IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 May 2023, accepted 5 June 2023, date of publication 9 June 2023, date of current version 19 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3284550

RESEARCH ARTICLE

The Application of Augmented Reality Technology in Urban Greening Plant Growth State Detection

HAOHAO XING[1], FEI DENG[1], YUN TANG[1], QINGQING LI[2], AND JING ZHANG[3]
[1]College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China
[2]Sichuan Tianyi Ecological Garden Group Company Ltd., Chengdu, Sichuan 610093, China
[3]China Southwest Architectural Design and Research Institute Company Ltd., Chengdu, Sichuan 610041, China

Corresponding authors: Jing Zhang (zhangjing_cswadi@163.com) and Fei Deng (dengfei@cdut.edu.cn)

ABSTRACT The current target detection network in deep learning has been widely used in plant growth state detection. However, with the development of deep learning, within the field of plant growth state detection, the performance of the detection network is no longer the primary factor limiting the detection accuracy and model generalization ability. The construction of high-quality and large-scale plant datasets is more significant for the improvement of model detection accuracy and generalization ability. However, traditional methods for building deep learning datasets for plants have a large time span and low efficiency. And it is difficult to construct and expand the dataset for plants with complex growth environments and difficult image acquisition by existing methods. To address this problem, this paper proposes a method for constructing plant datasets based on augmented reality techniques. The method proposed in this paper allows for the rapid and efficient construction of large-scale field datasets that match the actual inspection environment in the lack of data. Meanwhile, this paper proposes an automatic annotation method for datasets in conjunction with the imaging environment in virtual space. In this paper, we experimentally compare the proposed method with the method of expanding the dataset using GAN networks. Using the virtual dataset constructed by the method proposed in this paper as the training set, the trained YOLOv5 model achieves an average accuracy (@0.5:0.95) of 0.71 for the three detection categories on the test set. The detection accuracy of the six mixed datasets constructed using the two data expansion methods on the test set was experimentally tested. The proposed method in this paper improved the accuracy by 2.2%, 3.1%, and 7.0%, respectively. The smaller the percentage of real images, the greater the accuracy improvement. Experiments show that the method proposed in this paper can well solve the problems faced in the field of plant growth state detection, such as the lack of data, and provides a new idea for the production and expansion of datasets in plant detection tasks.

INDEX TERMS Deep learning, object detection, plant growth state detection, augmented reality, YOLOv5, dataset augmentation, dataset construction.

I. INTRODUCTION

Plant growth state detection is a method for monitoring and predicting the growth state of plants. Deep learning can extract features from large amounts of data to build models for classification and detection. Many scholars have applied it to plant growth state detection because of its rapid computer vision and image recognition development. Deep learning has various applications in detecting the growth stage of plants. Researchers typically focus on adapting and utilizing network models to match targeted datasets for better results. For instance, Fuentes et al. [1] improved the Backbone part of the original Faster-RCNN [2] network and achieved an average detection accuracy (mean Average Precision, mAP) of 85.98% on the tomato disease dataset; Ozguven and Adem [3] achieved automatic detection of beet leaf spots by

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

improving the Faster-RCNN network, obtaining 95.48% detection accuracy on 155 beet leaf photographs. However, Mohanty et al. [4] found that when the convolutional neural network developed based on the PlantVillage dataset was used to identify other tomato leaf disease datasets, the identification accuracy dropped sharply from 99% to 31%. This suggests that within the field of plant growth state detection, the performance of deep learning is mainly dependent on the dataset used [5].

A. CHALLENGES FACING

To address the challenges encountered in deep learning plant detection, improving the methods for building plant datasets may be more critical than optimizing the detection models themselves. Thakur et al. [6], in their paper, pointed out that the need for large-scale field public datasets is one of the major bottlenecks in model development for the detection of various plant diseases. Not only the number of datasets but also the consistency of the detection environment has an important impact on the detection accuracy. For example, Yuan et al. [7] conducted experiments on various image datasets with network models for crop pest recognition. They showed that their detection accuracy is higher when the background environment of the test images is consistent with the training images. Some scholars have tried to expand the plant dataset with Generative Adversarial Networks (GAN) [8]: for example, Barth et al. [9] used Cycle GAN [10] to synthesize crop images, trained the model with synthetic images, and fine-tuned the network with authentic images to improve the model detection accuracy. However, GAN generate high-quality images based on sufficient training images; for plant growth state detection applications (e.g., plant disease detection) with only a limited number of training images, it is challenging to train GAN models useful for downstream deep learning tasks [11]. For instance, Zhu et al. [12] trained GAN networks to generate orchid seedlings using different numbers of training images. They found that the GAN models trained using lower numbers of training images generated images that lacked texture details compared to the actual samples and could not capture the detailed structures of the roots and leaves.

B. CONTRIBUTION

The current primary approach to building deep learning plant datasets is to use GAN networks to expand the actual photographed and collected images. This approach makes it challenging to construct a large-scale field public dataset in the presence of variable detection environments and sparse samples. Therefore, in this paper, we propose an augmented reality-based [13] image data generation method as a way to construct a deep learning dataset of urban greenery plants Da Wu Feng Cao. Firstly, we mapped the Da Wu Feng Cao leaf blade and constructed a 3D model that represented the plant's shape and texture. We augmented these 3D models with rendering techniques to generate high-quality training images that capture detailed pose and texture details of the actual leaf blade.Additionally, we ensured that the background of the training images matched the actual growth environment of Da Wu Feng Cao to maintain consistency with the actual detection environment. To prevent the homogenization of image data in our constructed dataset, we employed a patchwork approach using randomly selected Da Wu Feng Cao leaf blades and various randomized models. This approach ensured that the training dataset contained a diverse range of pose, texture, and environmental variations. Additionally, we utilized an automated sample annotation approach based on virtual imaging environments that enhance the speed and accuracy of dataset construction by minimizing errors that can occur with manual annotation. The method described in this paper is experimentally proven to be superior to other plant dataset construction methods. The main contributions of this paper are as follows:

1) An augmented reality-based dataset construction method is proposed to build a plant dataset that fits the actual application situation through plant model construction and augmented reality technology, which provides a new way of thinking for producing deep learning datasets in plant growth state detection.

