

Research Article

Deep Learning Image Feature Recognition Algorithm for Judgment on the Rationality of Landscape Planning and Design

Bin Hu 

Xinyang Vocational and Technical College, Xinyang 464000, Henan, China

Correspondence should be addressed to Bin Hu; hubin@xyvtc.edu.cn

Received 17 March 2021; Revised 8 April 2021; Accepted 17 April 2021; Published 27 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Bin Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper uses an improved deep learning algorithm to judge the rationality of the design of landscape image feature recognition. The preprocessing of the image is proposed to enhance the data. The deficiencies in landscape feature extraction are further addressed based on the new model. Then, the two-stage training method of the model is used to solve the problems of long training time and convergence difficulties in deep learning. Innovative methods for zoning and segmentation training of landscape pattern features are proposed, which makes model training faster and generates more creative landscape patterns. Because of the impact of too many types of landscape elements in landscape images, traditional convolutional neural networks can no longer effectively solve this problem. On this basis, a fully convolutional neural network model is designed to perform semantic segmentation of landscape elements in landscape images. Through the method of deconvolution, the pixel-level semantic segmentation is realized. Compared with the 65% accuracy rate of the convolutional neural network, the fully convolutional neural network has an accuracy rate of 90.3% for the recognition of landscape elements. The method is effective, accurate, and intelligent for the classification of landscape element design, which better improves the accuracy of classification, greatly reduces the cost of landscape element design classification, and ensures that the technical method is feasible. This paper classifies landscape behavior based on this model for full convolutional neural network landscape images and demonstrates the effectiveness of using the model. In terms of landscape image processing, the image evaluation provides a certain basis.

1. Introduction

With the advancement of science and technology and the substantial increase in computer processing capabilities, artificial intelligence has achieved exponential development. Technologies related to artificial intelligence have widely affected the Internet, medical treatment, manufacturing, and other fields, greatly reducing labor costs and production cycles, and also facilitating people's work and life [1]. Artificial intelligence optimizes the combination of the four elements of landscape, namely, structure, system, service, and management, so that ordinary landscapes can be developed into more humane and intelligent landscapes, creating a convenient and comfortable living environment. However, the application of artificial intelligence in the landscape field is more

biased toward hardware intelligent solutions, and there is still a lack of more innovative applications such as the design and creation of landscape patterns [2]. This is because the design and creation of landscape patterns require creative thinking and unique image thinking in artistic creation. Therefore, people's understanding and reproduction of a large number of landscape features are the keys to landscape pattern design and creation [3]. In contemporary landscape dynamic characteristics and landscape planning and design, although the thoughts and design concepts of different landscape architects are different, what they cannot be separated from is the objective and stable law of landscape form, which is the law of common features in the landscape pattern. Landscape features to explore and discover the law has been a landscape hot research topic area [3]. With the

improvement of computer performance and the development of remote sensing technology, the feature extraction of landscape objects is no longer limited to human eye detection. The use of computer vision, remote sensing camera measurement, artificial intelligence, and other technologies to study landscape features has gradually become a trend [4]. There are many hidden features of landscape patterns, and traditional landscape feature extraction methods are not suitable for the extraction and reconstruction of common features of the landscape under big data. Therefore, the current mainstream methods for extracting common features of the landscape are still focused on human visual recognition and detection, and human subjectivity is relatively strong [5]. For example, according to the characteristics, the literature summarizes the landscape as scattered, lattice combination, staggered substitution, ring combination, parallel strip combination, and axial symmetry. The literature points out that the main way of landscape group layout is counterpoint, dislocation, periphery, center concentration, and free combination [6]. It can be seen that in the classification of landscape patterns summarized by characteristics, people have strong subjectivity, and the classification of irregular layouts is more general. Only obvious pattern characteristics can be seen and the hidden landscape patterns and small-scale rules under big data are ignored. The layout consists of a more macro-irregular layout.

The self-attention generative confrontation network introduces the self-attention module in natural language into the generative confrontation network. By improving the drawbacks of convolutional networks, such as limited perceptual fields and difficulty in extracting features from long distances, it allows one to observe the drivers of image generation tasks. Long correlation modeling can use cues from all feature locations to generate details, which solves the problem of feature memory [7]. The existing technology of image generation has achieved a substantial breakthrough. The introduction of the self-attention mechanism enables the network to truly capture and correlate long-distance spatial information while ensuring the efficiency of calculation. The use of spectral normalization in the discriminator and generator network not only reduces the computational cost of training but also improves the quality of the generated results. Feature integration theory divides visual attention into two stages: pre-attention and concentrated attention. In the pre-attention stage, the feature extraction of multiple parallel vision channels is realized. In the subsequent concentrated attention stage, the visual features are selectively reorganized and searched for each visual attention target in turn [8]. On this basis, a guided search theory is proposed, which believes that the visual system searches the region of interest through two parallel pathways, namely, selective pathways and nonselective pathways. Among them, the nonselective path is mainly used to extract the global features of the scene, and can quickly count the scene structure information related to the spatial layout of the scene and the semantic information

related to the task. The nonselective path is mainly used to extract local features in the scene, and can obtain detailed information such as color, brightness, texture, etc., and then use the global information as a priori to guide the salient area search process of the selective path from top to bottom. The gist descriptor represents the global information of the scene image and guides the classification and search of visual objects. According to the pixel scale applied by the algorithm, such as a pixel or super pixel, the algorithm can be divided into the following categories [9]. The model first performs preprocessing operations including image averaging and Gaussian blur on the original color image and then calculates the color difference corresponding to the two images in the Lab color space to obtain a significant area map. This method is to process the image from frequency and get an approximate salient target area. Since this model only uses color differences and does not consider other saliency-related visual features such as orientation and texture, theoretically, this method does not apply to scenic images with small color differences [10]. The algorithm first extracts three visual saliency features, namely, scale contrast features, color features, and central-peripheral contrast features, and then uses CRF to effectively fuse the extracted visual saliency features, and finally produces a significant target detection result. Region-based methods are generally based on super pixels, and regression models for saliency detection are based on super pixels [11].

Landscape images are a new research area in the field of image retrieval. With this technology, more and more landscape architects are beginning to research and study in this area and gradually expanding their research [12]. Using landscape images as the medium, this article can use the classification of landscape elements as the criteria for landscape evaluation, to realize the full automation of landscape evaluation, which not only saves time but also saves costs. This is also a qualitative breakthrough in computer technology for landscape architecture. We use the advantages of interdisciplinarity to complete this research [13]. First, perform semantic segmentation on the landscape image with known scene element category information, make it into a training set, use the built training set to train the model to identify the pictures to be tested, and use the fully convolutional neural network algorithm to effectively improve the classification accuracy of landscape images [14]. At the same time, the color, shape, and texture features of the scene elements in the image are extracted as the multi-features for identifying the scene elements in the landscape image. The category information of the sample image is introduced into the fully convolutional neural network algorithm, which can realize the effective clustering of the landscape image [15, 16]. Thereby, it effectively improves the classification rate of landscape elements in landscape images. The convolutional neural network algorithm in the deep learning method is analyzed in detail, the CNN algorithm is improved in some aspects, the dimensionality reduction method is applied to image recognition, and related experiments are completed.

2. Deep Learning Image Feature Recognition Design

2.1. Optimized Deep Learning Algorithm. Since the landscape pattern data are a remote sensing satellite image, the convolutional neural network is the most commonly used deep neural network structure for image feature recognition. Therefore, the discriminator that generates the confrontation network also uses the convolutional neural network structure to extract the features of the landscape. This section will introduce in detail the principle of deep convolutional generation network and the process of extracting landscape features. The difference between the convolutional neural network and the original neural network is that the neurons in each layer of the convolutional neural network are arranged in three dimensions: width, height, and depth. Assuming that the size of the landscape pattern data is 643, the convolution kernel will only connect to a small area in the previous layer of the picture, instead of adopting a full connection method. As shown in Figure 1, each convolution kernel will perform a point multiplication operation on the data in an area of the training image data instead of all the data. This region, also known as the receptive field, greatly reduces the computational effort and effectively avoids overfitting due to too many parameters. The last part of the convolutional neural network structure will compress the full-scale landscape data into a vector containing the classification score.

