IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 December 2023, accepted 23 February 2024, date of publication 4 March 2024, date of current version 14 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373548

RESEARCH ARTICLE

A Deep Learning Image Augmentation Method for Field Agriculture

KUNLIN ZOU[1], YI SHAN[2], XUN ZHAO[2], DE CAI RAN[2], AND XIAOXI CHE[2]
[1]School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China
[2]Youan International Building, Sinochem Agriculture Holdings, Fengtai, Beijing 100054, China

Corresponding author: Xiaoxi Che (chexiaoxi@sinochem.com)

ABSTRACT Vision-based smart agriculture is an important way to improve the efficiency of agricultural production. Labeling images for deep learning in complex field photos is a difficult task. In this paper, an image augmentation method that can help reduce the workload of image labeling based on synthetic images is proposed. The synthetic images consisted of three parts: crop, weed, and soil. The crop and weeds were obtained automatically by Excess Green (ExG) and minimum error threshold segmentation. The data augmentation method was tested on image classification, object detection, and semantic segmentation tasks by Resnet, YOLOV5, and DeeplabV3. The accuracy of the classification model reached 0.99. The IoU of object detection and semantic segmentation were 0.98 and 0.96, respectively. The results showed that the method in this paper was acceptable despite slight overfitting. This method was proposed based on the characteristics of field images, and it was meant for reducing the workload of labeling images.

INDEX TERMS Data augmentation, deep learning, field, synthetic samples.

I. INTRODUCTION

Agriculture plays a vital role in food security and social development [1]. Traditional agriculture has problems such as sloppy management, low production efficiency, and pesticide pollution [2]. Precision agriculture is an important direction for upgrading the the industry [3]. Machine vision-based precision agriculture is of high practical value [4]. Therefore, the development of machine vision technologies that can be applied in the agricultural industry has a crucial role in upgrading traditional agriculture.

In recent years, many scholars have studied the application of machine vision technology in agriculture, including disease detection, plant recognition, weed discrimination, and harvesting of fruits [5]. Three-dimensional cameras, spectral cameras, and thermal cameras are new technologies that have been developed in recent years. These techniques have also been studied by some scholars in precision agriculture [6], [7], [8], [9]. However, the high price of such devices, and the requirement for a stable working environment, make it difficult to use them in complex agricultural industries nowadays [10]. Digital cameras are the most common cameras. After decades of development, they have many advantages such as low price, high stability, and low requirements for working environment [11]. However, these cameras can only acquire the color information of three channels, red, green, and blue. Therefore, it usually needs to be used combined with an excellent algorithm.

Generally, artificial intelligence algorithms combined with machine vision can achieve good results [5]. Before the popularity of deep learning, machine learning algorithms, represented by neural networks, support vector machines, decision trees, and other algorithms, achieved state-of-the-art results [12]. These algorithms usually had to manually define image features such as color, texture, and contours. Then, machine learning models were constructed using the image features as input. Such algorithms achieved better applications in pest and disease identification, seedling and grass classification, and fruit harvesting [13], [14], [15], [16]. With the advent of convolutional neural networks, deep learning neural networks have gradually dominated the state-of-the-art models [5]. It replaces the manual definition of image features with convolutional layers. This greatly improves the adaptability of artificial intelligence algorithms.

The associate editor coordinating the review of this manuscript and approving it for publication was Rahim Rahmani.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Deep learning is highly applicable to complex agricultural environments. Numerous scholars have applied convolutional neural networks to solve difficult problems in agricultural production [17], [18], [19], [20]. However, deep learning network models are usually more complex and have more parameters. This requires more computing power and a larger number of samples to train the model. Therefore, developing a technique that can reduce the number of training samples is crucial for the application of deep learning in agriculture.

Data augmentation is a suite of techniques that enhance the quantity and quality of training datasets. It makes deep learning models can be trained better [21]. Early data augmentation was dominated by simple graphical transformations such as Geometric transformations, Flipping, Color space, Rotation [22], [23], [24], [25], [26], [27]. These image data enhancement methods were simple and effective. They are still widely used today. But these data augmentation methods are only graphically enhanced. Therefore, their effect is limited. Another important class of image data augmentation methods is based on the cropping of images [28]. This type of method initially crops off an arbitrary part of the image. Later on, the method of Cut-Mix was developed. In this method, a part of the image is cropped out and pasted on top of another image to form a new image. This method replaces the background and noise to the detection object and therefore can achieve good results. This method was applied to plant disease detection with good results by Douarre et al [29]. In the field, target, noise, and background are relatively fixed. The background of farmland is usually soil. Soil is relatively stable. Soil types in the world are fixed. The noise, in the farmland environment, is generally weeds, weeds are also generally fixed about a dozen species of weeds occupy more than 90% of the total weeds. Therefore, this method has a greater potential for application in agricultural fields.

In this paper, a data augment method for agriculture environments was investigated. The field image was divided into three parts, object, noise, and background. The main research target of this paper is to investigate the cut-mix method for data augment in the field environment. The specific research targets are: (1) to study the implementation of cut-mix in field images; (2) to investigate the augment effect of the method on classification, object detection, and semantic segmentation tasks.

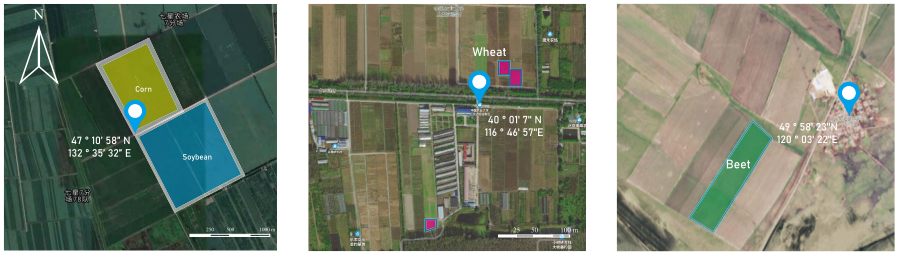
II. MATERIALS AND METHODS

A. EXPERIMENTAL IMAGES ACQUISITION

In this study, digital cameras (RERVISION USB8MP02G, China) were mounted on the tractor and sprayer equipment to capture images. It was controlled by embedded devices (Raspberry Pi 4B, IMX6ULL, IMX8) when capturing images. The cameras were 140 cm from the ground. The size of the image acquisition area was 90 × 120cm. There were no sundries and shadows in the collection area.