2) A leaf blade random patchwork and model randomization generation method is adopted, and an automatic sample annotation function is implemented. It accelerates the construction of datasets and solves the problems of lack of datasets, long collection periods, and high costs faced in plant growth state detection.

3) It solves the problem that it is difficult to expand the plant dataset when the GAN network faces the lack of actual data, and effectively improves the accuracy and generalization performance of the training model.

II. RELATED WORK

The superiority of deep learning in the field of vision tasks has provided new ideas in the field of plant growth state detection. With the rise of Convolutional Neural Networks (CNN) [14], many classical CNNs architectures such as AlexNet [15], VGG [16], and ResNet [17] are widely used in plant growth state detection. In the beginning, researchers extracted plant features from images of plant diseases using Convolutional Neural Networks (CNNs) and then employed classifiers to classify the different diseases. Later, as the demand for vision tasks upgraded, some classical target detection networks such as YOLO [18] and Faster-RCNN were proposed one after another. Researchers tried to apply the target detection networks to the field of plant growth state detection with success. The overall framework for the plant growth state detection using a target detection network is shown in Fig.1.

As shown in Fig.1, to use images taken from the field for plant disease detection, several preprocessing steps are necessary. This includes image filtering, cropping, data augmentation, and expansion to improve the quality of the dataset. The target detection network is generally divided into two modules: the backbone feature extraction network (Backbone)

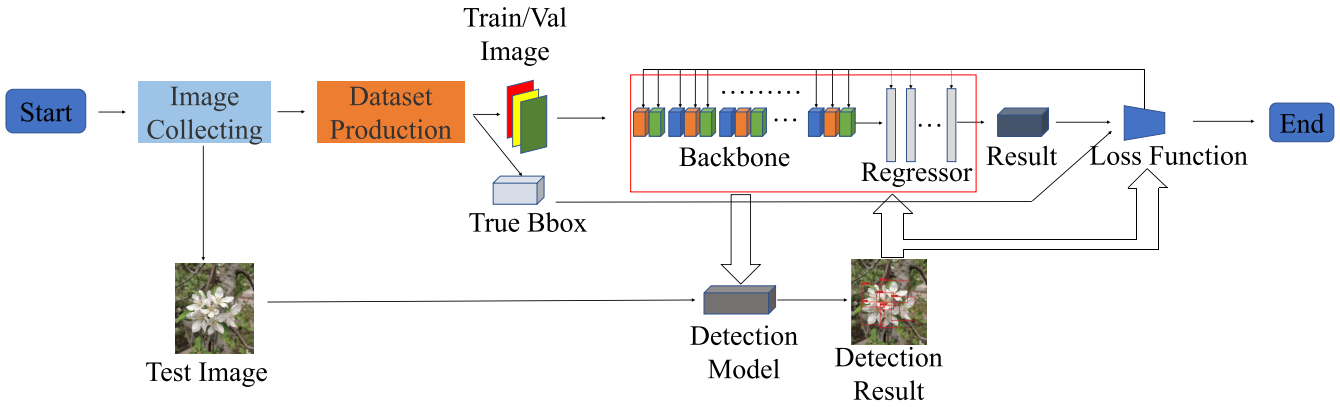


FIGURE 1. The detection framework contains two parts: the model training and the actual detection parts. The model training process is shown in the upper part of Figure 1: it includes the steps of dataset acquisition and production, feature extraction, prediction boxes regression, loss calculation, and model parameter update.

and the Regression prediction layer (Regression): The function of the Backbone part is to extract the deep semantic information of the image, and the implementation method is generally to convert the feature information of the image in W, H dimensions to the feature information on the channel by CNN, and fed into the Regression layer of prediction boxes and prediction of categories. The prediction boxes of the Regressor layer regression are fed into the loss function along with the labeled actual detection boxes for loss calculation, and finally, the parameters of the network model are iteratively updated based on the loss results. The best network model after the iterative update is generally selected for the actual detection, as shown in the red-bordered part in Fig.1. Finally, the network model structure and loss function are adjusted according to the detection result, and the modified network is trained again.

Standard target detection networks can be classified into single and two-step methods. In this paper, we focus on the single-step method and present results from several improved YOLO models that were trained on the COCO dataset and tested on the COCO test-dev dataset in recent years. As shown in Table 1.

In Table 1, FPS refers to the number of images that the network model can detect in one second, while AP (Average Precision) represents the accuracy of the model. AP is calculated based on Intersection over Union (IoU), which is a metric used to determine whether the predicted bounding box accurately captures the object. It measures the overlap between the predicted box and the ground truth box and provides a value between 0 and 1. TP (True Positive) refers to the number of correctly identified targets, while FP (False Positive) is the number of incorrectly identified targets. AP_{50} means that the threshold value of IoU is 50%, the detection boxes more excellent than this threshold are considered as TP, and the proportion of the number of TPs in this category to the total ground truth is calculated and recorded as the detection accuracy of this category. The average of the detection accuracy of multiple categories is taken under the COCO dataset, and Box AP indicates the average detection accuracy under each IoU threshold.

TABLE 1. Detection accuracy and detection rate achieved by the improved YOLO model on the COCO test-dev dataset in recent years.

Author&Year	Model	FPS	AP_{50}(%)	Box AP(%)
Redmon et al. 2018 [20]	YOLOv3 (Darknet-53)	19.6	57.9	33.0
Liu et al. 2019 [21]	YOLOv3 (@800+ASFF)	29.4	64.1	43.9
Wang et al. 2021 [22]	YOLOv4-P6	32	72.3	54.3
Ge et al. 2021 [23]	YOLOv5	62.5	68.8	50.4
Xu et al. 2022 [24]	PP-YOLOE-m	-	66.5	48.9

Table 1 shows that improving the network model is an effective way to improve the detection capability. However, the improvement in detection accuracy brought by the model improvement is limited in the target detection problem, and the performance of the detection accuracy depends mainly on the training data. For example, with detecting diseased tomato leaves, Fuentes et al. achieved an average detection accuracy of over 80% using the Faster-RCNN network in 2017. However, in 2020, Liu [19] achieved an average detection accuracy of 91.23% for this task using the relatively lightweight YOLOv3 network on own dataset. The above indicates that a superior dataset is more potent for improving detection accuracy than model improvement. Therefore, a more practical approach for tasks involving the detection of plant growth states is to construct a dataset that is adequate and suitable for the actual detection environment.

