

Three-dimensional image recognition of soybean canopy based on improved multi-view network

Xiaodan Ma, Wenkang Xu, Haiou Guan^{*}, Xi Zhang

College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Da Qing 163319, China



ARTICLE INFO

Keywords:

Soybean canopy
Three-dimensional images
Multi-view network
Recognition models

ABSTRACT

Rapid and effective identification and diagnosis of soybean drought conditions is crucial for soybean yield and quality. Due to the complexity and diversity of agricultural environments, deep learning models based on three-dimensional data suffer from low accuracy and slow efficiency in practical applications, this paper proposes a three-dimensional image recognition method for soybean canopy based on an improved multi-view network. A lightweight network Res2net was used to reconstruct the feature extraction skeleton network in the MVCNN model, and the group convolution module of the network was optimized by embedding the ECA attention mechanism to propose a new three-dimensional image recognition model based on multi-view network (ECA-MVRes2net). In the study, drought soybeans were used as an example to obtain projected images of soybean canopy in six viewpoints using three-dimensional rotation and image feature theory, and the proposed **ECA-MVRes2net** was applied to carry out three-dimensional image recognition experiments of drought soybeans, and its recognition accuracy, F1 value and Kappa coefficient reached 96.665 %, 96.7 % and 0.924, respectively, compared with MVCNN, MVResnet, Pointnet++ and PointConv models with 3 evaluation metrics average improved by 17.289 %, 17.43 % and 0.356, respectively. The result realized a lightweight fast and accurate network model suitable for three-dimensional image recognition, which provides a theoretical foundation and technical support for the rapid recognition and accurate management of crops based on three-dimensional image processing.

1. Introduction

Soybean was widely planted in all parts of China as a grain and oil crop and was also the most important source of plant protein for human beings, which occupies an important position in China's agricultural production (He et al., 2023; Ma et al., 2017). However, at present, China's soybean production is far from being able to meet the self-supply, 80 % of soybeans originated from imports from abroad, seriously threatening the national food and oil security. In the process of soybean growth and development, water demand was large, but the drought resistance was relatively weak, and it was sensitive to the water requirements, insufficient water would adversely affect its growth and development and yield, quality, and so on. At the same time, China's agricultural irrigation water utilization rate was low, and water-saving irrigation technology was not developed (Jia et al., 2019), so using deep learning technology to detect the phenomenon of water stress in the growth process of soybeans, how to improve the efficiency of water utilization in the case of shortage of water in agriculture, but also for the

cultivation of soybeans and to ensure that soybeans were high quality and high yield was of great significance (Guo et al., 2015). The traditional way of diagnosing drought stress was mainly through the detection of water content in the soil and the use of chemical analysis (Wang et al., 2018; Yang et al., 2020), these traditional methods of drought detection were inefficient and very time-consuming, which could cause irreversible damage to the crop. Moreover, these traditional methods were complicated and inefficient in detecting drought stress in the early stage of crops and were also susceptible to the influence of human subjective factors leading to misdiagnosis (Wang et al., 2016). When plants were under drought stress, they would exhibit corresponding morphological changes, such as wilting of leaves, curling and yellowing of the leaf surface, and other morphological features (Fahad et al., 2017; Manuela et al., 2003). Therefore, by using techniques such as machine learning and deep learning to learn and mine these phenotypic morphological features of plants at a deep level (Wang et al., 2023), it is possible to realize rapid and non-destructive detection and identification of drought stress in plants based on phenotypic morphological features

* Corresponding author.

E-mail addresses: mxd@cau.edu.cn (X. Ma), bynd_xwk@163.com (W. Xu), gho@cau.edu.cn (H. Guan), bynd_zx@163.com (X. Zhang).

(Liu et al., 2023). Traditional machine learning methods were developed by extracting features such as the color and texture of plants (Wang et al., 2019), and then performing further analysis of the intrinsic connection between the features and drought stress. Yue et al. (2018) used to represent the morphological information of maize plants by means of artificially designed features and used machine learning to classify the screened features for the purpose of fast and non-destructive detection of drought in maize. Gutierrez et al. (2018) used image data acquired through thermography as input to a machine learning model as a way to detect drought stress status in grapes. Although the use of machine learning technology in the drought detection of organisms was effective, it requires human design to extract the feature values, so it is susceptible to the influence of the external environment and human subjective factors, resulting in a low accuracy rate in practical applications. At present, the generally popular deep learning technology, compared with machine learning technology, deep learning processing technology does not require human analysis and design features, can be through the way of transfer learning, to solve the small data volume sample task, in the process of practical application was more convenient, and performance was also better. Zhao et al. (2020) realized the classification of drought stress levels of tomatoes in the greenhouse by improving the Densenet169 network model. Zhu et al. (2018) proposed a high-performance method for vegetable image classification based on a deep learning framework using AlexNet convolutional neural network to train the vegetable image dataset. The results show that AlexNet accuracy was up to 92.1 %, which is a big improvement compared to the BP neural network 78 % and SVM classifier 80.5 % methods. With the continuous expansion of three-dimensional data, traditional two-dimensional deep learning cannot directly deal with these three-dimensional data. Early point cloud processing used hand-designed rules for feature extraction followed by machine learning-based classification. However, hand-designed rules extracted features with limited expressive power, especially in complex scenarios, the accuracy and generalization ability of the model could not meet the requirements of the application, and the method relied heavily on researchers with specialized knowledge and experience (Gan et al., 2015). Currently, 3D object representations can be learned by converting point clouds to mesh, voxel, or multi-view data through indirect methods, but these methods are prone to problems such as loss of 3D geometric information about the object or excessive memory consumption. Before PointNet, deep learning techniques could not directly process point clouds due to their disorder and irregularity. With the proposal of a series of 3D deep learning models such as PointNet (Qi et al., 2017), PointNet++ (Chen et al., 2021), DGCNN(Wu et al., 2018), and Piont-Conv (Wu et al., 2019), the models can deal with disordered point cloud data directly (Shin et al., 2022; Sia et al., 2023), but these 3D deep learning models require high quality of data during testing and are not suitable for generalized application. Moreover, due to the complexity and variability of the external environment in which the crop grows, this also brings troubles and difficulties to the recognition of 3D point clouds, and although some methods have been applied in outdoor scenarios (Zhou et al., 2023), the efficiency and accuracy are yet to be further improved.

Aimed at the current lack of deep learning models for three-dimensional image recognition of soybean canopy under drought stress. In this paper, a three-dimensional image recognition method for soybean canopy was proposed based on an improved multi-view convolutional network model. A multi-view convolutional network soybean drought recognition model (ECA-MVRes2net) was established by combining the idea of feature extraction-feature aggregation-feature classification in MVCNN convolutional neural network, introducing the attention mechanism module ECA (Mao et al., 2021) to improve the structure of multi-scale convolutional module of Res2net network (Gao et al., 2021), and constructing a new type of convolutional module (Econv). Effective recognition of drought soybean canopy three-dimensional image was realized, which provides theoretical

foundation and technical support for deep learning-based crop three-dimensional image drought recognition.

2. Improved multi-view convolutional network models

2.1. Constructing a novel multi-view convolutional network structure

Multi-view convolutional networks efficiently supported the task of three-dimensional object recognition by integrating input features from different viewpoints and aggregating them into a comprehensive three-dimensional feature descriptor. However, existing network architectures were not sensitive enough to the edge regions of the image when extracting features, resulting in the inability to adequately capture the edge detail features in each view. Therefore, in this paper, a multi-view convolutional network model ECA-MVRes2net based on the Res2net network was proposed to solve the problem of three-dimensional image recognition of drought soybean canopy. The feature extraction capability of MVCNN (Multi-View Convolutional Neural Network) was enhanced by the lightweight Res2net architecture, and the ECA (Efficient Channel Attention) attention mechanism was introduced to optimize the weight distribution between channels during feature extraction, ensuring that the model focuses more on the features that have the greatest impact on the final recognition result. The structure of the multi-view network model was shown in Fig. 1.

Fig. 1 showed the main structure of the multi-view convolutional network, which consists of three main parts: a feature extraction layer, a feature fusion layer, and a feature classification layer. The n independent view data were used as inputs to the model, and feature extraction was performed for each input view. After each view passes through the designed feature extraction layer, the corresponding feature vector was output, thus capturing the unique information of each view. Subsequently, these features were integrated through a feature fusion layer, which aggregates the feature vectors from different views into a unified 3D feature descriptor, retaining the key information in each view and eliminating redundant or irrelevant features in the layer. Finally, the fused features were fed into a classifier that maps the aggregated feature vector to m categories.

2.1.1. Introduction of attention mechanisms

The ability to express multi-scale feature information during network model training has a significant impact on model performance, with greater reliance on contextually informative features at different scales in less varied classification and semantic segmentation tasks. Res2net network was a new type of multi-scale network, which reconstructed its residual module on the basis of the Resnet network, set multiple channel convolution groups in the residual module, and connected them in a similar way to the residuals, which made the network fuse multiple information features on different scales in finer-grained hierarchies to improve the expression ability of the features. Therefore, this study adopted Res2net as a baseline model to construct a multi-scale feature extraction backbone network for multi-view network architecture. In order to ensure the effective identification of sample images with less differentiated soybean canopy projections and less prominent drought-stressed canopy projection features, a new type of convolution module Econv was constructed by introducing an efficient channel ECA attention mechanism after the multi-scale convolution in the Res2net block module, which enables the Res2net block module to effectively capture the interaction information between multi-scale channels, and this optimization strategy significantly enhances the model's ability to capture the detailed features on the edges while avoiding the problems brought by the reduced dimensionality.

Fig. 2 showed the Res2net residual block improve the 3×3 convolutional layers with 3×3 convolutional groups, while connecting the outputs after the different convolutional groups using hierarchical residuals. This allows the Res2net residual block to modify the field at a finer-grained layer level, thus learning deeper details and global

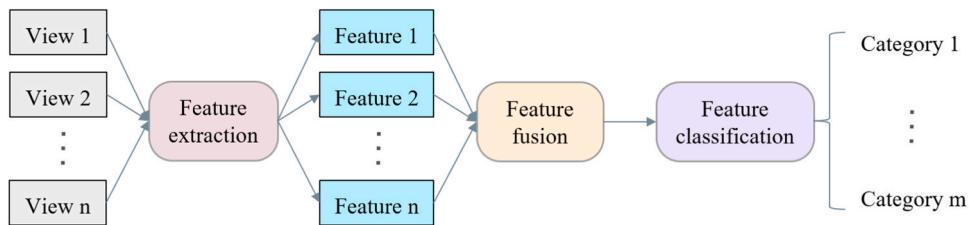


Fig. 1. Multi-view convolutional network model structure.

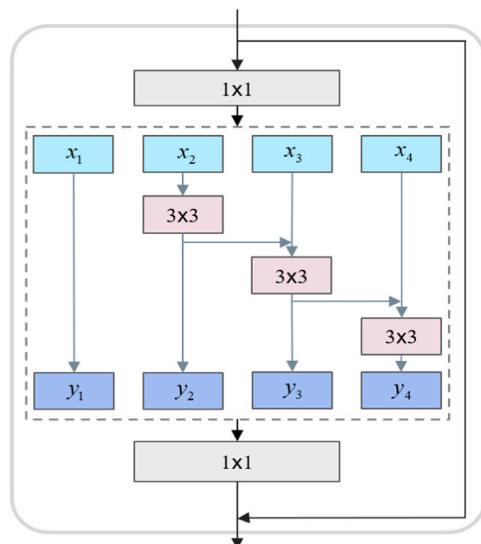


Fig. 2. Res2net block module.

features of the features and improving the performance of the model. The input features were divided into 4 groups by 1×1 convolution, while the direct mapping for x_1 can effectively reduce the number of parameters, the remaining three groups x_2, x_3 , and x_4 are processed by 3×3 convolution operation, the former group and the latter group of convolved features were connected with the residuals in the channel dimensions, and finally, the feature maps after the multiscale convolution were connected by the 1×1 convolution.