There were three experimental sites in this study. Experimental site A is located at Qixing Farm, Jiansanjiang, Jiamusi, Heilongjiang Province, PR China(132°35' 32'' E, 47°10' 58'' N), shown in Figure 1(a). Corn and soybean were cultivated at this experimental site. The main species of weeds growing here are Green bristlegrass, Thistle, Acalypha copperleaf, Equisetum arvense, Goosefoots. The soil here is fertile black soil. Experimental site B is located at Shangzhuang Experimental Station, China Agricultural University, No.2 Yuanmingyuan West Road, Haidian District, Beijing, PR China (116°46' 57'' E, 40 °01' 7'' N, shown in Figure 1(b)). Wheat is cultivated here. The main weeds in this area are Horsetail, Plantain, Parsnip, Endive, and Ashwort. The soil of this place is mainly cinnamon soil. Experimental site C is located at Unit 11, Hadatu Ranch, Hailar City, Inner Mongolia Autonomous Region, China(120°03' 22'' E, 49 °58' 23'' N, shown in Figure 1(c)). Beet is cultivated here. The main weeds in this area are Barnyard, Chenopodium, Amaranth, Cichorium, Endive, Millet, Ductars grass, Setargrass. The soil of this place is mainly chernozem.

In the corn field at experimental site A, 700 images containing weeds and corn were collected, 700 images containing only corn were collected, 700 images containing only weeds were collected, and 700 images containing only soil were collected. In the soybean field at test site A, 700 images containing weeds and soybeans were taken, 700 images containing only soybeans were taken, 700 images containing only weeds were taken, and 700 images with only soil were taken. In the wheat field at experimental site B, 700 images containing weeds and wheat were taken,



(a) (b) (c)

FIGURE 1. Experiment site.

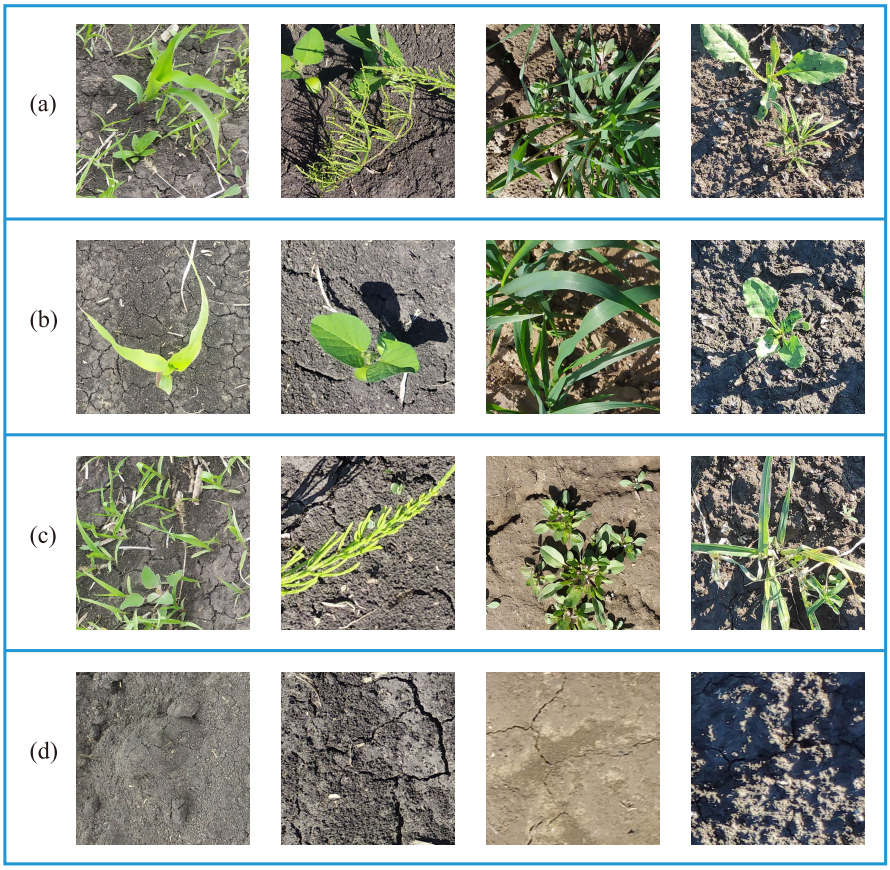


FIGURE 2. Sample images: (a)Crop and weed; (b)Crop; (c)Weed; (d)Soil.

700 images containing only wheat were taken, 700 images containing only weeds were taken, and 700 images with only soil were taken. In the sugar beet field at experimental site C, 700 images containing weeds and sugar beet were taken, 700 images containing only sugar beet were taken, 700 images containing only weeds were taken, and 700 images with only soil were taken. In total, 11,200 images were taken at the three experimental sites(shown in Figure 2).

For each type 700 images containing weeds and crops were manually labeled with crop classification, object detection, and semantic segmentation. Of each type of crop image, 400 were used as the training set, 100 as the validation set, and 200 as the test set. The 700 images containing only crops were also manually labeled for crop classification and object detection.

B. DATA AUGMENTATION METHOD

In this paper, the data augmentation method was implemented by artificially synthesizing new images. The artificially synthesized image consists of three main parts: target, noise, and background. Targets in the field were crops. In the field, the noise was normally weeds. The background in the farmland was usually the soil. Synthesizing images in this paper consisted of two main steps: (1) segmenting crops and weeds in the image; (2) attaching crops and weeds to the soil image to form synthetic images. The flow chart is shown in Figure3.

