Research Article

Received: 10 December 2025 Revised: 11 July 2026 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/ps.71193

SporaScan: cost-effective, high-precision leaf-disc disease severity assessment for grapevine downy mildew

Qi Tian,[a,b†] Junjie Qu,[c†] Ling Yin,[c] Shumei Wei,[c] Changqing Yan,[d] Zeyun Liang,[d] Bin Chen,[a] Linjia Yao,[b] Hui Fang,[e] Mohamed Amine Benaly,[f] Jing Wang,[g] Qiang Yu[a,b*] and Gang Zhao[a,b*]

Abstract

BACKGROUND: Accurate assessment of disease severity is essential for evaluating fungicide performance and breeding disease-resistant crop varieties. Manual scoring of infection on individual leaf discs is labor-intensive and variable, while traditional computer vision methods require manual parameter tuning and lack robustness. Existing deep learning approaches often struggle to simultaneously localize leaf discs and accurately segment disease symptoms, limiting their practical application.

RESULTS: We developed SporaScan, an automated pipeline combining YOLO v8n for leaf disc localization, Mobile SAM for background removal, and UNet for sporulation segmentation. It achieved high accuracy (mAP@50 >99%, mIoU@50 >96%), with background removal reducing misclassification (0.21% for sporulation and 2.75% for leaf discs). Severity estimates showed strong agreement with manual annotations ($R^2 = 0.99$). In a blind test, technicians selected SporaScan as superior in 37.2% of cases, manual annotation in 26.2%, and equal performance in 36.6% ($P < 0.001$). Compared with Mask R-CNN, the sequential design improved accuracy and efficiency.

CONCLUSION: These results demonstrate that SporaScan provides an efficient and practical approach for automated assessment of downy mildew severity, supporting applications in disease evaluation, breeding, and fungicide assessment (http://116.10.197.212:9060/segment/#/).

Supporting information may be found in the online version of this article.

Keywords: grape downy mildew; severity assessment; image segmentation; resistance breeding; deep learning

* *Correspondence to: G Zhao or Q Yu, College of Soil and Water Conservation Science and Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: gang.zhao@nwafu.edu.cn (Zhao); yuq@nwafu.edu.cn (Yu)*

† *These authors contributed equally to this work.*

a *College of Soil and Water Conservation Science and Engineering, Northwest A&F University, Yangling, China*

b *State Key Laboratory of Soil and Water Conservation and Desertification Control, Northwest A&F University, Yangling, China*

c *Guangxi Crop Genetic Improvement and Biotechnology Key Lab, Guangxi Academy of Agricultural Sciences, Nanning, China*

d *College of Intelligent Equipment, Shandong University of Science and Technology, Taian, China*

e *Guangxi Key Laboratory of Biology for Crop Diseases and Insect Pests/Plant Protection Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China*

f *International Water Research Institute, Mohammed VI Polytechnic University (UM6P), Benguerir, Morocco*

g *College of Resources and Environmental Sciences, China Agricultural University, Beijing, China*

1 INTRODUCTION

Crop diseases pose a significant threat to global food security, leading to substantial yield losses and increased reliance on fungicides, which incur economic costs and environmental risks.[1,2] Developing disease-resistant cultivars is therefore essential, requiring accurate assessment of disease severity under both laboratory and field conditions.[3] Under *in vitro* conditions, leaf disc assays are widely used, where multiple discs are arranged in Petri dishes for severity evaluation.

Traditionally, disease severity is assessed manually from images, a process that is labor-intensive, time-consuming, and prone to subjective bias.[4] To improve efficiency, various image-based methods have been developed, including traditional computer vision (CV) approaches and convolutional neural networks (CNNs).[5] Traditional CV methods, such as Hough circle detection and thresholding, rely heavily on manual parameter tuning, limiting their robustness under varying imaging conditions.[6,7] CNN-based approaches improve generalization and accuracy but often process cropped patches or focus on whole-leaf segmentation, making it difficult to localize and quantify disease severity for individual leaf discs within a Petri dish.[8–10] These limitations highlight the need for methods capable of both precise localization and fine-grained segmentation.

Although Mask R-CNN enables simultaneous object detection and segmentation, its application in plant disease analysis has primarily focused on disease identification or leaf segmentation rather than severity quantification.[11,12] More recently, object detection models such as YOLO and foundation models such as SAM have shown strong performance in phenotyping tasks.[13,14] In particular, Mobile SAM provides efficient mask generation based on prompt inputs but remains limited in fine-grained disease detection.[15] Integrating complementary models has therefore emerged as a promising strategy to improve both accuracy and efficiency in agricultural image analysis.[16,17]

This study introduces SporaScan, an automated pipeline for precise severity assessment of grapevine downy mildew at the individual leaf disc level. A dataset of 432 smartphone images with varying disease severity was manually annotated. The pipeline integrates YOLO v8n for leaf disc localization, Mobile SAM for background removal, and encoder–decoder (UNet) models for sporulation segmentation. The performance of SporaScan is evaluated against Faster R-CNN and Mask R-CNN, and severity predictions are validated against manual annotations. In addition, model performance is specifically analyzed for samples with low disease severity (<5%) to assess its capability in early-stage detection. SporaScan is further deployed on an online platform (http://116.10.197.212:9060/segment/#/), providing an efficient and practical tool for grapevine downy mildew assessment with potential applications in breeding and phenotyping.

2 MATERIALS AND METHODS

2.1 Grapevine leaf disc assay and image acquisition

The dataset used in this study was generated from leaf disc inoculation assays conducted to evaluate the resistance of grapevine varieties from the breeding program at the Guangxi Academy of Agricultural Sciences. Technicians captured images of the leaf discs in Petri dishes using smartphones after infection from multiple viewing angles. Figure 1 illustrates the main process of the leaf disc assay and image acquisition. The detailed procedure is described in detail below.

Grapevine leaves, which were healthy or only slightly infected, were collected from the experimental vineyard at the Grape and Wine Research Institute of the Guangxi Academy of Agricultural Sciences, located in Nanning, Guangxi, China (22.47° N, 108.21° E; altitude 142.3 m). The grapevine varieties evaluated in this study included Chardonnay (susceptible), Ye Niang No. 6 (susceptible), Guipu No. 6 (moderately resistant), and Xia Hei (susceptible).

The collected grapevine leaves were first rinsed with distilled water to remove surface contaminants and then gently dried using absorbent paper under laboratory conditions. Leaf discs, about a diameter of 1.5 cm, were excised from fully expanded, healthy leaves using a sterile hole punch. Depending on the experimental design, between six and 25 leaf discs were arranged in Petri dishes with their abaxial side upward. *Plasmopara viticola* sporangia were originally collected from naturally infected grapevine leaves in the field, stored at −80 °C. Before inoculation, samples were thawed and used to prepare the inoculum. Sporangia were collected by gently rinsing infected tissue with sterile distilled water and filtering the suspension to remove debris. The concentration of the spore suspension was adjusted to 1×10^5 sporangia/mL using a hemocytometer. Detached leaf discs were inoculated using the droplet inoculation method by applying a single 35-μL droplet of the spore suspension onto each leaf disc.

The Petri dishes were incubated in a plant growth chamber at 22 °C under a 16-h light/8-h dark cycle post inoculation. The incubation period lasted for 5–7 days, allowing sporulation to appear on the leaf discs. The uninoculated leaf discs remained symptom-free.

The images of these leaf discs were captured using smartphones to document the infection status. A total of 432 images were collected, representing 6972 leaf discs across multiple experimental replicates. Due to variations in the camera-to-subject distance during image capture, the resolution of the leaf discs differed. The image resolution distribution of the leaf discs is shown in Fig. 2. The distribution of disease severity across samples is shown in Fig. S1.

2.2 Manual annotation for leaf disc and downy mildew sporulation

We used an open-source tool, AnyLabeling (available at https://anylabeling.nrl.ai/), to annotate the leaf discs and the corresponding downy mildew sporulation. This software, equipped with pretrained models such as Mobile SAM, facilitates semi-automated annotation. In this process, a rectangular bounding box was manually delineated around the leaf disc, and its coordinates were then automatically input into Mobile SAM to generate the corresponding contour of the leaf disc mask. Downy mildew sporulation, with irregular contours, were manually annotated by line strings composed of multiple connected points. All other regions were classified as background within an image. Annotated images were assigned numerical labels: 0 for background, 1 for leaf discs, and 2 for downy mildew sporulation. The masks of manual annotation for leaf discs and downy mildew sporulation were generated using a Python script (available at https://github.com/wkentaro/labelme). To ensure consistency in annotation quality and minimize personal bias, eight technicians participated in the annotation process, with images randomly exchanged among them. Leaf discs with unclear sporulation were excluded from annotation due to blurring caused by variations in focal length, making accurate identification challenging even for trained

