Few-Shot Grape Leaf Diseases Classification Based on Generative Adversarial Network

Mingzhao Zeng¹², Huiyi Gao¹, Li Wan¹

¹ Institute of Intelligent Machines, Hefei Institutes of Physical Science, CAS, Hefei 230031, China
² University of Science and Technology of China, Hefei 230026, China
*e-mail: hygao@iim.ac.cn; wanli@iim.ac.cn

Abstract. The treatment and prevention of crop diseases have an extremely important impact on the yield and quality of crops. In recent years, with the development of computer vision and deep learning technology, research on crop disease recognition based on leaf images has received extensive attention. In the field of grape disease recognition, the lack of large-scale diseased leaf labeling data sets limits the accuracy of recognition, and obtaining professional grape disease data sets requires a lot of manpower and material resources. Aiming at the problem of the lack of grape leaf data set, this research proposes a data generation model based on the cycle Generative Adversarial Network model which introduced an leaf foreground module (LFM) block. Experiments show that the model can generate high-quality grape leaf disease images, which can improve the accuracy of grape disease recognition task in a Few-Shot Grape Leaf Diseases Classification task.

1. Introduction

Crop diseases caused by various fungi, bacteria and pests, etc. are one of the main factors affecting crop yield and agricultural economy. In order to minimize the production losses caused by pests, it is especially important to prevent and implement appropriate control measures for agricultural pests.

Diseases of grapes tend to cause plant leaves to appear different color textures and protoplast feature information of shape size, so artificial grape illness recognition methods require agricultural experts to observe and judge the grape leaves according to long-term accumulation. When the researchers choose a characteristic object, there is a lack of recognized basis. There is a real-time sexuality, low work efficiency, and labour-consuming problems, etc., and it is impossible to provide key early information for the decision-making process. Therefore, the technique of implementing automatic identification of crop disease type is recognized by computer. Automatic diagnostic techniques can not only increase the speed of identification, but the accuracy is higher than the accuracy of artificial diagnosis.

In recent years, more and more artificial intelligence image recognition techniques have been applied to the field of agricultural engineering, and the early disease testing of crop has provided great help. Its core is based on deep learning and image recognition technology of convolutional neural networks.

Therefore, the accuracy of image recognition will directly affect the prevention effect of crop diseases. At present, deep learning has achieved great progress in the field of smart agriculture, far exceeding the traditional agricultural disease recognition method. However, deep learning is a data-driven technology that requires sufficient amount of data to make neural network models to learn possible distribution. However, it is difficult to obtain enough diseases in actual work to ensure the
training of neural networks. Traditional data enhancements have multiple, such as color jitter, contrast enhancement, random zoom, random crop, shift, horizontal / vertical flipping, etc. However, the characteristics of the blade disease tend to be related to the color distribution of the image, contrast. When the number of image samples is small, the operation such as a random crop and flip is difficult to supplement the potential distribution law of the data implicit [1,2].

2. Related Works

In 2014, Ian Goodfellow et al. For the first time, the concept of generating adversarial network (GAN) [3], its network training does not rely on any prior hypothetical assumption, not using great likelihood estimates, and transmits the network forward from one of its composition The model can generate a sample that approximates the real distribution of training data, thereby implementing analog to the existing category data. The original GAN network has a series of problems such as difficult to converge, training instability and model uncontrollable. DCGAN [4] is the first network that combines convolutional neural networks with GAN, and its generator G and Judge D are implemented using convolutional neural networks, which successfully solves the problem of instability of the original GAN network training. Pix2Pix [5] generative adversarial network make the image to the image translation field have developed greatly. Image translation is a generated task, generating an image in this task is based on an additional condition input image. It is difficult to obtain a dataset with the input and output samples and the images that cannot be collected to be paired. Zhu et al. Proposed a unsupervised image translation model CycleGAN [6]. The model uses two group generators and discriminations to learn mappings and reverse mappings between source domain images and target domain images. The non-supervised model does not require image data to be paired one by one, and only the picture belonging to an image domain has the same properties.

3. Proposed method

3.1 Traditional CycleGAN model

The CycleGAN model consists of two group generator mappings: \( G: X \to Y \) and \( F: Y \to X \), with each associated discriminator \( D_Y \) and \( D_X \). CycleGAN's core idea is to introduce a cycle consistency loss function, which is equivalent to content loss in neural style transfer, which ensures that the generated image must retain the characteristics of the original image. Its mathematical formula is expressed as follows:

\[
\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim \text{data}(x)}[\| F(G(x)) - x \|_1] + \mathbb{E}_{y \sim \text{data}(y)}[\| G(F(y)) - y \|_1]
\]

(1)

\[
\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim \text{data}(y)}[\log D_Y(y)] \\
+ \mathbb{E}_{x \sim \text{data}(x)}[\log(1 - D_Y(G(x)))]
\]

(2)

Where \( X, Y \) are healthy grape leaf images and diseased leaf images, \( G \) and \( F \) are generators with same parameters and architecture. \( D_Y \) and \( D_X \) are two discriminator identifying real samples or fake. The schematic diagram of the model is shown in Figure 1:
3.2 leaf foreground module
The attention mechanism is an important concept in the field of deep learning and neural network. It is initially used in the field of machine translation. It is now widely used in the computer visual field. The attention mechanism can use the human visual mechanism to interpret it. Its basic thinking is to let the neural network model can ignore the relevant information like human vision. When the convolutional neural network is doing an image classification task, the last layer is usually a softmax layer, and the maximum value corresponding is a classification category. If back propagation from the node of the maximum probability classification category, the gradient of the last convolution layer can be obtained, then the local area of the image attention to the neural network discrimination category can be observed. Two discriminators need to learn whether the data generated by the corresponding domain generator is the domain data by learning the real data of the respective domain. Briefly, the discriminator is a two-class neural network model distinguishes between real data and generating data.