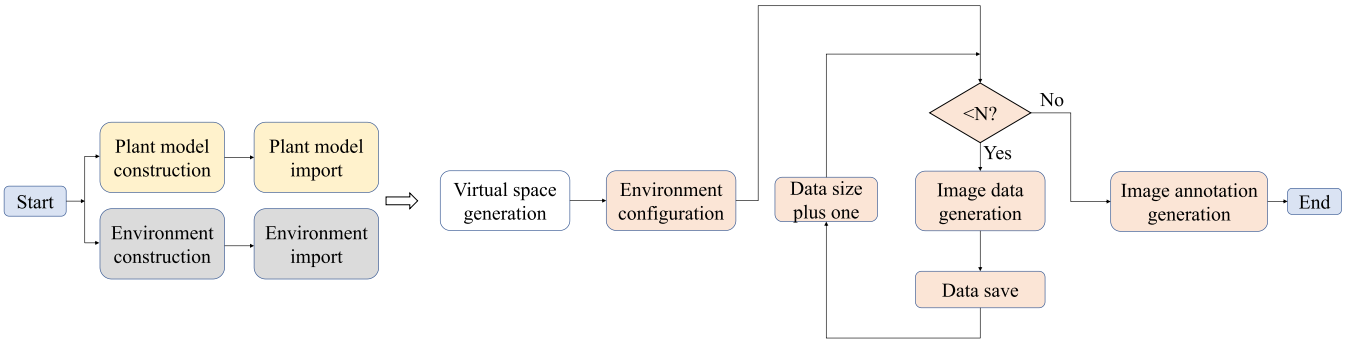


FIGURE 2. The virtual dataset construction process based on augmented reality technology is divided into three parts: Plant model construction and import, Environment construction and import, Sample data generation.

However, the construction of plant datasets also encounters many problems.For example, it can be challenging to collect image data for plants that grow in harsh environments and are scarce in nature, making it difficult to capture them in their natural settings. Plant samples cultivated artificially in the laboratory often lack detection accuracy during the collection process due to problems such as significant differences with the actual detection background. By contrast, cultivating plants in a laboratory requires significant expertise and complex equipment, which can drive up the cost of producing datasets. Therefore, this paper proposes a dataset construction method to construct plant datasets by augmented reality technology to train deep learning target detection networks. After being trained, the network model is utilized to detect "Da Wu Feng Cao" in its actual environment, with experimental results demonstrating the effective performance of the method. This paper has the following advantages over the traditional plant dataset construction methods.

1) Ability to construct plant datasets consistent with the actual detection environment and with sufficient data.

2) It is less expensive and generates datasets quickly and efficiently without relying on laboratory-grown plant samples or taking numerous repeat photographs of the samples.

3) Adopt the strategy of randomized model space to world space model generation to suppress the homogeneity of image data, implement the function of automatic dataset annotation to reduce the errors caused by mislabeling and omission in manual annotation, and accelerate the process of dataset construction.

4) Compared with the generative image expansion method can rely on a small number of plant image data to complete the expansion of plant data.

III. METHOD

The process of traditional plant dataset construction methods is rough as follows: taking plant images in the field, then select the captured images, and finally cropping and manually labeling the images, etc. Field capture requires setting up long-time cameras for on-time capture, and the construction of datasets spans a wide range of time and is inefficient. Therefore, this paper builds a dataset construction platform based on augmented reality technology,the detailed flow is shown in Fig.2.

Fig.2 illustrates the process of building a virtual dataset using the Virtual Dataset Building Platform. Firstly, the 3D plant model is constructed by manual modeling and mapping acquisition, which mainly includes the steps of plant geometry model reconstruction and plant mapping overlay. In order to complement the biological heterogeneity missing from the single modeling, the plant model is reconstructed by random mapping stitching technique in the construction of the plant model. At the same time, photogrammetry and aerial triangulation techniques are used to collect positioning attitude data from multiple perspectives in the actual background environment. The collected data was used to build a 3D environment model. Furthermore, a mapping patch for delicate parts is used to fill in the texture details in the environment model. The platform will then slice and dice the resulting model to prevent memory crashes caused by too much data. Then the cut model is imported into the platform database, and the platform calls the model in the database to generate the virtual space. The platform will randomly generate a specified number of plant models in the plant generation area based on the settings of environmental variables in the virtual space (environmental variables will control the change of weather in the virtual space, the number of normal and diseased plants generated in the generation area, etc.). To ensure the diversity of the collected data, the plant models are periodically regenerated. Next, the platform simulates the camera parameters of the actual data acquisition process for the samples in the virtual space. The camera view and coordinates are also changed during the acquisition process with reference to the real camera orientation and view. Save the captured image data to the data storage side and add one to the amount of data in the environment variable until the amount of data reaches the set value. Finally, after exiting the loop, the image annotation module will be called to automatically generate and save the coordinates of the smallest outer rectangle of the target plant in the virtual image to complete the automatic annotation of the dataset.

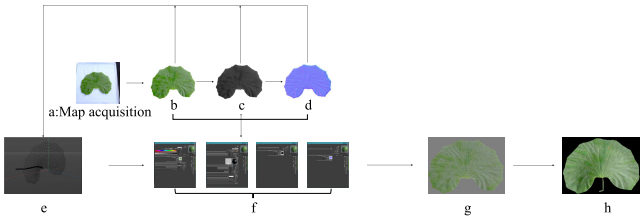


FIGURE 3. Artificial modeling process: (a) Mapping acquisition, (b) original mapping, (c) height map, (d) normal map, (e) model-creation, (f) material-creation, (g) UV show mapping, (h) model generation.

A. PLANT MODEL CONSTRUCTION METHOD

Plant model construction is an essential part of the dataset construction method and has the critical impact on the subsequent model training effect. Plants are highly biologically heterogeneous compared to other detection targets, so the impact of repeated models on the detection effect must be considered in the modeling process. This paper uses artificial modeling supplemented with random leaf blade patchwork to construct the plant model. The artificial modeling process is shown in Fig.3.

