



## OPEN ACCESS

## EDITED BY

Md. Motaheer Hossain,  
Bangabandhu Sheikh Mujibur Rahman  
Agricultural University, Bangladesh

## REVIEWED BY

Shaikhul Islam,  
Bangladesh Wheat and Maize Research  
Institute (BWMRI), Bangladesh  
Hasan M. Abdullah,  
Bangabandhu Sheikh Mujibur Rahman  
Agricultural University, Bangladesh

## \*CORRESPONDENCE

Qinjian Zhang

✉ zhangqinjian@bistu.edu.cn

†These authors have contributed equally to  
this work

RECEIVED 18 November 2023

ACCEPTED 11 March 2024

PUBLISHED 27 March 2024

## CITATION

Wu Y, Xi Z, Liu F, Hu W, Feng H and  
Zhang Q (2024) A deep semantic  
network-based image segmentation  
of soybean rust pathogens.  
*Front. Plant Sci.* 15:1340584.  
doi: 10.3389/fpls.2024.1340584

## COPYRIGHT

© 2024 Wu, Xi, Liu, Hu, Feng and Zhang. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# A deep semantic network-based image segmentation of soybean rust pathogens

Yalin Wu<sup>1†</sup>, Zhuobin Xi<sup>2†</sup>, Fen Liu<sup>1†</sup>, Weiming Hu<sup>1</sup>,  
Hongjuan Feng<sup>3</sup> and Qinjian Zhang<sup>2\*</sup>

<sup>1</sup>Lushan Botanical Garden, Jiangxi Province and Chinese Academy of Sciences, Jiujiang, China,

<sup>2</sup>Mechanical Electrical Engineering School, Beijing Information Science & Technology University, Beijing, China, <sup>3</sup>Zhongzhen Kejian (ShenZhen) Holdings Co., Ltd, Shenzhen, China

**Introduction:** Asian soybean rust is a highly aggressive leaf-based disease triggered by the obligate biotrophic fungus *Phakopsora pachyrhizi* which can cause up to 80% yield loss in soybean. The precise image segmentation of fungus can characterize fungal phenotype transitions during growth and help to discover new medicines and agricultural biocides using large-scale phenotypic screens.

**Methods:** The improved Mask R-CNN method is proposed to accomplish the segmentation of densely distributed, overlapping and intersecting microimages. First, Res2net is utilized to layer the residual connections in a single residual block to replace the backbone of the original Mask R-CNN, which is then combined with FPG to enhance the feature extraction capability of the network model. Secondly, the loss function is optimized and the CloU loss function is adopted as the loss function for boundary box regression prediction, which accelerates the convergence speed of the model and meets the accurate classification of high-density spore images.

**Results:** The experimental results show that the mAP for detection and segmentation, accuracy of the improved algorithm is improved by 6.4%, 12.3% and 2.2% respectively over the original Mask R-CNN algorithm.

**Discussion:** This method is more suitable for the segmentation of fungi images and provide an effective tool for large-scale phenotypic screens of plant fungal pathogens.

## KEYWORDS

Asian soybean rust, *Phakopsora pachyrhizi*, deep learning, instance segmentation, mask R-CNN

## 1 Introduction

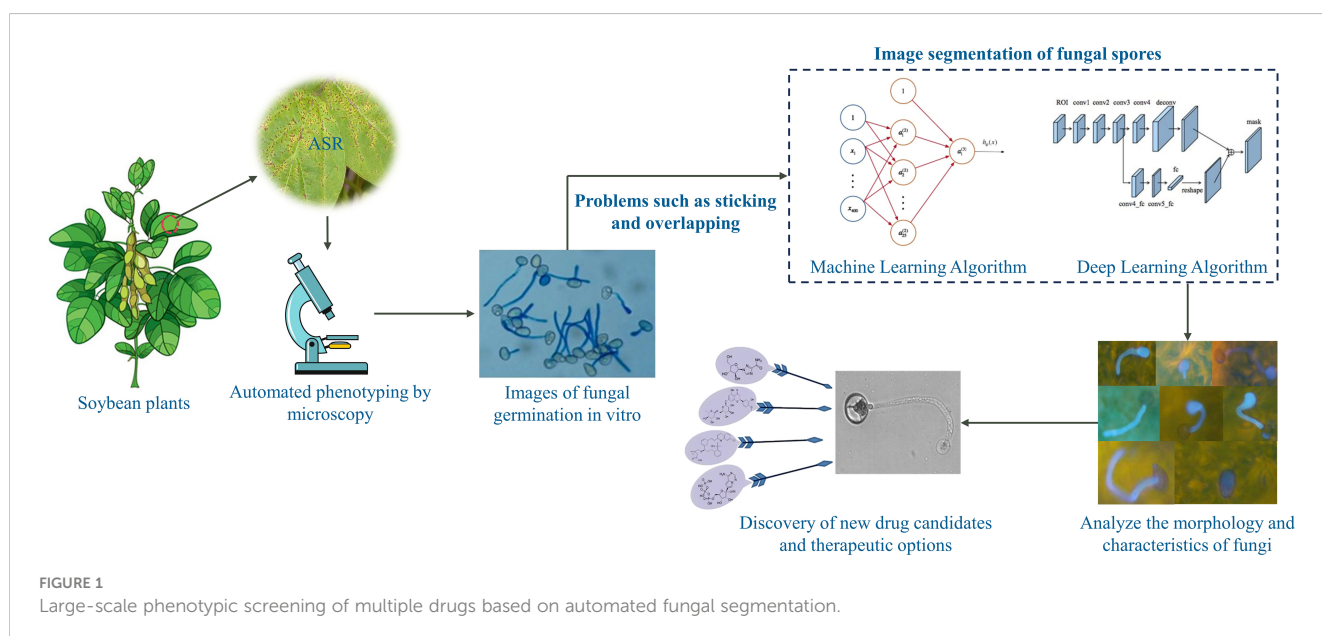
Soybean (*Glycine max*) is one of the most economically efficient crops since it is an important source of food, protein, and vegetable oil. Asian Soybean Rust (ASR) is a globally aggressive foliar disease of soybean plants that can cause up to 80% losses and have a significant impact on production costs in various geographical areas invaded by the pathogen (Lorrain et al., 2019). Fungus *Phakopsora pachyrhizi* is the causal agent of ASR. Infection begins with the deposition of uredospores on soybean leaves, where the rust fungus invades the epidermal cells of the host through the appressorium formed during spore germination and extracts nutrients from the host body (Goellner et al., 2010; Loehrer and Schaffrath, 2011). This fungus can defoliate soybean fields and accelerate maturation with a reduction of seed size and weight and may lead to complete crop failure within a few days (Hartman et al., 2015). Currently, timely fungicide application is the only means of controlling ASR (Saito et al., 2021).

It is important to analyze the characterization of fungi germinating *in vitro* for ASR disease control and research. Researchers have found that automated microscopy-based phenotyping is typically used under genetically or environmentally sensitive conditions to probe the relationship between cell structure and function by unbiased quantification of phenotypic changes in response to perturbations of interest (Liberali et al., 2015; Usaj et al., 2016). Some studies have applied fungal images to the field of drug discovery and development (Calderone et al., 2014; Carolus et al., 2020). By analyzing the morphology and characteristics of fungi, researchers are able to better understand the structure and function of fungi. Statistical information on fungal spores can reveal the degree of resistance and activity of spores to discover new drug candidates and therapeutic options. Large-scale phenotypic screening of multiple compounds acting on fungal spores can identify suitable fungicides and drugs for ASR. However, these drugs usually have to be screened from hundreds of compounds by expert labor, which requires huge

processing time. These image-based methods have found their greatest application in the pharmaceutical industry, where they have been used to primary screening stages of drug discovery, drug target validation, early evaluation of toxicity properties and complex multivariate drug profiling (Zanella et al., 2010; Reisen et al., 2015). For example, haploid yeast was treated with drugs that perturb cell wall and the dose-dependent changes in morphology were analyzed to identify drugs that interfere with cell wall synthesis (Okada et al., 2014).

