

# Landscape image recognition and analysis based on deep learning algorithm

Nong Limei<sup>a,\*</sup>, Wu Dongfan<sup>b</sup> and Zhang Bo<sup>a</sup>

<sup>a</sup>*School of Architecture & Art, North China University of Technology, Beijing, China*

<sup>b</sup>*School of Architecture, Tsinghua University, Beijing, China*

**Abstract.** Garden landscape is the combination of nature and humanity, with high aesthetic value, ecological value and cultural value, has become an important part of people's life. Modern people have a higher pursuit for the spiritual food such as garden landscape after the material life is satisfied, which brings new challenges to the construction of urban garden landscape. As an advanced type of machine learning, deep learning applied to landscape image recognition can solve the problem of low quality and low efficiency of manual recognition. Based on this, this paper proposes a garden landscape image recognition algorithm based on SSD (Single Shot Multibox Detector), which realizes accurate extraction and recognition of image features by positioning the target, and can effectively improve the quality and efficiency of landscape image recognition. In order to test the feasibility of the algorithm proposed in this paper, experimental analysis was carried out in the CVPR 2023 landscape data set. The experimental results show that the algorithm has a high recognition accuracy for landscape images, and has excellent performance compared with traditional image recognition algorithms.

Keywords: Deep learning, garden landscape, image recognition, target detection; image analysis

## 1. Introduction

Deep learning is a kind of neural network technology with powerful data fitting ability and automatic feature learning, which is the premise and foundation of artificial intelligence. Due to the strong automatic learning ability and feature extraction ability of deep learning, the application of deep learning to image recognition has become a hot research field in recent years. Among all the image recognition tasks, landscape recognition has become the most complex one due to its small inter-class differences, which has been studied extensively by the academic circles at home and abroad.

From the current stage, experts and scholars at home and abroad mainly apply machine learning methods for landscape image recognition research. Compared with domestic academic circles, foreign

researches on landscape image recognition are earlier. Jose et al. [1] described a new statistics-based method to retrieve images of natural scenes, which combined feature extraction, automatic clustering, automatic indexing and classification techniques, and could classify images into coast, mountain, forest and plain with high accuracy. Sowmya et al. [2] proposed DBN (Deep Belief Networks) according to BoW bag of Words model by using existing feature extraction techniques, and applied SVM (Support Vector Regression). To learn features from the output layer of the proposed DBN structure in order to improve the performance of the proposed scene classification system. The color scene classification system can classify 8 kinds of scene images, such as coast, forest and mountain, with an average accuracy of 94.1%. Krstinic et al. [3] used CNN (Convolutional Neural Networks) to learn the features of 12 natural landscapes, including clouds, sunrise and sunset, sea and sky, collected by the forest fire detection tower. Combined with transfer learning, the average classifi-

\*Corresponding author. Nong Limei, E-mail: nong.limei@998741000@hotmail.com.

cation accuracy reaches 92.3%. Zhou Yunlei et al. [4] used PCA (Principal Component Analysis) to carry out weighted fusion extraction of various underlying features of natural landscape images, and then combined with support vector machine to complete the classification of landscape images, and the final classification accuracy was about 80%.

Deep learning neural networks offer several advantages in recognizing landscape images [1–8]. In terms of feature learning, deep learning models can automatically learn hierarchical representations of features from raw data. In the context of landscape images, this means that the network can learn to identify and combine simple features (such as edges, textures) to more complex ones (objects, scenes) without explicit feature engineering. In the case of complex pattern recognition, landscape images often contain intricate patterns, textures, and structures [8–12]. Deep neural networks, especially convolutional neural networks (CNNs), excel at capturing these complex patterns, making them well-suited for recognizing diverse landscapes. In scale invariance, deep learning models can be designed to be invariant to scale variations. This is beneficial for recognizing landscapes where the size of objects or scenes may vary. Convolutional layers in CNNs, for example, automatically handle scale invariance by using filters that operate over local regions of the input. In the case of transfer learning, transfer learning allows pre-trained models on large datasets (e.g. ImageNet) to be fine-tuned for specific tasks like landscape recognition [13–18]. This leverages the knowledge gained from one domain to another, even if the target domain has limited labeled data. In addition, deep learning models have high potential of adaptability to different lighting conditions, weather, and seasonal changes. This adaptability is crucial for recognizing landscapes in various environments and under different circumstances. Turning to End-to-End Learning, deep learning models can be trained in an end-to-end fashion, meaning that the network learns to map input images directly to output labels. This eliminates the need for manual extraction of features or intermediate representations, simplifying the overall pipeline. Moreover, landscapes can vary significantly in terms of terrain, vegetation, and other elements. Deep learning models, especially with large and diverse training datasets, can learn to generalize well and recognize landscapes with a wide range of variations. Furthermore, deep learning models, when appropriately trained with sufficient data and computational resources, can achieve high levels of accuracy in

image recognition tasks. This is crucial for reliable landscape recognition, especially in applications such as autonomous vehicles, satellite image analysis, and environmental monitoring [19–25].

