Image Recognition and Empirical Application of Desert Plant Species Based on Convolutional Neural Network Postprint

Authors: LI Jicai

Date: 2022-12-12T00:00:00+00:00

Abstract

In recent years, deep convolutional neural networks have demonstrated excellent performance in computer vision, yielding far-reaching impacts. Traditional plant taxonomic identification demands high expertise and is time-consuming. Most nature reserves face challenges such as incomplete species surveys, inaccurate taxonomic identification, and untimely updates to status data. Applying convolutional neural network technology to explore optimal network models can enable simple and accurate recognition of plant images. This study takes 24 typical desert plant species widely distributed in nature reserves of China's Xinjiang Uygur Autonomous Region as research objects. Using deep learning, we established an image database and selected an optimal network model for desert plant species recognition, providing decision support for fine-grained management activities in Xinjiang's nature reserves, such as species investigation and monitoring. Since desert plant species were not included in public datasets, the images used in this study were obtained primarily through field photography and downloads from the Plant Photo Bank of China (PPBC). After sorting and statistical analysis, a total of 2,331 plant images were collected (2,071 from field collection and 260 from PPBC), comprising 24 plant species belonging to 14 families and 22 genera. We conducted extensive numerical experiments comparing 37 convolutional neural network models with strong performance across different perspectives to identify the optimal network model most suitable for desert plant species recognition in Xinjiang. The results revealed 24 models with recognition Accuracy greater than 70.000%. Among these, Residual Network X_8GF (RegNetX_8GF) performed best, achieving Accuracy, Precision, Recall, and F1 (the harmonic mean of Precision and Recall) values of 78.33%, 77.65%, 69.55%, and 71.26%, respectively. Considering hardware requirements and inference time, Mobile Network V2 (MobileNetV2) achieved the best

Full Text

Preamble

LI Jicai[1], SUN Shiding[2], JIANG Haoran[2], TIAN Yingjie[1],[3],[4]*, XU Xiaoliang[5]

[1] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[2] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China
[4] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[5] Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

Abstract

In recent years, deep convolutional neural networks have demonstrated excellent performance in computer vision, yielding far-reaching impacts. Traditional plant taxonomic identification demands high expertise and is time-consuming. Most nature reserves face challenges such as incomplete species surveys, inaccurate taxonomic identification, and untimely updates to status data. Applying convolutional neural network technology to explore optimal network models can enable simple and accurate recognition of plant images. This study takes 24 typical desert plant species widely distributed in nature reserves of China's Xinjiang Uygur Autonomous Region as research objects. Using deep learning, we established an image database and selected an optimal network model for desert plant species recognition, providing decision support for fine-grained management activities in Xinjiang's nature reserves, such as species investigation and monitoring. Since desert plant species were not included in public datasets, the images used in this study were obtained primarily through field photography and downloads from the Plant Photo Bank of China (PPBC). After sorting and statistical analysis, a total of 2,331 plant images were collected (2,071 from field collection and 260 from PPBC), comprising 24 plant species belonging to 14 families and 22 genera. We conducted extensive numerical experiments comparing 37 convolutional neural network models with strong performance across different perspectives to identify the optimal network model most suitable for desert plant species recognition in Xinjiang. The results revealed 24 models with recognition Accuracy greater than 70.000%. Among these, Residual Network X_8GF (RegNetX_8GF) performed best, achieving Accuracy, Precision, Recall, and F1 (the harmonic mean of Precision and Recall) values of 78.33%, 77.65%, 69.55%, and 71.26%, respectively. Considering hardware requirements and inference time, Mobile Network V2 (MobileNetV2) achieved the best balance among Accuracy, number of parameters, and number of floating-point operations. MobileNetV2's number of parameters is 1/16 that of RegNetX_8GF, and its number of floating-point operations is 1/24. Our findings can facilitate efficient decision-making for species survey, cataloging, inspection, and monitoring management in Xinjiang's nature reserves, providing a scientific basis for the protection and utilization of natural plant resources.