Attentional mechanisms were a method of mimicking biological attentional behavior, a technique in dealing with the efficient and accurate filtering and acquisition of critical or relevant and effective information from a large amount of information. Since there was little difference between the feature performance of soybean canopy three-dimensional image projection in the early stage of drought stress and that in normal water supply, subtle differences in drought soybean canopy projection images and normal soybean canopy projection images

would be ignored in feature extraction, leading to misrecognition. To further improve the model performance, a new convolutional module Econv, which was composed of a convolutional layer, an ECA module and a Relu function, was proposed with a new design and improvement of the Res2net model. By introducing the ECA attention mechanism, the ability to extract finer-grained feature layers from the convolutional group in the Res2net model is optimized. The new convolutional module Econv structure and ECA attention mechanism module were shown in Fig. 3.

Fig. 3(a) showed the Econv convolution module consists of a Conv convolution layer with a convolution kernel size of 3×3 , an ECA attention mechanism module, and a Relu activation function. Where the ECA attention mechanism can be expressed by the formula :

$$\omega = \sigma(W_k) \quad (1)$$

Where W_k was the channel attention band matrix with the following equation :

$$\begin{bmatrix} w^{1,1} & \dots & w^{1,k} & 0 & 0 & \dots & \dots & 0 \\ 0 & w^{2,2} & \dots & w^{2,k+1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & w^{c,c-k+1} & \dots & w^{c,c} \end{bmatrix} \quad (2)$$

The ECA attention mechanism only considered the information interaction between X and its k neighboring channels, which made the model more efficient in recognition, and this attention mechanism shared the same learning parameters for all channels during the computation, realizing a more efficient approach. The calculation formula was as follows :

$$\omega_i = \sigma\left(\sum_{j=1}^k w_j^i y_j^i\right), y_i^j \in \Omega_i^k \quad (3)$$

where Ω_i^k represents the set of k neighboring channels of y_i .

Finally, a one-dimensional convolution with a convolution kernel size of 1×1 is used for the connection of multi-scale features. The computational formula was as follows :

$$\omega = \sigma(C1D_k(y)) \quad (4)$$

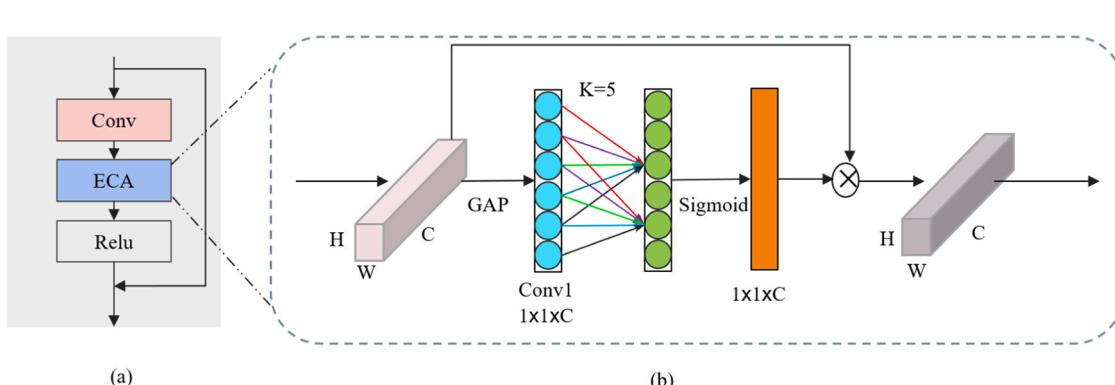


Fig. 3. Econv module and ECA module structure.

where C1DW was the 1-dimensional convolution and $k = 5$ was the number of parameters in the module. The performance and efficiency were ensured by 1-dimensional convolution, and the complexity of the model was also reduced.

The ECA attention mechanism module was used to reconfigure the 3×3 convolutional group in the Res2net block module structure to construct a new convolutional module Econv. The structure of the Res2net block module before and after the introduced ECA attention mechanism was shown in Fig. 4.

Fig. 4(a) showed the structure of the Res2net block module before improvement. Fig. 4(b) showed the structure of the Res2net block module after the ECA attention mechanism was introduced to improve the Res2net block module. The ECA attention mechanism has scanned the input image, filtered the features of the region of interest, and automatically redistributed the weights of each channel to increase the attention to the region of interest and to expand the global sensory field of the feature extraction network. This method was able to improve the accuracy of recognizing the three-dimensional images of drought-prone soybean canopies as well as save a lot of computational resources.

2.1.2. Novel multi-view convolutional networks

In this paper, the multi-scale feature extraction network Res2net was used as the backbone extraction network, and the ECA attention mechanism module was introduced in the multi-scale group convolution module in Res2net block to construct a new multi-view convolutional network ECA-MVRes2net. The network structure is shown in Fig. 5.

Fig. 5(a) showed the overall structure of the ECA-MVRes2net model including the input layer, the conv block module, the Layer (Layer1, Layer2, Layer3, Layer4) network layer, the view pooling layer, and the fully connected layer. The input image was first processed through a convolutional and pooling layer of size 7×7 and then through Layer1, Layer2, Layer3, and Layer4 network layers. The Layer network layer was composed of multiple Res2net block stacks, the structure of which was shown in Fig. 5(b). In the Res2net block, the feature map was partitioned into different subsets (x_1, x_2, x_3, x_4), which were successively subjected to different convolution operations, where the subsequent subsets would receive the outputs of the previous subsets as additional inputs for

feature accumulation and aggregation. Without adding additional parameters and computational complexity, it was achieved to capture richer scale information within the same receptive field, thus effectively enhancing the ability of feature expression. After being processed by the layer, the features were transported to the View Pooling layer (Fig. 5(c)) for feature fusion, where the features extracted from multiple views were aggregated into a set of features, and the feature maps were converted into vectors through the global average pooling layer. Finally, the final classification result was output through the fully connected layer and softmax layer.

2.2. Learning algorithms

In this paper, the Adaptive learning rate optimizer (Adam) was selected for regulating and optimizing the parameters of the deep learning network model, which integrates the advantages of the momentum gradient descent algorithm and RMSProp algorithm, and the core idea of Adam's algorithm was to adaptively and dynamically adjust the learning rate of the model by calculating the first-order matrix estimation of the parameters and second-order matrix estimation, to adapt to the different parameter variations and to improve the effectiveness of the model training. The steps for the Adam optimizer to update the network parameters were shown in Table 1.

3. Model simulation and analysis of results

3.1. Simulation experiment flow

The basic recognition process of the model on a three-dimensional image of the soybean canopy was shown in Fig. 6. Self-collected three-dimensional images of soybean canopies were used as objects. First, conditional filtering, threshold segmentation, statistical filtering, data enhancement, and three-dimensional spatial transformation were used for preprocessing operations. Then, the Res2net network was optimized into the MVCNN network to improve its feature extraction structure, and the attention mechanism was introduced to establish a novel multi-view convolutional model adapted to soybean canopy 3D image recognition

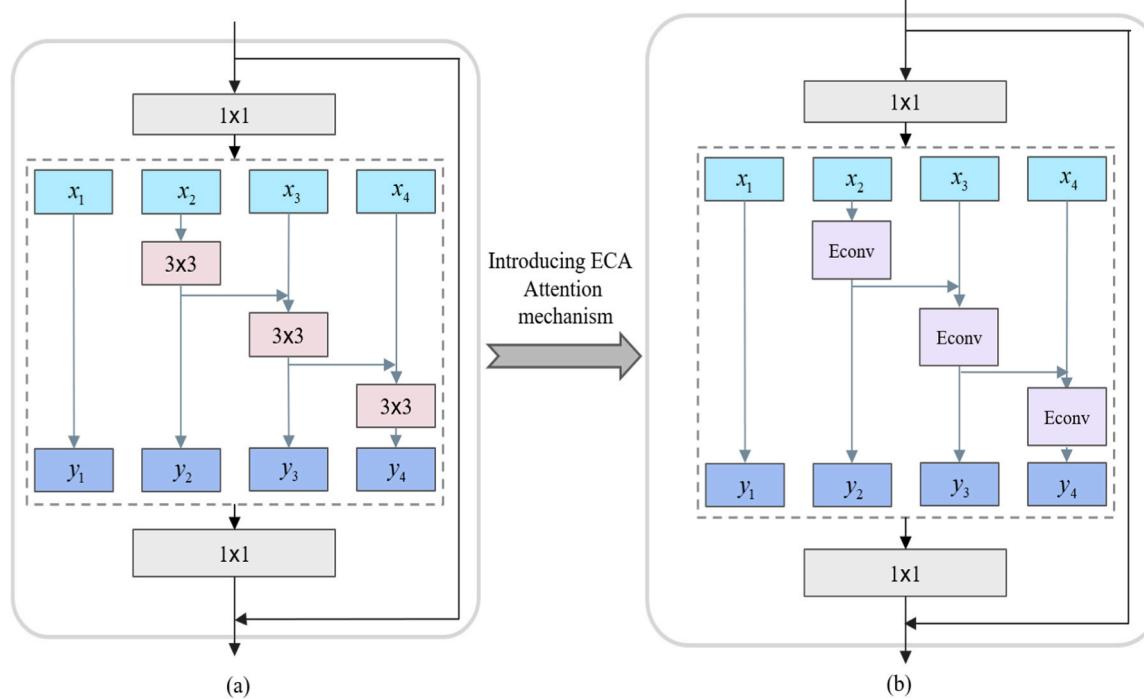


Fig. 4. Comparison of ECA module before and after embedding Res2net block.

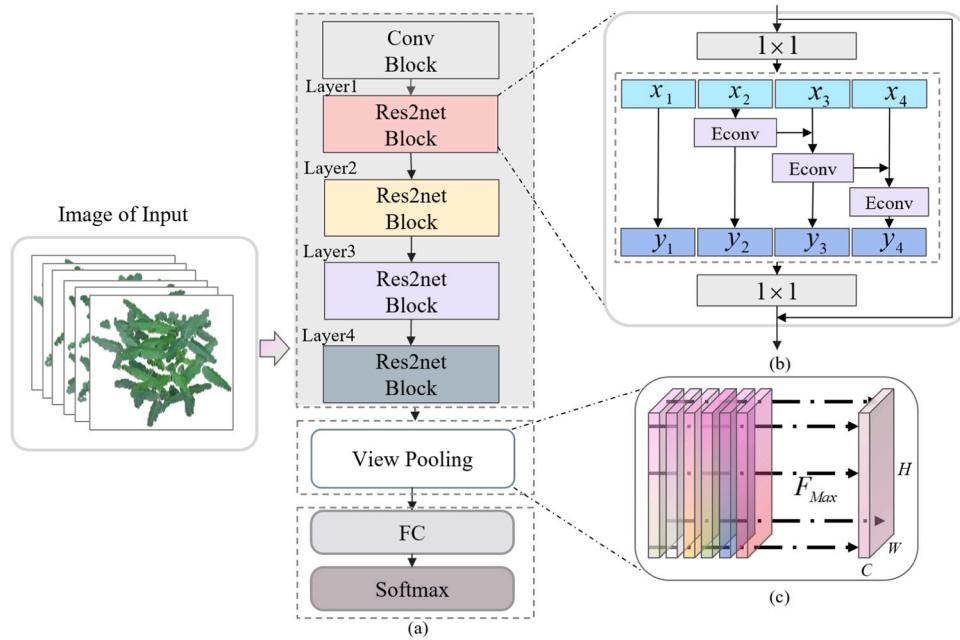


Fig. 5. ECA-MVRes2net network structure.

Table 1

Adam optimization algorithm process.

Input: $\alpha, \beta_1, \beta_2 \in [0,1], \theta_0, f(\theta)$

Initialize: $m_0 \leftarrow 0$ (*first moment*), $v_0 \leftarrow 0$ (*second moment*), $t \leftarrow 0$

while θ_t not converged **do**

- $t \leftarrow t + 1$
- $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ $\triangleright g_t$ Represents the gradient of time step
- $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ $\triangleright m_t$ Represents first-order momentum
- $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ $\triangleright v_t$ Represents second-order momentum
- $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$
- $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$
- $\theta_t \leftarrow \theta_{t-1} - \alpha \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$

end while

return θ_t

(ECA-MVRes2net). Finally, simulation validation was performed, where the training and validation sets were input to the model for training and parameter tuning, and then, classification performance was tested on the test set.

