Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3107

RECOGNITION OF CITRUS PLANTING AREA BY INTEGRATING PYRAMID BOTTLENECK RESIDUAL NETWORK AND DECISION TREE ALGORITHM

Yue Deng, KaiMing Zeng*
Jiangxi Institute of Territorial Space Surveying and Planning, Nanchang 330029, Jiangxi, China.
**Corresponding Author: KaiMing Zeng*

Abstract: As the agricultural cornerstone of Ganzhou, Jiangxi Province, China, citrus production necessitates rapid and precise mapping of orchard spatial distribution for agricultural management, resource assessment, ecological conservation, and science-driven industry development. Remote sensing has emerged as a vital agricultural informatics tool in China due to its non-contact, large-scale data acquisition capabilities. Recent advances in deep neural networks enable state-of-the-art solutions for image processing and computer vision tasks. Capitalizing on this, we propose a multispectral remote sensing framework integrating a Pyramid-Shaped Residual Network with a decision tree algorithm customized for citrus cultivation patterns. The results indicate that the new method demonstrates extremely high accuracy, with an estimation precision exceeding 80% when compared to the statistical yearbook. This approach demonstrates significant potential for citrus planting area identification, providing a valuable reference for precision agriculture applications.
Keywords: Citrus plantation area identification; Multi-spectral; Deep neural network; Pyramid residual

1 INTRODUCTION

Citrus, as the world's largest fruit crop, leads in both planting area and production. China, as one of its primary origins, boasts abundant resources and diverse varieties. The Gannan region, known as the "World Orange Town," has vigorously developed citrus cultivation since the 1980s. Today, its planting area and annual output consistently rank among the top in the country, making it a dominant force in the Gannan fruit market and a significant contributor to the economic development of southern China[1]. It serves as a backbone industry for rural revitalization and farmer income growth. In recent years, the government has actively promoted the upgrading of the citrus industry through policies and economic measures. However, the expansion of planting areas driven by growing market demand has also triggered ecological and production challenges. Citrus is cultivated in red soil, a typical ecologically fragile zone. Many new orchards have been converted from forests and farmland, where tree felling and farmland encroachment threaten biodiversity and impact grain yields[2]. Additionally, frequent natural disasters such as pests, diseases, and frost damage have led to tree deaths and fluctuations in planting areas, resulting in unstable production and prices, which in turn affect the economic livelihoods of residents. Against this backdrop, scientific monitoring of citrus cultivation has become particularly urgent.

Remote sensing technology, with its advantages of wide coverage, rapid information acquisition, and low cost, has been widely applied in agriculture[3]. It plays a crucial role in extracting area information, monitoring growth conditions, estimating yields, and detecting and preventing crop pests and diseases[4]. In citrus planting information extraction, compared to traditional field surveys, remote sensing enables large-scale dynamic monitoring, quickly and accurately obtaining orchard spatial distributions, significantly improving the efficiency and scientific rigor of information extraction. However, Ganzhou City is located in the hilly and mountainous regions of southern China, characterized by significant topographic undulations and diverse landforms, including mountains, hills, gorges, and rivers[5]. The fragmentation of land parcels results in weaker spatial connections between land objects, uneven distributions, and diverse vegetation communities, which complicates crop classification. Traditional machine learning extraction methods, such as Random Forest and Support Vector Machine models, are severely affected by the same object, different spectra effect and interference from multi-source background information. The extraction results often exhibit salt-and-pepper noise, with low accuracy. Object-oriented methods require frequent human-computer interaction, making it difficult to determine the segmentation scale for fragmented and irregular citrus planting areas. These methods inadequately mine complex features in the imagery, leading to limited extraction accuracy that struggles to meet practical application requirements. Additionally, cloudy and rainy weather in southern China affects the revisit cycles of medium- and high-resolution satellite data[6].

Deep neural networks are characterized by high automation and wide applicability, enabling automatic and in-depth extraction of image feature information with minimal interference from background information. In recent years, deep learning-based deep neural networks have emerged as advanced machine learning techniques, capable of addressing various tasks in image processing, computer vision, signal processing, and natural language processing. Deep learning has made significant progress in image classification tasks. In 2012, Krizhevsky et al.[7] proposed the convolutional neural network (CNN) architecture AlexNet[8], which won the ImageNet Large-Scale Visual Recognition Challenge.