In the context of semantic understanding, deep learning models can capture semantic relationships within the landscape, understanding not just individual objects but also their contextual relevance. This semantic understanding contributes to more meaningful and accurate recognition of entire scenes. When it comes to continuous improvement, deep learning models can benefit from continuous learning and improvement. As more labeled data becomes available, models can be retrained to further enhance their performance in recognizing landscapes [10].

Through the collation and summary of the literature on the application of machine learning to landscape image recognition methods at home and abroad, it can be seen that most scholars have adopted various types of machine learning models (i.e., DBN, SVM, CNN, Least absolute shrinkage and selection operator [LASSO] and random forest [RF]) to carry out image recognition research on the coast, mountains, sky, clouds and other natural landscapes, and have achieved fruitful research results [5–10]. But at the same time, it can be seen that domestic and foreign experts and scholars have used machine learning to carry out research on landscape image recognition of natural landscape and cultural landscape integration, and the image recognition research in this field is still in its infancy. The innovation of this paper lies in the establishment of a local multi-level classification model based on the optimization of landscape image target detection and local feature extraction algorithm, which realizes the accurate recognition of landscape images. The research process of this paper is shown in Fig. 1. Based on this, this paper uses deep learning algorithm to systematically study the image recognition algorithm of landscape architecture. The key contributions of this paper are presented as follows:

- (1) The SSD algorithm is introduced to detect the landscape image target. Combined with the research purpose of this paper, the training and prediction strategies of SSD are optimized and adjusted. On this basis, the process of landscape image recognition algorithm based on SSD is developed.
- (2) Gaussian distribution modeling is used to extract the features of second-order information in landscape images. The Stacking fusion

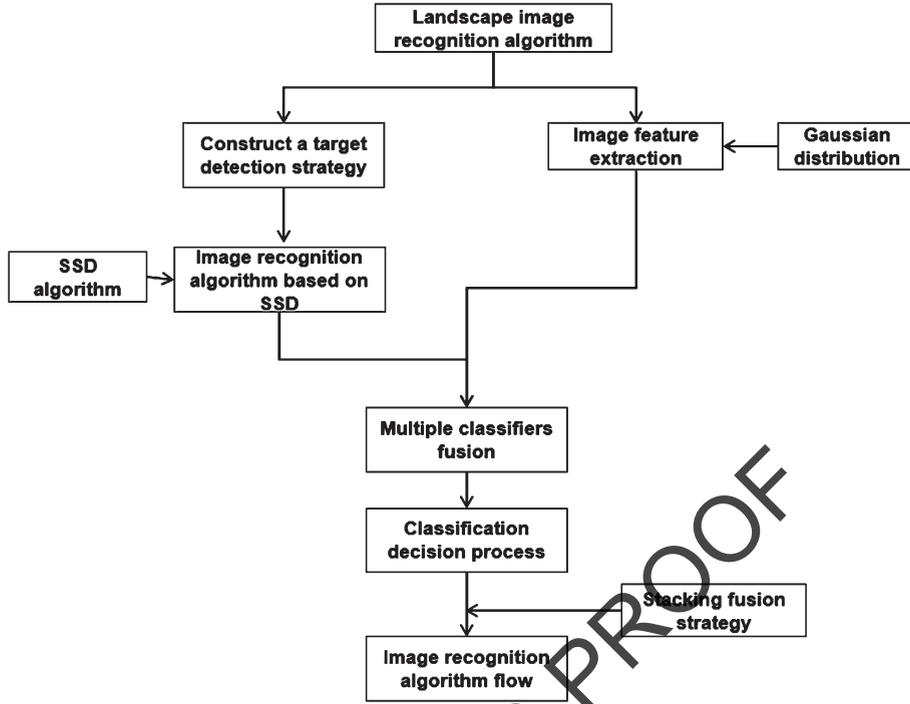


Fig. 1. Research flow chart.

strategy is introduced to solve the feature extraction problem of parallel multi-branch structures. The application flow of the Stacking fusion strategy in this paper is given.

- (3) Combined with the above optimization of target detection and local feature extraction algorithm of landscape images, the overall position-based multi-level classification model is presented. Finally, in the CVPR 2023 landscape data, the accuracy of the landscape image recognition algorithm constructed in this paper is analyzed experimentally.

