RESEARCH

Cucumber pathogenic spores' detection using the GCS-YOLOv8 network with microscopic images in natural scenes

Xinyi Zhu[1], Feifei Chen[1], Chen Qiao[1], Yiding Zhang[2*], Lingxian Zhang[1,3*], Wei Gao[4] and Yong Wang[4]

Abstract

Fungal diseases are the main factors affecting the quality and production of vegetables. Rapid and accurate detection of pathogenic spores is of great practical significance for early prediction and prevention of diseases. However, there are some problems with microscopic images collected in the natural environment, such as complex backgrounds, more disturbing materials, small size of spores, and various forms. Therefore, this study proposed an improved detection method of GCS-YOLOv8 (Global context and CARFAE and Small detector-optimized YOLOv8), effectively improving the detection accuracy of small-target pathogen spores in natural scenes. Firstly, by adding a small target detection layer in the network, the network's sensitivity to small targets is enhanced, and the problem of low detection accuracy of the small target is effectively improved. Secondly, Global Context attention is introduced in Backbone to optimize the CSPDarknet53 to 2-Stage FPN (C2F) module and model global context information. At the same time, the feature up-sampling module Content-Aware Reassembly of Features (CARAFE) was introduced into Neck to enhance the ability of the network to extract spore features in natural scenes further. Finally, we used an Explainable Artificial Intelligence (XAI) approach to interpret the model's predictions. The experimental results showed that the improved GCS-YOLOv8 model could detect the spores of the three fungi with an accuracy of 0.926 and a model size of 22.8 MB, which was significantly superior to the existing model and showed good robustness under different brightness conditions. The test on the microscopic images of the infection structure of cucumber down mildew also proved that the model had good generalization. Therefore, this study realized the accurate detection of pathogen spores in natural scenes and provided feasible technical support for early predicting and preventing fungal diseases.

Keywords Spore detection, In field, Small object detection, Global context attention, Feature upsampling

*Correspondence:
Yiding Zhang
zhangyiding117@163.com
Lingxian Zhang
zhanglx@cau.edu.cn
[1]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[2]College of Engineering, China Agricultural University, No.17 Qinghua East Road, Haidian District, Beijing 100083, China
[3]Key Laboratory of Agricultural Informationization Standardization, Ministry of Agriculture and Rural Affairs, Beijing 100083, China
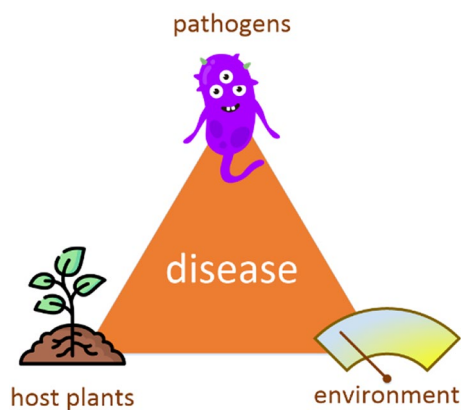[4]Tianjin Academy of Agricultural Sciences, Tianjin 300384, China

Background

Cucumber is one of the world's most widely cultivated vegetable crops [1], which people love because of its rich nutrition and considerable economic benefits [2]. However, the occurrence of the disease will lead to reduced cucumber yield and food safety problems [3, 4]. Pathogens, host plants, and environmental conditions are the three essential elements of plant disease. When the environmental condition is advantageous to the growth of pathogens, they will infect the host plant pathogens, causing disease [5], as shown in Fig. 1. Among them, fungi are the most crucial pathogenic microorganism, and fungal diseases are the most in all diseases, such as powdery mildew, fusarium wilt, and gray mold [6]. Therefore, preventing and controlling fungal diseases is very important for improving vegetable quality and reducing economic losses [7, 8].

The process from the invasion of fungal pathogens into host plants to the appearance of disease can be divided into four periods: contact period, invasion period, latent period, and onset period [9]. In the contact period, invasion period, and latent period, the pathogen was the pathogen conidium. The pathogen conidium had contact with the crop at this time, but no disease appeared on the crop surface. When in the onset period, the surface of the crop will appear green, with discoloration and other symptoms [10]. Therefore, compared with the diagnosis and prediction of disease conditions through observation of disease sites during the onset period [11–13], direct detection of pathogen spores can move the prevention and control threshold forward to achieve early accurate prediction and risk assessment of diseases [14], thus effectively reducing losses. Therefore, rapid and accurate detection of fungal spores is needed.

Traditional spore identification mainly relies on visual inspection and manual counting. Although this method is simple to operate, it is time-consuming and subjective [15–17]. With the development of artificial intelligence technology, researchers began to use computer-aided technology to detect fungal spores to achieve early diagnosis of diseases [18]. Among them, the detection method based on machine learning needs first to pre-process the image, such as graying and de-noising, and combine the segmentation method to obtain the target image, and then manually extract the features such as size, shape, and texture based on the characteristics of the spores to identify the spores [19–21]. This method can accurately extract the features of specific spores in a relatively simple scene, but because the features extracted by hand are for particular species, they are not universal.

Moreover, when different spores have similar size and morphological characteristics, and when the spores in the actual scene have morphological diversity due to different growth stages, the characteristics of the spores are not universal. The accuracy of spore detection using machine learning methods decreases. In addition, using machine learning methods to detect spores in natural scenarios accurately has also proved challenging [22]. Compared with machine learning methods, deep learning has strong learning ability and can automatically learn the high-dimensional features of the target with better adaptability [10, 23, 24]. At present, the mainstream deep learning detection methods mainly include semantic segmentation and object detection. The method based on semantic segmentation is to make pixel-level predictions by separating the target spores from the background image. This method can detect pathogen spores in natural scenes, but the detection accuracy will be reduced when there is spore adhesion in the image. In addition, the practicability of this method needs to be improved due to the time-consuming and laborious process of classifying the spore images at the pixel level [25, 26]. The method based on target detection can realize accurate location and recognition of pathogen spores. It only needs to mark the target spores in a rectangular box, without pixel-level labeling, so it has been widely used in fungal spore detection. For example, Zhang et al. proposed a YOLOv5 network based on ECA attention and the adaptive feature fusion mechanism ASFF, which enhanced the network's ability to extract features of different dimensions of spores and realized high-precision detection of wheat scab spores [23]. Li et al. constructed an MG-YOLO model to rapidly detect fungus spores [10], MHSA enhanced the processing of global image information, and a BiFPN-enhanced network was introduced to extract features of spores of different scales. A lightweight network was adopted to optimize the Neck, which maintained high precision and reduced the model size. However, most of the spore images collected in the current spore research are based on the fungi cultured in the laboratory environment. The background of the images is relatively simple; there are few disturbing materials, and only one kind of pathogen spores are targeted.

Fig. 1 Disease triangle schematic diagram

The detection accuracy rate in natural scenes and the generalization ability of other pathogen spores detection must also be verified.

Considering that pathogen spores are small targets with diverse forms, and there are many interferors and adhesions such as mycelia, leaf tissue, impurities, and bubbles in the actual images collected in the field, it brings a series of challenges to the detection of pathogen spores small targets. On this basis, this study proposes an improved identification method of GCS-YOLOv8, which combines the small target detection layer, the feature up-sampling module CARAFE, and the Global Context attention module based on the YOLOv8s benchmark model, effectively combining shallow features with deep features. The global context information was modeled to improve the feature extraction ability of the network and realize the accurate and rapid identification of cucumber pathogen spores. The main contributions of this study are as follows:

1. A small target fungal spore identification method, GCS-YOLOv8, was proposed in the natural scene, which was interpreted by an XAI method, its robustness under dim and light conditions was tested. Transfer learning was performed on microscopic images of cucumber downy mildew infection structure and its generalization was tested;
2. Add a small target detection layer to the network and detect the shallow and deep features of the network after fusion to enhance the ability of the network to extract small target spore features;
3. Use the Global Context attention mechanism to optimize the C2F module and model long-term dependency by capturing global context information;
4. CARAFE, a content-sensing feature recombination module, was used in Neck to conduct feature up-sampling and aggregate feature information in a large receptive field, effectively improving the

accuracy of spore detection based on lightweight and speed.

Materials and methods
Image acquisition and annotation
In this study, we built the microscopic image dataset obtained from the Plant Protection Research Institute of Tianjin Academy of Agricultural Sciences. After collecting several cucumber powdery mildew leaves, gray mold leaves, and wilt leaves in the field, take a slide, drop 1–2 drops of sterile water on the slide, scrape the mycelia spores from the mycelia parts of the sick leaves with a fungus-scraping stick and observe them under a microscope.

This study collected microscopic images of spores under different light sources. The microscope model used for image collection was the Olympus microscope BX51. Eight fields of view were observed on each slide, the magnification was $200\times$, and the resolution of the collected images was 2748×2220 pixels. A total of 484 images with 31,629 spores of powdery mildew, gray mold, and fusarium wilt were collected. The image samples of three fungal spores are shown in Fig. 2.

