



Article

Extracting Citrus in Southern China (Guangxi Region) Based on the Improved DeepLabV3+ Network

Hao Li ^{1,2}, Jia Zhang ^{1,2}, Jia Wang ^{1,2,*}, Zhongke Feng ^{1,2} , Boyi Liang ^{1,2} , Nina Xiong ^{1,2}, Junping Zhang ^{1,2}, Xiaoting Sun ^{1,2}, Yibing Li ^{1,2} and Shuqi Lin ^{1,2}

¹ Beijing Key Laboratory of Precision Forestry, Beijing Forestry University, Beijing 100083, China; lihao1012@bjfu.edu.cn (H.L.); zj3210001@bjfu.edu.cn (J.Z.); zhongkefeng@bjfu.edu.cn (Z.F.); liangboyi@bjfu.edu.cn (B.L.); xiongnina@bjfu.edu.cn (N.X.); zjp5901@bjfu.edu.cn (J.Z.); sunxiaoting@bjfu.edu.cn (X.S.); liyibing@bjfu.edu.cn (Y.L.); linshuqi77@bjfu.edu.cn (S.L.)

² Institute of GIS, RS & GPS, Beijing Forestry University, Beijing 100083, China

* Correspondence: wangjia2009@bjfu.edu.cn

Abstract: China is one of the countries with the largest citrus cultivation areas, and its citrus industry has received significant attention due to its substantial economic benefits. Traditional manual forestry surveys and remote sensing image classification tasks are labor-intensive and time-consuming, resulting in low efficiency. Remote sensing technology holds great potential for obtaining spatial information on citrus orchards on a large scale. This study proposes a lightweight model for citrus plantation extraction that combines the DeepLabV3+ model with the convolutional block attention module (CBAM) attention mechanism, with a focus on the phenological growth characteristics of citrus in the Guangxi region. The objective is to address issues such as inaccurate extraction of citrus edges in high-resolution images, misclassification and omissions caused by intra-class differences, as well as the large number of network parameters and long training time found in classical semantic segmentation models. To reduce parameter count and improve training speed, the MobileNetV2 lightweight network is used as a replacement for the Xception backbone network in DeepLabV3+. Additionally, the CBAM is introduced to extract citrus features more accurately and efficiently. Moreover, in consideration of the growth characteristics of citrus, this study augments the feature input with additional channels to better capture and utilize key phenological features of citrus, thereby enhancing the accuracy of citrus recognition. The results demonstrate that the improved DeepLabV3+ model exhibits high reliability in citrus recognition and extraction, achieving an overall accuracy (OA) of 96.23%, a mean pixel accuracy (mPA) of 83.79%, and a mean intersection over union (mIoU) of 85.40%. These metrics represent an improvement of 11.16%, 14.88%, and 14.98%, respectively, compared to the original DeepLabV3+ model. Furthermore, when compared to classical semantic segmentation models, such as UNet and PSPNet, the proposed model achieves higher recognition accuracy. Additionally, the improved DeepLabV3+ model demonstrates a significant reduction in both parameters and training time. Generalization experiments conducted in Nanning, Guangxi Province, further validate the model's strong generalization capabilities. Overall, this study emphasizes extraction accuracy, reduction in parameter count, adherence to timeliness requirements, and facilitation of rapid and accurate extraction of citrus plantation areas, presenting promising application prospects.



Citation: Li, H.; Zhang, J.; Wang, J.; Feng, Z.; Liang, B.; Xiong, N.; Zhang, J.; Sun, X.; Li, Y.; Lin, S. Extracting Citrus in Southern China (Guangxi Region) Based on the Improved DeepLabV3+ Network. *Remote Sens.* **2023**, *15*, 5614. <https://doi.org/10.3390/rs15235614>

Academic Editor: Jochem Verrelst

Received: 31 October 2023

Revised: 29 November 2023

Accepted: 30 November 2023

Published: 3 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: citrus extraction; DeepLabV3+; CBAM; lightweight; Guangxi region

1. Introduction

China is one of the countries with the widest areas of citrus cultivation [1], primarily concentrated in the southern regions, especially Guangxi Province. The citrus industry has emerged as a cornerstone industry for poverty alleviation and rural revitalization in the region, attracting considerable attention due to its substantial economic benefits.