The whole artificial modeling process is divided into a total of five parts, first, need to shoot the flattened blade to collect the original mapping of Da Wu Feng Cao leaves and then use PhotoShop software to capture the blade surface texture and bump to generate the height map, the height map for different color depth rendering, to generate with virtual bump texture normal map. Subsequently, the blade model was constructed using Cinema4D, and the blade model, including the blade contour and blade veins, required texture information such as a height map and normal map captured using the texture map; furthermore, through the acquired original mapping, height map, normal map on the blade material color, bump, reflection, normal line, etc. create, and according to the generated blade material and original mapping to create a mapping, use the UV mapping tool to spread mapping to the blade model, and finally generate the complete model.

Although artificial modeling can simulate leaf shape and texture information fairly realistically, a single leaf profile shape will lead to an overly homogeneous leaf model, resulting in the homogenization of the collected data. Therefore, in this paper, we construct multiple categories of shape contours and generate the mapping of leaf blades by random leaf blade patchwork. Taking diseased leaves as an example, if the diseased leaves are too homogeneous, the plants will lack biological heterogeneity, which will cause the trained network to be less effective in detection. Therefore, we intercept the diseased parts of the mapping for random patchwork and mapping on the surface of the leaf model, and the specific effect is shown in Fig.4.

A diseased leaf model was constructed in Fig.4, whose disease areas were randomly generated to compare individual diseased leaf samples, ensuring the validity of the data while significantly reducing the homogeneity of the samples.

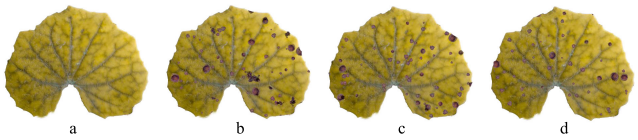


FIGURE 4. (a) Original leaf model. (b,c,d) Randomly generated diseased leaf model.

FIGURE 5. (a) Real environment. (b) Environment under virtual space.

B. ENVIRONMENT CONSTRUCTION METHOD

Even humans use the relationship between background and foreground to understand objects, a relationship known as background bias [25]. However, there needs to be more background bias in many public plant datasets, such as the PlantVillage dataset, so the accuracy of the model trained on the PlantVillage dataset decreases when detected. In order to overcome the effect of background bias, it is required to make the backgrounds of training and test datasets consistent. However, it is challenging to construct a training dataset consistent with the actual detection background in real situations due to time and space variations. Therefore, this paper improves the traditional dataset construction method and simulates the actual detection environment by augmented reality technology. Augmented reality is a technology that combines the real world with virtual information based on real-time computer computing and multi-sensor fusion [26]. The leading technology used in this paper is visual augmented reality, the core of which lies in matching and visualizing virtual information and the real world in physical space [27]. This paper utilizes photogrammetry to acquire multi-angle images and obtain positioning attitude data for environment model construction. This method offers several advantages, including reduced time consumption, low cost, high accuracy, and realistic texturing. The specific approach is first to set the heading of the Unmanned Aerial Vehicle(UAV), select a heading with 80% overlap with the actual site for aerial photogrammetry to obtain image files with site location information, and then use the camera to capture the delicate mapping part. Finally, ContextCapture was used to read the image file to generate the 3D mapping model, as shown in Fig.5.

The positional attitude parameters from the image files alone are insufficient for 3D reconstruction and need to be complemented using the Bundle Aerotriangulation method. To do this, it is first necessary to mathematically model the

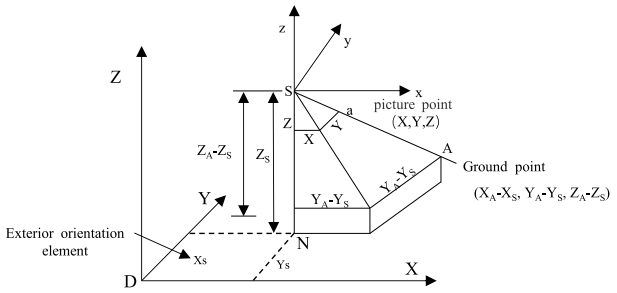


FIGURE 6. Central projection conformational relationship.

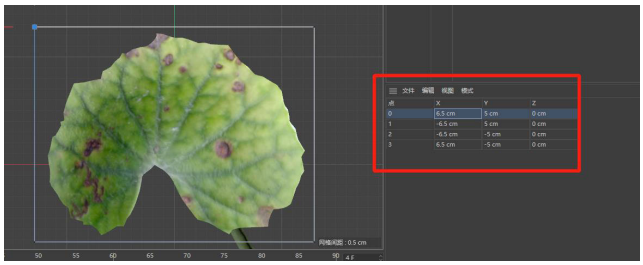


FIGURE 7. Coordinates of the smallest external rectangle in the model space.

photogrammetric data.As shown in Fig.6, the two parallel coordinate systems are the ground photogrammetric coordinates and the image-space auxiliary coordinates. When the image point a, the camera center point S, and the ground point A of the measured image are in the same line, the co-linear equation(1) can be deduced:

$$\begin{aligned}(x-x_0) &= -f\frac{a_1(X_A-X_S)+b_1(Y_A-Y_S)+c_1(Z_A-Z_S)}{a_3(X_A-X_S)+b_3(Y_A-Y_S)+c_3(Z_A-Z_S)}\\ (y-y_0) &= -f\frac{a_2(X_A-X_S)+b_2(Y_A-Y_S)+c_2(Z_A-Z_S)}{a_3(X_A-X_S)+b_3(Y_A-Y_S)+c_3(Z_A-Z_S)}\end{aligned} \tag{1}$$

In equation(1) (x, y) are the coordinates of the image point a in the image plane with the principal image point (x_0, y_0) as the origin; (X_A, Y_A, Z_A), (X_S, Y_S, Z_S) are the coordinates of ground point A and camera center point S in the object space, respectively; f is the image principal distance; (a_i, b_i, c_i) is the cosine of the angular elements of the external orientation of the image in 9 directions.

