IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 June 2022, accepted 8 July 2022, date of publication 13 July 2022, date of current version 26 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190347

RESEARCH ARTICLE

Semantic Segmentation of Plant Leaves Based on Generative Adversarial Network and Attention Mechanism

LIYING CAO[1], HONGDA LI[1], XUERUI LIU[1], GUIFEN CHEN[3], AND HELONG YU[1,2]
[1]College of Information and Technology, Jilin Agricultural University, Changchun 130118, China
[2]Jilin Precision Agriculture and Big Data Engineering Research Center, Changchun 130118, China
[3]Changchun Humanities and Sciences College, Changchun 130117, China

Corresponding authors: Guifen Chen (chenguifen@jlau.edu.cn) and Helong Yu (yuhelong@jlau.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U19A2061, in part by the 2021 Jilin Provincial Budget Construction Fund (Innovation Capacity Development) under Grant 2021C044-4, and in part by the 2021 Science and Technology Research Project of Jilin Provincial Department of Education under Grant JJKH20210337KJ.

ABSTRACT Due to the rapid growth of the population, the pressure on food is increasing and the demand for higher crop yields is rising. It is crucial to periodically monitor plant phenotypic traits, and deep learning has a good effect on image recognition and segmentation. This paper proposes a method based on generative adversarial network and attention mechanism to improve the accuracy of semantic segmentation of plant leaves. First, the data set is standardized and divided into a training set and test set. The generator that produces the confrontation network uses Segnet as the backbone network and adds an attention mechanism to extract the phenotypic characteristics of plants. The discriminator utilizes a dual-input fully connected layer for true and false estimate. The experimental results show that compared with the original Segnet segmentation network, the proposed strategy improves the precision of pixel recognition PA. Also, the suggested technique has a high level of robustness and feature extraction precision. In addition to providing technical assistance for future crop cultivation and breeding, monitoring crop growth, ensuring yields as well as solving the food shortage problem.

INDEX TERMS Semantic segmentation, attention mechanism, generative adversarial network, Segnet, phenotypic characteristics.

I. INTRODUCTION

In recent years, the problem of food crisis has become serious due to global warming and population growth [1]. Growing high-yielding crop varieties is considered as a solution to this problem [2]. Research on cultivation phenotypes is necessary in the breeding process [3]. Plant foliage is usually studied as the primary functional trait [4]. The monitoring of plant growth and health status provides a scientific basis for the selection and cultivation of new species and the rational layout of varieties.

In the semantic segmentation of plant leaves for phenotype study [5], the learning of the model can be enhanced with data enhancement methods such as: image fusion [6], rotation, movement, and adjustment of luminance [7], since the background of the plant is complex. Many techniques exist on plant leaf segmentation, including traditional threshold segmentation [8], edge detection [9], unsupervised classification [10], etc.

In the semantic segmentation of plant leaves for plant phenotype studies, deep learning methods in supervised classification have achieved excellent results [6], [11]–[13]. The main convolutional neural segmentation networks includes SegNet [14], Deeplab [15], RefineNet [16],PSPNet [17], GoogleNet [18], U-Net [19],Graph-FCN [20], HMANet [21], EVS [22], GPS-Net [23] etc. SegNet segments the building image to obtain accurate classification results. The network consists of a coding layer and a decoding layer. The pooling operation of the coding layer records the original coding position and restores the original feature pixels during

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

decoding, which improves the feature output of the image. Owing to the local perception ability, the U-Net network and can obtain good classification results with few training samples, making a great contribution to medical impact segmentation. However, the U-net network must balance the accuracy of local markings and the certainty of semantics. Also, it needs to perform multiple detection with the pixel as the center area. This requires many repetitions and slowdowns in the operation. Graph-FCN strengthens the feature segmentation of local location information and extend image grid data to graph-structured data for classification. EVS has fast optimization speed and takes less running time. GPSNet combines pixel-by-pixel and full-pixel methods to study inter-image and intra-image information, making the network better adaptable from sparse to dense light source distributions.

For traditional methods of image data enhancement of plant leaves, it is time-consuming and laborious to add annotations manually. In addition the segmentation model has the problems of unstable recognition accuracy and slow training speed for plant leaves. The model improves the classification precision of the network model by adding generative adversarial network and attention mechanism, thus achieving a better image classification effect.

In the case of plant culture and breeding, it is necessary to observe the growth of plants, In this paper, the traditional Segnet segmentation model is combined with the discriminative principle of generative adversarial networks, training the classifier secondly in order to enhance the accuracy of leaf semantic segmentation of plants. The attention mechanism is integrated into the space and channels during classification to improve the learning efficiency. This study provides technical support for the extraction of plant phenotypic features and lays a solid foundation for the selection and breeding of high-yielding varieties.