Accurate acquisition of planting areas and spatial distribution information of citrus orchards holds critical importance in guiding the scientific development of the citrus industry. Remote sensing technology offers tremendous potential for large-scale acquisition of spatial information for citrus orchards. Threshold-based and edge detection methods are well-established traditional approaches to image segmentation. In threshold segmentation [2], the grayscale levels of an image are divided based on one or multiple thresholds, assigning pixels with similar grayscale values to the same class according to specific rules. The selection of appropriate threshold values significantly influences the effectiveness of the segmentation algorithm, underscoring the importance of threshold selection. In 2021, Houssein et al. introduced a threshold image segmentation method based on the black widow optimization algorithm [3], utilizing maximum entropy thresholding and Otsu's method to determine the optimal threshold for the image. Although numerous effective threshold segmentation algorithms are in use, traditional methods may struggle to meet the requirements of more detailed and precise segmentation, making multidimensional threshold segmentation more suitable for efficient targeting of images. In 2019, Qin et al. proposed a multi-level thresholding method for images based on subspace elimination optimization [4]. Zhao et al. introduced a chaos-randomized ant colony optimization approach, employing two-dimensional maximum entropy for multi-threshold image segmentation [5]. In 2020, Di Martino et al. employed the particle swarm optimization method (PSO) for multi-level threshold extraction of compressed images during threshold segmentation through fuzzy transformations [6]. However, as the number of thresholds for image feature acquisition increases, the time complexity grows exponentially. Therefore, Eusuff et al. proposed a novel image segmentation method based on the shuffled frog leaping algorithm (SFLA) [7], demonstrating significant performance in breast cancer segmentation. In 2022, Chen et al. integrated the SFLA algorithm with maximum entropy thresholding for multi-threshold image segmentation [8], resulting in improved time complexity. Edges, as fundamental image features, reflect variations in grayscale values among different regions and expose abrupt changes in image characteristics along these boundaries, thus serving as a basis for image segmentation. One commonly employed method in practical research for edge detection and segmentation is the parallel edge detection technique that harnesses differential operators. In 2020, Chen et al. devised a remote sensing image road recognition algorithm by integrating recursive operators and wavelet transform [9]. In 2021, Xu et al. segmented remote sensing images captured by drones and employed traditional edge operators to determine suitable thresholds for segmentation [10], thus demonstrating the applicability of conventional edge operators in high-definition image segmentation tasks. While traditional operators can achieve satisfactory edge detection speed, they may fall short in distinguishing multiple categories of edge features within complex scenes. In 2021, Chetia et al. proposed an edge detection algorithm that enhanced the quantum representation of the Sobel operator [11], leveraging non-maximum suppression and multi-threshold methods. Jan designed a novel edge detection algorithm (CLOG) based on the Canny and LoG operators [12]. In 2023, Roy et al. introduced an unsupervised edge detection method grounded on local standard deviation [13], attaining precise detection of cell nuclei during the segmentation of stained tissue pathology images. These improved algorithms have all demonstrated enhanced detection performance.

Although traditional methods are capable of segmenting the geometric or color features of images, they exhibit low accuracy in segmentation and poor resilience to noise. Furthermore, they are only suitable for relatively simple images and perform inadequately when confronted with complex imagery, such as remote sensing images. Consequently, machine learning methods grounded in mathematical statistics have gained popularity. In 2022, Ali et al. introduced a novel method based on data-driven principles [14], offering automatic determination of the optimal number of clusters (k). They applied this method to satellite images of the Islamabad region captured by the Sentinel-2B satellite. This technique provided concrete evidence for urban and forest planning and demonstrated promising time complexity when applied to large satellite image datasets. Furthermore, in 2020,

