



OPEN ACCESS

EDITED BY

Guoxiong Zhou,
Central South University Forestry and
Technology, China

REVIEWED BY

Pappu Kumar Yadav,
University of Florida, Gainesville,
United States

Mingfang He,
Central South University Forestry and
Technology, China

*CORRESPONDENCE

Le Yang

✉ jxnzhyangle@163.com

RECEIVED 13 April 2023

ACCEPTED 16 June 2023

PUBLISHED 07 July 2023

CITATION

Yang L, Xu S, Yu X, Long H, Zhang H
and Zhu Y (2023) A new model
based on improved VGG16 for
corn weed identification.
Front. Plant Sci. 14:1205151.
doi: 10.3389/fpls.2023.1205151

COPYRIGHT

© 2023 Yang, Xu, Yu, Long, Zhang and Zhu.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A new model based on improved VGG16 for corn weed identification

Le Yang^{1*}, Shuang Xu², XiaoYun Yu¹, HuiBin Long¹,
HuanHuan Zhang¹ and YingWen Zhu¹

¹School of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang, China,

²Software College, Jiangxi Agricultural University, Nanchang, China

Weeds remain one of the most important factors affecting the yield and quality of corn in modern agricultural production. To use deep convolutional neural networks to accurately, efficiently, and losslessly identify weeds in corn fields, a new corn weed identification model, SE-VGG16, is proposed. The SE-VGG16 model uses VGG16 as the basis and adds the SE attention mechanism to realize that the network automatically focuses on useful parts and allocates limited information processing resources to important parts. Then the 3×3 convolutional kernels in the first block are reduced to 1×1 convolutional kernels, and the ReLU activation function is replaced by Leaky ReLU to perform feature extraction while dimensionality reduction. Finally, it is replaced by a global average pooling layer for the fully connected layer of VGG16, and the output is performed by softmax. The experimental results verify that the SE-VGG16 model classifies corn weeds superiorly to other classical and advanced multiscale models with an average accuracy of 99.67%, which is more than the 97.75% of the original VGG16 model. Based on the three evaluation indices of precision rate, recall rate, and F1, it was concluded that SE-VGG16 has good robustness, high stability, and a high recognition rate, and the network model can be used to accurately identify weeds in corn fields, which can provide an effective solution for weed control in corn fields in practical applications.

KEYWORDS

attention mechanism, corn weed, deep convolutional neural network, global average pooling, Leaky ReLU

1 Introduction

China is a large agricultural country with a population of 1.4 billion, and the development of agriculture affects all aspects of life. Weeds can significantly affect crop yield and quality, and weed control is a laborious and tedious task. Herbicide spraying and manual weeding are the most common methods of weed control; however, they are not desirable from an economic or environmental point of view, and manual weeding is more costly and less efficient. With the modernization of our countryside, combining computers

with agricultural production has become an inevitable trend (Sujaritha et al., 2017; Jin et al., 2021; Rong et al., 2022). Weed identification using deep learning (Jin et al., 2012; Xu et al., 2021) can not only improve the weed identification rate and rationalize the use of weed control methods; but also make efficient use of herbicides to protect the environment. As innovations in deep learning theory and hardware conditions continue to develop, people can construct deeper network models to extract more features (Ye et al., 2019; Wagle and Harikrishnan, 2021), and an increasing number of network models are being constructed for use in various aspects of agricultural production (Chakraborty et al., 2021). Deep learning has been widely applied in recent years, especially in smart agriculture fields, such as pest and disease detection (Mique and Palaoag, 2018; Liu et al., 2022; Wu et al., 2022), plant and fruit recognition (Jaiganesh et al., 2020; Bongulwar Deepali, 2021), and crop and weed detection and classification (Pando et al., 2018; Jin et al., 2021). Image recognition technology has long been used for weed recognition applications (Jiang et al., 2020). In 2019, Jiang et al. (Jiang, 2019) proposed a new model to identify field weeds by adding transfer learning to VGG16. The final model achieved good results on 12 weed images with 91.08% accuracy in the validation set. Liang et al. (Liang, 2019) subsequently constructed a new network model for weed identification by adding transfer learning to Inceptionv3 network, and the final training accuracy was over 99%. Subsequently, Barbedo (2019) proposed a method to increase the image dataset. Multiple diseases in different plants were considered in this experiment. Using the proposed method, the accuracy increased by 12%. In 2020, Fu et al. (2020) proposed a new model based on a VGG network to identify weeds in fields. Using the Kaggle image dataset, the detection accuracy of field weeds reached over 98%, and in real fields, the accuracy reached 80%. In 2022, Shundong et al (Fang et al., 2022). proposed the HCA-MFFNet for maize leaf disease identification. To validate the feasibility and effectiveness of the model in complex environments, it was compared with existing methods and the average detection accuracy of the model was 97.75%. In 2022, Subeesh et al. (2022) constructed a deep learning method to detect pepper weeds and achieved 98.5% accuracy. Later, Najmeh et al. (Razfar et al., 2022) proposed a weed identification model to improve the efficiency of weed detection in soybean plantation forests using several network models and three custom network models for comparison. The custom CNN architecture exhibited a detection accuracy of 97.7%. In 2023, Zhang et al. (2023) proposed a new model for tomato leaf disease detection by introducing an asymptotic non-local mean algorithm (ANLM) and a multi-channel Automated Oriented Recursive Attention Network (M-AORANet) to extract rich disease features. Experimental results on 7,493 images showed that the recognition accuracy of M-AORANet reached 96.47%.

