Image Recognition of Crop Disease Based on Generative Adversarial Networks

Tianci Liu[1,a] and Xueyun Chen[1,b]

[1]College of electrical engineering, Guangxi University, Nanning, Guangxi Province, China
a. 2873098858@qq.com, b. cxy177@163.com

Keywords: Convolutional neural network, Generative Adversarial Network(GAN), deep learning, crop disease.

Abstract: At present, the mainstream methods for detecting crop diseases rely mainly on using color and artificially designed feature operators, and the detection accuracy is not high. The state-of-the-art deep learning methods require a large number of training sample images. Relying on the traditional Generative Adversarial Network is difficult to generate images that meet the quality requirements. This paper proposes a light mask Generative Adversarial Network(LMGAN) to identify generate grape leaf diseases images with controllable shape and light intensity. By extracting texture and brightness information of the leaf, the corresponding morphological mask label and light mask label are made and used as the training input of the network. Using the convolutional neural network to detect grape leaves disease. The experimental results show that the proposed algorithm achieves better accuracy than the existing mainstream algorithms.

1. Introduction

Crop diseases detection has a significant meaning to precision agriculture. Due to the influences from various illumination, quantity, shape, and light, it remains far away from being solved.

Current various image detection methods for crop diseases and insect pests can be roughly divided into three types, image-based, optical, and biological detection [1-3]. Image-based color segmentation methods are widely used in most image recognition tasks. Lanlan [4] proposed a detection method based on color segmentation, by extracting the image color, brightness and direction features for identification and detection. Pareek et al. [5] proposed an image segmentation method based on chromatic aberration for detecting diseased areas of rice, in order to spray pesticides on diseased rice plants in time. Later, Fan et al. [6] used the normalized segmentation algorithm based on the spectrogram theory to segment the disease and pest images. Solemane et al. [7] proposed a method combining transfer learning and feature extraction to identify mildew in pearl millet. Later, Tian Youwen et al [8] achieved better results than neural network recognition by extracting the color and texture features of grape diseased leaves and using Support Vector Machine (SVM). Wang Xianfeng et al. [9] extracted color, shape, texture and other features of leaf lesions, combined with environmental information. Zhang et al. [10] also extracted the color, shape and texture features of the lesions after segmentation of the spots, and then identified the five corn leaves through the K-Nearest Neighbor (KNN) classification algorithm. Yuan Lin [11] proposed a method for distinguishing wheat diseases and insect pests based on imaging hyper spectral

technology from optical point of view. Tan Feng et al. [12] established a multi-layer BP (Back Propagation) neural network model to realize the disease identification of soybean leaves. Deep learning methods represented by deep convolutional neural networks are used in image recognition [13-17]. Mehedi et al. [18] used the most advanced face recognition depth model for automatic recognition in cloud environments such as video surveillance, with an accuracy of 95.84%. Yingchao et al. [19] in high-resolution aerial photography images have achieved very good results for target recognition. However, in the field of crop disease detection, it has two defects: 1. Deep learning methods cannot give an precise result in the case of insufficient samples; 2.The computing of convolutional neural network must be iterated many times.

Generative adversarial network (GAN) has made great progress in image generation. To address these problems, we put forward a light mask feature fusion convolutional network structure based on GAN, named as LMGAN. The main contributes are listed as:

1) A light mask generative adversarial network (LMGAN) is proposed to generate grape disease image of different features from the mask image.

2) The special design of the discriminator fixed convolutional-kernels make it suitable for reducing training time.

2. Related Work

2.1.Convolutinal Neural Network Generative Adversarial Network

A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units.

2.2. Generative Adversarial Network

The traditional GAN[20] consists of a generative network(G) and a discriminant network(D). Suppose that the real image x satisfies the probability distribution P_{data}, random noise z is passed into the generating network to generate a fake image, and z satisfies the probability distribution P_z. Discriminant accepts real images and generates images and judges whether they are true or false. Define the objective function of GAN as V (G, D), it can be expressed as:

$$\min_G \max_D (G, D) = E_{x \sim P_{\text{data}}} [\log D(x)] + E_{z \sim P_z} [\log(1 - D(G(z)))] \tag{1}$$