The images to make synthetic samples in this paper have two components, including plant and soil. The plant contained weeds and crops. There is a clear difference between plants and soil in terms of color. Plants are usually green. The color of the soil varies depending on the soil type and may be yellow or black. But it cannot be green. Therefore, the green color can be used to segment plants and soils. Excess Green (ExG) is a common color index used to extract the green color [30]. It can amplify the differences in green features without being disturbed by red and blue. It is the green channel two times minus the red channel and then minus the blue channel. It is calculated by the formula shown in 1. ExG was calculated for each image in this paper. The optimum segmentation threshold was calculated on the ExG image using minimum error method.

$$ExG = 2 \times g - r - b \tag{1}$$

The minimum error method is a threshold segmentation method based on the normal distribution [31]. When two probability density distributions present normal distributions, the intersection of the two normal distribution curves is the

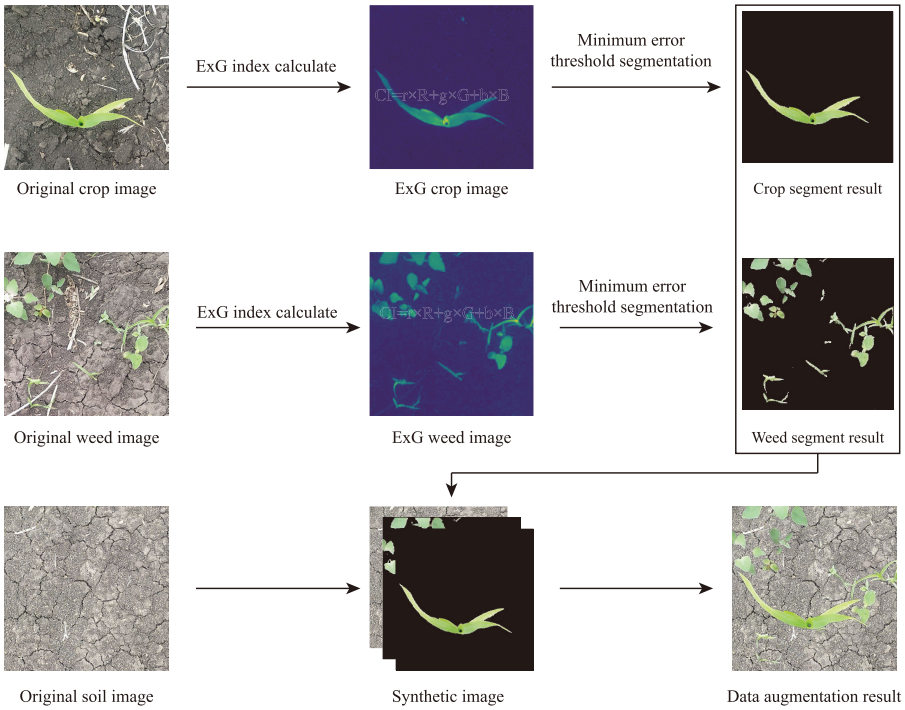


FIGURE 3. Data augmentation process.

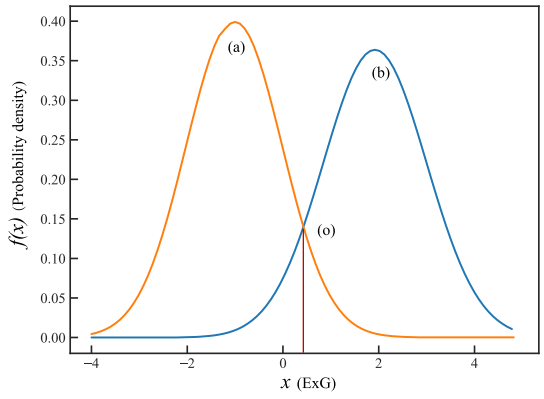


FIGURE 4. Diagram of minimum error method: (a) Normal distribution curve of foreground; (b) Normal distribution curve of background; (o) Intersection point of the two normal distributions curves.

theoretical optimal segmentation threshold. The schematic diagram of this threshold segmentation method is shown in Figure 4. Where the Figure 4(a) curve represents one class of samples, the Figure 4(b) curve represents the other class of samples, Figure 4(o) is the intersection point, and the x value corresponding to "o" is the optimal segmentation threshold.

The error could be defined as Equation 2. In order to obtain the minimum value of $E(t)$, let the partial derivative of $E(t)$ with respect to t be 0, and the conditions for minimizing the error fraction were as Equation 3. When both weeds and bare land obey normal distribution, the equation to solve for o could be expressed as Equation 4. The t value of Equation 4 could be solved as Equation 5 and Equation 6:

$$E(t) = w\int_{-\infty}^{t} f_a(x) + (1-w)\int_{t}^{\infty} f_b(x) \tag{2}$$

where w represents the proportion of one kind of samples in the whole image, while $1-w$ represents the proportion of the other kind of samples in the whole image.

$$wf_a(t) = (1-w)f_b(t) \tag{3}$$

$$w\frac{1}{\sqrt{2\pi}\sigma_a}\exp\left[-\frac{1}{2}\left(\frac{t-\mu_a}{\sigma_a}\right)^2\right] = (1-w)\exp\left[-\frac{1}{2}\left(\frac{t-\mu_b}{\sigma_b}\right)^2\right] \tag{4}$$

$$o = \frac{-2\left(\sigma_a^2\mu_b - \sigma_b^2\mu_a\right) \pm \sqrt{4\left(\sigma_a^2\mu_b - \sigma_b^2\mu_a\right)^2 - 4c\left(\sigma_b^2 - \sigma_a^2\right)}}{2\left(\sigma_b^2 - \sigma_a^2\right)} \tag{5}$$

$$c = \mu_a^2\sigma_b^2 - \mu_b^2\sigma_a^2 + 2\sigma_a^2\sigma_b^2\left(\ln\frac{\sigma_a}{w} + \ln\frac{1-w}{\sigma_b}\right) \tag{6}$$

where μ_a and μ_b are the mean value of two kind of smple pixels, σ_a and σ_b are the root means square error of two kind of smple pixels.

In this paper, 5 images of each crop were randomly selected. Five images of weeds on different lands were also randomly selected. A total of 10 images were selected. For each of these 10 images, 75 points were randomly selected

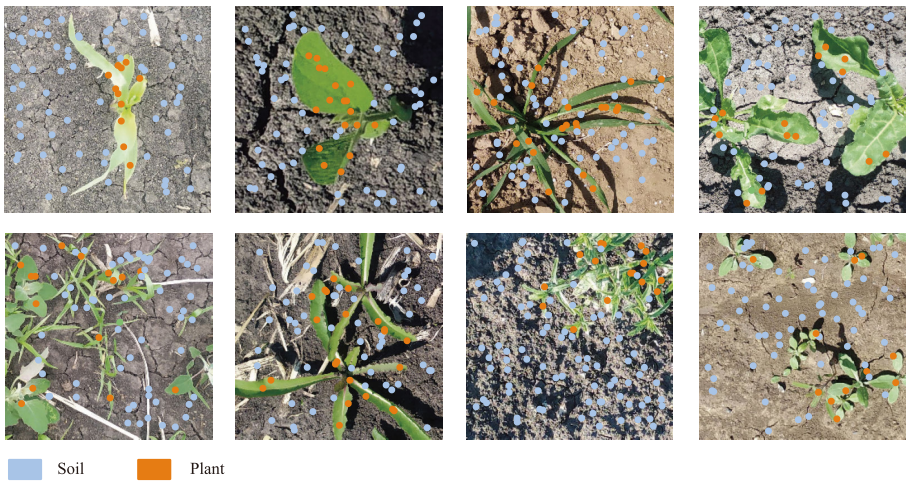


FIGURE 5. Part of the sample points for green plant segmentation.