II. THE PROPOSED METHOD

As a crucial element of deep neural networks, attention mechanism [24] was initially exploited to machine translation. soft attention focuses on significant pixels. Semantic segmentation has greater criteria for pixel categorization. In Convolutional Block Attention Module (CBAM) [25], which combines channel attention [26] and spatial attention mechanisms [27] to raise the weight of the segmentation target in the convolution operation.

A. NETWORK STRUCTURE

SAM [28] (Spatial Attention Module): After the maximum and average pooling of the input convolution block, the position of the entire feature area can be determined through the convolution operation, and the influence of noise and rotation on the image can be filtered out. The definition of SAM is as follows:

$$\begin{aligned} M_s(F) &= \sigma\left(f^{5\times5}([AvgPool(F), MaxPool(F)])\right) \\ &= \sigma\left(f^{5\times5}\left(\left[F^S_{avg}; F^S_{max}\right]\right)\right) \end{aligned} \quad (1)$$

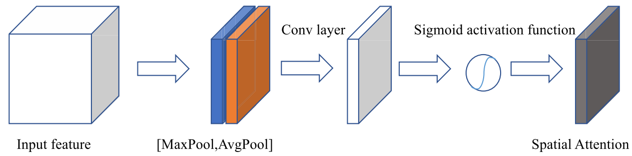

FIGURE 1. Spatial attention function module.

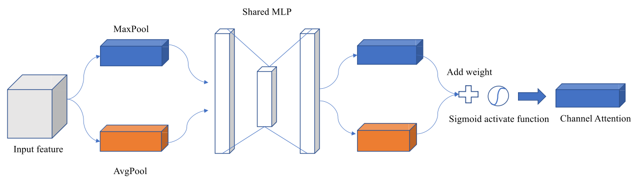

FIGURE 2. Channel attention function module.

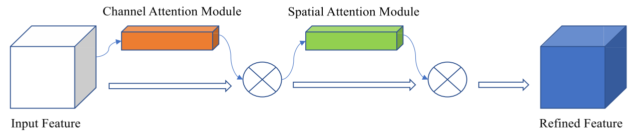

FIGURE 3. CBAM function module.

For a feature input F with a size of H×W×C, average pooling and maximum pooling are performed first in the channel dimension to obtain a channel description with a size of H×W×1. Then, the two channels are spliced through a 5 × 5 convolutional layer. Next, a sigmoid activation function is applied to obtain the weight coefficient *Ms*. Finally, the input feature F is multiplied by the *Ms* to obtain a new characteristic. The diagram of the spatial attention module is shown in Figure 1.

CAM [25] (Channel Attention Module): The input of CAM is a feature F with a size of H×W×C. The global maximum pooling and average pooling of space is respectively performed to obtain two-channel descriptions of a size of 1 × 1×C. MLP(w) is fully connected, Then, the channel descriptions are sent to the neural network consisting of two layers. After the two obtained features are summed, the weight Mc is gotten by a sigmoid activation function. Next, the Mc and the input feature are multiplied to obtain the channel attention weight characteristic. The diagram of CAM is shown in Figure 2 and the formula is defined as follows:

$$\begin{aligned} Mc(F) &= \sigma\,(MLP(AvgPool(F)) + \sigma\,(MLP(MaxPool(F)) \\ &= \sigma\left(w1\left(w_0\left(F^c_{avg}\right) + w_1\left(w_0\left(F^c_{max}\right)\right)\right.\right. \end{aligned} \quad (2)$$

CBAM[25] combines the advantages of channel attention and spatial attention, and it adaptively optimizes the convolution process of eigenmap by multiplying the original feature input and the attention weight. CBAM is integrated into the coding process of the segmentation network by adding it between each convolution block. The diagram of CBAM is shown in Figure 3.

Generative Adversarial Network (GAN) [29] is a deep learning model that is utilized to generate images for