Subsequently, VGG and GoogleNet surpassed AlexNet's record on ImageNet. However, when attempting to use deeper networks to improve classification accuracy, issues such as gradient explosion, gradient vanishing, and network convergence arise, making it difficult to increase network complexity and leading to degraded classification performance. To address these problems, the residual network (ResNet) introduced the concepts of residual learning and identity shortcut connections, effectively mitigating the aforementioned issues[9]. These methods have made it possible to train deeper network architectures. Paoletti et al.[10] proposed the Pyramid-Shaped Residual Network (PSSRN) by improving the SSRN. This method groups convolutional layers into pyramid residual modules and gradually increases the feature map dimensions across all convolutional layers, allowing the diversity of high-level spectral-spatial features to increase layer by layer, thereby achieving precise image classification.

This study selected Xingguo County, Ganzhou City, Jiangxi Province, China, as the research area. Utilizing multispectral remote sensing imagery and addressing challenges such as fragmented citrus planting plots, uneven distribution of land objects, and diverse vegetation communities, we conducted information extraction for citrus planting. We proposed a method combining a deep pyramid residual network with a decision tree tailored to citrus planting characteristics, achieving the identification of citrus planting areas in southern Jiangxi. The aim is to accurately grasp the spatial distribution of citrus planting, providing support for ecological protection, scientific yield estimation, and government decision-making.

2 MATERIALS AND METHODS

2.1 Study Area

Xingguo County (26°03′–26°44′N, 115°01′–115°52′E) is located in the south-central part of Jiangxi Province and the northern part of Ganzhou City, covering an area of 3,215 km². The topography is dominated by low mountains and hills, with red soil as the primary soil type. It belongs to a subtropical monsoon climate, with an average annual temperature of 18.9°C and an average annual precipitation of 1,522.3 mm (Figure 1).

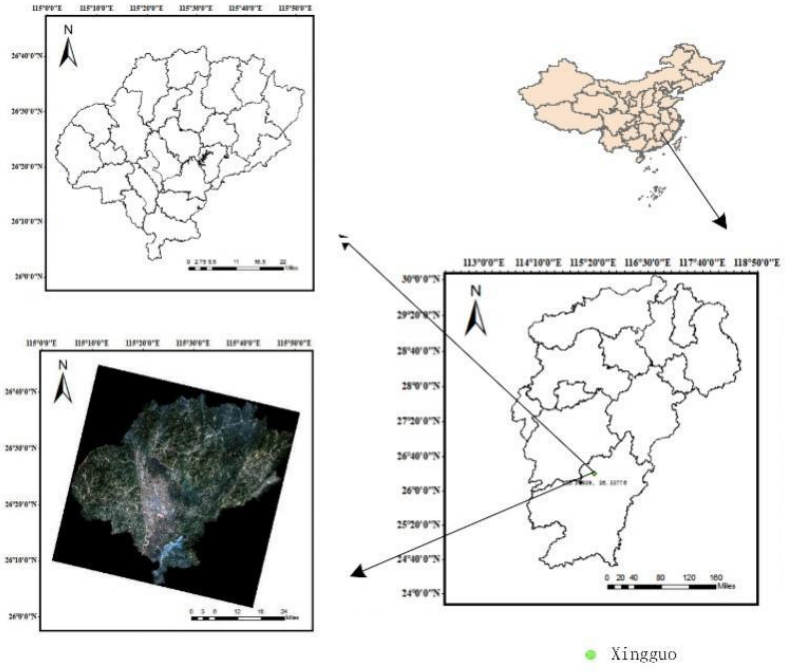

Figure 1 Location of the Study Area
Source: It was obtained from the Standard Map Service of the Ministry of Natural Resources of China. (available at: http://bzdt.ch.mnr.gov.cn/)

2.2 Data Source and Preprocessing

The remote sensing data of GF-1 were obtained from the China Centre for Resources Satellite Data and Application. GF-1 can acquire three types of remote sensing images: 2 m panchromatic (PAN), 8 m multispectral (MS), and 16 m multispectral (WFV) images. The revisit cycles for PAN and MS images are 41 days, while that for WFV images is 4 days. This study utilized WFV images acquired on December 6, 2020, November 27, 2021, December 21, 2022, and November 27, 2023. These data underwent preprocessing, including registration and clipping (Table 1).