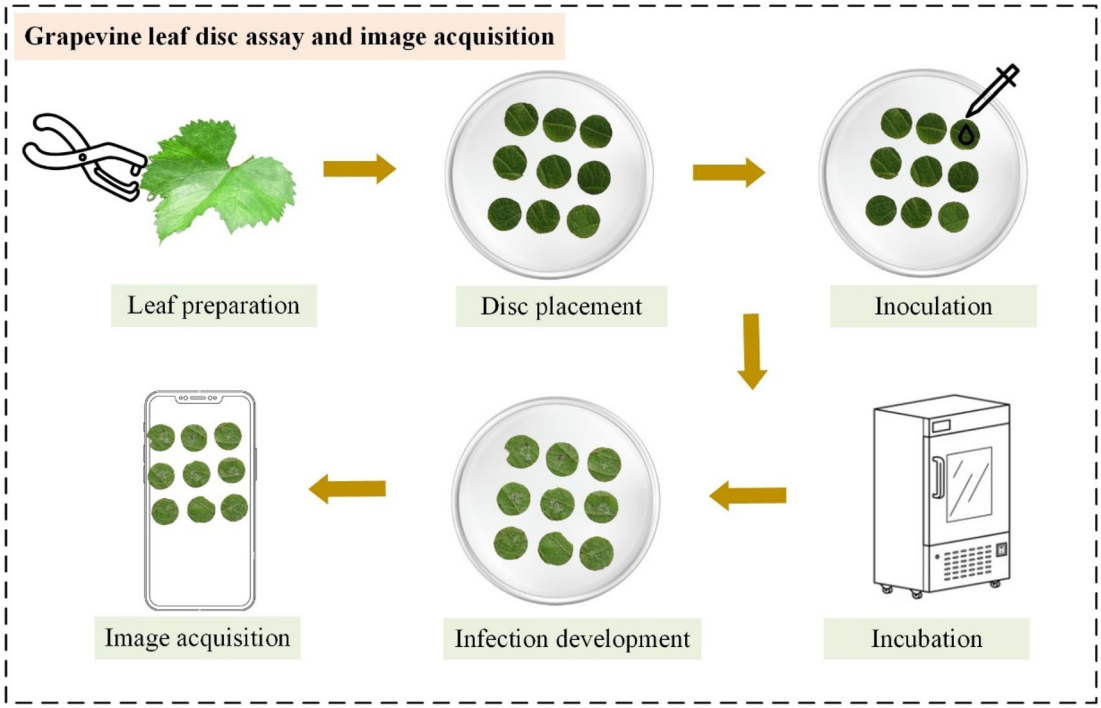

Figure 1. Schematic representation of the leaf disc assay and image acquisition, illustrating the sequential steps of leaf preparation, disc placement, inoculation with *Plasmopara viticola* sporangia, incubation under controlled conditions, infection development, and image acquisition for disease severity assessment.

technicians. Representative samples of annotated leaf discs and sporulation are shown in Fig. 3.

2.3 SporaScan: a pipeline for severity assessment

2.3.1 Overall workflow

We developed an automated pipeline, *SporaScan*, to facilitate the severity assessment of downy mildew on grapevine leaf discs. The complete workflow is illustrated in Fig. 4. The primary objectives of *SporaScan* development were as follows: (1) to identify the most suitable detection model for leaf disc localization from the YOLO v8 series and Faster R-CNN, (2) to evaluate the impact of background removal using Mobile SAM on segmentation performance, and (3) to select an appropriate encoder-decoder model for the segmentation of both leaf discs and downy mildew sporulation. The details of each stage are described in the following subsections.

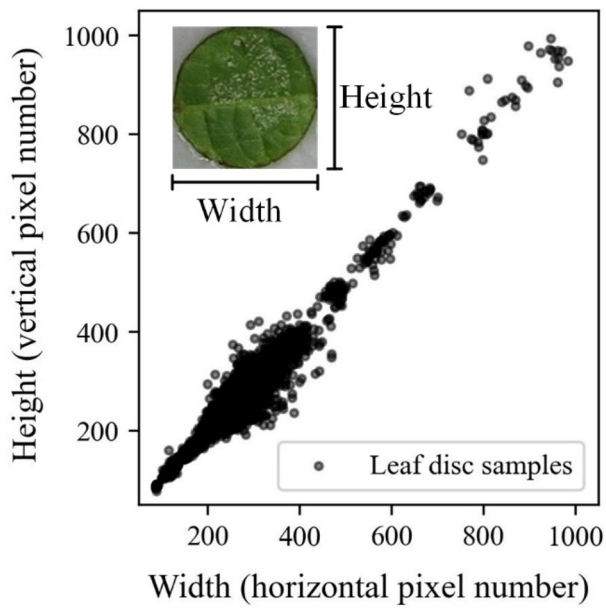

Figure 2. Different resolutions of grapevine leaf discs in the dataset. 'Width' represents the horizontal pixel number, and 'Height' represents the vertical pixel number.

2.3.2 Leaf disc localization

To accurately and efficiently localize each leaf disc with varying infection statuses, we conducted a comparative evaluation between the YOLO v8 series and Faster R-CNN. Both models are widely employed in object detection tasks in agriculture, and have demonstrated high precision.[18] The YOLO v8 series, as a one-stage model, is designed to effectively balance inference time and accuracy.[19] In contrast, Faster R-CNN, a two-stage model, typically provides higher precision under more complex conditions but comes with a larger parameter size and slower inference time.[20]

To train and evaluate these models, 432 images were randomly split into training (260), validation (86), and test (86) sets, following a 6:2:2 ratio. The YOLO v8 series adopted CSPNet as backbone network for feature extraction, while Faster R-CNN employed ResNet50. All other parameters were kept consistent across models. Specifically, the initial weights were pre-trained from the COCO dataset.[21] Stochastic Gradient Descent (SGD) was utilized as the optimizer, with an initial learning rate of 0.001 and a weight decay of 0.0005. The training process ran for 150 epochs, with early stopping implemented after 10 epochs of no improvement. A batch size of 16 was utilized. The models' weights that yielded the best performance on the validation set were retained for further evaluation.

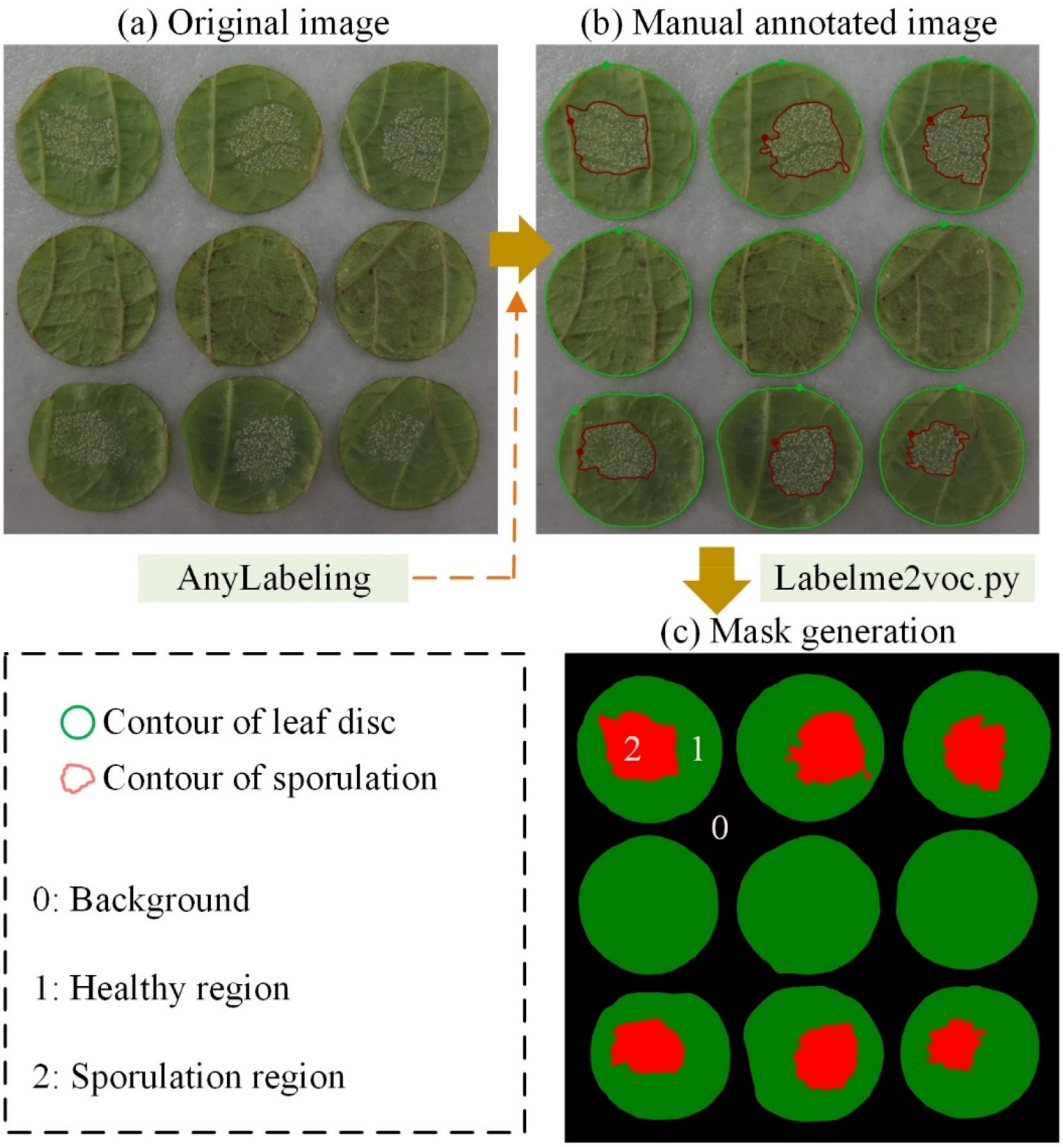

Figure 3. Overview of the annotation and mask generation process for leaf discs and downy mildew sporulation. (a) The original image depicts healthy or infected leaf discs. (b) Annotation for each leaf disc and sporulation using AnyLabeling (https://anylabeling.nrl.ai/). Green contours for leaf discs were generated by Mobile SAM, while red contours were manually annotated. (c) Generated segmentation masks with numerical labels assigned to different classes.