The above is the co-linear equation with the main distance of the image film. The Bundle Aerotriangulation is precisely based on the similarly projected beam in each image film as the leveling unit, the co-linear equation based on the central projection as the mathematical model of leveling, the coordinates of the image point as the observation value, according to the condition that the coordinates of the common intersection point of adjacent images are equal. The encrypted coordinates of the control point are equal to the ground coordinates, and the coordinates of the outer orientation elements and encrypted points of each image film are solved. These coordinates complement the positional attitude parameters of the model.

C. AUTOMATIC ANNOTATION METHOD FOR DATASET

The dataset labeling is the most time-consuming and laborious step in the data pre-processing process, and it is also the one that has a significant impact on the network training. The traditional data annotation method uses tools such as Labelimg, and it is difficult for human annotation to achieve standard and uniform, accurate annotation on datasets with small targets and high overlap. This section proposes a method for automatically annotating datasets in a virtual spatial imaging environment, which achieves better results on plant virtual datasets. As shown in Fig.7, the coordinates of the sample under the model space and the coordinates p_m of the minimal outer rectangle of the sample can be obtained when the target sample is generated.

The coordinates in the model space must be left multiplied with the M matrix to get the coordinates p_w in the world space. As shown in equation(2):

$$p_w = M_T M_R M_S p_m \tag{2}$$

The equation(2) represents the *translation*, *rotation*, and *scaling* of the coordinates of the object in model space to obtain the coordinates of the object in world space. The coordinates of the objects in world space need to be transformed by V and P matrices to get the screen coordinates presented in front of the screen p_S. V matrix is a transformation matrix from world space to camera space, which maps the world coordinates to the lens space coordinates of the current position and pose [28]. P matrix is a transformation matrix from camera space to a two-dimensional plane. Objects in camera space are mapped into the two-dimensional plane by perspective projection and eliminating the coordinates of those invisible range points. Specifically, as shown in equation (3):

$$\begin{aligned}p' &= R^T T^{-1} p_w\\ p_s &= M_{project} p'\end{aligned} \tag{3}$$

The equation(3), describes the mapping process from the world coordinate system to the camera coordinate system, and p' is the coordinates in camera space at this time. represents the matrix that maps from camera space to the 2D plane, and p_s is the coordinates of the resulting 2D plane. The effect of the actual labeled sample is shown in Fig.8.

IV. EXPERIMENT

In order to verify the effectiveness of the proposed plant dataset construction method, a real image dataset is constructed in this paper, as shown in Fig.9(a).And based on part of the real image data, a virtual dataset is constructed using the dataset construction method proposed in this paper, as shown in Fig.9(b). Meanwhile, the GAN network for generating virtual images is trained based on real images in this paper, and the training data is expanded by the GAN network, as shown in Fig.9(c).

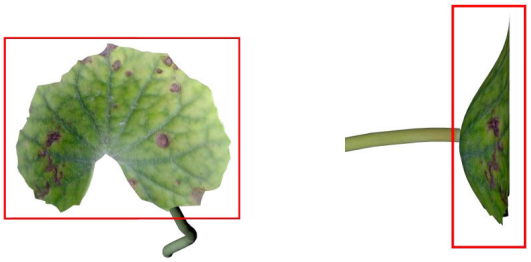


FIGURE 8. Actual labeling sample.

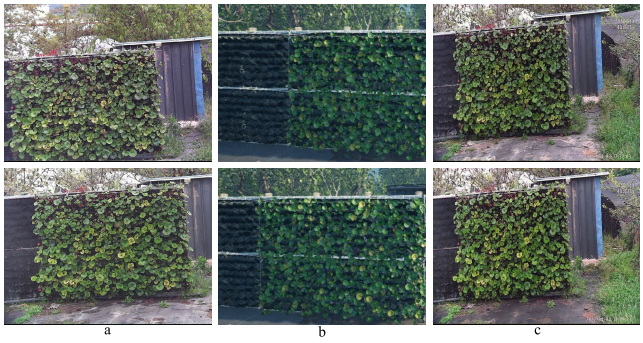


FIGURE 9. (a)Real image data,(b)Virtual images generated using the method proposed in this paper,(c)Virtual images generated using GAN networks.

As shown in Fig.9(a), the real picture collection site is located in the Tian Yi Ecological Park in southwest China. Through human intervention to control the growth state of green plants in fixed areas, using fixed cameras for image acquisition, and using intelligent wide-angle dome-type network cameras, the camera's highest resolution can reach 2560 × 1440, the highest frame rate of 25FPS. A total of three groups of camera positions are set up, respectively, front view camera group 1, left view camera group 2, right view camera group 3. Front view, left view and right view images of the growing area of greenery are captured by the camera. The field of view of the left- and right-view camera groups is 60.2-3.4°. The collection time is February 28, 2022 - March 31, 2022, daily 7:00-19:00 time period, which includes all kinds of light conditions and weather conditions such as sunny days, cloudy days, and haze. In order to build a real dataset with clear images and easy to annotate, a total of 1100 clear images with different types of lighting and weather were selected as the real image dataset from the 11920 images collected in this paper, with 1920×1080 pixels. In order to improve the detection accuracy, the detection targets are divided into three categories in the process of labeling the diseased leaves, from low to high, according to the severity of the leaf disease as l_y, l_i, and l_d.

Fig.9(b) shows the virtual dataset constructed by the dataset construction method proposed in this paper. In order to ensure that the generated virtual images can highly simulate the real detection environment, this paper sets the environmental parameters in the virtual space by imitating the lighting and weather conditions in the real environment. Firstly, we count the percentage of various kinds of weather in the real picture acquisition process, and generate virtual images with the same weather percentage in the virtual environment according to that percentage. First, the percentage of various kinds of weather during the real picture acquisition is counted, and the weather changes in the virtual environment are adjusted according to that percentage. Secondly, the light level in the real environment is collected by sensors, and the change in light level during the collection is simulated using the dataset construction platform. For example, the illumination level of a sunny day in the real environment gradually rises from 30,000-130000lux in the morning from 7:00 am to 12:00 am. At this time, the illumination level under the virtual platform will also be set to rise gradually according to this rule, to simulate the lighting conditions in the real environment. Finally, the number of plants generated in the virtual space is set to imitate the number of each type of plant in the real environment. At this point, the environmental parameters in the virtual space have been set. Then the virtual data are collected by simulating the imaging angles and camera parameters of the three groups of cameras in the real acquisition process through UE4.