Under the guidance of experts in this field, the image annotation software LabelImg was used to label the spores of three kinds of fungus in rectangular boxes. The annotation file was saved in the ".xml " format of the PASCAL Visual Object Class (VOC) dataset. In order to make the model play the best performance, we use the smallest rectangular frame outside the spore as the label frame to mark artificially, so as to reduce the interference of complex background and make the model extract the characteristic information of the spore more accurately. For the spores at the edge of the image, it is judged that the spore area accounts for more than half of the overall area. Overlapping and adherent spores are labeled as long as they can be distinguished by the naked eye. The annotated dataset contained 484 images and 31,629 data
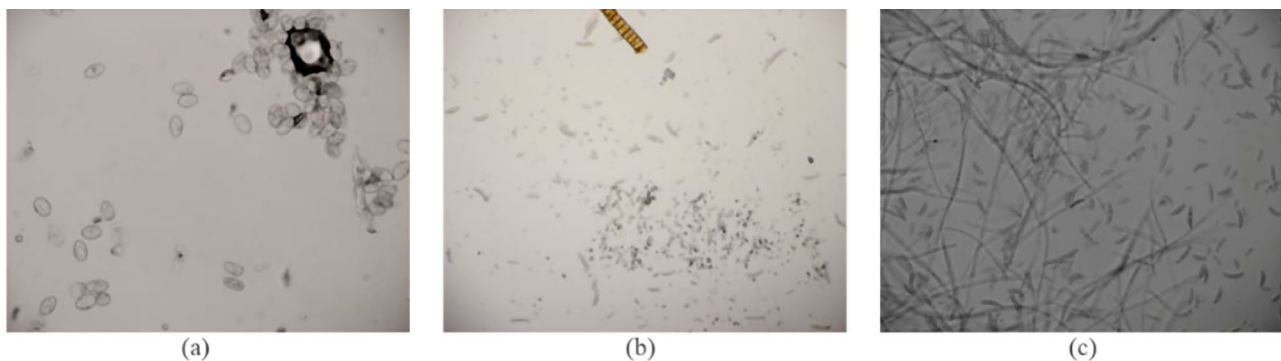
Fig. 2 Examples of microscopic images of pathogen spores（where (**a**) are examples of microscopic images of Powdery mildew spores, and (**b**) and (**c**) are examples of microscopic images of Fusarium wilm spores

Table 1 Details of images and spore samples in the dataset

Dataset	Images	Spores	Powdery mildew		Gray mold		Fusarium wilt	
			Images	Spores	Images	Spores	Images	Spores
Training set	338	22,258	110	2791	114	8468	114	10,999
Validation set	72	4742	22	656	27	1943	23	2143
Test set	74	4629	31	806	20	1602	23	2221
Total	484	31,629	163	4253	161	12,013	160	15,363
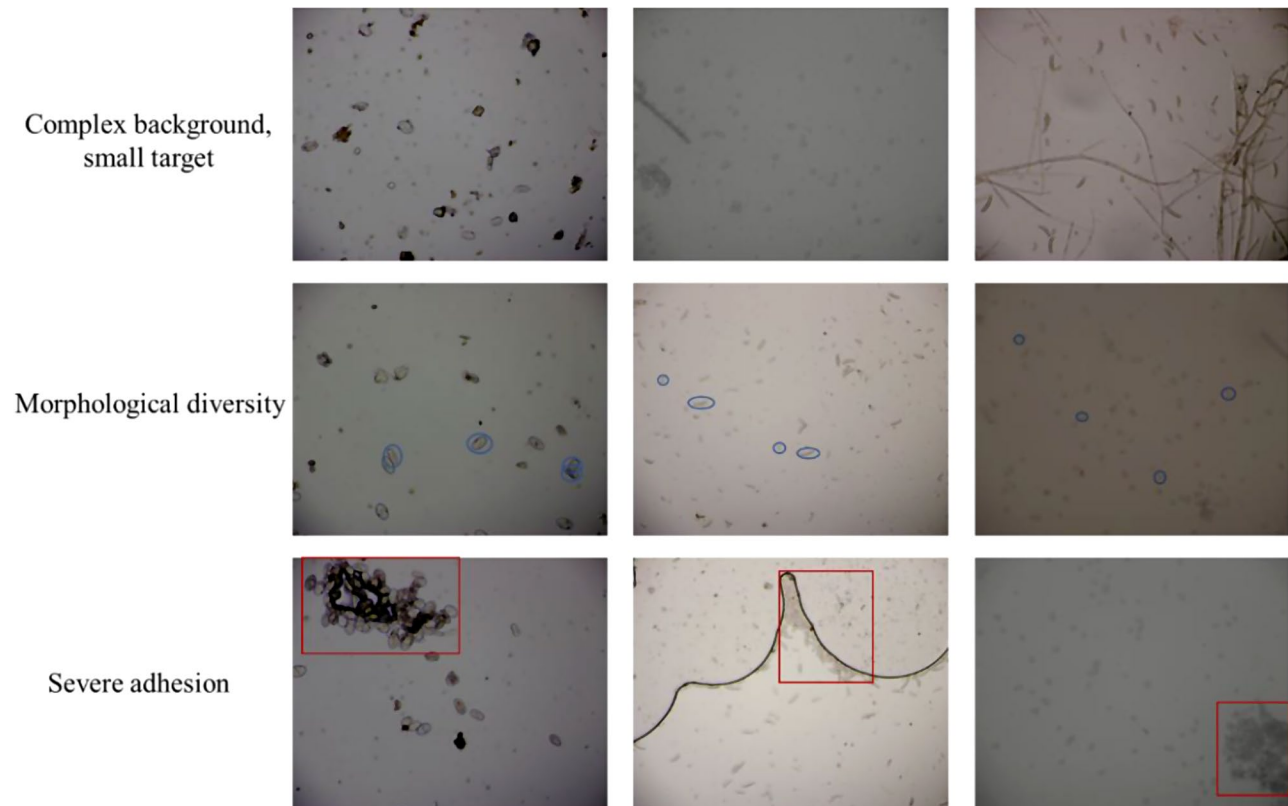
Fig. 3 Object detection difficulties in datasets (The shape diversity and adhesion are indicated by the blue oval box and the red rectangular box, respectively)

sets of spore samples. The training set, validation set, and test set are divided according to the ratio of 7:1.5:1.5. The number of spore image samples and spore numbers of the three categories in the dataset are shown in Table 1.

Difficulties in object detection

Since the data are directly scraped from the leaves without any processing after picking the fresh diseased leaves and the slides were directly observed under the microscope, this method is simple to operate. Still, the collected images have the following characteristics (Fig. 3), which makes the detection of pathogen spores challenging.

Small target

It can be seen from the figure that the size of the pathogen spores in the microscopic image is small, belonging to small targets. Such small targets have low resolution,

fewer pixels to cover, and weak feature expression ability, so it is easy to miss detection.

Complex image background and many disturbing materials

The background of images directly collected from leaves is complex and diverse, and there are mycelia, infectious microbe, leaf tissue, bubbles, and other disturbing materials, which brings difficulty to the identification of pathogen spores.

The individual morphology of the same pathogen varies greatly in different periods

In different periods, spores of the same pathogen will show different forms and be affected by the external environment such as operation during image collection. The morphology of spores will also be deformed to a certain extent. For example, the spores of powdery mildew show an elliptic cylinder or oval shape in the early stage, and

the cells are full of protoplasm, while the center of the cells begin to become empty and transparent until hollow after isolation, and the spores begin to appear concave and deformed. The conidium of fusarium fusarium is oval or sickle type, with two forms of large and small spores. The large spores have 3–4 septa and the two ends are fusiform, while the small spores have no septa and the two ends are blunt round. The spores of gray mold will appear convex and concave deformation due to germination in different periods.
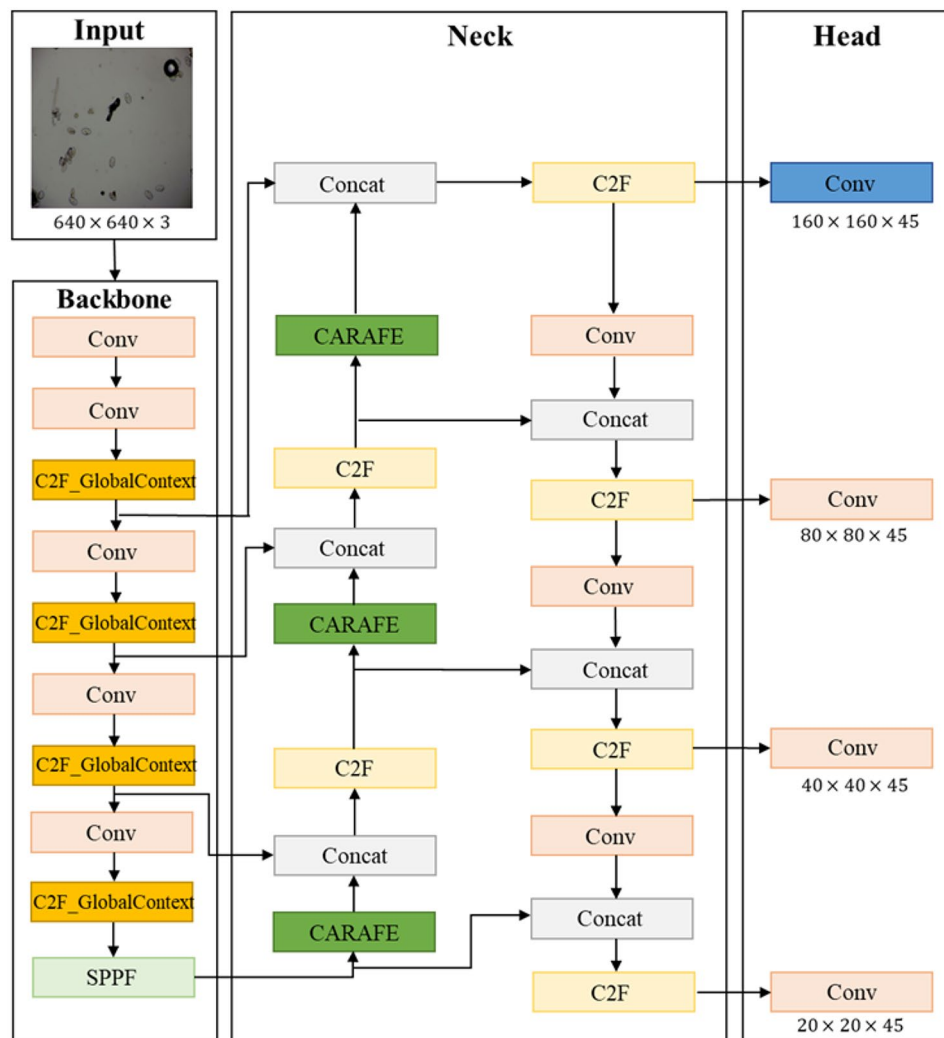
The proposed GCS-YOLOv8 spores detection model
To accurately detect small-target pathogen spores in natural scenes, this study constructed a GCS-YOLOv8 model, as shown in Fig. 4. First of all, given the problem that the sample size of pathogen spores is small and the subsampling multiple of YOLOv8 is large, it is challenging to learn the feature information of small targets from deep feature maps. This study adds a small target