The convolutional layer is the layer used by the convolutional neural network to extract the features of landscape pattern picture data, and it is the most important layer. We undertake the task of transforming from image to high-level semantics [17]. Convolution layer can view and extract a rectangular area of the size of the characteristic value; extracting a feature of each rectangle becomes convolution kernel convolution operation and is performed by the sliding window that is similar to the step size to obtain a map sheet and perform calculations. The calculation formula of convolution is

$$G_j = \frac{\sum_{i,j=1}^{m,n} R_i * U_{ij}}{\sum_{i,j=1}^{m,n} E_i * U_{ij}}. \quad (1)$$

There are usually multiple convolution kernels, and each convolution kernel has a receptive field corresponding to the number of channels in the landscape pattern picture. These receptive fields, respectively, perform dot multiplication operations on one-dimensional arrays of a certain length on different channels in the area of the same convolution kernel size on the same landscape pattern data, and then add the obtained values to obtain the output value. There are several convolution kernels to get several output matrices [18]. The output data will eventually become the extracted landscape feature value. The pooling layer can perform down-sampling operations on the input landscape pattern data. The purpose is to remove excess information, improve training effects, increase receptive fields, increase translation invariance, and reduce the optimization difficulty and parameters, and reduce the computational cost [19]. The method is similar to

the window sliding of the convolution kernel, and there is also an area range to obtain the value of the same size area of the input matrix. Unlike the previous example, we are using the maximum set level as an example [20]. The pooling layer obtains the maximum value in the area. In addition to maximum pooling, pooling also means pooling and random pooling. Mean pooling is the summing and averaging of the values in the acquisition area, and random pooling is the random selection of the values in the acquisition area. In practical applications, maximum pooling is used the most.

Each node of the fully connected layer is connected to all the nodes in the previous layer and is used to integrate the features extracted from the front [21]. The fully connected layer is the classifier in the entire convolutional neural network. If operations, such as the convolution player, pooling layer, and activation function layer, map the original data to the hidden layer feature space, the fully connected layer maps the learned distributed feature representation of the sample labeled space. The calculation process formula of the fully connected layer is

$$f^n(t) = \lim_{h \rightarrow 0} \frac{1}{h^n} \sum_{m=0}^n (-1)^m \binom{n}{m} f(t - mh). \quad (2)$$

It can be seen that the essence of the fully connected layer is a linear transformation from one feature space to another feature space. Convolution and pooling are equivalent to the feature engineering of the landscape. The subsequent full connection is equivalent to the dimensional transformation after feature weighting. In particular, data features extracted by convolution can be changed from high-dimensional to low-dimensional, and the landscape of each area. Put the extracted features back together while retaining useful information. Its features are just a bunch of codes. To make it easier to understand, we have visualized the result as a picture.

Standardization is generally to map the extracted landscape feature data to a specified range, so that the output maintains a relatively stable distribution, and is used to remove the dimensions and dimensional units of the landscape data of different dimensions [22, 23]. The use of standardization in convolutional neural networks can speed up the convergence, have the effect of regularization, improve the generalization ability of the model, and allow a higher learning rate to accelerate the convergence. Commonly used standardization includes regional standardization, back standardization, and spectrum normalization. The back standardization and spectrum normalization are commonly used in deep convolutional neural networks. Among them, spectral normalization is often used in the discriminator in the generation of confrontation networks:

$$F = \frac{Ml \tan^2 2\alpha}{4 \sin \alpha} \theta(0)^2. \quad (3)$$

The principle of batch standardization is to use regularization to reduce the deviation of the internal correlation variable distribution, thereby improving the robustness of the algorithm. Back normalization consists of two parts. The

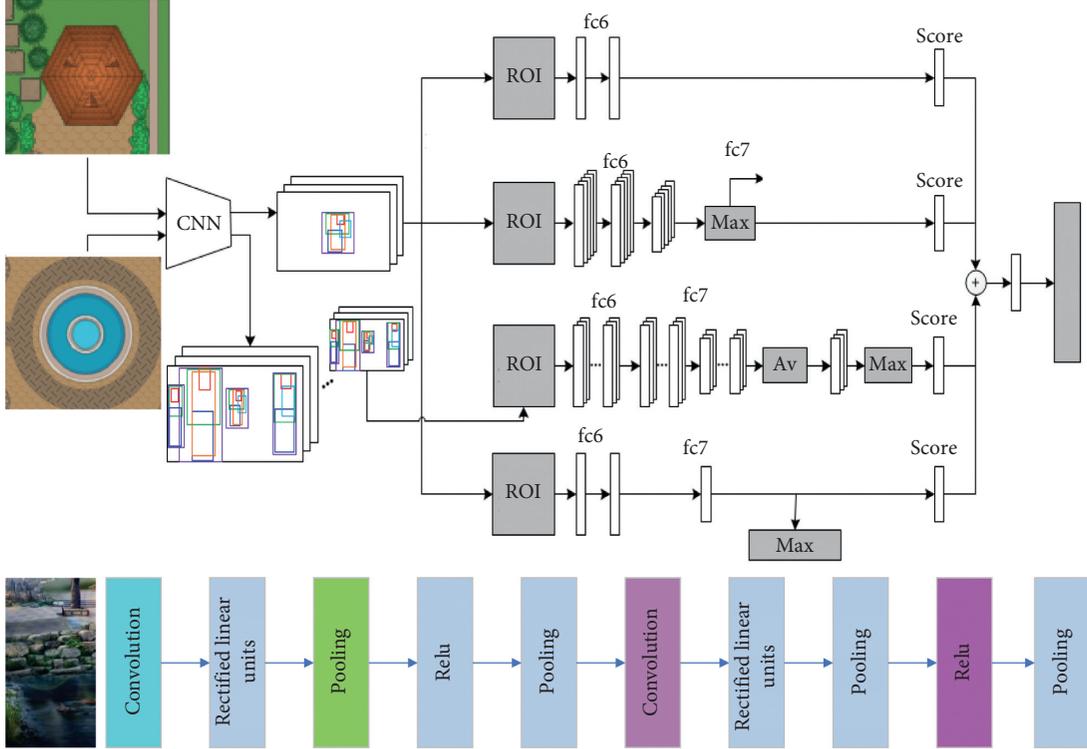


FIGURE 1: Improved convolutional neural network model.

first part is zooming and translation and the second part is training the parameters of zooming scale and translation. The algorithm steps are as follows:

$$\begin{aligned}
 b_{ij} &= \frac{b_{ij}}{\sum_{k=1}^n b_{kj}(i, j = 1, 2, \dots, n)}, \\
 \omega_i &= \sum_{j=1}^n b_{ij}(i, j = 1, 2, \dots, n), \\
 g_i &= \frac{\omega_i}{\sum_{j=1}^n \omega_j}, \\
 \sum_{i=1}^n g_i &= 1, \\
 \lambda &= \frac{1}{n} \sum_{i=1}^n \frac{(AW)_i}{\omega_i}.
 \end{aligned} \tag{4}$$

It can be seen from the above algorithm that batch standardization can normalize the data to form a normal distribution so that the data will not be too scattered and the activation function can better obtain more meaningful values. The objective function of generating the confrontation network model is as follows:

$$\ddot{\varphi} = (1 - \cos \varphi) \ddot{\theta} + \sin \varphi \dot{\theta} \dot{\varphi} - \frac{g}{l} \sin(\varphi - \theta) - \frac{k}{ml^2} \varphi. \tag{5}$$

The task of the discriminator is to learn the characteristics of the real sample set of the landscape to distinguish the

authenticity of the generator sample. Its structure is a deep convolutional neural network structure, which classifies the real landscape data as 1 and classifies the generated landscape data as 0. The task of the generator is to fool the discriminator as much as possible. When the generator generates realistic landscape pattern data, the discriminator will not be able to distinguish its authenticity [23]. If the discriminator judges its result to be approximately 1, it means that the generator's landscape pattern has the main features that are the same as the real samples, which successfully deceived the discriminator.