Fig.9(c) shows the image data generated by the GAN network. The training of the GAN network requires a large number of real images. In order to make the virtual images generated by GAN network can be effectively used for the expansion of the dataset, 8,000 of the 11,920 real images collected by real are used for the training of the network in this paper. The network is trained by setting different numbers of training sets,and comparing the generated results after training until the model performance is optimal. The effect of the GAN network trained with different number of training sets is shown in Fig.10.Due to the large number of samples, the time required to train the GAN network is longer, and the requirements for the equipment are higher. The number of training sets has an important impact on the quality of the images generated by the GAN network, as seen in Fig.10. In the lack of real data, higher quality virtual data cannot be generated through GAN networks.

The experimental environment in this paper is NVIDIA GeForce RTX3060, using Pytorch1.8.0 to build the detection network. Using the average accuracy rate @0.5:0.95 as a model accuracy measure, it measures the detection ability of the model under different IoU thresholds. A higher average accuracy indicates that the regression boxes of the model are more accurate, and the detection results are more accurately fitted to the original labels. The experiment divides the real dataset into a training set, validation set, and test set in the ratio of 8:1:1. The total number of training sets was kept constant in the experiment, and the total number of training sets was 880. Subsequently, different amounts of real data were assigned to each training set. Finally, the training set is expanded by two different dataset expansion methods, bringing the total number of training sets to 880 for each group. The YOLOv5s network was used to train each dataset,

FIGURE 10. (a)Real samples of input GAN networks,(b)The result of the GAN model trained with 500 images as the training set,(c)The result of the GAN model trained with 1000 images as the training set,(d)The result of the GAN model trained with 3000 images as the training set,(e)The result of the GAN model trained with 8000 images as the training set.

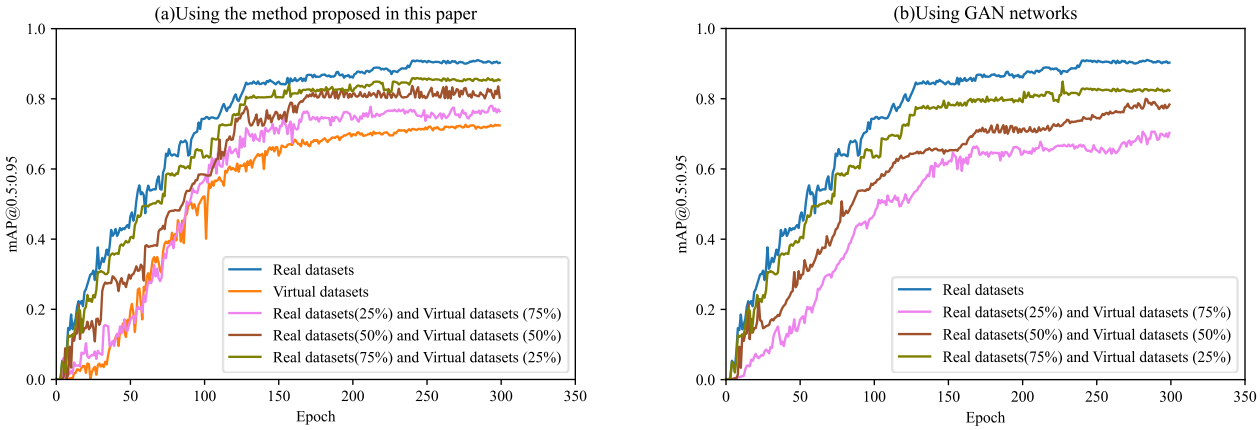


FIGURE 11. Accuracy of models trained on different training sets.(a)The dataset constructed and expanded by the method proposed in this paper is used as the training set.(b)Use the dataset expanded by GAN network as the training set.

and then the resulting models were tested on the test set. Fig.11 shows the accuracy curves of the trained models on different datasets on the validation set.

Fig.11(a) shows the accuracy curves on the validation set for the dataset constructed according to the method described in this paper. The real dataset is the actual acquired real images and the virtual dataset is the virtual images generated using the method proposed in this paper. The percentage of real data in the mixed dataset is 25%, 50%, and 75%, respectively. Fig.11(b) shows the accuracy curves of the dataset constructed using the GAN network on the validation set. The mixed dataset is formed by expanding the real data using the GAN network, and the number of real images in each group is 220, 440, and 660 in turn. The real images are fed into the GAN network to generate virtual images, and then the generated virtual images are mixed with the real images to form a mixed dataset. From the accuracy curves of the model trained on the three mixed datasets in Fig.11(a), it can be seen that the proposed method in this paper has good results for the expansion of the dataset, although it cannot perfectly fit the complexity of real images. Even without mixing real data, the network trained on the virtual data is still able to achieve an average detection accuracy of more than 0.7 for the three detection categories. Since GAN networks need to use existing real images to generate virtual images, it is not possible to construct datasets that are all virtual images. From Fig.11(b), it can be seen that the GAN network can achieve good results for data expansion when the dataset is sufficient. However, in the lack of real data compared to the proposed method in this paper, expanding the dataset by GAN network leads to too much similarity of images and lack of diversity of samples, resulting in the reduction of detection accuracy.