3.2. Data preprocessing

The experiment was carried out at Bayi Agricultural Reclamation University, Daqing City, Heilongjiang Province (Fig. 7), and Suinong 26 (Glycine max (L.) Merr.) from the Suihua Branch of the Heilongjiang

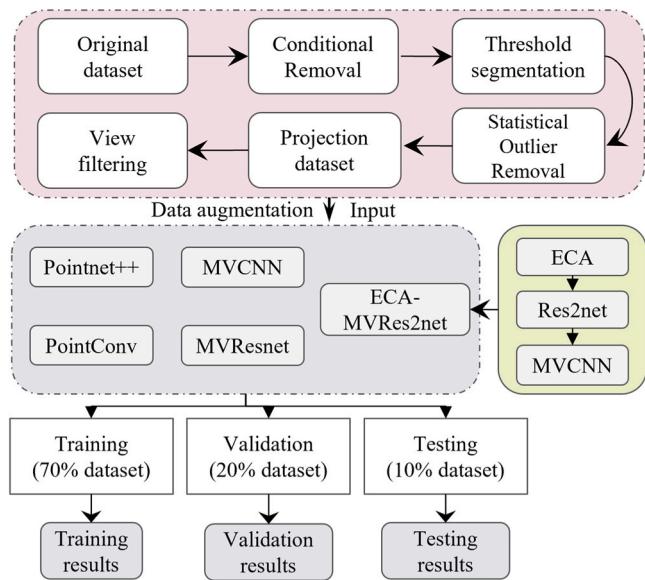


Fig. 6. Simulation experiment flow.

Academy of Agricultural Sciences was selected as the research subject, aiming to study the effect of drought stress on soybean canopy morphology. The experiment was started in late May of each year and grown using the potting method at outdoor temperatures ranging from 20 to 35°C. After soybean germination and seedling emergence, each potted plant needs to be fixed to achieve precise water control. This was done by selecting two seedlings in each pot with the closest growth rate and condition and removing the others to ensure the consistency and effectiveness of the drought treatment. The drought treatment was then initiated by gradually reducing the water given so that the soil moisture content of each pot was slowly reduced to the established drought level. External factors such as the growing environment and treatments should be consistent when growing and managing soybeans to ensure the

scientific rigor and validity of the experiment. In this study, a total of 24 pots of soybean plants were selected as experimental subjects, 12 pots of normal water supply and 12 pots of drought stress treated soybean plants, in which the soil water content of normal water supply soybean was maintained at 75 % and that of drought stress soybean was maintained between 50 % and 75 %, including the early and middle stages of drought stress. A total of 240 sets of data were collected in five periods, from the seedling stage (V2), meristematic tissue stage (V3, V4, V5) to the early stage of reproductive growth (V6), using the constructed three-dimensional image dynamic acquisition system based on Kinect V2 camera. Agronomically, it was able to meet the experimental requirements of the control group (Zou et al., 2019; Dong et al., 2019) and also reflect the physiological and morphological changes of soybean plants under different treatments of normal water supply and drought stress. In the growth cycle of soybean, stages V2 through V6 were extremely critical, covering from the seedling stage (V2), meristematic stage (V3, V4, V5) to the early stage of reproductive growth (V6). Therefore, three-dimensional image data acquisition of soybean canopy under drought stress and normal water supply conditions during these critical periods was essential to understanding the effects of water stress on soybean growth and development. In this experiment, a total of 240 sets of 3D image data of soybean canopy under drought stress and normal water supply were collected from the growth period of soybean from V2 to V6.

The comprehensive data set on growth periods allows the model to identify and analyze the specific response patterns to drought stress at various stages of soybean development. This enables more precise prediction and recognition of drought conditions across different growth states. The integration of data encompassing a range of drought levels enhances the model's comprehension and resilience to diverse stress intensities. Furthermore, this approach strengthens the model's capacity to adapt to complex environmental changes and facilitates broader generalization and portability when presented with novel datasets or real-world scenarios.

3.2.1. Canopy extraction

In this experiment, a Kinect V2.0 camera was used to collect depth

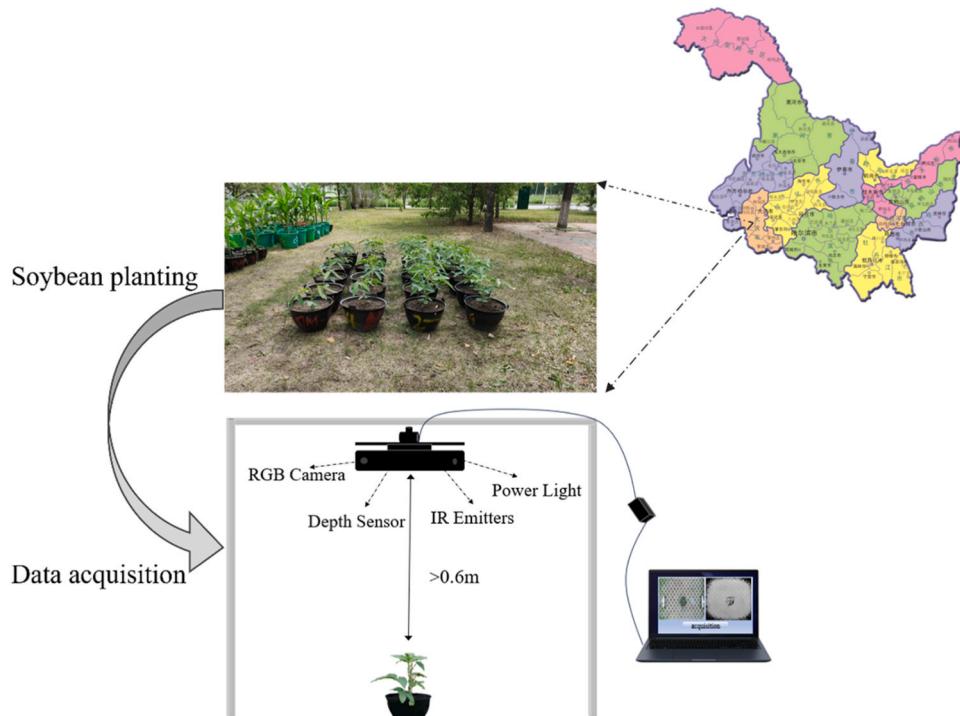


Fig. 7. Soybean planting and data acquisition.

images and color images of soybean plants, and discrete point cloud data were generated by establishing a mapping relationship between the obtained color image information and depth space information through the mapping function contained in the camera. Soybean plants were reconstructed based on the generated point cloud, and the reconstructed point cloud data of soybean with color information included spatial location, density, texture, and color (Yang et al., 2019). When collecting the canopy data of soybeans, the vertical shooting method was used, so while acquiring the soybean canopy point cloud, a large number of noisy spots on the ground, weeds, etc., which are not related to the soybean plants, were acquired, which caused a large impact on the recognition of the subsequent three-dimensional image, and thus the original soybean point cloud data need to be pre-processed by the operation. In this paper, conditional filtering, Otsu threshold segmentation and statistical filtering methods were used to streamline and denoise the raw soybean point cloud data (Zhu et al., 2023; Bhandari et al., 2020; Yang et al., 2020). The processing flow and effect were shown in Fig. 8.

Fig. 8 showed a large number of background point clouds and discrete noise points in the original soybean point cloud data. To obtain a complete soybean canopy, the first choice was to use conditional filtering to remove a large number of ground background point clouds, but there were some stray color point clouds inside the acquired canopy point cloud data. Secondly, to obtain more complete soybean canopy point cloud data, the Otsu threshold segmentation method based on the ultra-green index (ExG) was used to remove the stray color point cloud inside the canopy point cloud data. Finally, there were isolated noise point clouds in the soybean canopy point cloud after threshold segmentation, and to further smooth the morphological structure of the soybean canopy point cloud, a statistical filtering method was used to realize the removal of discrete noise points.

To evaluate the effectiveness of soybean canopy extraction and denoising, the noise point cloud recognition accuracy, target point cloud recognition accuracy, denoising accuracy (Guo et al., 2023), and streamlining rate (Ge et al., 2012) were used in this study as the evaluation indexes of data preprocessing.

Noise recognized accuracy P_n represents the ratio of the noise points

of soybean plants eliminated after preprocessing to the total noise points of soybean plants. In the conditional filtering and Otsu threshold segmentation of the point cloud, the target soybean canopy point cloud will be stored incorrectly segmented, and the higher the noise recognition accuracy represents the stronger ability of the algorithm to reject the noise points, and the higher the noise recognition accuracy, the better the ability of the algorithm. The formula was defined as follows :

$$P_n = \frac{N_T}{N_F + N_T} \quad (5)$$

where N_T represents the number of noise points correctly removed during preprocessing, and N_F represented the number of targeted soybean canopy point clouds removed.

The target point cloud recognition accuracy P_r represents the ratio of the number of target soybean canopy point clouds retained after the preprocessing operation to the number of complete soybean canopy point clouds. The higher target point cloud recognition accuracy represents the better ability of the algorithm to recognize the target soybean canopy point cloud. The formula was defined as follows :

$$P_r = \frac{R_T}{R_F + R_T} \quad (6)$$

where R_T represents the final retained point cloud of the target soybean canopy and R_F represents the complete point cloud of the target soybean canopy.

The denoising accuracy P represents the combined evaluated value of noise identification and target point cloud identification during the preprocessing of soybean plant point cloud data. The formula was defined as follows :

$$P = \frac{N_T + R_T}{N_T + N_F + R_T + R_F} \quad (7)$$

The streamlining rate R represents the ratio of the number of point clouds after preprocessing to the number of original point clouds. The higher streamlining ratio of the point cloud means that fewer point clouds were retained after preprocessing, which can effectively reduce

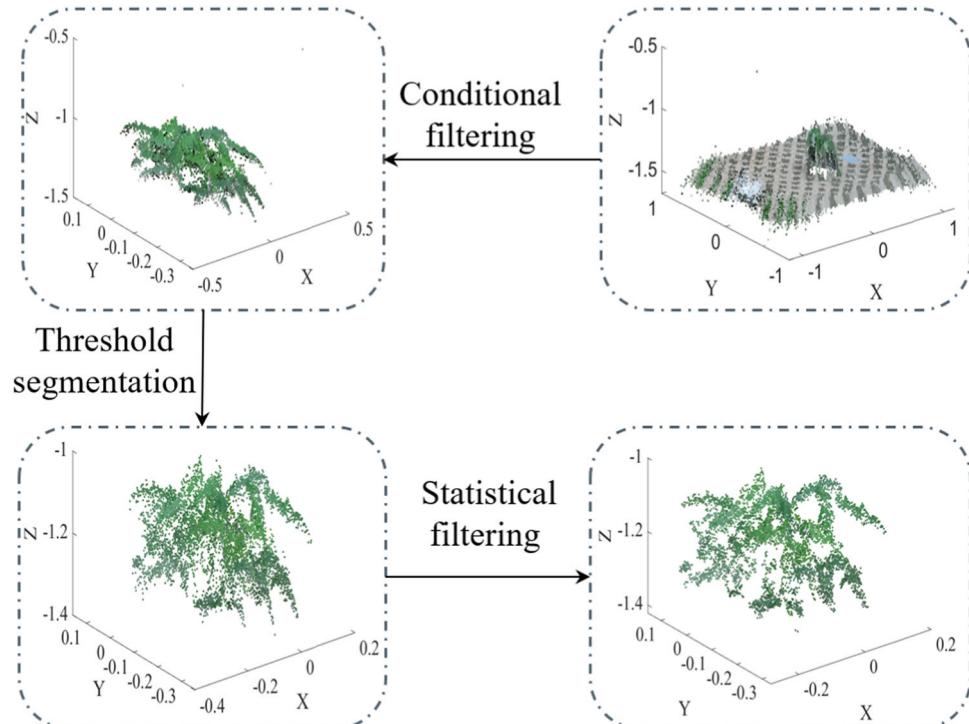


Fig. 8. Effect of pretreatment process.

the number of point clouds, thus improving the speed and reducing the computation amount of the point cloud processing. The formula was as follows :

$$R = \frac{TP}{TR} \quad (8)$$

Where TP represents the number of original point clouds without pre-processing, and TR represents the number of target soybean canopy point clouds extracted after data preprocessing. For the four evaluation indexes P_n, P_r, P and R (Table 2), the higher value indicates the better denoising of the soybean plant point cloud and segmentation of the target soybean canopy, and the evaluation indexes were shown in Table 2.