detection layer to the network to fuse shallow and deep feature maps for detection. Secondly, the C2F module is optimized by the Global Context attention module in the Backbone layer, which enables the network to model the global context effectively. Finally, to solve the problem of dense distribution of pathogen spores, the lightweight up-sampling operator CARAFE was introduced in the Neck layer of the network to enhance the network's ability to perceive image details and further improve the accuracy of spore detection.

Optimized detection head
The output of the YOLOv8 model uses P3, P4, and P5 level feature maps by default, and the corresponding detection feature maps of these three layers can detect targets with sizes of 8×8, 16×16, and 32×32 at multiple scales. However, in the microscopic image of pathogen spores, due to the small size of the spore sample and the large downsampling multiple of YOLOv8, it is
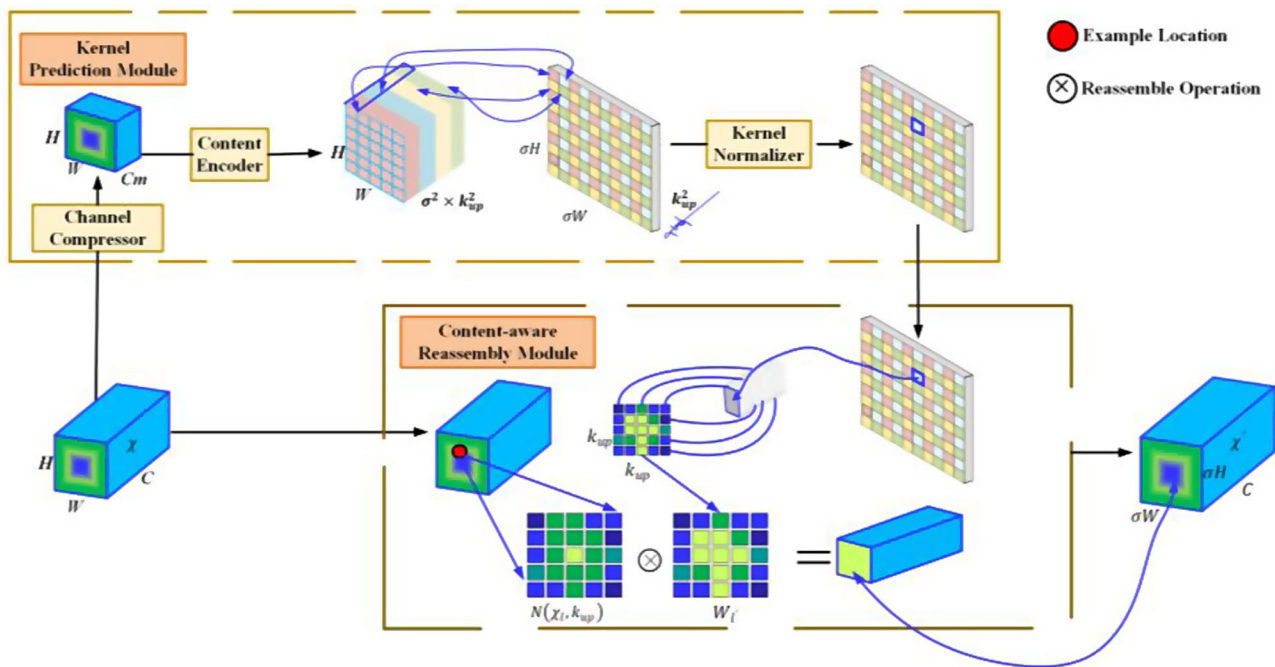

Fig. 4 The structure diagram of the proposed GCS-YOLOv8 model
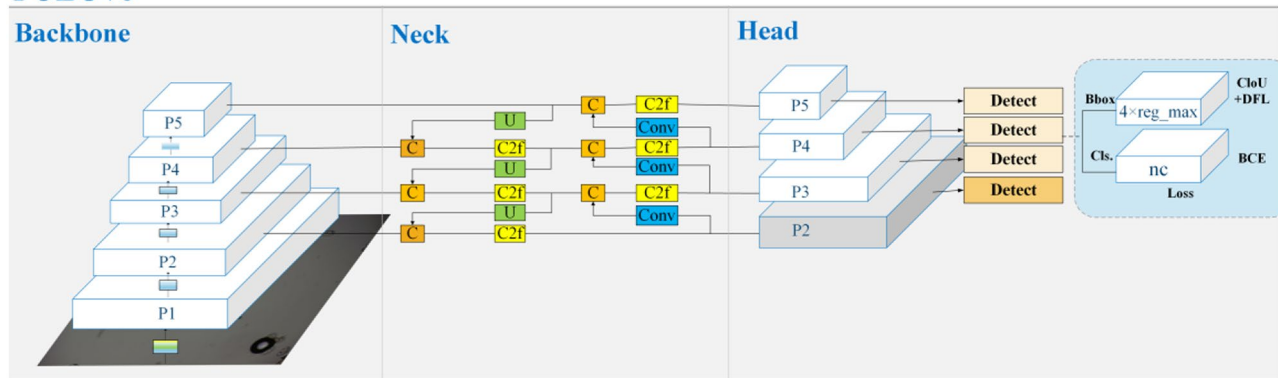
Fig. 5 The structure of YOLOv8 adding a small object detection layer

Fig. 6 The structure of the content-aware reassembly features module

challenging to learn the feature information of small target pathogen spores in the deep feature map. Therefore, given the small target of pathogen spores, this study added a small target detection layer to the network, as shown in Fig. 5. The network's shallow feature map and deep feature map were joined to make the network more sensitive to small targets and improve the identification effect of small target pathogen spores. The experimental results showed that although the calculation and storage cost were increased to a certain extent, the identification performance of the small target of pathogen spores was significantly improved after adding the detection head.

Neck using content-aware reassembly features

Feature upsampling is a key operation in convolutional neural networks, which can help the network extract feature information effectively and improve the recognition accuracy. CARAFE is a lightweight universal upsampling operator, which can use adaptive and optimized recombination cores at different locations to obtain better performance than the original upsampling operation, and increase the acceptance field to ensure lightweight, so that the network can aggregate context information in a large receiving domain [27]. Zeng et al. introduced lightweight conttion-sensing recombination module operators into the decoding module for upsampling to improve the feature extraction capability of rice diseases and image segmentation accuracy [28]. Mbouembe et al. used

CARAFE to replace the conventional upsampling operator in the model to generate feature maps containing more semantic information, improving the accuracy of the model to distinguish overlapping tomatoes [29]. Due to the complex background of pathogenic spores microscopic images collected in natural scenes such as bubbles, leaf cells and fungi, this study introduced CARAFE in the neck layer of the network to enhance the ability of the network to extract pathogenic spores under complex background.

As shown in Fig. 6, CARAFE is mainly divided into two main modules, namely the upsampled kernel prediction module and the feature recombination module. Assuming the upsampling rate is σ, given an input feature graph with the shape of $C \times H \times W$, first use the upsampled kernel prediction module to predict the upsampled kernel, and then use the feature recombination module to complete the upsampled kernel, and obtain the output feature graph with the shape of $C \times \sigma H \times \sigma W$. In the upsampled kernel prediction module, the channel compression of the feature graph is first carried out. For the input feature graph with shape $C \times H \times W$, a 1×1 convolution is used to compress the channel number to C_m, which is mainly to reduce the calculation amount of the following steps. Second, assuming that the size of the upsampled kernel is $k_{up} \times k_{up}$, if we want to use a different upsampled kernel for each position of the output feature map, we need to predict the shape of the upsampled kernel as $\sigma H \times \sigma W \times k_{up} \times k_{up}$. For the compressed input feature graph in the first step, it uses a convolution layer of $k_{encoder} \times k_{encoder}$ to predict the upsampled kernel, the number of input channels is C_m, the number of output channels is $\sigma^2 k_{up}^2$, and then we expand the channel dimensions in space dimension to get the upsampled kernel of shape $\sigma H \times \sigma W \times k_{up}^2$. Finally, we use softmax to normalize the upsampled kernel obtained in the second step, so that the weighted sum of the convolution kernel is 1. In the feature recombination module, for each position in the output feature map, we map it back to the input feature map, take out the $k_{up} \times k_{up}$ region centered on it, and dot the predicted upper sample kernel of the point to get the output value. For a given feature map X of size $C \times H \times W$ and an upsampling rate σ, CARAFE will generate a new feature map $X\prime$ of size $C \times \sigma H \times \sigma W$. For any target location $l' = (i'\ j')$ of output X, there is a corresponding source position $l = (i, j)$ in the input X, in which $i = [i'/\sigma]$, $j = [j'/\sigma]$. Among them, the formula of the reorganization step and the formula of the reorganization is given below as (1–2).

$$w_{l'} = \psi\left(N\left(x_l, k_{encoder}\right)\right) \tag{1}$$

$$x'_{l'} = ?\left(N\left(x_l, k_{up}\right), w_{l'}\right) \tag{2}$$

Where $N(x_l, k)$ denotes the $k \times k$ subregion of X centered at position l.