as sample points, for a total of 750 sample points. Each pixel was manually labeled soil pixes(the value set as 0) or plants (the value set as 1). In these pixels, 500 pixels were used to calculate the segmentation threshold, and 250 pixels were used to test the accuracy of segmentation. Part of the sample points and labels are shown in Figure 5.

C. PERFORMANCE EVALUATION

In this study, three types of tasks, including image classification, object detection, and semantic segmentation, were used to test the effectiveness of the proposed data augmentation method. The first task was crop classification. There were images of four crops: corn, soybeans, sugar beets, and wheat. The image classification network was used to discriminate which crop was in the image. This task Resnet50 network was selected [32]. The second task was object detection. The location of the crop was marked in the image. For this task, YOLOV5 was selected. The last task was the image semantic segmentation task [33]. The crop images were segmented out of the mixed crop and grass images. Deeplabv3 was used in the semantic segmentation task [34].

There were two training methods. The first was real image training. This stage used real images. The training result was used for comparison with the synthetic image. The second time, the images of the training set were replaced with composite images, and the validation set and the test set were unchanged. In the image classification task, real images were first used for training. There were 1600 images in the training set, 400 images in the validation set, and 800 images in the test set. The model was trained with a total of 100 epochs. In the synthetic image training stage, 1600 synthetic images consisting of crops, weeds, and soil were randomly selected. These synthetic images were used as the training set. Both the validation set and the test set were the same as in the real image training. In the object detection task, only the soybean dataset was used. In the real image training, the training set of each kind of crop had 400 images, the validation set had 100 images, and the test set had 200 images. In the synthetic image training stage, object detection labels were consistent with the object detection labels of the crop images. The training was performed for 20 epochs each. In the semantic detection task, the soybean was used. In the real image training, the training set for each kind of crop had 400 images, the validation set had 100 images, and the test set had 200 images. In the synthetic image training stage, the semantic segmentation labels were derived from the crop images automatically segmented in the synthetic image stage. The training was performed for 20 epochs each. In the image classification and semantic segmentation tasks, the loss values and accuracies of the training data and validation data were recorded during the training process. In the object classification and semantic segmentation task, only the loss values of the training data and validation data were recorded. The hardware environment was Intel Core i7-9700 K CPU, 16 GB memory, NVIDIA GeForce RTX 2080 Super. The software environment was Windows 10, CUDA 10.1, Python 3.6, and Tensorflow 2.3.

After the training, accuracy, recall, and precision were used to evaluate the effectiveness network. Intersection over Union(IoU) was used to evaluate the object detection task and the semantic segmentation task. The calculation formula is as Equations 7-10.

$$Acc = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \times 100\% \quad (7)$$

$$Pr = \frac{\sum TP}{\sum TP + \sum FP} \times 100\% \quad (8)$$

$$Re = \frac{\sum TP}{\sum TP + \sum FN} \times 100\% \quad (9)$$

$$IoU = \frac{\sum TP}{\sum TP + \sum FN + \sum FP} \times 100\% \quad (10)$$

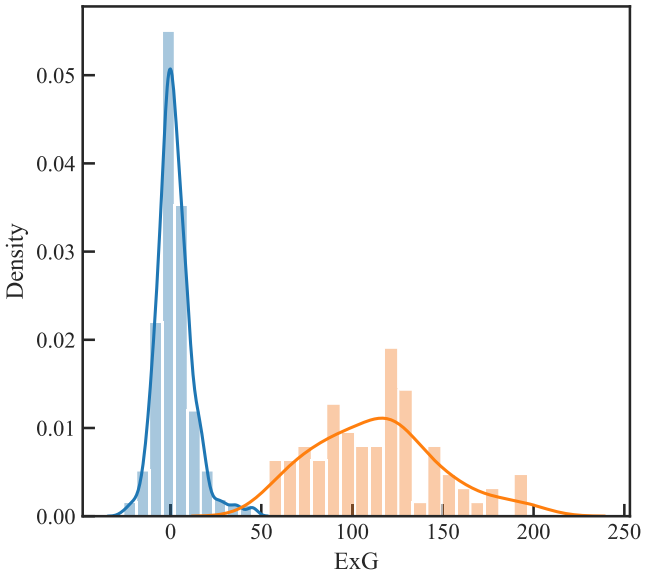


FIGURE 6. Histograms and fitted probability density function curves.

where: TP is true positive; TN is true negative; FP is false positive; FN is false negative.

III. RESULTS AND DISCUSSION

A. SYNTHETIC SAMPLE RESULTS

After calculation, the μ_a of soil sample points was 2.97 and the σ_a was 11.46. The μ_b of plant sample points was 112.30 and the σ_b was 34.07. The best segmentation threshold was calculated to be 30. Histograms and fitted probability density function curves are shown in Figure6 The error of the results obtained using this threshold to segment the test set was 4%. Some of the synthetic images and their semantic segmentation label are shown in Figure7.

The results show that ExG can effectively enlarge the color difference between plants and soil. It can make threshold segmentation easier. The segmentation thresholds obtained using the minimum error method can effectively segment plants and soil in images. This method had a 4% error, but it can save a lot of labor on manually marking the samples.

B. IMAGE CLASSIFICATION RESULTS

The change of loss and accuracy with epochs during the training process using real samples is shown in Figure 8. From Figure 8a, it can be seen that in the first 20 epochs, accuracy fluctuates and rises as the training proceeds. Compared with the training set, the fluctuation of the validation set was larger. This was because the training set had a smaller number of samples and was less stable and more prone to fluctuations. Also, the accuracy of the validation set never exceeds that of the training set. After 20 epochs, the whole training process was gradually stabilized. The training set and validation set accuracy gradually stabilized at 100%. The change of loss and accuracy with epochs during the training process using real samples is shown in Figure 8b. Loss values showed the same trend.