The nonconvergence phenomenon in the landscape generation training is caused by the disappearance and collapse of the gradient so that the model cannot obtain an effective gradient. This phenomenon of gradient disappearance has always been the main problem faced by various fields in deep learning; when the deep network and loss function in the deep neural network are not suitable, the gradient will disappear and collapse, and it belongs to the generation of the deep learning generation field. This problem also exists in the confrontation network. The unbalanced training of the generator and the discriminator will cause the gradient to disappear or collapse. The data are from real sample data X , and the data distribution function is P . The data are from real sample data X , the data distribution function is P . The real data are used to generate pseudo-random probability sampling noise under normal Z . The generated probability distribution function is P , the generated network is G , and the network is determined as D . The training process of generating adversarial neural networks can be regarded as an optimization task.

$$P(y_i) = p_i^{y_i} (1 - p_i)^{1 - y_i}. \quad (6)$$

The goal of generating a confrontation network is to hope that the trained discriminator can make the first term to be 1. Generally speaking, the discriminator can make the discriminator from the real distribution. The data are judged as 1 and the data from the pseudo data distribution are judged as 0. Overall, the training goal is to make a V maximum. For the generator, the training goal of the network is to minimize V , which means that the discriminator cannot determine the difference between the real data and the fake generated data. Therefore, the optimization of GAN requires the generator and the discriminator to achieve Nash equilibrium, but because the discriminator D and the generator g are trained separately, the Nash equilibrium may not be achieved. Through the analysis, it can be known that when the parameters of the generator g are fixed, it only trains the discriminator. When a true or false random sample x is input, the formula is

$$GJ_{ik} = \frac{GJ_{ik}^{jint}}{\sum_{i=1}^d GJ_{ik}^{jint}}. \quad (7)$$

The goal of training the discriminator is to obtain the discriminator in the optimal state; so, to obtain the extreme point of the formula, let the derivative of D be 0, and enter the formula:

$$k_{m'}(x) = \sum_{i=1}^N w_i s_i(x) = W^T S(x). \quad (8)$$

The optimized discriminator after the calculation is

$$GJ_{ik}^{jint} = \sum_{i=1}^N GJ_{ik}^{jint}, \quad (9)$$

$$\partial_1 \leq \int_{t_0}^{t_0+T_0} |S^T(\tau)c| d\tau \leq \partial_2, \quad \forall t_0 \geq 0.$$

It can be seen from the analysis that the primary problem when training a generative adversarial network is to maintain the stability of the training gradient. The generator and the discriminator are related to each other: if one does not perform well, the other does not perform well. If the discriminator is much better than the generator in the training, it will take a long time for the generator to make a little progress. If the discriminator is adjusted too hard to distinguish right from wrong, then the generator has no gradient to learn. Therefore, the training of the discriminator and generator cannot be too good or too bad and must be maintained within a range allowed by the balance to make the training converge. However, this balance is difficult to master, and the pooling layer and activation function in the deep network increases the uncertainty of the network parameters, so there may be an imbalance in the training of the generator and the discriminator in each training.

2.2. Landscape Image Feature Recognition Design. Features learned and extracted by deep learning are all mapped in high-dimensional manifolds. Manifolds can be

seen as the promotion of three-dimensional surface space and the geometric concept of high-dimensional space. The idea of manifold learning for feature learning of deep neural network data can be explained like that in the properly trained deep neural network [24]. Each category of image data is mapped on a different dimensional space, such as 0–9 handwritten image data. Each type of number should be mapped on a different manifold space. When a new handwritten number is to be classified as input, the data to be classified can be mapped on the manifold space surface of its correct number classification through neural network calculations [25]. The theory is also an attention theory involving automatic processing. They distinguish between Object and Feature, considering a feature as a specific value of some dimensional quantity, and an object as a combination of some features. They argue that features are analyzed by a functionally independent subsystem of perception and that this processing takes place in a parallel manner, whereas the recognition of objects requires the involvement of focused attention and is entirely the result of serial processing; the role of focused attention is similar to that of “glue,” which allows for the combination of features into a single object.

The significance of the geographic database is not only to provide a convenient data management model but also to realize the true sense of paperless and digital storage. The change of the thematic map storage medium has made it the basis for real-time data sharing and regional operations among multiple users. As a type of big data, landscape big data not only has the five characteristics of big data, namely, volume, velocity, variety, veracity, and timeliness, but also has the characteristics of space, time, diversity, and complexity. Summarizing the characteristics of landscape big data can be introduced from six aspects [26].

The objectivity of landscape big data means that the data are natural or objective data, which are not affected by human intervention and the basic data for landscape planning and design. The objectivity of the big data of the landscape can be reflected in the spatial data, which is the basis for respecting the principles of the site in planning and design [27]. Objective data can accurately reflect the actual situation of the site and are important considerations for the feasibility of planning implementation and the cost of planning. For example, the mold type of three-dimensional space can reflect the slope and gradient of the site, which is convenient for analyzing the direction of the catchment of the site, deciding the location of landscape points, the setting of scenic spots, etc., and carrying out drainage design to avoid a lot of soil cutting and filling works. The vegetation and soil data in the natural resource data are important parameters for site design and construction, which determine the level and quality of construction.

The diversity of landscape big data is reflected in two aspects. One is the diversity of data sources. As shown in Figure 2, more than 30 types of data come from drawings, tables, survey records, official websites, local chronicles, network data, etc. The same type of data can be obtained in various ways. The second is the diversity of data types;

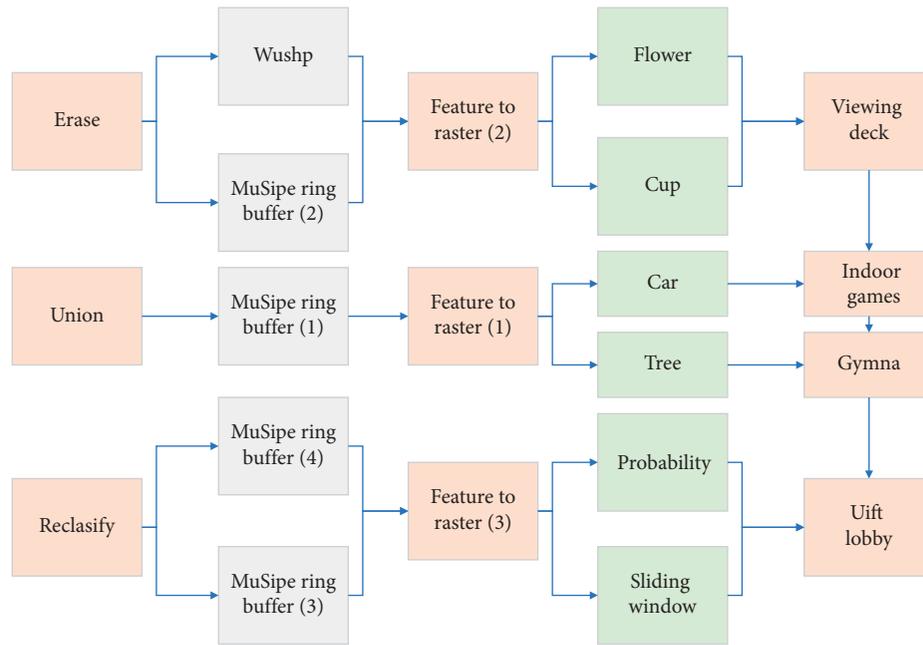


FIGURE 2: Landscape image features recognition design.

the data storage of landscape big data is diverse, which can be traditional data forms such as drawings, tables, photos, and document records. It can also be perception, mood, behavior, experience, video, etc. With the development of big data technology, more and more content is converted into data, which promotes the integration and innovation of disciplines. The multisource nature of landscape big data reflects the multidisciplinary nature of landscape planning and design, as well as its wide-ranging characteristics.

The dynamic nature of landscape big data is a common feature of big data. With the rapid development of the Internet, data have exploded, and dynamic changes of data have become the norm. It changed with time and crowd activities. By locating the communication data, we can analyze the direction of the flow of people in different periods and the location and time of the point of interest, which is helpful to analyze people's behavior and activities and the space usage of the venue. Online media data transform people's interests and hobbies into data. Through dynamic monitoring of online media data, it is possible to understand the need of the crowd, as well as people's evaluation of the social service functions of the venue before and after the scenic reconstruction, which is conducive to public participation in design. Corresponding to the dynamic nature of landscape big data is the current nature of the data. With the rapid update of the data, the interception of the data can only replace the changes in the performance stage, reflecting the latest natural and human conditions of the site, the flow of people, and the evaluation information, etc. Such as remote sensing image data, generally showing the latest site images, using socio-economic data, through the latest population distribution and community data, can effectively allocate landscape resources, improve resource

utilization, and avoid waste of resources. The current trend of big data also reminds us that big data is a resource and an analysis tool. What big data provides is reasoning based on the current situation, which is only a reference answer, not the final answer.