Optimized C2F module with global context attention module

The C2F module is one of the core components of YOLOv8, which plays a key role in improving model performance and accuracy. It mainly adopts the design concept of BottleNeck, dividing characteristics into two parts. One part reserves and does not handle any BottleNeck, and the other deals with them through a number of bottlenecks, each of which is divided into two channels. One is to pass processed characteristics to the next BottleNeck. One BottleNeck is reserved for subsequent concat, and all features are merged after n bottlenecks. The C2F module enables the model to capture the complex features of the image better. Due to the rich feature information of pathogen spores, it is necessary to maintain these rich features and identify spore features in the global information in the detection task. However, noise and impurities in the image will hinder the extraction of context features. In addition, convolution has the feature of local perception, which can only model the context of local regions, resulting in limited receptive fields. Global context attention is a new attention mechanism proposed by the GCNet network [30], which is lightweight and can effectively model the global context. The main idea is to calculate a general attention map and apply it to all positions on the input feature map, which greatly reduces the computation and ensures good performance. Choi et al. introduced Global Context attention into the model to extract and summarize global scene information to improve the robustness of the model [31]. Therefore, this study introduces the Global Context module to optimize the C2F block to overcome the shortcomings of convolution focusing on local features and enhance the aggregation of global context information in the extracted spore features, which is conducive to improving the network's ability to maintain the detailed features of pathogen spores.

The structure diagram of the Global Context (GC) module is shown in Fig. 7. Firstly, the global feature description is obtained through the context modeling part Wk, the 1×1 convolution and softmax function are used to extract the attention weight, the attention modeling is realized by matrix multiplication, and then the interchannel dependence is captured through the convolution layer $Wv1$ and $Wv2$. Layer $Normalization + ReLU$ is inserted between the two 1×1 convolutions for feature exchange. Finally, the residual connection is used to fuse the global context with the features of each position to obtain the context feature information. Adding a GC module can effectively model global context information while ensuring lightweight and improving the network's ability to extract spore features.
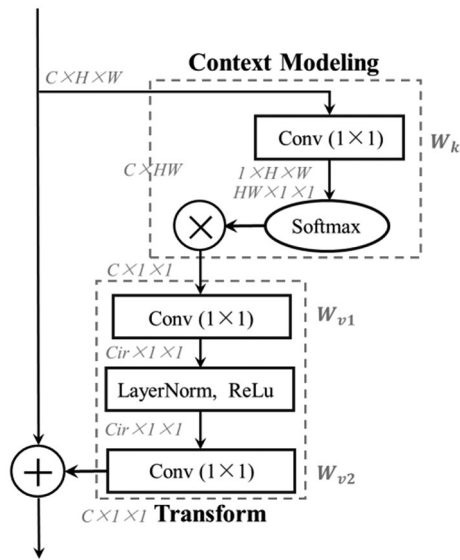
Fig. 7 The structure of the global context attention module(Where C represents the number of channels of the feature map, H and W are the height and width of the feature map, respectively, and r is the channel compression ratio)

Evaluation metrics

In this study, Mean Average Precision (mAP), Model size, floating point operations (FLOPs), and Parameters were used as evaluation indicators to evaluate the ability of different models to detect pathogen spores in natural scenes.

Average Precision (AP) reflects the mean value of each point in the interval of $0 - 1$ recall rate, and the area enclosed by the curve and the two coordinate axes on the P-R curve is the AP value of the detection target. P is the precision rate as shown in (3), R is the recall rate as shown in (4), and the AP value is calculated as shown in (5). True Positive (TP) and False Positive (FP) are the number of true and false positive samples, respectively, and False Negative (FN) and True Negative (TN) are the number of false and true negative samples.

$$P = \frac{TP}{TP + FP} \qquad (3)$$

$$R = \frac{TP}{TP + FN} \qquad (4)$$

$$AP = \int_0^1 (P \cdot R) dR \qquad (5)$$

mAP represents the average AP of n categories, and the value of mAP ranges between [0, 1]. The closer the mAP is to 1, the higher the model's accuracy. The mAP value is calculated as shown in (6).

Table 2 Experiment configuration

Experimental environment	Configuration
operating system	Windows 11
processor	13th Gen Intel Core i9-13900KF
memory	64 GB
graphics card	NVIDIA GeForce RTX4090
GPU acceleration library	CUDA 11.7
deep learning framework	Pytorch1.13.0

$$mAP = \sum_1^n (\frac{AP_i}{n}) \qquad (6)$$

Model size refers to the storage capacity required to store a trained model, generally measured in Mega. The model size depends on the architecture of the network, the number of parameters, and the format in which the model is saved. A larger model size usually means a more complex network structure or a larger number of parameters.

FLOPs refer to the total number of floating-point operations performed during the reasoning or training of a network and are a measure of the number of floating-point operations performed during the execution of a computer program or algorithm. In deep learning, FLOPs are usually used to measure the computational complexity or computation amount of a neural network model and are one of the important indicators to evaluate the computational efficiency and speed of the model.

The parameter refers to the number of trainable parameters in the neural network, which is used to represent the spatial complexity of the model. It will directly affect the complexity and storage requirements of the model and also affect the training and reasoning speed of the model. The parameter number includes the weight parameters and bias parameters of each layer.

Experimental results and discussion
Experimental setting details

The hardware configuration used in this study is shown in Table 2. The number of samples in each batch was set to 32, and 300 iterations were carried out. SGDmomentum optimizer was adopted, and the momentum value was 0.937. The initial learning rate was 0.01. To prevent instability of the model due to a larger learning rate, Warmup was selected to Warmup the learning rate so that the learning rate of several epochs at the beginning of training was smaller, and warmup_epochs=3.0 was set. In addition, HSV (Hue, Saturation, Value) color gamut transformation combined with random left and right flips was used for random online data enhancement of the dataset, and the HSV three-channel values were 0.015, 0.7, and 0.4, respectively.

Table 3 Detection results of YOLOv8 series

Model	mAP(%)	Model size(MB)	FLOPs(G)	Parameters($\times10^6$)
YOLOv8n	87.3	6.2	8.1	3.0
YOLOv8s	88.0	22.5	28.4	11.1
YOLOv8m	88.3	52.0	78.7	25.8
YOLOv8l	88.4	87.6	164.8	43.6
YOLOv8x	89.7	136.7	257.4	68.1

Detection results of YOLOv8 series

The YOLOv8 network consists of five versions of different sizes, namely YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. In this study, the five models were trained under the condition that the data set and parameters were the same, and the test results were shown in the table. As seen from Table 3, pathogen spores' detection accuracy gradually increases as the model scale increases, and the model size and operation time also increase. After a comprehensive analysis of mAP, model size, FLOPs, and parameters, YOLOv8n model size, FLOPs, and parameters are the smallest. However, the detection accuracy is also the lowest; the mAP is only 87.3%. YOLOv8x achieved the highest mAP at 89.7%, but its model size, FLOPs, and parameters reached 136.7 M, 257.4G, and 68.1×10^6, respectively. Compared with YOLOv8n, which has the smallest model and lowest complexity, the accuracy of YOLOv8s is improved by 0.7%, and the model size is increased by 16.3 MB. Compared with YOLOv8x, which has the highest accuracy, although the accuracy of YOLOv8s is 1.7% lower, the model size, FLOPs, and parameters are reduced by 114.2 MB, 229G, and 57×10^6. Compared with YOLOv8m and YOLOv8l, YOLOv8s reduces the model size by 30.5 MB and 65.1 MB, respectively, at 0.3-0.4% mAP loss. Therefore, in the case of balancing detection accuracy and model complexity, YOLOv8s was selected as the benchmark model for pathogen spore detection in this study.
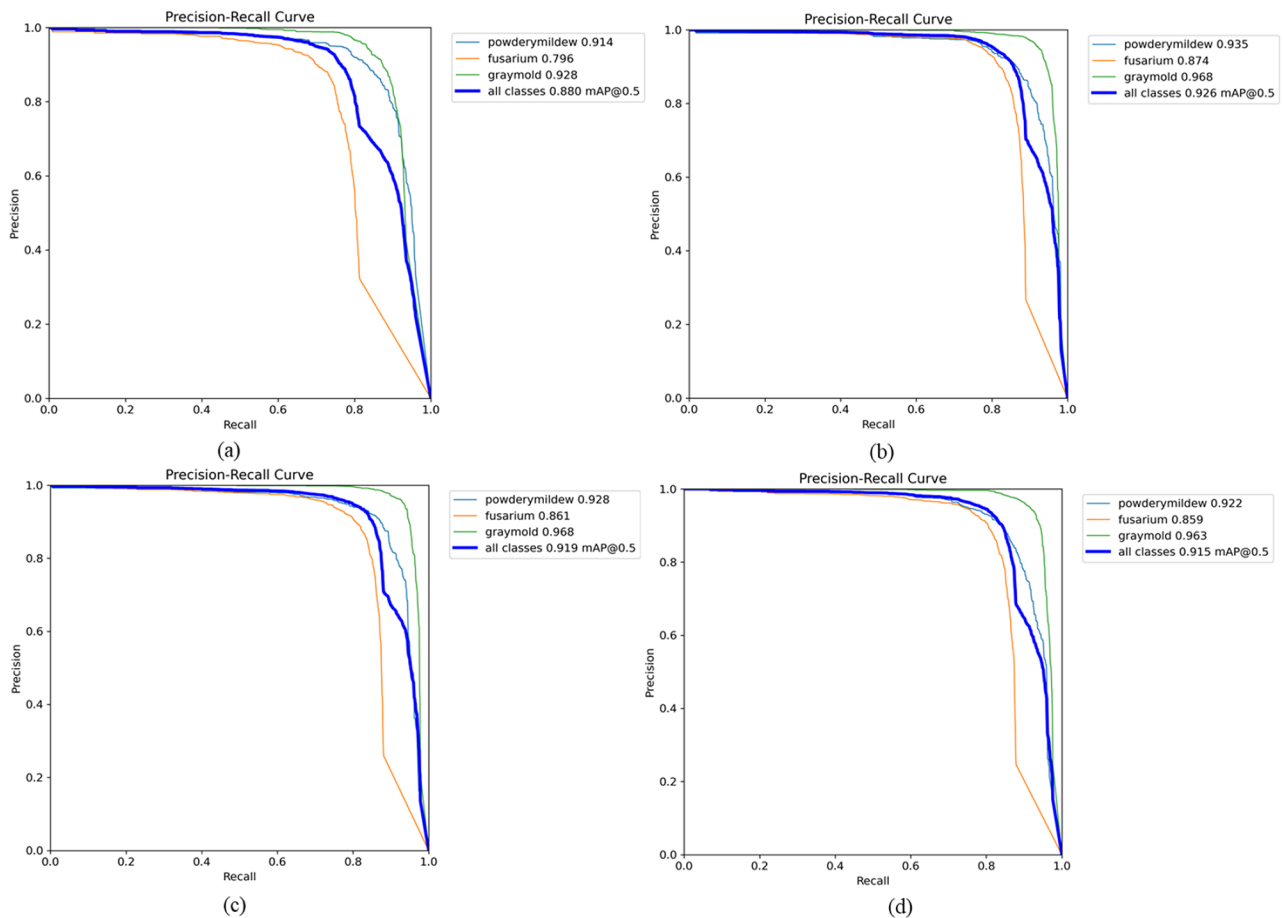
Ablation experiment