The accuracy and loss of training with synthetic samples are shown in Figure 9. From Figure 9a, it could be seen that the accuracy increases in fluctuation in the training of 25 epochs. The latter gradually tended to be stable. The change of loss value is shown in Figure 9b. The overall change trend was basically the same. The accuracy of the training set with synthetic samples eventually tended to 1, but the accuracy of the validation set stops rising after reaching 99%. There is a little bit of overfitting here. This was due to the difference between the synthetic samples and the real samples. The neural network was only trained by the synthetic samples, not by the real samples, and some of the features in the real samples could not be recognized correctly. This led to some errors.

The evaluation results of the real sample-trained and synthetic sample-trained models are shown in Table 1 and Table 2. From Table 1, it could be seen that the accuracy of the real value trained samples was better. However, the precision, recall, and F1-score of soybean and beet were 0.99. By looking at the confusion matrix Figure 10, it could be found that one sample of soybean and beet was classified wrongly. Despite this, the classification results were satisfactory.

TABLE 1. Classification results of real sample training.

Sample Type	Precision	Recall	F1-score
Corn	1.00	1.00	1.00
Soybean	0.99	0.99	0.99
Wheat	1.00	1.00	1.00
Beet	0.99	0.99	0.99
Overall accuracy			1.00

TABLE 2. Classification results of synthetic sample training.

Sample Type	Precision	Recall	F1-score
Corn	0.97	0.99	0.98
Soybean	0.99	0.99	0.99
Wheat	1.00	1.00	1.00
Beet	1.00	0.97	0.99
Overall accuracy			0.99

It could found from Table 2 very good results were also achieved with synthetic sample training. The accuracy rate reached 99%. However, the precision, recall, and F1-score of corn were 0.97,0.99,0.98, respectively. precision, recall, and F1-score of soybean were all 0.99. recall of beet was 0.98 and the F1-score was 0.99. By looking at the Figure 10b, it could be found that one corn was classified as soybean. A soybean was classified as corn. A soybean was classified as wheat. Five beets were classified as corn.

Both synthetic and real samples achieve good results in the classification task. Although the synthetic samples achieved

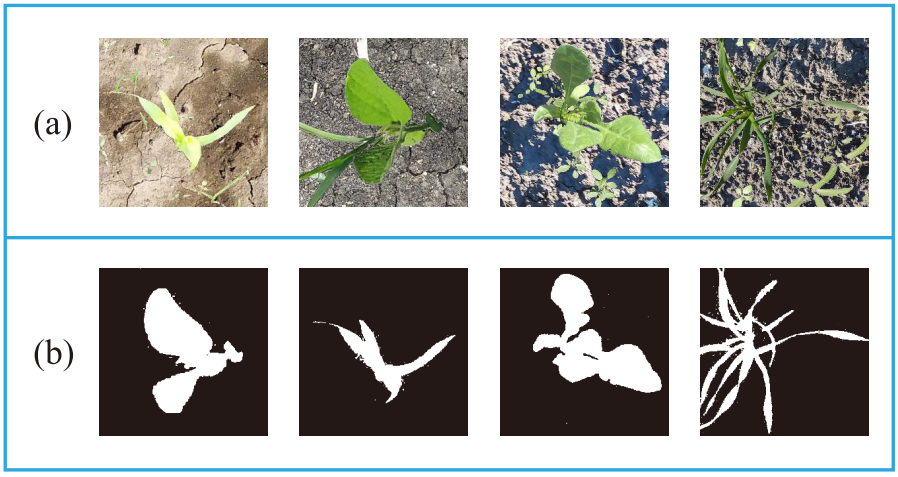


FIGURE 7. Artificially synthesized images and semantic segmentation labels: (a) Synthesized images; (b)Labels.

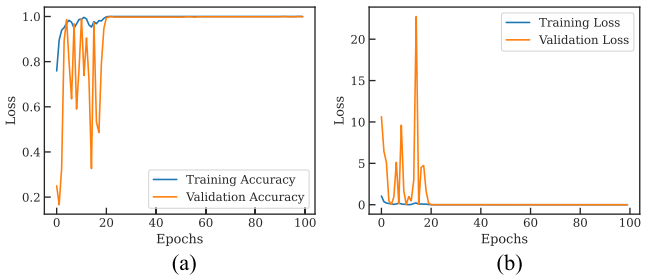


FIGURE 8. The loss and accuracy during training with real samples in classification task: (a)Accuracy; (b) Loss.

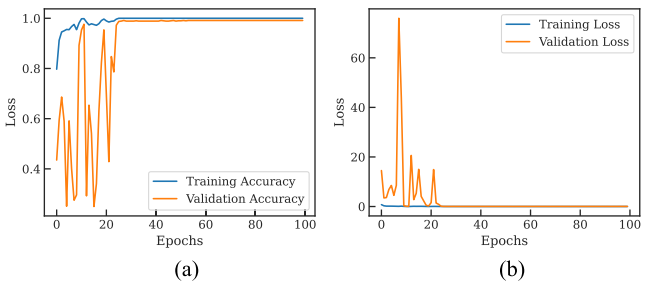


FIGURE 9. The loss and accuracy during training with synthetic samples in classification task: (a)Accuracy; (b) Loss.

poorer results than real samples. But this difference was not significant. However, the synthetic samples could be applied to more fields by changing the soil type and weed species. This was the advantage of the synthetic sample approach proposed in this problem.

C. OBJECT DETECTION RESULTS

The IoU of the object detection neural network trained on real samples was 0.96, and the result of training on synthetic samples was 0.94. The loss values of the training process for the object detection task change with training as shown in Figure 11. From Figure 11a, it could be seen that the loss of the real sample training gradually decreases in the first 10 epochs as the training progresses. It gradually stabilized around 1. The difference between the training and test set samples was less than 0.05. This level of overfitting was acceptable.