Table 1 Parameters of GF-1 Satellite Data

GF-1 Data Type	PAN	MS	WFV
Spatial Resolution	2m	8m	16m

revisit period	41day	41day	4day
Spectral Range	450nm-900nm	Blue: 450nm-520nm Green: 520nm-590nm Red: 630nm-690nm Near-infrared: 770nm-890nm	

The digital elevation data (DEM) was sourced from NASA's Earthdata platform, with a spatial resolution of 12.5 m × 12.5 m. This resolution was resampled to match that of the WFV imagery, and the data was then integrated into the input of the deep neural network through band synthesis to utilize slope information in citrus planting area extraction.
The citrus planting area data of Xingguo County, Ganzhou City, Jiangxi Province, were sourced from the Jiangxi Statistical Yearbook (2020–2023). This data was used for classifier accuracy validation and annual assessment of citrus planting area extraction.

2.3 Method

2.3.1 Pyramid-Shaped Residual Network (PSSRN)

Paoletti et al. first introduced the Pyramid-Shaped Residual Network (PSSRN) for spectral-spatial classification of hyperspectral imagery. In the PSSRN structure, the number of feature maps output by each residual unit is not fixed but gradually increases, resembling a pyramid that widens from top to bottom[11]. The network consists of five modules: an input convolutional module (C), three pyramid modules (M1, M2, and M3), a pooling module (P), and a fully connected module (FC). Each pyramid module comprises three pyramid bottleneck residual units. The pooling module uses average pooling to downsample the output feature maps of the last pyramid module (M3) for the final time, reshaping them into vectors to be fed into the fully connected layer at the end of the network for classification. The input to the PSSRN is a multi-band multispectral image patch, and the output is the classification category. Excluding the input layer, output layer, BN layer, and ReLU layer, the entire network has a depth of 30 layers.
The PSSRN was designed for hyperspectral images, and the differences in spectral and spatial resolutions make it suboptimal for direct application to multispectral classification. To address this, we modified the model to adapt it to multispectral classification, considering the differences between the two types of images.
On one hand, multispectral images have superior spatial resolution compared to hyperspectral images. Due to the lower spatial resolution of hyperspectral data, smaller input image patches are sufficient to extract texture features effectively, achieving precise classification. In contrast, the high spatial resolution of multispectral images requires larger input image patches to fully extract texture features and enhance the distinction between different land cover types, thereby capturing neighborhood and non-local similarities within the images. Therefore, we increased the input image patch size to 27×27 to extract image features over a larger area. Additionally, considering the richness of details in multispectral images, we also reduced the filter size to extract details around each pixel, further improving classification accuracy.
On the other hand, multispectral images have far fewer bands than hyperspectral images. To match the scale of the input data and prevent overfitting, we reduced the number of convolutional kernels, decreasing the initial layer's convolutional kernel count from 103 to 64. The detailed architecture and parameters of the final network are illustrated in Table 2, where "conv" represents the convolutional layer, "BN" represents the batch normalization layer, and "ReLU" represents the activation layer.
To achieve classification, auxiliary preprocessing steps are required. First, the original multispectral image is normalized to zero mean and unit variance. Second, to utilize spectral-spatial information, a 3D neighboring block around each pixel is extracted, denoted as $M_{ij} \in R^{N \times d \times d}$, where N is the number of bands and d is the block size. The cubic block is then input into the model as the input data, with the boundary areas padded using the mirror method. The fully connected layer of the network uses softmax to output class labels.