Since the original GAN could not control the image categories generated by the generator, some researchers later proposed a conditional generation adversarial network [21] (Conditional Generative Adversarial Network, CGAN). This work proposes a conditionally constrained GAN, which introduces a condition vector y in the modeling of generative model D and discriminant model G. Using y to add information to the model can guide the data generation process. This simple and direct improvement has played a certain role in controlling the category of the generated image, and it objective function is as equation (2):

$$\min_G \max_D (G,D) = E_{x\sim P_{\text{data}}}[\log(D(x,\boldsymbol{y}(x)))]+ \\ E_{x\sim P_{\text{data}},z\sim P_z}[\log(1-D(G(z,\boldsymbol{y}(x)),\boldsymbol{y}(x)))] \tag{2}$$

In order to get deeper image features and improve the convergence ability of the model. In 2017, Isola et al [22] proposed an image-to-image translation with Conditional Adversarial Networks (Pix2Pix) structure based on CGAN [14]. Here comes a better effect. So Pix2Pix also added L1loss [22] to the generator G to minimize it, and it improved objective function is shown in equation (3).

$$\min_G \max_D V = E_x[\log(D(x,\boldsymbol{y}(x)))]+ \\ E_{x,z}[\log(1-D(G(z,\boldsymbol{y}(x)),\boldsymbol{y}(x)))]+ \\ E_{x,z}[(\| x-G(z,\boldsymbol{y}(x)),\boldsymbol{y}(x))\|)] \tag{3}$$

From the experimental results obtained by Pix2Pix, although the images generated by the generator have certain constraints on the shape category, all the images still cannot be effectively controlled in terms of brightness. However, in the field of crop disease detection, whether it is possible to obtain disease images of different shapes and light brightness is very important for experimental research.

3. Proposed Method

We propose the framework LMGAN to tackle the diseased grape leaves image generation problem. Firstly, it uses the canny arithmetic to extract the shape and texture of the leaf image. Second, the light mask is realized by calculating the gray average value of the original image. Inputting both as condition information at the same time, which provide a great help for subsequent experiments. The network can learn the corresponding relationship between the input mask and the original images, so as to ensure the detail clarity of Generated images.

3.1. Objective Function

The objective function of the LMGAN network is defined as equation (4):

$$\min_G \max_D (G,D) = E_{x\sim P_{\text{data}}}[\log(D(x,\boldsymbol{y}(x),\boldsymbol{l}(x)))]+ \\ E_{x\sim P_{\text{data}},z\sim P_z}\log[(1-D(G(z,\boldsymbol{y}(x),\boldsymbol{l}(x)),\boldsymbol{y}(x),\boldsymbol{l}(x)))] \tag{4}$$

Where the noise z, condition information y, and illumination mask l are input to the generator at the same time, and the three are jointly input as the hidden layer. Similarly, in the discriminator, the image x, the condition information y, and the light mask l are also connected together as the hidden layer input. Generator continuously tries to maximize the objective function, and D continuously iterates to maximize and minimize the objective function, so as to achieve the purpose of generating images of grape leaf black rot disease with artificially controllable shape and brightness.

3.2. System Structure

In Figure 1, we can see thatThe LMGAN proposed in this paper contains two parts of the training process: generator and discriminator. For the generator G, on the one hand, by inputting the shape

texture mask and the illumination brightness mask, a corresponding natural leaf image is generated at the output end. Compare the manually processed standard blade image with it, calculate the error and use it to correct the parameters and weights of the generator G. On the other hand, the generated image and the corresponding shape texture mask and illumination mask are simultaneously input to the discriminator. Because the purpose of the generator is to produce a more realistic image so that it can fool the discriminator, so after comparing and calculating the error between the output of the discriminator and the real result, back propagation is carried out, so that the generator is networked again. Structural parameter adjustment and weight optimization.