The LeakyReLU activation function is currently used in a wide range of fields and has been shown to be effective in improving the accuracy of the models. In 2022, Yang et al. (2023) proposed a RE-GoogLeNet network model to identify rice leaf diseases. By replacing the ReLU activation function with the LeakyReLU activation function, RE-GoogLeNet exhibited better classification performance for rice leaf diseases than the other models, with an

average accuracy of 99.58%. The SE attention mechanism is also widely used in various networks and has been shown to help models extract image features and improve the network performance. In 2013, an improved network for rock recognition based on the SE attention mechanism was proposed (Huang et al., 2023), yielding an accuracy of 93.2%. Owing to the redundancy of parameters in the fully connected layer, some recent high-performance network models use global averaging pools rather than FCs to fuse learned depth features (Zhao, 2017). In view of these advantages, these methods are used in the proposed model. Starting from the theme of accurate weed control in the field, this study used common weeds in corn fields as the main research objects. Through several groups of controlled experiments, an efficient and stable detection model SE-VGG16 with high detection accuracy was constructed, and the SE-VGG16 model was applied to weed detection of other species with the aim of constructing a weed detection model with the ability to generalize weed control in the agricultural production process. The SE-VGG16 model proposed in this study makes the following five major contributions to existing models:

1. A new model, SE-VGG16, was proposed for identifying weeds in agricultural crops.
2. The number of parameters in the model is reduced from 70M to 15M by replacing the 3×3 convolutional kernel in the first block of VGG16 with a 1×1 sized convolutional kernel.
3. The SE attention mechanism is added to the VGG16 model to obtain more important feature information using a weight matrix that gives different weights to different positions of the image from the perspective of the channel domain.
4. ReLU is replaced by Leaky ReLU to obtain more image features and reduce the sparsity of ReLU.
5. Using a global average pooling layer to replace two fully connected layers can better unify the global spatial information corresponding to the last convolutional layer of the category and feature map, thereby integrating the global spatial information and improving the robustness of the model.

2 Materials and methods

2.1 Image collection

Images of corn seedlings and weeds were collected from Gitee (<https://gitee.com/Monster7/weed-datase/tree/master/>) via the Internet, and the corn weed dataset was taken from fields of corn seedlings in their natural environment. A Canon PowerShot SX600 HS camera was used, with the camera pointing vertically towards the ground to reduce the effect of sunlight reflections. After expert identification and manual screening, a total of 6,000 images were obtained, including images of one corn seedling and four corn weed species, with the weed categories of Bluegrass, Chenopodium album, Cirsium setosum, and Sedge. Some examples of images are shown in Figure 1, and the data for different image categories in the dataset are shown in Table 1.

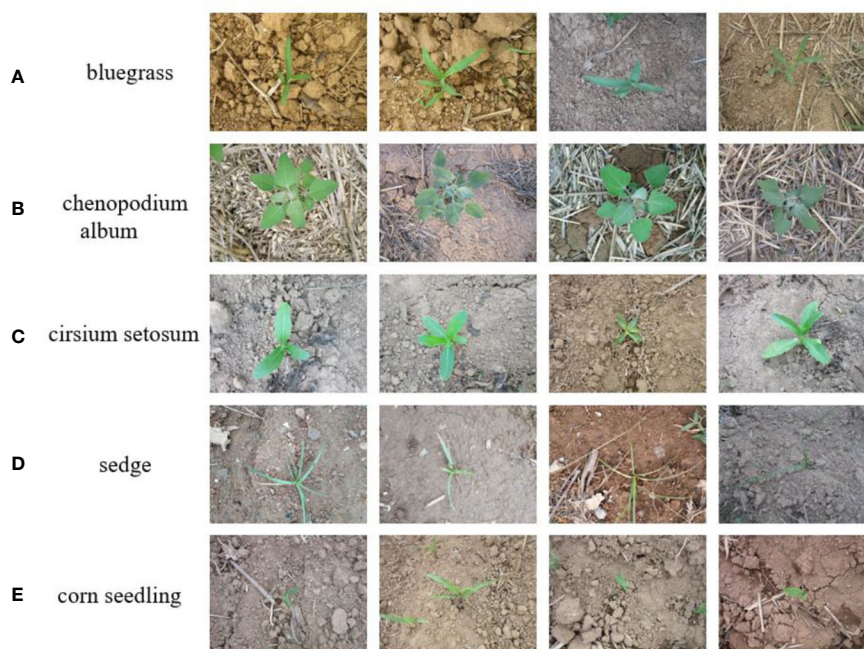


FIGURE 1 Images of some samples, (A–E) of Bluegrass, Chenopodium album, Cirsium setosum, Sedge and corn seedling, respectively.

2.2 Model building

2.2.1 VGG16

VGGNet (Simonyan and Zisserman, 2014) achieved the second place in ImageNet image classification in 2014, with VGG16 being a particularly high-performance network in VGGNet. VGG16 means that there are 16 layers containing parameters in the model, here they are divided into five blocks and a section of fully connected layers, the structure of the model is shown in Figure 2, there are two convolutional layers in block1 and block2, three convolutional layers in block3, block4, and block5, the size is 3×3 for the convolution kernel, and the ReLU activation function is used after each convolution, each block has a maximum pooling layer at the end of the block, and the size is 2×2 for the pooling kernel. The final segment of the VGG16 network consisted of three fully connected layers, and the final output was obtained using the softmax function. The mathematical expression for the softmax function is given by Equation 1. The implication of softmax is that instead of uniquely determining a maximum value, each output classification is assigned a probability value, indicating the

likelihood of belonging to each category, where Z_i is the output value of the i th node and C is the number of output nodes, that is, the number of categories in the classification. The Softmax function transforms the output of multiple categories into a probability distribution ranging from $[0, 1]$.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}} \quad (1)$$