With the improvement of equipment performance, observable and collected data become more and more refined. The time and space resolution of data have been improved. The high time resolution can reflect the subtle dynamic changes of things and phenomena, and the high spatial resolution can see the details of things clearly, which is of great significance for identifying and analyzing the natural and human elements of the site. For example, the data of the heat map is changed from the original one hour to automatic statistics every fifteen minutes, which greatly shortens the time interval and improves the accuracy of analyzing the trend. It increases the resolution of remote sensing images from 30 meters to 50 centimeters now, allowing planners and designers to see more detailed features. Such an increase in resolution facilitates the excavation of cultural and natural features in the site. Traditional on-site inspections may encounter impassable problems. With the improvement of accuracy, researchers can fully understand the current situation of the site without leaving home, thereby improving work efficiency.

Landscape planning and design need to pay attention to the role of human beings and give full play to the human subjective initiative. Human nature is reflected in the big data, including human nature and social attributes, through which effective and targeted planning can be carried out. Mobile positioning data can reflect the activities of the crowd in a specific period and a specific place, and reveal people's temporal and spatial behavior patterns. Many data of landscape big data are related to people, such as network media data, communication data, activity behavior data, etc.

These data can reflect the human nature of landscape big data and provide support for people-oriented planning and design concepts.

3. Reasonable Landscape Design

3.1. The Rationality of Landscape Design Process Design. In this paper, the specific technical workflow is divided into six steps, such as data acquisition and preprocessing, data cleaning, data analysis, data mining, and result visualization. Among them, data cleaning and data analysis include two stages, that is, a rough classification and type division of technical framework. Data mining and visualization include the fuzzy-type identification and type definition phases of the technical framework.

First, in the data acquisition step, the basic description of the street view image data, including the specific method and source of acquisition, and the relevant parameters in the acquisition process are explained, and the research scope of this research is clarified, that is, the downtown area of Nanjing. Subsequently, in the process of data cleaning and analysis, InfoGAN's unsupervised classification method is used to roughly classify the style data. The InfoGAN generative confrontation network model will first generalize the image features of the picture and try to classify it. After classifying the overall characteristics of the street view data, MobileNets is used to perform semantic segmentation of the street view data to extract the landscape elements from the street view, complete the basic data cleaning, and enter the preliminary style classification stage. In the preliminary classification stage, the segmented data are interpolated to form a dataset in three directions, and then they are again entered into the InfoGAN model for classification, and the preliminary results of classification under different directions are obtained. It is necessary to enter the stage of supervised learning. Therefore, the results of different orientations are manually labeled, the labels are labeled according to the theoretical model described above, and the training set and test set of the supervised classification model ResNet are constructed, and the classification results according to the label type are generated [28]. However, the classification results of label types are purely visual localization and need to be tested in reverse at both the location distribution and functional fit levels. Finally, the classification and type definition of the entire streetscape landscape is completed.

The previous technical framework clarifies several target stages of data processing. On this basis, this article further improves the framework in the form of a flowchart, clarifying relevant elements such as human annotations, external parameters, etc., and then guides subsequent experimental operations, as shown in Figure 3.

As a kind of city image database, street view image data refer to image data that is close to the human eye, and the content of the photo is image data of urban street scenes. To a certain extent, it can more truly reflect the urban street scenes seen from a human perspective. Compared with ordinary photos, it contains more and more complete urban elements and has a richness of information. Street view

image data are based on the image characteristics of the street view. To extract the landscape from the streetscape, the data must be cleaned first. The entire cleaning process is divided into two processes. The first is to roughly classify the entire collected street view data. It separates the landscape-based street view from other street view data to reduce the overall data sample size. The cleaning method is to first use the information generation confrontation network for rough classification. The network model is written in the Python programming language, mainly using the PyTorch library and the Numpy library-related functions. We set different discrete dimensions, that is, the different number of categories, to train multiple sets of models on the street view data set.

Through the comparison of multiple groups of different classification numbers, the phenomenon of the unclear meaning of dimensions caused by too few classification numbers can be avoided. There are too many categories, which lead to the classification of details. From the model output results, the model settings of 25 types of discrete dimensions have relatively better interpretability. Subsequently, the model of network classifier trained to 7 million images one by one score data, and take the highest-scoring dimension to the final classification, and finally sort out the results, with cxv recording different types of picture data files. In the entire actual operation, the original data are used as the training set and the test set, and the effect of clustering the original data set is finally realized [29]. Each classification number has multiple iterations. Generally speaking, the higher the iteration number, the better the effect. However, the number of iterations is limited by the number of classifications. From the data of the last three iterations, it can be seen that the overall data are more evenly distributed in different types of data.

To perform separate operations on the landscape element pixels and nonlandscape element pixels in the picture, the next step is to make an image mask. Image making is a process of excluding selected images, graphics, or objects, and images to be processed to control the area or process of image processing. In digital image processing, the mask is usually a two-dimensional matrix array, and sometimes multi-value images are also used. The main purpose of the image mask is to extract the part of the interest in the picture. The premade mask of the area of interest is multiplied by the image to be processed to obtain the image of the area of interest so that the image value of the area of interest remains unchanged. The values of the images outside the area are all 0, and the purpose of screening is achieved. The position of the image that people are interested in is entered into the subsequent algorithm model to reduce the interference of the noninterest area. Therefore, the processing of non-landscape elements in street view data will affect the machine learning model's extraction of image features to a certain extent. The next step of the experiment will be to conduct contextual processing on these nonlandscape elements.

The semantic segmentation of the landscape data goes to the next model analysis. Because the size and shape of different landscape elements will lead to inconsistent dimensions between the data, it will greatly affect the model results. Therefore,

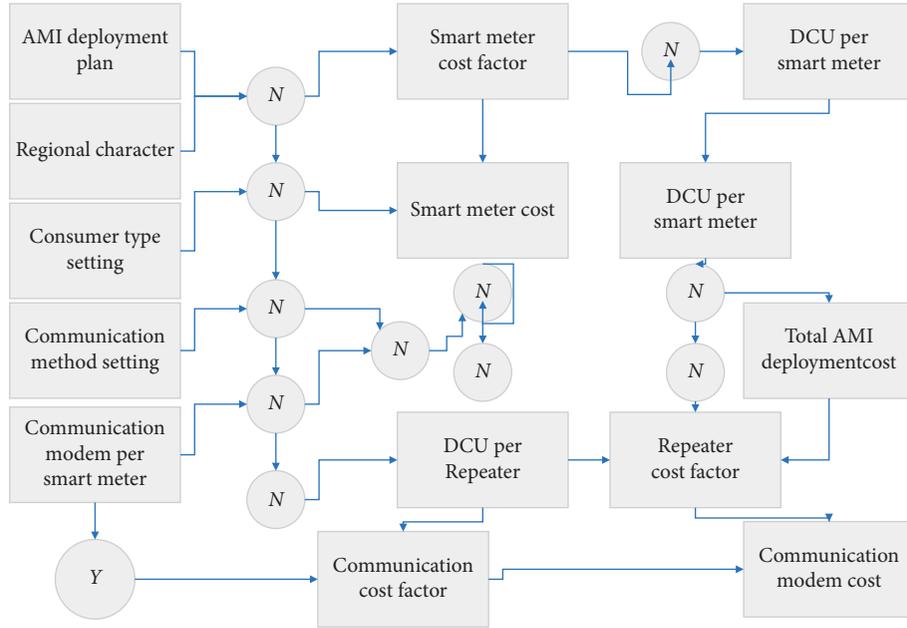


FIGURE 3: Flow chart of landscape data operation technology.

how to deal with elementary pixels other than landscape elements is very important. This article uses three processing methods at this stage: the first is a mask that is not processed, that is, all pixels in the landscape image that are not part of the landscape are filled with pure black. This type of operation will remain the perspective elements of the landscape elements, and due to the gradient changes of colors, the edge characteristics of the landscape elements will be strengthened. This operation will preserve the perspective elements of the landscape elements, and the edge features of the landscape elements will be enhanced due to the different variations of colors. Therefore, the new InfoGAN model analysis based on these data will strengthen the classification caused by retaining the perspective of the landscape. This method firstly performs an average operation on the pixel values of all the landscape parts in the street view image after semantic segmentation. Because most of the images are masked, the proportion of suoyi landscape elements is relatively small. This article wants to expand the proportion of the landscape and preserve the neighborhood relationship between the original pixels as much as possible, so I thought of RF transformation to regularize the pixel data. Rastafarian extracts each pixel of the landscape element in the landscape image data, and calculates the number of pixels N , as shown in Figure 4.