Mean Average Precision at a threshold of 0.5 (mAP@50), was used as the evaluation metric for leaf disc localization. The calculation of mAP@50 involves the following steps: (1) Precision-Recall curves are generated based on the Intersection over Union (IoU) at a threshold of 0.5 for each category, (2) Average Precision (AP) is computed as the area under the interpolated Precision-Recall curve for each category. (3) AP values for all categories are averaged to obtain mAP@50. As only a single category (leaf disc) is considered in this study, mAP@50 is equivalent to the AP of the leaf disc. The formula for mAP@50 is expressed as Eqn (1):

$$mAP@50 = \frac{1}{N}\sum_{i=1}^{N} AP_i \times 100\,\% \qquad (1)$$

where N represents the total number of categories, and AP_i denotes the Average Precision for the ith category. The mAP@50 ranges from 0% to 100%, with higher values indicating improved performance in leaf disc localization. Additionally, frames per second (FPS) was utilized to evaluate inference times and computational efficiency.

2.3.3 *Background removal*

Accurate background removal is crucial to enhance precision of downstream analysis.[16] To separate leaf disc from their backgrounds, Mobile SAM was employed directly without additional training. Mobile SAM is a lightweight visual foundation model optimized for mobile devices and derived from the Segmentation Anything Model.[22] It can encode prompt inputs, such as points, bounding boxes, and masks, with image embeddings, enabling it to perform target segmentation with strong zero-shot generalization capabilities across various tasks.

In this study, the bounding box coordinates, generated by YOLO v8n, were used directly as bounding boxes prompts for Mobile SAM. This enabled the accurate generation of leaf disc contours

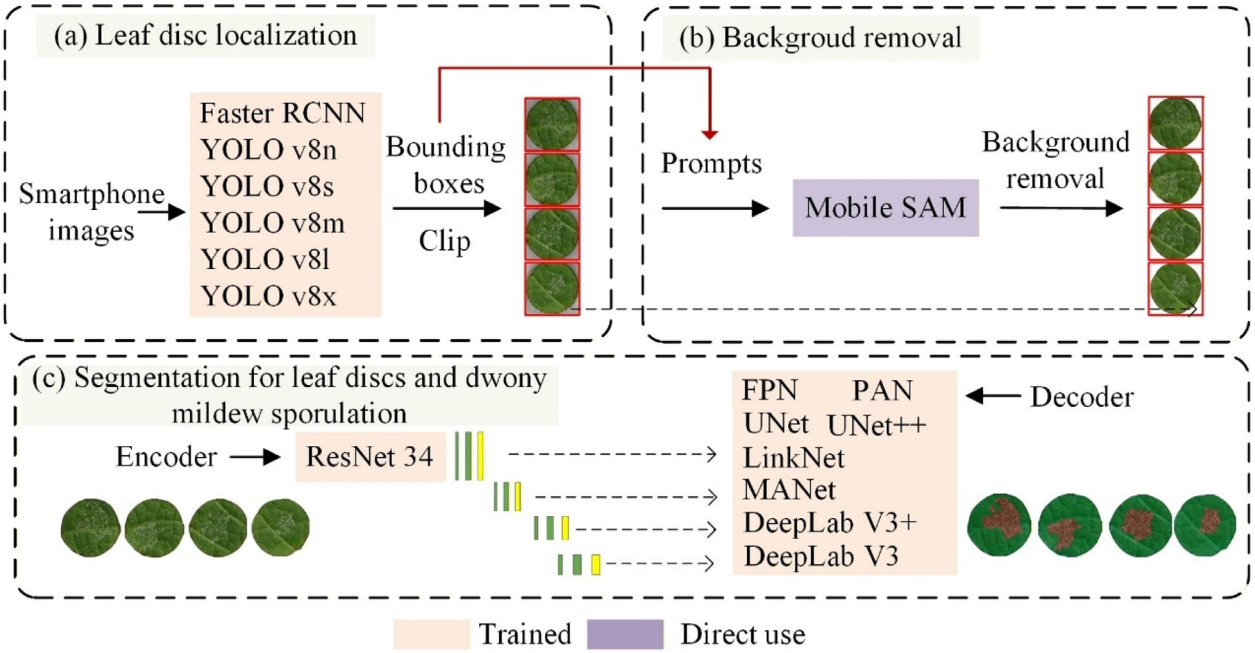

Figure 4. The complete workflow for automated severity assessment of grapevine downy mildew on leaf discs. (a) Models employed for leaf discs localization in smartphone images, (b) Background removal using Mobile SAM, with bounding boxes provided as input prompts, (c) Models utilized for segmenting leaf discs and downy mildew sporulation.

Table 1. Overview of the eight segmentation models employed for leaf disc and downy mildew sporulation segmentation, highlighting their key features and corresponding references

Model	Key Feature	References
UNet	A foundational encoder-decoder architecture integrating feature maps from the encoder with those in the decoder at corresponding depths.	24
UNet++	Enhanced UNet with densely connected pathways facilitating feature propagation between different depths.	25
DeepLabV3	Using dilated convolutions to expand the receptive field and Atrous Spatial Pyramid Pooling (ASPP) for multi-scale feature fusion.	26
DeepLabV3+	Incorporates Xception and depth-wise separable convolution into the ASPP module.	27
FPN	Feature Pyramid Network that constructs multi-scale feature maps.	28
PAN	Extends FPN by introducing attention modules before feature map fusion at different scales.	29
MANet	Introduces multiple kernel attention modules to refine feature maps.	30
LinkNet	Lightweight segmentation model optimized for efficiency while maintaining good performance.	31

(Fig. S2). These contours were subsequently utilized to remove the background through bitwise operations, implemented using OpenCV 4.2.[23] As a result, each leaf disc, with a clean background, was effectively isolated from the input image.

2.3.4 *Segmentation for leaf disc and downy mildew sporulation*

To select the most suitable model, we employed eight state-of-the-art models for segmentation of leaf disc and sporulation. The evaluated models include UNet, UNet++, DeepLabV3, DeepLabV3+, FPN, PAN, MANet, and LinkNet. These models were implemented using encoder-decoder architecture, which had been proven effective in accurately identifying sporulation on leaf discs.[9] Key features and corresponding references for these models are summarized in Table 1.

Furthermore, to assess the impact of the background removal, these models were trained on two distinct datasets of leaf disc images: one containing images with a clean background, and the other with the original background. Images without annotated sporulation were excluded to ensure accurate model evaluation. As a result, 2658 leaf disc images were included in each dataset.

For both datasets, the leaf disc images, along with their corresponding masks, were randomly split into a training (1595 images), validation (531 images), and test sets (532 images) following a 6:2:2 ratio, with grapevine varieties represented in each subset to minimize potential bias associated with varietal differences. All eight segmentation models were configured with the same hyperparameter settings. Specifically, ResNet-34 was utilized as the backbone network for feature extraction, with the initial pre-trained weights on the ImageNet dataset.[32,33] The image encoders were configured with a depth of 5. The number of epochs for model training was set to 300, with early stopping after 10 epochs without improvement. A batch size of 64 was used. The

Adam optimizer was used with an initial learning rate of 0.0001, and the Dice coefficient was selected as the loss function. The weights yielding the best performance on the validation set were retained.

Mean Intersection over Union at a threshold of 0.5 (mIoU@50) was utilized to evaluate the performance of these segmentation models. The Intersection over Union (IoU) is defined as the ratio of the overlap area to the union area between the corresponding masks for manual annotation and model prediction. Specifically, the overlap area was the true positive pixels, and the union area is the sum of true positive pixels (TP), false positive pixels (FP), and false negative pixels (FN). Consequently, mIoU@50 represents the average IoU across all classes. The formula for mIoU@50 is expressed as Eqn (2):

$$\text{mIoU@50} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \frac{\text{TP}_i}{TP_i + FP_i + FN_i} \times 100\,\% \tag{2}$$

where N represents the total number of categories, TP_i denotes the number of true positive samples correctly predicted for the ith category, FP_i represents the number of false positive samples predicted for the ith category, and FN_i represents the number of false negative samples for the ith category. The mIoU@50 ranges from 0% to 100%, with higher values indicating superior segmentation performance. Similarly, FPS was also utilized to compare the inference times and computational efficiency.

2.4 Performance comparison with end-to-end mask R-CNN

As previously discussed, Mask R-CNN is widely utilized as one of the most effective models for instance segmentation, and has demonstrated end-to-end potential for automating disease severity assessments.[11] Therefore, the performance of SporaScan was further evaluated in comparison with Mask R-CNN for both leaf disc localization and the segmentation sporulation. Mask R-CNN (approximately 44 million in parameter size) was trained using the original 432 images, with the model initialized from the weights pre-trained on COCO dataset.[21] The SGD (Stochastic Gradient Descent) was set as the optimizer, with an initial learning rate of 0.001 and a momentum of 0.9. A batch size of 2 was used due to memory limitations. For the evaluation, mAP@50 was used in leaf disc localization, and mIoU@50 for segmentation. It is important to note that the segmentation performance of Mask R-CNN may be lower than that of SporaScan due to the absence of annotated contours for some samples of unclear sporulation. However, accurate leaf disc localization is a prerequisite for effective segmentation, and therefore, the primary focus of this comparison is the performance of localization.