Table 2 showed the raw point cloud data were removed from the complex background and outlier points after the data preprocessing method, and a more complete soybean canopy point cloud was obtained, and the extracted drought stress and normal water supply soybean canopy point cloud achieved an average of more than 95 % in terms of the noise point cloud recognition accuracy, target point cloud recognition accuracy, denoising accuracy, and streamlining rate, which was able to satisfy the needs of this study.

3.2.2. Canopy multi-view projection

Three-dimensional models have diverse formats and complex representations, and the amount of computation will increase exponentially if three-dimensional models are directly used as objects for deep learning. To improve the computational efficiency and reduce the complexity, it is first necessary to process the 3D models to reduce their dimensionality, so that they become 2D images with a uniform format and easy to process. The operation of dimensionality reduction makes the processing objects more unified and allows for better portability of the objects to convolutional neural networks, which have excellent performance in the field of two-dimensional image processing (Dong et al., 2020). Capturing the characteristics of the soybean canopy from different perspectives resulted in a more accurate characterization of the subtle changes in the morphology and structure of the soybean three-dimensional canopy during the early reproductive process, and it was easier to differentiate between the drought-stressed soybean three-dimensional canopy image and the three-dimensional image of the soybean canopy with normal water supply. The spatial transformation of the soybean canopy point cloud was realized by three-dimensional matrix rotation, and the two-dimensional projection image of the rotated soybean canopy in the xoy plane was obtained. The principle was shown in Fig. 9.

The pre-processed soybean canopy point cloud was first rotated counterclockwise by 30° with the z-axis as the rotation axis, and the formula was as follows :

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos 30\theta & -\sin 30\theta & 0 & 0 \\ \sin 30\theta & \cos 30\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (9)$$

Where (x', y', z') were the coordinates of the point cloud after each clockwise rotation around the z-axis, where $\theta \in [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]$.

The point cloud after each rotation around the z-axis was then rotated by $30\alpha^\circ$ clockwise with the x-axis as the axis of rotation, the

Table 2
Evaluation of soybean canopy point cloud extraction.

| Categories | Evaluation index | | | |
|------------|------------------|---------|---------|---------|
| | P_n | P_r | P | R |
| Drought | 97.53 % | 98.74 % | 98.62 % | 95.36 % |
| Nomal | 97.49 % | 99.17 % | 98.63 % | 97.73 % |

equation was as follows :

$$\begin{bmatrix} x'' \\ y'' \\ z'' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 30\alpha & -\sin 30\alpha & 0 \\ 0 & \sin 30\alpha & \cos 30\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \quad (10)$$

Where (x'', y'', z'') represents the three-dimensional coordinates of the point transformed by rotating counterclockwise around the z-axis and then clockwise around the x-axis, $\alpha \in [1, 2]$. After the above steps, a total of 24 point cloud transformed projection coordinates of the soybean canopy at different angles were acquired as shown in Figs. 10 and 11, and the xoy plane projection image of the soybean canopy was acquired for each rotation as shown in Fig. 12.

Fig. 12(1)~(12) showed the planar projection images after 12 different angles of $30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ, 210^\circ, 240^\circ, 270^\circ, 300^\circ, 330^\circ$, and 360° were rotated counterclockwise around the z-axis under the condition of $\alpha = 1$, respectively. Fig. 12(13)~(24) showed the planar projection images after rotating the above 12 angles counterclockwise around the z-axis, under the condition of $\alpha = 2$, respectively.

3.2.3. Filtering critical view projection

Although the three-dimensional point cloud data were reduced by projection to greatly reduce the complexity of the model and improve the accuracy of recognition to a certain extent (Wang et al., 2019). However, too many images directly as the input of the model have the problem of data information redundancy and will consume a large amount of computational and storage resources. Therefore, in this paper, five metrics including Entropy, Edge Information, Contrast, Homogeneity, and Dissimilarity were selected from the information entropy, edge information, and Gray Level Coevolution Matrix (GLCM) of the images for the key view optimization of the obtained 24 projected images to maximize the retention of the image containing the most geometric feature information while reducing the input images to the model. The principles and formulas for the five indicators were as follows :

(1) Entropy was a measure of the amount of information in the image, reflecting the complexity of the image content. In image processing, the higher the entropy, the richer the effective soybean canopy information contained in the image. The formula was as follows :

$$\text{Entropy} = - \sum_{i=0}^{L-1} p(i) \log_2 p(i) \quad (11)$$

Where L represents the gray level of the image and $p(i)$ represents the probability of occurrence of gray level (i).

(2) Edge Information represents the characteristics of edges in an image. It was a basic feature in image analysis. The edge was the place in the image where the brightness change was obvious, and it was the key area to separate the object from the background. Edge detection was commonly used in tasks such as image segmentation, feature extraction, and image recognition. The formula was as follows :

$$\text{Edge} = \sum_{i=1}^m \sum_{j=1}^n \text{Canny}(P, x, X)_{ij} > 0 \quad (12)$$

Where P was the input image, x and X were the low and high thresholds for dual threshold detection respectively.

(3) Contrast was an important index to evaluate the brightness of the image difference, to a certain extent, can reflect the soybean canopy projection image texture characteristics of the trend. The higher the contrast, the clearer the color level of the image, which will appear clear and three-dimensional, the formula was as follows.

$$\text{GLCM}(i, j) = \sum_{p=1}^W \sum_{q=1}^H \begin{cases} 1, \text{if } I(p, q) = i \text{ and } I(p + Vx, q + Vy) = j \\ 0, \text{otherwise} \end{cases} \quad (13)$$

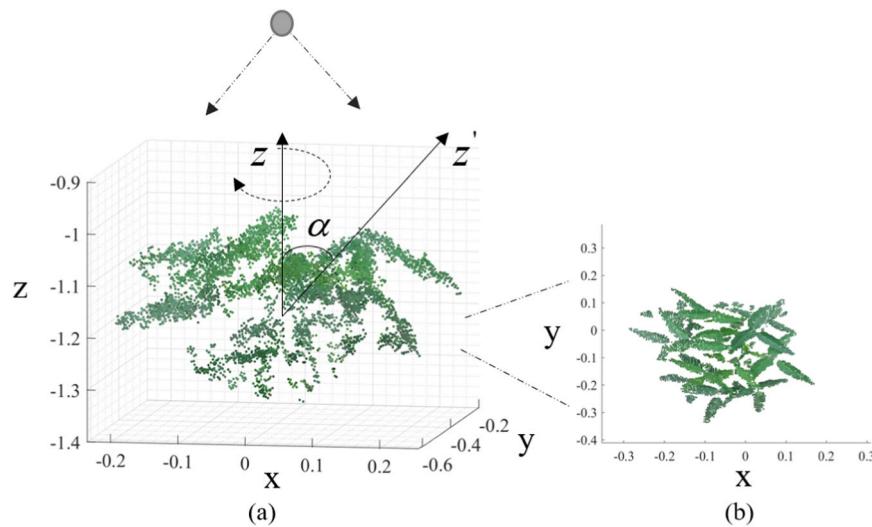


Fig. 9. Schematic of soybean canopy rotation.

Where W, H was the width and height of the image, respectively, and $I(p, q)$ represents the gray value at the position (p, q) . V_x, V_y were the offsets in the direction of the x, y coordinate axes, respectively.

$$\text{Contrast} = \sum_{ij=0}^{L-1} (i-j)^2 GLCM(i,j) \quad (14)$$

(4) Homogeneity can be used as an index to evaluate the variation of pixel values inside the image, and to some extent, it can reflect the variability of soybean canopy projection images among different gray levels. The higher the homogeneity, the smaller the change of pixel value varied within the image, and the color or grayscale of the image changed gently, the calculation formula was as follows.

$$\text{Homogeneity} = \sum_{ij=0}^{L-1} \frac{GLCM(i,j)}{1 + (i-j)^2} \quad (15)$$

(5) Dissimilarity was the degree of variation of pixel values in a localized area. The higher the dissimilarity, the greater the variation of pixel values in this local area, indicating that the projected image of the soybean canopy was richer in texture features in the local area, which was calculated by the formula as follows.

$$\text{Dissimilarity} = \sum |i-j| GLCM(i,j) \quad (16)$$

In order to eliminate the effect of magnitude between different features, the calculated image features were selected to be normalized. The calculation formula was as follows.

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (17)$$

The entropy, edge information, contrast, homogeneity, and dissimilarity of the multi-view projected image of the soybean canopy can be calculated according to Eqs. (11) to (17), and the results of the evaluation indexes of the projected image of each view were shown in Fig. 13.

Fig. 13 showed the 5-dimensional feature variation curves of entropy, edge information, contrast, homogeneity, and dissimilarity of the projected images of different views of the soybean canopy. Among them, entropy was an important index to indicate the information content in the image, as can be seen in Fig. 13, the entropy values of view projection images 3, 8, 9, 10, 16, 17, and 20 were the largest, and the edge information, contrast and dissimilarity of these seven view projection images were mostly at the peak point, so these seven view projections were prioritized in the preferred projected images. Since the homogeneity (Homogeneity) value of view 16 was higher than the homogeneity

of view projection 20, which indicates that the texture features of the projected image 20 are more prominent, the view projections 3, 8, 9, 10, 17 and 20 were selected in the preferred view projection, as shown in Fig. 14.

3.2.4. Data extension and segmentation

Data augmentation helps to solve the problem of small data samples and improves the recognition ability of the model (Kusrini et al., 2020). Augmenting the original data, allows the model to be exposed to more data in different states during the training process, which leads to better learning of universal patterns and features in the data. For all 240 data samples, the canopy extraction and view screening methods described above were used for canopy projection key view optimization, and the canopy projection images of the same data sample with 6 different viewpoints were a group, and based on the data enhancement method in literature (Liu et al., 2020; Zhao et al., 2020), the data samples obtained were expanded to obtain 1440 projection images in groups, and divided into training set (1008 groups), validation set (288 groups) and test set (144 groups) in the ratio of 7:2:1. An example of the projected soybean canopy after data enhancement was shown in Fig. 15.

3.3. Model simulation experiment

3.3.1. Analysis of training results

In this paper, five different networks were selected for experiments, namely MVCNN, MVResnet, PointNet++, PointConv, and ECA-MVRes2net models. The specific hyperparameter settings were shown in Table 3.

The training was performed on the divided training set and validation set, and the change in the accuracy of the five network models on the training set and validation set was shown in Fig. 16.

Fig. 16: The accuracy curve of MVResnet model on the training set varied from 0.8218 to 0.9238, and began to have a convergence trend at Epoch=24, and the final training accuracy was 92.38 %, and the accuracy curve on the validation set varied from 0.8362 to 0.9187, and the final validation accuracy was 91.87 %; The MVCNN model's training accuracy curve on the training set varied from 0.7365 to 0.8362, with a convergence trend starting at Epoch=16, and the final training accuracy was 83.62 %, and the accuracy curve on the validation set varied from 0.7012 to 0.8306, with a final validation set accuracy of 83.06 %; the Pointnet++ model The accuracy curve on the training set varied from 0.4186 to 0.8424, with a convergence trend starting at Epoch=19, and the final training accuracy was 81.24 %, and the accuracy curve on the validation set varied from 0.3457 to 0.6012, with a final validation set accuracy of 60.12 %.