Table 2 Proposed Deep Network Architecture

Modules	Units	Layers	Kernel size	Kernel numbers	Stride
C		conv	3×3	64	1
Pyramid Module M_1	Pyramid Bottleneck Residual Block $B(1)_1$	BN			
		conv	1×1	64	1
		BN, conv	5×5	32	1
		BN, conv	1×1	84	1
		Relu			
	Pyramid Bottleneck Residual Block $B(1)_2$	BN			
		conv	1×1	84	1
		BN,conv	5×5	42	1
		BN,conv	1×1	104	1
		Relu			
	Pyramid Bottleneck Residual Block $B(1)_3$	BN			
		conv	1×1	104	1
		BN, conv	5×5	52	1
		BN, conv	1×1	124	1
		Relu			

		BN			
	Pyramid Bottleneck Residual Block $B(2)_1$	conv	1×1	124	1
		BN,conv	6×6	62	2
		BN,conv	1×1	144	1
		Relu			
		BN			
Pyramid Module M_2	Pyramid Bottleneck Residual Block $B(2)_2$	conv	1×1	144	1
		BN,conv	5×5	72	1
		BN,conv	1×1	164	1
		Relu			
		BN			
	Pyramid Bottleneck Residual Block $B(2)_3$	conv	1×1	164	1
		BN, conv	5×5	82	1
		BN, conv	1×1	184	1
		Relu			
		BN			
	Pyramid Bottleneck Residual Block $B(3)_1$	conv	1×1	184	1
		BN,conv	6×6	92	2
		BN,conv	1×1	204	1
		Relu			
		BN			
Pyramid Module M_3	Pyramid Bottleneck Residual Block $B(3)_2$	conv	1×1	204	1
		BN,conv	5×5	102	1
		BN,conv	1×1	224	1
		Relu			
		BN			
	Pyramid Bottleneck Residual Block $B(3)_3$	conv	1×1	224	1
		BN,conv	5×5	112	1
		BN,conv	1×1	244	1
		Relu			
Pyramid Pooling Module P		Average Pooling			
Fully Connected FC	——	Fully Connected, soft max			

2.3.2 Combination algorithm of PSSRN and decision tree

Based on field surveys and literature reviews, we found that citrus cultivation is constrained by slope and elevation, which are influenced by topography. Therefore, DEM data can be incorporated into the PSSRN input. Citrus orchards exhibit a certain degree of sparsity, allowing vegetation coverage to serve as a decision-making reference. Additionally, factors such as slope and illumination shadows affect citrus planting. For example, the DN values of GF-1 imagery in the blue and near-infrared bands can be utilized as references. Consequently, we integrated the citrus extraction results from the optimized pyramid bottleneck residual network with a decision tree for comprehensive evaluation, ultimately forming a complete citrus planting area map for the study region (Figure 2).