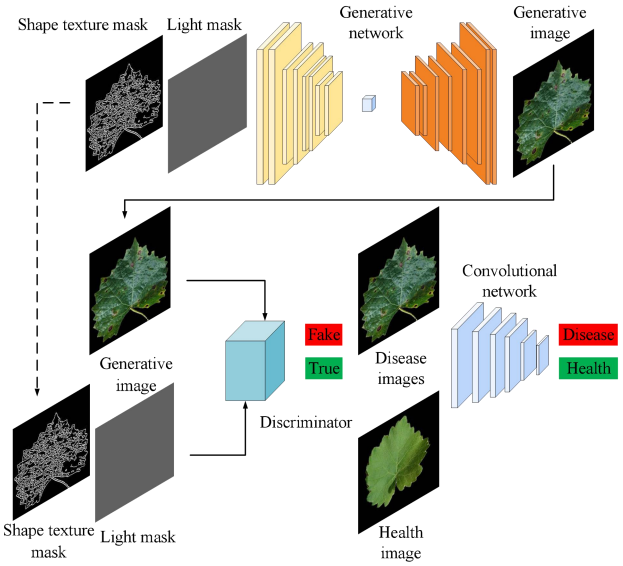

Figure 1: System structure.

The generator of LMGAN contains 8 convolution layers and 8 deconvolution layers. The size of each convolution kernel is 5 × 5, and the step size of each convolution or deconvolution operation is 2. The LeakyReLu function is used as the activation function of each layer in the convolution part. In the deconvolution part, the ReLu function is used as the activation function of each layer. And after each layer, batch normalization is used to improve the convergence performance of the network. After generating a 256 × 256 × 3 grape leaf image input by the network, after 8 layers of convolution, a 1 × 1 × 512 leaf feature map (Featuremap) was finally obtained. Afterwards, an 8-layer deconvolution operation is performed, but after each deconvolution operation, the network combines the output of the previous deconvolution layer with the output of the corresponding convolution layer. The network ends with a fully connected layer using Tanh. For the discriminator D, since the role of the discriminator is to judge that the real picture is true and the generated picture is false under the conditions of the input shape texture mask and the light mask. Therefore, on the one hand, the discriminator needs to perform error calculation on whether the discrimination result is true under the condition of inputting the real image and the mask image. At the same time, it must adjust various parameters and weights to optimize its discrimination effect in time. On the other hand, under the condition of inputting the generated image and the mask image, it is necessary to perform error calculation on whether the authentication result is false, and again optimize the weights and parameters of the authentication network to improve the discriminator's Discrimination ability.

In addition, in the network of discriminators, the experiment found that the training process is prone to generate network capabilities and identify the imbalance of network capabilities. Therefore,

by limiting the network depth of the discriminating network, the ability to discriminate the network and the training capacity of the generating network are balanced.

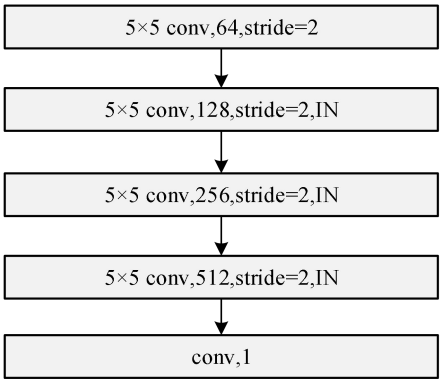

Figure 2: Discriminator network structure.

The discriminant network structure is composed of five convolutional layers as shown in Figure 2, without using fully connected and pooling layers. BatchNormal layer is replaced with InstanceNormal layer. This is because the experiment found that removing the BatchNormal layer helps to make the training network more stable and at the same time reduces the computational complexity, and the addition of BatchNormal makes the generated pictures have a fuzzy effect, making it difficult to achieve network training. stable. The InstanceNormal layer is the depth information in the network input direction minus the average divided by the standard deviation to speed up the network training speed, so that the network's nonlinear fitting ability has been greatly improved, so it is replaced by the InstanceNormal layer.

In this paper, an independent volume integral network is used to classify the final disease. The network consists of 6 convolutional layers with a kernel of 3 × 3. The first 5 convolutional layers are followed by a batch normalization layer and the LeakyReLU activation function to obtain a 1 × 1 × 256 feature vector. The shared vector is input to the first Six convolutional layers are connected to the Softmax layer to classify grape leaf diseases.

In the LMGAN generation adversarial network, the purpose of the generator is to allow the generated image to "deceive" the discriminator, and the purpose of the discriminator is to generate a disease and pest image that is as similar as possible to the original image data distribution and the controllable light intensity Find out "fake pictures". The detailed algorithm flow is as follows.