Keywords: desert plants; image recognition; deep learning; convolutional neural network; Residual Network X_8GF (RegNetX_8GF); Mobile Network V2 (MobileNetV2); nature reserves

Citation: LI Jicai, SUN Shiding, JIANG Haoran, TIAN Yingjie, XU Xiaoliang. 2022. Image recognition and empirical application of desert plant species based on convolutional neural network. Journal of Arid Land, 14(12): 1440–1455. https://doi.org/10.1007/s40333-022-0086-9

*Corresponding author: TIAN Yingjie (E-mail: tyj@ucas.ac.cn)

Received 2022-09-01; revised 2022-10-14; accepted 2022-10-16

1 Introduction

Wild plants constitute the primary component of ecosystems in nature reserves, making species investigation and classification a fundamental task for managers (Li et al., 2020). Traditional plant classification and recognition demand high professional expertise and are time-consuming and inefficient (Cao et al., 2018). Consequently, most nature reserves suffer from incomplete species surveys, inaccurate taxonomic identification, and untimely data updates, preventing administrative agencies from implementing timely and effective protection, management, and ecological restoration countermeasures (Liu et al., 2018; Wang et al., 2019; Xiao, 2019). In June 2019, the General Office of the State Council of the People's Republic of China issued a document establishing national parks as the main component of the nature reserve system. Based on ecological environment regulation and big data platforms, information technologies such as cloud computing and the Internet of Things will be employed to comprehensively understand ecosystem composition, distribution, and dynamic changes in nature reserves, thereby providing scientific support for management decisions (http://www.gov.cn/zhengce/2019-06/26/content_{5403497}.htm). Therefore, developing efficient and accurate methods to obtain plant-related data using new information technologies has become an urgent challenge for researchers and managers.

Convolutional neural networks, with their powerful feature extraction capabilities, offer significant advantages for recognizing and analyzing high-dimensional data such as images, sounds, and texts. These networks can reduce damage to fragile plant resources from field specimen collection, decrease the difficulty of identifying and classifying similar plant species, and improve work efficiency (Mikolov et al., 2011). In recent years, intelligent recognition of plant images has gradually become a research hotspot (Liu, 2020). Previous literature has applied convolutional neural networks to recognize images of plant organs such as leaves, fruits, and flowers under simple backgrounds (Hall et al., 2015; Abdullahi et al., 2017; Bargoti and Underwood, 2017; Coulibaly et al., 2019; Cao et al., 2020). Researchers have also classified and recognized images of five crops and 100 different ornamental plant species in various natural scenes using convolutional neural networks (Simonyan and Zisserman, 2014; Kussul et al., 2017; Liu, 2018). Several mature convolutional neural network image recognition systems, such as "Xingse APP" and "Aiplants APP," have been widely used in wild plant resource surveys (Gao et al., 2020). However, the accuracy of these systems is generally low for classifying and recognizing desert plant images in complex natural scenes (Jin, 2020). For example, Zhang and Huai (2016) used hierarchical deep learning to train and recognize leaf images of plants in simple versus complex scenes, finding that recognition rates reached 91.11% for single-scene plants but only 34.38% for complex-scene plants.

The main problems are twofold. First, the number of images for the same desert plant species across different natural scenes is too small, with few images capturing the salient classification characteristics of desert plant species. Second, previous image recognition systems were based on datasets of urban and rural cultivated plants, specific plant organs, or simple background images. However, through long-term adaptation to the special desert environment, different desert plant species have developed similar external morphological characteristics (homogenization of plant and branch characteristics, similarity in branch morphology and color, highly degraded leaf patterns, etc.), which increases the difficulty of machine learning-based visual recognition and makes misjudgment more likely (He et al., 2006). To address the first problem, researchers have proposed methods for obtaining large numbers of plant images that meet technical requirements (Jin, 2020). The second problem represents the key scientific and technical issue that this study must address: how to significantly improve image recognition accuracy for similar plant species in complex natural scenes and select the optimal network model suitable for desert plant species recognition. This is a challenging task with broad practical applications.

Given the lack of research on desert plant species image recognition, this study takes panoramic image sets of major desert plant species distributed in Xinjiang's nature reserves as research objects. We integrated 37 non-lightweight and lightweight models across eight categories that are widely used today, including Visual Geometry Group Network (VGG), Residual Network (RegNet), and Mobile Network (MobileNet) (Krizhevsky et al., 2012; He et al., 2016; Howard et al., 2017). Using grid search to find optimal hyperparameters and comparing performance, we identified the network model best suited for desert plant species image recognition, enabling convenient and accurate classification and recognition of desert plant species and providing solutions for large-scale field plant background investigations in Xinjiang's nature reserves.