2.5 Evaluation of predicted severity with manual annotations

To assess the efficiency of SporaScan, severity was calculated based on the predicted masks of leaf disc and sporulation. Severity is defined as the proportion of the leaf disc area covered by sporulation, serving as a key indicator for evaluating the resistance of grapevine varieties.[7,10] It can be calculated using Eqn (3):

$$\text{Severity} = \frac{\text{Area of sporulation}}{\text{Area of leaf disc}} \times 100\,\% \tag{3}$$

The two masks for each image – one for the leaf disc and one for the sporulation – were firstly converted into binary masks using a threshold of 0.5, with pixels classified as 1 indicating correct classification and those classified as 0 indicating incorrect classification. The area of sporulation was determined by the sum of pixels correctly classified as itself, while the area of the leaf disc by the sum of the pixels correctly classified as both leaf disc and sporulation. The severity from manual annotation, was calculated directly based on the contour areas of the annotated leaf disc and sporulation.

The coefficient of determination (R^2) was set to evaluate the consistency between manual annotation and model prediction. The formula for R^2 is expressed in as Eqn (4):

$$R^2 = 1 - \frac{\sum_{\text{i}=1}^{\text{n}} (y_i - \hat{y_i})^2}{\sum_{\text{i}=1}^{\text{n}} (y_i - \overline{y})^2} \tag{4}$$

where y_i represents the severity of manual annotation for each leaf disc, $\hat{y_i}$ represents the severity of model prediction for each leaf disc, $\overline{y}$ represents the mean severity of manual annotation across all leaf discs, and n represents the sample size. Only images in the test set were considered for evaluation.

2.6 Blind test of SporaScan predictions *versus* manual annotations

To evaluate the practical performance of SporaScan relative to manual annotations, a blind test was conducted to assess technician preferences under realistic conditions.

An online platform was developed to display three types of images for each sample: (1) the original leaf disc image, (2) the manually annotated mask overlaid on the image with 50% transparency, and (3) the SporaScan-predicted mask overlaid on the image with 50% transparency. The latter two were presented in random order to avoid bias.

A total of 11 grape breeding technicians participated in the evaluation. Each technician independently assessed the same set of 100 randomly selected images from the test dataset, resulting in 1100 individual evaluations. For each image, technicians were asked to select the most accurate annotation among three options: (1) manual annotation is superior, (2) model prediction is superior, or (3) both annotations are equally accurate.

The proportion of each response category was calculated across all evaluations. To statistically assess differences in preference distribution, a Chi-square test for proportions was applied. All analyses were performed at a significance level of $P < 0.05$.

2.7 Computing environments

The experiments were conducted on hardware consisting of an Intel (R) Core (TM) i7-13700KF CPU, 128G memory, and an NVIDIA GeForce GTX 4090 GPU with CUDA 12.2 driver and 24G memory. Training and validation for YOLO series were conducted using scripts implemented by Ultralytics (https://docs.ultralytics.com/), while Faster R-CNN was implemented in TorchVision (https://github.com/pytorch/vision). The scripts for training Mask R-CNN are available at https://github.com/WZMIAOMIAO/deep-learning-for-image-processing. The SporaScan web application is available at http://116.10.197.212:9060/segment/#/. To improve service availability, an alternative mirror server is also provided at http://118.89.50.72:18009/segment/#/. The source code and trained model weights are publicly available at https://github.com/weathergit/LeafDiscDiagnosis, and a step-by-step video tutorial is available at https://www.youtube.com/watch?v=2TJMRtRuWIE.

Table 2. Models' performance for grapevine leaf discs localization.

Models	mAP@50^{val}	mAP@50^{test}	Params (M)	FPS
Faster R-CNN	99.8%	99.9%	41.3	19
YOLO v8 series				
YOLO v8n	**99.4%**	**99.5%**	**3.0**	**213**
YOLO v8s	99.4%	99.4%	11.1	196
YOLO v8m	99.4%	99.5%	25.8	185
YOLO v8l	99.4%	99.5%	43.6	110
YOLO v8x	99.4%	99.5%	68.2	120

Note: mAP@50 represents the mean of average precision for each class at a threshold of 0.5. Params denotes the size of models' parameters in million. FPS refers to the number of frames processed per second. Bold values indicate YOLOv8n as the selected optimal detection model considering accuracy, model size, and inference speed.

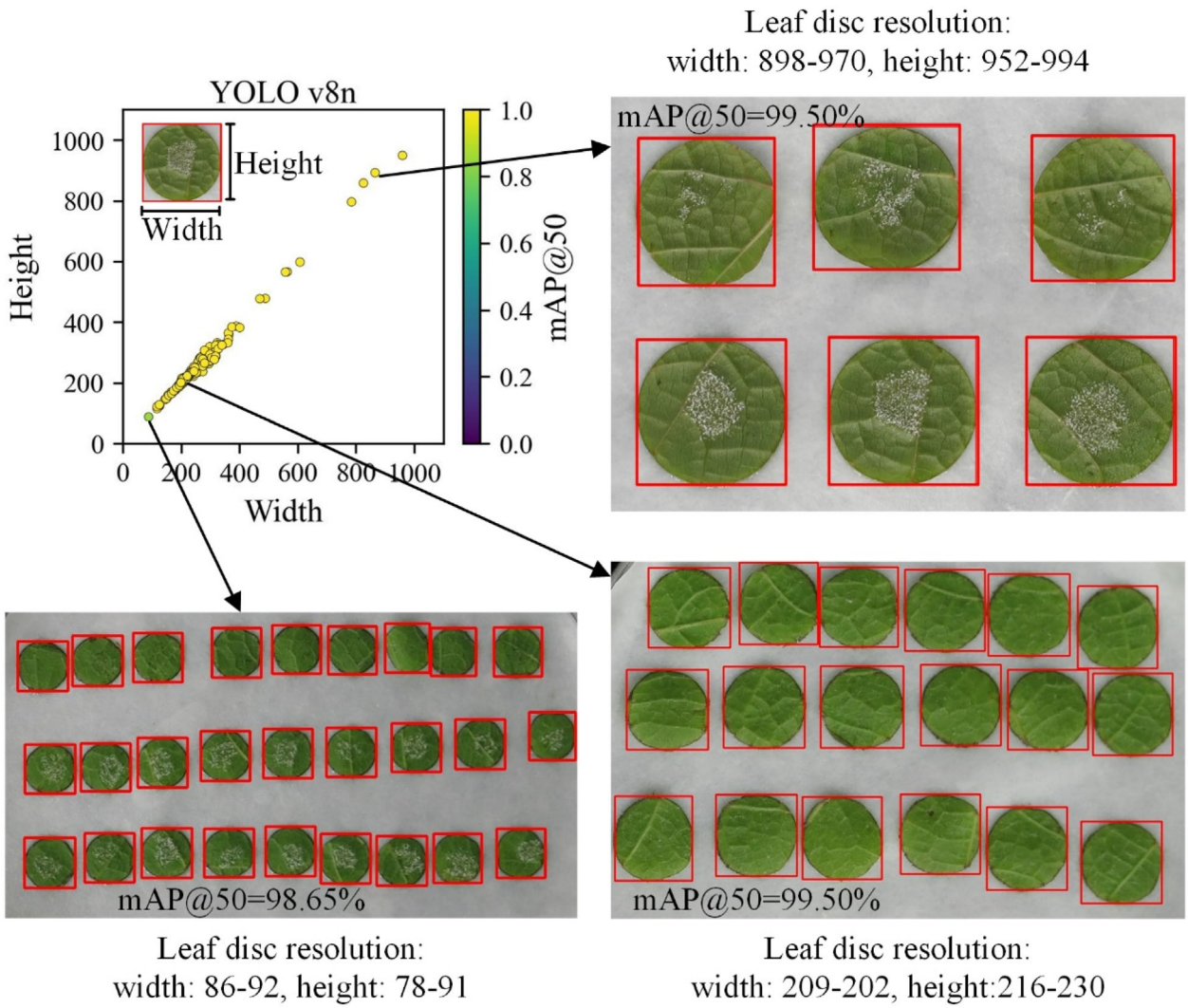

Figure 5. Predicted localization of grapevine leaf discs at different resolutions by the YOLO v8n. Red boxes represent the predicted bounding boxes of grapevine leaf discs. 'Width' is the horizontal pixel number, and 'Height' is the vertical pixel number. The background is a Petri dish.

3 RESULTS

3.1 Performance of YOLO v8 series in leaf disc localization

Table 2 summarizes the performance of the YOLO v8 series and Faster R-CNN in leaf disc localization. All models achieved a mAP@50 exceeding 99% on both the validation and test sets, indicating exceptional accuracy and robust generalization. Among the YOLO v8 series, YOLO v8n, which has the smallest parameter size (3.0 M), achieved a comparable mAP@50 to YOLO v8x (68.2 M). Furthermore, YOLO v8n exhibited the highest inference

Table 3. The performance of the eight segmentation models in segmenting grapevine leaf discs and downy mildew sporulation.