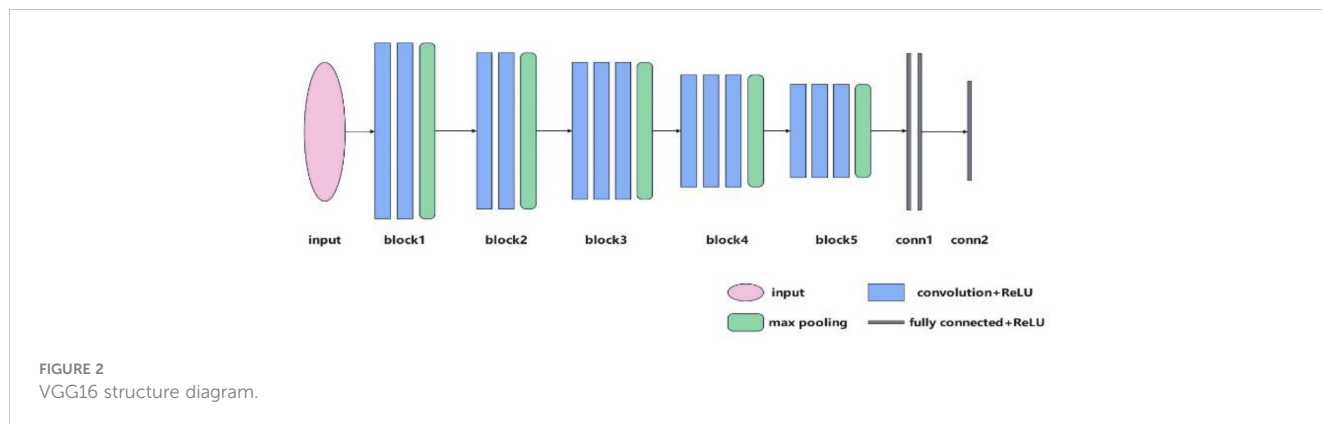
2.2.2 SENet

Attention mechanisms (Bahdanau et al., 2014; Vaswani et al., 2017) aim to achieve a network that automatically focuses on useful parts and allocates limited information processing resources to important parts. Attention mechanisms include soft and hard, and soft attention mechanisms include spatial (Jaderberg et al., 2015), channel (Hu et al., 2019), and mixed domain (Fe et al., 2017) attention mechanisms.

SENet (Hu et al., 2019) was proposed in 2017 and won the image classification task in the ImageNet 2017 competition, with the structure shown in Figure 3. It comprises two main components, squeezing and excitation. In the Squeeze operation, the global average pooling of C feature maps of size $H \times W$ is performed, and the feature maps of $C \times H \times W$ are compressed into one-dimensional feature maps of size $1 \times 1 \times C$, i.e., the global information of one-dimensional $H \times W$ is obtained with a global perceptual field. In the excitation operation, $1 \times 1 \times C$ one-dimensional features obtained from the squeeze operation are added to the fully connected layer, and the significance for each channel is obtained by the parameter W , which generates weights for each feature channel to show the correlation between each channel. Finally, the weights of the excitation output are multiplied

TABLE 1 Corn seedling and weed data set.

Type	Number
bluegrass	1,200
chenopodium album	1,200
cirsium setosum	1,200
sedge	1,200
corn seedling	1,200



with the previous features by channel-by-channel multiplication to achieve recalibration of the original features.

2.2.3 Global average pooling

The global average pooling layer (Lin et al., 2013) was first proposed in Network in Network in 2013 and is widely used in large convolutional neural networks. Traditional neural networks often have one or two fully connected layers; however, the number of parameters used is very large, which tends to cause overfitting. Global average pooling is an important component of the network, and the specific implementation is to calculate an average value for all pixels of the feature map of each channel of the output, to obtain a feature vector with a dimension equal to the number of categories, and then directly input to the softmax layer, which can achieve the effect of dimensionality reduction, thus reducing overfitting and improving the recognition accuracy of the network. In addition, by adding global averaging pooling, the model can have a global perceptual field so that the underlying network can also use global information to achieve better results. A comparative diagram of these results is shown in Figure 4.

2.2.4 Leaky ReLU activation function

The ReLU activation function (Liang and Xu, 2021) is a modified linear unit with a mathematical expression, as shown in Equation 2, and input x . If $x > 0$, its gradient is positive and can be used for weight updating, defined as the active state; if $x < 0$, its gradient is 0, the weights cannot be updated and the neuron is in a non-learning state, defined as the inactive state. To solve this problem, a linear unit function with leakage correction, that is,

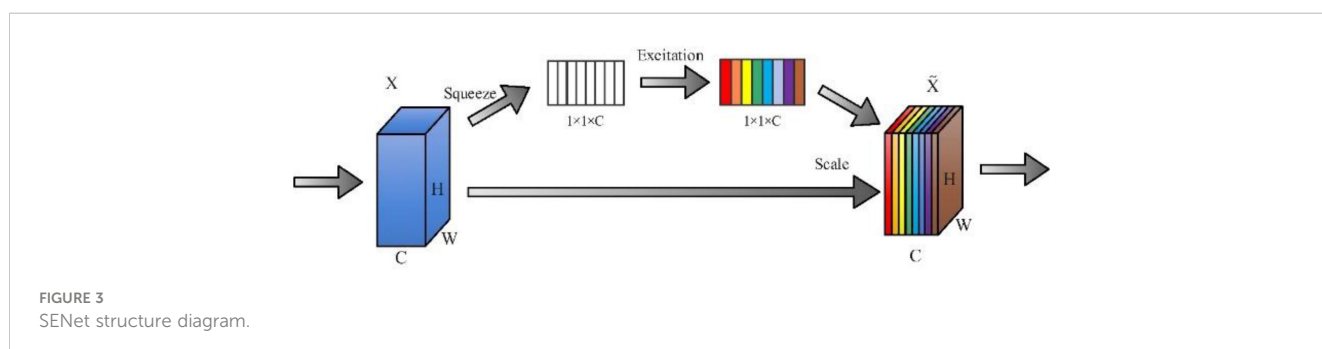
the Leaky ReLU function, is introduced. Compared with the ReLU function, the Leaky ReLU function retains a very small constant a in the negative axis; therefore when the input information is less than 0, the information is not completely lost and is retained accordingly, solving the problem that the neurons are not activated in the negative interval of the ReLU activation function. The mathematical expression of the Leaky ReLU function is shown in Equation 3, and Figure 5 shows the plots of the ReLU and Leaky ReLU activation functions.

$$\text{ReLU}(x) = \max(0, x) = \begin{cases} 0, & \text{if } x \leq 0 \text{ (Inactive state)} \\ x, & \text{if } x > 0 \text{ (Active state)} \end{cases} \quad (2)$$

$$\text{LeakyReLU}(x) = \max(ax, x) = \begin{cases} ax, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \quad (3)$$

2.2.5 SE-VGG16

In this study, based on VGG16, the SE attention mechanism was added after the first and second blocks, respectively, to enable the network to automatically focus on the useful parts and allocate the limited information processing resources to the important parts; then the 3×3 convolutional kernels in the first block were reduced to 1×1 convolutional kernels, and 1×1 convolution was used to reduce the number of channels in the deep neural network to reduce the number of channels without reducing the computational complexity without degrading the model performance, replacing ReLU with the Leaky ReLU activation function. Leaky ReLU can improve the accuracy of the model in some cases, as it allows some