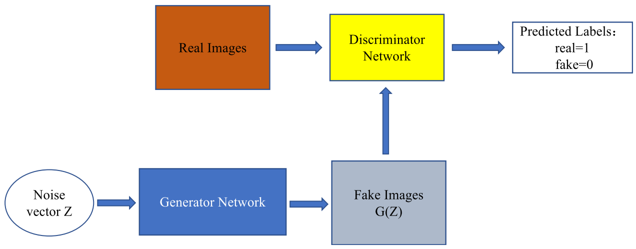

FIGURE 4. **Generative adversarial network.**

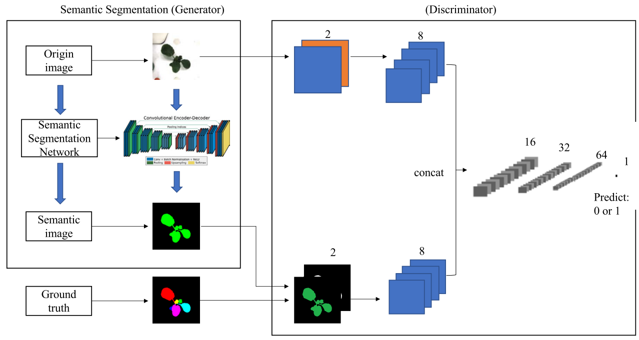

FIGURE 5. **Generative adversarial semantic segmentation network.**

data enhancement. Unstable training is a common problem of GAN, and the best result are to find the Nash equilibrium point of the generator and the discriminator. The improvement of the generative confrontation networks includes DCGAN[30], LSGAN[31],WGAN[32], and WGAN-GP[33]. DCGAN changes the network structure of GAN to a convolutional neural network, which greatly keeps the stability training and the quality of the generated results. WGAN improves the loss function of GAN, but there are still many issues, such as training difficulties and slow convergence speed. WGAN-GP [34], [35] has a great advantages effect in terms of convergence speed, training difficulty, and prevention of gradient disappearance.

GAN is a model against generated data, and it consists of two parts: generator network(G) and discriminator network(D). G randomly produces fake pictures according to the noise z, and D judges whether the generated pictures are true. There are two situations for training the input of D: The original image output is true, and the generated image output is false. The diagram of GAN is shown in Figure. 4.

Initially, a confrontation network was created in order to expand the variety of the data collection. The concepts of segmentation network and generative confrontation network are integrated in this study. The classic segmentation network is employed as the generator of the generation confrontation network, while the fully connected layer is applied as the discriminator. There are two possibilities for the discriminator's input: In one scenario, the input is the original picture and the label image, and the discriminator's output is true; in the other, the input is the original image and the segmentation image generated by the generator, and the discriminator's result is false. The structure of the generative adversarial semantic segmentation network is shown in Figure 5. The original image and the label image are put in the discriminator. Since the original image is a binary image, the label image has two channels, i.e., the background and the plant leaves. The two images are converted into the same matrix by convolution. Through the splicing operation, multiple neurons similar to the perceptron are obtained, and then judgements are made through a fully connected layer.

As a component of the generative adversarial network, the generator that inputs the images into Segnet with attention mechanism and outputs the segmented image is invoked. The discriminative network receives the original picture, the label image, and the segmented image. The input consists of two pictures with outputs of 1 for (original image, labeled image) and 0 for (original image, segmented image). Training in classification is carried out. The feedback improves the picture segmentation.

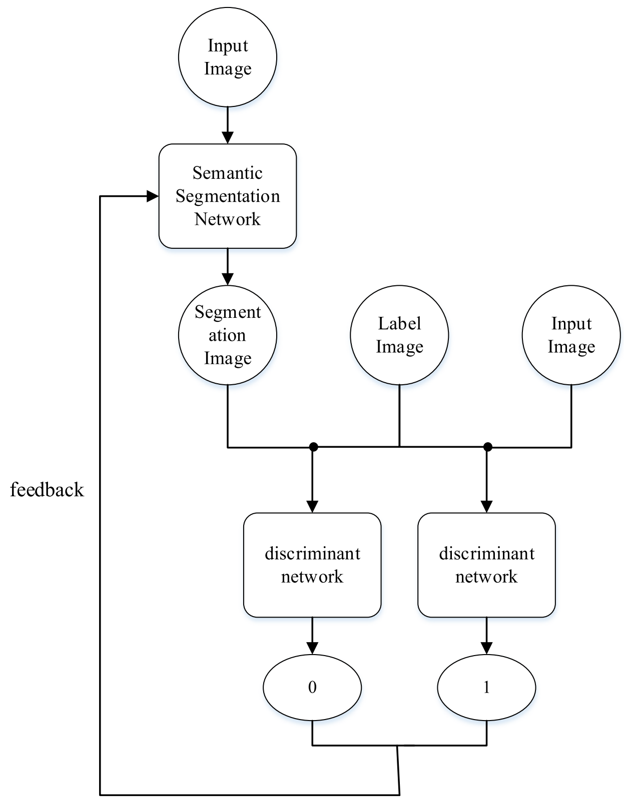

FIGURE 6. **Neural network flowchart.**

This study serves the Segnet as a generator of the generative adversarial semantic segmentation network. Segnet consists of an encoder and a decoder. One-to-one correspondence between encoder and decoder. The transformation portion saves and restores the original feature locations during upsampling. The softmax classifier is exploited by the decoder to identify the category of each pixel. The structure of the Segnet semantic segmentation network is shown in Figure 7.