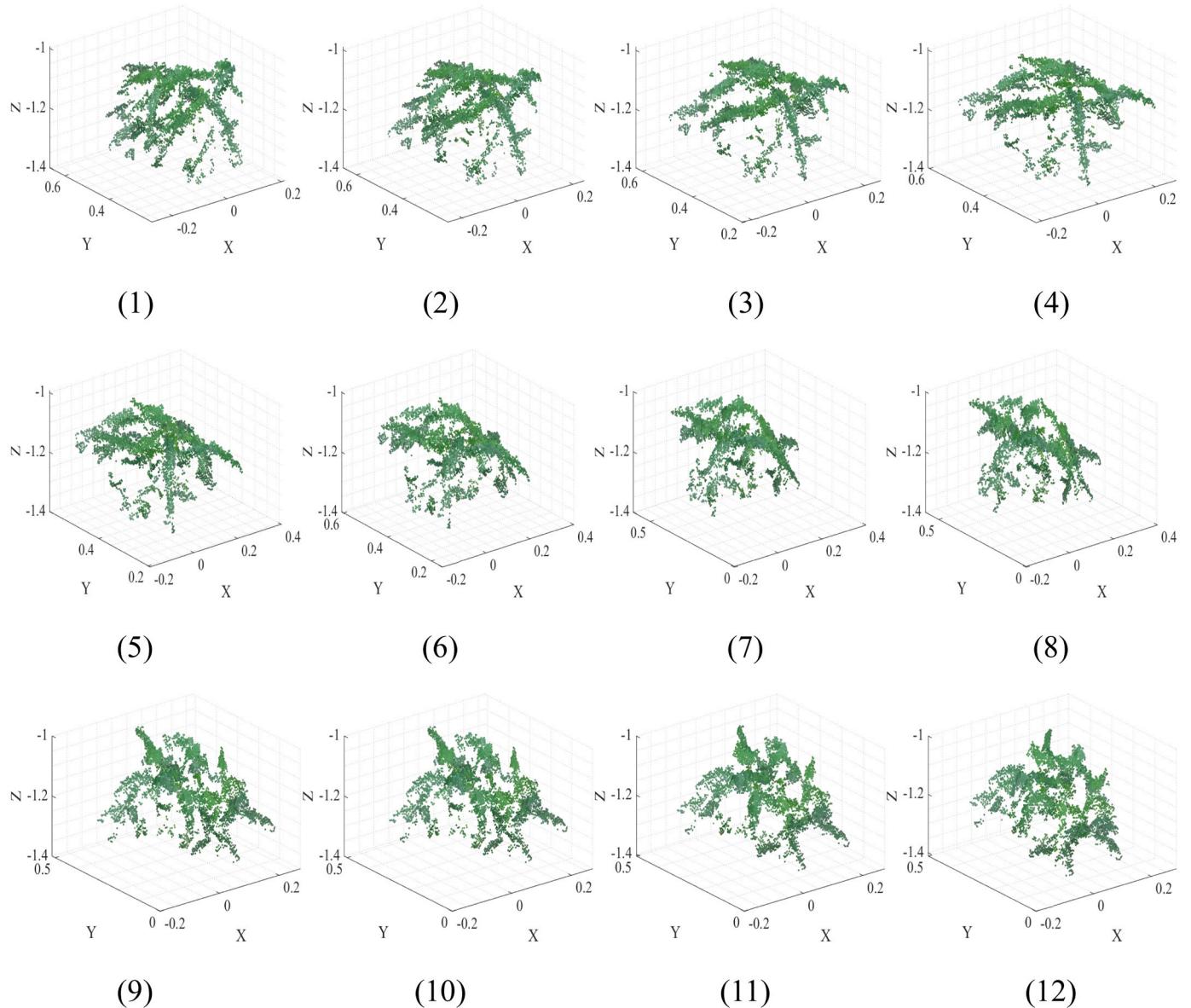


Fig. 10. The effect of rotating around the z-axis by 12 different angles when $\alpha = 1$.

accuracy of 60.13 %; the accuracy curve of the PointCov model on the training set varied from 0.3641 to 0.7835, with a convergence trend at Epoch=93, and the final training accuracy was 78.35 %; the accuracy curve of ECA-MVRes2net model proposed in this paper on the training set varied from 0.8826 to 0.9621, with a convergence trend at Epoch=18, and the final training accuracy was 96.2 %, the accuracy curve on the validation set varied from 0.8823 to 0.9613, and the final validation set accuracy was 96.1 %. The higher accuracy rate of the model during the training process indicates that the model was recognized better, the smaller variation indicates that the model was more stable, and the smaller Epoch number indicates that the model converged faster in the training set. From the change curves of the accuracy rate in the training and validation sets in Fig. 16(a) and (b), it can be seen that the ECA-MVRes2net model proposed in this paper has the highest recognition accuracy and the strongest stability.

Based on the above analysis of the changed curves of training and validation accuracy in Fig. 16, the optimal network model in training can be comprehensively evaluated as the SE-MVCNN network. For Pointnet++ and PointConv, the two 3D deep learning models did not

perform well during training, as can be seen in Fig. 16, there was an overfitting problem in the Pointnet++ model and a low recognition accuracy in the PointConv model, which was due to the non-uniform distribution of the soybean canopy point cloud data collected through the Kinect camera, and the low recognition accuracy of the vertical soybean canopy point cloud collected through The vertical soybean canopy point cloud acquired by the top view method was characterized by near dense and far sparse, and there was a leaf occlusion problem, which made the acquired point cloud data of poor quality and led to the underfitting problem when the model was trained. The results showed that the soybean canopy 3D image recognition model ECA-MVRes2net proposed in this paper achieved the best accuracy of 96.2 %. In this paper, the Res2net network was used as the skeleton network for feature extraction, which was more effective in recognizing small target features, and introduced the ECA attention mechanism, which further enhances the expression of fine-grained feature information of multi-scale features, highlighting the differences between drought stress and normal water supply soybean canopies.

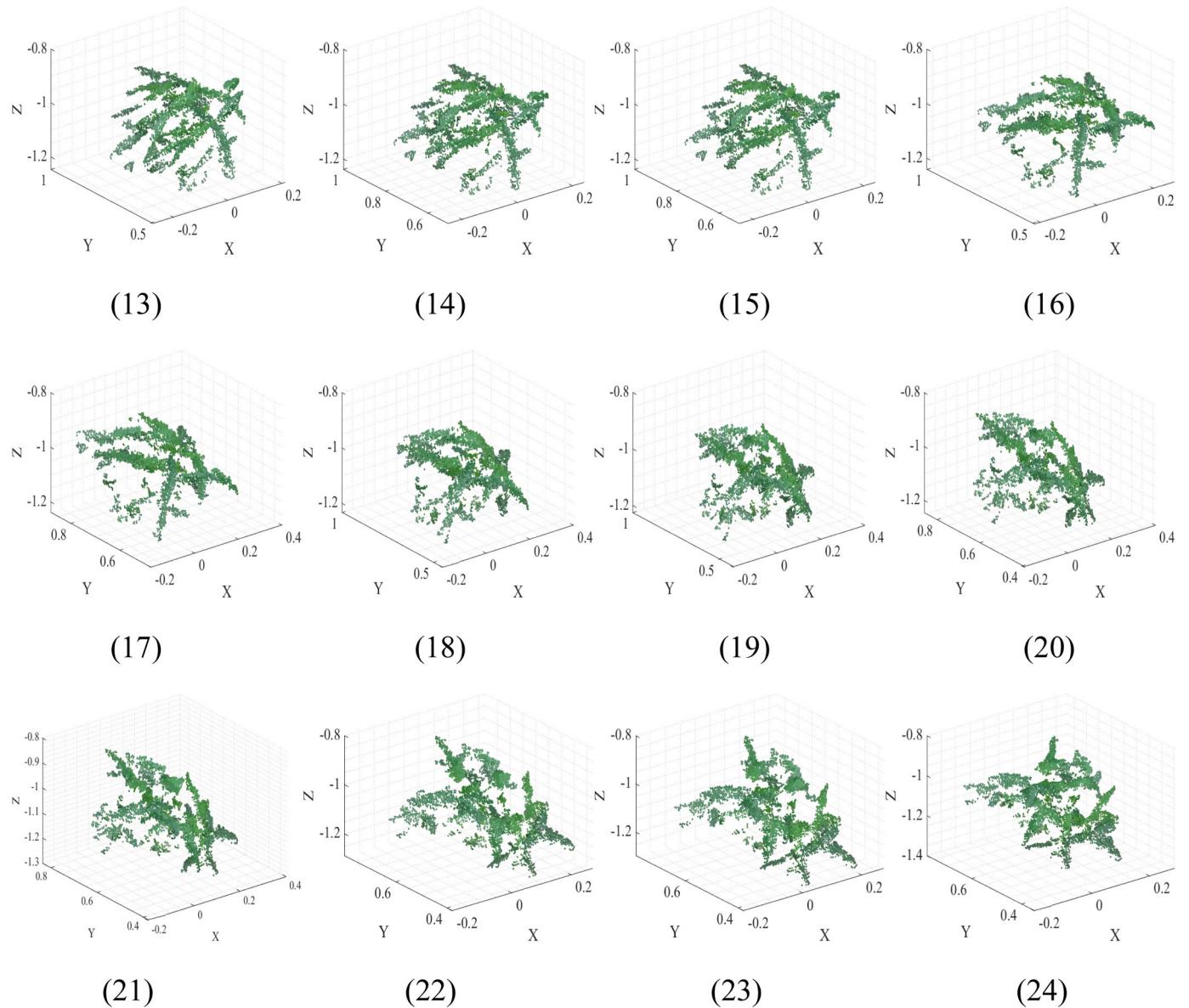


Fig. 11. The effect of rotating around the z-axis by 12 different angles when $\alpha = 2$.

3.3.2. Analysis of test results

In this paper, a total of 864 images in 144 groups each of drought-stressed soybean and normal water-supplied soybean canopy projection images were selected as the test set of the model, and precision, recall, F1 value, Kappa coefficient, and average accuracy (OA) were used as important indexes to evaluate the recognition performance of the model. The calculation formula was as follows.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (19)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

Where TP was the number of samples where the prediction results correctly identified drought soybeans, FP was the number of samples where the prediction results incorrectly identified normal soybeans as drought soybeans, and FN was the number of samples where the pre-

diction results incorrectly identified drought soybeans as normal soybeans.