In the form of the cluster-based structure, the InfoGAN model further subdivides the form. According to the shape ratio of the landscape window, some dimensions are classified. The proportions of viewing windows or landscape elements are grouped in longitudinal longer landscape styles through 20 dimensions, and the common features of such styles in the landscape, a small number of vertical opening windows residential landscape. In this paper, this type of cluster structure is defined as the longitudinal line emphasized cluster style. The functional type tends to be mainly industrial and residential landscapes. This article defines this type as horizontal. The lines emphasize the group style.

3.2. Model Parameter Design. The structure of the fully convolutional neural network used in this article is shown in Table 1. The convolutional neural network mainly includes a convolutional layer and a maximum pooling layer. We use rectified linear units with sparse activation as the activation function. Conv-5 uses a stacked convolution method. After each convolution layer, one or two identical convolution layers are stacked. In order to prevent overfitting and improve the robustness of the model, a dropout layer is added after the Conv6 and Conv7 layers, which is not involved in network propagation, so that the output of some neurons is 0. This improves the generalization ability of the model. The detailed parameters of each layer are shown in Table 1.

Because deep learning fully convolutional neural network is very complicated, it will take a lot of time to train, and the network level of the fully convolutional network is very deep, so the final result of the model is possible while conducting specific experiments. It will stay near the optimal solution, which may affect the experimental results. On the other hand, it will make the convergence time of the entire model longer.

Therefore, in the training of the actual full convolutional neural network model, it is common to use the parameters of a better convergence model to initialize the initial parameters of the new model. On the deep executive network and autoloader network, the method has been widely used recently. We use relatively few training samples to train the parameters of each layer in the network, and then use these parameters to initialize the trained model, and finally train the formal model [30]. If the method of random initialization is used, it has carried out experimental verification in a paper published in 2010, which will cause the model to easily fall into the local minimum and fail to get the global minimum. However, the pretraining method can make the model of this article convenient to



FIGURE 4: Sample image of landscape.

TABLE 1: Model parameter design.

Name	Type	Area	Step	Number	Size
1, 1, 1, 1	Convolution	64	1	[0101]	(54, 54, 54)
1, 2, 1, 3	Convolution	128	2	[0000]	(14, 14, 14)
2, 1, 2, 1	Convolution	256	3	[1010]	(54, 67, 65)
2, 2, 2, 2	Pooling	126	4	[1001]	(43, 54, 32)
3, 1, 3, 1	Convolution	215	3	[0110]	(55, 55, 55)
3, 2, 3, 2	Max	1024	2	[1111]	(67, 43, 23)
3, 3, 3, 3	Pooling	128	4	[1010]	(56, 21, 12)

achieve better performance here. Existing convolutional neural network models all adopt a pretraining strategy. Based on the above model, this article also refers to the idea of pretraining and adopts the method of model two-stage training [31–34].

In this paper, the model is initialized and trained so that the parameters of the convolutional layer can be shared, and also to improve the effect of using random initialization alone [35, 36]. The research in this paper adopts the following optimizations for the training of the model. This article manually selects some 500 simpler images uses these 500 images to train the model separately, and waits for the model to converge. In the future, the parameters in the model can be saved and downloaded. Because the selected images in this article are simple, the convergence speed of the model will be very fast. We perform a second training of the model on all training sets in this paper. All the network weights and parameters are updated, which also reduces training time to a great extent on the model, and the model performance will be enhanced. It can also speed up the convergence speed of the model, and finally, a good model can be obtained. After these models are trained, a lot of Caffe model files will be generated. The model represents the documents FCN network parameters of the network model stored in different iterations. Finally, these models can be used in a landscape image. The two-stage training process of the model is shown in Figure 5.

4. Results and Analysis

4.1. Model Training Effect Analysis. The model operating mechanism of this experiment is shown in Figure 6, and will not be described in detail here. After repeated experiments, this article determines that the learning rate of the weight parameter is 1010, and the weight attenuation coefficient is 0.005. According to the two-stage training method of the model given in Figure 5, the two-stage training method of the model is no longer used for comparison with the experiment. In the process of the model two-stage training method, the previous section pointed out that 500 images of the training set were used as the pretraining data set in the early stage. After that, model two losses and iteration number class l were used, and the object type stage was used. The relationship between the training method and the first-order training method using the model is shown in Figure 6.

The value of the loss function represented by the loss is to weigh the probability of which type of data the test data belong to. The lower the loss functions, the faster the network convergence. It can be seen that the loss values of both the two-stage training with model and the two-stage training without model decrease rapidly in the number of iteration steps from 300 to 500. The loss of secondary training reached 0.43, while only 0.47 was used without secondary training. The convergence speed of the second-stage training with the model is significantly faster than the second-stage training without the model, indicating that the effect of the second-stage training with the model is better. As the iteration proceeds, the last two are in a relatively stable state. The loss value after stage training is stable at about 0.41, while the unused loss value is stable at about 0.45, which shows that the convergence speed of model training is becoming more and more gentle. The experimental results show that the two-stage training of the model can effectively speed up the convergence speed of the landscape image semantic segmentation model training process, and the data prove that the method is feasible.

According to the calculations of the three scenic element classification and evaluation formulas, the accuracy of the water scene element classification in the landscape image is shown in Figure 7, which can directly reflect the performance of the three up-sampling structures.

The accuracy rate of landscape element classification reflects the probability that the specific categories in the landscape image are correctly distinguished and segmented. As shown in Figure 7, the accuracy of FCN-32s is the lowest in the classification of water feature elements. FCN-16s is lower than FCN-8s in the classification accuracy of water features. Therefore, it can be concluded that FCN-8s has the best overall performance in the classification of water features of landscape images. Finally, the results were verified using the scene element classification and evaluation indicators mentioned, and finally, two values were obtained, namely, the pixel accuracy rate, the average accuracy rate.

To show that the contour-based mid-level features have a good effect in calculating visual saliency, this paper adds a comparative experiment, which is to apply the CS model and the IT model to the standard contour and the rough contour