To verify the effectiveness of the improved part in improving the accuracy of pathogen spores, this study conducted an ablation experiment analysis on the proposed network, and the results are shown in Table 4; Fig. 8. As can be seen from the table, after adding a small detection layer (SD) (YOLOv8s-a), although the running cost is slightly increased and FLOPs is increased by 8.2G, the corresponding mAP is increased by 3.5%,

and the model size and parameters are also reduced by 0.8 M and 0.5×10^6 respectively. This shows that adding a small target detection layer can effectively increase the network's sensitivity to small targets and enhance the feature extraction ability for small target spores. On this basis, after using the Global Context module to optimize C2F (YOLOv8s-b), the mAP of the network increases by 0.4% while maintaining the lightweight, which indicates that the Global Context block effectively improves the detection accuracy of the network for pathogen spores by modeling global context information. After introducing CARAFE (GCS-YOLOv8), the detection accuracy of pathogen spores was further improved at the cost of increasing the model size by 1.1 MB and FLOPs by 4.1G. Finally, a mAP of 0.926 was obtained, indicating that the CARAFE block enhances the recognition ability of features in a larger receptive field. In summary, the gradual introduction of SD, GC attention module, and CARAFE upsampling module into the model can significantly improve the detection accuracy of the model against the three pathogens, and the performance of GCS-YOLOv8 is the best, which also proves the effectiveness of the addition of SD, GC module and CARAFE for the detection of pathogen spores and the complementarity of the three components.

As can be seen from Fig. 8, our proposed model has a high recognition performance for powdery mildew spores and gray mold spores, with AP reaching 93.5% and 96.8% respectively, among which gray mold spores have the highest detection accuracy. This is because gray mold spores have more easily distinguishable morphological characteristics, and gray mold spores have fewer morphological changes in different growth stages. In contrast, the variation of powdery mildew spores and fusarium spores was greater. Due to the loss of water in the air, the conidia of powdery mildew will have great morphological changes, and the center of the spores will become hollow and transparent until they are completely hollow, and the spores will begin to deform. Spores of fusarium wilt have different sizes and forms at different growth stages. Morphological differences between small conidia and large conidia may confuse the model, leading to some missed detection. It can be seen that spore morphological differences will have a certain impact on model detection, and the identification accuracy of spores with large characteristic changes and high diversity is low. The identification

Table 4 Detection results of the ablation experiment

Model	SD	GC	CARAFE	mAP(%)	Model Size(MB)	FLOPs(G)	Parameters($\times10^6$)
YOLOv8s				88.0	22.5	28.4	11.1
YOLOv8s-a	√			91.5	21.7	36.6	10.6
YOLOv8s-b	√	√		91.9	21.7	36.7	10.6
GCS-YOLOv8(proposed)	√	√	√	92.6	22.8	40.8	11.1

Fig. 8 The detection results of the ablation experiment of various fungal spores

Table 5 Detection results of the identification experiment

| Model	mAP(%)	Model Size(MB)	FLOPs(G)	Parameters(×10^6)
Faster R-CNN	67.4	313.8	108.38	82.3
YOLOX	50.0	13.0	26.8	9.0
YOLOv5	81.6	14.4	16.0	7.0
YOLOv7	86.2	74.8	103.2	36.4
YOLOv8s	88.0	22.5	28.4	11.1
GCS-YOLOv8(proposed)	92.6	22.8	40.8	11.1

accuracy of pathogen spores with little characteristic change was relatively high.

Compared to the classical model

Compared with traditional detection methods, the one-stage target detection model has greater advantages in detecting pathogen spores with higher accuracy and speed. To further verify the detection performance of GCS-YOLOv8 for pathogen spores, several classical target detection models such as Faster R-CNN [32], YOLOX [33], YOLOv5, YOLOv7 [34], and YOLOv8 were trained in this study. At the same time, to ensure the fairness of

the experimental results, the same hardware configuration is used for training during the experiment, and the obtained results are shown in Table 5; Fig. 9.

In terms of detection accuracy, the model proposed in this study performs the best, which is 42.6%, 11%, 6.4%, and 4.6% higher than YOLOX, YOLOv5, YOLOv7, and YOLOv8s in mAP, respectively, and has the highest mAP. At the same time, the model size and FLOPs are only 22.8 MB and 40.8G, respectively. This is because the small target detection layer can enhance the characteristic information of the network on the small target pathogen spores, and the C2F module optimized by the GC block enables the network to model the global context information and improve the detection accuracy of the pathogen spores in complex scenes. The introduction of the CARAFE upsampling module enables the network to further focus on extracting the correct characteristics of pathogen spores while maintaining light weight. In addition, the performance of YOLOv8s is second only to that of GCS-YOLOv8, which is also why YOLOv8s is selected as the base model for optimization in this study. In terms of overall performance, although the number of parameters and computation of GCS-YOLOv8 has
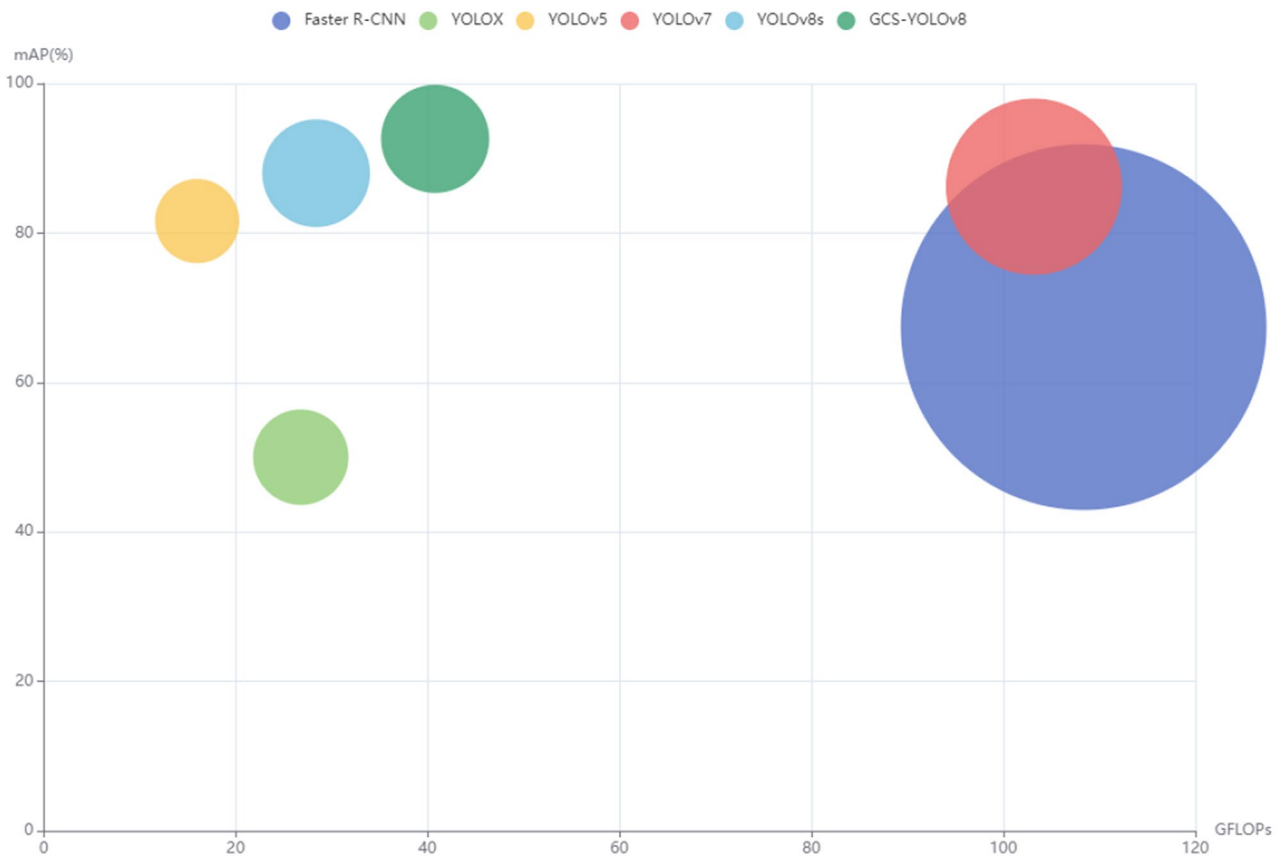
Fig. 9 Bubble diagram of the effect of different models in recognizing pathogen spores in natural scenes (bubble area represents model size)