## 2. The target detection strategy of landscape based on SSD algorithm

### 2.1. SSD algorithm

As one of the current mainstream detection box algorithms, SSD takes VGG-16 as the reference network, replaces the two fully connected layers FC6 and FC7 of VGG-16 with  $3 \times 3$  convolution Conv6 and  $1 \times 1$  convolution Conv7, and modifies the pooling layer Pool5 from  $2 \times 2$  pooling to  $3 \times 3$  pooling. Then remove the Dropout layer and the full-connection layer FC8, and add four new convolutional layers:

Conv8, Conv9, Conv10, Conv11. The basic structure of an SSD is shown in Fig. 2.

The main idea of SSD algorithm is to use CNN to extract image features, and carry out intensive sampling at different locations of the image. Different scales and aspect ratios can be used in sampling. These sampling locations are called default boxes and are used to predict the presence or absence of objects as well as the category of objects, with object classification occurring simultaneously with regression of the prediction boxes. The advantages of SSD are obvious. The multi-scale feature fusion strategy is used, and the output of multiple convolutional layers is used to participate in the prediction, which improves the detection accuracy. The default box is used directly for detection, and there is no process of generating candidate regions, which improves the detection speed. However, the disadvantage is that the size of the default box needs to be manually set, and the size and shape of the default box used by each layer are different, resulting in the network debugging process is very dependent on experience.

### 2.2. Training and prediction strategies

The SSD needs to input the image and the actual mark box for each target in the image during training.

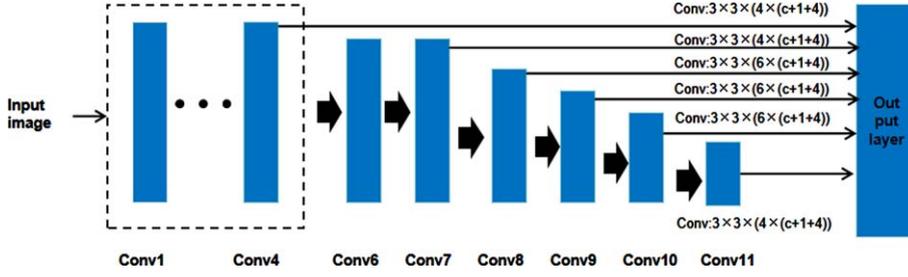


Fig. 2. Structure of SSD.

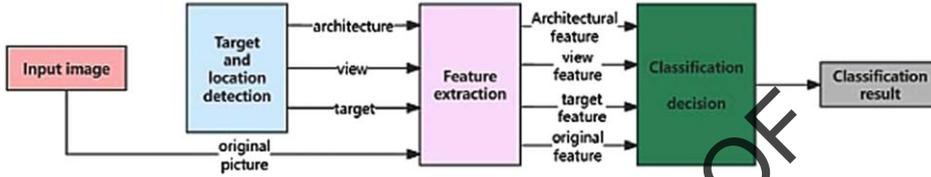


Fig. 3. Algorithm flow chart.

In the training process, the definition of positive and negative samples should be determined first. At this time, the concept of Intersection over Union (IoU) should be introduced. IoU is the overlap rate between two rectangular boxes, that is, the ratio between the intersection part and the union part. The larger the IoU, the larger the overlap rate of the two boxes. Thus, it can be considered that the two are close. In order to ensure that the number of positive and negative samples is basically balanced, SSD adopts the difficult negative sample mining strategy, that is, the negative samples are arranged in descending order according to the confidence error, and several samples with the largest error are selected as difficult negative samples to participate in the training, and the ratio of positive and negative samples in the training is ensured to be about 1:3. The second important problem in training is the loss function. The loss function of SSD is the weighted sum of the confidence error and the position error. Among them, the confidence error adopts cross entropy loss, and the position error adopts Smooth L1 loss.

The prediction process of SSD is relatively simple: For each prediction box, first determine its category according to the principle of maximum confidence, filter out the background class and the prediction box whose confidence is lower than the threshold value (usually 0.5), and then sort the remaining detection results in descending order of confidence. After retaining several prediction boxes with the highest confidence (usually 400), the NMS (Non-Maximum Suppression) algorithm is used to remove the predic-

tion results with large overlap, and the last remaining prediction box is the detection result.

In this paper, when SSD is applied to the detection of landscape targets and parts, the training and prediction strategies are adjusted according to the actual situation. During the training, the object in the garden landscape is taken as the key part, and the garden landscape target, architectural landscape and natural landscape are sent into the network as different categories for training. That is, by adjusting the number of output channels of the convolutional predictor, SSD can complete the detection of original drawing, object, architectural landscape and natural landscape. When forecasting, the threshold value is no longer used to filter the prediction box with low confidence, and the prediction box with the highest confidence for each category of target, building, and natural landscape is selected as the detection result. In addition to the target and location, the algorithm also uses the original image to participate in classification and recognition. Therefore, after processing by this module, the recognition system has a parallel multi-branch structure, which is respectively the original image, architectural landscape, and natural landscape 4 branches.

