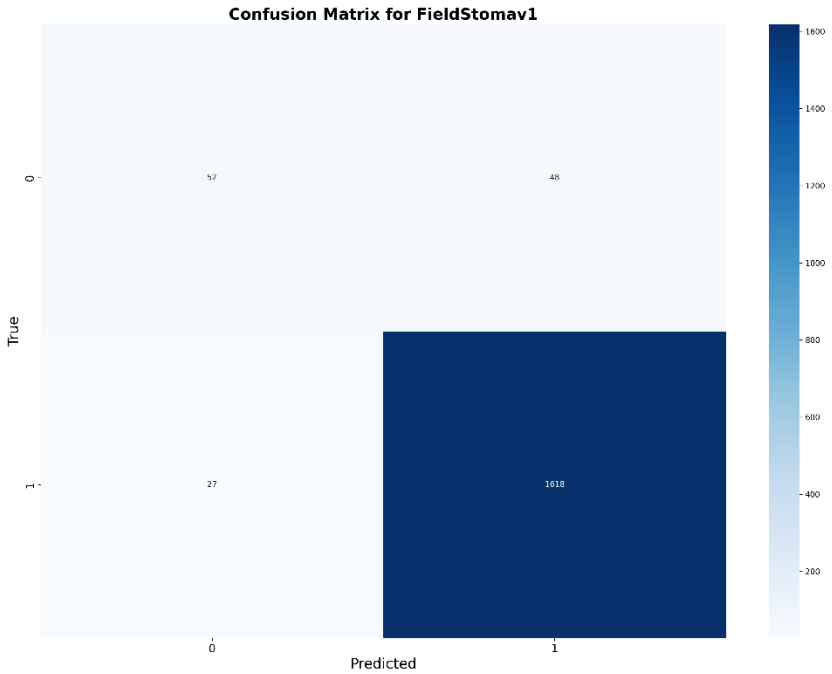

Fig. A3 Confusion Matrix for FieldStomav1