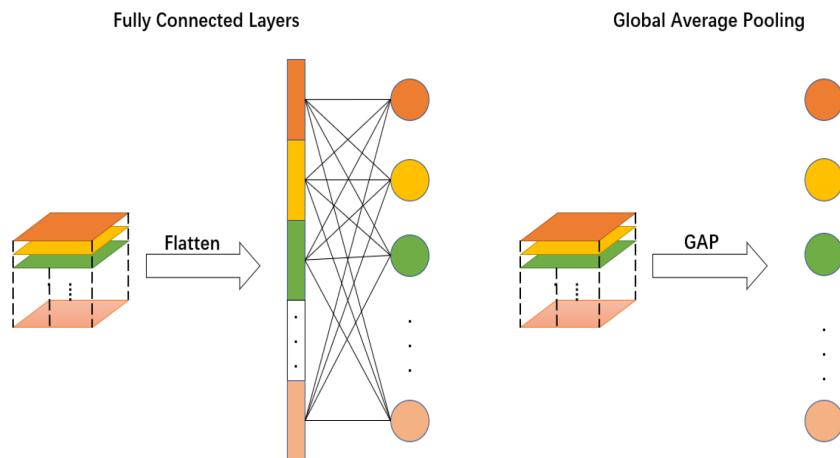


FIGURE 4 Comparison of fully connected layer and global average pooling layer.

negative outputs that may contain useful information, replacing the fully connected layer in VGG16 with global average pooling to reduce the risk of overfitting the model and speed up the training speed, and finally outputting with softmax, whose structure is shown in Figure 6.

3 Experimental results

The experiments were conducted on a Alienware computer (Dell, Inc, Round Rock, USA) configured with AMD Ryzen 7 5800H; 3.20 GHz, 24.0 GB RAM; NVIDIA GeForce RTX 3060 graphics; Windows 10 64-bit operating system. CUDA version 11.2, Cudnn version 8.1, Tensorflow 2.5.0, Python 3.7. The software used mainly included OpenCV image-processing software, with the parameters listed in Table 2.

3.1 Experimental results and design

For this experiment, the dataset was divided into a training set, validation set, and test set in a ratio of 6:2:2, and then the images were input to a uniform size of 224 × 224. Figure 7 shows the

accuracy curve and Figure 8 shows the loss rate curve of SE-VGG16 training. The figure shows that the model started to converge after 25 iterations. As the number of iterations increases, the two curves gradually fit together, and the difference between the two accuracies is close to 0, indicating that the model has reached the fit state and achieves a good training effect.

To further validate the generalization ability of SE-VGG16, other experiments were conducted in this study, namely, different activation function comparison experiments, different attention mechanism result comparison experiments, comparison tests of multiple SE modules added to the model, ablation experiments, model validation experiments in other datasets, and comparison of results with other models.

3.2 Evaluation metrics

In this study, the Precision, Recall, Accuracy, and F1 were used to measure the performance of SE-VGG16 in identifying corn weeds, calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

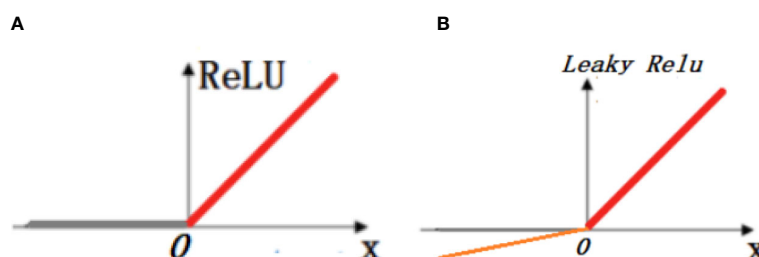


FIGURE 5 Plot of ReLU and Leaky ReLU activation functions. In the figure, (A) shows the linear representation of the ReLU activation function and (B) shows the linear representation of the Leaky ReLU activation function.

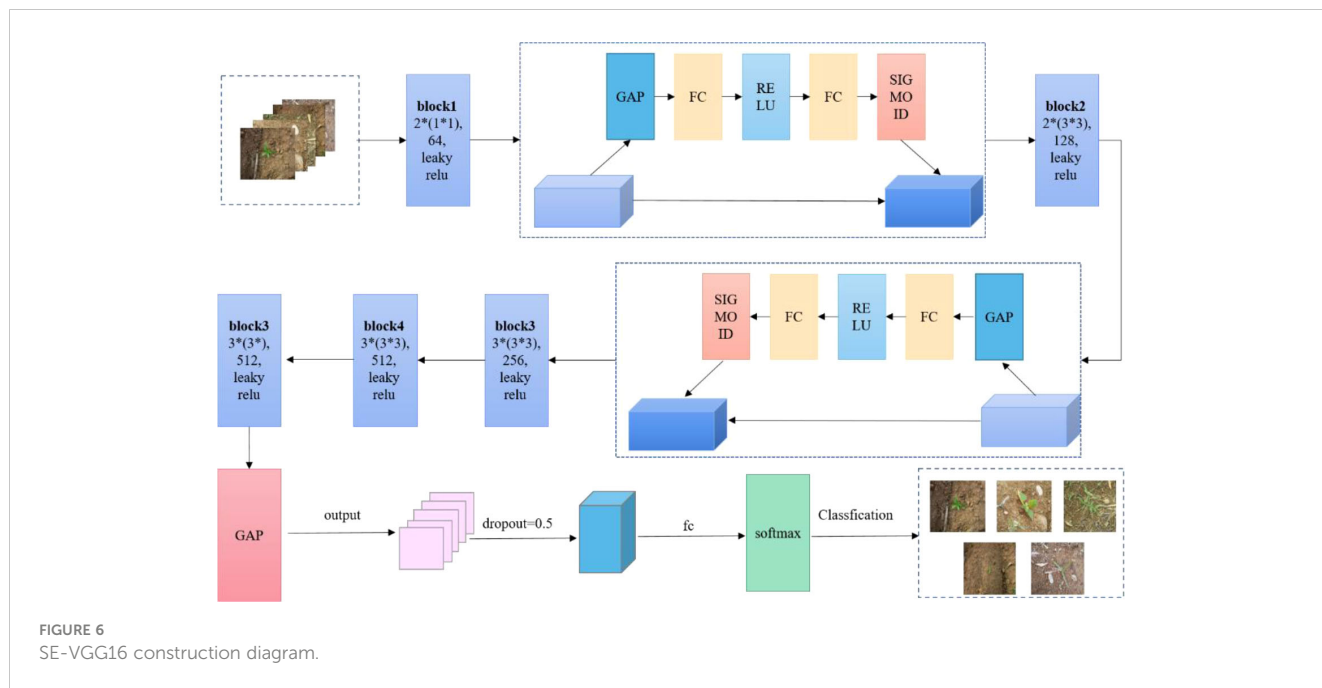


FIGURE 6 SE-VGG16 construction diagram.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{6}$$