Comparing the accuracy of the detection networks trained above, it can be seen that the data expansion method proposed in this paper is better than the GAN network. Moreover, the method can quickly and efficiently construct virtual datasets in the absence of real data, solving the problem of not being able to use GAN networks to expand datasets in the absence of real data. To verify the generalization performance of the resulting models and the correctness of the experiments, the average accuracy of each model was tested on the test set again, as shown in Table 2:

The accuracy comparison of the test set in Table 2 shows that the dataset expanded by the method proposed in this paper is superior to the dataset expanded using the GAN network in terms of accuracy. Comparing the accuracy errors on the test set, the expansion effect of the two methods on the dataset is in accordance with the performance on the validation set. The virtual datasets constructed by both methods can expand the datasets to some extent, but the method proposed

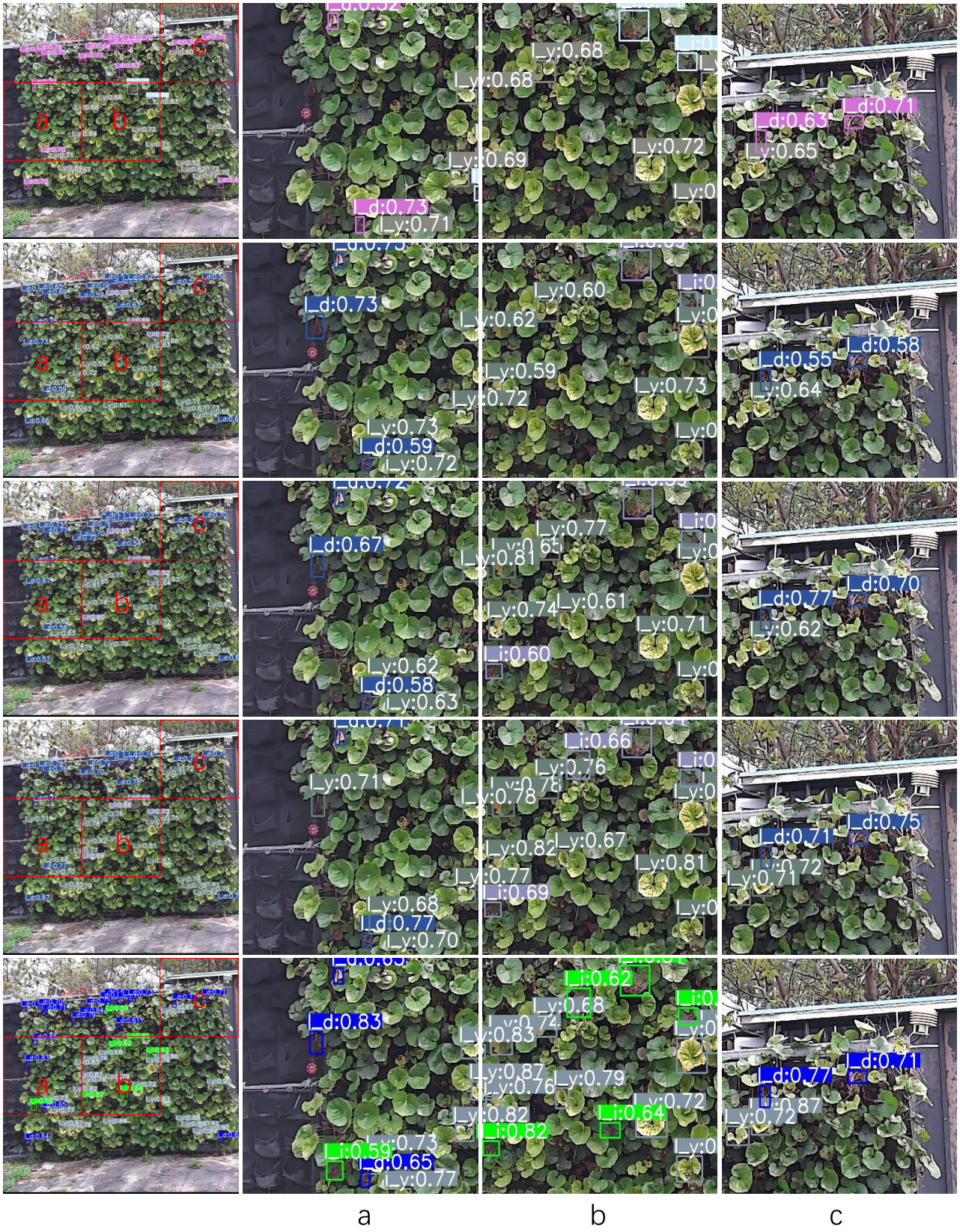


FIGURE 12. Actual detection results.The leftmost column shows the global detection results. In contrast, the right-hand columns (a), (b), and (c) show the local detection effects. Different rows indicate the detection results of the trained models for different datasets.

TABLE 2. Detection accuracy on the test set.

Dataset	Proportion of Virtual data/%	Ours @0.5:0.95/%	GAN @0.5:0.95/%
Real dataset	0	89.2	89.2
Virtual dataset	25	84.6	82.4
Mixed dataset 1	50	80.1	77.0
Mixed dataset 2	75	76.3	69.3
Mixed dataset 3	100	70.1	-

in this paper can generate a large number of virtual samples quickly and efficiently with less real samples. Comparing the detection accuracy of the six mixed datasets on the test set, the proposed method in this paper improves 2.2%, 3.1% and 7.0%, respectively. This indicates that the method proposed in this paper is more superior than the GAN network, and it is still able to construct and expand the dataset better in the lack of real data. The actual detection results of the model trained based on this method are shown in Fig.12:

Fig.12 shows the actual detection results of the trained model on the five training sets, with the first column showing the overall detection results and the three columns on the right side showing the local detection results. The training sets of the detection model from top to bottom are: virtual dataset, mixed dataset 3 (75% virtual data), mixed dataset 2 (50% virtual data), mixed dataset 1 (25% virtual data), and real dataset sequentially. It can be seen from the figure that the virtual dataset does not achieve the detection effect of the actual data, but the model trained by the virtual dataset can still accurately complete some of the detection tasks. Comparing the detection results with the mixed data, we can conclude that expanding the training set by the proposed dataset construction method is effective.

V. CONCLUSION

In this paper, we propose an augmented reality-based dataset construction method for plant growth state detection tasks due to the difficulty and high cost of data collection, the separateness of detection targets and environments, and the lack of uniform standards for manual annotation, which have the following main advantages.

1) Augmented reality technology can highly fit the growth environment as well as the shape of the plant to build a training dataset consistent with the actual detection environment.

2) Realistic mapping and physical rendering techniques are used, which can fit the plant shape contour and leaf details well. The plant sample data can be expanded and constructed efficiently by leaf blade random patchwork and sample randomization generation.