Algorithm 1:

Input: Real image dataset x, shape mask y, light mask l; Generative image $G(y, l)$, Parameters to be optimized ω; Loss functions used include $Loss_g$, $Loss_d$; Maximum times t_1 in first stage; Maximum times t_{max} in total.

1. For $t < t_{max}$ do
2. Obtain the original blade image, corresponding shape and texture, and light mask image from the real blade image data set in batches;
3. Take x, y, l in as input and get output through network
4. If $t < t_1$ do

Calculate $Loss_g(x, G(y, l))$

```
Else do
Calculate Loss_d (x,G (y, l) )
5. Update ω, return to step 1
6. End for
```

4. Experiment

In this section, we use experiments to validate the effectiveness of the proposed method. We use mini-batch SGD and apply the Adam solver, with a learning rate of 0.0002, and momentum parameters β1 = 0.5, β2 = 0.998. The mini-batch for each training is set to 2, and Epoch is set to 500.

We evaluate our model on the diseased grape leaves database, which contains more than 50,00 photos, and each leaf includes different texture, light, and poses. In this paper, we select 3000 diseased photos of black rot in the dataset, and normalize them to the size of 256×256. The training set and the test set are separated according to the ratio of 8: 2.

To evaluate the performance of diseased grape leaves generate, visualization comparison and objective indicators are used as the quality evaluation indicators of diseased grape leaves generate.

4.1. Visualization Comparison with Other Models

The designed network model reached convergence after several iterations. Figure 3 is the corresponding image generated after the network is input with the corresponding shape texture and light mask when the number of iterations of the network during the training process is 1, 50, 100, and 500, respectively. As can be seen from Epoch10 in Figure 3, after 10 iterations of the entire training set, the blade contour has not been fully displayed; through Epoch50, it can be seen that after 50 iterations of the entire training set, the blade images generated by the model are clear, but black The rot disease area and the corresponding texture details are still not well displayed; according to Epoch100, after 100 iterations of the entire training set, partially unstable images are gradually determined, and the details are not stable; Finally, through Epoch500 It can be seen that after 500 iterations of the entire training set, the model is basically stable, and the resulting black rot grape leaves are clearly visible.

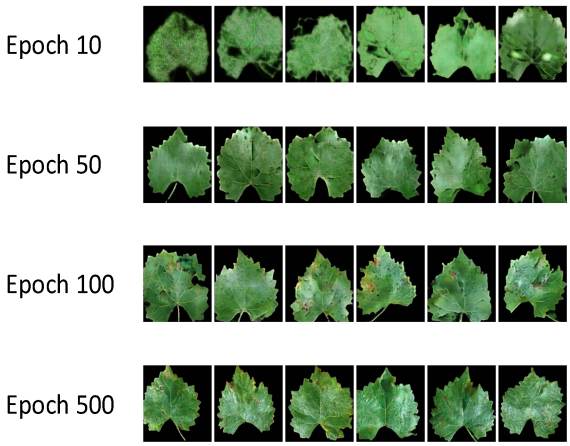

Figure 3: LMGAN Training output image.

In order to verify the quality of the black rot image of raw grape leaves by the LMGAN method, the generation method proposed in this paper is compared with the generation effects of traditional GAN and DCGAN.

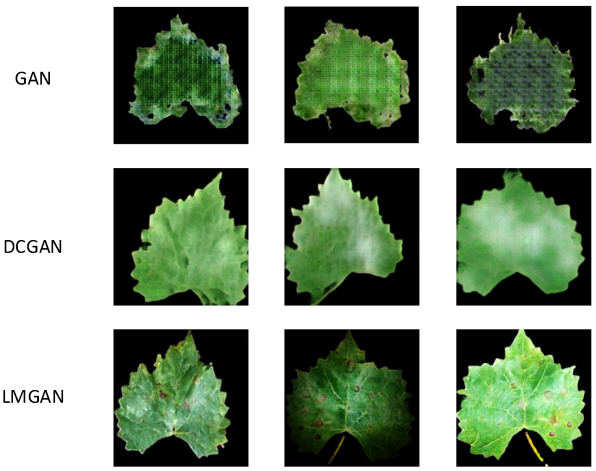

Figure 4: Different models generate grape leaf black rot leaf images.

