ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

Research on Data Enhancement Method for Complex Pavement

Tingting Zhang[1], Shuyan Ren[1,*], Hailong Duan[1], Shuoyi Wen[1]

[1]School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin, China
[*]Corresponding author

Abstract: *In the era of intelligent vehicle development, environmental awareness technology as the basis of vehicle decision-making and route planning, affected by the weather, accurate road image recognition is particularly important. Aiming at this problem, this paper constructs a large scale pavement image dataset and proposes an improved image generation method based on CycleGAN data enhancement. The core innovation of the method is the improvement of the model structure, introduce Wassertein distance, and implement global Lipschitz constraint with spectral constraint. The experimental results demonstrate that the accuracy of the data set is up to 91.85%, which is 6.41% higher than the original CycleGAN, and the convergence speed is faster.*

Keywords: *Pavement Image, Dataset, CycleGAN, Image Generation*

1. Introduction

In general, the model should have enough data sets to cover as many cases and features as possible, so that the model can be more generalized, and the model can address the needs of people. But it is sometimes difficult to obtain the data set, so we use the method of data enhancement to shorten the time of data acquisition. Literature [1] presents an image data enhancement method combining sharpening and noise reduction, but whether this method is suitable for single channel image data and three-channel images remains to be investigated. Restricted Boltzmann Machines (RBM) and Deep Belief Network (DBN) are traditional generation models, but these methods are computationally complex and have limited generation effects. Generative Adversarial Net (GAN), as one of the more advanced technologies in the field of deep learning, has been receiving extensive attention since its introduction. The development of GAN is rapid, and a series of improved models have appeared on the basis of the original GAN. The Least Squares Generative Adversarial Networks (LSGAN) proposed in [2] mainly replaces the cross-entropy loss function with the least squares loss function, which further improves the quality of the generated picture, but it does not fundamentally solve the oscillation problem of GAN in the training process. Document [3] suggests that Wasserstein Generative Adversarial Network (WGAN) replaces the JS (Jensen-Shannon) distance with Wasserstein distance to better measure the distance between the two distributions and to some extent alleviate the problem of GAN training instability, but this algorithm does not allow the discriminator to really be confined within the Lipschitz function, nor does it rigorously give the Wasserstein distance calculation method. The large GAN (Large Scale GAN (Big GAN), proposed by Google's Deep Mind research team [4], uses the Res Net architecture, which generates realistic samples with a more natural contour, but the algorithm requires a lot of hardware and takes a long time to train.

This paper offers an improved data enhancement method for Cyc1eGAN. The model can generate more abundant images, greatly improve the recognition ability of the model, and effectively alleviate the problem of uneven sample distribution. Experimental results show that the quality of the proposed method is better than that of the original Cyc1eGAN and the customary data enhancement method.

2. Data Set

With the rapid development of intelligent vehicle, environmental perception has gradually become a significant research direction. Real-time recognition of road type can be invoked as a basis for real-time adjustment of intelligent vehicle control strategy, and also improve the safety and comfort of the vehicle. In the field of meaningful learning, the size and quality of datasets are essential for algorithmic models.

ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

2.1. Dataset Construction

The data set [5] is generally subdivided into data collection. Data processing and data annotation. Going down this pattern, this article builds a complete road data set flow as shown in Figure 1.

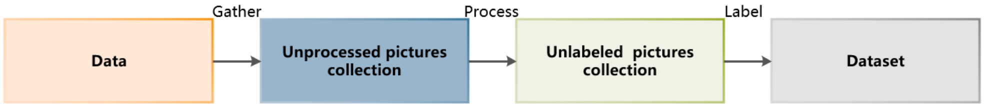

Figure 1: Pavement data set construction process

The main task of the data acquisition phase is to define the data source and obtain data from the data source.

The principal problems in the data processing stage are data filtering and data noise removal.

Data annotation stage is the highest stage of data set construction, is the manual use of annotation tools for data classification and information calibration.

This paper identifies 6 typical pavement images, as shown in Figure 2, including asphalt pavement, muddy pavement, gravel pavement, snow pavement, grassland pavement, block pavement.