### 2.3. SSD-based garden landscape image recognition algorithm

Considering that garden landscape is a combination of nature and humanity, this paper defines the image view of garden landscape as composed of

humanistic architecture and natural scenery. On this basis, three modules of target and location detection, feature extraction and classification decision are designed by using the position-based classification model and modularization idea. Considering that the difficulty of garden landscape image recognition is that there are large differences between the categories and small differences within the categories, and the distinguishing features often exist in some subtle local areas in the images, this paper takes object and part detection as the first module of the recognition system, and firstly locates the foreground target and key parts before classification and recognition. In order to achieve accurate extraction of image features in the subsequent modules. After the processing of this module, the recognition system has a parallel multi-branch structure, that is, the original image, the target image, the building and the landscape branch. The specific algorithm block diagram is shown in Fig. 3.

### 3. Image second-order information feature extraction based on Gaussian distribution modeling

Considering the complex structure of garden landscape images, simple features are difficult to express their information, this paper introduces the image second-order information based on Gaussian distribution modeling into the feature extraction module, and combines it with deep learning. Using CNN features as local features of images, Gaussian distribution modeling is carried out to extract higher-order information, so as to obtain more distinguishable image features.

The BoVW (Bag of Visual Words) model can obtain better classification performance than the global descriptor through the extraction and modeling of image local features, which mainly includes the steps of image local feature extraction, visual dictionary training, feature coding and normalization. However, considering that the feature space constructed by the BoVW model through the codebook will bring some quantization errors in the segmentation, this paper adopts the codebook-free model based on probability distribution modeling to conduct statistical modeling of the local features of the image and obtain the statistical information of the features to solve the error problem. Among the many probability models, Gaussian distribution is the most common probability distribution in nature, and it is also one

of the most commonly used probability models in the field of computer vision and pattern recognition. Its high-order properties make Gaussian distribution have a powerful ability of information expression. Set image local features:

$$X = \{x_1, x_2, \dots, x_N | x_i \in R^{d \times 1}\} \quad (1)$$

Then the expression of probability density is:

$$p(x) = |2\pi \Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (2)$$

In the formula,  $\Sigma$  is the covariance matrix,  $\mu$  is the mean vector, and the  $\Sigma$ MLE (Maximum Likelihood Estimation) formula of the pair is as follows. For the Gaussian distribution, solving the optimization problem can obtain the estimation of the covariance  $\Sigma = S$ . In this paper, the output of CNN is used as the local feature, and the second-order information of CNN and images based on Gaussian distribution modeling is introduced into the feature extraction module.

$$\min_{\Sigma} \frac{N}{2} \log |\Sigma| + \frac{1}{2} \text{tr}(\Sigma^{-1}S) \quad (3)$$

Covariance matrix is a common second-order information based on Gaussian distribution modeling, and its high order property makes it play an obvious role in the expression of image region information. The covariance matrix describes the association between data points by calculating the covariance between features such as pixel intensity, color, or texture of an image. For Gaussian distributions, the covariance matrix is a key property because it determines the shape of the distribution. In image processing, since the distribution of image pixel intensity or color is often close to Gaussian distribution, covariance matrix provides an effective way to describe the properties of these distributions. However, the ability of covariance descriptors constructed by original low-dimensional feature space to obtain more distinguishable information is limited. Therefore, this paper introduces RAID-G (Robust Approximate Infinite Dimensional Gaussian), the specific application process is as follows: First, the covariance and mean value in the original feature Space are obtained, and the Kernel function is used to map the original features to Reproducing Kernel Hilbert Space (RKHS) to obtain the high-dimensional features, and the approximation of the high-dimensional Gaussian distribution is realized in this space. Remember the mapping function corre-

sponding to kernel function as  $\phi$ , then the mean vector and sampling covariance matrix in RKHS can be calculated by the following formula:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \phi(x_i) \quad (4)$$

$$\hat{\Sigma} = \Phi(X)J\Phi(X)^T \quad (5)$$

where,  $\Phi(X)$ ,  $J$ ,  $I_N$  is the identity matrix of order  $N$ . In practical application, RAID-G can be combined with CNN to model Gaussian distribution with convolutional layer output as local feature.

#### 4. Fusion of multiple classifiers based on ensemble learning

##### 4.1. Classification decision process

After the processing of the above modules, the identification system has a parallel multi-branch structure. This method has the disadvantages of high feature dimension, difficult classifier training and weak information expression ability. In this paper, a decision level fusion strategy is proposed in the classification decision module based on the idea of ensemble learning. Firstly, all branches are trained separately, and then the classification results of multiple branches are integrated to get the final prediction. This module focuses on two tasks, namely the selection of base classifier and the fusion strategy of multiple classifiers. The specific diagram is shown in Fig. 4.