As show in Figure 1, Segmentation is one of the significant steps of phenotypic screening (Cabre et al., 2021). Accurate image segmentation of fungal spores can characterize phenotypic changes during fungal growth, and accurate segmentation significantly determines the efficiency and effectiveness of drug screening, contributing to the discovery of drugs and agricultural fungicides using large-scale phenotypic screening, as well as to the development of strategies for the control of ASR using biotechnological approaches. Manual segmentation of images is cost-ineffective and time-consuming for expert annotation, and thus is impractical for large data segmentation. More importantly, due to the variability of individuals, manual segmentation can introduce large segmentation errors and biases, so there is a need to find an accurate and efficient automatic segmentation method. Due to the different degree of response of the fungi to different drugs, it appears that the spore morphology of fungi in the process of reproduction appears to have a large morphological variability. Moreover, the interaction between fungi, many fungi overlap, distort and adhere to each other, which can make accurate segmentation difficult. Finally, the collected microscope images have low contrast, and the fungal edges are very blurred and difficult to identify accurately, while some of the images have impurities.

Some microscopy applications use machine learning algorithms, such as those for range thresholding, simple filters, and edge detection based on intensity changes are now widely used (Melo et al., 2019). Traditional image segmentation methods



include Otsu's thresholding (Otsu, 1979), watershed algorithm (Beucher, 1992) and clustering (Coleman and Andrews, 1979). For instance, a Gaussian Separate Degree is used for Otsu method, called as G-Otsu, is proposed to segment anthrax spore images (Zhao et al., 2020). Korsnes et al. (2016) used methods such as mean gradient and morphological processing to detect spore boundaries for spore segmentation, followed by egg shape-fitting techniques to fit spore perimeter. Using K-means method to segment spores of *Puccinia striiformis* f. sp. *tritici* (*Pst*) by clustering pixel values, and isolate touching spores based on the shape and area factors (Lei et al., 2018). However, these classical machine vision methods are sensitive to noise and lack robustness, and usually cannot realize the segmentation of complex shapes.

Deep learning can learn how to extract the features from a large number of samples. Zhao et al. took anthrax spores as the research objects and applied CFL (Constrained Focal Loss) Loss function to DeeplabV3+. Experimentally, this proposed CFLNet\* can achieve better performance than original DeepLabv3+ (Zhao et al., 2019). Yang et al. (2020) proposed a Nuclear Segmentation Tool (NuSeT), which assimilates the advantages of semantic segmentation (U-Net) and instance segmentation (Mask R-CNN) and can work with both fluorescent and histopathology image samples. Xie et al. (2021) used Mask Scoring R-CNN network to detect mango disease spores to control and prevent mango disease. Li et al. (2023) proposed an MG-YOLO detection algorithm that introduces Multi-head self-attention in the YOLO backbone and optimizes the network neck and pyramid structure for fast and accurate gray mold spores detection, with a detection accuracy of 0.983 for the improved model and a time spent of 0.009 seconds per image. Zhang et al. (2023) introduced the attention mechanism module (ECA-Net) and adaptive feature fusion mechanism (ASFF) into the feature pyramid structure of YOLO to detect *Fusarium germinate* spores of small targets, and the average recognition accuracy of this model was 98.57%.

Image segmentation includes semantic and instance segmentation. The task of semantic segmentation is to classify each pixel in the image without separating the objects (Long et al., 2015), but this does not apply to our fungal segmentation task because there are a large number of *Phakopsora pachyrhizi* adhering or overlapping in the image, which can cause the touching fungus to not be segmented from each other and cause under-segmentation problems. Instance segmentation is a combination of the object detection and the semantic segmentation, where the object is detected in the image and then each pixel is labeled. Identifying *Phakopsora pachyrhizi* in an image is best viewed as an instance segmentation task (Hariharan et al., 2014). In this paper, we propose an improved Mask R-CNN spore segmentation method to solve these problems and improve the accuracy of spore segmentation. The main objectives include:

Optimization of backbone using Res2net block. by hierarchizing the residual connections in a single residual block, it is possible to achieve a multi-scale characterization of the fine-grained layers and, at the same time, increase the size of the sensory field at each level of the network. The use of Feature Pyramid Grids (FPGs) highlights the importance of deep pyramid representations by improving single-path feature pyramid networks by significantly improving

their performance at a similar computational cost. Use CIOU as a bounding box regression loss function to reduce the error.

A new method for fungal spore segmentation is proposed; extensive experiments show that this method achieves better segmentation performance under high density and overlapping conditions.

## 2 Materials and methods

In this section, we first summarize the problems and challenges faced in segmenting target images. Then, the segmentation model is designed and optimized, including the optimization of backbone and the optimization of mask branch.

### 2.1 Data source

In this paper, a spore image dataset was established to characterize the phenotypic transformation of fungi during *in vitro* growth in the presence of different fungicides. We used PerkinElmer's Opera QEHS high content rotary confocal system to observe fungal spores at different stages of growth and extract high quality images. The rotary disk confocal microscope can scan multi-channel fluorescence signals in a short time, and reduces the influence of detection environment on cells through extremely sensitive confocal imaging and synchronous acquisition.

Fresh leaves with rust organelles that had broken through the leaf epidermis and yellow rust spores were collected from the experimental field and brought back to the laboratory. Firstly, the surface of the diseased leaves was rinsed with running water, and then the leaf surface around the rust organelles was wiped with 75% ethanol, and then the diseased leaves were put into a petri dish with a wet filter paper at the bottom to keep humidity, and then fresh spores scattered around the rust organelles were collected after 1 d. Fresh spores collected were put into a 2 mL EP tube with appropriate amount of sterile water containing 0.3% Tween 80, and shaken well to make a spore suspension. The collected fresh rust spores were put into 2 mL EP tubes, and the spore suspension was made by adding appropriate amount of sterile water containing 0.3% Tween 80 and shaking well. 100 $\mu$ l of the sample was inoculated into the wells of a 96-well plate, and the spores from the 96-well plate were mixed in batches with six solvents: Carbendazim (1 ppm), DMSO (0.1%), PIK-75 (3.3 ppm), Solatenol (0.041 ppm), Solatenol (10 ppm) and TOU-951 (1.1 ppm). After 90 minutes, spores were stained using Calcofluor White solution with KOH and imaging of spores was recorded every 15 minutes one the Opera QEHS at a magnification of 10x to track spore growth status, with a total of 9 time-state data recorded. Hundreds of images of fungal spores at different stages of growth were collected under each time period for each chemical treatment.

Each image has a size of 685  $\times$  503 pixels and a pixel range of 8bit, saved in TIF format. We randomly selected 300 images as our dataset which fully contains the various morphology of fungal spores under the action of fungicides, as is show in Figure 2. (Due to the different efficacy of drugs acting on fungi, fungal spores vary greatly in phenotype, such as size, length, and number.) The

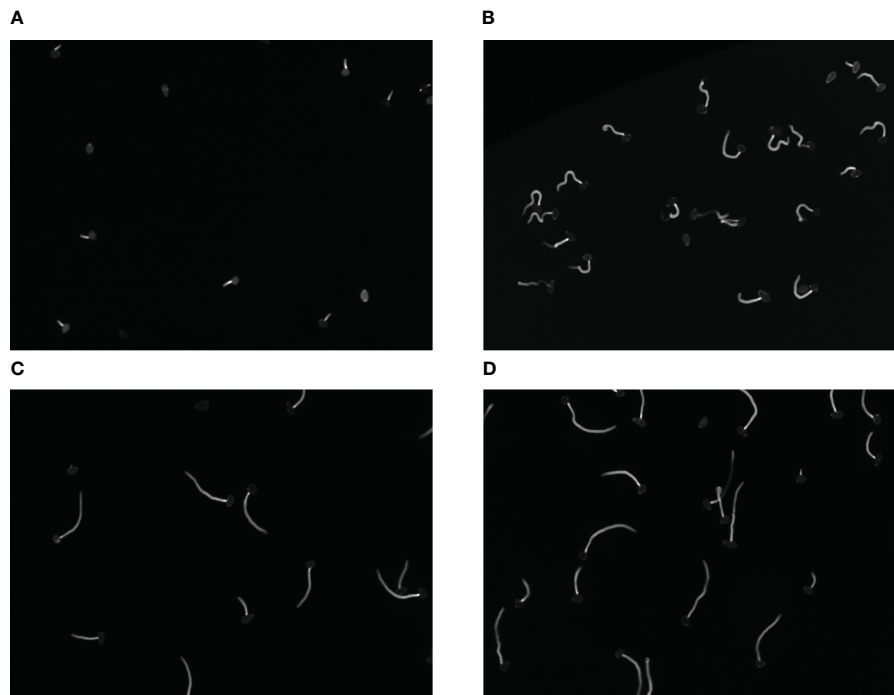


FIGURE 2  
(A–D) Respectively represent spores in different growth states.

diversity of fungi images collected improves the generalization ability and robustness of algorithm.