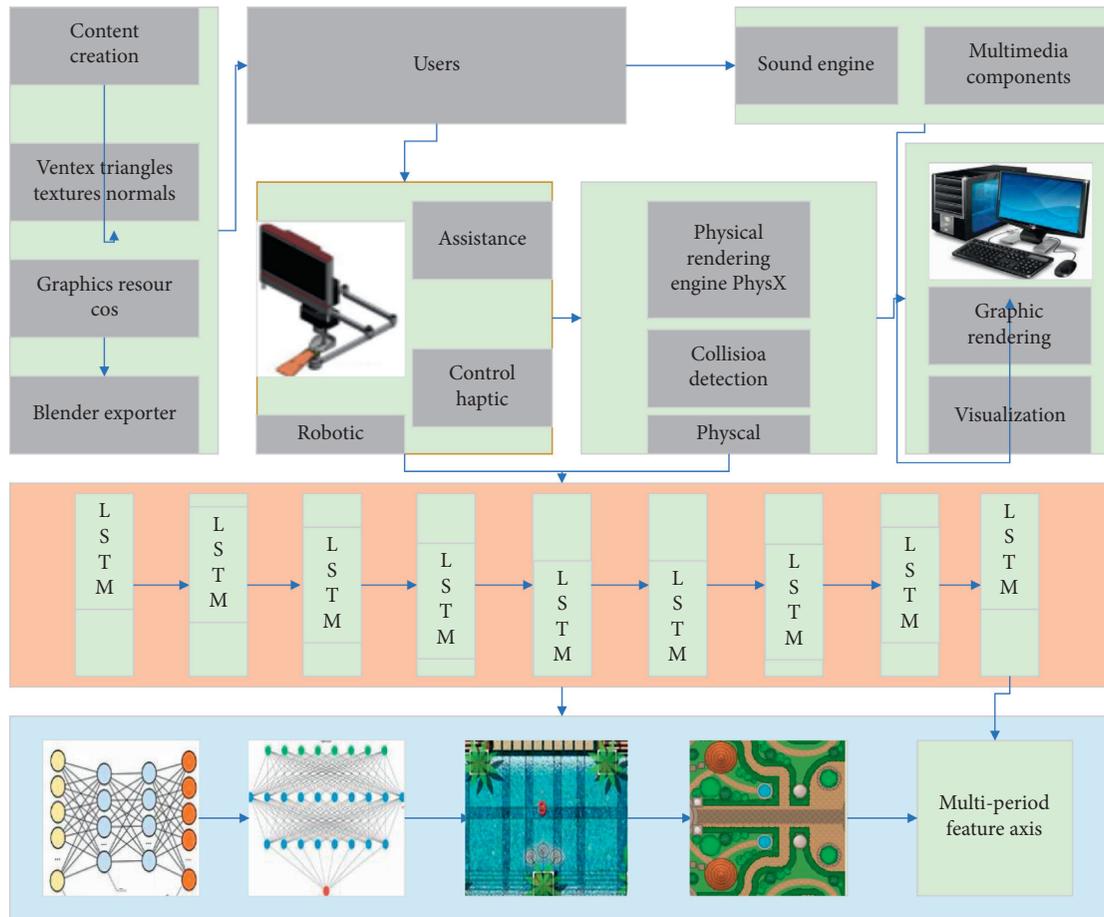


FIGURE 5: Two-stage training process of the model.

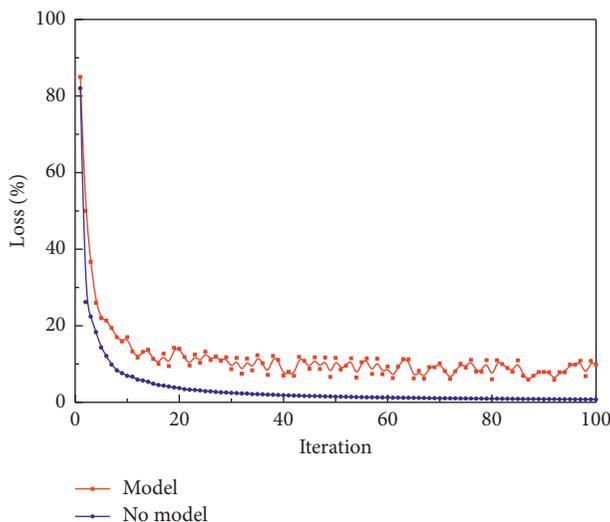


FIGURE 6: The model's two-stage training results.

in the BSDS500 data set, and quantify them through related indicators. It analyzed the performance of the two models on the two contours. The ROC curve drawn in the experiment is shown in Figure 8. It can be seen from the graph that the

performance of the contour-based CS model proposed in this paper is significantly better than that of the IT model, whether it is on the standard contour or the rough contour. The model's expectations are very consistent.

The reason why the performance of the IT model is inferior to the CS model is that the visual saliency features used by the IT model are mainly based on the local bottom layer, and the CS model in this article is based on the contour features of the middle layer, which reflects the middle layer features of the scene from the side. It does play a very important role in visual saliency. It shows the significance detection results of some samples of the BSDS500 data set in this experiment.

As shown in Figure 9, we can see that the performance of the CS model and its model proposed in this paper on the standard profile is significantly better than the performance of the two models on the rough profile. This is because there are some models in the rough profile. As mentioned above, the fragmented edge line segment noise will affect the results of the contour-based saliency detection method to a certain extent. Besides, when only contour information is used, the performance of the CS model is significantly better than that of the classic IT model when the scene image lacks visual information such as color and brightness. This shows that the middle-level visual cues of the scene image are better at

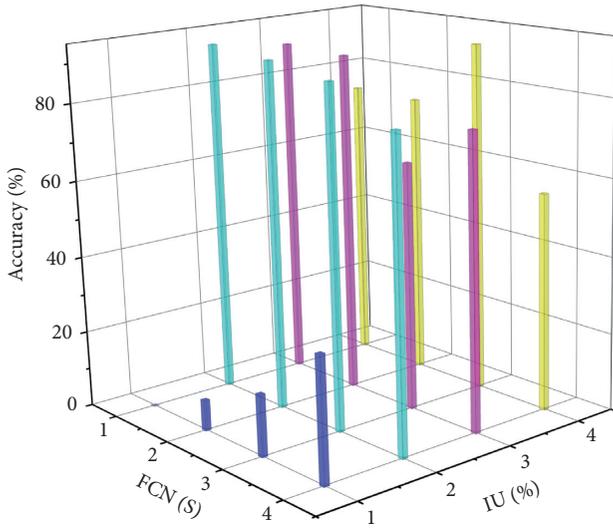


FIGURE 7: The results of three up-sampling semantic segmentation on landscape images.

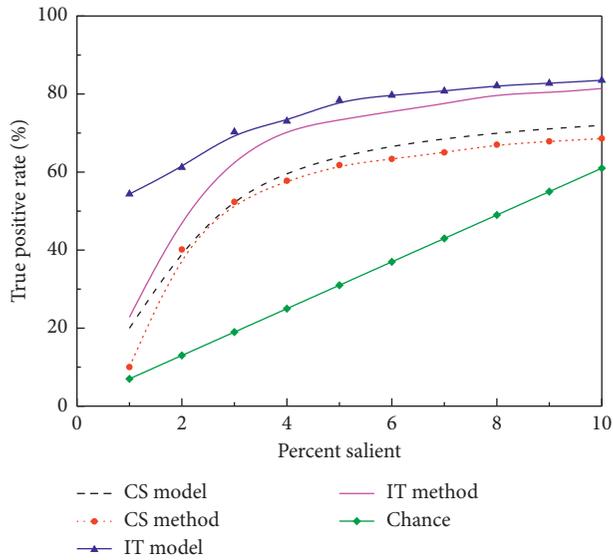


FIGURE 8: The ROC curves of CS and IT models on standard contour and rough contour, respectively.

guiding visual attention. The lower level visual cues have a stronger effect. To make the above analysis and related conclusions more convincing, the performance indicators of the related models are calculated in the experiment and presented in the form of a table.

4.2. Analysis of Evaluation Results of the Recognition Model.

In this paper, the weight information sharing mechanism is added based on SAGAN, and the training of the poor discriminator and the better generator is selected by the alternating training of two sets of adversarial pairs. Due to the relationship between hardware facilities and computational cost, the experiment chooses to train 25,000 times, and the loss value of the traditional SAGAN structure discriminator

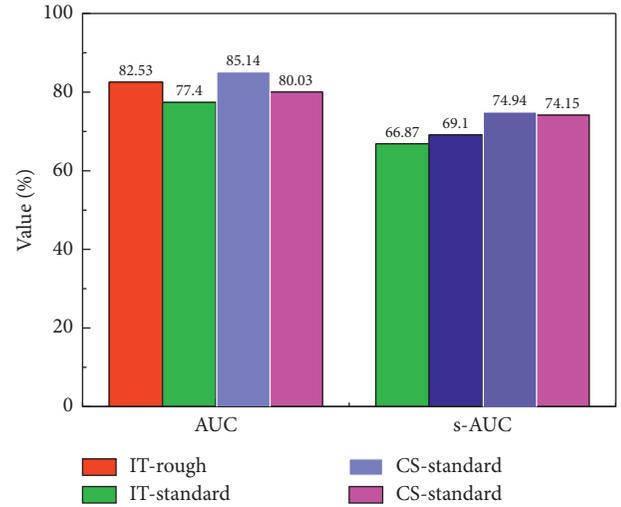


FIGURE 9: The AUC and s-AUC values of the CS model on the BSDS500 data set.

generator is recorded every 100 times. The loss value of the discriminator and generator of the improved model proposed in this paper is compared. The dotted curve is the traditional self-attention confrontation neural network, and the blue line curve is the multiple confrontational information sharing generation network proposed in this paper. Figure 10 shows the discriminator loss curve. It can be seen from the first half that the loss function of the discriminator is close to 0, indicating that the better the discriminator is trained, the more likely it is to have gradient collapse. The generator loss function is close to 0, indicating that the better the generator training, the better the training effect. The closer the confrontation balance is 0, the more stable the training of the model.