2.1 General Survey of Nature Reserves in Xinjiang

Currently, Xinjiang has 201 protected nature reserves covering an area of 2.51$×10^{5}$ km^2, accounting for 15.07% of Xinjiang's total land area (Fig. 1 [Figure 1: see original paper]). These include one World Natural Heritage Site, 28 nature reserves, 24 scenic spots, 13 geological parks, 57 forest parks, 51 wetland parks, and 27 desert parks. Regionally, there are 63 reserves in southern Xinjiang and 138 in northern Xinjiang, accounting for 31.00% and 69.00% of the total, respectively.

Fig. 1 Overview of Xinjiang and spatial distribution of nature reserves in Xinjiang. Note: The figure is based on the standard map (新 S(2021)023) from the Map Service System (https://xinjiang.tianditu.gov.cn/main/bzdt.html) marked by the Xinjiang Uygur Autonomous Region Platform for Common Geospatial Information Services, and the standard map has not been modified. Satellite image source: Geospatial Data Cloud (http://www.gscloud.cn/).

2.2 Dataset

Based on the "List of National Key Protected Wild Plants" (Ming, 2021), this study selected 24 representative xerophytic desert plant species distributed in Xinjiang's nature reserves as identification targets (Fig. 2 [Figure 2: see original paper]). Since desert plant species were not included in public datasets, images were obtained primarily through field photography and downloads from the Plant Photo Bank of China (PPBC; http://ppbc.iplant.cn/sp/12519). Field collection spanned from 2019 to 2021, with rangers in nature reserves commissioned to take pictures using digital cameras or mobile phones in natural environments. These RGB true-color images were saved in JPG format. Collected plant images were verified by experienced plant experts and manually labeled, with unclear images deleted directly. After sorting and statistical analysis, a total of 2,331 plant images were collected (2,071 from field collection and 260 from PPBC), comprising 24 plant species belonging to 14 families and 22 genera (Table 1). The training, validation, and test sets were allocated in a 3:1:1 ratio. Plant species information can be found in the *Flora of Xinjiang* (Xinjiang Flora Editorial Committee, 1992–2004) and the Red List of Chinese Biodiversity: Higher Plant Volume (http://www.iplant.cn/rep/protlist/4).

Fig. 2 Images of the selected 24 desert plant species in nature reserves in Xinjiang

2.3 Methods

Convolutional neural network is a branch of deep learning, representing a feed-forward neural network structure containing convolutional computation with deep architecture. In recent years, it has been widely used in image recognition (Lecun and Bengio, 1998). Convolutional neural networks include convolutional layers, pooling layers, and fully connected layers (Fig. 3 [Figure 3: see original paper]). The mathematical expression of the network is as follows:

$$F(x) = f_N(f_{N-1}(...(f_2(f_1(x)))))$$

where x represents the input image; $F(x)$ represents the network output, such as the corresponding class or probability of input image x; N represents the number of hidden layers; and f_i represents the function of layer i.

In the convolutional layer, f consists of multiple convolution kernels $(g_1, ..., g_{k-1}, g_k)$, with common kernel sizes of 1×1, 3×3, 5×5, etc. Each g_k represents a linear function in the kth kernel, expressed as:

$$g_k(x, y, z) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} \sum_{w=0}^{d-1} W_k(u, v, w) \cdot I(x - u, y - v, z - w)$$

where (x, y, z) represents the pixel position in input image I; $W_k(u, v, w)$ represents the weight of kernel k; and m, n, and d represent the height, width, and depth of the convolution kernel, respectively.

In the activation layer, f is a pixel-wise nonlinear function, specifically a rectified linear unit, represented by:

$$f(x) = \max(0, x)$$

In the pooling layer, f is a layer-wise nonlinear down-sampling function that gradually reduces the size of feature representations.

The fully connected layer can be considered a convolutional layer with a 1×1 kernel size. For classification tasks, a prediction layer (i.e., softmax layer) is typically added to the last fully connected layer to calculate the probability that input images belong to different classes. If the number of neurons in the prediction layer is C (i.e., the number of categories is C): $p_1, p_2, ..., p_C$, these C values can be converted to probability values through the softmax layer (Eq. 4):

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad \text{for } i = 1, 2, ..., C$$

Finally, the loss function is calculated, and parameters are updated through the stochastic gradient descent algorithm. The cross-entropy loss function is one of the most commonly used loss functions in deep learning, measuring the difference between true and predicted values (Li et al., 2020). It is calculated as:

$$loss = -\sum_{i=1}^{C} y_i \log(p_i)$$

where y_i and p_i represent the expected label value and predicted output value of sample i, respectively.