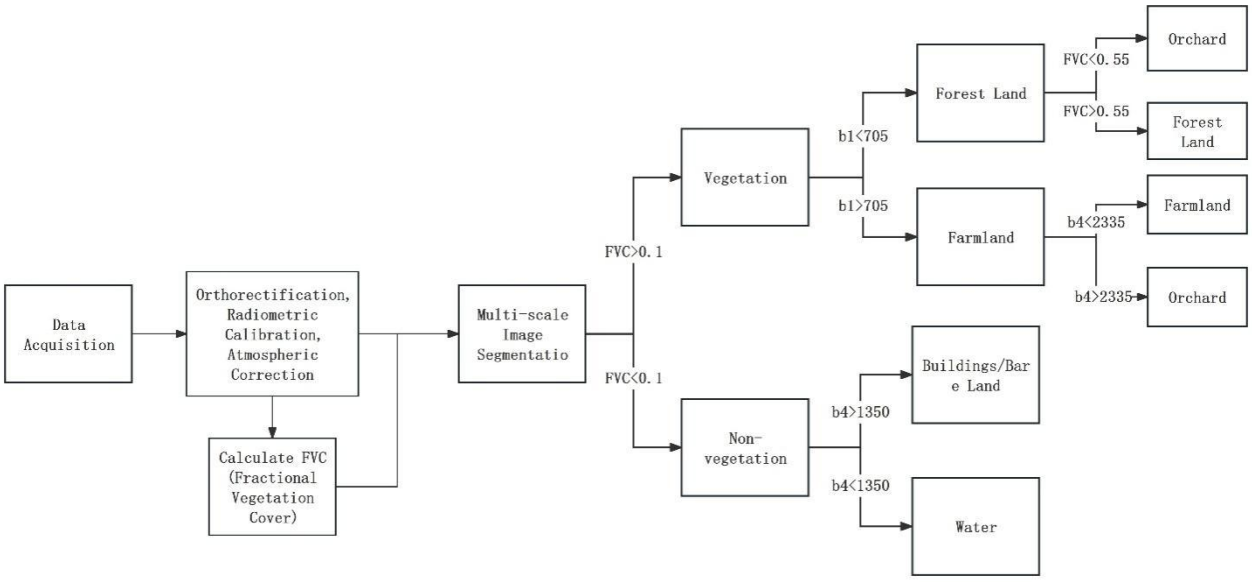

Figure 2 Citrus Planting Extraction Algorithm Combining PSSRN and Decision Tree

3 EXPERIMENTAL RESULTS

3.1 Recognition Result

The experimental setup includes an Intel(R) Xeon(R) E5-2620 v2 processor with 6M cache and 2.10 GHz (6-core/12-thread multitasking), 32GB RAM, an NVIDIA Tesla K20Xm GPU with 6GB VRAM, an 8TB Toshiba hard drive, and a Supermicro X9DRG-QF motherboard. The system runs on Red Hat Enterprise Linux Server release 6.4 x64, CUDA 9, cuDNN 7.2, Python 3.5, and a Keras framework.

In the first step, a pyramid bottleneck residual optimization network was used for citrus extraction. Randomly selected 10% of the labeled data was used as the training set, while the remaining labeled pixels served as the test set. After hyperparameter tuning, the network optimization algorithm employed stochastic gradient descent (SGD) with 200 training epochs. A variable learning rate was used during training, and the input multi-band image patch size was set to 27×27. In this study, we conducted a first-level classification, categorizing land cover types into forest, farmland, water, artificial building (including roads, etc.), bare land classes. To extract citrus planting areas, we added the "orchard" class, resulting in a total of six categories. The final classification results are as follows (Figure 3).