##### 4.2. Stacking based fusion strategy

The integration strategies of ensemble learning can be divided into average method, voting method and learning method. Because voting method will lose some score information, which is not conducive to the

accurate judgment of the system, average method is often insufficient or noisy in task training. Therefore, this paper adopts learning method as fusion strategy. Take class C classification problem as an example, assuming that the number of base classifiers is  $N$ , the input samples are represented by  $(x, y)$ ,  $x$  is the data,  $y$  is the label,  $R(x)$  is used to represent the fusion result of the multi-classifier system,  $res$  is the system prediction result, the output category score of the base classifier is:  $s_i(x)$ ,  $i = 1, 2, \dots, N$ ,  $s_i(x)$ ,  $R(x) \in R^{C \times 1}$  and the fusion strategy of the learning method is shown in Fig. 5.

The algorithm obtains a primary learner from the initial training data and uses the output of the primary learner as the input of the secondary learner to construct a training set to train the secondary learner. Compared with the average fusion strategy, the fusion strategy based on stacking does not assign weights to each classifier, but considers the interrelationships between different categories. MLR (Multi-response Linear Regression) is one of the best choices for the learning algorithm of the secondary learner when the category probability output of the primary learner is used as the input of the secondary learner. MLR is a linear classifier that performs linear regression for each class separately. During training, the corresponding position output of the training sample belonging to the class is set to 1, and the rest position is 0; When testing, the category with the largest output value is taken as the predicted result. In fact, MLR is a neural network with no activation function and no hidden layer, and the number of nodes in the output layer is equal to the number of categories. The Stacking diagram is shown in Fig. 6.

When applied to this problem, the input images are first fed into the recognition system, and the image features of each branch are obtained after being processed by the target and part detection and feature extraction modules, and then sent into the trained corresponding base classifier respec-

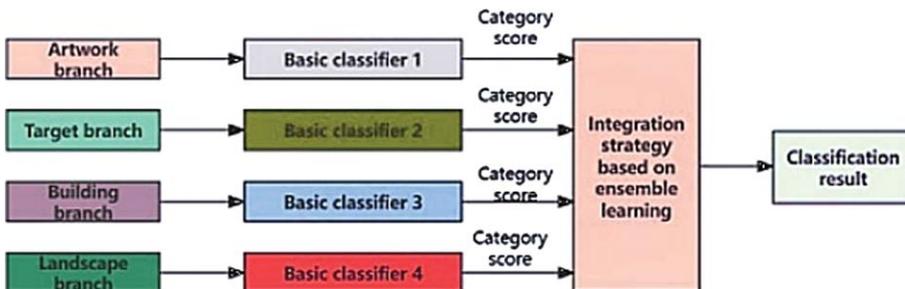


Fig. 4. Diagram of classification module.

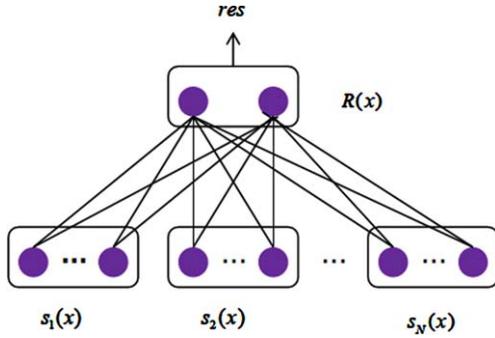


Fig. 5. Fusion strategy of learning method.

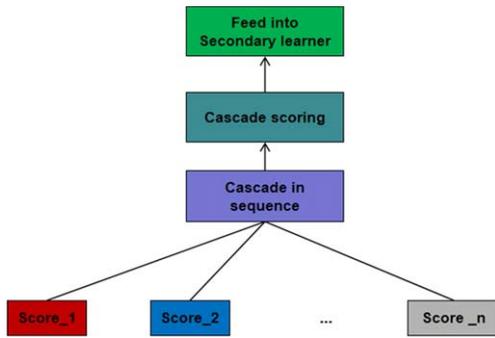


Fig. 6. Diagram of stacking.

tively to obtain multiple groups of category scores. The Stacking algorithm cascades the set of scores sequentially and serves as training samples for secondary learners. The data label of this sample is the label of the current input sample. The above operations are performed on all training samples during training to obtain the training data set of the secondary learner. MLR is used as the learning algorithm to obtain the secondary learner. During the test, the test samples are sent to the recognition system, the class score is obtained according to the above steps, and the new samples are cascaded in the same order, and the classification decision can be realized by sending the trained MLR.