Generally, convolutional neural network performance improves as network depth increases, as seen in VGG with 16 layers, Google Inception Network (GoogLeNet) with 22 layers, and Residual Network (ResNet) with 152 layers (Simonyan and Zisserman, 2014). However, research shows that no single network structure guarantees superior performance across all datasets (Liu and Luo, 2019). For a specific dataset, the network structure with the best performance must be selected based on experimental results.

Therefore, this study adopted 37 common non-lightweight and lightweight network structures across eight categories, including Visual Geometry Group Network (VGG), ResNet, Dense Convolutional Network (DenseNet), Squeeze Network (SqueezeNet), MobileNet, Shuffle Network (ShuffleNet), Efficient Network (EfficientNet), and RegNet. We adjusted model parameters to find the best-performing network structure. The experimental environment consisted of: Intel(R) Xeon(R) CPU E5-2640 v3@2.60GHz, NVIDIA GeForce GTX 2080Ti, Ubuntu 18.04.1. We used the PyTorch 1.6 deep learning framework with batch sizes set to 4, 8, 16, and 32. Using the Stochastic Gradient Descent (SGD) optimization algorithm (Li et al., 2021), we set a learning rate size, momentum of 0.9, and weight decay rate of 0.005.

2.4 Accuracy Evaluation

This study used Accuracy, Precision, Recall, and F1 (the harmonic mean of Precision and Recall) to evaluate model performance (Cai, 2020). Accuracy measures the ratio of all correct classification results to total samples. Precision is the proportion of results predicted as positive that are actually positive. Recall is the proportion of all actual positive samples that are correctly predicted as positive.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

$$Precision = \frac{TP}{TP + FP} \times 100$$

$$Recall = \frac{TP}{TP + FN} \times 100$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100$$

where TP, TN, FP, and FN represent the numbers of true positive, true negative, false positive, and false negative samples in prediction results, respectively.

Model complexity was measured by the number of parameters and floating-point operations (Shen, 2021). The number of parameters refers to the total parameters requiring training in the network model, measuring model size. The number of floating-point operations refers to operations per second, measuring algorithmic complexity. Higher floating-point operation counts indicate slower convolutional neural network operation speeds.

3.1 Recognition Results and Comparative Analysis of Multi-Models

The model recognition results for plant images are presented in Table 2 . Thirteen models showed Accuracy below 70.000%, with three below 55.000%: SqueezeNet1_0, SqueezeNet1_1, and ShuffleNetV2_X0_5. Twenty-four models exceeded 70.000% Accuracy, with nine surpassing 75.000%: EfficientNet_B1, EfficientNet_B3, RegNetX_{400}MF, RegNetX_{800}MF, RegNetX_{3_2}GF, RegNetX_8GF, RegNetX_{16}GF, RegNetY_{3_2}GF, and RegNetY_{16}GF. RegNetX_8GF outperformed other networks, achieving Accuracy, Precision, Recall, and F1 values of 78.333%, 77.654%, 69.547%, and 71.256%, respectively.

We also compared the number of parameters and floating-point operations across different network structures. For parameter count, 16 models were smaller than 10.000 M (megabytes, referring to storage space occupied by model parameters; 1 M = 1024 kilobytes) and four models exceeded 100.000 M (VGG11, VGG13, VGG16, and VGG19). For floating-point operations, 15 models were smaller than 1.000 G (model operation speed; 1 G = 10^9/s). We quantified relationships between Accuracy and both parameter count and floating-point operations (Fig. 4 [Figure 4: see original paper]). Among models with Accuracy higher than 70.000%, MobileNetV2 achieved the best balance among Accuracy, parameter count, and floating-point operations. For MobileNetV2, Accuracy reached 71.429%, parameter count was only 2.255 M, floating-point operations were only 0.313 G, the Accuracy-to-parameter ratio was 31.676, and the Accuracy-to-floating-point-operations ratio was 228.206. Although RegNetX_8GF exhibited the best performance, its parameter count and floating-point operations were 16 and 25 times higher, respectively, compared to MobileNetV2.

3.2 Optimal Model for Image Recognition of Desert Plant Species

Based on comparative analysis of these results and considering hardware equipment and inference time, MobileNetV2 demonstrated the best comprehensive performance for desert plant species image recognition and shows strong application prospects for practical work. The classification results of MobileNetV2 and RegNetX_8GF are shown in Table 3 , with confusion matrices displayed in Figure 5 [Figure 5: see original paper].