	Without background removal		With background removal			
Model	mIoU@50^{val}	mIoU@50^{test}	mIoU@50^{val}	mIoU@50^{test}	Params (M)	FPS
DeepLab V3	95.84%	96.02%	96.22%	96.37%	26.00	234
DeepLab V3+	95.68%	95.79%	96.08%	96.25%	22.44	280
FPN	95.84%	95.95%	96.25%	96.44%	23.15	289
LinkNet	95.77%	95.93%	96.17%	96.31%	21.77	274
MANet	95.83%	95.94%	96.33%	96.53%	31.78	237
PAN	95.68%	95.82%	96.21%	96.38%	21.47	262
UNet++	95.96%	96.10%	96.33%	96.48%	26.08	236
UNet	95.82%	96.07%	**96.42%**	**96.54%**	24.44	276

Note: mIoU@50 represents the mean of IoU (Intersection over Union) for each class at a threshold of 0.5. Params denotes the size of models' parameters in million. FPS refers to the number of frames processed per second. Bold values indicate the best segmentation performance among the evaluated models.

speed, recording 212.8 FPS, while YOLO v8l had the lowest (109.8 FPS). Although Faster R-CNN slightly outperformed the YOLO v8 series in terms of mAP@50, it demonstrated a significantly lower inference speed due to its larger parameter size and computational complexity. Figure 5 illustrates the accurate localization of leaf discs under varying resolutions and conditions, confirming the model's capability to precisely detect both healthy and infected leaf discs.

3.2 Performance of segmentation models: Mobile SAM and UNet

Table 3 summarizes the performance of eight segmentation models in segmenting grapevine leaf discs and downy mildew sporulation. With background removal, all segmentation models demonstrated consistent performance improvements (+0.4%) on both the validation and test sets, achieving high mIoU@50 exceeding 96%, along with reliable generalization performance. The average mIoU@50 reached 96.2% on the validation set and 96.4% on the test set. The performance differences were minimal among these models. UNet slightly outperformed the other models, achieving the highest mIoU@50 on validation (96.42%) and test sets (96.54%). Most models, with parameter sizes between 21.47 M and 26.08 M, demonstrated closely comparable performance. In terms of inference time, FPN achieved the highest FPS at 289, while DeepLab V3 exhibited the slowest FPS at 234 FPS. Despite a similar mIoU@50 to UNet, UNet++ demonstrated significantly lower FPS. UNet balanced high mIoU@50 with a competitive FPS of 276.

Figure 6 presents the confusion matrix illustrating the classification performance among the background, grapevine leaf disc, and downy mildew sporulation. With background removal, the misclassification rates between categories generally decreased, except that between leaf disc and sporulation. The misclassification rate between the leaf disc and the background decreased by 2.77%. The misclassification rate between sporulation and background was reduced to zero. However, the misclassification rate between sporulation and leaf disc increased approximately by 1%. The differences were minimal in misclassification rates between other categories.

Figure 7 presents the segmentation instances of leaf disc and downy mildew sporulation produced by the eight segmentation models. These examples highlight the strengths of these models.

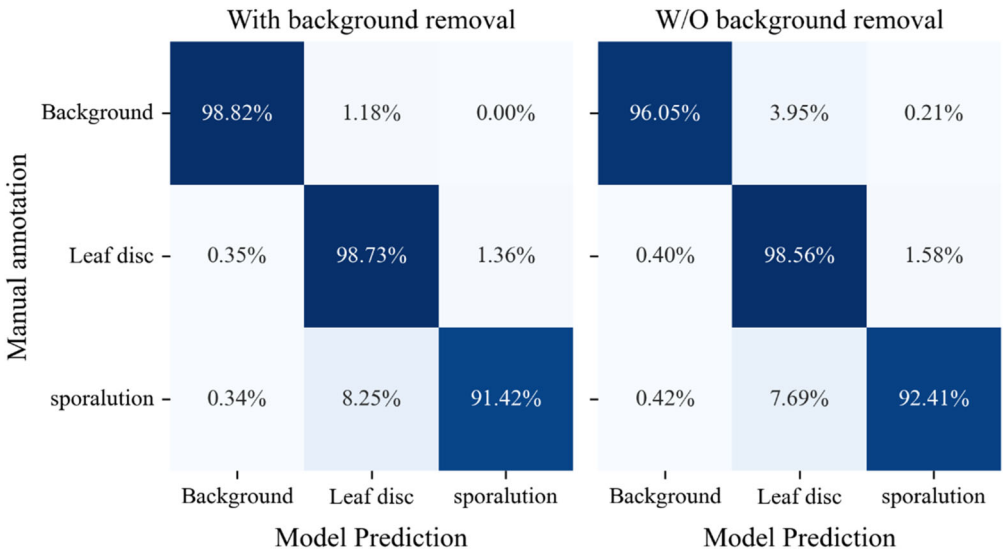

Figure 6. The confusion matrix of the classification performance among the three categories: background, leaf disc and sporulation with and w/o background removal by Mobile SAM. The values in each cell are normalized by row, representing the inter-class misclassification rates.

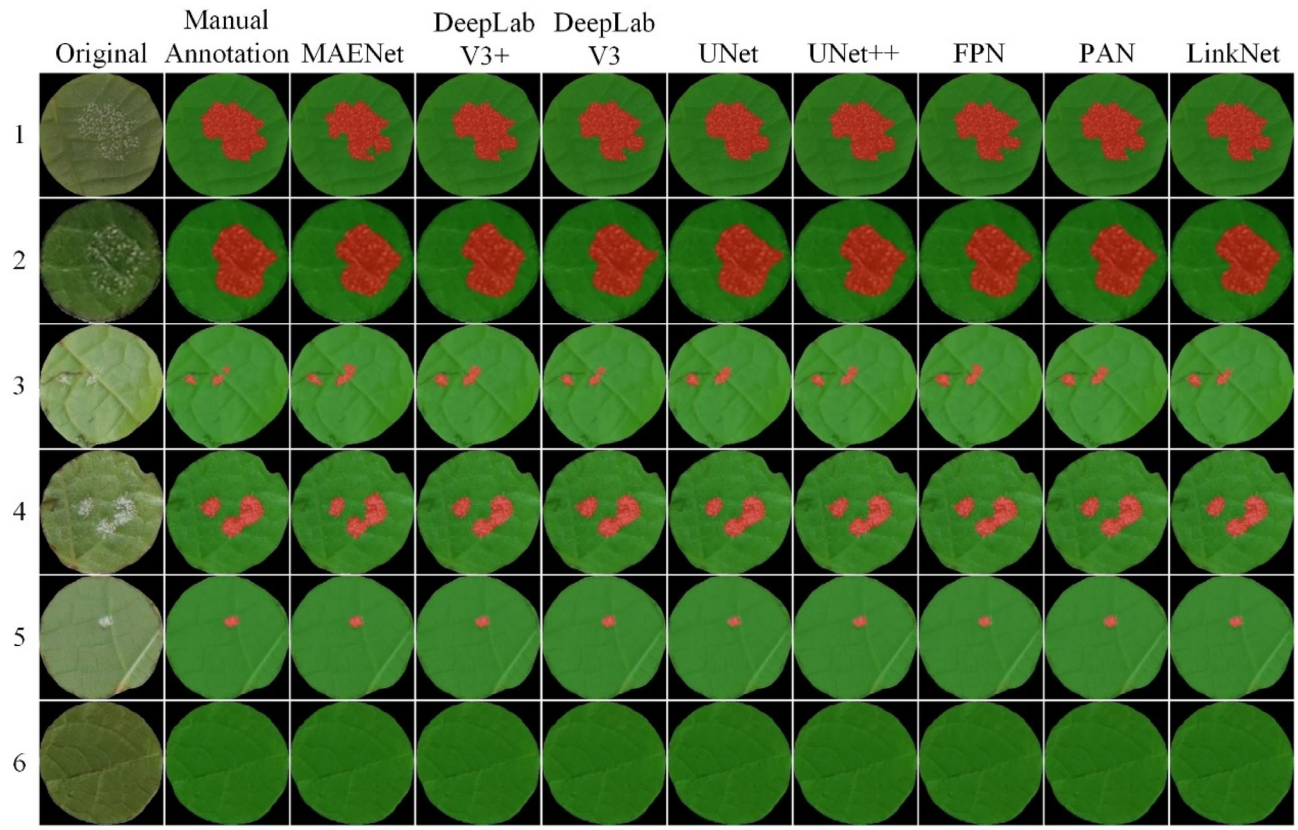

Figure 7. Representative examples of downy mildew sporulation segmentation on leaf discs. The 'Origin' column displays the original image, while 'Manual annotation' column shows the masks from manual annotation. The other columns present the predicted masks generated by the eight segmentation models.

The first two rows emphasize the model's ability to accurately segment diseased regions on leaf discs. Rows 3 to 5 showcase the model's ability to segment both large and multiple small patches of downy mildew sporulation. The final row demonstrates the accurate segmentation of healthy leaf discs. Throughout all instances, the models consistently provided accurate segmentation, with only slight visual differences observed when compared to manual annotations in the segmentation of leaf discs and downy mildew sporulation.