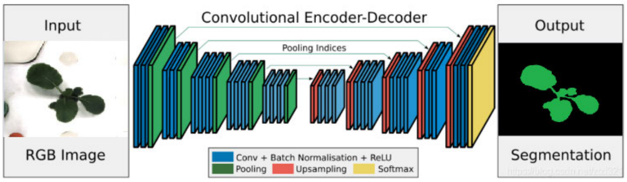

FIGURE 7. The structure of the segnet semantic segmentation network.

III. NETWORK TRAINING

A. EXPERIMENTAL PARAMETER SETTINGS

In this experiment, the software configurations are Ubuntu 18.04 TLS operating systems, python3.6 development language, and Keras deep learning framework; the hardware configurations are three NVIDIA GTX1080 Ti GPUs. Besides, the batch size is 4; the generator learning rate is 0.0001; the discriminator learning rate is 0.0001, and the number of iterations is 500.

B. TRAINING STEPS

First, the original image and the label image are taken as input to train the discriminator to mark the sample as true. The generator in the semantic segmentation networks creates a segmented images. Then, The segmented picture and the original image are then sent into the discriminator, which is then trained to flag the sample as false.

Next, the network generates semantic segmentation images with insufficient accuracy. The parameters of the generator are optimized according to the results of the discriminator. At this time, there is no requirement to modify the parameters of the discriminator, and iterative training is then performed.

However, during the training process, the value of the loss function of the generator is extremely small, and the accuracy of the entire model is not the best. This is because the optimal solution to generate the confrontation network is that the generator model and the discriminator model reach the Nash equilibrium.

C. LOSS FUNCTION

Since the discriminator can make a two-category judgment on whether the generated image can be changed from false to true, the cross-entropy loss function is applied in the generator to classify the pixels:

$$Loss = -\hat{y}\log y \tag{3}$$

where y is the correct value, and $\hat{y}$ and is the predicted value.

The overall GAN loss function is:

$$\min_G \max_D V(D, G) = E_{x\sim Pdata}(x)[logD(x)] + E_{z\sim p_{z(z)}}[(1 - logD(G(z)))] \tag{4}$$

The optimized discriminator D is:

$$\max_D V(D, G) = E_{x\sim Pdata}(x)[logD(x)]$$

The optimized generator G is:

$$\min_G V(D, G) = E_{z\sim p_{z(z)}}[(1 - logD(G(z)))]$$

where, D is the discriminator, G is the generator, z is the noise, and x is the input data. $Pdata(x)$ is the sample data distribution, and $Pz(z)$ is the noise prior distribution. In the first step of training, the discriminator D needs to be trained, and G is kept unchanged as much as possible, and the loss of D is too large. The second step confuses the discriminator so that the output of $D(G(z))$ tends to 1, which does not cause the loss of the generator to be biased.

The original GAN is difficult to train, and the training level of the generator and the discriminator cannot be balanced. In this study, the loss function of the discriminator is changed to the optimized Wasserstein distance of WGAN[32].

$$L(D) = -E_{x\sim P_r}[D(x)] + E_{x\sim P_g}[D(x)] \tag{5}$$

$$W(P_r, P_g) = \inf_{\gamma \backsim \pi(p_r, p_g)} E_{(x,y)\sim\gamma}[||x - y||] \tag{6}$$

where (P_r, P_g) is the possible joint distribution of P_r and P_g. For each possible joint distribution γ, $(x, y) \sim\gamma$ can be sampled to obtain a real sample x and a generated sample y and calculate the sample Distance $||x - y||$. Then, the expected distance of the sample under the joint distribution γ can be calculated.

D. EVALUATION

Pixel Accuracy (PA), the ratio of correctly categorized pixels to all of them, is taken as an evaluation metric for semantic segmentation. MPA (Mean Pixel Accuracy) calculates the percentage of correct pixels for each class to the number of all pixels in that category. MIOU (Mean Intersection over Union) calculates the IoU for each type. FWIoU (Frequency Weighted Intersection over Union) is a weighted summation of the frequencies of each type of IoU.

$$PA = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} P_{ij}} \tag{7}$$

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{8}$$

$$MIOU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{9}$$

$$FWIOU = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \times \sum_{i=0}^{k}\frac{p_{ii}\sum_{j=0}^{k} p_{ij}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{10}$$

In the above formulas, k represents the number of times the category is divided; p_{ii} indicates whether the original type i is predicted as type i, that is, the true-positive (TP) and true -negative (TN) of the confusion matrix; p_{ij} indicates whether the original type i is predicted as type j, that is, false-positive (FP) and false-negative (FN).