Each image in the dataset was manually annotated for network training by the open-source software Labelme (Russell et al., 2008) which labels the pixels of each class. Specifically, to instance segmentation, each single fungus segmentation served as an instance. As is shown in Figure 3.

The *Phakopsora pachyrhizi* dataset was divided into training and test sets in a ratio of 80:20, where 240 images were used as the training set (containing 5526 spores) and 60 images were used as the training set (containing 1621 spores). Thereafter, the labeled image instance information is stored in test set and training set json files respectively. The dataset is shown in Table 1.

## 2.2 Instance segmentation methods for spores

Microscopic image segmentation is an intricate task, with the target spores in *Phakopsora pachyrhizi* images encompass fungal spores with variable shapes and the same image mixed with multiple different growth states. An integral fungus consists of 2 parts, the germ tube and the spore. Due to the defects of light and the different sensitivity of different parts of the spore to light, the image of the stained spore has a low contrast, and it is difficult for the naked eye to detect its edges, which improves the difficulty of accurate segmentation. Additionally, growth phenotypes, such as fungal germination count, germ tube length, and growth direction,

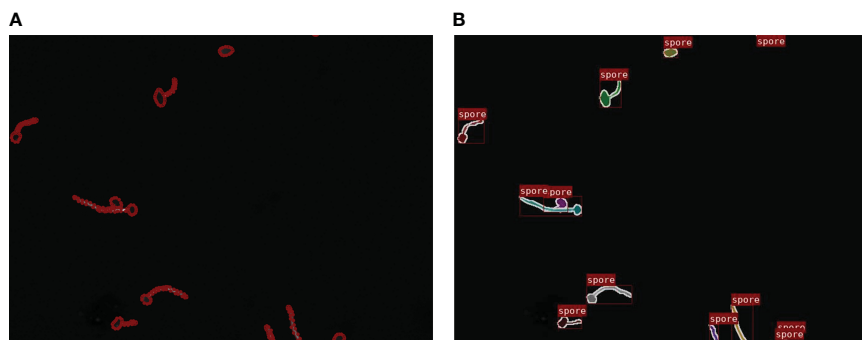


FIGURE 3  
Marking process: (A) Data annotation. (B) Visualization of the mask image.



exhibited significant variations under distinct medicinal treatments. Furthermore, fungi within the same image often appear densely populated, particularly during the later growth stages when exposed to certain solutions. Germ tubes tend to spread across a wide area, resulting in increased instances of crossing, overlapping, and clumping. These challenges have posed difficulties in achieving precise segmentation of *Phakopsora pachyrhizi*.

Image segmentation entails the meticulous classification of pixels into specific categories within an image. In contrast to semantic segmentation, instance segmentation not only segregates diverse objects within an image but also goes a step further by assigning a distinct classification to each individual pixel within the identified instances. This approach facilitates precise localization and differentiation of individual entities within the input image, leading to heightened accuracy in the process.

In the context of microscopic image spore segmentation, the application of instance segmentation techniques offers multifaceted benefits. Beyond effectively addressing challenges arising from spore intersections, overlaps, and adhesions, instance segmentation ensures the isolation of each spore as an autonomous entity, thereby averting any potential information ambiguities. Moreover, this approach excels at precisely determining the spatial coordinates of each spore, encompassing crucial details such as spore boundaries and internal structures. Such precision assumes paramount significance in comprehending the spatial arrangement and density distribution of spores within the image. Through instance segmentation, a nuanced understanding of the spore layout and distribution emerges, thereby enabling more informed analyses and interpretations. Instance segmentation offers the capability to establish distinct units of analysis for each spore, enabling segmentation at an individual level. This approach facilitates the quantification of various attributes like size, shape, color, and additional characteristics inherent to each spore. Consequently, this yields a more comprehensive dataset, enabling a deeper exploration into the intricate nuances of spore variation, interactions, and other pertinent traits. The detailed data acquired through instance segmentation serves as a foundation for conducting exhaustive investigations into the diverse aspects of spore behavior, facilitating enhanced insights and understanding.

Mask R-CNN is a classical top-down two-stage instance segmentation network, which can be considered as the extension of the Faster R-CNN architecture. This network builds upon the original network structure, incorporating additional branches to facilitate the prediction of segmentation masks for each ROI, all the while concurrently performing classification and bounding box regression. The process begins with the input image being fed through the backbone network, resulting in a feature map. This map is then utilized in the Region Proposal Network (RPN) to generate the corresponding anchor boxes. Subsequently, the feature maps linked to each anchor box are homogenized to a consistent size through RoIAlign, ensuring compatibility for further processing. Eventually, this standardized feature map is introduced to a fully connected layer, followed by anchor position refinement executed through regression layers, and class probabilities estimation performed through classification layers. The combination of these processes yields accurate instance

segmentation results, with the model generating both precise boundaries and segmentation masks for identified objects.

Researchers are constantly exploring new ways to combine Mask R-CNN with other techniques to improve the performance of segmentation tasks. For example, Seki and Toda (2022) utilized Mask R-CNN to segment lettuce seeds and extract their morphological parameters. Jia et al. (2022) Optimizing Mask R-CNN using the lightweight backbone network MobileNetv3 speeds up the model and meets the storage resource requirements of mobile robots. Chen et al. (2023) incorporated the attention mechanism into the backbone network of Mask R-CNN, which can better detect and segment the tapping area of natural rubber trees under different shooting conditions. Although Mask R-CNN has demonstrated excellent performance in the field of instance segmentation, it still faces great challenges when dealing with data such as spore images, which are characterized by a high degree of overlap and adhesion. In view of this, it is particularly crucial to develop an efficient segmentation strategy for spore image characteristics.

In order to enhancing the precision of spore segmentation, this research introduces an enhanced methodology for spore segmentation using Mask R-CNN. This approach integrates a variant of the Mask R-CNN architecture by incorporating Feature Pyramid Grids (FPG). The single-path feature pyramid network is improved by using FPG, where the feature scale space is represented as a regular lattice of parallel pathways and the pathways are fused together through multidirectional transversal links, which significantly improves the performance of the network with similar computational cost. The backbone network was optimized using an improved backbone, and using Res2Net module by layering residual connections in a single residual block allows for multiscale features at fine-grained layers while increasing the size of the perceptual field at each level of the network.

### 2.2.1 Feature pyramid grids

Feature Pyramid Grid (FPG) is an FPN-derived deep multi pathway network as shown in Figure 4. The feature scale space of this deep multi pathway feature pyramid network is a fusion of multidirectional lateral connections between parallel paths for information exchange at all levels to build a robust network with high discriminatory power and fine resolution across spatial dimensions. The single pyramid path back-propagates semantic information into the network by successively up-sampling the feature maps. FPG is a parallel extension of the single pyramid, which enriches the multidirectional (semantic) information in the scale space through the lateral connections between feature maps, allowing complex hierarchical features to be learned across scales. Lateral connection has 4 categories. Among them, AcrossSame is the fusion of features of the same level with those in the neighboring paths after using 1\*1 convolution. AcrossUp uses a convolution of 3\*3 stride of 2 to fuse the low-level features of the previous pathway with the high-level features of the next neighboring pathway. AcrossDown fuses the high-level features of the previous pathway to the low-level features of the next neighboring pathway by nearest interpolation convolution with a scaling factor of 2. AcrossSkip uses 1\*1 convolutional skip connections between same-level features.