## 5. Algorithm flow

The algorithm proposed in this paper adopts a position-based multi-level classification model and modular design, the output of the previous module is used as the input of the next module, which is functionally related to each other and separated from each other in training. Although the model structure can

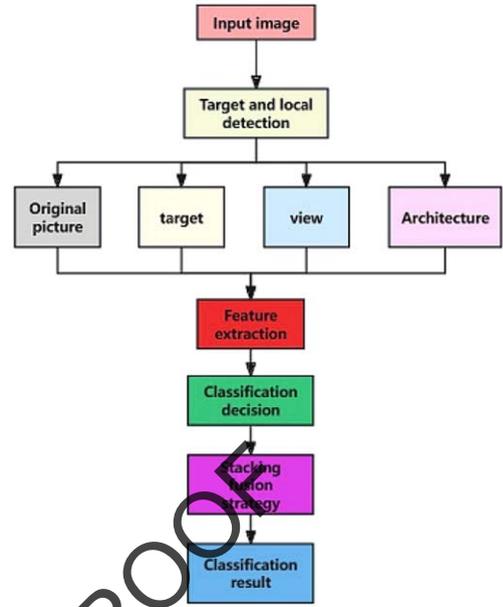


Fig. 7. Algorithm flow.

not bring end-to-end characteristics to the algorithm, the modular design makes the classification system more flexible and convenient for future updating and upgrading. The specific algorithm flow is as follows:

- (1) Given the input image, it is first sent to the target and location detection module to realize the detection and positioning of the target and key parts (buildings and landscapes), and the image block of landscape garden targets and parts is obtained. Then the recognition system is divided into four branches: original image, target, building and landscape.
- (2) The images of each branch are sent to the feature extraction module respectively to get the image features.
- (3) The classification decision module first trains the base classifier for different branches, and then realizes the fusion of multiple classifiers based on the idea of ensemble learning. After integrating the classification information of multiple branches, the category with the highest score is obtained as the classification result.

It is represented by a flow chart as Fig. 7. During the training phase, the SSD in the target and position detection module, the RAID-G and MPN-COV benchmark networks in the feature extraction module, the base classifier in the classification and decision module, and the secondary classifier MLR in the Stacking fusion strategy need to be trained respec-

tively. In the test phase, images are sent to the recognition system, and classification and recognition can be realized with only one forward propagation.

## 6. Experimental analysis

The experiment was carried out in the CVPR 2023 landscape data set. The samples of the database are clear and well marked, and the data division given by the official makes the comparison between algorithms more convincing. It has become the most commonly used standard database for fine-grained image recognition, especially in the field of landscape recognition. The images used for the experiment are shown in Fig. 8. In order to carry out this experiment, this paper is implemented on the CPU+GPU heterogeneous platform, and its specific parameter configuration is shown in Table 1.

### 6.1. Evaluation indicators

For image recognition system, this paper uses recognition accuracy as an evaluation index, that is, the ratio of the number of correct samples to the total number of samples. For the object and part detection module, this paper uses MIoU (Intersection over Union) as the evaluation index and the average value of the intersection over union of all categories. The intersection ratio is an index to measure the similarity between the predicted results of the model and the real label. It is calculated by dividing the intersection size between the predicted results and the real label by the size of the union.

The object and location detection module uses SSD-300 to realize the object and location detection, and uses the annotated information given in the database for training, and sends the garden landscape objects, buildings and landscapes as different categories into the model to realize the detection task. The target marking frame given by the database can be directly used for SSD training, while the location marking points need to be manually processed to obtain the location marking frame.

### 6.2. Analysis of results

In this paper, MIoU is used as the evaluation index of the object and part detection module. This paper presents an SSD-based landscape image recognition algorithm and a region-based Convolutional Networks with Deformable Part-based Pose based

Table 1  
Parameters of CPU+GPU heterogeneous platform

CPU Configuration	Model number	Intel®Core™ i7-7700@3.6GHz
GPU Configuration	Internal memory	32 GB
	Model number	NVIDIA GeForce GTX 1080Ti
	Core frequency	1582MHz
	Video memory frequency	11000MHz
	Video memory capacity	11GB
	Video memory width	352bit
	CUDA Core	3,584
	CUDA run version	9.0
	CUDA driver version	9.0

Table 2  
Experimental results of object and location detection module

Methods	Pixel accuracy (%)		
	Garden	Buildings	Scenery
Part-based R-CNN	—	68.5	78.9
Deep LAC	—	74.0	96.0
Mask-CNN	—	83.2	91.6
SSD	99.5	99.2	97.0

on R-CNN (Region Based convolutional Networks with deformable Part-based pose) Prior), Deep LAC (Deep Learning for Audio Coding) and Mask-CNN (Masked Convolutional Neural Networks) were compared and analyzed. The specific experimental results are shown in Table 2. It can be seen from the experimental results that the image recognition algorithm constructed in this paper can obtain the highest positioning accuracy among the listed methods in both architecture and landscape positioning, and can also achieve good results in landscape target positioning. The calculation formula of MIoU is:

$$MIoU = \frac{1}{k} \sum_{i=1}^k \frac{P \cap G}{P \cup G} \quad (6)$$

In the formula,  $P$  is the predicted value,  $G$  represents the true value, and  $K$  represents the total number of categories. The formula represents the intersection of the predicted and true values for each category divided by the union and then averaged.