Due to limited data availability for *Ammodendron bifolium*, division into training, validation, and test sets yielded inconsistent results lacking analytical value. Examining the remaining 23 species from a Precision perspective, MobileNetV2 achieved Precision higher than 60.000% for all species except *Oxytropis bogdoschanica* and *Haloxylon persicum*, demonstrating its ability to identify various plant types with high Accuracy. For misclassifications, the confusion matrix reveals that one *Caragana polourensis* image, one *Helianthemum songaricum* image, one *Tamarix taklamakanensis* image, and two *Corydalis kashgarica* images were recognized as *Oxytropis bogdoschanica*. Additionally, one *Salsola junatovii* image, one *Populus pruinosa* image, one *Tamarix taklamakanensis* image, one *Calligonum ebinuricum* image, and one *Haloxylon ammodendron* image were recognized as *Haloxylon persicum*.

Eremosparton songoricum showed the lowest Recall at only 33.000%. The confusion matrix indicates that two *Eremosparton songoricum* images were recognized as *Haloxylon ammodendron* in the test set. *Corydalis kashgarica* had the next lowest Recall at 38.500%. The confusion matrix shows that in the test set, one *Corydalis kashgarica* image each was recognized as *Ammopiptanthus nanus*, *Lagochilus lanatonodus*, and *Haloxylon ammodendron*, two images were recognized as *Oxytropis bogdoschanica*, and three images were predicted as *Caryopteris mongholica*.

Inspecting original images of wrongly classified species (Fig. 6 [Figure 6: see original paper]) reveals that due to long-term adaptation to harsh environments, desert plants exhibit similar shape characteristics or highly degraded leaves that are scaly or cylindrical, with plant shapes approximating round spheres. *Eremosparton songoricum* and *Haloxylon persicum* images from spring and summer are very similar in branch shape, branch color, and branching pattern. For example, vertically-viewed images of *Corydalis kashgarica*, *Ammopiptanthus nanus*, *Lagochilus lanatonodus*, *Haloxylon ammodendron*, *Oxytropis bogdoschanica*, and

Caryopteris mongholica are nearly spherical. Desert plants have different adhesive types for stem smoothness or leaf distribution, with scaly or cylindrical leaves leading to significantly different taxonomic characteristics. During computer vision recognition, fine-grained recognition of these subtle attributes is unclear, resulting in low similarity of external morphological features, low image recognition sensitivity, and high false positive rates for higher plants. This indicates that MobileNetV2's performance still needs improvement in recognizing similar but distinct species, and the image dataset requires enhancement. F1 values show that performance for *Eremosparton songoricum*, *Corydalis kashgarica*, *Oxytropis bogdoschanica*, and *Haloxylon persicum* is inadequate, affected by low Precision and Recall values. Considering all factors, plants with flowers, fruits, or distinct crown shapes and colors—such as *Populus euphratica*, *Cistanche deserticola*, *Calligonum ebinuricum*, *Prunus tenella*, *Lagochilus lanatonodus*, and *Caryopteris mongholica*—obviously differ from others lacking these characteristics and show better recognition performance (all indicators exceeding 80.000%). In conclusion, without expert intervention, the lightweight network MobileNetV2 achieves automatic classification of plant images accurately and quickly.

3.3 Verifying the Research

To validate the optimal model discovered in this study, we selected Tianchi Bogda Peak Nature Reserve and Ebinur Lake Wetland National Nature Reserve—both rich in desert plants—for empirical verification. Ebinur Lake Wetland National Nature Reserve hosts over 90.00% of desert plant species in the Junggar Basin, including some endangered and endemic species. It is one of the regions with the most abundant desert plant populations in China's inland river basins, with plant species accounting for approximately 64.00% of the country's total desert plant species (Yang et al., 2009). Tianchi Bogda Peak Nature Reserve covers the area where Bogda Peak—the main peak of the eastern Tianshan Mountains—is located. Within a horizontal distance of 80 km from south to north, it features a complete vertical mountain belt spectrum. With about 700 plant species, the area represents the most typical example of vertical mountains in the world's temperate arid regions and is included in UNESCO's Man and the Biosphere Programme (Su and Niu, 2016). Among the 24 desert plant species selected in this study, six are present in Ebinur Lake Wetland National Nature Reserve, and nine are distributed in Tianchi Bogda Peak Nature Reserve.