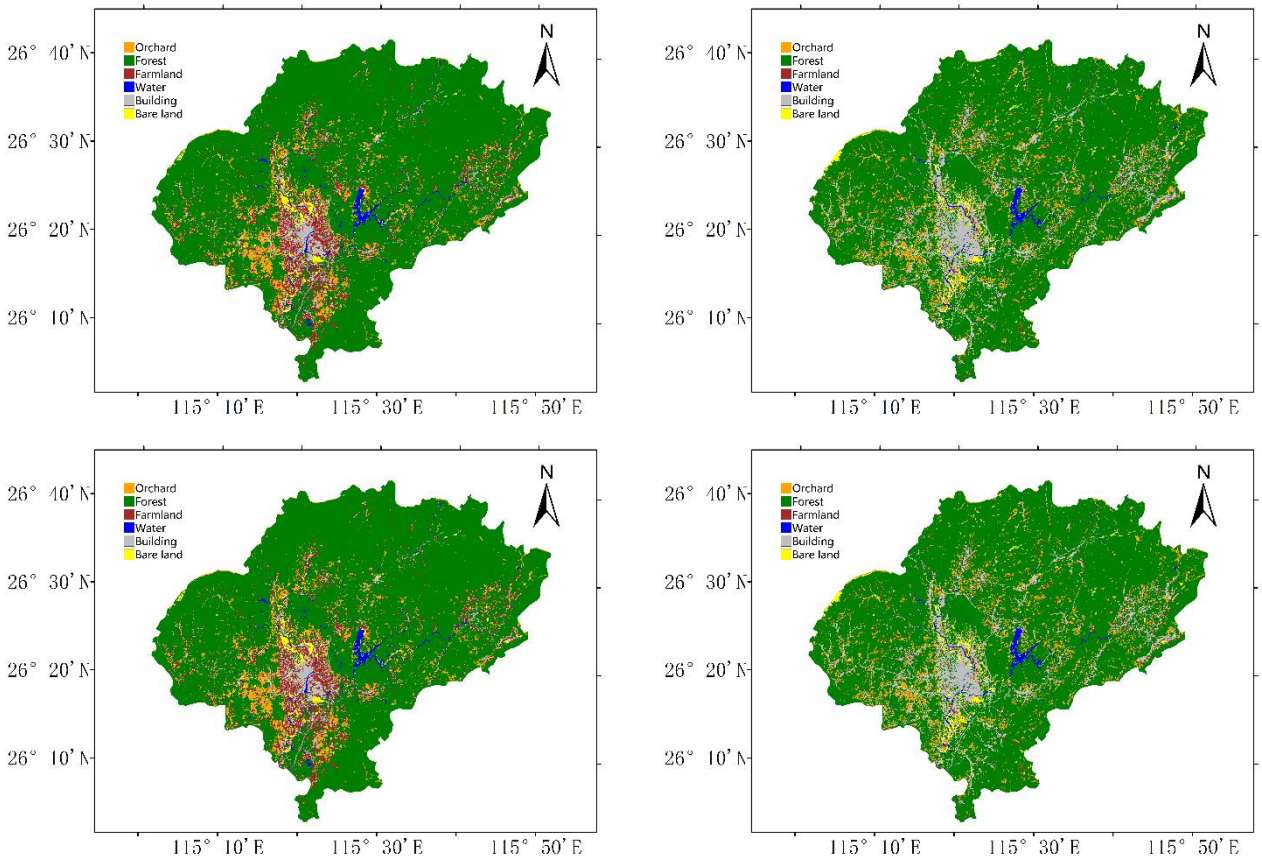

Figure 3 Results of Citrus Planting Areas in Xingguo County (2020–2022) Using GF-1 Satellite Data
Source: It was obtained from the Standard Map Service of the Ministry of Natural Resources of China. (available at: http://bzdt.ch.mnr.gov.cn/)

3.2 Accuracy Calculation

In this study, we adopted accuracy, precision, overall classification accuracy, and the Kappa coefficient as metrics to evaluate classification performance (Table 3).

Table 3 Accuracy and Precision Metrics

TP	The true class is positive, and the model predicts it as positive.
FP	The model predicts positive, but the true class is negative (inconsistent prediction).
FN	The model predicts negative, but the true class is positive (inconsistent prediction).
TN	The true class is negative, and the model predicts it as negative.

Accuracy represents the proportion of correctly predicted samples (TP and TN) among all samples. When the dataset is imbalanced, accuracy may not effectively reflect model performance. High accuracy can coexist with all minority class samples being misclassified; in such cases, other model evaluation metrics should be considered.The accuracy is calculated as follows:

$$\text{Accuracy} = (TP + TN) / \text{all data} \quad (1)$$

Precision for the positive class indicates the proportion of true positive samples among all samples predicted as positive. The precision is calculated as follows:

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Overall accuracy represents the proportion of correctly classified pixels among the total pixels, where the ground truth image or region of interest defines the true classification. Correctly classified pixels are distributed along the diagonal of the confusion matrix, indicating the number of pixels classified into their correct ground truth categories. The total number of pixels equals the sum of pixels across all ground truth categories. Overall accuracy is calculated as:

$$OA = (TP + TN) / (TP + FN + FP + TN) \quad (3)$$

The Kappa coefficient is another method for calculating classification accuracy. It is derived by subtracting the sum of the products of the ground truth pixel counts and the classified pixel counts for each category from the product of the total number of pixels (N) and the sum of the diagonal elements of the confusion matrix (X_{kk}), and then dividing by the difference between the square of the total number of pixels and the sum of the products of the ground truth pixel counts and the classified pixel counts for all categories. The Kappa coefficient is calculated as follows:

$$\text{Kappa}=\frac{N\sum_{i=1}^{r} x_{ii}-\sum_{i=1}^{r}(x_{i+}\times x_{+i})}{N^2-\sum_{i=1}^{r}(x_{i+}\times x_{+i})} \quad (4)$$

The classification results for 2023 were validated through field sampling and image screening, with the accuracy as shown in the following Table 4.