$$F1 = \frac{2TF}{2TP + FP + FN} \tag{7}$$

TP denotes predicting a positive sample that is positive, i.e., a correct prediction; FP denotes predicting a negative sample that is positive, i.e., an incorrect prediction; FN denotes predicting a positive sample that is negative, i.e., an incorrect prediction; and TN denotes predicting a negative sample that is negative, i.e., a correct prediction.

3.3 Comparison of the results of different activation functions

The original VGG16 model uses the ReLU activation function, but the ReLU activation function can lead to “necrosis” in some neurons (Xu et al., 2020). This means that the neuron stops responding to any input during training and permanently outputs

TABLE 2 Parameter settings.

Parameter	Value
Batch size	16
Optimizer	Adam
Learning rate	0.0001
Epochs	200
Activation	Leaky ReLU
Dropout rate	0.5

0. This is because, if the value of the input is less than 0, the gradient is 0, and the neuron does not update its weights during backpropagation and continues to output 0 without learning any useful features from the data. To avoid the above problem, this study additionally selected four common activation functions to investigate their effects on network performance: Leaky ReLU, Swish, Elu, and Selu; the results are shown in Table 3. By comparing the evaluated values of the four activation functions, it was found that the accuracy of the model using the other three activation functions was higher than that of the model using ReLU, while Leaky ReLU worked best with an accuracy of 99.33%. Finally, we visualize the feature layers of ReLU and Leaky ReLU separately, as shown in Figure 9, which shows that Leaky ReLU has better performance for images. Figure 10 shows the confusion matrix for each experimental group.

3.4 Comparison of the results of different attention mechanisms

The three commonly used attention mechanisms are CBAM, ECA, and SE. In this study, the three attention mechanisms are added to the VGG16 network separately, keeping each parameter the same and using Leaky ReLU for the activation function, and the values are displayed in Table 4. Experiments showed that the SE attention mechanism performed well in this network and was best for classifying corn seed-lings with weeds.

3.5 Comparison test of multiple SE modules added to the model

To further verify the effect of the attention mechanism in the model, we added one to five SE modules to the model, and the results are presented in Table 5. The best results were achieved when the SE

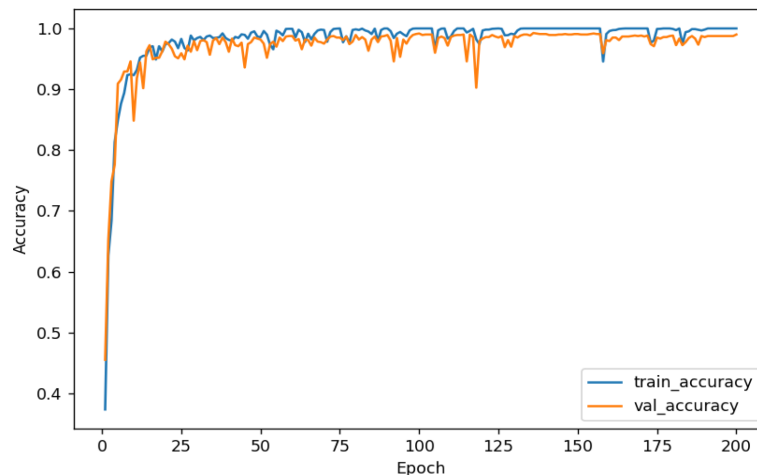


FIGURE 7
SE-VGG16 training set and validation set accuracy curves.

module was added to the first and second blocks of the model, with an accuracy of 99.67% and almost no change in the size of the model.

3.6 Ablation experiment

To verify the improved effect of the model, this study trains each classification network under the same condition, which is the network with only SE added, the network with only global average pooling added, the network with SE and global average pooling added, and the VGG16 network. The obtained results are shown in Table 6. These data also revealed that the network using SE with global average pooling was the best for classifying corn seedlings and weeds with an accuracy of 99.67%. Figure 11 displays the confusion matrix plots for the above four networks, which show that SE-VGG16 successfully recognized most of the sample images for each type, and the superiority of SE-VGG16 is again certified in terms of performance.

3.7 Model validation experiments on other datasets

To validate the generalization ability of SE-VGG16, we obtained datasets of soybean seedlings and weeds from the internet. Typically, convolutional neural networks require many images for learning, and given the existing conditions, a larger number of images cannot be collected. The existing dataset is small and has a large gap with ImageNet. To prevent overfitting, the original dataset was expanded in this study, and the expanded dataset included 1,000 images of soybean seedlings and 1,000 images of each of the two types of soybean weeds: broadleaf and grass, respectively. As in the previous experiments, the dataset was split into a training set, validation set, and test set with a ratio of 6:2:2, and the same parameters as in the previous experiments were kept, and the training was performed for VGG16 and SE-VGG16, respectively; the results are presented in Table 7. The results indicated that the

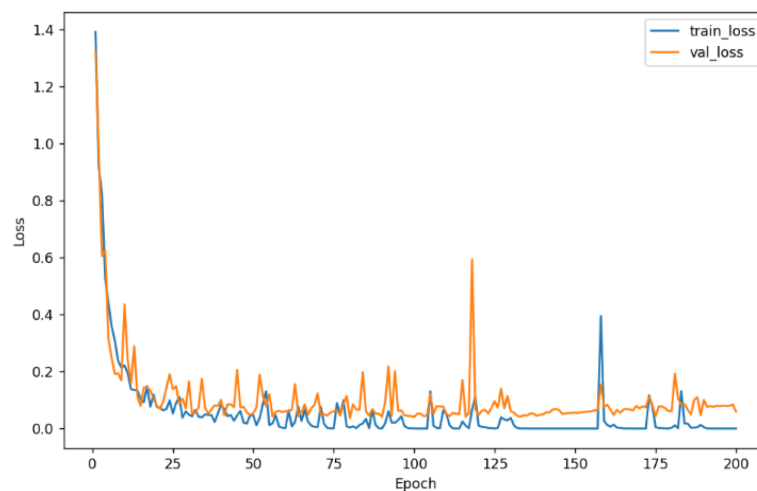


FIGURE 8
SE-VGG16 training set and validation set loss rate curves.