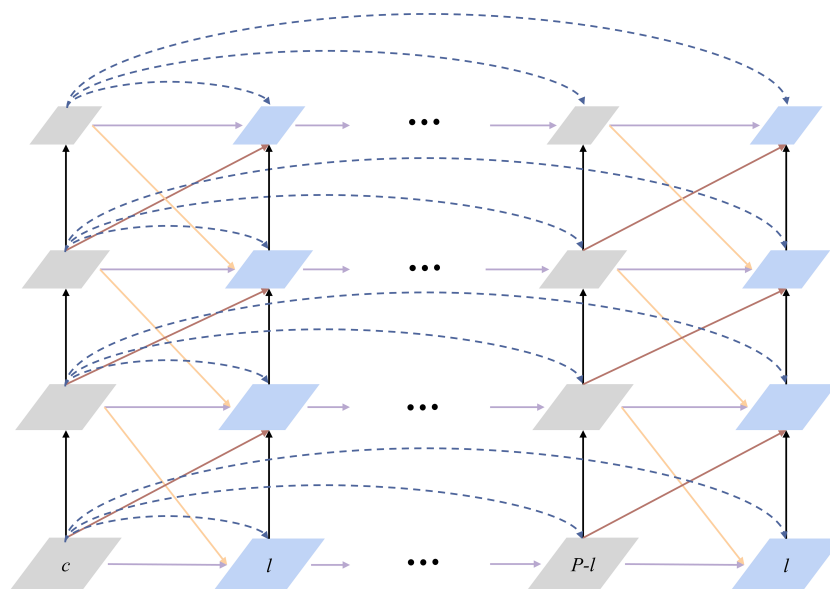


FIGURE 4  
A Feature Pyramid Grid (FPG).

Each convolutional block consists of a ReLU, a convolutional layer and a BatchNorm layer, and the fusion function uses element-wise Sum.

### 2.2.2 Composite backbone

The backbone network in instance segmentation is used to extract features from the input image and determines the feature representation capability of the model. For densely distributed fungi in the image, the large appearance of spores especially at the late stage of chemical treatments processing elevates the segmentation difficulty. In order to obtain better segmentation performance under high density and overlapping conditions, we optimize the backbone network based on FPN using Res2Net (Gao et al., 2019) fusion FPG as a composite backbone. Compared with ResNet, Res2Net adds small residual blocks to extract features with different receptive fields and multiple scales, so that the network can learn multiple features with different scales, in order to promote the communication of multi-scale features.

The Res2Net structure is shown in Figure 5. In the Res2net module, the input features are categorized into  $s$  subset, denoted as  $x_i, i \in \{1, 2, \dots, s\}$ ; the number of channels of the feature map in each group is  $1/s$  of the number of channels of the input feature map. Then, each set of feature maps undergoes a  $3 \times 3$  convolution (denoted as  $K_i$ ), except for  $x_1$ . Starting from  $x_3$ , the feature map  $x_i$  of the  $i$ th group is first summed with the  $K_{i-1}$  output of the previous group, and the result of the sum is subjected to the  $K_i$  operation. The whole process is represented in Formula (1).

$$y_i = \begin{cases} x_i & i = 1 \\ K_i(x_i) & i = 2 \\ K_i(x_i + y_{i-1}) & 2 < i \leq s \end{cases} \quad (1)$$

where  $y_i$  is the output of the module that is fed into the next convolutional layer.  $s$  is *scal*, which serves as the number of parameters controlling the dimensionality of the dimensions, and the larger  $s$  is, the better the multidimensional characterization. In

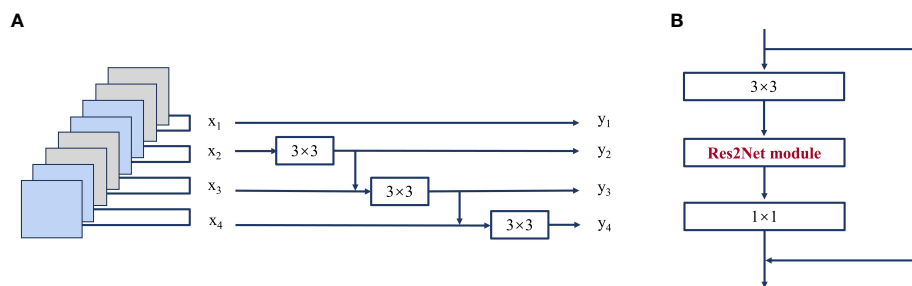


FIGURE 5  
(A) Res2Net module (scale=4); (B) Res2Net block.

this study, the s value size of 26 is used as the Res2Net block, which is used to modify the ResNet structure in the FPN.

Meanwhile, drawing on the idea of FPG, the improved backbone is combined with FPG to represent the feature scale space as a regular lattice of parallel pathways, and the pathways are fused together by multidirectional transverse links, improving the single-path feature pyramid network, which significantly improves the performance of the network with similar computational cost.

The design of Backbone is shown in Figure 6. The bottom-up C2, C3, C4 and C5 are Res2Net module layers, and the stride of each layer is 2. The number of channels of the structural layers output from C2 to C4 are adjusted using 1\*1 convolution, which produces P21, P31, P41 and P51, respectively, with P61 being derived from P51. To enhance computational efficiency, some simplifications were made in the lateral connections of FPG. Specifically, the AcrossSkip connection was removed, and a subset of the AcrossSkip, SameUp, and AcrossDown connections were omitted. We retained half of the triangular structure in the lateral connections, and in our experiments, we opted for P=9 paths to enrich the network's capabilities.

### 2.2.3 Loss function

In the whole improved network structure, the corresponding loss function consists of five parts, which are: the RPN's classification result prediction  $L_{R-cl_s}$ , its bounding box regression prediction  $L_{R-box}$ , alongside the final classification result prediction  $L_{cls}$ , final bounding box regression prediction  $L_{box}$  and the final mask image prediction  $L_{mask}$ . Loss function is calculated by Formula (2).

$$L = L_{R-cl_s} + L_{R-box} + L_{cls} + L_{box} + L_{mask} \quad (2)$$

Classification loss  $L_{cls}$  computes the loss of class probability using Cross Entropy.

$L_{mask}$  uses the Binary crossentropy loss function, calculated by the Formula (3).