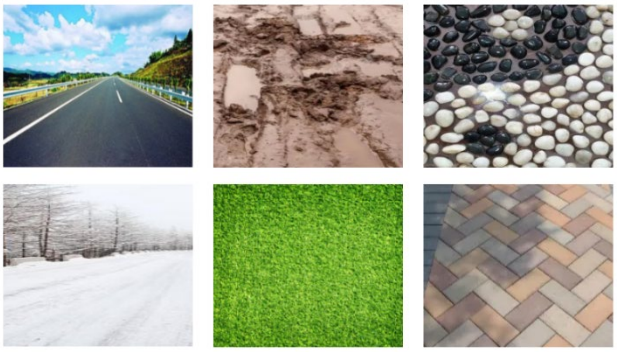

Figure 2: Partial data set picture

2.2. Data Acquisition and Feature Analysis

2.2.1. Collection Method

Shooting in kind: Select the actual scene suitable for the light through the camera, including pictures and videos.

Web access: Baidu, Google, Bing, Web Crawler.

Baidu is a Chinese search engine prepared in 2000. Compare to the other two search engines, Baidu satisfies common entertainment life related searches and is intended to provide people with "simple and reliable" information access. At the same time, many pages of Baidu through manual finishing, there is a certain degree of professionalism and good fault tolerance.

The Google search engine is the world's more commonly used search engine. It is a dataset search by name toolbox, but it is also updated more timely.

Bing is Microsoft's new search engine, opened on May 28, 2009. Bing image search has been one of the most frequently utilized vertical search products. In order to help users find the most apt exquisite pictures, Bing took the lead in implementing the Chinese input global search. Users are not required to search in English, but just input Chinese. Bing will automatically match the user's English, to assist users find suitable images from around the world.

Search engines use web crawlers to crawl web pages, documents and even images, audio, video and other resources, through the corresponding index technology to organize this information to search for users.

ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

2.2.2. Typical Characteristic

1) Color Features

The most intuitive feature in the image is the color feature of the signal, which is characterised by high resolution and easy identification between different colors. And the surface areas are often presented as clusters of the same or similar color blocks. Therefore, color feature is an important indicator of ground type. Among them, the color characteristics of snow, grass, mud and asphalt are observable in the data set. Typical road RGB values are shown in Table 1.

Table 1: Typical road RGB values

Category	R	G	B
Snow pavement	255	250	250
Grassland pavement	110	204	18
Muddy pavement	185	155	144
Asphalt pavement	75	84	98

2) Texture Features

Another feature corresponding to the color feature is texture feature, which can indicate the texture direction of the target unit and distinguish different objects by the difference of the delicate texture. Texture feature is a worldwide feature that describes the surface properties of an image or area of an image. It is just an effective method to use texture features in texture images with large differences in thickness and density. Most of the texture features of the gravel pavement are elliptical and most of the masonry pavement is rectangular, as shown in Figure 3.

Figure 3: Pavement texture map

2.3. Data Enhancement

Data enhancement technology has traditionally been an important means to overcome data shortage by synthesizing or converting to generate new data from limited data. Traditional image enhancement techniques are based on a series of known affine transformations such as rotation, scaling, displacement, and simple image processing methods such as illumination color transformation, contrast transformation, and the addition of noise. These changes are premised on not changing the label of the image and can only be limited to the image domain. This data enhancement method based on geometric transformation and image operation can alleviate the problem of neural network over fitting to some extent and improve the generalization ability. However, compared with the original data, the increase of data points does not solve the problem of insufficient data. At the same time, this kind of data enhancement method needs to set up the conversion function and corresponding parameters, usually based on empirical knowledge, the optimal data enhancement is often difficult to achieve, so the generalization performance of the model can only be limited.

Some of the contemporary generation models have attracted a lot of attention because of their excellent performance. Examples include Variable Auto-Encoding Network (VAE) and Generative Adversarial Network (GAN), whose methods of generating samples can be used for data enhancement. This method is more complex than traditional data enhancement techniques, but the resulting samples are more diverse and can be used in a variety of scenarios, including image editing and image denoising. This paper mainly introduces the data enhancement technique based on the improved generation and confrontation network and applies this method to the classification task of pavement dataset.

GAN is the foundation of CycleGAN. The improved CycleGAN gets the advantages of stable training, fast convergence and high quality image.