The training process using synthetic samples is shown in Figure 11b. The process was basically the same as the real sample. However, its loss value was always greater than the value of the real sample, both in the training and validation sets. The YOLOV5 network used in this paper was pre-trained. The pre-trained data were real data. Therefore, the pre-trained network was less adaptable to the synthetic data. This was an important reason why the training effect of synthetic samples was worse than the training effect of real samples. The results of some test set samples are shown in Figure 12. The precision, recall, and F1-score of of the real sample-trained and synthetic sample-trained object detection models are shown in Table 4 and Table 3. Comparing Figure 12 and Figure 12, it could be seen that although there was some difference in IoU, this difference was not

TABLE 3. Object detection results of synthetic samples training.

Sample Type	Precision	Recall	F1-score
Background	0.98	0.99	0.98
Soybean	0.87	0.84	0.85
Overall accuracy			0.97

TABLE 4. Object detection results of real samples training.

Sample Type	Precision	Recall	F1-score
Background	1.00	0.99	0.99
Soybean	0.89	0.95	0.92
Overall accuracy			0.99

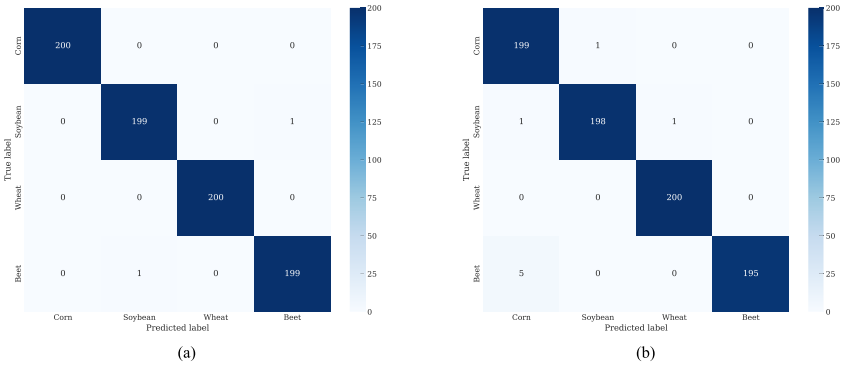


FIGURE 10. Confusion matrix: (a)Training by real sample; (b)Training by synthetic sample.

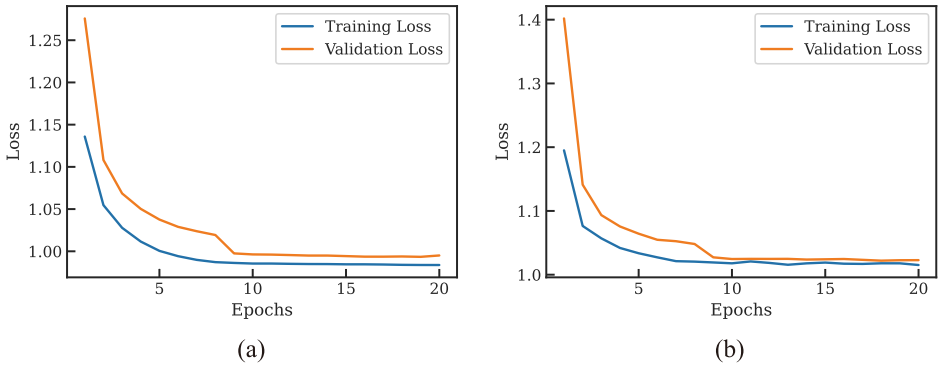


FIGURE 11. The accuracy and loss during training in object detection: (a)Training by real sample; (b)Training by synthetic sample.

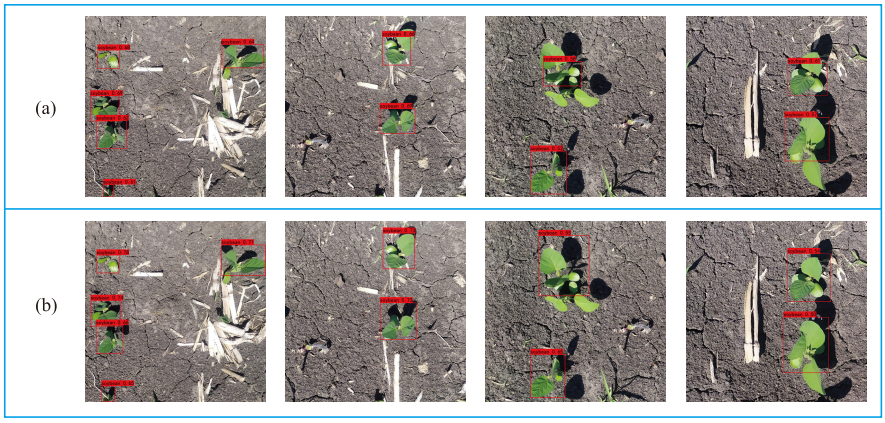


FIGURE 12. Results of object detection: (a)Training by real samples; (b)Training by synthetic samples.

significant. The difference mainly came from the edges. The edges of the target detection were more accurate for the model trained with real samples. The edges of the synthetic sample training were relatively poor. If the application was more sensitive to edges, the model can be improved by fine-tuning.

D. SEMANTIC SEGMENTATION RESULTS

The training process in the semantic segmentation problem is shown in Figure 13 and Figure 14. The evaluation results of the real sample-trained and synthetic sample-trained semantic segmentation models are shown in Table 5 and Table 6. In the training of real samples, the first 4 epochs rise faster. The decreasing process of loss value was basically the same. The overall accuracy of the model without the final model was 0.99. The precision, recall, and F1-score of the background segmentation were all 0.99. The precision, recall, and F1-score of soybean were 0.95, 0.93, and 0.94, respectively. The average IoU was 0.92.

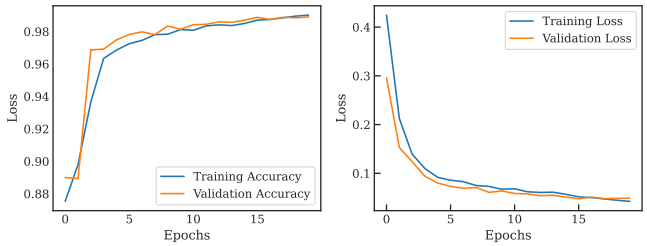


FIGURE 13. The loss and accuracy during training with real samples in semantic segmentation: (a)Accuracy; (b) Loss.