TABLE 1. The segmentation accuracy of different methods.

	PA	MPA	MIOU	FWIOU
FCN8	0.9881	0.9510	0.9062	0.9773
FCN32	0.9746	0.8839	0.8206	0.9527
Segnet	0.9974	0.9922	0.9774	0.9951
U-Net	0.9983	0.9927	0.9849	0.9967
Pspnet-resnet50	0.9893	0.9762	0.9136	0.9797
Pspnet-mobilenet	0.9916	0.9580	0.9248	0.9836
Deeplabv3plus	0.9855	0.8661	0.8531	0.9715
CBAM-Unet	0.9971	0.9895	0.9745	0.9942
CBAM-Segnet	0.9978	0.9941	0.9801	0.9957
GAN-CBAM-Segnet	**0.9985**	**0.9964**	**0.9860**	**0.9969**
GAN-CBAM-UNet	0.9962	0.9921	0.9652	0.9926

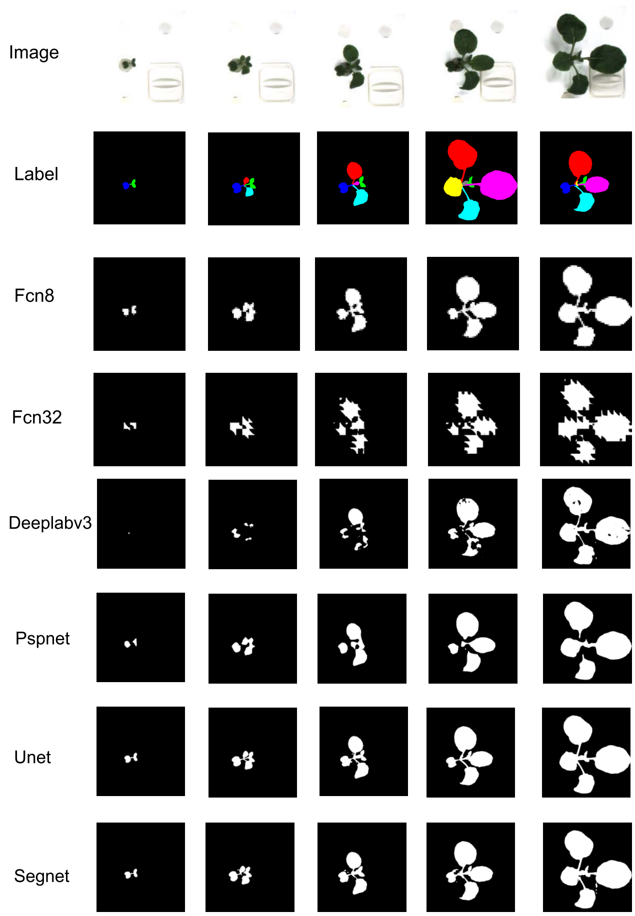

FIGURE 8. The plant leaves segmentation image of different models.

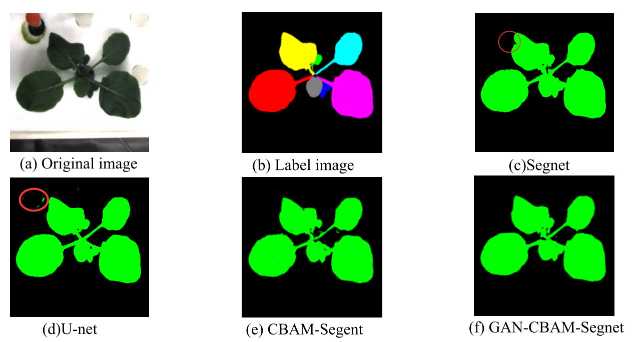

FIGURE 9. The segmentation results of a phenotypic leaf.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET

The KOMATSUNA dataset (Hideaki Uchiyama, Kyushu University) is used in this study for semantic segmentation, tracking, and reconstruction of leaves in depth images. There are a total of 900 pictures, and the picture size is 480 $\times$ 480 pixels. Meanwhile, the label images are layered according to the background and the target area. Then, the resulting segmented picture labels are converted into binary images. Finally, the images are divided into a training set and a test set at a ratio of 7:3.