3.3 Performance comparison between SporaScan and mask R-CNN

Table 4 presents a comparative analysis of the performance of Mask R-CNN and SporaScan in leaf disc localization and segmentation of leaf discs and downy mildew sporulation. SporaScan achieved higher mAP@50 in leaf disc localization and higher mIoU@50 in segmentation. The training process of Mask R-CNN showed stable convergence in both mAP@50 and loss (Fig. S4), indicating that its inferior performance was not due to insufficient training. Specifically, SporaScan outperformed Mask R-CNN in leaf disc localization by 4.51% and 4.90% on the validation and test sets, respectively. In the segmentation of leaf discs and sporulation, SporaScan demonstrated an improvement of 8.04% on the validation set and 9.27% on the test set. This highlights its superior performance and enhanced generalization. In terms of computational efficiency, SporaScan achieved an inference speed of 0.6 FPS, which was markedly lower than that of Mask R-CNN (10.0 FPS).

Figure 8 presents representative instances of model predictions by Mask R-CNN and SporaScan. The primary failure cases observed in Mask R-CNN included incomplete leaf disc localization and redundant segmentation of sporulation. Notably, the

Table 4. The performance of Mask R-CNN and the SporaScan. mAP@50 represents the mean average precision at a threshold of 0.5.

	Leaf disc location		Segmentation for leaf disc and sporulation		
Model	$mAP@50^{val}$	$mAP@50^{test}$	$mIoU@50^{val}$	$mIoU@50^{test}$	FPS
Mask R-CNN	94.89%	94.60%	88.36%	87.23%	10.0
SporaScan	**99.40%**	**99.50%**	**96.40%**	**96.50%**	0.6

Note: mIoU@50 represents the mean of IoU (Intersection over Union) at a threshold of 0.5. FPS refers to the number of frames processed per second. Bold values indicate the superior performance of SporaScan compared with Mask R-CNN.

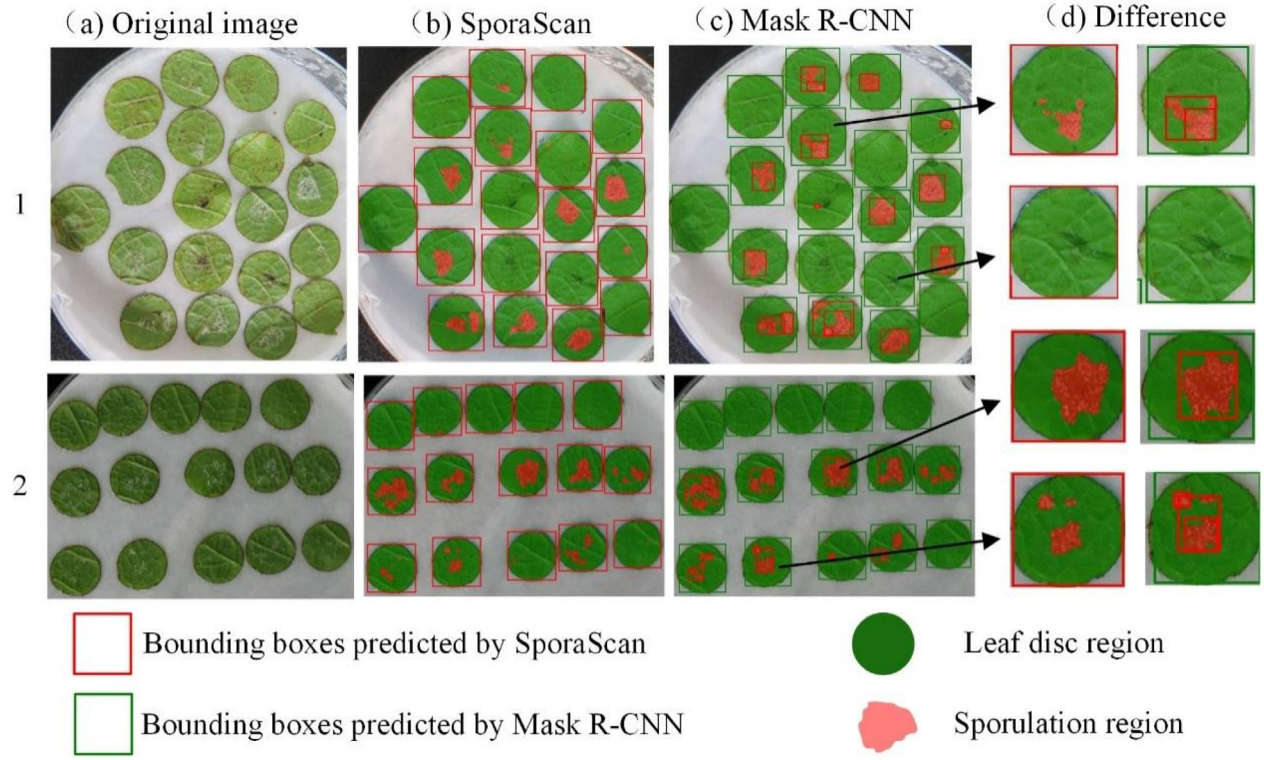

Figure 8. The instances of grapevine leaf disc localization and the segmentation of both leaf discs and sporulation. (a) original images, (b) predicted bounding boxes and masks by SporaScan, (c) predicted bounding boxes and masks by Mask R-CNN, (d) differences between the two models.

unlocated portions of leaf discs were predominantly positioned on the left side of the images. In contrast, SporaScan, employing a sequential processing approach, accurately localized leaf discs while effectively preventing redundant sporulation masks, demonstrating its robustness and precision in segmentation.

3.4 Performance of the SporaScan pipeline in disease severity prediction

Figure 9 illustrates the relationship between manually annotated severity and SporaScan-predicted severity for grapevine downy mildew, together with representative visualization results. In the test dataset, most severity values were concentrated between 0% and 40%, with relatively fewer samples distributed between 40% and 60% (Fig. S1). As shown in Fig. 9(a), SporaScan predictions exhibited a strong agreement with manual annotations, achieving a coefficient of determination (R^2) of 0.992, indicating high consistency between the two approaches. Representative examples are presented in Fig. 9(b), demonstrating that SporaScan can accurately capture both the spatial distribution and extent of sporulation across leaf discs. To further evaluate model performance under early-stage infection conditions, a subset of samples with severity lower than 5% was analyzed (Fig. 9(c)). Although the dynamic range of severity was substantially reduced in this subset, the model still maintained a reasonable level of agreement with manual annotations ($R^2 = 0.941$), suggesting that SporaScan is capable of detecting and quantifying subtle disease symptoms.

3.5 Blind test results: SporaScan *vs.* manual annotations

Figure 10 shows the distribution of evaluators' preferences among the three categories: manual annotation better, equal accuracy, and SporaScan better. SporaScan was preferred in 37.2% of cases (409/1100), whereas manual annotation was preferred in 26.2% (288/1100), and equal performance was observed in 36.6% (403/1100). A chi-square goodness-of-fit test revealed that the distribution differed significantly from an equal distribution ($\chi^2 = 25.37$, $P < 0.001$). Specifically, cases favoring manual annotation were less frequent than expected, while both SporaScan-preferred and equal-performance cases occurred more frequently. These results indicate that SporaScan achieved performance comparable to or better than manual annotation in most cases.

4 DISCUSSION

Accurate assessment of disease severity is essential for evaluating the effectiveness of newly developed fungicides and resistant cultivars. However, traditional severity assessments using leaf disc assays are labor-intensive, time-consuming, and prone to variability and bias arising from differences in technician skill and subjective judgment. In this study, we present SporaScan, a deep learning-based pipeline that integrates YOLOv8n for leaf disc detection, Mobile SAM for background removal, and UNet for sporulation segmentation. Utilizing a cost-effective dataset captured with smartphones, SporaScan achieves superior accuracy compared to end-to-end models such as Mask R-CNN and even surpasses manual assessments in a blinded evaluation. This practical tool improves the consistency and efficiency of disease severity assessment, supporting breeding programs focused on disease-resistant cultivars and fungicide evaluation.

4.1 Sequential SporaScan outperforms end-to-end mask R-CNN

Mask R-CNN is a widely adopted end-to-end instance segmentation model that, in principle, can compute disease severity by simultaneously detecting and segmenting leaves and disease