2.3.1. GAN

GAN consists of a generator G and a discriminator D[6], where the generator takes random noise as input, outputs the images generated, and constantly learns the distribution of real data; the discriminator takes the images generated by the generator and the real images as input, and the output samples belong to the probability of real images. The two alternate training, and finally achieve global optimization [7-8]. As shown in Figure 4, is the structure diagram of GAN.

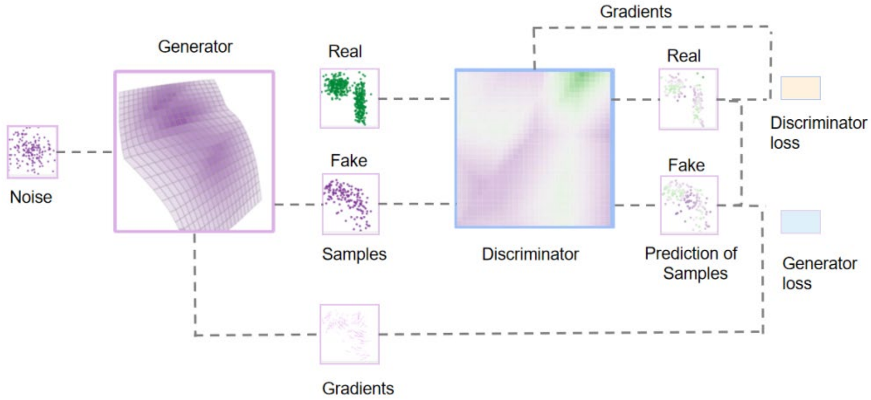

Figure 4: GAN structure diagrams

Its objective functions are:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim pdata(x)}[\log D(x)] + E_{z\sim pz(z)}[\log\left(1 - D\big(G(z)\big)\right)] \tag{1}$$

In equation(1). $P_{data(x)}$ is the data distribution of the real image. $P_{z(z)}$ is the data distribution of random noise; and E denotes the mathematical expectation.

2.3.2. CycleGAN

CycleGAN is essentially a ring network of two GANs, each sharing two generators and each having a discriminator that maps x to y. Suppose the generator G is mapped from X to Y, the discriminator can judge whether y is generating samples or real samples. The target function of generator G and discriminator is:

$$L_{GAN}(G, D_Y, X, Y) = E_{y\sim pdata(y)}[\log D_Y(y)] + E_{x\sim pdata(x)}\left[\log\left(1 - D_Y\big(G(x)\big)\right)\right] \tag{2}$$

Similarly, if generator F implements the mapping from Y to X, the discriminator determines whether X generates samples or real samples. The target function for generator F and discriminator is:

$$L_{GAN}(F, D_X, Y, X) = E_{x\sim pdata(x)}[\log D_X(x)] + E_{y\sim pdata(y)}\left[\log\left(1 - D_X\big(F(y)\big)\right)\right] \tag{3}$$

Cycle-Loss is also included in CycleGAN to ensure that the generated image retains the characteristics of the original image, which prevents multiple original images from being mapped to a generated image. Cycle-loss is:

$$L_{cyc}(G, F) = E_{x\sim pdata(x)}\left[\left|F\big(G(x)\big) - x\right|_1\right] + E_{y\sim pdata(y)}\left[\left|G\big(F(y)\big) - y\right|_1\right] \tag{4}$$

The total loss function is:

$$L(G, F, D_x, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \tag{5}$$

2.3.3. Improvements To The CycleGAN Loss Function

1) Introduction of Wasserstein distance instead of the original loss function

The main reason for the difficulty in GAN training is the gradient disappearance, that is, the loss function of the optimized generator is equal to the JS divergence between the optimized and the real data. The Wasserstein distance is introduced to replace the original loss function to solve the gradient disappearance problem. The improved objective functions are:

$$\min_{G}\max_{f,\|f\|L\leq 1} E_{x\sim p(x)}\,[f(x)] - E_{z\sim q(z)}\big[f\big(G(z)\big)\big] \tag{6}$$

In formula (6): $f(x)$ is a discriminator function. this function needs to satisfy the Lipschitz constraint, that is, the spectral constraint is used to implement the Lipschitz constraint. In this paper, Wasserstein distance is introduced into two sets of discriminators and two sets of discriminators. The objective

ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

functions of G and discriminator D_Y are shown in formula (7) and F and discriminator D_X are shown in formula (8).

$$L'_{GAN}(G, D_Y, X, Y) = \min_G \max_{f,\|f\|L\leq 1} E_{y\sim P(y)}\ [f(y)] - E_{x\sim q(x)}\big[f\big(G(x)\big)\big] \tag{7}$$

$$L'_{GAN}(F, D_X, Y, X) = \min_F \max_{g,\|g\|L\leq 1} E_{x\sim q(x)}\ [g(x)] - E_{y\sim p(y)}\big[g\big(F(y)\big)\big] \tag{8}$$

In formula (9): $f(y)$ is the discriminator D_X function; $g(x)$ is a discriminator D_Y to function in. The improved loss function is shown in formula (9).

$$L'(G, F, D_x, D_Y) = L'_{GAN}(G, D_Y, X, Y) + L'_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \tag{9}$$

2) Implementing Global Lipschitz Constraints with Spectral Constraints

The physical meaning of the matrix spectral norm [9] is that any vector that has been transformed into a matrix is less than or equal to the length of the product of the vector and the norm of the matrix spectral norm. Namely:

$$\frac{|f(x+\delta)-f(x)|_2}{|\delta|_2} = \frac{|W\delta|_2}{|\delta|_2} \leq \sigma(W) \tag{10}$$

In formula (10): $\sigma(W)$ represents the spectral norm of the weight matrix, x the input vector of the layer, and δ the variable of x. Lipschitz's constraints can be realized by controlling the spectral norm of the weight matrix, so the spectral norm is added to the loss function as a regular term, namely:

$$\min_\theta \frac{1}{K}\sum_{k=1}^{K} L(f_\theta(x_k), y_k) + \frac{\lambda}{2}\sum_{i=1}^{N} \sigma(W^i)^2 \tag{11}$$

In formula (11), the latter half is the canonical term of the spectral norm of the weight matrix, and the discriminator satisfies the Lipschitz constraint by punishing the sum of the spectral norm of each layer.

2.3.4. Improvements to the CycleGAN Network Structure

A U-Net skip structure is added to the network structure of the generator[10-11] to preserve pixel level image information at different resolutions. As shown in Figure 5, the U-Net network structure diagrams.

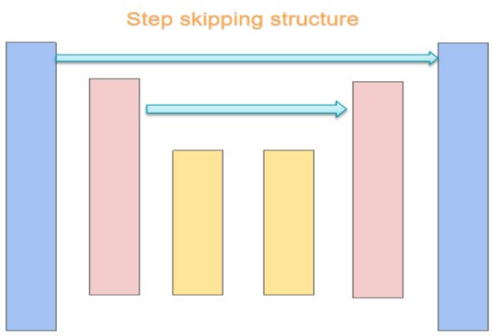

Figure 5: U-Net network structure diagrams

In order to enhance the feature representation ability of a convolution neural network, the traditional deep residual neural network introduces more parameters and more time overhead. The spatial pyramid [12] makes full use of the resolution details on the large picture and the macro-dimension features on the small picture in the dimension of feature extraction to achieve the effect of variable receptive field, that is, making full use of the information of the original input picture. Therefore, the spatial pyramid network is in addition to the generator for better performance.

The space pyramid structure is illustrated in figure 6. The behavior of global average cooling is similar to that of a structural regulator and prevents over fitting. The spatial pyramid structure includes three adaptive averaging pools of different sizes, and the structure normalization and structure information are to be incorporated into the attention path. Adaptive and average mapping of input features to 3 scales: 4x4, 2x2, and 1x1. Resize these three outputs into three one-dimensional vectors, and through the connection of them together to generate a one-dimensional feature vector. The spatial pyramid structure can preserve the feature representation and inherit the advantage of global average pooling. Space pyramid formula:

$$S(x_l) = C\left(R\left(P(x_l, 4)\right), R\left(P(x_l, 2)\right), R\left(P(x_l, 1)\right)\right) \tag{12}$$

Formula (12) :x_l: output of $l \in [1, L]$layer; $P(\cdot, \cdot)$: Adaptive Average pooling; $R(\cdot)$:Resize function; $C(\cdot)$:join operation.