TABLE 3 Comparison of different activation functions.

Type	ReLU	Elu	Swish	Selu	Leaky ReLU
Batch size	16	16	16	16	16
Epochs	200	200	200	200	200
Acc%	97.75	98.83	98.75	98.25	99.33
Pre%	97.78	98.88	98.78	98.26	99.36
Rec%	97.76	98.84	98.76	98.26	99.34
F1%	97.77	98.86	98.77	98.26	99.35

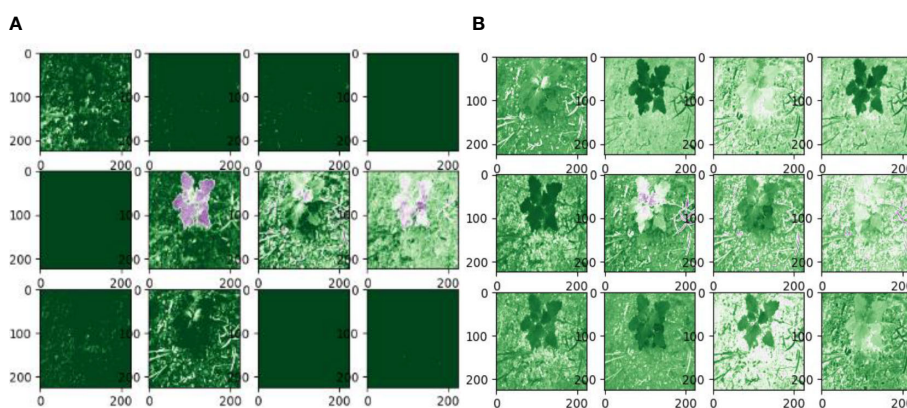


FIGURE 9 Partial convolutional layer visualization. In the figure, (A) is the feature layer visualization of the ReLU activation function, and (B) is the feature layer visualization of the Leaky ReLU activation function.

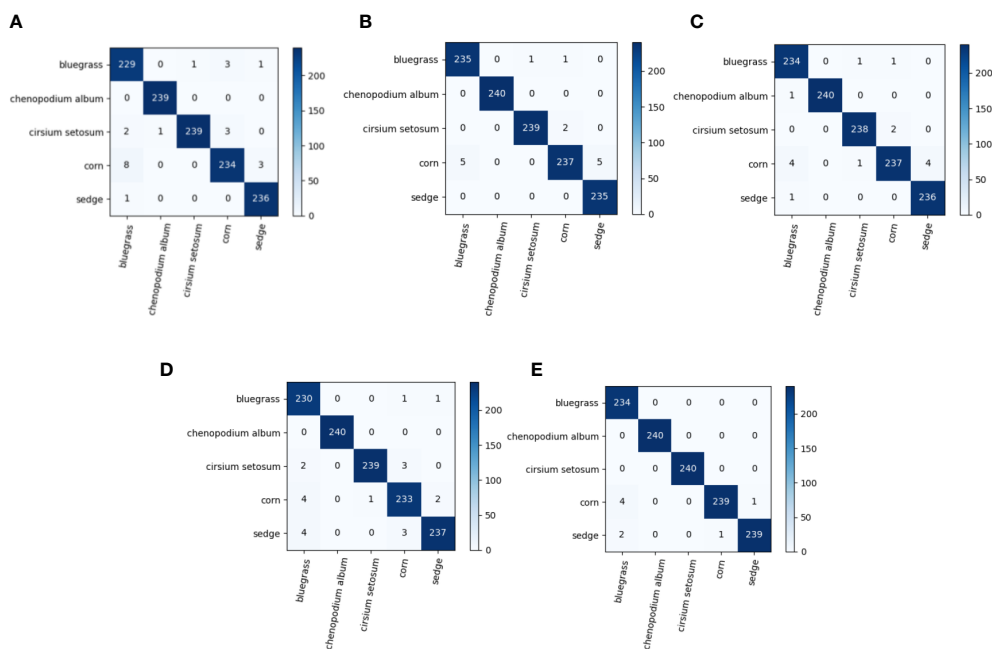


FIGURE 10 (A–E) show the confusion matrix for the ReLU, Elu, Swish, Selu, and LeakyReLU activation functions, in that order.

TABLE 4 Comparison of different attention mechanisms.

	Eps	Acc%	Pre%	Rec%	F1%
Add CBAM	200	98.67	98.68	98.66	98.67
Add ECA	200	99.17	99.16	99.18	99.17
Add SE	200	99.33	99.36	99.34	99.35

TABLE 5 Comparison test of multiple SE modules added to the model.

	Eps	Acc%	Pre%	Rec%	F1%
1 SE module	200	99.33	99.36	99.34	99.35
2 SE module	200	99.67	99.68	99.68	99.68
3 SE module	200	99.58	99.60	99.60	99.60
4 SE module	200	99.41	99.44	99.44	99.44
5 SE module	200	99.25	99.26	99.28	99.27

accuracy of the SE-VGG16 model reached 99.83% in the soybean dataset, which was higher than the accuracy of the VGG16 model (98.00%), indicating that SE-VGG16 was also applicable to other crops, proving that the model has good generalization ability.

3.8 Comparison with other model results

To further validate the performance of SE-VGG16, the SE-VGG16 model was compared with other convolutional neural network models, including AlexNet, VGG13, VGG19, EfficientNet, and MobileNetV3, and the results are shown in Table 8. Compared with AlexNet, VGG13 and VGG19, SE-VGG16 has a slightly higher accuracy, but the number of parameters is significantly lower, which makes training easier. Compared with lightweight models, such as MobileNetV3, the number of parameters still needs to be further reduced, but the accuracy of SE-VGG16 is 2.59% higher than that of MobileNetV3, further demonstrating the good performance of the SE-VGG16 network.