As shown in Figure 10, compared with traditional SAGAN, the discriminator of the new model proposed in this paper can maintain a higher level of loss of the discriminator. After SAGAN is trained for 10,000 times, the loss of the discriminator gradually approaches 0. Therefore, the gradient disappears, and the generator cannot obtain effective gradient information, so the generation result is getting worse and worse. In the adversarial information-sharing network proposed in this paper, when the discriminator is trained well, the weight parameters will be adjusted in time to obtain poor weight parameters, and the discriminant ability of the discriminator will be pulled back to a more appropriate level, avoiding the occurrence of gradient descent.

As shown in Figure 11, compared with the traditional SAGAN, the generator of the new model proposed in this paper can maintain a lower loss value. After SAGAN is trained for 10,000 times, the discriminator has disappeared because of the gradient. The detector creates a bias in the direction of gradient descent, causing the loss value of the generator to rise, producing any result that can be recognized by the discriminator, and this causes the discriminator to lose its ability to compete. The result generated by the generator quickly developed a good result with a low loss value due to the inability to generate the correct result, and

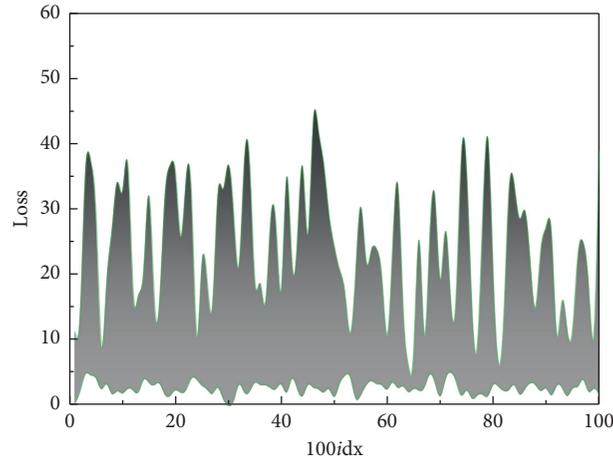


FIGURE 10: Discriminator loss curve.

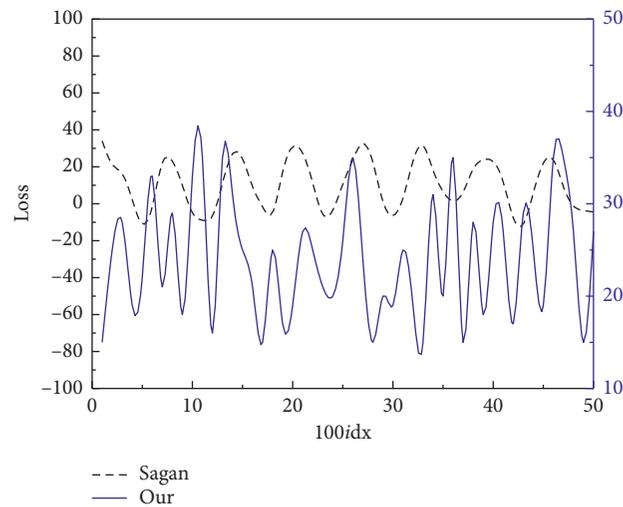


FIGURE 11: Generator loss curve.

finally, the training failed. It can be seen from the new model curve proposed in this paper that the level of the discriminator has been at a low and reasonable level, and a realistic fake image is finally generated.

As shown in Figure 12, it reflects the training balance of the discriminator and generator. It can be seen that because the discriminator loss of traditional SAGAN is constantly approaching 0 during the training process, and the generator loss is getting larger and larger, its balance is restricted below 0. Equilibrium curve of the new model proposed is 0 near the point of hovering, indicating that the generator and the discriminator maintained good training practice status, against the balance.

In summary, the training stability of the method proposed in this article is much higher than that of traditional SAGAN during the training process. In the later stage of training, SAGAN has the problem of gradient disappearance, and the model does not appear to have problems such as gradient disappearance at the end of the iteration. The superiority of the model is significantly higher than that of the generative adversarial network that does not use the

method in this paper. In the later stage of training, due to the relatively small number of data sets, the discriminator of the traditional generative adversarial network is easy to be trained well, which leads to the problem of gradient disappearance, which means the generator can only generate random noise after $300 \times 50idx$. Using the generative adversarial network incorporating the information sharing mechanism of multiple adversarial weights proposed in this paper, it is superior to train better quality generative data with only two adversarial network pairs.

As shown in Figure 13, when the discriminator gradient is close to 0, the discriminator training has a risk of overfitting, the training gradient fed back to the generator disappears, and then the information monitoring mechanism obtains the ill-conditioned feedback information and selects the appropriate one from the shared information. The information re-updates the discriminator and generator, strengthens the generator while restraining the discrimination to a certain extent, and the generator regains the gradient. From the formula, it can be seen that when the loss value of the discriminator and generator is maintained

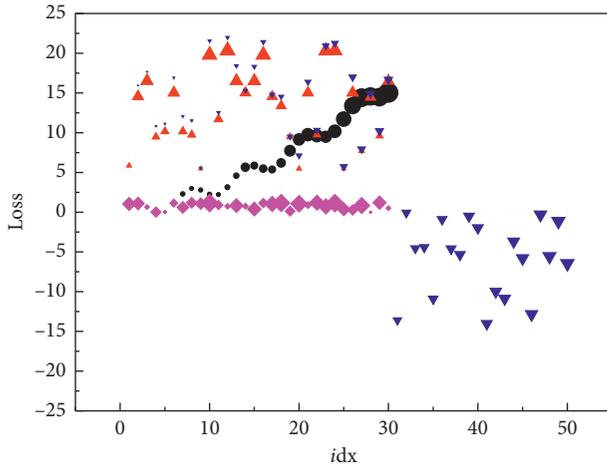


FIGURE 12: Training balance performance graph.

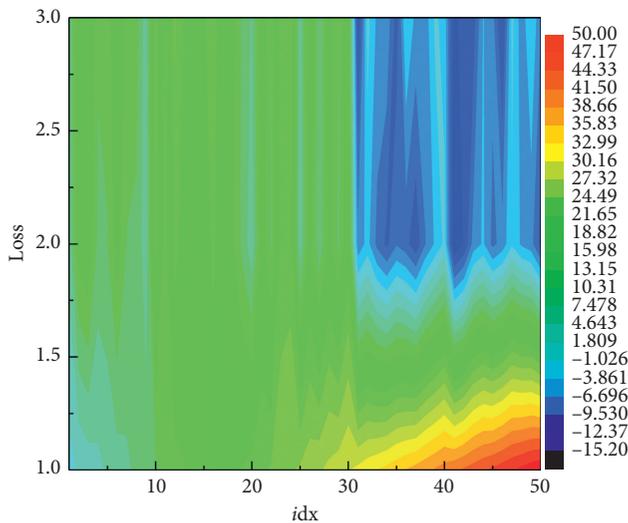


FIGURE 13: The gradient disappears and pulls back.

within 0.5, the ideal Nash balance will be achieved. Therefore, it can be seen that although the loss curve of the discriminator is suppressed above 0 points, it gradually moves closer to 0.5. Therefore, the training of the model proposed in this article is reasonable and stable.

Spectral normalization serves to limit the Lipschitz of the network function by limiting the spectral norm of the weight matrix, thus allowing fewer discriminator updates to be performed per generator update. It analyzes the reasons for the disappearance of gradients and unstable training in the landscape generation of the generation confrontation network and introduces the core ideas of the WGAN model to solve these problems. It introduces the innovatively designed multiple confrontational information-sharing generation network for landscape pattern generation and proposes new ideas, new network structure, and specific methods to solve the problems of gradient disappearance and training instability. Finally, to prove the advantages and versatility of the model proposed in this paper, landscape pattern data, and two representative classical data sets, are used for

training generation and result verification. The experimental results show that the multiple confrontational information-sharing generation network proposed in this paper has greater advantages and good versatility than the traditional generation confrontation network, and the method proposed in this paper is superior in the generation of landscape patterns. The performance evaluation of this new model is further carried out by analyzing the experimental process.