Table 4 Classification Accuracy

class	Accuracy	Precision
Forest	0.9107	0.9082
Farmland	0.7599	0.9156
citrus orchard	0.8847	0.8059
Water	0.9953	0.9738
Building	0.8831	0.9401
Bare land	0.8527	0.7537
OA	0.8839	
Kappa	0.8594	

The statistical results of citrus planting areas from 2020 to 2023 were compared with the official statistical yearbook data, and the results are presented in the following Table 5.

Table 5 Comparison of Extracted Areas (2020–2023) with Statistical Yearbook Data

	Statistical Yearbook Data	Extracted Areas	error
2020	14022 hm^2	16257.03 hm^2	15.94%
2021	14784 hm^2	17610.55 hm^2	19.12%
2022	15274 hm^2	14356.91 hm^2	-6.00%
2023	16173 hm^2	15137.23 hm^2	-6.40%

4 CONCLUSION

This study addresses the challenges of large-scale citrus planting area extraction, such as fragmented plots, uneven distribution of ground objects, and diverse vegetation communities. We propose a method combining a deep pyramid-shaped residual network with a decision tree tailored to citrus planting characteristics, achieving citrus planting area identification in southern Jiangxi Province. The key findings are as follows:

(1) The proposed method effectively utilizes spectral and spatial features of multi-spectral image patches, incorporating citrus planting environmental characteristics, such as vegetation coverage, DEM, and illumination shadows for comprehensive evaluation, achieving a classification accuracy of 88%. However, the presence of a small amount of cloud cover in the imagery slightly impacts classification accuracy.

(2) The extraction accuracy of citrus planting areas in Xingguo County, Ganzhou City (2020–2023) was compared with the final statistical yearbook. The results indicate overestimation in 2020 and 2021, primarily due to the misclassification of grasslands and mountains as citrus. In 2022 and 2023, the error was below 6.5%, with an overall estimation accuracy exceeding 80%. This provides a reference for automated monitoring and planting planning of citrus orchards at the county level.

(3) Although the proposed method demonstrates excellent results, several issues remain to be addressed. Our future work will focus on the following two directions. First, we will test multi-spectral imagery from more diverse regions and types to validate the robustness of the method. Second, we will further integrate SAR imagery, such as Sentinel-1, to mitigate the impact of clouds and rain on optical imagery, compensating for the limitations of optical data.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The study is supported by the project of Jiangxi Provincial Department of Natural Resources ZRKJ20242501.

REFERENCES

[1] Li SY, Zhou XJ, Yin JM et al. Comprehensive division of climate, topography and soil for navel orange planting in Xunwu County. Acta Agriculturae Jiangxiensis, 2007, 19(5): 40–43.
[2] Xu H, Qi S, Gong P, et al. Long-term monitoring of citrus orchard dynamics using time-series Landsat data: a case study in southern China. International Journal of Remote Sensing, 2018, 39(22): 8271-8292.
[3] Chen Y, Hu J, Cai Z, et al. A phenology-based vegetation index for improving ratoon rice mapping using harmonized Landsat and Sentinel-2 data. Journal of Integrative Agriculture, 2024, 23(4): 1164-1178.
[4] Parente L, Taquary E, Silva A P, et al. Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. Remote Sensing, 2019, 11(23): 2881.
[5] Tu A, Xie S, Mo M, et al. Water budget components estimation for a mature citrus orchard of southern China based on HYDRUS-1D model. Agricultural Water Management, 2021, 243: 106426.
[6] Xiao C, Li P, Feng Z, et al. Sentinel-2 red-edge spectral indices (RESI) suitability for mapping rubber boom in Luang Namtha Province, northern Lao PDR. International Journal of Applied Earth Observation and Geoinformation, 2020, 93: 102176.
[7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84-90.
[8] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International journal of computer vision, 2015, 115(3): 211-252.
[9] Zhong Z, Li J, Luo Z, et al. Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. IEEE transactions on geoscience and remote sensing, 2017, 56(2): 847-858.
[10] Paoletti M E, Haut J M, Fernandez-Beltran R, et al. Deep pyramidal residual networks for spectral–spatial hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(2): 740-754.
[11] Han D, Kim J. Deep Pyramidal Residual Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5927-5935.