The Kappa coefficient was a measure of classification accuracy (Schieck et al., 2023). It was obtained by multiplying the total number of pixels in all surface-true classifications by the sum of the diagonal of the confusion matrix, subtracting the result of summing the product of the total number of surface-true pixels in a particular category and the total number of classified pixels in that category overall categories and dividing by the square of the total number of pixels minus the result of summing the product of the total number of surface-true pixels in a particular category and the total number of classified pixels in that category overall categories. In the field of machine learning, the Kappa coefficient was used to evaluate the performance of classification models, providing a more comprehensive performance evaluation than simple accuracy. The formula was calculated as follows.

$$k = \frac{P_0 - p_e}{1 - p_e} \quad (21)$$

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_n \times b_n}{N \times N} \quad (22)$$

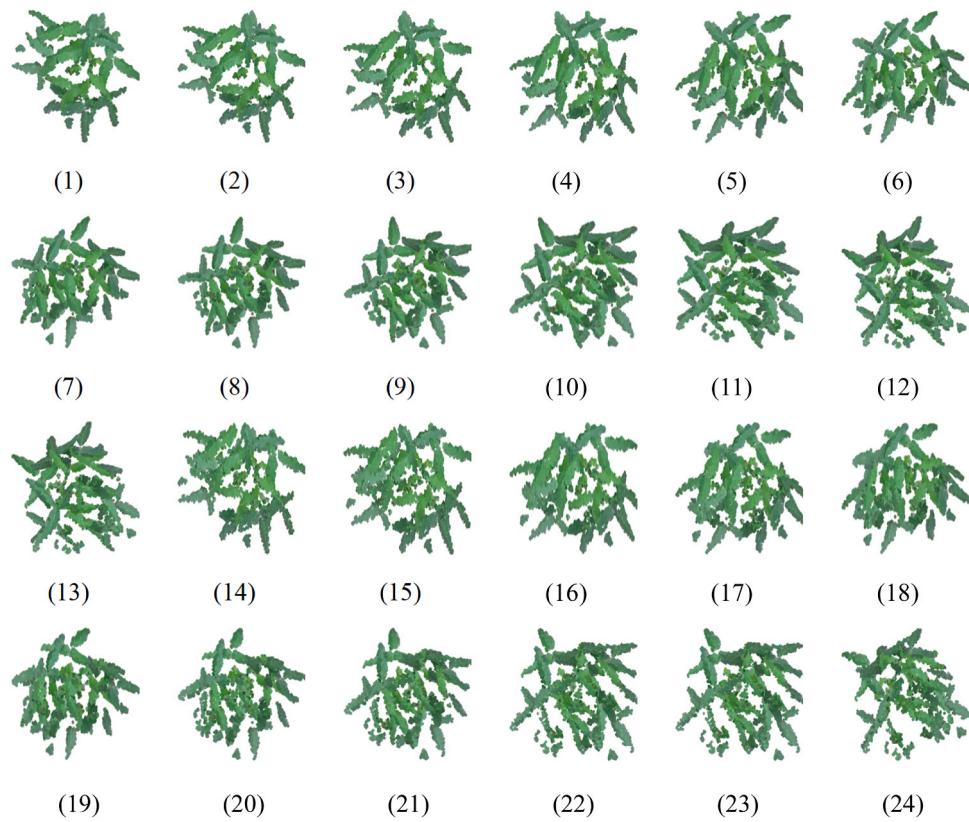


Fig. 12. 24-view canopy projection.

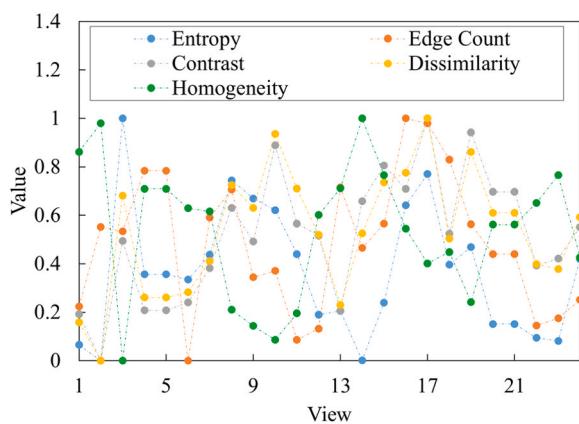


Fig. 13. 5-dimensional features of the projected images of different views.

Where p_e represented the kappa value, p_0 represented the ratio of the number of correctly identified samples to the total number of samples, a_i represented the number of true samples in each category, b_i represented the number of samples predicted for each category, and N represented the total number of samples, and the higher kappa coefficient represented the better the recognition performance of the model.

According to Eq. (18)~Eq. (22), the value of each evaluation index of the model can be calculated, and the performance results of different models on the testing set were shown in Table 4.

Table 4 showed that in terms of the recall and precision of the model, the ECA-MVRes2net model improved the recall and precision by an average of 5–32.08 %, and 4.97–32.075 % compared to MVCNN, MVResnet, Pointnet++, and PointConv, respectively. In terms of F1 value, Kappa coefficient, and overall accuracy, the F1 value, Kappa coefficient, and average accuracy of the ECA-MVRes2net model were

improved by 5.01–32.11 %, 0.714–0.088, and 5–32.8 %, respectively, in comparison with the other models. The results showed that the multi-view Res2net model based on the ECA attention mechanism proposed in this paper outperforms MVCNN, MVResnet, Pointnet++, and PointConv models in terms of accuracy, recall, F1 value, and Kappa coefficient, and has better classification performance.

The model performance was further evaluated by calculating the model's confusion matrix. The confusion matrix was commonly used to measure the difference between the predicted and true values and can be used to evaluate the recognition accuracy of the model. The category classification performance of each network model was described by comparing the true labels with the predicted labels, and the difference between the true values and the predicted values was indicated by the color shades of the confusion matrix, where darker colors on the diagonal of the matrix indicate better classification performance, and conversely, lighter colors indicate worse classification performance. The confusion matrix of the model was calculated from the recall of the model in Table 4, as shown in Fig. 17.

Fig. 17(a), (b), (c), (d), and (e) showed that the recognition error rates of drought soybean canopies for the five network models MVCNN, MVResnet, Pointnet++, PointConv, and ECA-MVRes2net were 12.5 %, 6.67 %, 23.33 %, 33.33 % and 2.5 %, respectively; the recognition error rates for normal soybean canopy were 20.83 %, 10 %, 20.83 %, 37.5 % and 4.17 %, respectively. Compared with other models, the identification error rate of the ECA-MVRes2net model proposed in this paper was reduced by 4.17–30.8 % for drought soybean canopies and 5.83–33.33 % for normal soybean canopies. The results showed that after having introduced the attention mechanism, the network model differentiated more strongly between drought stress and normal water supply soybean canopy features, enabling it to extract multi-scale fine-grained color, texture, and other feature information of the projected soybean canopy images, and to improve the recognition accuracy of the stress state of the soybean canopy.

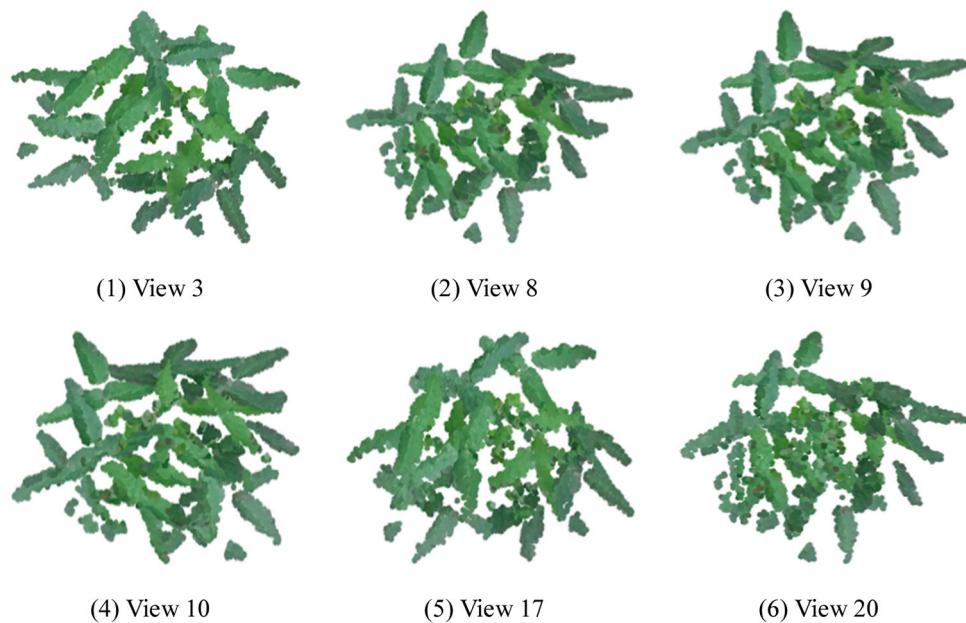


Fig. 14. Projected image of screened soybean canopies.

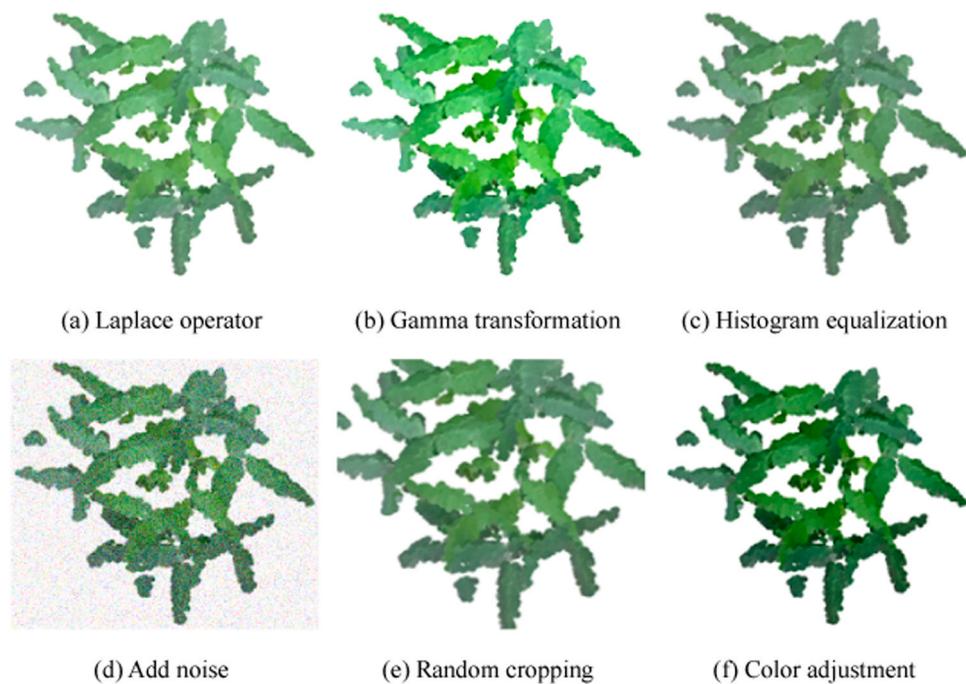


Fig. 15. Data enhancement effect.

Table 3
Training hyper-parameters settings.

| Experimental environment | Method | Epochs | View num | Input size | Input points | Learning rate | Optimizer | Batch size |
|--|---------------|--------|----------|------------|--------------|---------------|-----------|------------|
| NVIDIA GeForce GTX 1050 Ti (4096MB), Anaconda3, CUDA 11.7, Pytorch-GPU 2.0 | MVCNN | 100 | 6 | 224×224 | - | 0.001 | Adam | 64 |
| | MVResnet | 100 | 6 | 224×224 | - | 0.001 | Adam | 64 |
| | PointNet++ | 100 | - | - | 1024 | 0.001 | Adam | 32 |
| | PointConv | 100 | - | - | 1024 | 0.001 | Adam | 32 |
| | ECA-MVRes2net | 100 | 6 | 224×224 | - | 0.001 | Adam | 64 |

In order to visually analyze the model's attention regions for the input images before and after the introduced attention mechanism, this paper adopts Grad-CAM (Selvaraju et al., 2020), a class heat map used as

a visualization tool for the model learning process, to generate the class heat map visualization images for the improved Res2net block module after the ECA attention mechanism was deployed to construct the new

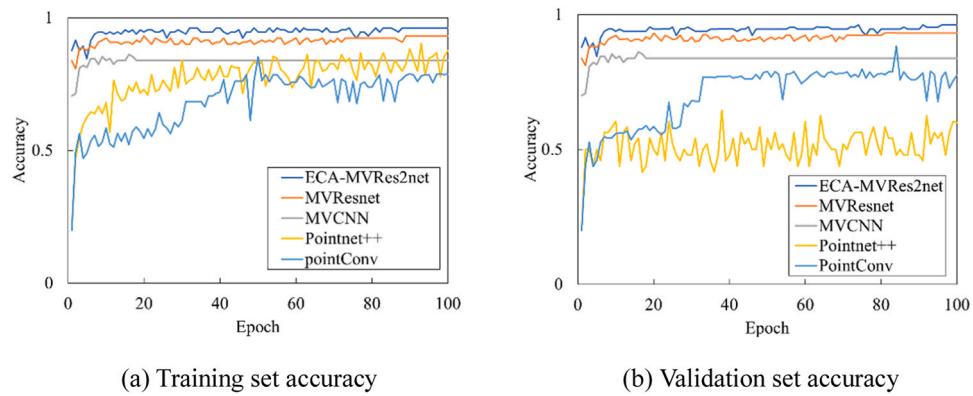


Fig. 16. Accuracy change curves.