When extracting RAID-G features, VGG-16 is used as the reference network, the last three fully connected layers are removed, and the output of the last convolutional layer Conv5\_3 is used as the local feature. The full convolutional structure enables CNN to accept any size of image input. In this paper, the RAID-G method based on VGG-16 was used



Fig. 8. Experimental images.

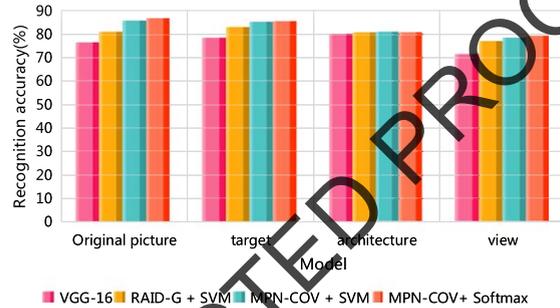


Fig. 9. Feature extraction accuracy of garden landscape images by different algorithms.

to extract the features of garden landscape images. In order to investigate its classification performance, this experiment used the above features to train one-to-many SVM, and compared it with the classification performance of its benchmark CNN model. In the experiment, the RAID-G based classifier is trained directly using extracted RAID-G features. Softmax classifier based on MPN-COV takes the Softmax layer of the network as the classifier according to its end-to-end features. The MPN-COV based one-to-many SVM first obtains the output feature of the network at the MPN-COV layer, and then uses this feature to train the classifier. The specific results are shown in Table 3.

In order to compare the recognition accuracy results more intuitively, the above table is converted into a histogram, as shown in Fig. 9. It can be seen that MPN-COV feature with end-to-end training characteristics can achieve the highest recognition accuracy on each branch, and its performance is better than MPN-COV based on one-to-many SVM, and it is far better than RAID-G. This shows that the image second-order information based on Gaussian distri-

Table 3  
Experimental results of feature extraction module

Methods	Recognition accuracy (%)			
	Original image	Target	Buildings	Scenery
VGG-16	76.5	78.4	79.9	71.2
RAID-G + SVM	80.9	83.0	80.6	76.9
MPN-COV + SVM	85.9	85.3	80.9	78.5
MPN-COV+ Softmax	87.0	85.6	80.8	79.2

bution modeling is a more distinguishable image feature, which proves the effectiveness of the feature extraction module.

## 7. Conclusion

To sum up, with the development of artificial intelligence, deep learning is applied more and more widely in the field of computer vision, and how to use deep learning to recognize images has become a research hotspot. However, landscape recognition

is a familiar but unfamiliar research direction. In the context of people's pursuit of higher spiritual satisfaction, it is of great significance to conduct in-depth research on landscape image recognition. This paper constructs an SSD-based landscape image recognition algorithm, introduces Gaussian distribution to extract image higher-order features, uses the fusion strategy of Stacking to train the classifiers, and establishes the final algorithm flow. The experimental results show that the landscape image recognition algorithm based on SSDS constructed in this paper has the highest positioning accuracy and recognition accuracy compared with traditional recognition methods. On the whole, although this paper has achieved certain research results, the object and part detection module built in this paper uses the strong supervised detection model SSD, which requires a lot of information annotation and consumes a lot of time and energy. In the following work, the weak supervised positioning algorithm can be considered to improve it, so as to reduce the dependence of the recognition system on annotated information.

## Funding

This study was supported by a grant from China National Social Science Foundation (No. 23BH158 to NONG Limei) and Beijing Social Science Foundation (No. 22YTC032 to NONG Limei).