4 Conclusion

To address the low accuracy of the original VGG16 model, the SE attention mechanism is added, which allows the model to focus

more on features that are more critical to the classification task while reducing interference from noisy features, thus improving the accuracy and generalization ability of the model. The 3×3 convolutional kernels in the first block of VGG16 are then reduced to 1×1 convolutional kernels to reduce computation and increase nonlinearity, thereby reducing the number of parameters to 20% and speeding up the computation. The Leaky ReLU activation function was used instead of the ReLU activation function, and feature extraction was performed with reduced dimensionality to achieve a 99.67% accuracy. Compared with the models proposed by other researchers in the literature cited in this paper, our proposed model performs better. However, the number of SE-VGG16 parameters remains high and cannot be easily applied to portable devices. The next step in this research will be to reduce the number of parameters in portable devices to automatically track and identify plant seedlings with extensive weed-related knowledge. Combining deep learning methods with weed control in the field improves the efficiency of weed control, saving labor and time costs, while accurate control can also help protect the soil and environment. The research in this paper provides theoretical support for the precise application of herbicides in modern agriculture, as well as for the informatization and intelligence of agriculture. In the future, deep learning combined with agricultural production will be applied to other areas to support the entire process of agricultural production and improve the efficiency of agricultural production.

TABLE 6 Comparison of results of ablation experiments.

	VGG16	Add SE only	Add GAP only	SE-VGG16
Accuracy%	97.75	97.75	99.25	99.67
Precision%	97.78	97.76	99.28	99.68
Recall%	97.76	97.74	99.26	99.68
F1%	97.77	97.75	99.27	99.68

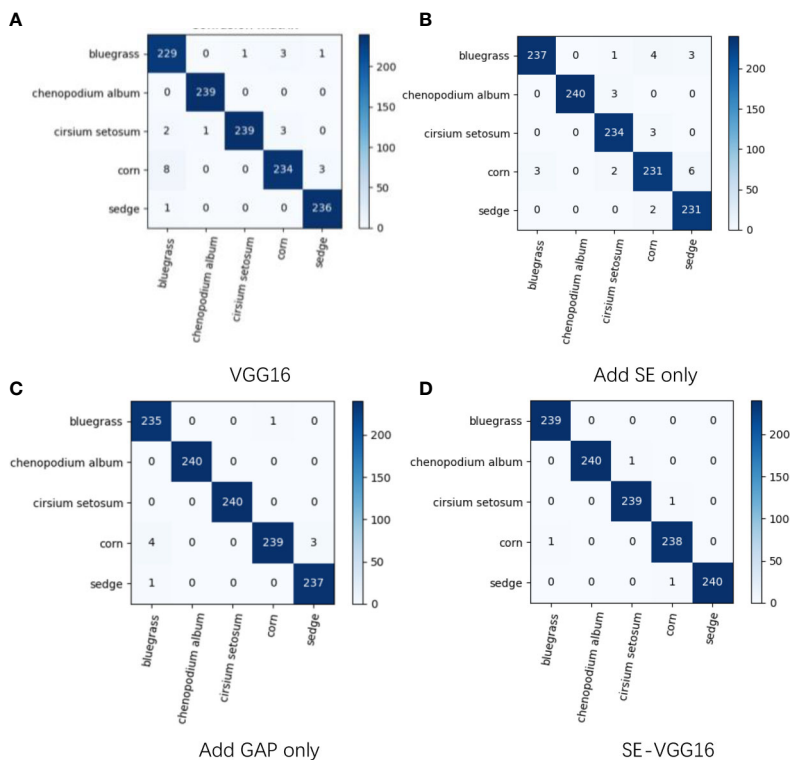


FIGURE 11 Confusion matrix comparison chart. The figure shows (A) the confusion matrix of the VGG16 recognition results, (B) the confusion matrix of the model recognition results when only the SE module is added, (C) the confusion matrix of the recognition results when only the global average pooling layer model is added, and (D) the confusion matrix of the SE-VGG16 model recognition results.

TABLE 7 Table of results for soybean data set.

Models	Type	Pre%	Rec%	F1%	Acc%
VGG16	broadleaf	100	96.50	98.22	98.00
	grass	98.50	98.00	98.25	
	Soybean	95.70	99.50	97.56	
SE-VGG16	broadleaf	100.00	100.00	100.00	99.83
	grass	100.00	99.50	99.75	
	soybean	99.50	100.00	99.75	

TABLE 8 Comparison results with other models.

No.	Models	Acc%	Pre%	Rec%	F1%
1	AlexNet	97.17	97.18	97.14	97.16
2	VGG13	97.92	97.90	97.92	97.91
3	VGG19	98.67	98.68	98.68	98.68
4	MobileNetV3	97.08	97.10	97.08	97.07
5	EfficientNet	98.00	98.02	97.98	98.00
6	SE-VGG16	99.67	99.68	99.68	99.68