B. RESULTS AND ANALYSIS

Deep learning semantic segmentation networks come in a variety of flavors. Representative numerous models are selected for this research. U-net is appropriate for dense and small sample data. Its training accuracy is greater than the precision of the Segnet network. However, training speed of the U-net is slow, and too many pooling layers are required, reducing the local accuracy. Segnet network can not only extract the two-dimensional information on the surface but also the feature of spatial information. By combining the original Segnet network with CBAM in this paper, the accuracy has been significantly improved, Meanwhile, through combining the generative confrontation network with the semantic network, the precision of phenotypic of plant leaves has reached 99.85%. The accuracy of different segmentation methods on the test set is listed in Table 1.

The semantic segmentation effects of several selected representative deep learning models for plant leaves are shown in Figure 8. We attempt to segment the dataset. The emerging models with better results are Segnet and Unet networks, which accurately segment the overall outline of plant leaves. The outcomes of the FCN network model have teeth at the edges of the leaves. The performance of the FCN32 model is more evident. The results of PSPNET and Deeplab failed to produce satisfactory of plant rhizomes.

The phenotypic segmentation results of different methods are shown in Figure 9. It can be observed in the figure that the leaf image obtained by Segnet is not good for local feature processing, and the U-net network has error classification for the background. The integration of CBAM into the Segnet network ignores the connection between pixels in the stem of the leaf. In comparison, the integration of GAN retains the accuracy of segmentation and the feature connection between pixels.

The perceptual field of view of spatial attention is determined by the size of the convolutional kernel. Therefore, different dimensions of spatial attention convolutional kernels were set for experimental comparison, and the test convolutional kernel size-change was 3×3, 5×5 and 7×7. The results of the study showed that the input image size of

TABLE 2. The segmentation accuracy of different receptive fields.

	PA	MPA	MIOU	FWIOU
Spatial Attention Module**3 × 3**	0.9976	0.9946	0.9783	0.9953
Spatial Attention Module**5 × 5**	0.9778	0.9941	0.9801	0.9957
Spatial Attention Module**7 × 7**	0.9968	0.9950	0.9711	0.9938

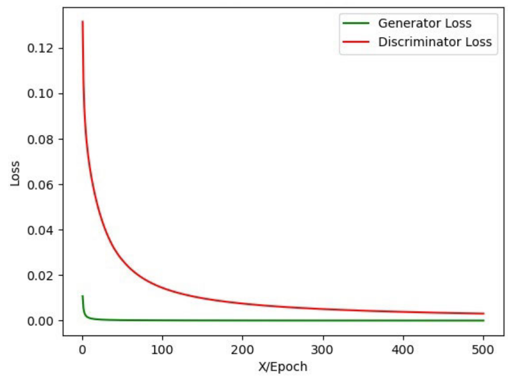

FIGURE 10. The training loss of the generated adversarial network.

480 × 480 pixels and the spatial attention convolutional kernel dimension of 5 × 5 helped achieve the best segmentation accuracy. Besides, a larger size of the spatial attention field does not necessarily lead to better efficient. The accuracies of different receptive fields are listed in Table 2.

The method of transfer learning is adopted in this study. The parameters of VGG-16 pre-trained with ImageNet are taken as initial weight; the Adam optimization algorithm is availed. The generator is trained with the manually annotated images. Then, the obtained segmented images and label images are then passed to the discriminator for judgement. Take into account this, a well-trained segmentation network model can be generated. As shown in Figure. 10, the loss of generator and discriminator gradually decreases as the number of iterations increases.

By changing the loss function of WGAN to Wasserstein, Thus, the change of the loss function completely solves the problem of instability in GAN training, and there is no need to carefully balance the generator and the discriminator in training processes. Finally, the training progress can be viewed based on the cross-entropy loss, The more the value tends to 0.5, the better the GAN training. Meanwhile, the higher the quality of the image produced by the generator, the better the split.

The segmentation accuracy of iterative training is shown in Figure.11. It can be seen from the figure that PA improves to a stable position as the number of training rises. The maximum PA reaches 99.85%. But the optimal results are not achieved at the end of training, because when the generator and discriminator reaches an equilibrium, the discriminator's recognition ability is enhanced and the segmentation requirements of the generator are more stringent, leading to overfitting of the model.

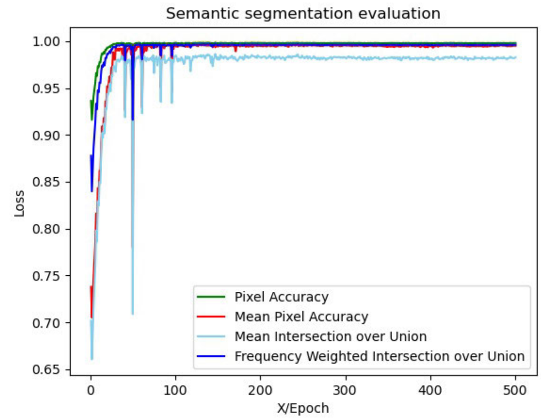

FIGURE 11. The number of iterations and segmentation accuracy.