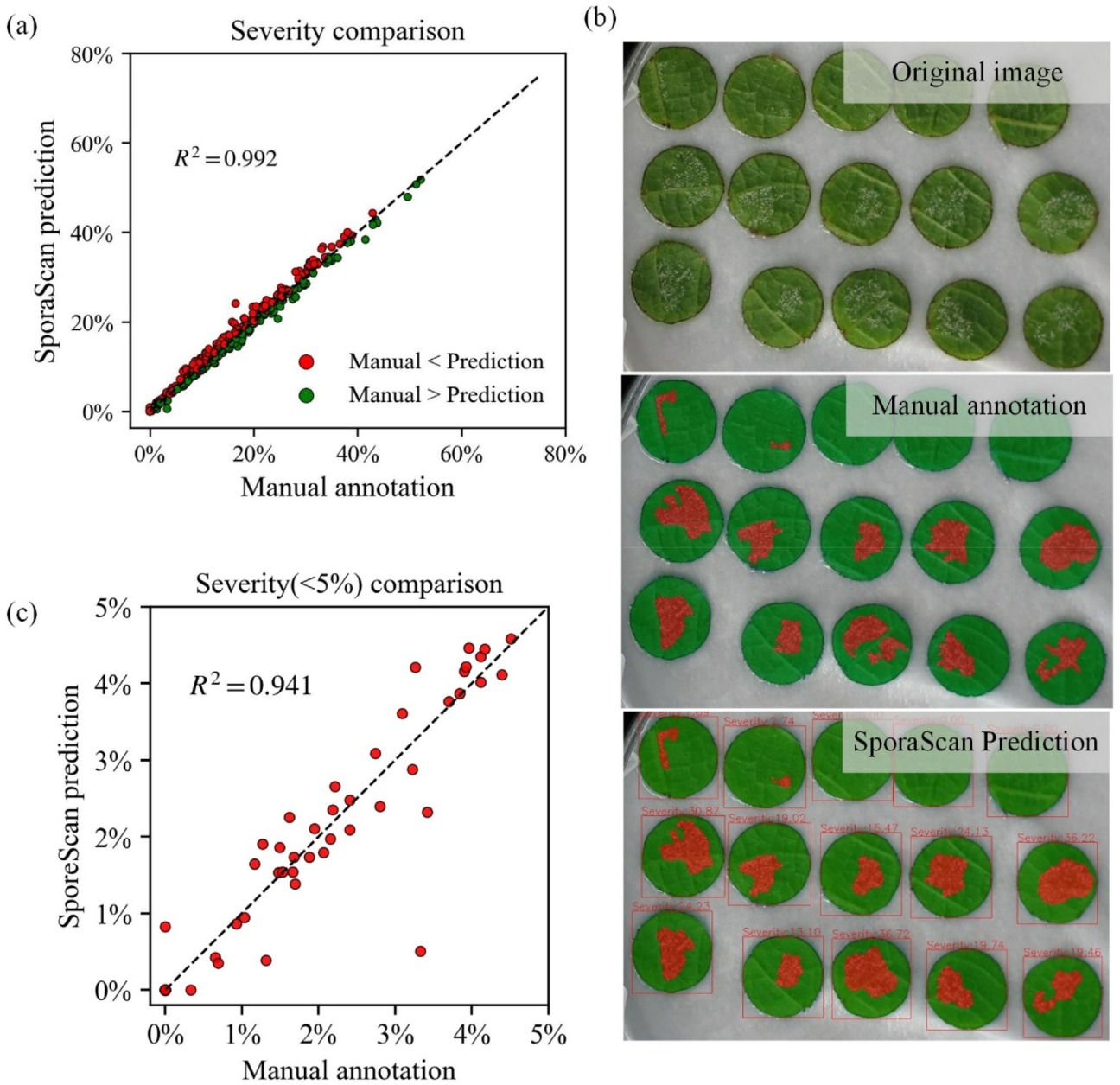

Figure 9. Comparison between manual annotation and SporaScan prediction for downy mildew severity, along with representative visualization examples. (a) Relationship between manually annotated severity and SporaScan-predicted severity across all samples, with the dashed line indicating the 1:1 reference. (b) Representative examples including the original image, manual annotation, and SporaScan prediction. (c) Relationship between manual annotation and model prediction for samples with severity lower than 5%, highlighting model performance under early-stage infection conditions.

regions.[11,34] Although this approach offers an attractive end-to-end solution, experiments in this study reveal that Mask R-CNN underperforms compared to the sequential pipeline implemented in SporaScan (Table 4). Despite stable training convergence (Fig. S4), this performance gap suggests inherent limitations of the end-to-end architecture, potentially arising from multiple technical factors.

One key issue is that Mask R-CNN relies on a region proposal network for object detection. While the distinct color contrast between leaf discs and the background favors detection, the complex background and subtle boundaries between leaf discs and sporulation often confound the region proposals, leading to less precise localization of the leaf discs. In contrast, SporaScan employs YOLOv8n, which is specifically optimized for single-object detection, achieving a mAP@50 exceeding 99% through robust localization that minimizes misclassification.

Another factor is the inherent challenge in task specialization within an integrated end-to-end framework. Mask R-CNN is required to perform detection, background differentiation, and fine-grained segmentation simultaneously, which prevents the model from optimizing any single task.[17] In particular, Mask R-CNN segment the sporulation after detecting them. For discrete but closely located sporulation regions, it may detect a single larger region and multiple smaller regions, leading to duplicated segmentations of the sporulation (Fig. 8). Conversely, SporaScan circumvents this limitation by decoupling the workflow: it uses

Mobile SAM for effective background removal and UNet for accurate segmentation. This modular approach allows each component to focus on its specific function, resulting in a segmentation mIoU@50 exceeding 96% and reducing the uncertainty that typically arises when multiple tasks are handled by a single model.

Furthermore, the complexity of Mask R-CNN's architecture, with its larger parameter size and increased computational overhead, is generally associated with slower inference in many applications. However, in this study, SporaScan exhibited a lower inference speed compared to Mask R-CNN due to its sequential pipeline design, where multiple models are executed consecutively. Despite this limitation, SporaScan achieved substantially higher accuracy in both localization and segmentation.

In summary, while Mask R-CNN offers the convenience of an end-to-end severity assessment, its limitations in localization precision and task specialization hinder its overall performance. The sequential design of SporaScan, by separating and optimizing each individual task, provides a more accurate and robust solution for disease severity assessment, albeit at the cost of reduced computational efficiency.

4.2 Superiority of SporaScan in automated grapevine severity assessment

Accurate and automated quantification of disease severity is essential for breeding resistant varieties, necessitating an efficient and robust model to minimize bias introduced by manual annotations. In this study, SporaScan's severity assessments exhibit excellent agreement with manual annotations (Figs 7 and 9), and even surpass them in a blind preference test conducted among technicians (Fig. 10). This performance suggests that SporaScan not only replicates manual evaluations with high accuracy but also overcomes the inconsistencies and subjective biases inherent in manual annotation.

A key advantage of SporaScan lies in its ability to perform precise, pixel-level segmentation – an accuracy level that technicians cannot consistently achieve through manual annotation. Manual methods often yield approximations of disease regions, whereas models like UNet are designed to delineate these regions with high fidelity. As a result, SporaScan effectively isolates and quantifies disease symptoms from structured input data, particularly under controlled conditions with minimal background interference. Recent studies have demonstrated that CNN-based models can achieve accuracy comparable to manual annotations when assessing disease severity using high-quality microscopic image.[9,35]

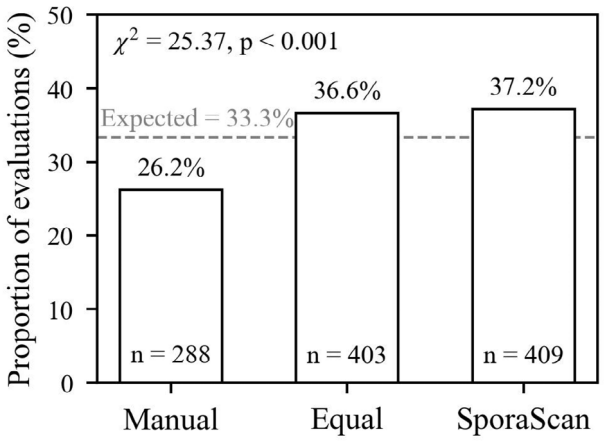

Figure 10. Distribution of option preferences among three categories: manual better, equal, and SporaScan better. Bar heights indicate the proportion of evaluations in each category, and sample sizes are shown inside the bars. The dashed line represents the expected proportion under a uniform distribution (33.3%). The category distribution differed significantly from the expected equal distribution ($\chi^2 = 25.37$, $P < 0.001$).

From a practical standpoint, these findings underscore both the advantages and limitations of CNN-based models in disease severity assessment. Technicians showed a preference for model predictions over manual annotations (Fig. 10), likely due to the lower error rates and full automation associated with CNN-based models, even when applied to low-cost images captured *via* smartphones. Observed limitations may stem from noise introduced by small, discrete sporulation and variable leaf surface illumination (Fig. S3). Future research could focus on advancing disease severity assessment by addressing challenges such as complex backgrounds, overlapping leaves, and diverse plant growth stages.[36]

4.3 Limitations and future work

This study has several limitations related to data size, imaging conditions, and varietal representation. The dataset, consisting of 432 images, may constrain model generalizability, as large-scale datasets for leaf disc assays are difficult to obtain due to controlled experimental requirements and labor-intensive annotation, particularly for small and irregular sporulation regions. To address this limitation, transfer learning was adopted to enhance model performance under limited data conditions. Variability in smartphone-based image acquisition, including differences in resolution, focal length, and illumination, may influence segmentation accuracy, especially for subtle disease symptoms. However, such variability reflects non-standardized conditions in practical breeding scenarios. The stable performance observed in this study indicates that the proposed pipeline can accommodate this variability to a certain extent, although its specific impact warrants further investigation.

In addition, the dataset includes a limited number of grapevine varieties, which may not fully capture the diversity of grapevine germplasm in terms of leaf morphology, pubescence, and disease expression. This limitation is particularly relevant for early-stage disease assessment, where symptom expression is subtle and may vary across genetic backgrounds. To further examine model performance under such conditions, an additional analysis focusing on samples with severity lower than 5% was conducted, demonstrating the model's capability in detecting and quantifying early-stage disease symptoms.

Further validation using larger datasets, more diverse imaging conditions, and a wider range of grapevine varieties is needed to comprehensively assess the robustness and applicability of the proposed approach.