Mahata et al. developed an unsupervised segmentation model that combined K-means and cellular automata algorithms for land cover segmentation in satellite images [15]. A decision tree is a supervised classification algorithm that partitions data into multiple subgroups to construct a classifier. The partitioning strategy is based on the criterion that yields the largest heterogeneity. In 2020, Pastorino et al. introduced a novel approach that combines Markov models with decision algorithms for processing multi-resolution remote sensing images [16]. Experimental results demonstrated the superior segmentation effectiveness of this new method compared to traditional approaches. Random Forest (RF) is an ensemble of decision trees that are randomly constructed during training for the purpose of learning [17]. By reducing the sample size compared to individual decision trees, Random Forest effectively decreases the risk of overfitting. In 2019, Dong et al. proposed a method that integrates Random Forest with Convolutional Neural Networks [18]. Experimental results demonstrated the accurate segmentation of bamboo forests from other subtropical vegetation in high-resolution remote sensing images using this method. The SVM (Support Vector Machine) is a prominent supervised machine learning classifier widely employed in image recognition [19]. It establishes a discriminative hyperplane to identify the optimal classification boundary that maximizes the margin between data points of different classes in a higher-dimensional space, thus capturing more distinctive features within the high-dimensional feature space. In 2021, Razaque et al. introduced an enhanced SVM approach combining the Radial Basis Function (RBF) and linear SVM for land classification in remote sensing imagery [20]. Empirical results validated the efficacy of this improved SVM method in remote sensing classification. However, when handling high-resolution remote sensing images, the SVM algorithms focusing solely on Euclidean distance might disregard the global distribution of samples.

In 2012, the emergence of AlexNet popularized Convolutional Neural Networks (CNNs) and other deep learning methods within the academic community [21]. Researchers have also introduced this technology into the field of image segmentation to address the limitations of traditional segmentation methods. CNNs can accurately extract surface information, including buildings, vegetation, waterways, and roads, from remote sensing images. Several studies have demonstrated significant achievements of CNNs in remote sensing image segmentation [22]. Yang et al. successfully employed CNNs to extract mature rice fields and automatically estimate rice yield [23], while Su et al. improved the CNN method for identifying rice in agricultural remote sensing imagery [24]. As one of the pioneering deep learning networks applied in image segmentation, Fully Convolutional Neural Networks (FCNs) have the capacity to accomplish end-to-end segmentation while overcoming the limitations associated with input image sizes [25,26]. FCNs have demonstrated promising performance in handling larger remote sensing datasets and complex semantic segmentation scenarios [27]. However, the series of convolutional and pooling operations employed by FCNs may potentially degrade the resolution of feature maps, resulting in a loss of spatial information. With regard to capturing comprehensive contextual information of images, FCNs may struggle to effectively leverage global information and certain local features, thereby leading to diminished performance in extracting small targets. In order to address the limitation of FCNs in acquiring comprehensive global information, Chen et al. introduced the DeepLab model [28,29]. This model enhances the accuracy of pixel positioning and expands the receptive field through the utilization of dilated convolutions and conditional random fields. Building upon the DeepLab framework, Chen et al. subsequently proposed DeepLabV2, DeepLabV3, and DeepLabV3+, progressively improving the algorithm's segmentation performance. Wang et al. utilized DeepLabV3+ to perform segmentation on forest remote sensing images [30], successfully achieving accurate segmentation in forest fire scenarios. Wang et al. improved the overall segmentation accuracy by incorporating an attention mechanism into DeepLabV3+ [31]. Additionally, Du et al. achieved high precision by combining DeepLabV3+ with object-based image analysis methods for annotating remote sensing images [32]. In addition to the DeepLab series of networks, several other deep learning models have shown promising

performance in remote sensing image segmentation. One such model is UNet, which follows an encoder-decoder architecture for image semantic segmentation. UNet introduces a unique U-shaped structure, where the initial half serves as the feature extraction component, while the latter half employs upsampling on deep features to restore unutilized spatial feature information. Gu et al. combined the strengths of Transformer and CNNs, proposing an adaptive enhanced Swin Transformer known as AESwin-UNet for remote sensing segmentation [33]. This model exhibits commendable performance in semantic segmentation tasks.