$$L_{mask} = -\sum y \log(1 - \hat{y}) + (1 - y) \log(1 - \hat{y}) \quad (3)$$

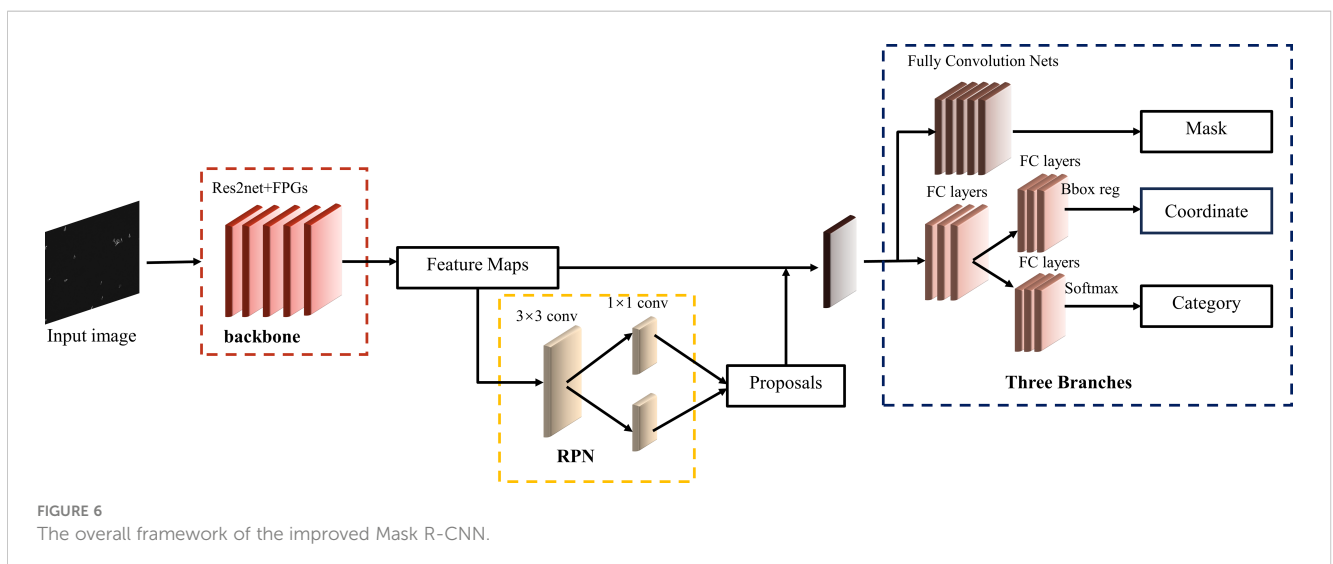
where  $y$  denotes the binarized ground truth,  $\hat{y}$  denotes the predicted segmentation result after binarization.

Edge information is very important for instance segmentation, and they can characterize the instance well. Mask R-CNN begins by utilizing the smooth $L_1$  function for calculating edge loss in target detection. Within this approach, losses for the four coordinate points are computed separately and aggregated to derive the ultimate edge loss. Despite assuming independence among the four points, there exists a certain degree of correlation among them in reality. The process of assessing box detection involves employing Intersection over Union (IoU), which differs from the regression coordinate box derived from the four points. Multiple detection boxes might yield identical smooth $L_1$  Loss values despite differing IoU values. To address this disparity, IoU Loss was introduced as a solution.

However, researchers and scholars soon found that IoU Loss has a drawback: when the prediction box does not intersect the target box, the loss function is not derivable. This problem makes the boundary information ignored in the prediction, and inaccurate edge detection occurs in the experiment, which affects the accuracy of segmentation. In order to meet the accurate segmentation of high-density spore images and improve the sensitivity of boundary segmentation, this paper optimizes the loss function and adopts the  $L_{CIoU}$  loss function as the loss function of the bounding box regression prediction, which accelerates the convergence speed of the model and makes the results of boundary segmentation more accurate.

The  $L_{CIoU}$  calculates the discrepancy between the predicted bounding box and the ground truth. Its definition is outlined in Formula (4).

$$\begin{cases} L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{e^2} + \frac{v^2}{(1 - IoU) + v} \\ IoU = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \end{cases} \quad (4)$$



Here,  $v$  signifies the alignment of the two frame aspect ratios, while  $b$  and  $b^{st}$  denote the center coordinates of the prediction and actual boxes respectively.  $\rho$  represents the Euclidean distance between their center points, indicative of the diagonal span of the smallest enclosed area containing both boxes. IoU stands for the Intersection over Union, representing the ratio between the shared area and the combined area of the predicted and actual bounding boxes.

In summary, the loss function used in this paper takes into account the error factor between the predicted value and the true value, which improves the convergence rate of the model, and the optimized network is more accurate in terms of error, and more flexible and feasible.

## 3 Results and discussion

### 3.1 Evaluation metrics

In this section, we present the key metrics used to measure the performance of spore instance segmentation. By leveraging these metrics, we can objectively analyze and showcase the strengths of our method, while also enabling a comprehensive comparison with existing instance segmentation techniques.

Average precision (AP) and average recall (AR) are the main evaluation metrics currently used in the field of object detection and instance segmentation. These metrics depend on two different segmentation masks: a ground truth segmentation mask labelled by experts and an output segmentation mask predicted by the network. Calculating AP and AR requires first calculating Precision, Recall, and IoU (Intersection over Union), as shown in [Formulas \(5–7\)](#).

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (6)$$

TP is true positive which means the number of correctly detect fungal areas, FP as false positive means the number of incorrectly detect fungi areas and FN is false negative which means the number of fungal areas incorrectly detected as background. Precision represents the proportion of TP in the predicted fungal areas and Recall means the proportion of TP in the true fungal areas.

$$IOU = \frac{\text{target} \cap \text{prediction}}{\text{target} \cup \text{prediction}} \times 100\% \quad (7)$$

IoU is the metrics to evaluate segmentation accuracy in one category, which calculate the intersection over union between predicted object and ground truth object.

Since the spore image is a small object, we select four metrics, AP, AP50, AP75, and AR, for network performance evaluation. Where AP is io ranging from 0.5 to 0.95 with a step rate of 0.05, AP50 is an iou threshold of 0.50, and AP75 is an iou threshold of 0.75. The higher these values are, the more desirable the instance segmentation model is ([Tong et al., 2020](#)).

AR calculates the average recall at different thresholds, how many real objects are correctly detected by the model. AR is the maximum recall of a given fixed number of detections per image, averaged over all IoU and all categories. Since there is only one type of spore, the category is 1. In this study, AR was calculated and averaged over 10 IoU thresholds between 0.5 and 0.95, as shown in [Formula \(8\)](#).

$$P_{\text{interp}}(r) = \max_{\tilde{r} \geq r} p(\tilde{r}) \quad (8)$$

where  $p(\tilde{r})$  is the measured precision at recall  $\tilde{r}$ .

AP is calculated by the [Formula \(9\)](#).

$$AP = \frac{1}{11} \sum_{r \in \{0.0, 0.1, \dots, 1\}} P_{\text{interp}}(r) \quad (9)$$

### 3.2 Implementation details

The computational environment for this study utilizes Python 3.7.3 and Ubuntu 18.04 LTS, employing Jupyter Notebook as the editor. The integrated model outlined above was constructed on an Intel(R) Core (TM) i7-12700H (20 CPU) with a 2.30GHz processor, 16 GB of DDR4 RAM, and three graphics cards: two discrete graphics cards (NVIDIA GeForce RTX 3060 laptop GPU with 6023 MB, NVIDIA GeForce RTX 3080 Ti with 12108 MB) along with one integrated graphics card (Intel(R) Iris(R) Xe graphics card with 128 MB), which were utilized for training and testing.

The instance segmentation model was trained with stochastic gradient descent (SGD) method, batch size was set to 4, momentum factor was 0.9, the initial learning rate was 0.08, and for each epoch, the learning rate changed to 0.9 times of the previous one. The total number of epochs for model training is 100, and when training for the first 60 epochs, the pre-feature extraction network is frozen, and only the neck network and the detection head network are trained in order to improve the training speed of the network model.

TABLE 1 Number of images and spores in the training and test sets in the dataset under different solution treatments, respectively.