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

SX and LY contributed to the conception of the study. SX and XY performed the experiment. HZ and HL contributed significantly to analysis and manuscript preparation. YZ and SX performed the data analyses and wrote the manuscript. LY and SX helped perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No. 61862032), and the Project of Natural Science Foundation of Jiangxi Province (No. 20202BABL202034).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *Comput. Sci.* 1–15.
- Barbedo, J. A. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* 180, 96–107. doi: 10.1016/j.biosystemseng.2019.02.002
- Bongulwar Deepali, M. (2021). Identification of fruits using deep learning approach. *IOP Conf. Series: Materials Sci. Eng.* 1049 (1), 012004. doi: 10.1088/1757-899X/1049/1/012004
- Chakraborty, K. K., Mukherjee, R., Chakraborty, C., and Bora, K. (2021). Automated recognition of optical image based potato leaf blight diseases using deep learning. *Physiol. Mol. Plant Pathol.* 117, 1–10. doi: 10.1016/j.pmp.2021.101781
- Fang, S., Wang, Y., Zhou, G., Chen, A., Cai, W., Wang, Q., et al. (2022). Multi-channel feature fusion networks with hard coordinate attention mechanism for maize disease identification under complex backgrounds. *Comput. Electron. Agric.* 203, 1–13. doi: 10.1016/j.compag.2022.107486
- Fe, I. W., Jiang, M., Chen, Q., Yang, S., and Tang, X. (2017). Residual attention network for image classification. *IEEE*, 6450–6458.
- Fu, L., Lv, X., Wu, Q., and Pei, C. (2020). Field weed recognition based on an improved VGG with inception module. *Int. J. Agric. Environ. Inf. Syst. (IJAEIS)* 11 (2), 1–13. doi: 10.4018/IJAEIS.2020040101
- Hu, J., Shen, L., Sun, G., and Wu, E. (2019). Squeeze-and-Excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (8), 2011–2023.
- Huang, Z., Su, L., Wu, J., and Chen, Y. (2023). Rock image classification based on EfficientNet and triplet attention mechanism. *Appl. Sci.* 13 (5), 1–20. doi: 10.3390/app13053180
- Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). *Spatial transformer networks* (MIT Press), 2017–2025.
- Jaiganesh, M., SathyaDevi, M., Srinivasa Chakravarthy, K., and Sarada, C. (2020). Identification of plant species using CNN- classifier. *J. Crit. Rev.* 7 (3), 923–931.
- Jiang, Z. (2019). A novel crop weed recognition method based on transfer learning from VGG16 implemented by Keras. *IOP Conf. Series: Materials Sci. Eng.* 677 (3), 1–8. doi: 10.1088/1757-899X/677/3/032073
- Jiang, H., Zhang, C., Qiao, Y., Zhang, Z., Zhang, W., and Song, C. (2020). CNN Feature based graph convolutional network for weed and crop recognition in smart farming. *Comput. Electron. Agric.* 174 (C), 1–11. doi: 10.1016/j.compag.2020.105450
- Jin, X., Che, J., and Chen, Y. (2021). Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access* 9, 10940–10950. doi: 10.1109/ACCESS.2021.3050296
- Jin, X., Chen, Y., Hou, X., and Guo, W. (2012). Weed recognition of the machine vision based weeding robot. *J. Shandong Univ. Sci. Technology(Natural Science)*, 104–108.
- Liang, S. (2019). “Weed identification in crops based on convolutional neural Networks[C]//,” in *Proceedings of 2019 2nd International Conference on Information Science and Electronic Technology(ISET 2019)*. 309–314.
- Liang, X. L., and Xu, J. (2021). Biased ReLU neural networks. *Neurocomputing* 423, 71–79. doi: 10.1016/j.neucom.2020.09.050
- Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv e-prints*, 1–10.
- Liu, Y., Zhang, X., Gao, Y., Qu, T., and Shi, Y. (2022). Improved CNN method for crop pest identification based on transfer learning. *Comput. Intell. Neurosci.*, 101781. doi: 10.1155/2022/9709648
- Mique, E. L., and Palaoag, T. D. (2018). Rice pest and disease detection using convolutional neural network. *Inf. Sci. Syst.* 147–151. doi: 10.1145/3209914.3209945
- Pando, F., Ignacio, H., Aedo, C., Velayos, M., Iglesias, L. L., and Calvo, J. (2018). Deep learning for weed identification based on seed images. *Biodivers. Inf. Sci. Stand.* 2. doi: 10.3897/biss.2.25749
- Razfar, N., True, J., Bassiouny, R., Venkatesh, V., and Kashef, R. (2022). Weed detection in soybean crops using custom lightweight deep learning models. *J. Agric. Food Res.* 8, 1–10. doi: 10.1016/j.jafr.2022.100308
- Rong, M., Wang, Z., Ban, B., and Guo, X. (2022). Pest identification and counting of yellow plate in field based on improved mask R-CNN. *Discrete Dynamics Nat. Sci.* 2022, 1–9. doi: 10.1155/2022/1913577
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for Large-scale image Recognition[C]// computer vision and pattern recognition. 1–14.
- Subeesh, A., Bhole, S., Singh, K., Chandel, N. S., Rajwade, Y. A., Rao, K. V. R., et al. (2022). Deep convolutional neural network models for weed detection in polyhouse grown bell peppers. *Artif. Intell. Agric.* doi: 10.1016/j.iaia.2022.01.002
- Sujaritha, M., Annadurai, S., Satheeshkumar, J., Kowshik Sharan, S., and Mahesh, L. (2017). Weed detecting robot in sugarcane fields using fuzzy real time classifier. *Comput. Electron. Agric.* 134 (6), 160–171. doi: 10.1016/j.compag.2017.01.008
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *arXiv*, 6000–6010. doi: 10.48550/arXiv.1706.03762
- Wagle, S. A., and Harikrishnan, R. (2021). A deep learning-based approach in classification and validation of tomato leaf disease. *TS* 38 (3), 699–709. doi: 10.18280/ts.380317
- Wu, T., Wu, G., and Tao, J. (2022). A research review of pest identification and detection based on deep Learning[C]// *Proc. 34th China Conf. Control Decision-making*, 325–329. doi: 10.1109/CCDC55256.2022.10034017
- Xu, J., Li, Z., Du, B., Zhang, M., and Liu, J. (2020). Reluplex made more practical: leaky ReLU[C]// 2020 IEEE symposium on computers and communications (ISCC). *IEEE*. 1–7.
- Xu, Y., Zhai, Y., Zhao, B., Jiao, Y., Kong, S., Zhou, Y., et al. (2021). Weed recognition for depthwise separable network based on transfer learning. *Intelligent Automation Soft Computing* 27 (3), 669–682. doi: 10.32604/iasc.2021.015225
- Yang, L., Yu, X., Zhang, S., Long, H., Zhang, H., Xu, S., et al. (2023). GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Comput. Electron. Agric.* 204, 1–11. doi: 10.1016/j.compag.2022.107543
- Ye, H., Han, H., Zhu, L., and Duan, Q. (2019). Vegetable pest image recognition method based on improved VGG convolution neural network. *J. Physics: Conf. Ser.* 1237 (3), 032018(10pp).
- Zhang, Y., Huang, S., Zhou, G., Y. H., and L. L. (2023). Identification of tomato leaf diseases based on multi-channel automatic orientation recurrent attention network. *Comput. Electron. Agric.* 205 (1), 107605. doi: 10.1016/j.compag.2022.107605
- Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning. *AIP Conf. Proc* 36 (03), 14–17. doi: 10.1063/1.4992835