been improved compared with YOLOX, YOLOv5, and YOLOv8, it can meet the requirements of lightweight and real-time detection required for pathogen spore detection within an acceptable range. Therefore, with a complex background and small pathogen spore target, GCS-YOLOv8 has the best detection effect. In summary, it can provide reliable theoretical support for the early detection and prevention of fungal diseases.

Visualization of detection models and results
In order to verify the effectiveness of the model proposed in this study, the detection results of the model GCS-YOLOv8 proposed in this study were visually compared and displayed with the benchmark model YOLOv8s when the dataset was the same (Fig. 10). As can be seen from the figure, compared with YOLOv8s, the model GCS-YOLOv8 proposed in this study can efficiently detect more pathogen spores in natural scenes, and the accuracy of small targets has been improved. YOLOv8s has different degrees of missed detection for pathogen spores at the edge, and the model GCS-YOLOv8 proposed in this study can significantly reduce the missed detection rate of spores and effectively improve the accuracy of fungal spores detection. Due to the complex background and excessive impurities in the microscopic

images of pathogen spores collected in the natural scene, YOLOv8s makes it difficult to extract accurate features of spores, resulting in misdetection and missed detection. However, the model proposed in this study can identify pathogen spores with unclear features and detect adhered pathogen spores, and the number of missed detections is correspondingly reduced. For the pathogen spores with large individual morphological differences, the model proposed in this study can also achieve a better recognition effect. Therefore, in the natural scene, our proposed model has better spore detection ability, especially in the face of complex background, diverse spore individual morphology, and serious adhesion of pathogen spores, which can effectively improve recognition accuracy. Therefore, the model proposed in this study provides a feasible method for identifying pathogen spores in natural scenes.

To make deep convolutional neural networks interpretable, an XAI method is used in this study to interpret the model's predictions [35]. The heatmaps of YOLOv8s and GCS-YOLOv8 were mapped using Gradient-weighted Class Activation Mapping (Grad CAM) to visualize areas of interest and considered important to the network. As can be seen from Fig. 11, compared with GCS-YOLOv8, YOLOv8 cannot accurately focus on the region
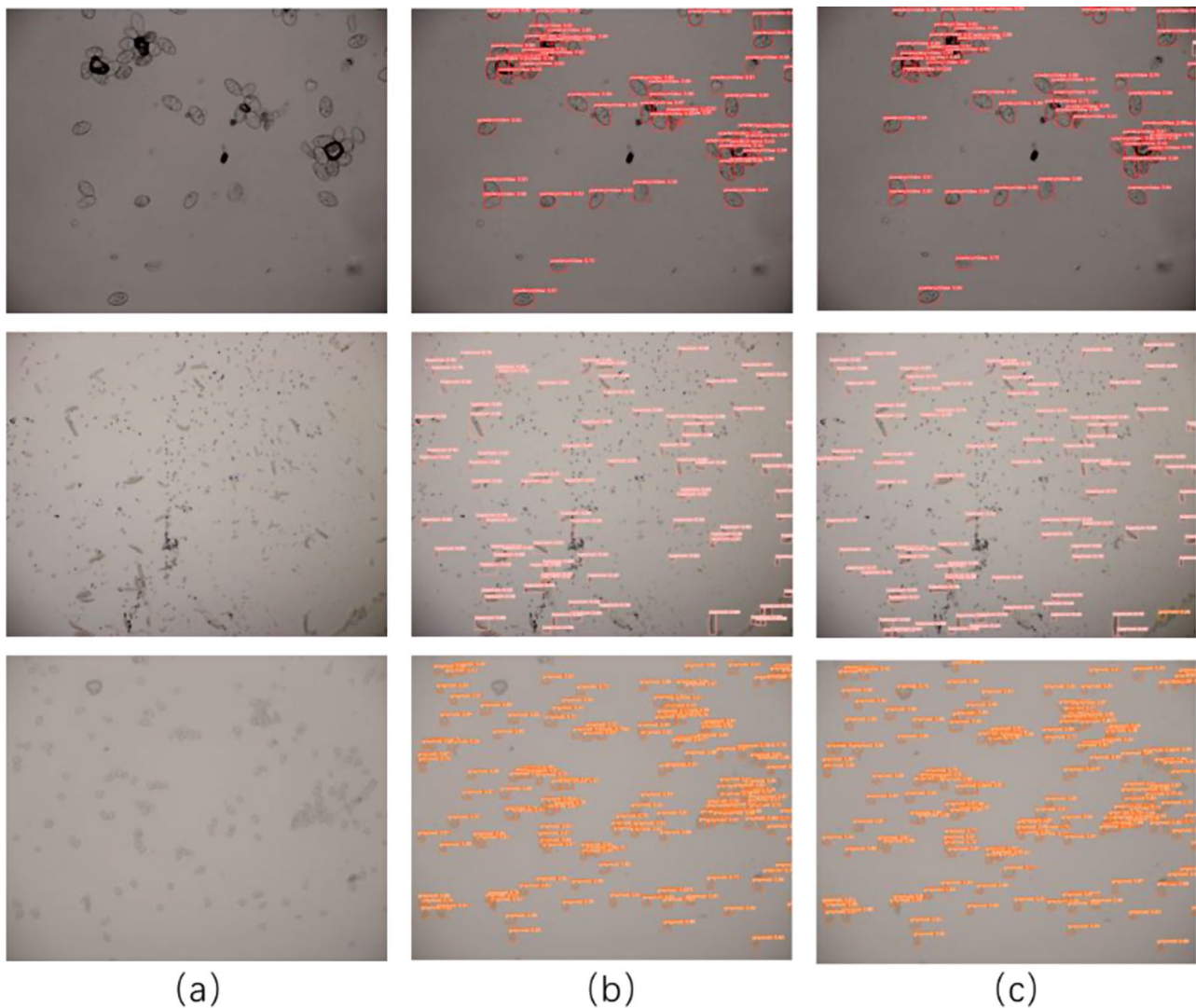
Fig. 10 Visual display of pathogen spore recognition results in a natural scene (where (**a**) are microscopic image examples of three kinds of pathogen spores, (**b**) are detection effects of YOLOv8s, and (**c**) are detection effects of GCS-YOLOv8)