The above results indicate that the method proposed in this paper can effectively segment the plant leaves, guiding semantic segmentation of plant leaves by computer vision.

V. CONCLUSION

In this study, Semantic segmentation of leaves was performed using deep learning methods to extract the phenotypic features of plants, and better classification results were obtained. To classify pixel categories in image segmentation, the pixel-level accuracy is high, but the relationship between pixels is ignored. Generative adversarial networks can augment data and enhance the relationship between pixels. For semantic segmentation, long and complex input sequences lose original informative features during the compression process. The attention mechanism allows the model to focus dynamically on a specific input, while the combination of it allows the segmentation network to investigate more about the characteristics of the image, Generative adversarial networks strengthen the generality of the model and the pixel-wise feature connections of images, thereby improving the accuracy of image segmentation. The model enhances the classification precision of plant leaf images, provides technical support for phenotypic feature extraction, and can be applied as a valuable tool for generalization.

FUNDING

This research received no external funding.

CONFLICTS OF INTEREST

The authors declare no competing interests.

DATA ACCESSIBILITY

The raw datasets and code are available in the Supporting Information.

REFERENCES

[1] M. Nishida, M. Namikawa, and T. Takahashi, "Climate warming has affected rice production: A proof from a long-term paddy field experiment," in *Proc. 7th Int. Congr. Crop Sci.*, Beijing, China, 2016, p. 1.

[2] S. K. Panguluri and A. A. Kumar, *Phenotyping for Plant Breeding Applications of Phenotyping Methods for Crop Improvement*. Springer, 2013.

[3] L. Silberstein, I. Kovalski, Y. Brotman, C. Perin, C. Dogimont, M. Pitrat, J. Klingler, G. Thompson, V. Portnoy, N. Katzir, and R. Perl-Treves, "Linkage map of cucumis melo including phenotypic traits and sequence-characterized genes," *Genome*, vol. 46, no. 5, pp. 761–773, Oct. 2003.

[4] R. A. T. Fischer and G. O. Edmeades, "Breeding and cereal yield progress," *Crop Sci.*, vol. 50, no. S1, pp. S-85–S-98, 2010.

[5] H. Scharr, M. Minervini, A. P. French, and C. Klukas, "Leaf segmentation in plant phenotyping: A collation study," *Mach. Vis. Appl.*, vol. 27, no. 4, pp. 585–606, 2016.

[6] D. Ward, P. Moghadam, and N. Hudson, "Deep leaf segmentation using synthetic data," 2018, *arXiv:1807.10931*.

[7] D. Kuznichov, A. Zvirin, Y. Honen, and R. Kimmel, "Data augmentation for leaf segmentation and counting tasks in rosette plants," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019.

[8] N. Valliammal and S. N. Geethalakshmi, "Hybrid image segmentation algorithm for leaf recognition and characterization," in *Proc. Int. Conf. Process Autom., Control Comput.*, Jul. 2011, pp. 1–6.

[9] J. Bell and H. M. Dee, "Leaf segmentation through the classification of edges," 2019, *arXiv:1904.03124*.

[10] N. M. Al-Shakarji, Y. M. Kassim, and K. Palaniappan, "Unsupervised learning method for plant and leaf segmentation," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2017, pp. 1–4.

[11] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "Tomato leaf segmentation algorithms for mobile phone applications using deep learning," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105788.

[12] S. Bhagat, M. Kokare, V. Haswani, P. Hambarde, and R. Kamble, "Eff-UNet++: A novel architecture for plant leaf segmentation and counting," *Ecol. Inform.*, vol. 68, May 2022, Art. no. 101583.

[13] S. Talasila, K. Rawal, and G. Sethi, "PLRSNet: A semantic segmentation network for segmenting plant leaf region under complex background," *Int. J. Intell. Unmanned Syst.*, 2021.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[15] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[16] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1925–1934.

[17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2881–2890.

[18] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[20] Y. Lu, Y. Chen, D. Zhao, and J. Chen, "Graph-FCN for image semantic segmentation," in *Proc. Int. Symp. Neural Netw.* Springer, 2019, pp. 97–105.

[21] R. Niu, X. Sun, Y. Tian, W. Diao, K. Chen, and K. Fu, "Hybrid multiple attention network for semantic segmentation in aerial images," 2020, *arXiv:2001.02870*.

[22] M. Paul, C. Mayer, L. Van Gool, and R. Timofte, "Efficient video semantic segmentation with labels propagation and refinement," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2873–2882.