## References

- [1] J.F. Serrano-Talamantes, C. Aviles-Cruz, J. Villegas-Cortez and J.H. Sossa-Azuela, Self organizing natural scene image retrieval, *Expert Systems with Applications* **40** (7) (2013), 2398–2409.
- [2] C. Chao, Application of machine learning in image analysis and recognition, *China New Technology and New Products* (24) (2023), 27–29.
- [3] D. Krstinic, M. Braovic and D. Boi-Tuli, Convolutional neural networks and transfer learning based classification of natural landscape images, *Journal of Universal Computer Science* **26** (2) (2020), 244–267.
- [4] Z. Yunlei, G. Jiechang, Z. Rong, L. Qingqing and J. Xiaofei, Landscape image classification based on multi-feature and support vector machine, *Application of Computer Systems* **25** (05) (2016), 135–141.
- [5] W. Bingnan, X. Fanjiang, Z. Quan et al., A survey on facial image deblurring, *Computational Visual Media* **10** (1) (2023), 3–25.
- [6] P. Siyue, Z. Fan, S. Yuexuan et al., Microscopic image recognition of diatoms based on deep learning, *Journal of Psychology* **59** (6) (2023), 1166–1178.
- [7] F. Xie and H. Wei, BesNet: Binocular ferrographic image recognition model based on deep learning technology, *Industrial Lubrication and Tribology* **75** (6) (2023), 714–720.
- [8] O.Y. Shaobo, H. Shuqin, S. Dan et al., The preliminary in vitro study and application of deep learning algorithm in cone beam computed tomography image implant recognition, *Scientific Reports* **13** (1) (2023), 18467–18467.
- [9] H.A. Zurqani, High-resolution forest canopy cover estimation in ecodiverse landscape using machine learning and Google Earth Engine: Validity and reliability assessment, *Remote Sensing Applications: Society and Environment* **33** (2024), 101095.
- [10] X. Zhenchang, L. Kuirong, G. Bill et al., Image recognition model of pipeline magnetic flux leakage detection based on deep learning, *Corrosion Reviews* **41** (6) (2023), 689–701.
- [11] M. Ding, J. Zhang, G. Shen et al., From photographic images to hierarchical networks—Color associations of a traditional Chinese garden, *Color Research & Application* **48** (6) (2023), 735–747.
- [12] J. Jayachitra, K. Suganya Devi, S.V. Manisekaran et al., An optimal deep learning model for recognition of hidden hazardous weapons in terahertz and millimeter wave images, *Earth Science Informatics* **16** (3) (2023), 2709–2726.
- [13] Z. Zhiyi, W. Biao, C. Wenwen et al., Recognition of abnormal individuals based on lightweight deep learning using aerial images in complex forest landscapes: A case study of pine wood nematode, *Remote Sensing* **15** (5) (2023), 1181–1181.
- [14] L. Man, H. Xianli, Y. Yinxue et al., Multiple factors influence coal and gangue image recognition method and experimental research based on deep learning, *International Journal of Coal Preparation and Utilization* **43** (8) (2023), 1411–1427.
- [15] X. Wang, X. Jin, Y. Feng et al., Landscape reconstruction of traditional village couplets based on image recognition algorithm, *Journal of Optics* **52** (1) (2022), 1–9.
- [16] J. Shi, T. Honjo, Y. Yazawa et al., Recognition and classification of homogeneous landscape with visitor-employed photography and cloud image annotation API—An example of the Riverscape in Nihonbashi, Tokyo, Japan, *Landscape Architecture Frontiers* **9** (5) (2021), 12–31.
- [17] Y. Yan-Fei, C. Li and H. Xiang-Yi, Recognition and classification of Rail Internal Damage B-display image Based on Improved YOLO v5, *Foreign Electronic Measurement Technology* **42** (12) (2023), 70–76.
- [18] H. Xiaohu and W. Hai, Research on welding seam image recognition based on deep learning, *Wireless Internet Technology* **20** (24) (2023), 126–132.
- [19] A.V. Nartova, A.V. Matveev, M.Y. Mashukov et al., iOk platform for automatic search and analysis of objects in images using artificial intelligence in the study of supported catalysts, *Kinetics and Catalysis* **64** (4) (2023), 458–465.
- [20] K. Vaishnavi, G. Pranay Reddy, T. Balaram Reddy et al., Real-time object detection using deep learning, *Journal of Advances in Mathematics and Computer Science* **38** (8) (2023), 24–32.
- [21] S.F. Wang, X.J. Xie, L. Zhang et al., [Research on multi-class orthodontic image recognition system based on deep learning network model], *Zhonghua kou qiang yi xue za zhi = Zhonghua kouqiang yixue zazhi = Chinese Journal of Stomatology* **58** (6) (2023), 561–568.
- [21] A.Q. ul Ain, R. Mudassar, K. Wazir Zada et al., Anomalous situations recognition in surveillance images using deep learning, *Computers, Materials & Continua* **76** (1) (2023), 1103–1125.

- [22] P. Siyue, Z. Fan, S. Yuexuan et al., Microscopic image recognition of diatoms based on deep learning, *Journal of Phycology* **59** (6) (2023), 1166–1178.
- [23] F. Xie and H. Wei, BesNet: binocular ferrographic image recognition model based on deep learning technology, *Industrial Lubrication and Tribology* **75** (6) (2023), 714–720.
- [24] Z. Wu and B. Xie, Fine-grained pornographic image recognition with multi-instance learning, *Computer Systems Science and Engineering* **47** (1) (2023), 299–316.
- [25] S. Hussein and S. Ali Salem Bin, Optimized deep learning vision system for human action recognition from drone images, *Multimedia Tools and Applications* (2023), 21–22.

CORRECTED PROOF