of pathogen spores, especially in the case of complex background, so it cannot learn the correct characteristics of pathogen spores well. At the same time, the heatmap of GCS-YOLOv8 shows that it can focus and cover the region of small target pathogen spores well. This indicates that it is more sensitive to small goals and thus has better small-goal learning ability. Therefore, the reliability of the model GCS-YOLOv8 proposed in this study is also verified.
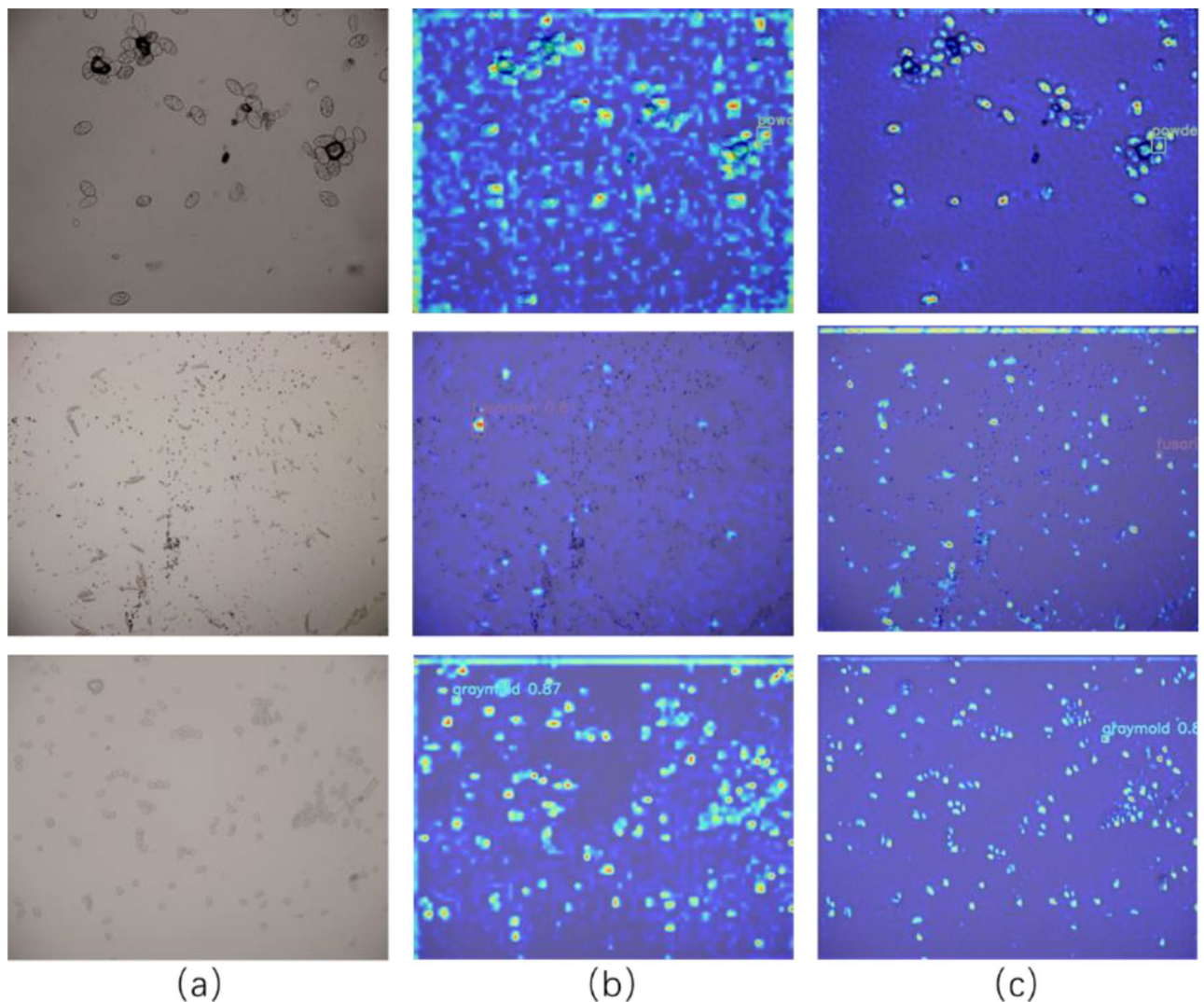
Robustness test

Due to the different environmental states in the shooting process, the brightness of the image will change significantly in the actual process of image acquisition. In order to further evaluate the robustness of the model, this study generated images of different brightness by adjusting the RGB channel ratio of microscopic images of

pathogen spores in the test set, so as to verify the recognition performance of pathogen spores under both dim and bright conditions (Table 6). As shown in Fig. 12, the recognition effect of the model is better in dim scenes. In contrast, the recognition accuracy of the model was slightly reduced in the high-light scene, because the shape and texture features of the pathogen spores in the microscopic images were weakened (Fig. 13). However, in general, light conditions have little effect on spore detection, and the model proposed in this study can meet the needs of accurate recognition under various light conditions in natural scenes.

Generalization test

There are many kinds of fungal diseases in the natural environment. In order to verify the effectiveness of the model in identifying other types of fungus and the

Fig. 11 Heatmap of YOLOv8s and GCS-YOLOv8 drawn by Grad CAM (where (**a-c**) shows the microscopic images of spores of three bacteria, the heatmap of YOLOv8s and the heatmap of GCS-YOLOv8 respectively)

**Table 6** Identification results of the robustness test

Pathogen AP(%)	natural scene	dusk	light
AP of powdery mildew pathogen	93.5	92	91.3
AP of the fusarium wilt pathogen	87.4	87.3	81.7
AP of gray mold pathogen	96.8	96.2	95.6
mAP	92.6	91.8	89.5

generalization ability of the model, this study conducted transfer training on 709 microimage data of cucumber downy mildew infection structure collected in the actual scene and tested on the test set. As can be seen from Table 7, compared with the YOLOv8s benchmark model, the model proposed in this study can achieve a better recognition effect and has good generalization.

Conclusions and future work

In this study, we built a microscopic image dataset of cucumber spores. The dataset contained 484 images and 31,629 spore samples, including 163 images of powdery mildew, 161 images of gray mold, and 160 images of fusarium wilt. Aiming at the problems of small targets, diverse morphology, complex background, and serious adhesion of pathogen spores, the pathogen spore detection model GCS-YOLOv8 was constructed. In this study, a new Backbone module is constructed, which uses the Global Context block to optimize the C2F module, effectively models global context information, uses the CARAFE up-sampling module in the Neck, and adds a small target detection layer in the Head layer. The network's performance to detect the small target and polymorphic pathogen spores in the actual scene was effectively improved. The experimental results show that
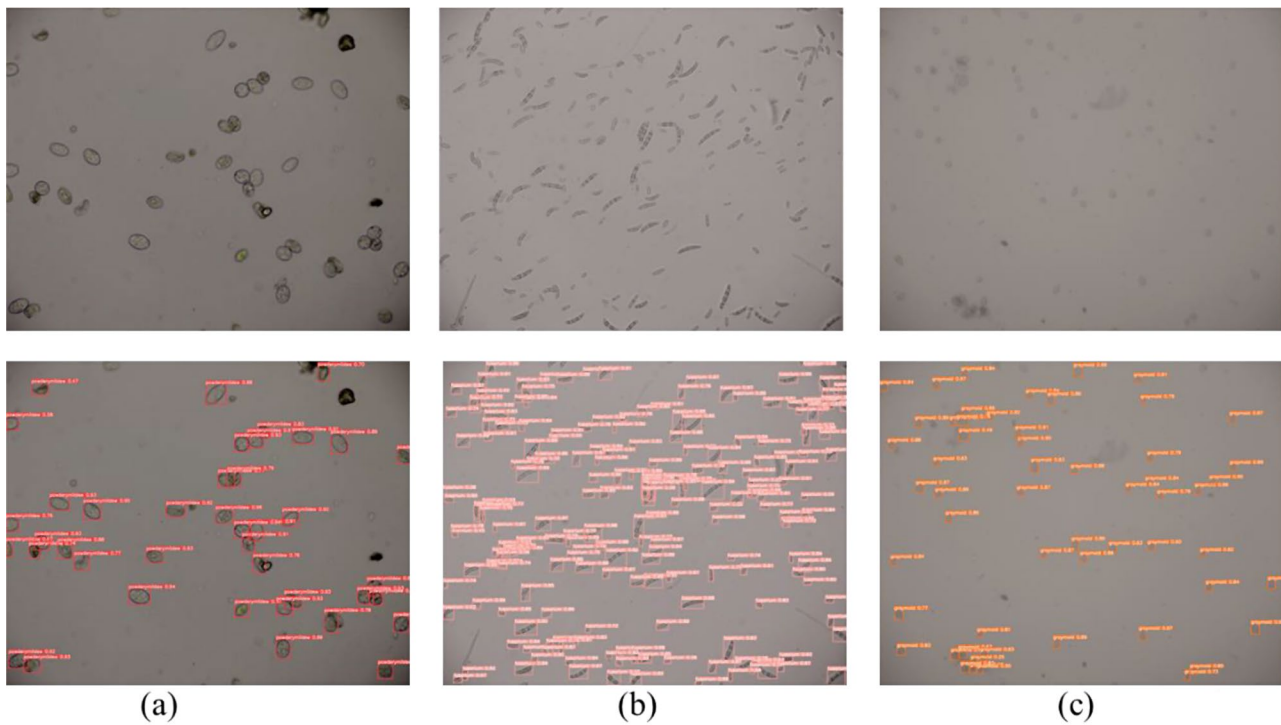
Fig. 12 Robustness test results under dim conditions. (**a**) powdery mildew spores (**b**) fusarium spores (**c**) gray mold spores
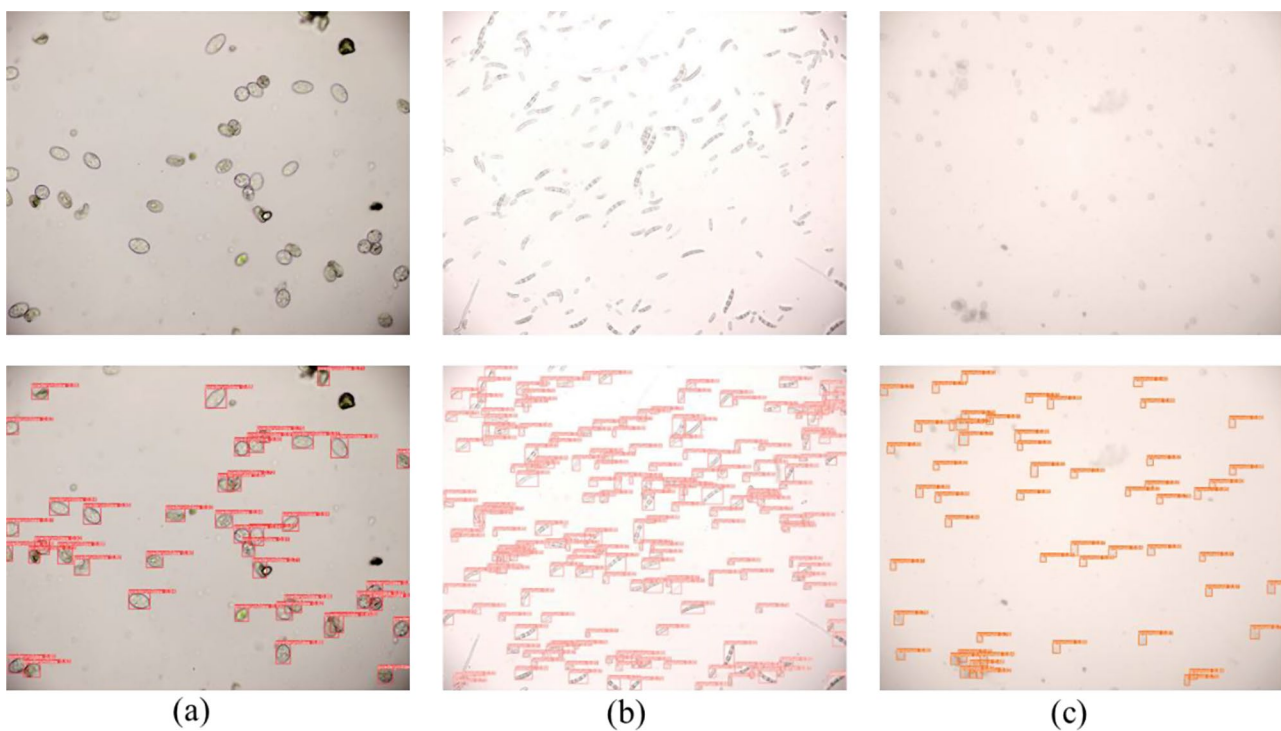
Fig. 13 Robustness test results in bright conditions. (**a**) powdery mildew spores (**b**) fusarium spores (**c**) gray mold spores