[23] Q. Geng, H. Zhang, X. Qi, G. Huang, R. Yang, and Z. Zhou, "Gated path selection network for semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 2436–2449, 2021.

[24] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 5, pp. 1–32, Oct. 2021.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[26] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[27] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 6688–6697.

[28] B. Chen, Y. Huang, Q. Xia, and Q. Zhang, "Nonlocal spatial attention module for image classification," *Int. J. Adv. Robotic Syst.*, vol. 17, no. 5, Sep. 2020, Art. no. 172988142093892.

[29] I. J. Goodfellow, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[30] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[31] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[34] L. Yu, X. Long, and C. Tong, "Single image super-resolution based on improved WGAN," in *Proc. Int. Conf. Adv. Control, Automat. Artif. Intell. (ACAAI)*, Shenzhen, China, 2018, pp. 21–22.

[35] L. Yu, X. Long, and C. Tong, "Single image super-resolution based on improved WGAN," in *Proc. Int. Conf. Adv. Control, Automat. Artif. Intell. (ACAAI)*, Shenzhen, China, 2018, pp. 21–22.

LIYING CAO was born in Dehui, Jilin, China, in 1978. She received the Ph.D. degree from Anhui Agricultural University, China.

She currently works with the College of Information and Technology, Jilin Agricultural University. She is a Postdoctoral Fellow at the School of Resources and Environment, Jilin Agricultural University, a Visiting Scholar at Texas Tech University (TTU), and a member of the Digital Agriculture Branch of the Chinese Agricultural Society. She has published more than 40 articles and written six academic monographs. Her research interests include data mining and agricultural informatization.

HONGDA LI was born in Chaoyang, Liaoning, China, in 1998. He received the degree in computer science and technology from Shenyang Aerospace University, China, in 2020. He is currently pursuing the master's degree with Jilin Agricultural University.

He was qualified as a Software Designer, in 2019. His research interest includes artificial intelligence. He has won several provincial and national awards.

XUERUI LIU was born in Dongjin, Hubei, China, in 1996. He received the bachelor's degree from the North China Institute of Aerospace Engineering, China. He is currently pursuing the master's degree with Jilin Agricultural University.

His research interest includes agricultural informatization. He was awarded the title of the Outstanding Graduate of Hebei Province in 2019 and passed the CET-6.

GUIFEN CHEN was born in Changchun, Jilin, in 1956. She received the B.S. degree in geochemistry from the Department of Geochemistry, Jilin Agricultural University, in 1982, and the M.S. and Ph.D. degrees in computer application technology from the School of Computer Science, Jilin University, in 1999 and 2009, respectively.

Her research interests include application theory and technology of computer agriculture, such as expert systems, data mining, precision agriculture, and the Internet of Things agriculture.

Dr. Chen received several awards, including the First Prize of Science and Technology Award of China Business Federation, Research on New Technology of Intelligent Decision Making and Agricultural Application, in December 2014, the First Prize of Science and Technology Progress Award of Jilin Province, Research and Development of Large-scale Modern Agricultural Digital Technology Application, in 2006, and the First Prize of Jilin Provincial Science and Technology Progress Award, Research and Application of Spatio-Temporal Reasoning, Data Mining and Decision Making System, in 2008.

HELONG YU was born in October 1974. He received the Ph.D. degree.

He is currently the President of the Department of Information Technology, Jilin Agricultural University, a Postdoctoral Fellow with Tsinghua University, and a Visiting Scholar with Massey University, New Zealand. He has presided over more than 20 projects at all levels, such as the National Natural Science Foundation of China, published more than 70 articles at all levels, such as SCI and EI, published five books, and won five first and second prizes for scientific and technological progress at provincial and ministerial levels. He has guided students to participate in various provincial competitions and won two gold medals, one silver medal, and one bronze medal.

Dr. Yu is a member of the 6th Batch of Jilin Province Top Innovative Talents and a member of the Youth Committee of the Provincial Democratic League. He is a member of the Jilin Provincial Computer Teaching Steering Committee, the Leader of the Provincial Advantageous Characteristic Discipline (computer science and technology), the Head of the Provincial First-Class major (computer science and technology), the Head of the Provincial Excellent Engineer Program Pilot major (Internet of Things engineering), the Head of the Provincial Excellent Class (computer network), and the Deputy Director of the Changchun Higher Education Working Committee of the Democratic League. He is the Head of the Jilin Precision Agriculture and Big Data Engineering Research Center, the Director of the Jilin Computer Society, the Vice Chairperson of the CCT YOCSEF Changchun, and the Expert of Jilin "12316" New Rural Hotline. Executive Vice President of the Institute of Intelligent Agriculture of Jilin Agricultural University.