Table 4
Comparison of the performance of each model testing set.

| Method | Class | Precision(%) | Recall(%) | F1-score(%) | Kappa | Time(s) | OA (%) |
|---------------|---------|--------------|-----------|-------------|-------|---------|--------|
| MVCNN | Normal | 79.17 | 86.36 | 82.6 | 0.667 | 0.425 | 83.335 |
| | Drought | 87.5 | 80.77 | 84 | | | |
| MVResnet | Normal | 90 | 93.1 | 91.52 | 0.836 | 0.3957 | 91.665 |
| | Drought | 93.33 | 90.32 | 91.8 | | | |
| Pointnet++ | Normal | 62.5 | 65.21 | 63.83 | 0.21 | 2.584 | 64.585 |
| | Drought | 66.67 | 64 | 65.29 | | | |
| PointConv | Normal | 79.17 | 77.24 | 78.2 | 0.56 | 2.465 | 77.92 |
| | Drought | 76.67 | 78.63 | 76.67 | | | |
| ECA-MVRes2net | Normal | 95.83 | 97.46 | 96.64 | 0.924 | 0.2865 | 96.665 |
| | Drought | 97.5 | 95.9 | 96.7 | | | |

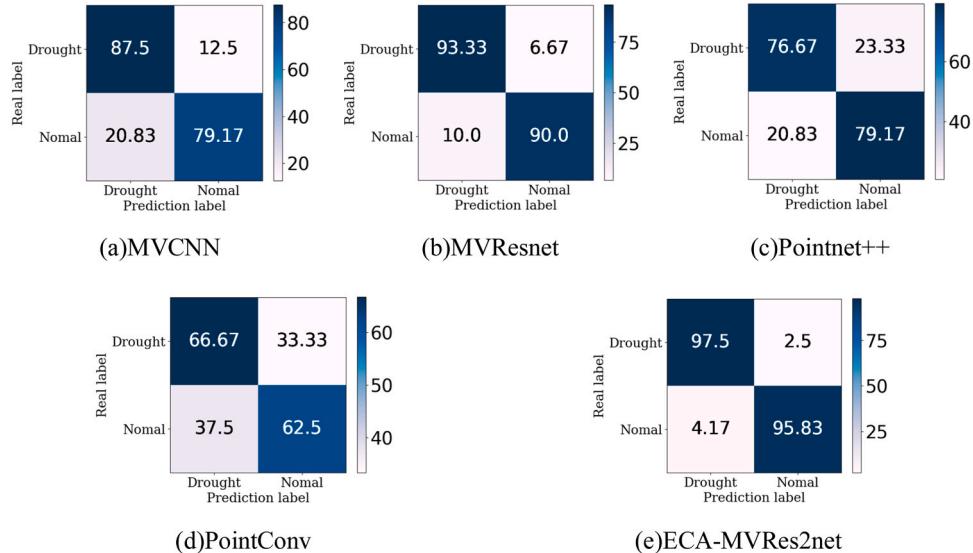


Fig. 17. Confusion matrix for each model.

Econv module. The class heat map for six different input views of the model were shown in Fig. 18.

Fig. 18 showed that the feature learning region of the model was significantly enlarged after the addition of the attention mechanism. This indicates that the improved ECA-MVRes2net model has enhanced the resolution of fine-grained features, and by expanding the sensory field, it has improved the model's attention to the edge of the soybean canopy leaves to a certain extent.

4. Discussion

In this paper, a novel multi-view convolutional neural network model was proposed, and the overall accuracy rate of simulation experiments reached 96.665 %, which achieved better classification results in drought stress recognition of soybean canopy three-dimensional images. To verify the scientific validity and generalization ability of the model, a new dataset was selected and constructed from the previous year's drought stress soybean 3D point cloud data for model training and validation, and the recognition accuracy reached more than 93 % on the training set and more than 92 % on the test set. The drought recognition

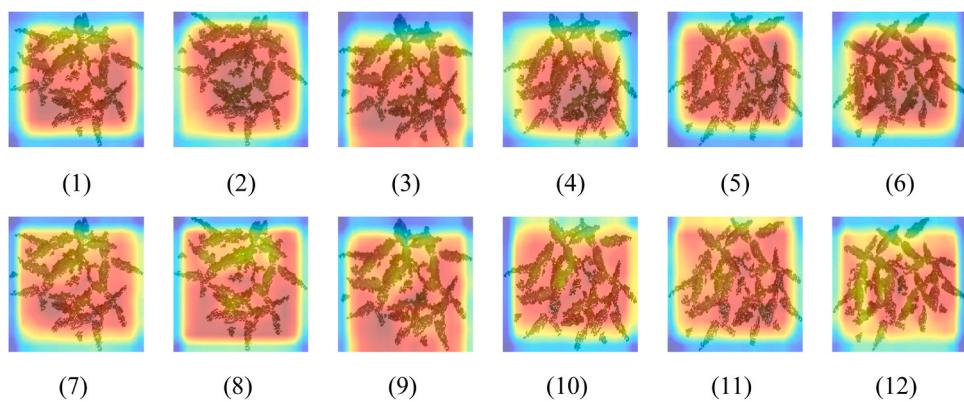


Fig. 18. Characteristic thermograms of the model before and after improvement.

model proposed in this paper has excellent generalization ability to maintain high accuracy and stability over different data samples. The specific experimental results of the model were specifically analyzed as follows:

4.1. Analysis of relevant studies

At present, three-dimensional image recognition technology was widely used, but it was still in the theoretical stage of agricultural production, and it lacked a three-dimensional image recognition model for soybean canopy. Therefore, this paper proposed a three-dimensional image recognition method for soybean canopy based on a multi-view convolutional neural network, and the average accuracy, F1 value, and Kappa coefficient of the model recognition of this method reached 96.665 %, 96.67 %, and 0.924, respectively. Compared with the traditional 2D image-based recognition methods (Zheng et al., 2019; Gai et al., 2021), 3D images can provide more comprehensive and rich spatial, color, and texture, etc., which enables the model to capture the subtle features and 3D structure of the image more intuitively, and to some extent can improve the accuracy and efficiency of recognition. Compared with the latest proposed drought recognition model based on multispectral images of soybean (Zhang et al., 2023), the model proposed in this paper improved the recognition speed by 0.21 s, which greatly improved the efficiency of model drought recognition. Compared with the proposed KSR-based 3D recognition method in literature (Paigwar et al., 2019), the recognition speed of the 3D image recognition method proposed in this paper was improved by 2.7135 s, and the recognition accuracy of the model proposed in this paper was improved by 2.1 % compared with the DCNN-based soybean drought recognition model proposed in the literature (Chandel et al., 2021). Through three-dimensional modeling and analysis of soybean canopy images, this achievement provided a new idea for drought identification in early soybean canopy three-dimensional images, verified the potential and effectiveness of the model in three-dimensional image recognition tasks in simulation experiments, and provided a theoretical basis and technical support for rapid detection of drought in soybean plants and scientific water control.

4.2. Experimental error analysis

In the acquisition method used in this paper, there will be noise points unrelated to the target data when acquiring data, through conditional filtering, Otsu threshold segmentation, and statistical filtering can eliminate most of the noise points, but due to the defects generated by the equipment and the algorithm itself, the operation of the pre-processing is relatively complex, and there are over-segmentation and under-segmentation problems. If the data is collected against a solid color background, the generation of noise points can be greatly reduced, and data processing will be simpler and more efficient. However, in

practical applications for crop data in the field, it is often difficult to collect in a solid color background, so the pre-processing method used in this paper is necessary and more advantageous to provide technical support for the collection and pre-processing of crop data in large fields. In the collection of soybean canopy images, external environmental factors caused color and geometric distortion problems (Mou et al., 2020; Hu et al., 2018). In this regard, the use of higher resolution acquisition equipment can be considered to obtain more detailed and accurate image data, and the use of sparse optimization, curvature diffusion, and edge reconstruction to weaken the influence of external environmental factors when processing data, to improve the accuracy of the data and reduce the error in the training process. For the error problem existing in the model itself, the problem of insufficient recognition accuracy and slow convergence can be avoided by adjusting the training method or replacing different optimization algorithms thus finding the most suitable algorithm for the problem (Tong et al., 2018).

4.3. Application and promotion

In this study, the improved multi-view convolutional network 3D image recognition model of the soybean canopy proposed by this study has the advantages of fast recognition speed, high accuracy, and strong generalization ability, and there was no requirement for human-designed filtering in feature extraction, and deep learning model was used to automate the computation of features for the recognition of the 3D image of soybean canopy, and the model is small in computation and able to be embedded into a portable mobile end device. Future research will further develop based on the existing works and comprehensively consider more indicators, for example, the change of water content in the stem directly reflects the water storage of the whole canopy (Zhao et al., 2023; Zhang et al., 2023), which provides an important basis for researchers to deeply understand the water status of plants under different environmental conditions, and it is especially important for the detection of drought stress of plants. Concurrently, the incorporation of environmental sensors, including temperature and soil moisture sensors (Zhao et al., 2015), facilitates the acquisition of more comprehensive and abundant data. Such sensors afford the monitoring of soil moisture conditions, temperature fluctuations, and other crucial environmental parameters in real-time, thereby furnishing a robust data basis for the development of a more comprehensive and precise drought identification model.

5. Conclusion

In this paper, a three-dimensional image recognition method for soybean canopy is proposed based on an improved multi-view convolutional network model. Taking the three-dimensional images of drought soybean and normal soybean canopy as the object, the soybean canopy was extracted by conditional filtering, threshold segmentation,

and SOR filtering algorithms, and the multi-view projections were obtained by three-dimensional rotation, and the six-view projection images were preferred by calculating the features such as texture, edge, and information entropy. The image enhancement technique was then applied for sample expansion and sample random division, and the network model of multi-view input was constructed based on Res2Net network according to the characteristics of the enhanced data set, and then the multi-scale grouped convolution module in Res2Net network was reconstructed by introducing the attention mechanism, and a new multi-view convolutional network model for soybean canopy three-dimensional image recognition was established, ECA-MVRes2Net, which solved the problem of accurate and fast recognition of 3D images of crops in practical applications. It performed well in simulation experiments, with higher recognition accuracy and faster recognition speed compared to MVCNN, MVResnet, Pointnet, and PointConv network models. The results can be used to develop an embedded real-time crop 3D image recognition system to monitor crop growth in real-time.

Funding

This study was funded jointly by the Natural Science Foundation of Heilongjiang Province, China (Funding code: LH2021C062), the Heilongjiang Bayi Agricultural University Support Program for San Heng San Zong, China (Funding code: TDJH202101), the Postdoctoral Scientific Research Developmental Fund of Heilongjiang Province, China (LBH-Q20053).