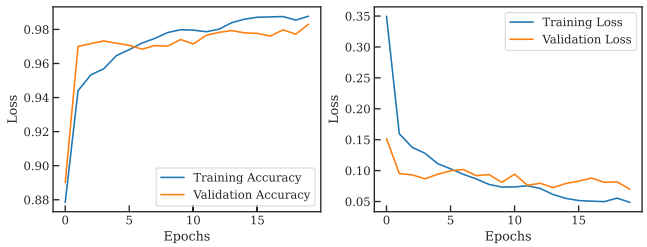


FIGURE 14. The loss and accuracy during training with synthetic samples in semantic segmentation: (a)Accuracy; (b) Loss.

TABLE 5. Semantic segmentation results of real samples training.

Sample Type	Precision	Recall	F1-score
Background	0.99	0.99	0.99
Soybean	0.95	0.93	0.94
Overall accuracy			0.99

TABLE 6. Semantic segmentation results of synthetic samples training.

Sample Type	Precision	Recall	F1-score
Background	0.99	0.99	0.99
Soybean	0.90	0.94	0.92
Overall accuracy			0.98

The training process with synthetic samples was basically the same as that with real samples. After the first four epochs, it gradually tended to stabilize. But the validation set of synthetic samples fluctuated more. Meanwhile, there were some differences between the validation set and the training set. This difference was more obvious compared to the image classification task and the target detection task. The results on the final test set showed that The overall accuracy of the model without the final model was 0.98. The precision, recall, and F1-score of the background segmentation were all 0.99. The precision, recall, and F1-score of soybean were 0.90, 0.94, and 0.92, respectively. The average IoU was 0.92.

The results of the partial test set are shown in Figure 15. From the figure, it could be found that the data trained by the real samples had higher accuracy for edge segmentation. The errors of the synthetic samples also mainly originated from the edges. This was because the edges in the synthetic samples were not naturally transitioned and were raw and directly pasted. There was no such image in the real world. After the model was trained with such samples when it predicted a real image, the processing of edges was in error. The model also had some degree of overfitting. Despite the degree of overfitting, the results were satisfactory.

E. DISSCUSSION

In this paper, a method of artificially synthesizing training samples based on ExG and the minimum error threshold was proposed. Although there was some error in synthesizing, this error was acceptable. The method could synthesize the training samples effectively. Compared with other methods, this method fully addressed the compositional characteristics of the images of the field by dividing the field images into three parts, including target, noise, and background. These three parts were segmented and combined to form a new image training network. It could achieve better results in an agriculture environment, although the method does not necessarily work well in other task types.

There were certain differences between the artificially synthesized samples and the real sample. This results in the models trained on synthetic data being slightly less effective on the test set compared to the models trained on real images. But this difference was not significant. If the accuracy requirement was not required particularly high, the model trained by artificial synthetic samples could also meet the demand. In the case of particularly high accuracy requirements, synthetic samples could be used for pre-training and real samples for fine-tuning. This could effectively reduce the need for real samples and avoid overfitting to some extent.

The error of training models with synthetic samples varies depending on the type of task. From the classification task to the object detection task to the semantic segmentation task, the accuracy of manual model training was getting worse due to the progressively higher requirements of the tasks on image processing accuracy. This was due to the fact that the subject image of the synthetic sample was consistent with the real image, but the details such as edges were not processed carefully enough. For tasks with higher accuracy requirements, such as semantic segmentation, this difference in details also had a significant impact.

Despite the advantages of the method described in this paper, there was still much further research work to be done in the future. The Synthetic method used in this paper was a direct paste method, and this method leads to a difference between the edges of the synthetic image and the real image. Therefore, there was a need to investigate the implementation of sample synthesis with the help of GAN. It could synthesize closer to the real samples. In this paper, the number of neural networks to validate the method was limited. Only Resnet, YOLOv5, and deeplabv3 were validated. More networks needed to be validated in the future. It also needs to be tested in more complex images.

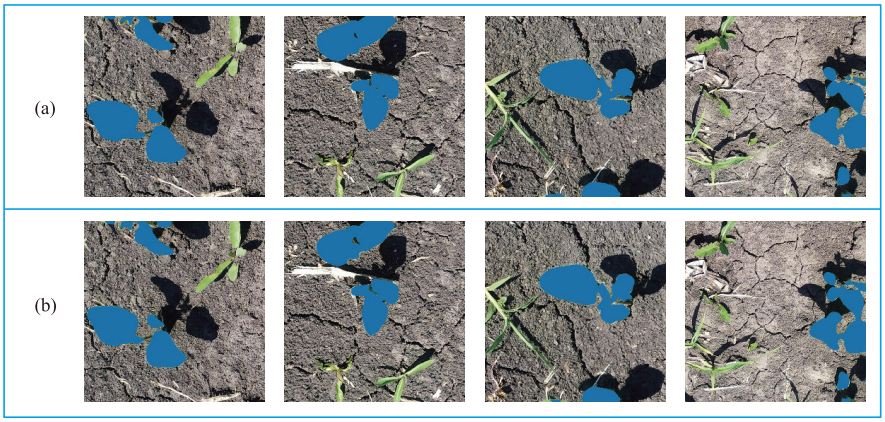


FIGURE 15. Semantic segmentation results.(a) Training by real images; (b) Training by synthetic images.

IV. CONCLUSION

In this paper, an image augmentation method based on target, noise, and background was investigated for the characteristics of the field. Synthesized images were used to train the neural network. Tests were performed on image classification, objection detection, and semantic segmentation. By analyzing the results, it could be found that:

1 The accuracy in Resnet, YOLOv5, and Deeplabv3 trained with synthetic samples were 0.99,0.98 and 0.96 respectively;

2 There was slight overfitting of the model trained by this method. But it could greatly reduce the workload of sample labeling;

3 The method was designed for the agriculture industry, and the method had wide practicality for complex field environments.

REFERENCES

[1] N. Patyka, O. Gryschenko, A. Kucher, M. Hełdak, and B. Raszka, "Assessment of the degree of factors impact on employment in Ukraine's agriculture," *Sustainability*, vol. 13, no. 2, p. 564, Jan. 2021.

[2] C.-H. Chen, Y.-C. Wu, I. Ou-Yang, and J.-J. Chen, "Unmanned self-propelled vegetable seedling planting vehicle based on embedded system," *Sensors Mater.*, vol. 34, no. 5, p. 1803, 2022.

[3] I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," *Comput. Electron. Agricult.*, vol. 176, Sep. 2020, Art. no. 105626.