We add attention mechanisms in the CycleGAN model called leaf foreground module (LFM), want to generate an image feature that is able to pay attention to the image characteristics that you need to convert, making the model more pay more attention to the conversion of the foreground target rather than the background. The LFM block could produce an activation map $A(x)$ after input a leaf image $x$. The activation map $A(x)$ represents the image area that the generator should pay attention to. The improved LFM GAN structure is shown in Figure 2.

3.3 loss functions
The generated network module is decomposed into a classic generator module and a LFM module. The activation map generated by the X-domain image sample $X$ should be consistent with the output image $G(x)$ of the generator $G$. Therefore, the focus activation map generated by constraining two attention modules is satisfied the constraint: $A_X(x) \approx A_Y(G(x))$ and $A_Y(y) \approx A_X(F(y))$. To do this, add a LFM activation loss function on the loss function of CycleGAN:

$$
\mathcal{L}_A(A_X, A_Y) = \mathbb{E}_{x \in X} [\| A_X(x) - A_Y(G(x)) \|_1] + \mathbb{E}_{y \in Y} [\| A_Y(y) - A_X(F(y)) \|_1].
$$

The total loss function of this paper is as shown in the formula (4):

$$
\mathcal{L}(X, Y) = \mathcal{L}_{CE}(X, Y) + \lambda \mathcal{L}_{cycle}(X, Y) + \mathcal{L}_A(A_X, A_Y)
$$

(3)
\[
\mathcal{L}(G, F, D_X, D_Y, A_X, A_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) + \lambda \mathcal{L}_{\text{cyc}}(G, F) + \lambda_A \mathcal{L}_A(A_X, A_Y)
\] (4)

4. Experiments

We set up a unbalanced few-shot grape disease subset using the PlantVillage dataset. The number of corresponding test sets per class is 100 sheets. The size of each picture is 128×128 pixels.

| Class         | Numbers |
|---------------|---------|
| Esca          | 10      |
| Leaf blight   | 10      |
| Black rot     | 300     |
| Healthy       | 300     |

Table 1. Grape leaf disease training set

This experiment uses the grape health leaf images as input, generates black rot disease, leaf blight disease and black spot leaf images. The baseline model is used is a classic CycleGAN model, and then compares the experiment using the improved CycleGAN model based on the LFM activation map based CycleGAN model. Means of we only use 10 different disease images as a target domain to perform image translation experiments.

This experiment employs the quality of the generated image by Fréchet Inception Distance (FID) [8]. FID distance measures the real picture and generating the distribution distance of the picture in high-dimensional features, reflects the similarity of the two types of pictures. FID measures the similarity of the two sets of images from the statistical aspect of the computer visual features of the original image, which is obtained by using the INCEPTION V3 [9] image classification model. The lower the score, the more similar images, or the more similar quantities of both, the FID is divided into 0.0 in the best case, indicating that the two sets of images are the same. The FID score is used to evaluate the quality of the image generated by the generative counterfeit network, and the lower score is highly dependent on the higher quality image. The results obtained by calculating the FID value are shown in Table 2.
Table 2. Different models generate image quality assessment

| Task                  | FID distance |
|-----------------------|--------------|
| CycleGAN Esca         | 220.37       |
| LFMGAN Esca           | **185.78**   |
| CycleGAN Leaf blight  | 175.78       |
| LFMGAN leaf blight    | **117.75**   |

The image generated by the two methods is shown in Figure 4:

![Figure 4. fake grape disease leaf images generated by two methods](image)

The ratio of unbalanced grape data sets was expanded to 300. Then use the ResNet18 residual network model [7] to identify. Compare the identification accuracy before its data augmentation. The result is shown in Table 3. The results showed that CycleGAN improved the accuracy of grape disease from 83.59% to 90.91%. The improved model LFMGAN of this paper further improved the accuracy of 95.50%.

Table 3. Comparison of the classification accuracy of this paper and baseline model

| Model                          | Accuracy |
|--------------------------------|----------|
| original dataset               | 83.59    |
| Augmentation dataset(CycleGAN) | 90.91    |
| Augmentation dataset(LFMGAN)   | **92.44**|

5. Conclusion
This paper uses the LFMGAN network to convert a healthy grape leaf image into the grape disease leaf image that is not easy to get, which greatly expands the scale of the training data of grape disease classification tasks. The leaf foreground module (LFM) block activation map is introduced. The auxiliary activation map generated by the LMF module enables the network more attention to the key area and texture details of the image conversion. The experimental results show that the improved LFMGAN model can effectively improve the accuracy of grape disease recognition with the lack of data. The results demonstrate that the proposed model can reach the expected data augmentation effect.
References

[1] Geetharamani, G., and J., A.P. (2019) Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers and Electrical Engineering, 76: 323–338.

[2] Ferentinos, K.P. (2018) Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145: 311–318.

[3] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In: Advances in Neural Information Processing Systems. Montreal. pp. 2672–2680.

[4] Wu, Q., Chen, Y., and Meng, J. (2020) Dcgan-based data augmentation for tomato leaf disease identification. IEEE Access, 8: 98716–98728.

[5] Isola, P., Zhu, J.Y., Zhou, T., and Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. Honolulu. pp. 5967–5976.

[6] Zhu, J.Y., Park, T., Isola, P., and Efros, A.A. (2017) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision. Venice. pp. 2242–2251.

[7] He, K., Zhang, X., Ren, S., and Sun, J. (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas. pp: 770–778.

[8] Chong, M.J., and Forsyth, D. (2020) Effectively unbiased FID and inception score and where to find them. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Seattle. pp. 6069–6078.

[9] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas. pp. 2818–2826.