CRediT authorship contribution statement

Wenkang Xu: Writing – original draft, Visualization, Validation, Methodology. **Xiaodan Ma:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Xi Zhang:** Visualization, Validation. **Haiou Guan:** Writing – review & editing, Resources, Methodology, Funding acquisition, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

- Bhandari, A., Kumar, I., Srinivas, K., 2020. Cuttlefish algorithm-based multilevel 3-D otsu function for color image segmentation. *IEEE Trans. Instrum. Meas.* 69 (5), 1871–1880. <https://doi.org/10.1109/tim.2019.2922516>.
- Chandel, N.S., Chakraborty, S.K., Rajwade, Y.A., Dubey, K., Tiwari, M.K., Jat, D., 2021. Identifying crop water stress using deep learning models. *Neural Comput. Appl.* 3535–5367. <https://doi.org/10.1007/s00521-020-05325-4>.
- Chen, Y., Liu, G., Xu, Y., Pan, P., Xing, Y., 2021. PointNet++ Network Architecture with Individual Point Level and Global Features on Centroid for ALS Point Cloud Classification. *Remote Sens.* 13 (3), 472. <https://doi.org/10.3390/rs13030472>.
- Dong, S., Jiang, Y., Dong, Y., Wang, L., Wang, W., Ma, Z., Liu, L., 2019. A study on soybean responses to drought stress and rehydration. *Saudi J. Biol. Sci.* 26 (8), 2006–2017. <https://doi.org/10.1016/j.sjbs.2019.08.005>.
- Dong, S., Li, W., Zhang, W., Zou, K., 2020. Recognition of 3D Object Based on Multi-View Recurrent Neural Networks. *J. Univ. Electron. Sci. Technol. China* 02, 269–275. <https://doi.org/10.12178/1001-0548.2019017>.
- Fahad, S., Bajwa, A.A., Nazir, U., Anjum, S.A., Farooq, A., Zohaib, A., Huang, J., 2017. Crop Production under Drought and Heat Stress: Plant Responses and Management Options. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2017.01147>.
- Gai, R., Cai, J., Wang, S., Cang, Y., Chen, N., 2021. Research review on image recognition based on deep learning. *J. Chin. Comput. Syst.* (09), 1980–1984. <https://doi.org/10.3969/j.issn.1000-1220.2021.09.030>.
- Gan, Z., Zhong, L., Li, Y., Guan, H., 2015. A random forest based method for urban object classification using lidar data and aerial imagery. In 2015 (June). 23rd International Conference on Geoinformatics. IEEE, pp. 1–4. <https://doi.org/10.1109/GEOINFORMATICS.2015.7378621> (June).
- Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2021. Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* 652–662. <https://doi.org/10.1109/tpami.2019.2938758>.
- Ge, Y., Li, R., Li, H., 2012. Study of point cloud data reduction algorithm integrating space partition and curvature. *Appl. Res. Comput.* (05), 1997–2000. <https://doi.org/10.3969/j.issn.1001-3695.2012.05.106>.
- Guo, C., Xia, Z., Li, C., Chen, H., Zhang, Y., 2023. A point cloud denoising algorithm combined with improved radius filtering and local plane fitting. *Laser Optoelectron. Prog.* 1–11. (<https://link.cnki.net/urlid/31.1690.in.20230920.1045.016>).
- Guo, S., Yang, K., Huo, J., Zhou, Y., Wang, Y., Li, Q., 2015. Influence of drought on leaf photosynthetic capacity and root growth of soybeans at grain filling stage. *Chin. J. Appl. Ecol.* (05), 1419–1425. <https://doi.org/10.13287/j.1001-9332.20150319.005>.
- He, H., Ma, X., Guan, H., Wang, F., Shen, P., 2023. Recognition of soybean pods and yield prediction based on improved deep learning model. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2022.1096619>.
- Hu, L., Li, Q., Liu, S., Guo, X., 2018. Color Texture Inpainting for 3D Shape Based on Signal Sparse Optimization. *J. Comput.-Aided Des. Comput. Graph.* 30 (04), 673–680. <https://doi.org/10.3724/S.P.J.1089.2018.16494>.
- Jia, S., Wan, N., Hao, X., Zong, Y., Zhang, D., Li, P., 2019. Effects of different drought stresses on growth and physiological properties of soybean. *Acta Agric. Borealis-Sin.* 34 (5), 137–144. <https://doi.org/10.7668/hbnxb.201751770>.
- Kusrini, K., Suputa, S., Setyanto, A., Agastya, I.M.A., Priantoro, H., Chandramouli, K., Izquierdo, E., 2020. Data augmentation for automated pest classification in Mango farms. *Comput. Electron. Agric.* 179, 105842 <https://doi.org/10.1016/j.compag.2020.105842>.
- Liu, F., Li, Y., Wu, L., 2023. Identification and classification of flax drought stress based on improved ResNet8l. *Acta Agric. Univ. Jiangxiensis* 45 (6), 1517–1527. <https://doi.org/10.13836/j.jaua.2023139>.
- Liu, X., Fan, C., Li, J., Gao, Y., Zhang, Y., Yang, Q., 2020. Identification Method of Strawberry Based on Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* (02), 237–244. <https://doi.org/10.6041/j.issn.1000-1298.2020.02.026>.
- Ma, X., Guan, H., Qi, G., et al., 2017. Diagnosis Model of Soybean Leaf Diseases Based on Improved Cascade Neural Network. *J. Agric. Mach.* 48 (1), 163–168. <https://doi.org/10.6041/j.issn.1000-1298.2017.01.021>.
- Manuela, M., João, P., João, S., 2003. Understanding plant responses to drought from genes to the whole plant. *Funct. Plant Biol.* 239. <https://doi.org/10.1071/fp02076>.
- Mao, Z., Du, Y., Xiao, S., Shi, S., 2021. Fine-grained image classification method based on ECA-Net and multi-scale. *Appl. Res. Comput.* (11), 3484–3488. <https://doi.org/10.19734/j.issn.1001-3695.2021.02.0074>.
- Mou, Q., Li, Z., Li, H., 2020. Depth image hole inpainting method using curvature diffusion and edge reconstruction. *J. Xi' Univ. Sci. Technol.* 02, 369–376. <https://doi.org/10.13800/j.cnki.xakjdxhb.2020.0225>.
- Paigwar, A., Erkent, O., Wolf, C., Laugier, C., 2019. Attentional PointNet for 3D-Object Detection in Point Clouds. *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*. <https://doi.org/10.1109/cvprw.2019.00169>.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (652–660) <https://doi.org/10.1109/cvpr.2017.16>.
- Schieck, M., Krajsic, P., Loos, F., Hussein, A., Franczky, B., Kozierkiewicz, A., Pietranik, M., 2023. Comparison of deep learning methods for grapevine growth stage recognition. *Comput. Electron. Agric.*, 107944 <https://doi.org/10.1016/j.compag.2023.107944>.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int. J. Comput. Vis.* 336–359. <https://doi.org/10.1007/s11263-019-01228-7>.
- Shin, Y.H., Son, K.W., Lee, D.C., 2022. Semantic segmentation and building extraction from airborne LiDAR data with multiple return using PointNet++. *Appl. Sci.* 12 (4), 1975. <https://doi.org/10.3390/app12041975>.
- Sia, C.W., Lim, K.H., Phang, J.T.S., 2023. Comparative Study of Point Cloud Classification Using Deep Learning Neural Networks. In 2023 International Conference on Digital Applications, Transformation & Economy (ICDATE). IEEE, pp. 1–5. <https://doi.org/10.1109/ICDATE58146.2023.10248818>.
- Tong, W., Li, M., Zhang, Y., 2018. Research on Optimization Algorithm of Deep Learning. *Comput. Sci.* 45 (S2), 155–159. <https://doi.org/10.11896/j.issn.1002-137X.2018.11A.029>.
- Wang, C., Guo, X., Du, J., Wen, W., Wu, S., 2016. Maize plant drought stress phenotype testing method based on time-series images. *Trans. Chin. Soc. Agric. Eng.* 32 (21), 189–195. <https://doi.org/10.11975/j.issn.1002-6819.2016.21.025>.
- Wang, D., Zheng, J., Nu, E., 2019. Study on plant leaf identification in arid area. *Comput. Eng. Appl.* 55 (13), 129–133. <https://doi.org/10.3778/j.issn.1002-8331.1803-0473>.
- Wang, M., Zhang, H., Fan, J., Chen, B., Yun, T., 2023. Detection and identification of tomato diseases and pests based on deep learning networks. *J. China Agric. Univ.* 28 (11), 165–181. <https://doi.org/10.11841/j.issn.1007-4333.2023.11.15>.
- Wang, X., Zhang, Y., Li, Y., Liu, T., Zhang, J., Qi, X., 2018. Effects of drought stress on growth and screening methods and indexes for drought-resistance in soybean. *J. Plant Genet. Resour.* (01), 49–56. <https://doi.org/10.1343/j.cnki.jpgr.2018.01.006>.
- Wu, B., Liu, Y., Lang, B., Huang, L., 2018. DGCNN: Disordered Graph Convolutional Neural Network Based on the Gaussian Mixture Model. *Neurocomputing* 346–356. <https://doi.org/10.1016/j.neucom.2018.09.008>.
- Wu, W., Qi, Z., Fuxin, L., 2019. Pointconv: Deep convolutional networks on 3d point clouds. *Proceedings of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 9621–9630. <https://doi.org/10.1109/cvpr.2019.00985>.

- Yang, P., Liu, D., Liu, J., Zhang, H., 2020. Mine Ground Point Cloud Extraction Algorithm Based on Statistical Filtering and Density Clustering. *Laser. Optoelectron. Prog.* 02, 237–249. <https://doi.org/10.3788/LOP57.021107>.
- Yang, S., Gao, W., Mi, J., Wu, M., Wang, M., Zheng, L., 2019. Method for Measurement of Vegetable Seedlings Height Based on RGB-D Camera. *Trans. Chin. Soc. Agric. Mach.* (S1) 128–135. <https://doi.org/10.6041/j.issn.1000-1298.2019.S0.021>.
- Yue, H., Li, M., An, J., 2018. Drought Identification of Maize Based on Color and Texture Features. *Chin. Agric. Sci. Bull.* 34 (24), 18–28. <https://doi.org/10.11924/j.issn.1000-6850.casb18030021>.
- Zhang, T., Guan, H., Ma, X., Shen, P., 2023. Drought recognition based on feature extraction of multispectral images for the soybean canopy. *Ecol. Inform.* 77, 102248 <https://doi.org/10.1016/j.ecoinf.2023.102248>.
- Zhao, L., Hou, F., Lyu, Z., Zhu, H., 2020. Image recognition of cotton leaf diseases and pests based on transfer learning[J]. *Trans. Chin. Soc. Agric. Eng.* 36 (7), 184–191. <https://doi.org/10.11975/j.issn.1002-6819.2020.07.021>.
- Zhao, Y., Li, N., Pi, T., 2015. Multi-parameter Compounded Testing System for Soil. *Trans. Chin. Soc. Agric. Mach.* 46 (8), 289–298. <https://doi.org/10.6041/j.issn.1000-1298.2015.08.040>.
- Zhao, Y., Huang, H., Zhao, Y., Liu, W., Mi, X., 2023. Study of Real-time Detection Sensor for Stem Moisture Status of Living Tree. *Trans. Chin. Soc. Agric. Mach.* 54 (7), 282–289. <https://doi.org/10.6041/j.issn.1000-1298.2023.07.028>.
- Zheng, Y., Li, G., Li, Y., 2019. Survey of Application of Deep Learning in Image Recognition. *Comput. Eng. Appl.* (12), 20–36. <https://doi.org/10.3778/j.issn.1002-8331.1903-0031>.
- Zhou, L., Liu, Y., Zhang, P., Bai, X., Gu, L., Zhou, J., Hancock, E., 2023. Information bottleneck and selective noise supervision for zero-shot learning. *Mach. Learn.* 112 (7), 2239–2261. <https://doi.org/10.1007/s10994-022-06196-7>.
- Zhu, L., Li, Z., Li, C., Wu, J., Yue, J., 2018. High performance vegetable classification from images based on alexnet deep learning model. *Int. J. Agric. Biol. Eng.* 11 (4), 217–223. <https://doi.org/10.25165/j.ijabe.20181104.2690>.
- Zhu, T., Ma, X., Guan, H., et al., 2023. A calculation method of phenotypic traits based on three-dimensional reconstruction of tomato canopy. *Comput. Electron. Agric.* 204, 107515 <https://doi.org/10.1016/j.compag.2022.107515>.
- Zou, J., Jin, X., Zhang, Y., Ren, C., Zhang, M., Wang, M., 2019. Effects of melatonin on photosynthesis and soybean seed growth during grain filling under drought stress. *Photosynthetica* 57 (2). <https://doi.org/10.32615/ps.2019.066>.
- Xiaodan Ma (1980 –), female, doctor, professor, Master Supervisor, she is mainly engaged in agricultural information technology.
- Wenkang Xu (1999 –), male, master, he is mainly engaged in digital information.
- Haiou Guan (1980 –), male, doctor, professor, Master Supervisor, he is mainly engaged in agricultural electrical and automation technology.
- Xi Zhang (2000 –), female, master, she is mainly engaged in digital information.