3) Reduce the cost of dataset production, the dataset generated by the virtual platform does not require artificial cultivation and manual annotation of the actual dataset, and the dataset production cycle is short and convenient.

The experimental results show that the proposed dataset construction method outperforms GAN networks and solves the problem that GAN networks cannot be used in the lack of real data. The actual detection results show that the method demonstrated in this paper can better meet the needs of non-high-precision plant detection. This method can build a high quality and large scale field dataset, and solve the problem of lack of data in the field of plant growth state detection.

REFERENCES

[1] A. Fuentes, S. Yoon, S. Kim, and D. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, Sep. 2017.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[3] M. M. Ozguven and K. Adem, "Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 535, Dec. 2019, Art. no. 122537.

[4] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, p. 1419, Sep. 2016.

[5] S. M. Saranya, R. R. Rajalaxmi, R. Prabavathi, T. S. Suganya, S. P. Mohanapriya, and T. V. G. Tamilselvi, "Deep learning techniques in tomato plant—A review," *J. Phys., Conf. Ser.*, vol. 1767, no. 1, 2021, Art. no. 012010.

[6] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Trends in vision-based machine learning techniques for plant disease identification: A systematic review," *Expert Syst. Appl.*, vol. 208, Dec. 2022, Art. no. 118117.

[7] Y. Yuan, L. Chen, Y. Ren, S. Wang, and Y. Li, "Impact of dataset on the study of crop disease image recognition," *Int. J. Agricult. Biol. Eng.*, vol. 15, no. 5, pp. 181–186, 2022.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 1–9.

[9] R. Barth, J. Hemming, and E. J. van Henten, "Improved part segmentation performance by optimising realism of synthetic images using cycle generative adversarial networks," 2018, *arXiv:1803.06301*.

[10] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[11] Y. Lu, D. Chen, E. Olaniyi, and Y.-Y. Huang, "Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review," *Comput. Electron. Agricult.*, vol. 200, Sep. 2022, Art. no. 107208.

[12] F. Zhu, M. He, and Z. Zheng, "Data augmentation using improved cDCGAN for plant vigor rating," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105603.

[13] L. Abdi and A. Meddeb, "Driver information system: A combination of augmented reality, deep learning and vehicular ad-hoc networks," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 14673–14703, Jun. 2018.

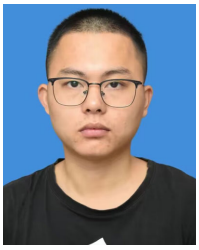
[14] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, pp. 1–14, Sep. 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[19] J. Liu and X. Wang, "Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model," *Plant Methods*, vol. 16, no. 1, pp. 1–16, Dec. 2020.
[20] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[21] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," 2019, *arXiv:1911.09516*.
[22] C. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13024–13033.
[23] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
[24] S. Xu, X. Wang, W. Lv, Q. Chang, C. Cui, K. Deng, G. Wang, Q. Dang, S. Wei, Y. Du, and B. Lai, "PP-YOLOE: An evolved version of YOLO," 2022, *arXiv:2203.16250*.
[25] M. A. Noyan, "Uncovering bias in the PlantVillage dataset," 2022, *arXiv:2206.04374*.
[26] M. Hu, D. Weng, F. Chen, and Y. Wang, "Object detecting augmented reality system," in *Proc. IEEE 20th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2020, pp. 1432–1438.
[27] S. Yang and W. Lin, "Application of augmented reality technology in smart cartoon character design and visual modeling," in *Proc. 4th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Jan. 2022, pp. 1434–1437.
[28] S. Xia, "Application of augmented reality technology in carton packaging structure design," in *Proc. 13th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA)*, Jan. 2021, pp. 9–13.



HAOHAO XING received the bachelor's degree from the Hubei Institute of Automotive Technology, in 2017. He is currently pursuing the master's degree in computer science with the Chengdu University of Technology. During his undergraduate studies, he mainly worked in the field of industrial automation and computer science and participated in nearly ten provincial and municipal projects as an assistant. During this period, he received many university and municipal scholarships. He is mainly engaged in research in the field of deep learning and computer vision.

FEI DENG received the Ph.D. degree in earth exploration and information technology from the College of Information Engineering, Chengdu University of Technology, China, in 2007. Since 2004, he has been with the College of Computer and Network Security, Chengdu University of Technology, where he is currently a Professor. His current research interests include artificial intelligence, deep learning, and computer graphics.

YUN TANG received the master's degree from the Chengdu University of Technology. Currently, he is mainly engaged in research in the fields of nonlinear computing, visualization, numerical computing, artificial neural networks, information security, and other scientific research. He has conducted several research projects, including national projects, such as the Land Survey Project of the Ministry of Land and Resources, Provincial and Ministerial Projects of the Department of Land and Resources, and several local horizontal projects. He has published more than ten scientific papers as the first author, including one EI, one ISTP, and six Chinese core journals.

QINGQING LI received the master's degree from the Landscape Architecture Department, Nanjing Agricultural University, in 2018. From 2018 to 2020, she was a Researcher with Sichuan Tianyi Ecological Garden Group Company Ltd. Since 2020, she has been the Department Manager of the Second Design Department and the Deputy Manager of the Space Green Ecosystem Research Center, Sichuan Tianyi Ecological Garden Group Company Ltd. Her research interests include space greening system research and flower border construction technology research. Her awards and honors include the Second Prize in the 2022 "Flower Brocade Official City, the Flowers Youth Season" Flower Border Competition, and the Third Prize in the 2022 First "Tianyi Cup" Micro-Landscape Competition.

JING ZHANG received the dual bachelor's degree from Sichuan University, in 2002. She is currently the Chief Architect of the Landscape Architecture Institute, China Southwest Architectural Design and Research Institute Company Ltd. She was committed to the theoretical research and practice of park city, combined with the application of landscape digitalization development direction, and promote the near-natural sustainable ecological landscape and low-cost maintenance management system. In 2019, she received two industry excellence awards from the China Survey Association. She received the Science and Technology Award of the Chinese Landscape Architecture Society, in 2020. She received the Zhan Tianyou Award, in 2021.