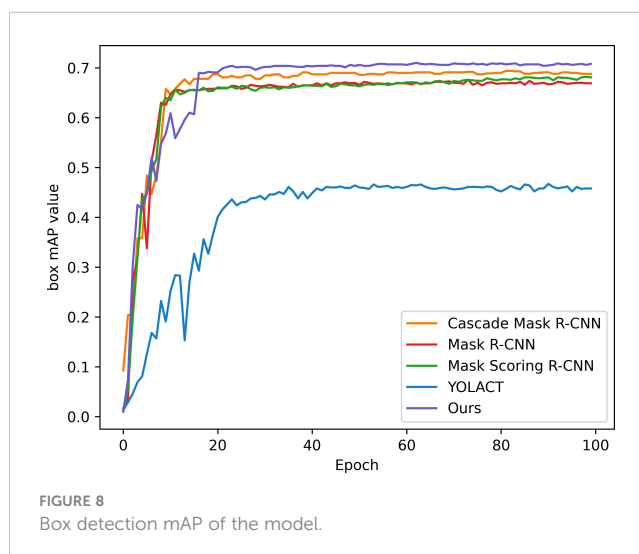
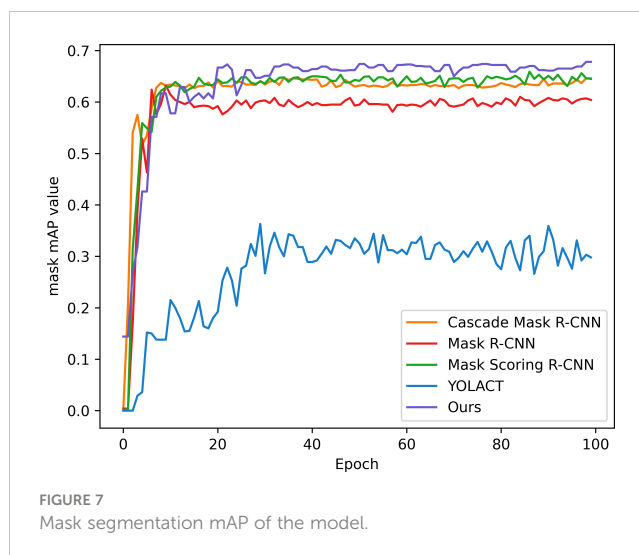
Dataset		Carbendazim	DMSO	Solatenol (0.041 ppm)	Solatenol (10 ppm)	TOU-951
Training set	Images	240	48	48	48	48
	Spore instance	5526	1023	1130	970	1112
Test set	Images	60	12	12	12	12
	Spore instance	1621	306	312	289	302

### 3.3 Performance comparison with state-of-the-art methods and visualization analysis

In order to verify the effectiveness and accuracy of the model in spore instance segmentation, we use the same spore dataset under the same training environment and experimental configuration, and analyze them in comparison with Mask R-CNN, Mask Scoring R-CNN (Huang et al., 2019), YOLACT (Bolya et al., 2019) and Cascade Mask R-CNN (Cai and Vasconcelos, 2019) models. The parameters introduced in 3.2 are used as evaluation indexes to compare the performance with several other methods, and the experimental results are shown in Table 2. All algorithms are trained for 100 epochs, and after each epoch is completed, the mAP values for mask segmentation and box detection are calculated, as shown in Figures 7 and 8.

In this application, detection accuracy refers to the detection of spore individuals from complex environments, detection accuracy refers to the model correctly identifying and localizing spore instances in complex environments, and segmentation accuracy is concerned with the model correctly segmenting each spore at the pixel level, and segmentation accuracy is as important as detection accuracy. As can be seen from the table, our improved algorithm outperforms these classical instance segmentation algorithms in both segmentation accuracy and detection accuracy. From the parameter comparison, the detection accuracy of the model is 0.712 and the segmentation accuracy is 0.618, which are 3.5% and 5% better than the existing optimal methods, respectively. As can be seen in Figures 7 and 8, when the training epoch is less than 20, the advantage of the method is not obvious. However, as the training epoch increases, the detection mAP of our proposed method clearly outperforms these state-of-the-art methods. Our model is more powerful because we not only optimize the backbone network to efficiently extract global and local features; we also introduce a deep multipath feature pyramid network to construct fine-resolution features with strong semantic information; all of these improvements greatly improve the robustness of the CNN to geometric transformations of the target. Our model is able to explore the complex nuances of spore variants, interactions and other related features in greater depth, understand the spatial arrangement and density distribution of spores in an image, effectively resolve spore crossings, overlaps and adhesions, and ensure that each spore is separated as an independent entity.

In order to highlight the superiority of the proposed architecture more intuitively, a visual comparative analysis



between the current networks and ours is carried out. As shown in Figures 9A–C, in the figure, are three different growth patterns of spores, where I is the original spore images, and II to VI are Cascade Mask R-CNN, Mask R-CNN, Mask Scoring R-CNN, YOLACT, and ours, respectively.

As can be seen from the confidence level of the anchor box and the box in the figure, the example segmentation results of these advanced networks for a single scattered distribution of spore

TABLE 2 Evaluation results of serious models.

Models	box_mAP	box_mAP_50	box_mAP_75	mask_mAP	mask_mAP_50	mask_mAP_75
Cascade Mask R-CNN(ResNet-101)	0.688	0.911	0.802	0.646	0.796	0.696
Mask R-CNN(ResNet-101)	0.669	0.938	0.754	0.604	0.757	0.619
Mask Scoring R-CNN(ResNet-101)	0.681	0.929	0.787	0.645	0.785	0.695
YOLACT(ResNet-101)	0.458	0.836	0.49	0.298	0.442	0.036
Ours	0.712	0.936	0.809	0.678	0.828	0.724



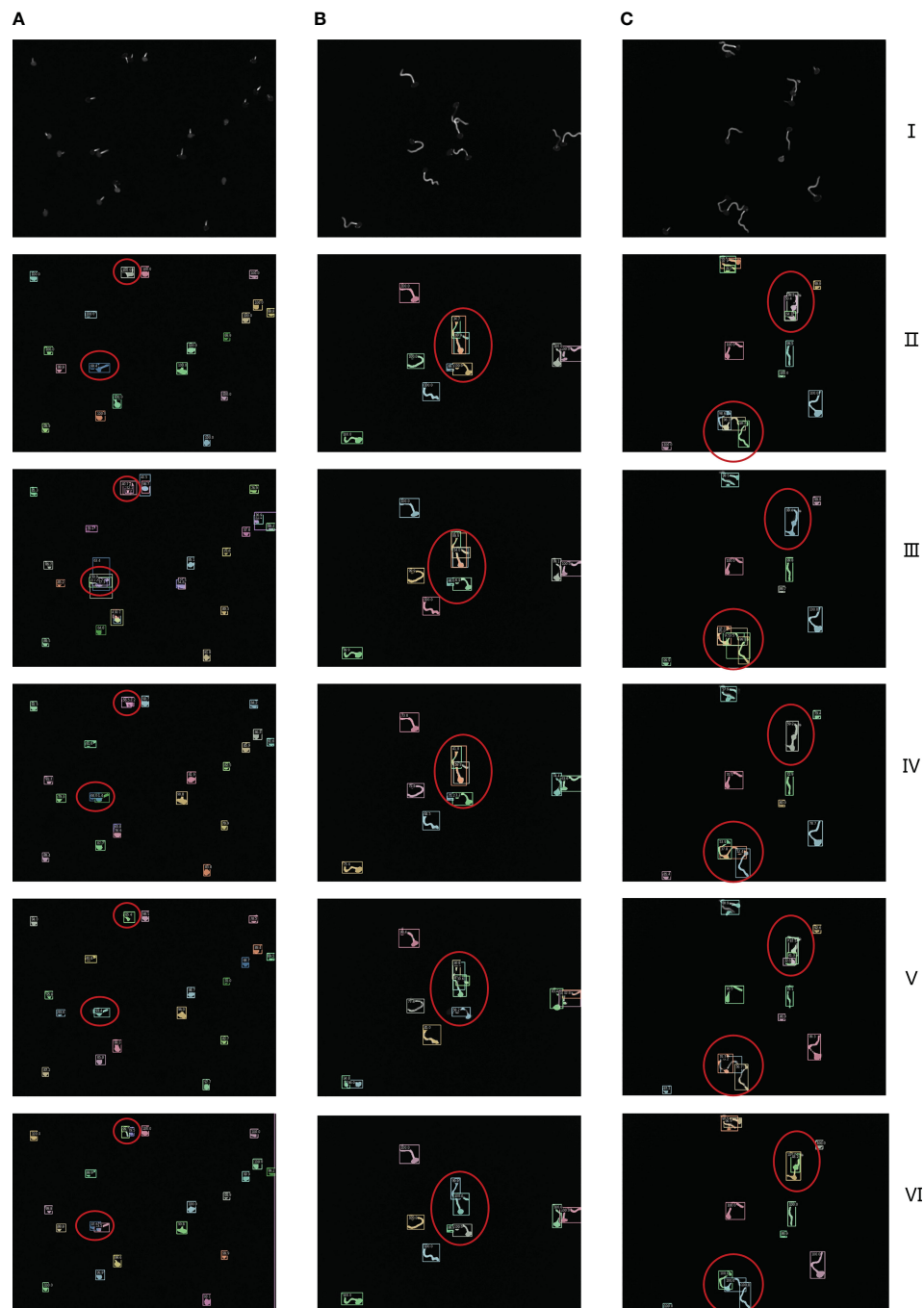


FIGURE 9

I is the original image of spores, II to VI are Cascade Mask R-CNN, Mask R-CNN, Mask Scoring R-CNN, YOLACT, ours network model visualization and analysis images respectively. (A–C) are the segmentation results of three randomly selected spore images under different networks.