Figure 4 is a comparison of grape leaf black rot image samples generated by traditional GAN and DCGAN and the LMGAN method proposed in this paper. Observing the image, we can see that the sample image generated by the traditional GAN can only vaguely see its approximate shape through the outline, which is very blurry. Although the sample image generated by DCGAN has a clear shape and contour, it loses too much texture, and the brightness information cannot be controlled. The LMGAN method proposed in this paper is superior to the first two methods. The generated image can not only accurately control the shape and texture information, but also effectively control the brightness part. Most of the images can achieve the effect of being false and can effectively expand the training samples for grape black rot leaf image recognition.

4.2. Quantitative Comparison

We measure the reconstruction effect from two objective evaluation indicators: structural similarity (SSIM) and peak signal-to-noise ratio (PSNR).

4.2.1. SSIM

SSIM can evaluate the similarity of two pictures effectively and objectively. It measures in terms of brightness, contrast and structure, which is more consistent with the visual recognition and perception characteristics of human eyes. Range of SSIM is between 0 and 1. The larger the SSIM value, the higher the similarity of the two pictures. When the two images are the same, the SSIM value is 1. The SSIM of the reconstructed image I and the original image I_{HR} can be obtained according to equation (8), where μ_I and $\mu_{I_{HR}}$ are the average of the images I and I_{HR}. σ_I^2 and $\sigma_{I_{HR}}^2$ are the variance of I and I_{HR}, $\sigma_{II_{HR}}$ is the covariance of I and I_{HR}. $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, they are constants. L is the dynamic range of image pixels, $k_1 = 0.01$, $k_2 = 0.03$.

$$\mathrm{SSIM}(I, I_{HR}) = \frac{(2\mu_I \mu_{I_{HR}} + c_1)(2\sigma_{II_{HR}} + c_2)}{(\mu_I^2 + \mu_{I_{HR}}^2 + c_1)(\sigma_I^2 + \sigma_{I_{HR}}^2 + c_2)} \tag{5}$$

Table 1: Comparison of SSIM values of various algorithms.

Factor SSIM	Algorithms		
	GAN	DCGAN	LMGAN
1	0.8023	0.9438	0.9825
2	0.7946	0.9116	0.9632
3	0.7754	0.8943	0.9337

Table 1 shows the SSIM values between the original images and the reconstructed images of various algorithms. As is observed from the table, compared with other algorithms, the SSIM values of this method are the largest. It indicates that the images generated by LMGAN are more realistic and have the highest similarity with the original images.

4.2.2. PSNR

PSNR is another objective indicator for evaluating image quality. It directly compares the pixel differences between the two pictures. The larger the value, the better the image quality. Assuming that both the sizes of two images I and I_{HR} are $m \times n$, the mean square error (MSE) between them is:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - I_{HR}(i,j)]^2 \tag{6}$$

Calculate PSNR through MSE, the formula is as equation (7):

$$PSNR = 10 \cdot \log_{10}(\frac{255^2}{MSE}) \tag{7}$$

Table 2: Comparison of PSNR values of various algorithms.

Factor PSNR (dB)	Algorithms		
	GAN	DCGAN	LMGAN
1	18.14	24.19	27.69
2	16.53	23.51	25.36
3	15.79	21.97	24.45

From Table 2, we can see that the algorithm we proposed has the largest PSNR value and the quality of the image is the best. It also indicates that the images generated by LMGAN are more realistic and have the highest similarity with the original images.

The light mask generation adversarial network model proposed in this paper improves the quality of the generated image by introducing light mask and shape texture mask. In addition, according to different needs, leaf images with different shapes, textures, and lightness can be generated manually. It is not difficult to see from Table 4 that, compared with the traditional grape leaf disease identification method based on artificial features, the addition of convolutional neural networks and the generation of adversarial networks simultaneously greatly improve the recognition accuracy of grape leaf black rot. It also proves that the generation of artificially controllable and high-quality

images plays an important role in the identification of crop diseases. These are of great significance for the realization of agricultural automation and the development of precision agriculture.