In addition to the commonly used methods mentioned above, there are specific models for remote sensing image segmentation that are designed for particular tasks. In their study conducted in 2022, Hou et al. found that most deep learning networks prioritize capturing comprehensive contextual information [34], which can lead to the loss of edge features in remote sensing images. To address this issue, the authors proposed the Boundary Sensitive Network (BSNet), which integrates the Dynamic Hybrid Gradient Convolution (DHGC) and the CSA attention mechanism. Experimental results on the Vaihingen, Potsdam, and iSAID remote sensing datasets demonstrated that BSNet produces clearer boundaries. Likewise, in 2020, Li et al. developed a post-processing CNN model (PP-CNN) to capture spatial information in winter wheat remote sensing images [35]. This approach involves statistical analysis to acquire prior knowledge of classes, thereby improving classification accuracy. The method achieved a 94.4% accuracy on a high-resolution remote sensing dataset in the Feicheng winter wheat production area of Shandong. Overall, recent studies on remote sensing image segmentation methods primarily focus on deep learning approaches based on convolutional neural networks. These approaches explore various aspects such as encoder-decoder structures, feature extraction, and attention mechanisms to enhance the segmentation performance of the models.

Despite the fact that multispectral remote sensing photos provide more spatial information and adequate feature scales, most current studies focus on hyperspectral remote sensing images. On the contrary, there have been few investigations into feature information extraction from multispectral pictures. In this respect, this study introduces the DeepLabV3+ network, which combines the attention mechanism module and the citrus feature channel to improve citrus recognition accuracy. The experimental findings demonstrate that the model developed in this paper can meet the accuracy requirements of feature information extraction and perform better for citrus classification. This study provides significant support for the use of multispectral images in citrus identification, as well as novel ideas for improving the accuracy and efficiency of remote sensing image processing. It also allows for regional precision agriculture and for more scientific policy implementation.

2. Materials and Methods

2.1. Study Area and Data

2.1.1. Study Area

The research was conducted in the Yangshuo County, located in Guilin City, Guangxi Province, China (24°28'N~25°4'N, 110°13'E~110°40'E), as indicated in Figure 1. Yangshuo County is situated in the northeastern part of Guangxi Province, adjacent to the urban area of Guilin [36]. It has an approximate area of 1436 km². The terrain in the study area is primarily characterized by karst hills and uneven topography. There are numerous rivers in the area, predominantly flowing from the northwest to the southeast, with elevations ranging from 200 to 500 m and a relative height difference of 50 to 300 m. Yangshuo County falls under the subtropical monsoon climate of Central Asia, exhibiting warm and sunny weather with relatively abundant rainfall. These climatic characteristics contribute to a moderate temperature in the area, which is favorable for crop growth and abundant vegetation.