images are more general. As can be seen from the red circles in the figure, under-segmentation and over-segmentation occur for spore cross, adhesion and overlapping part segmentation with low confidence and lack of refinement and edge processing. In sharp contrast, our network model generated finer detection segmentation images. To further demonstrate the visual analysis results of this network, we performed a zoom-in comparison, as shown in Figure 10.

We chose two images with sparse and tight spore distribution, and both images contain cross overlapping spore features. As can be seen from the figure, the model uses CIoU loss optimization to obtain the optimal prediction box, so that the box detection part can quickly and accurately find the differences between spore individuals with high confidence. At the same time, the model effectively solves the problem of poor robustness of fuzzy pixel segmentation, and the mask segmentation part not only refines the

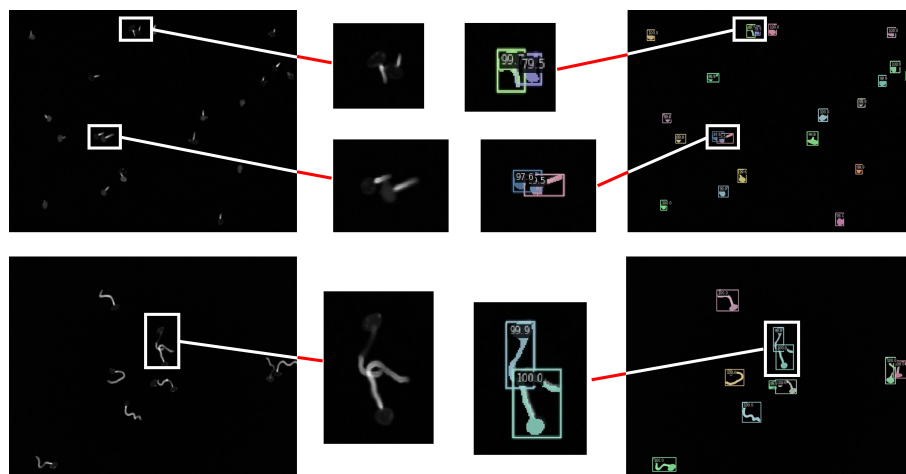


FIGURE 10  
Visualization results of the target network.

overall segmentation, but also greatly improves the edge segmentation accuracy, so that the cross-overlapping spore images are separated as separate individuals.

### 3.4 Ablation study of improved models

Based on the improved Mask R-CNN model proposed in this paper, ablation experiments were conducted to compare the experimental results of the original model with the improved ResNet-101, ResNet-101+FPG, Res2Net-101+FPG, and Res2Net+FPGs, as shown in Table 3. Table 3 shows the experimental results of different backbone networks: after adding FPG to the original network, the detection accuracy of the model decreases significantly and the instance segmentation accuracy decreases, but the segmentation accuracy is greatly improved; after replacing ResNet-101, the detection accuracy, segmentation accuracy and accuracy are significantly improved and exceed the original model, which indicates that each module of the improvement has a positive effect on instance segmentation. Finally, the best performing model was found to come from the combined effect of the two improved modules, which improved the detection accuracy, segmentation accuracy, and instance segmentation accuracy by 6.4%, 12.3%, and 2.2%, respectively, compared to the Mask R-CNN model, proving that these improvement strategies of the model are effective.

TABLE 3 Evaluation results of ablation experiments.

Models	box_mAP	mask_mAP	acc	#param.	GFLOPs
Original Mask R-CNN(ResNet-101)	0.669	0.604	95.99	63.388M	302
Mask R-CNN(ResNet-101+FPG)	0.621	0.67	94.72	63.274M	147
Mask R-CNN(Res2Net-101+FPG)	0.69	0.678	96.77	63.951M	152
Ours	0.712	0.678	98.144	63.951M	154

## 4 Conclusion

In this study, an improved Mask R-CNN method is proposed to accomplish the task of segmentation of densely distributed and overlapping crossed *Phakopsora pachyrhizi* micro-images. The method is optimized and improved from the original MaskR-CNN. Firstly, the res2net was used to replace the backbone network of the original Mask R-CNN by layering the residual connections in a single residual block and then combining it with FPG in order to improve the fine resolution and high-resolution capability, strengthen the feature extraction capability of the network model, and enhance the detection accuracy. Secondly, for the problem of inaccurate edge detection of the original model, the loss function is optimized, and the CIoU loss function is adopted as the loss function of the boundary box regression prediction, which accelerates the convergence speed of the model, meets the accurate segmentation of high-density spore images, and improves the sensitivity of boundary segmentation. Compared with the original model, it is more robust and further improves the accuracy of instance segmentation. In summary, the proposed model can better detect and segment spores under various conditions.

However, this study suffers from an insufficient number of samples in the dataset, and the accuracy of detection and segmentation in the case of spore stacking with a large number of anomalies needs to be further improved. In the follow-up work,

collecting and labeling more spore clusters with complex shapes should be considered to expand the spore dataset under different overlap types. Meanwhile, in order to improve the performance of the network, the effect of spore morphology such as length, width and area on segmentation can be deeply investigated, and its features can be fused with image information and input into the segmentation network. Finally, in future research, the scheme proposed in this paper needs to be installed and applied in real scenarios to validate the performance of the model and algorithm. The technique can be applied to perform automatic segmentation of images on microscopes to facilitate the discovery of new drug candidates and the discovery of therapeutic options. Simultaneously, it offers valuable insights to the fields of agriculture, ecology, and medicine, enhancing our understanding and management of fungal-related issues, including disease transmission and ecological balance.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

YW: Validation, Writing – review & editing. ZX: Writing – original draft. FL: Data curation, Writing – review & editing. WH: Investigation, Writing – review & editing. HF: Writing – review & editing. QZ: Supervision, Writing – review & editing.