5. Conclusions

In this paper, a light mask Generative Adversarial Network(LMGAN) is proposed. Experiments show that the LMGAN with Convolutional Neural Network not only improves the prediction results of the network, but also provides the optimization performance beyond the traditional conditional generative adversarial nets. It does not increase the scale of the network model and affect the running speed of the network model. In the future work, the loss function of the LMGAN should be further improved to obtain better performance.

References

[1] Low-energy electron beams for protection of grain crops from insect pests and diseases[J]. T V Chizh, N N Loy, A N Pavlov, M S Vorobyov, S Yu Doroshkevich. Journal of Physics: Conference Series. 2018 (2): 79–87.

[2] Parul Sharma,Yash Paul Singh Berwal,Wiqas Ghai. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation[J]. Information Processing in Agriculture,2019: 103–110.

[3] Chang Yue. Optimization of methods for identifying plant diseases and insect pests [C]. Chinese Plant Protection Society. Proceedings of the 2019 Annual Conference of Chinese Plant Protection Society. Chinese Plant Protection Society: Chinese Plant Protection Society, 2019,24: 86–92.

[4] Wu L. Detection of salient region of in-field rapeseed plant images based-on visual attention model[C]// 2017 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS). IEEE, 2017: 215–221.

[5] V.K. Tewari,C.M. Pareek,Gurdeep Lal,L.K. Dhruw,Naseeb Singh. Image processing based real-time variable rate chemical spraying system for disease control in paddy crop[J]. Artificial Intelligence in Agriculture, 2020: 525–536.

[6] Tongke Fan,Jing Xu. Image Classification of Crop Diseases and Pests Based on Deep Learning and Fuzzy System[J]. International Journal of Data Warehousing and Mining (IJDWM), 2020,16(2): 423–431.

[7] Solemane Coulibaly, Bernard Kamsu-Foguem,Dantouma Kamissoko, Daouda Traore. Deep neural networks with transfer learning in millet crop images[J]. Computers in Industry, 2019, 108: 367–371.

[8] Tian Youwen, Li Tianlai, Li Chenghua, et al. Method for recognition of grape disease based on support vector machine[J]. Transactions of The Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2007, 23(6): 175-180.

[9] Wang Xianfeng, Zhang Shanwen, Wang Zhen, et al. Recognition of cucumber diseases based on leaf image and environmental information[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(14): 148-153.

[10] Zhang S W, Shang Y J, Wang L.Plant disease recognition based on plant leaf image[J]. Journal of Animal & Plant Sciences, 2015, 25(3): 42-45.

[11] Yuan Lin. Multi-scale remote sensing recognition and discrimination method for wheat diseases and insect pests [D]. Zhejiang University, 2015.

[12] Tan Feng, Ma Xiaodan.The method of recognition of damage by disease and insect based on laminae[J]. Journal of Agricultural Mechanization Research, 2009, 31(6): 41-43.

[13] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[14] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes[C]//Computer Vision and Pattern Recognition. IEEE, 2014: 1891-1898.

[15] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[16] Tan Feng, Ma Xiaodan.The method of recognition of damage by disease and insect based on laminae[J]. Journal of Agricultural Mechanization Research, 2009, 31(6): 41-43.

[17] Monji Kherallah,Mohamed Elleuch. Boosting of Deep Convolutional Architectures for Arabic Handwriting Recognition[J]. International Journal of Multimedia Data Engineering and Management (IJMDEM),2019,10(4): 251–259.

[18] Mehedi Masud,Ghulam Muhammad,Hesham Alhumyani,Sultan S Alshamrani,Omar Cheikhrouhou,Saleh Ibrahim,M. Shamim Hossain. Deep learning-based intelligent face recognition in IoT-cloud environment[J]. Computer Communications, 2020,152: 77–81.

[19] Yingchao Feng,Wenhui Diao,Xian Sun,Menglong Yan,Xin Gao. Towards Automated Ship Detection and Category Recognition from High-Resolution Aerial Images[J]. Remote Sensing,2019, 11(16): 98–136.

[20] Wang K, Gou C, Duan Y , et al. Generative Adversarial Networks[J]. Acta Automatica Sinica: English version, 2017: 598.

[21] M Mirza, S. Osindero, Conditional Generative Adversarial Nets[J]. Computer Science, pp. 2672-2680, 2014: 205–209.

[22] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, "Imageto-image translation with conditional adversarial networks", 2016: 441–450.