[4] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Comput. Electron. Agricult.*, vol. 156, pp. 585–605, Jan. 2019.

[5] M. H. Saleem, J. Potgieter, and K. M. Arif, "Automation in agriculture by machine and deep learning techniques: A review of recent developments," *Precis. Agricult.*, vol. 22, no. 6, pp. 2053–2091, Dec. 2021.

[6] S. Campuzano and A. E. Pelling, "Scaffolds for 3D cell culture and cellular agriculture applications derived from non-animal sources," *Frontiers Sustain. Food Syst.*, vol. 3, p. 38, May 2019.

[7] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. Sousa, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.*, vol. 9, no. 11, p. 1110, Oct. 2017.

[8] L. Nhamo, G. Y. Ebrahim, T. Mabhaudhi, S. Mpandeli, M. Magombeyi, M. Chitakira, J. Magidi, and M. Sibanda, "An assessment of groundwater use in irrigated agriculture using multi-spectral remote sensing," *Phys. Chem. Earth, A/B/C*, vol. 115, Feb. 2020, Art. no. 102810.

[9] D. Li, P. Zhang, T. Chen, and W. Qin, "Recent development and challenges in spectroscopy and machine vision technologies for crop nitrogen diagnosis: A review," *Remote Sens.*, vol. 12, no. 16, p. 2578, Aug. 2020.

[10] C. Zhang, K. Zou, and Y. Pan, "A method of apple image segmentation based on color-texture fusion feature and machine learning," *Agronomy*, vol. 10, no. 7, p. 972, Jul. 2020.

[11] B. Alberton, R. D. S. Torres, L. F. Cancian, B. D. Borges, J. Almeida, G. C. Mariano, J. D. Santos, and L. P. C. Morellato, "Introducing digital cameras to monitor plant phenology in the tropics: Applications for conservation," *Perspect. Ecol. Conservation*, vol. 15, no. 2, pp. 82–90, Apr. 2017.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[13] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Comput. Electron. Agricult.*, vol. 150, pp. 220–234, Jul. 2018.

[14] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and one class classifiers," *Comput. Electron. Agricult.*, vol. 156, pp. 96–104, Jan. 2019.

[15] D. S. Jodas, N. Marranghello, A. S. Pereira, and R. C. Guido, "Comparing support vector machines and artificial neural networks in the recognition of steering angle for driving of mobile robots through paths in plantations," *Proc. Comput. Sci.*, vol. 18, pp. 240–249, Jan. 2013.

[16] J. Wu, B. Zhang, J. Zhou, Y. Xiong, B. Gu, and X. Yang, "Automatic recognition of ripening tomatoes by combining multi-feature fusion with a bi-layer classification strategy for harvesting robots," *Sensors*, vol. 19, no. 3, p. 612, Feb. 2019.

[17] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[18] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021.

[19] M. Fathi Kazerouni, N. T. Mohammed Saeed, and K.-D. Kuhnert, "Fully-automatic natural plant recognition system using deep neural network for dynamic outdoor environments," *Social Netw. Appl. Sci.*, vol. 1, no. 7, pp. 1–18, Jul. 2019.

[20] K. Zou, X. Chen, Y. Wang, C. Zhang, and F. Zhang, "A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106242.

[21] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[22] H. Bagherinezhad, M. Horton, M. Rastegari, and A. Farhadi, "Label refinery: Improving ImageNet classification through label progression," 2018, *arXiv:1805.02641*.
[23] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*.
[24] F. J. Moreno-Barea, F. Strazzera, J. M. Jerez, D. Urda, and L. Franco, "Forward noise adjustment scheme for data augmentation," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 728–734.
[25] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014, *arXiv:1405.3531*.
[26] A. Jurio, M. Pagola, M. Galar, C. Lopez-Molina, and D. Paternain, "A comparison study of different color spaces in clustering based image segmentation," in *Proc. Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst. (IPMU)*. Cham, Switzerland: Springer, Jun./Jul. 2010, pp. 532–541.
[27] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020.
[28] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.
[29] C. Douarre, C. F. Crispim-Junior, A. Gelibert, L. Tougne, and D. Rousseau, "Novel data augmentation strategies to boost supervised segmentation of plant disease," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104967.
[30] G. E. Meyer and J. C. Neto, "Verification of color vegetation indices for automated crop imaging applications," *Comput. Electron. Agricult.*, vol. 63, no. 2, pp. 282–293, Oct. 2008.
[31] C.-Y. Hsu, L.-J. Shao, K.-K. Tseng, and W.-T. Huang, "Moon image segmentation with a new mixture histogram model," *Enterprise Inf. Syst.*, vol. 15, no. 8, pp. 1046–1069, Sep. 2021.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[33] A. Kuznetsova, T. Maleva, and V. Soloviev, "Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images," in *Proc. Int. Symp. Neural Netw.* Cham, Switzerland: Springer, 2020, pp. 233–243.
[34] S. C. Yurtkulu, Y. H. Sahin, and G. Unal, "Semantic segmentation with extended DeepLabv3 architecture," in *Proc. 27th Signal Process. Commun. Appl. Conf. (SIU)*, Apr. 2019, pp. 1–4.

KUNLIN ZOU received the B.S. degree in mechatronic engineering from Tarim University, Alaer, China, in 2016, the M.S. degree in mechatronic engineering from Shihezi University, Shihezi, China, in 2019, and the Ph.D. degree in mechatronic engineering from the College of Engineering, China Agriculture University, Beijing, China, in 2023.

His research interests include image processing and machine learning.

YI SHAN is currently with Sinochem Modern Agriculture Company Ltd. His research interest includes smart agricultural applications. He is involved in product development for a number of digital farming applications.

XUN ZHAO is currently with Sinochem Modern Agriculture Company Ltd. His research interest includes application and development of agricultural IoT equipment. He has participated in the development of agricultural machinery digital products.

DE CAI RAN is currently with Sinochem Modern Agriculture Company Ltd. He is good at the innovative design and application of business models and solutions in the agricultural field and led a team to develop an APP software for farms.

XIAOXI CHE is currently with Sinochem Modern Agriculture Company Ltd. His research interests include the integration and innovative application of artificial intelligence and agricultural machinery equipment, has rich product experience in the field of intelligent agricultural machinery, and is responsible for the development of plant protection robots.