the proposed model is superior to the comparison detection model. On this basis, this study adopted an XAI method to explain the model's prediction. Grad CAM heatmaps showed that GCS-YOLOv8 could accurately focus and locate the spore region of small target fungus, thus extracting correct spore characteristics and information, which also verified the reliability of GCS-YOLOv8 in pathogen spore recognition. Finally, the robustness of

Zhu *et al. Plant Methods*   (2024) 20:131

Page 15 of 16

Table 7 Generalized test results

Pathogen category	model	AP(%)
Downy mildew spores	YOLOv8s	70
	GCS-YOLOv8	81.6
Downy fungus sporangium	YOLOv8s	83.9
	GCS-YOLOv8	91.2
Sporulation structure of downy mildew	YOLOv8s	72.6
	GCS-YOLOv8	95.2

the constructed model was tested under different brightness conditions, which verified that the model had good robustness, and it was also proved that the model had a good generalization effect on the microscopic images of spores, sporangium, and sporulation structure of downy mildew of cucumber. Therefore, this study provides a fast and accurate method of pathogen spore identification for microscopic images collected in actual field scenes and a feasible scheme for early prediction prevention and control of fungal diseases.

In the future, we will further optimize the model to identify the morphological characteristics of pathogen spores in the infection process to improve the practicability and generalization of the model.

Author contributions
X.Z., Y.Z., and L.Z.wrote the manuscript. X.Z. and F.C. designed and performed the field experiments. X.Z. designed and implemented the image processing and the deep learning models. W.G. and Y.W. performed data acquisition and data analysis. F.C. and C.Q. performed the data annotation and data analysis. X.Z., Y.Z., and L.Z. revised the manuscript. All authors read and approved the final manuscript.

Data availability
The dataset used during the current study can be accessed through https://github.com/fangtang-z/Dataset-of-Cucumber-spore-microscopic-images.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests

References
1. Zhang S, Zhu Y, You Z, Wu X. Fusion of superpixel, expectation maximization and PHOG for recognizing cucumber diseases. Comput Electron Agric. 2017;140:338–47.
2. Chomicki G, Schaefer H, Renner SS. Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. New Phytologist. Blackwell Publishing Ltd; 2020. pp. 1240–55.
3. Lin K, Gong L, Huang Y, Liu C, Pan J. Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. Front Plant Sci. 2019;10.
4. Li K, Zhang L, Li B, Li S, Ma J. Attention-optimized DeepLab V3 + for automatic estimation of cucumber disease severity. Plant Methods. 2022;18.
5. Velásquez AC, Castroverde CDM, He SY. Plant–Pathogen Warfare under changing Climate conditions. Current Biology. Cell; 2018. pp. R619–34.
6. Yang N, Chen C, Li T, Li Z, Zou L, Zhang R et al. Portable rice disease spores capture and detection method using diffraction fingerprints on microf luidic chip. Micromachines (Basel). 2019;10.
7. Zhao Y, Liu S, Hu Z, Bai Y, Shen C, Shi X. Separate degree based Otsu and signed similarity driven level set for segmenting and counting anthrax spores. Comput Electron Agric. 2020;169.
8. Zhang X, Song H, Wang Y, Hu L, Wang P, Mao H. Detection of Rice fungal spores based on Micro- Hyperspectral and Microfluidic techniques. Biosens (Basel). 2023;13.
9. Liu J, Wang X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. Plant Methods. 2020;16.
10. Li K, Zhu X, Qiao C, Zhang L, Gao W, Wang Y. The Gray Mold Spore detection of Cucumber based on microscopic image and deep learning. Plant Phenomics. 2023;5.
11. Bendel N, Kicherer A, Backhaus A, Klück HC, Seiffert U, Fischer M et al. Evaluating the suitability of hyper- and multispectral imaging to detect foliar symptoms of the grapevine trunk disease Esca in vineyards. Plant Methods. 2020;16.
12. Xu Y, Mao Y, Li H, Sun L, Wang S, Li X et al. A deep learning model for rapid classification of tea coal disease. Plant Methods. 2023;19.
13. McDonald SC, Buck J, Li Z. Automated, image-based disease measurement for phenotyping resistance to soybean frogeye leaf spot. Plant Methods. 2022;18.
14. Woyzichovski J, Shchepin O, Dagamac NH, Schnittler M. A workflow for low-cost automated image analysis of myxomycete spore numbers, size and shape. PeerJ. 2021;9.
15. Mah J-H, Kang D-H, Tang J. Morphological study of heat-sensitive and heat-resistant spores of Clostridium sporogenes, using transmission Electron Microscopy. J Food Prot. 2008.
16. Setyati D, Sulistyowati H, Rahmawati R, Ratnasari T. The spores structure of ferns growing in mountain Gumitir coffee plantation area Jember Regency. IOP Conf Ser Earth Environ Sci. IOP Publishing Ltd; 2021.
17. van den Brule T, Lee CLS, Houbraken J, Haas PJ, Wösten H, Dijksterhuis J. Conidial heat resistance of various strains of the food spoilage fungus Paecilomyces variotii correlates with mean spore size, spore shape and size distribution. Food Res Int. 2020;137.
18. Biermann R, Niemeyer L, Rösner L, Ude C, Lindner P, Bice I, et al. Facilitated endospore detection for Bacillus spp. through automated algorithm-based image processing. Eng Life Sci. 2022;22:299–307.
19. Prasobhkumar PP, Venukumar A, Francis CR, Gorthi SS. Pebrine diagnosis using quantitative phase imaging and machine learning. J Biophotonics. 2021;14.
20. Wang Y, Zhang X, Taha MF, Chen T, Yang N, Zhang J et al. Detection method of fungal spores based on fingerprint characteristics of diffraction–polarization images. J Fungi. 2023;9.
21. Zhang X, Guo B, Wang Y, Hu L, Yang N, Mao H. A detection method for crop fungal spores based on microfluidic separation Enrichment and AC Impedance characteristics. J Fungi. 2022;8.
22. Wang Y, Du X, Ma G, Liu Y, Wang B, Mao H. Classification methods for airborne disease spores from greenhouse crops based on multifeature fusion. Appl Sci (Switzerland). 2020;10:1–15.
23. Zhang DY, Zhang W, Cheng T, Zhou XG, Yan Z, Wu Y et al. Detection of wheat scab fungus spores utilizing the Yolov5-ECA-ASFF network structure. Comput Electron Agric. 2023;210.
24. Zhang Y, Li J, Tang F, Zhang H, Cui Z, Zhou H. An automatic detector for fungal spores in microscopic images based on deep learning. Appl Eng Agric. 2021;37:85–94.

25. Zhao Y, Lin F, Liu S, Hu Z, Li H, Bai Y. Constrained-focal-loss based deep learning for segmentation of spores. IEEE Access. 2019;7:165029–38.

26. Hoorali F, Khosravi H, Moradi B. Automatic Bacillus anthracis bacteria detection and segmentation in microscopic images using UNet++. J Microbiol Methods. 2020;177.

27. Wang J, Chen K, Xu R, Liu Z, Loy CC, Lin D. CARAFE: Content-Aware ReAssembly of FEatures. 2019; http://arxiv.org/abs/1905.02188.

28. Zeng W, He M. Rice disease segmentation method based on CBAM-CARAFE-DeepLabv3+. Crop Prot. 2024;180.

29. Touko Mbouembe PL, Liu G, Park S, Kim JH. Accurate and fast detection of tomatoes based on improved YOLOv5s in natural environments. Front Plant Sci. 2023;14.

30. Cao Y, Xu J, Lin S, Wei F, Hu H, GCNet. Non-local Networks Meet Squeeze-Excitation Networks and Beyond. 2019; http://arxiv.org/abs/1904.11492.

31. Choi J. Global context attention for Robust Visual Tracking. Sensors. 2023;23.

32. Ren S, He K, Girshick R, Sun J, Faster R-CNN. Towards Real-Time Object Detection with Region Proposal Networks. 2015; http://arxiv.org/abs/1506.01497.

33. Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: Exceeding YOLO Series in 2021. 2021; http://arxiv.org/abs/2107.08430.

34. Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2022; http://arxiv.org/abs/2207.02696.

35. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis. 2020;128:336–59.

Publisher's Note