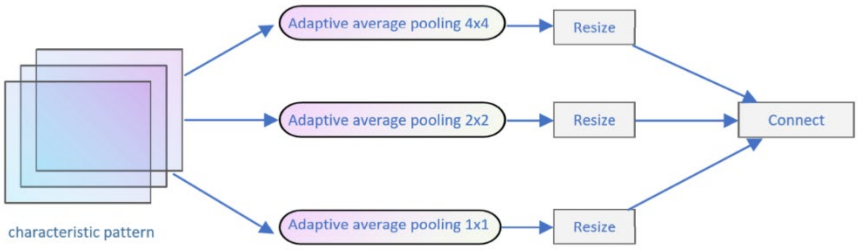

Figure 6: Schematic diagram of spatial pyramid structure

As shown in Figure 7, the original CycleGAN network residue block and the improved network residue block structure diagram are shown.

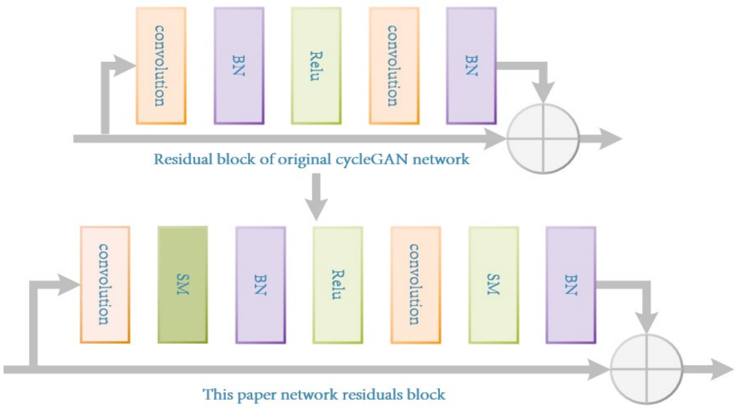

Figure 7: Residue block structure diagram with space pyramid

3. Experimental Validation And Result Analysis

3.1. Experimental setup

Using Win10 operating system, PyTorch using dynamic computing graph mechanism, easy to debug the process of this article, so this article experiment using PyTorch framework, and accelerated model training using GPU. Use the computer configuration as CPU ES 2670, RAM 128 GB, video card NVIDIA Tesla V100 16 GB.

In this paper, four scenarios are set up for pavement recognition and comparative experiments are conducted as shown in Table 2.

Table 2: Compares the experiments

Training Set	Operation	Category	Quantity	Total Quantity
Unsettled	/	6	250	1500
Traditional Data Enhancement	Flip, Rotate, Crop, etc.	6	1000	6000
CycleGAN	CycleGAN	6	1000	6000
This Article	G-CycleGAN	6	1000	6000

In order to ensure the validity of the experiment, the real road pictures were taken as a unified test set, including 110 pictures in each category, for a total of 660 pictures.

ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

The test network selects ResNet50; the model iteration is 200; the batch size is set to 8; and the learning rate is set to 0.0002.

3.2. Experimental analysis

Recent Inception distance (FID) distance was used to measure the quality of generated images [13]. FID distance measures the distribution distance between the real image and the generated image in the high-dimensional feature, and takes into account the similarity of the two types of images. FID determines the similarity of the two groups of images from the similarity of the computer vision features of the original image, which is computed using the Inception V3 image classification model. The consequences of FID calculation are given in Table 3. The FID value of the traditional method is the highest, and the FID value of the depth network is smaller than that of the traditional method.

Table 3: Assessment of the quality of different model generation

Model	FID Distance Value
Tradition	39.256
CycleGAN	25.227
G-CycleGAN	12.648

In order to verify the effectiveness of the proposed method, the ResNet50 model is directly trained with raw data, traditional data enhancement. CycleGan, G-cycleGan generated image data set, and the accuracy of the test set is evaluated from Figure 8.