Empirical results for MobileNetV2 and RegNetX_8GF in desert plant species image recognition are shown in Table 4 . In Tianchi Bogda Peak Nature Reserve, both models achieved Accuracy, Precision, and Recall of 83.000% or higher. MobileNetV2 reached 83.871% Accuracy—nearly 5.00% higher than its 78.33% Accuracy for recognizing all 24 desert plant species. MobileNetV2 demonstrates high accuracy in desert plant species image recognition and shows strong application prospects for practical work. In Ebinur Lake Wetland National Nature Reserve, both models achieved evaluation indicators above 60.000% for all metrics. Both MobileNetV2 and RegNetX_8GF show high accuracy values. Comparing the two models across both reserves for recognizing 24 desert plant species, empirical identification performance was poorer in Ebinur Lake Wetland National Nature Reserve than in Tianchi Bogda Peak Nature Reserve, with lower evaluation indicator values. Multiple factors may explain this difference. Images of the nine desert plant species in Tianchi Bogda Peak Nature Reserve were obtained from the *Color Atlas of Wild Vascular Bundle Plants in Bogda Biosphere* (Su and Niu, 2016), with pictures screened and cleared. Images of the six desert plant species in Ebinur Lake Wetland National Nature Reserve were collected from the field without clearing or other processing. These differences contributed to the comparative findings. This comparison also illustrates the importance of dataset quality and quantity for image recognition network models and implies that no single network structure guarantees superiority across other network structures or datasets. For specific datasets, experiments must be conducted to select the network structure with the best performance based on results, which represents the practical significance of this research.

---

4 Discussion

The optimal MobileNetV2 model's Accuracy in desert plant species image recognition did not reach the 90.000% level achieved by plant species recognition in single-background images, indicating that this study's findings remain far from practical application (Zhang and Huai, 2016).

First, from the perspective of dataset construction, increasing image quantity and enhancing image quality help improve recognition accuracy (Li, 2022). This study used 2,331 plant images for model training and testing. However, large differences in pixel size between field-collected images (using mobile phones and cameras) and PPBC images affected classifier performance to some extent. Future work can employ transfer learning, data expansion, and image cleaning technologies to address insufficient data and inconsistent image standards. Regarding complex image backgrounds, recognition results from Tianchi Bogda Peak Nature Reserve (Table 4) show that Accuracy can exceed 80.000% when objects are focused and features are prominent. Barbedo (2016) also demonstrated that background removal can improve recognition accuracy by 3.000%. However, background removal requires substantial work and professional expertise, which is often difficult to achieve in applications. Theoretically, the more differential features extracted from an image, the higher the recognition accuracy (Gai et al., 2021). Obviously, plant leaves, flowers, and fruits offer advantages in multiple shape features, high recognizability, and strong discriminative power. Future research can fully utilize multi-feature fusion methods combining panoramic plant images with organ images (such as flowers, fruits, and leaves) to further improve model accuracy and sensitivity.

Second, from the perspective of data processing and analysis, fine-grained or new network structures can be considered to learn and obtain more expressive deep features. Misclassified plant images (Fig. 6) show that due to similar morphological characteristics among desert plant species, a high misjudgment rate remains problematic. On one hand, collection environment complexity causes uncertainty in expert labeling. However, Bekker and Goldberger (2016) verified that deep convolutional networks maintain high reliability when the number of mislabeled samples is not excessive. On the other hand, some recognition errors likely occur because the model learning process neglects or cannot distinguish certain subtle attribute features. For such extreme cases that cannot be effectively distinguished visually, prior knowledge should be integrated into decision-making. Introducing existing plant family and genus classification labels as prior information to improve neural network generalization ability and make it more suitable for desert plant species image recognition in Xinjiang will become a future research focus (Cao et al., 2018).

Third, regarding learning algorithms, current convolutional neural network parameter adjustment relies primarily on experience and practical operation, requiring constant training, parameter tuning, and repeated trial-and-error, which consumes substantial time and energy (Tang, 2020). Auto Machine Learning (AutoML) has emerged as a popular research field in recent years (Liu and Luo, 2019). It automatically builds network structures that can guarantee the same accuracy as manually selected classic networks. Applied to species identification in natural protected areas, AutoML is expected to overcome subjective faults in manual selection, objectively select better network structures, and improve image recognition accuracy.