## References

- Beucher, S. (1992). The watershed transformation applied to image segmentation. *Scanning microscopy* 1992, 28.
- Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). “Yolact: Real-time instance segmentation,” in *Proceedings of the IEEE/CVF international conference on computer vision*. 9157–9166. doi: 10.1109/ICCV.2019.00925
- Cabre, L., Peyrard, S., Sirven, C., Gilles, L., Pelissier, B., Ducerf, S., et al. (2021). Identification and characterization of a new soybean promoter induced by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *BMC Biotechnol.* 21, pp.1–pp13. doi: 10.1186/s12896-021-00684-9
- Cai, Z., and Vasconcelos, N. (2019). Cascade R-CNN: High quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, pp.1483–1498. doi: 10.1109/TPAMI.2019.2956516
- Calderone, R., Sun, N., Gay-Andrieu, F., Groutas, W., Weerawarna, P., Prasad, S., et al. (2014). Antifungal drug discovery: the process and outcomes. *Future Microbiol.* 9, 791–805. doi: 10.2217/fmb.14.32
- Carolus, H., Pierson, S., Lagrou, K., and Van Dijck, P. (2020). Amphotericin B and other polyenes—Discovery, clinical use, mode of action and drug resistance. *J. Fungi* 6, 321. doi: 10.3390/jof6040321
- Chen, Y., Zhang, H., Liu, J., Zhang, Z., and Zhang, X. (2023). Tapped area detection and new tapping line location for natural rubber trees based on improved mask region convolutional neural network. *Front. Plant Sci.* 13, 1038000. doi: 10.3389/fpls.2022.1038000
- Coleman, G. B., and Andrews, H. C. (1979). Image segmentation by clustering. *Proc. IEEE* 67, 773–785. doi: 10.1109/PROC.1979.11327
- Gao, S. H., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., and Torr, P. (2019). Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 652–662. doi: 10.1109/TPAMI.34
- Goellner, K., Loehrer, M., Langenbach, C., Conrath, U. W. E., Koch, E., and Schaffrath, U. (2010). *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Mol. Plant Pathol.* 11, 169–177. doi: 10.1111/j.1364-3703.2009.00589.x
- Hariharan, B., Arbeláez, P., Girshick, R., and Malik, J. (2014). “Simultaneous detection and segmentation,” in *Computer Vision—ECCV 2014* (Cham: Springer International Publishing), 297–312. Available at: [http://dx.doi.org/10.1007/978-3-319-10584-0\\_20](http://dx.doi.org/10.1007/978-3-319-10584-0_20)
- Hartman, G. L., Chang, H. X., and Leandro, L. F. (2015). Research advances and management of soybean sudden death syndrome. *Crop Prot.* 73, 60–66. doi: 10.1016/j.cropro.2015.01.017
- Huang, Z., Huang, L., Gong, Y., Huang, C., and Wang, X. (2019). “Mask scoring r-cnn” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 6409–6418. doi: 10.1109/CVPR.2019.00657
- Jia, W., Wei, J., Zhang, Q., Pan, N., Niu, Y., Yin, X., et al. (2022). Accurate segmentation of green fruit based on optimized mask RCNN application in complex orchard. *Front. Plant Sci.* 13, 955256. doi: 10.3389/fpls.2022.955256
- Korsnes, R., Westrum, K., Fløistad, E., and Klingen, I. (2016). Computer-assisted image processing to detect spores from the fungus *Pandora neophidii*. *MethodsX* 3, 231–241. doi: 10.1016/j.mex.2016.03.011
- Lei, Y., Yao, Z., and He, D. (2018). Automatic detection and counting of urediniospores of *Puccinia striiformis* f. sp. *tritici* using spore traps and image processing. *Sci. Rep.* 8, 13647. doi: 10.1038/s41598-018-31899-0
- Li, K., Zhu, X., Qiao, C., Zhang, L., Gao, W., and Wang, Y. (2023). The gray mold spore detection of cucumber based on microscopic image and deep learning. *Plant Phenomics* 5, 0011. doi: 10.34133/plantphenomics.0011

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Grant No. 82272534), Shenzhen Key Laboratory of Bone Tissue Repair and Translational Research (NO. ZDSYS20230626091402006).

## Acknowledgments

This work was technically supervised by Professor Fu-Xin Wei of the Seventh Affiliated Hospital of Sun Yat-sen University in the field of cells.

## Conflict of interest

Author HF was employed by the company Zhongzhen Kejian ShenZhen Holdings Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Liberali, P., Snijder, B., and Pelkmans, L. (2015). Single-cell and multivariate approaches in genetic perturbation screens. *Nat. Rev. Genet.* 16, 18–32. doi: 10.1038/nrg3768
- Loehrer, M., and Schaffrath, U. (2011). Asian soybean rust—meet a prominent challenge in soybean cultivation. *Soybean-biochemistry Chem. Physiol. Rijeka Croatia: InTech*, 83–100. doi: 10.5772/15651
- Long, J., Shelhamer, E., and Darrell, T. (2015). “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440. doi: 10.1109/cvpr.2015.7298965
- Lorrain, Cécile, Gonçalves dos Santos, K. C., Germain, H., Hecker, A., and Duplessis, Sébastien (2019). Advances in understanding obligate biotrophy in rust fungi. *New Phytol.* 222, 1190–1206. doi: 10.1111/nph.15641
- Melo, C. A. D., Lopes, J. G., Andrade, A. O., Trindade, R. M., and Magalhaes, R. S. (2019). Semi-automated counting model for arbuscular mycorrhizal fungi spores using the Circle Hough Transform and an artificial neural network. *Anais da Academia Bras. Ciências* 91, e20180165. doi: 10.1590/0001-3765201920180165
- Okada, H., Ohnuki, S., Roncero, C., Konopka, J. B., and Ohya, Y. (2014). Distinct roles of cell wall biogenesis in yeast morphogenesis as revealed by multivariate analysis of high-dimensional morphometric data. *Mol. Biol. Cell* 25, 222–233. doi: 10.1091/mbc.e13-07-0396
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. systems man cybernetics* 9, 62–66. doi: 10.1109/TSMC.1979.4310076
- Reisen, F., De Chalon, A. S., Pfeifer, M., Zhang, X., Gabriel, D., and Selzer, P. (2015). Linking phenotypes and modes of action through high-content screen fingerprints. *Assay Drug Dev. Technol.* 13, 415–427. doi: 10.1089/adt.2015.656
- Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vision* 77, 157–173. doi: 10.1007/s11263-007-0090-8
- Saito, H., Yamashita, Y., Sakata, N., Ishiga, T., Shiraishi, N., Usuki, G., et al. (2021). Covering soybean leaves with cellulose nanofiber changes leaf surface hydrophobicity and confers resistance against *Phakopsora pachyrhizi*. *Front. Plant Sci.* 111, 146–147. doi: 10.3389/fpls.2021.726565
- Seki, K., and Toda, Y. (2022). QTL mapping for seed morphology using the instance segmentation neural network in *Lactuca* spp. *Front. Plant Sci.* 13, 949470. doi: 10.3389/fpls.2022.949470
- Tong, K., Wu, Y., and Zhou, F. (2020). Recent advances in small object detection based on deep learning: A review. *Image Vision Computing* 97, 103910. doi: 10.1016/j.imavis.2020.103910
- Usaj, M. M., Styles, E. B., Verster, A. J., Friesen, H., Boone, C., and Andrews, B. J. (2016). High-content screening for quantitative cell biology. *Trends Cell Biol.* 26, 598–611. doi: 10.1016/j.tcb.2016.03.008
- Xie, X., Wang, J., Hu, Z., and Zhao, Y. (2021). “Intelligent detection of mango disease spores based on mask scoring R-CNN,” in *2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT)*. 768–774 (IEEE), October. Available at: <http://dx.doi.org/10.1109/acait53529.2021.9731325>
- Yang, L., Ghosh, R. P., Franklin, J. M., Chen, S., You, C., Narayan, R. R., et al. (2020). NuSeT: A deep learning tool for reliably separating and analyzing crowded cells. *PLoS Comput. Biol.* 16, e1008193. doi: 10.1371/journal.pcbi.1008193
- Zanella, F., Lorens, J. B., and Link, W. (2010). High content screening: seeing is believing. *Trends Biotechnol.* 28, 237–245. doi: 10.1016/j.tibtech.2010.02.005
- Zhang, D. Y., Zhang, W., Cheng, T., Zhou, X. G., Yan, Z., Wu, Y., et al. (2023). Detection of wheat scab fungus spores utilizing the Yolov5-ECA-ASFF network structure. *Comput. Electron. Agric.* 210, 107953. doi: 10.1016/j.compag.2023.107953
- Zhao, Y., Lin, F., Liu, S., Hu, Z., Li, H., and Bai, Y. (2019). Constrained-focal-loss based deep learning for segmentation of spores. *IEEE Access* 7, 165029–165038. doi: 10.1109/Access.6287639
- Zhao, Y., Liu, S., Hu, Z., Bai, Y., Shen, C., and Shi, X. (2020). Separate degree based Otsu and signed similarity driven level set for segmenting and counting anthrax spores. *Comput. Electron. Agric.* 169, 105230. doi: 10.1016/j.compag.2020.105230