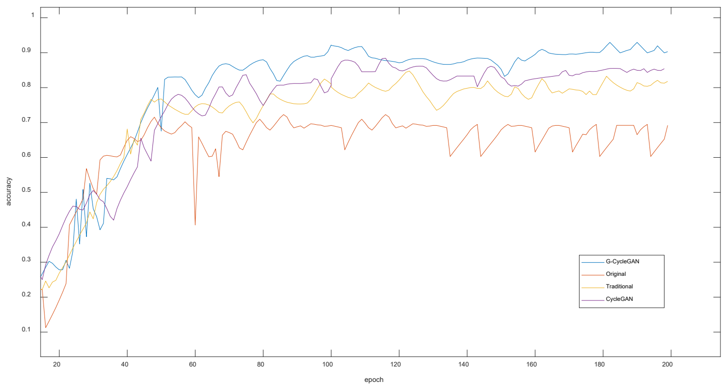

Figure 8: Test model training accuracy

As shown in Figure 8, the abscissa is equal to the number of iterations and the ordinate is the accuracy of the test set. Because of the small number of authentic data set images, it cannot be well generalized in the test. The accuracy is the lowest, the result is 69.02%. The accuracy of the traditional dataset is improved from 69.02% to 80.72%. Therefore, the dataset enhancement plays a major role in network training, but the accuracy is still low. The accuracy of road recognition is improved from 80.72% to 85.44%, and the accuracy is further improved to 91.85%, which is 6.41% higher than that of the original CycleGAN. The experimental results demonstrate that the G-CycleGAN model can generate data sets for road surface recognition and achieve the effect of data enhancement.

4. Conclusion

Road surface recognition system needs a lot of images, but in reality it is difficult to obtain images and the samples are unevenly distributed. Being aimed at the problems of small amount of data and poor picture quality, this paper improves CycleGAN model and constructs a set of pavement datasets which can meet the actual need. Four pavement datasets are tested on ResNet50 by FID. The experimental results demonstrate that the improved model is more accurate than the traditional data enhancement method, which verifies the validity of the dataset.

References

[1] Russo Fabrizio. An image enhancement technique is combining sharpening and noise reduction [J].

ISSN 2522-3488 Vol. 7, Issue 3: 28-35, DOI: 10.25236/IJNDES.2023.070305

IEEE Transactions on Instrumentation and Measurement, 2001, 51(4):824-828.
[2] Mao Xudong, Li Qing, Xie Haoran, et al. Least squares generative adversarial networks[C]. Proceedings of the IEEE international conference on computer vision, 2017:2794-2802.
[3] Arjovsky Martin, Chintala Soumith.Bottom Léon. Wasserstein generative adversarial networks[C]. International conference on machine learning, 2017:214-223.
[4] Brock Andrew, Donahue Jeff, Simonyan Karen. Large scale GAN training for high fidelity natural image synthesis[C]. International Conference on Learning representations, 2019.
[5] Collins B, Deng J, Li K, et al. Towards scalable dataset construction: an active learning approach [C]/European conference on computer vision. Springer, Berlin, Heidelberg, 2008: 86-98.
[6] Lin Jenn L., Dai Xingyuan, Li L., et al. A new frontier in artificial intelligence research: generative adversarial networks[J]. Journal of Automation, 2018, 44 (5): 775 -792
[7] Wang Kunfeng, Gou Chao, Duan Yanjie, et al. Research progress and prospect of generative adversarial network GAN [J]. Journal of Automation, 2017, 43 (3): 321 a 332
[8] Wang Gongming, Qiao Junfei, Wang Lei a generative adversarial network in the sense of energy function [J]. Journal of Automation, 2018, 44 (5): 793 I 803
[9] Yan Bei and Zhang Jianlin. Study on image data generation based on spectral constraint for generative adversarial networks [J]. Semiconductor Photonics, 201940 (6): 896 -901
[10] Tian L, Zheng Y, Cui Q. Research on Data Enhanced Ancient Pictogram Recognition Method Based on Convolutional Neural Network[C]// the 2019 3rd High Performance Computing and Cluster Technologies Conference. 2019.
[11] Peng Peng. Image style conversion based on CycleGAN [D]. Chengdu:University of Electronic Science and Technology, 2019.
[12] Guo Jingda, Ma Xu, Sansom Andrew, et al. Spanet: Spatial pyramid attention network for enhanced image recognition[C]. IEEE International Conference on Multimedia and Expo, 2020.
[13] Cai Zhiling, Weng Qian, Ye Shaozhen, et al. Scene classification of high-resolution remote sensing images based on Inception V3 model [J]. Remote Sensing of Land Resources, 2020, 32 (3): 80-89