Finally, regarding model application scope, rare desert plant species also include *Betula holophila*, *Reaumuria kaschgarica*, etc. (Yin, 1991). However, due to collection difficulties, this study selected only some desert plant species in Xinjiang as research objects. Additionally, this study is based on static image data processing and analysis. Currently, numerous video surveillance systems have been deployed in Xinjiang's nature reserves. Therefore, strengthening research on video image data recognition represents an important future direction.

---

5 Conclusions

Based on image processing and deep learning technology, this study adopted 37 commonly used non-lightweight and lightweight convolutional neural network models across eight categories to recognize images of 24 typical desert plant species distributed in Xinjiang. Results show that 24 models achieved Accuracy above 70.000%, with nine models exceeding 75.000%. Among these, RegNetX_8GF outperformed other network models, with Accuracy, Precision, Recall, and F1 values of 78.333%, 77.654%, 69.547%, and 71.256%, respectively, meeting conventional image recognition requirements. To further examine relationships between Accuracy and both parameter count and floating-point operations in models with Accuracy above 70.000%, we found that MobileNetV2 achieves the best balance among Accuracy, parameter count, and floating-point operations. MobileNetV2's parameter count is 1/16 that of RegNetX_8GF, and its floating-point operations are 1/24. Considering hardware equipment, inference time, and other factors, MobileNetV2 demonstrates optimal performance for desert plant species image recognition and is more suitable for field investigations. To verify this study's effectiveness, we empirically tested RegNetX_8GF and MobileNetV2 for desert plant species image recognition in Tianchi Bogda Peak Nature Reserve and Ebinur Lake Wetland National Nature Reserve, finding that MobileNetV2 shows strong application prospects for practical work.

Due to dataset limitations, image recognition accuracy still requires improvement. Future research will further enrich desert plant species image sets in Xinjiang through multiple methods and forms, optimize convolutional neural network models, improve test accuracy, and provide solutions for nature reserve administrative agencies to conduct large-scale field plant background investigations, thereby enhancing work efficiency and decision-making capacity.

Acknowledgements

This work was supported by the West Light Foundation of the Chinese Academy of Sciences (2019-XBQNXZ-A-007) and the National Natural Science Foundation of China (12071458, 71731009).

References

Abdullahi H S, Sheriff R E, Mahieddine F. 2017. Convolution neural network in precision agriculture for plant image recognition and classification. In: 2017 Seventh International Conference on Innovative Computing Technology (INTECH). New York: IEEE, 10: 256–272.

Barbedo J G. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, 144: 52–60.

Bargoti S, Underwood J. 2017. Deep fruit detection in orchards. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). New York: IEEE, doi: 10.48550/arXiv.1610.03677.

Bekker A J, Goldberger J. 2016. Training deep neural-networks based on unreliable labels. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2682–2686.

Cai Z. 2020. Research on deep learning model for Chinese herbal medicine planting process. MSc Thesis. Chengdu: University of Electronic Science and technology. (in Chinese)

Cao X J, Mo Y, Yan Y L. 2020. Convolutional neural network flower image recognition using transfer learning. *Computer Applications and Software*, 37(8): 142–148. (in Chinese)

Cao X Y, Sun W M, Zhu Y X, et al. 2018. Plant image recognition based on family priority strategy. *Journal of Computer Applications*, 38(11): 3241–3245. (in Chinese)

Coulibaly S, Kamsu F B, Kamissoko D, et al. 2019. Deep neural networks with transfer learning in millet crop images. *Computers in Industry*, 108: 115–120.

Gai R L, Cai J R, Wang S Y. 2021. Research review on image recognition based on deep learning. *Journal of Chinese Computer Systems*, 42(9): 1980–1984. (in Chinese)

Gao H Y, Gao X H, Feng Q S, et al. 2020. Approach to plant species identification in natural grasslands based on deep learning. *Pratacultural Science*, 37(9): 1931–1939. (in Chinese)

Hall D, McCool C, Dayoub F, et al. 2015. Evaluation of features for leaf classification in challenging conditions. In: 2015 IEEE Winter Conference on Applications of Computer Vision. New York: IEEE, 797–804.

He K M, Zhang X Y, Ren S Q, et al. 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 770–778.