5 CONCLUSIONS

This study introduces SporaScan, an automated pipeline that integrates YOLO v8n for leaf disc localization, Mobile SAM for background removal, and UNet for segmentation in grapevine downy mildew assessment. Utilizing 432 smartphone images encompassing over 6000 leaf discs, SporaScan achieved a mAP@50 exceeding 99%, an mIoU@50 above 96%, and an R^2 of 0.99 – demonstrating performance comparable to manual annotation while significantly reducing labor and subjectivity.

Additional analysis for samples with severity <5% further demonstrates its potential for early-stage disease detection.

These results underscore the potential of deep learning to streamline workflows in breeding diseases-resistant crop varieties, fungicide evaluation, and related applications by decomposing complex tasks into specialized modules. Future research should focus on extending this approach to diverse environmental conditions and crop types, improving segmentation accuracy for complex disease patterns, and incorporating higher-quality datasets and field trials to enhance robustness and generalizability.

ACKNOWLEDGEMENTS

This work was supported by the Guangxi Key Research and Development Program Project (Grant No. GuiKeAB21076001), and the Key Research and Development Program of Shaanxi (Grant No. 2023-ZDLNY-64).

DATA AVAILABILITY STATEMENT

All datasets, scripts, and model weights used in this study will be made available upon publication.

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

REFERENCES

1 Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N and Nelson A, The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* **3**:430–439 (2019). https://doi.org/10.1038/s41559-018-0793-y.
2 Wightwick A, Walters R, Allinson G, Reichman S and Menzies N, Environmental risks of fungicides used in horticultural production systems. *Fungicides* 273–304 (2010). https://doi.org/10.5772/13032.
3 Kunova A, Pizzatti C, Saracchi M, Pasquali M and Cortesi P, Grapevine powdery mildew: fungicides for its management and advances in molecular detection of markers associated with resistance. *Microorganisms* **9**:1541 (2021). https://doi.org/10.3390/microorganisms9071541.
4 Nutter FW, Esker PD and Netto RAC, Disease assessment concepts and the advancements made in improving the accuracy and precision of plant disease data. *Eur J Plant Pathol* **115**:95–103 (2006). https://doi.org/10.1007/s10658-005-1230-z.
5 Ngugi LC, Abelwahab M and Abo-Zahhad M, Recent advances in image processing techniques for automated leaf pest and disease recognition-a review. *Inf Process Agric* **8**:27–51 (2021). https://doi.org/10.1016/j.inpa.2020.04.004.
6 Hernández I, Gutiérrez S, Ceballos S, Palacios F, Toffolatti SL, Maddalena G *et al*., Assessment of downy mildew in grapevine using computer vision and fuzzy logic. Development and validation of a new method. *Oeno One* **56**:41–53 (2022). https://doi.org/10.20870/oeno-one.2022.56.3.5359.
7 Hernández I, Gutiérrez S and Tardaguila J, Image analysis with deep learning for early detection of downy mildew in grapevine. *Sci Hortic* **331**:113155 (2024). https://doi.org/10.1016/j.scienta.2024.113155.
8 Bierman A, LaPlumm T, Cadle-Davidson L, Gadoury D, Martinez D, Sapkota S *et al*., A high-throughput phenotyping system using machine vision to quantify severity of grapevine powdery mildew. *Plant Phenomics* **2019**:9209727 (2019). https://doi.org/10.34133/2019/9209727.
9 Lin K, Gong L, Huang Y, Liu C and Pan J, Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front Plant Sci* **10**:155 (2019). https://doi.org/10.3389/fpls.2019.00155.
10 Zendler D, Malagol N, Schwandner A, Töpfer R, Hausmann L and Zyprian E, High-throughput phenotyping of leaf discs infected with grapevine downy mildew using shallow convolutional neural networks. *Agronomy* **11**:1768 (2021). https://doi.org/10.3390/agronomy11091768.
11 Afzaal U, Bhattarai B, Pandeya YR and Lee J, An instance segmentation model for strawberry diseases based on mask R-CNN. *Sensors (Basel)* **21**:6565 (2021). https://doi.org/10.3390/s21196565.
12 He K, Gkioxari G, Dollár P and Girshick R, Mask R-CNN, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. IEEE, Piscataway, NJ, USA, pp. 2961–2969 (2017). https://doi.org/10.1109/ICCV.2017.322.
13 Karim MJ, Nahiduzzaman M, Ahsan M and Haider J, Development of an early detection and automatic targeting system for cotton weeds using an improved lightweight YOLOv8 architecture on an edge device. *Knowl-Based Syst* **300**:112204 (2024). https://doi.org/10.1016/j.knosys.2024.112204.
14 Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L *et al*., Segment Anything (2023).
15 Liu W, Chen Y, Lu Z, Lu X, Wu Z, Zheng Z *et al*., StripeRust-pocket: a Mobile-based deep learning application for efficient disease severity assessment of wheat stripe rust. *Plant Phenomics* **6**:0201 (2024). https://doi.org/10.34133/plantphenomics.0201.
16 Xu S, Shen J, Wei Y, Li Y, He Y, Hu H *et al*., Automatic plant phenotyping analysis of melon (cucumis melo L.) germplasm resources using deep learning methods and computer vision. *Plant Methods* **20**:166 (2024). https://doi.org/10.1186/s13007-024-01293-1.
17 Yang X, Lu X, Xie P, Guo Z, Fang H, Fu H *et al*., PanicleNeRF: low-cost, high-precision in-field phenotyping of rice panicles with smartphone. *Plant Phenomics* **6**:279 (2024). https://doi.org/10.34133/plantphenomics.0279.
18 Solimani F, Cardellicchio A, Dimauro G, Petrozza A, Summerer S, Cellini F *et al*., Optimizing tomato plant phenotyping detection: boosting YOLOv8 architecture to tackle data complexity. *Comput Electron Agric* **218**:108728 (2024). https://doi.org/10.1016/j.compag.2024.108728.
19 Jocher G, Qiu J and Chaurasia A, Ultralytics YOLO (2023).
20 Ren S, He K, Girshick R and Sun J, Faster R-CNN: towards real-time object detection with region proposal networks. arXiv:1506.01497 [cs] (2016). https://doi.org/10.48550/arXiv.1506.01497.
21 Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J *et al*., Microsoft COCO: Common objects in context, in *Proceedings of the European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science*, Vol. **8693**. Springer, Cham, pp. 740–755 (2014). https://doi.org/10.1007/978-3-319-10602-1_48.
22 Zhang C, Han D, Qiao Y, Kim JU, Bae S-H, Lee S *et al*., Faster segment anything: towards lightweight SAM for mobile applications. arXiv:2306.14289 [cs] (2023). https://doi.org/10.48550/ARXIV.2306.14289.
23 Bradski G, The OpenCV library. *Dr Dobbs J Softw Tools* (2000).
24 Ronneberger O, Fischer P and Brox T, U-net: convolutional networks for biomedical image segmentation. arXiv:1505.04597 [cs] (2015). https://doi.org/10.48550/arXiv.1505.04597.
25 Zhou Z, Siddiquee MMR, Tajbakhsh N and Liang J, UNet++: redesigning skip connections to exploit multiscale features in image segmentation. arXiv:1912.05074 [cs] (2020). https://doi.org/10.48550/arXiv.1912.05074.
26 Chen L-C, Papandreou G, Schroff F and Adam H, Rethinking atrous convolution for semantic image segmentation (2017). https://doi.org/10.48550/arXiv.1706.05587.
27 Chen L-C, Zhu Y, Papandreou G, Schroff F and Adam H, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation (2018).
28 Lin T-Y, Dollár P, Girshick R, He K, Hariharan B and Belongie S, Feature pyramid networks for object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA, pp. 2117–2125 (2017). https://doi.org/10.1109/CVPR.2017.106.
29 Li H, Xiong P, An J and Wang L, Pyramid attention network for semantic segmentation. arXiv preprint arXiv:180510180 (2018). Available: https://arxiv.org/abs/1805.10180.
30 Li R, Zheng S, Zhang C, Duan C, Su J, Wang L *et al*., Multiattention network for semantic segmentation of fine-resolution remote sensing images. *IEEE Trans Geosci Remote Sensing* **60**:1–13 (2022). https://doi.org/10.1109/TGRS.2021.3093977.

31 Chaurasia A and Culurciello E, LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation, in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, Piscataway, NJ, USA, pp. 1–4 (2017). https://doi.org/10.1109/VCIP.2017.8305148.
32 Deng J, Dong W, Socher R, Li L-J and Kai Li LF-F, ImageNet: a large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*. IEEE, Miami, FL, pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848.
33 He K, Zhang X, Ren S and Sun J, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA, pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90.
34 Kumar M, Chandel NS, Singh D and Rajput LS, Soybean disease detection and segmentation based on mask-RCNN algorithm. *J Exp Agric Int* **45**:63–72 (2023). https://doi.org/10.9734/jeai/2023/v45i52132.
35 Qiu T, Underhill A, Sapkota S, Cadle-Davidson L and Jiang Y, High throughput saliency-based quantification of grape powdery mildew at the microscopic level for disease resistance breeding. *Hortic Res* **9**: uhac187 (2022). https://doi.org/10.1093/hr/uhac187.
36 Barbedo JGA, Factors influencing the use of deep learning for plant disease recognition. *Biosyst Eng* **172**:84–91 (2018). https://doi.org/10.1016/j.biosystemseng.2018.05.013.