5. Conclusion

This paper uses deep learning algorithms to segment the landscape pattern and proposes training the local landscape pattern first and then stitching the results into a whole. The landscape in the local pattern data uses feature segmentation to reduce the repetitive features, accelerates the model training speed, and avoids unnecessary repetitive features. By improving the multiple confrontation information-sharing generation network proposed for training, a clear landscape pattern map is obtained more effectively, and it has obvious advantages compared with the traditional self-attention confrontation network. Splicing into a small landscape group landscape pattern figure has a stronger artistic creativity compared with the real landscape pattern. Then, the rationality of layout is judged, and its generated result is better than the traditional self-attention-generating confrontation network, which proves that it can be used not only for landscape data sets but also for other data sets. However, due to the idea of segmentation training, the longitudinal splicing marks caused by re-splicing are inevitable. Finally, this paper uses this innovative generative confrontation network model to conduct a preliminary study on the macro characteristics of landscape patterns. It is concluded that the use of big data and machine learning methods to comprehensively summarize the macro-features of the landscape is more scientific than human eye analysis, and can extract less-known hidden features.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Y. H. Wang, A. N. Phung, I. Mohaimenul, Y. C. Li, and H. C. Yang, "Development of deep learning algorithm for detection of colorectal cancer in EHR data," in *Proceedings of the world Congress on Medical and health informatics*, vol. 264, pp. 438–441, Lyon, France, August 2019.
- [2] M. Bergler, M. Benz, D. Rauber et al., "Automatic detection of tumor buds in pan-cytokeratin stained colorectal cancer sections by a hybrid image analysis approach," *Digital Pathology*, vol. 11435, pp. 83–90, 2019.

- [3] C. C. Y. Poon, J. Yuqi, Z. Ruikai et al., "AI-doscopist: a real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices," *Npj Digital Medicine*, vol. 3, no. 1, pp. 1–8, 2020.
- [4] K. Tetsushi, T. Tamaki, S. Yoshida et al., "Application of real-time image recognition system with machine and transfer learnings to computer-aided diagnosis for endoscopic images of colorectal cancer," *IEICE Technical Report*, vol. 117, no. 278, p. 19, 2017.
- [5] B. Korbar, A. M. Olofson, A. P. Mirafior et al., "Deep learning for classification of colorectal polyps on whole-slide images," *Journal of Pathology Informatics*, vol. 8, no. 1, p. 30, 2017.
- [6] R. Zhang, Y. Zheng, C. C. Y. Poon, D. Shen, and J. Y. W. Lau, "Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker," *Pattern Recognition*, vol. 83, pp. 209–219, 2018.
- [7] A. Kumar, A. Deep, R. K. Gupta, V. Atam, and S. Mohindra, "Brain microstructural correlates of cognitive dysfunction in clinically and biochemically normal hepatitis C virus infection," *Journal of Clinical and Experimental Hepatology*, vol. 7, no. 3, pp. 198–204, 2017.
- [8] D. H. Katz, R. C. Deo, F. G. Aguilar et al., "Phenomapping for the identification of hypertensive patients with the myocardial substrate for heart failure with preserved ejection fraction," *Journal of Cardiovascular Translational Research*, vol. 10, no. 3, pp. 275–284, 2017.
- [9] N. Boy, C. Mühlhausen, E. M. Maier et al., "Proposed recommendations for diagnosing and managing individuals with glutaric aciduria type I: second revision," *Journal of Inherited Metabolic Disease*, vol. 40, no. 1, pp. 75–101, 2017.
- [10] S. Montagnese, F. P. Russo, P. Amodio et al., "Hepatic encephalopathy 2018: a clinical practice guideline by the Italian association for the study of the liver (AISF)," *Digestive and Liver Disease*, vol. 51, no. 2, pp. 190–205, 2019.
- [11] L. M. Castello, M. Baldrighi, A. Panizza, E. Bartoli, and G. C. Avanzi, "Efficacy and safety of two different tolvaptan doses in the treatment of hyponatremia in the Emergency Department," *Internal and Emergency Medicine*, vol. 12, no. 7, pp. 993–1001, 2017.
- [12] C. Eggers, K. Hanh, G. Arendt et al., "HIV-1-associated neurocognitive disorder: epidemiology, pathogenesis, diagnosis, and treatment," *Journal of Neurology*, vol. 264, no. 8, pp. 1715–1727, 2017.
- [13] J. Hatvani, A. Basarab, J.-Y. Tournier, M. Gyongy, and D. Kouame, "A tensor factorization method for 3-D super resolution with application to dental CT," *IEEE Transactions on Medical Imaging*, vol. 38, no. 6, pp. 1524–1531, 2019.
- [14] J. Tao and G. Sun, "Application of deep learning based multi-fidelity surrogate model to robust aerodynamic design optimization," *Aerospace Science and Technology*, vol. 92, pp. 722–737, 2019.
- [15] Q. Qin, K. Wang, J. Yang et al., "Algorithms for immunochromatographic assay: review and impact on future application," *The Analyst*, vol. 144, no. 19, pp. 5659–5676, 2019.
- [16] Z. Wu, S. Jiang, X. Zhou et al., "Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications," *Computer Communications*, vol. 150, pp. 729–738, 2020.
- [17] Y. Han and J. C. Ye, "Framing U-net via deep convolutional framelets: application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.
- [18] D. Guest, K. Cranmer, and D. Whiteson, "Deep learning and its application to LHC physics," *Annual Review of Nuclear and Particle Science*, vol. 68, no. 1, pp. 161–181, 2018.
- [19] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018.
- [20] A. Shah, L. Zhou, M. D. Abrámoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images," *Biomedical Optics Express*, vol. 9, no. 9, pp. 4509–4526, 2018.
- [21] H. R. Roth, S. Chen, O. Hirohisa et al., "Deep learning and its application to medical image segmentation," *Medical Imaging Technology*, vol. 36, no. 2, pp. 63–71, 2018.
- [22] W. Wei, E. A. Huerta, B. C. Whitmore et al., "Deep transfer learning for star cluster classification: I. application to the PHANGS-HST survey," *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 3, pp. 3178–3193, 2020.
- [23] H. Palangi, L. Deng, Y. Shen et al., "Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 694–707, 2016.
- [24] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: an overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [25] Y. Arijji, M. Fukuda, Y. Kise et al., "Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 127, no. 5, pp. 458–463, 2019.
- [26] Y. Chen, K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7048–7066, 2019.
- [27] P. R. Jeyaraj and E. R. Samuel Nadar, "computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of Cancer Research and Clinical Oncology*, vol. 145, no. 4, pp. 829–837, 2019.
- [28] P. Haber, B. Bishkek, A. Dengel, and B. Damian, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [29] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, 2019.
- [30] M. E. Hirschtritt, M. H. Bloch, and C. A. Mathews, "Obsessive-compulsive disorder: advances in diagnosis and treatment," *Jama*, vol. 317, no. 13, pp. 1358–1367, 2017.
- [31] I. Solos and Y. Liang, "A historical evaluation of Chinese tongue diagnosis in the treatment of septicemic plague in the pre-antibiotic era, and as a new direction for revolutionary clinical research applications," *Journal of Integrative Medicine*, vol. 16, no. 3, pp. 141–146, 2018.
- [32] H.-J. Ma and L.-X. Xu, "Decentralized adaptive fault-tolerant control for a class of strong interconnected nonlinear systems via graph theory," *IEEE Transactions on Automatic Control*, vol. 99, 2021.
- [33] X. Zenggang, T. Zhiwen, C. Xiaowen et al., "Research on image retrieval algorithm based on combination of color and shape features," *Journal of Signal Processing Systems*, vol. 93, pp. 139–146, 2019.

- [34] H. Lu, Y. Guan, L. He et al., "Patch aggregation trends of the global climate landscape under future global warming scenario," *International Journal of Climatology*, vol. 40, no. 5, pp. 2674–2685, 2020.
- [35] K. Sim, J. Yang, W. Lu et al., "MaD-DLS: mean and deviation of deep and local similarity for image quality assessment," *IEEE Transactions on Multimedia*, vol. 99, 2020.
- [36] J. Yang, Y. Zhao, J. Liu et al., "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Transactions on Cybernetics*, vol. 99, 2020.