He M Z, Zhang J G, Wang H. 2006. Analysis of branching architecture factors of desert plants. *Journal of Desert Research*, (4): 625–630. (in Chinese)

Howard A G, Zhu M L, Chen B, et al. 2017. Efficient convolutional neural networks for mobile vision applications. [2022-09-01]. https://arxiv.org/abs/1704.04861v1.

Jin L T. 2020. Research on plant image recognition with complex background based on convolution neural network. MSc Thesis. Lanzhou: Lanzhou Jiaotong University. (in Chinese)

Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, 25: 1097–1105.

Kussul N, Lavreniuk M, Skakun S, et al. 2017. Deep learning classification of land cover and crop types using remote sensing data. In: 2017 IEEE Geoscience and Remote Sensing Letters. New York: IEEE, 14(5): 778–782.

Lecun Y, Bengio Y. 1998. Convolutional Networks for Images, Speech, and Time Series. Cambridge, MA: MIT Press, 255–258.

Li L P, Shi F P, Tian W B, et al. 2021. Wild Plant Image recognition method based on residual network and transfer learning. *Radio Engineering*, 51(9):

857–863. (in Chinese)

Li M M, Xia W C, Wang M, et al. 2020. Research on monitoring of Chinese nature reserves based on bibliometrics. *Journal of Ecology*, 40(6): 2158–2165. (in Chinese)

Li X H, Wu Z H, Liu H, et al. 2020. Species recognition of succulent plants based on convolutional neural network model. *Journal of Guizhou Normal University*, 36(3): 9–15. (in Chinese)

Li Y F. 2022. Research on image classification based on optimize on factors in convolutional neural network. *Journal of Jinling Institute of Technology*, 38(1): 26–31. (in Chinese)

Liu F Z, Du J H, Zhou Y, et al. 2018. Biodiversity monitoring technology and practice in nature reserves combining UAV and ground. *Biodiversity*, 26(8): 905–917. (in Chinese)

Liu H S. 2020. Panoramic plant recognition method based on CNN and GLCM fusion discrimination. MSc Thesis. Wuhan: Hubei University of Technology. (in Chinese)

Liu Y. 2018. Research on plant recognition based on deep learning. MSc Thesis. Beijing: Beijing Forestry University. (in Chinese)

Liu Y, Luo Z. 2019. Species recognition of protected area based on AutoML. *Computer Systems & Applications*, 28(9): 147–153. (in Chinese)

Mikolov T, Deo R A, Povey D, et al. 2011. Strategies for training large scale neural network language models. In: 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. New York: IEEE, 196–201.

Ming Y. 2021. The adjusted "List of National Key Protected Wild Plants" was officially announced. *Green China*, (19): 74–79. (in Chinese)

Shen T M J. 2021. Video denoising based on prior information and convolutional neural network. MSc Thesis. Chengdu: University of Electronic Science and technology. (in Chinese)

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. [2022-09-01]. https://arxiv.org/pdf/1409.1556.pdf.

Su H M, Niu S M. 2016. *Color Atlas of Wild Vascular Bundle Plants in Bogda Biosphere*. Beijing: China Forestry Press, 10–95. (in Chinese)

Tang M J. 2020. Research on fast prediction method of ship resistance performance based on convolutional neural network. MSc Thesis. Harbin: Harbin Engineering University. (in Chinese)

Wang Y W, Tang X L, Xu J P, et al. 2019. The use of big data in nature reserves. *China Forestry Economy*, (4): 16–20, 27. (in Chinese)

Xiao Z S. 2019. Application of infrared camera technology in wildlife inventory and assessment of natural reserves in China. *Biodiversity*, 27(3): 235–236. (in Chinese)

Xinjiang Flora Editorial Committee. 1992–2004. *Xinjiang Flora* (Volume I–Volume VI). Urumqi: Xinjiang Science and Technology Press. (in Chinese)

Yang X D, LV G H, Tian Y H, et al. 2009. Ecological grouping of plants in Lake Abby Wetland Nature Reserve in Xinjiang. *Journal of Ecology*, 28(12): 2489–2494. (in Chinese)

Yin L K, Pan B R, Wang Y, et al. 1991. Introduction and cultivation of rare and endangered plants in temperate desert. *Arid Zone Research*, (2): 1–8. (in Chinese)

Zhang S, Huai Y J. 2016. Leaf image recognition based on layered convolutions neural network deep learning. *Journal of Beijing Forestry University*, 38(9): 108–115. (in Chinese)

Note: Figure translations are in progress. See original paper for figures.