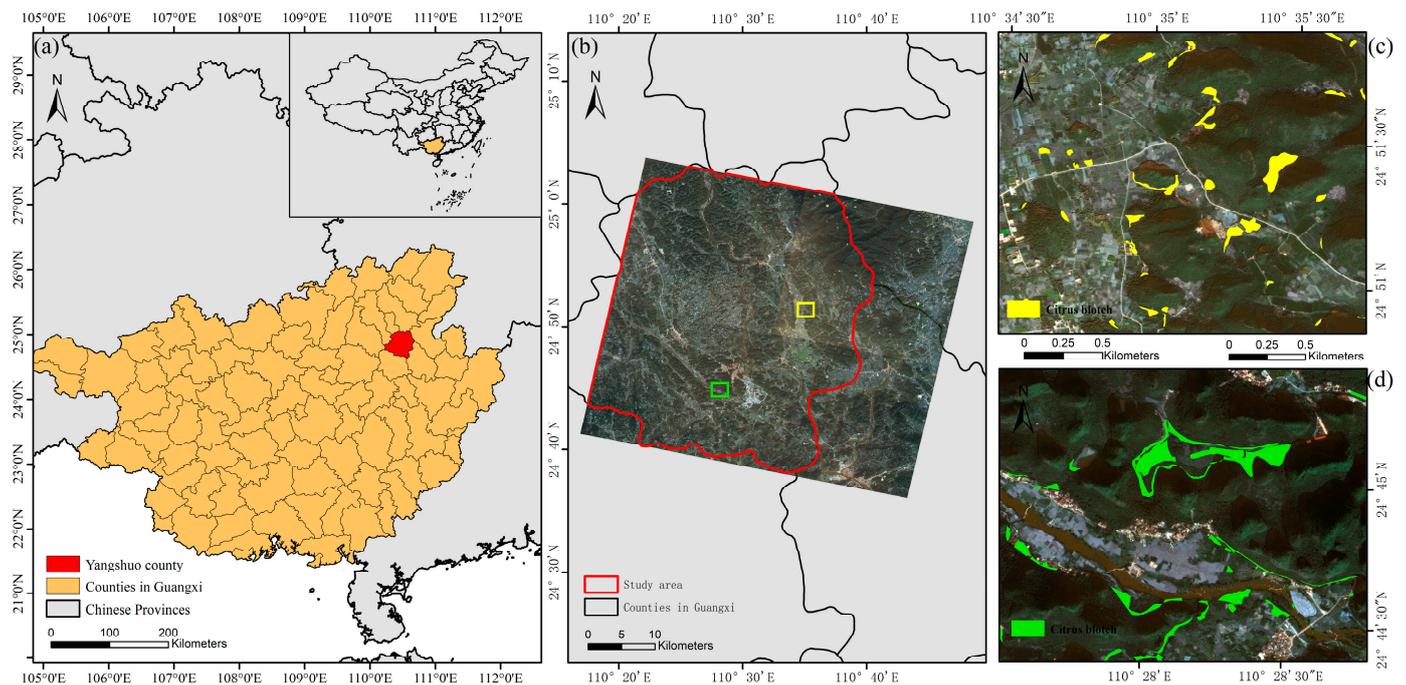


Figure 1. Study area. (a) Geographic location of the study area. (b) Main study area, i.e., Yangshuo County, Guangxi Province. (c,d) show the labeled areas of citrus samples (marked by yellow and green blocks). The images used are GF-2 images with pseudo-color components (R = near-infrared, G = red, B = green).

2.1.2. Data

This study utilized Gaofen-2 (GF-2) satellite imagery as the data source for remote sensing, aligning with the requirement for high spatial resolution. The parameters of GF-2 satellite data are provided in Table 1 [37]. GF-2 imagery consists of a panchromatic band and four multispectral bands: red, green, blue, and near-infrared. The panchromatic band offers a spatial resolution of 1 m, delivering precise image information. The multispectral bands exhibit a spatial resolution of 4 m, maintaining a relatively high resolution suitable for multispectral information extraction. Four cloud-free GF-2 images captured on 28 October 2021 were carefully collected and downloaded within the study area. This particular time period was selected due to the substantial growth and yellowing of citrus fruits. As a result, the canopy vegetation information weakens, therefore causing a decline in NDVI values [38].

Table 1. Parameters of GF-2 satellite data.

Parameters	Multispectral	Panchromatic
Spectral range	0.45~0.52 μm	0.45~0.90 μm
	0.52~0.59 μm	
	0.63~0.69 μm	
	0.77~0.89 μm	
Spatial resolution	4 m	1 m
width		
Side-swing capability		45 km
Revisit period		$\pm 35^\circ$
Coverage period		5 days
Orbital altitude		69 days
		631 km

The GF-2 image data from the study area underwent preprocessing using ENVI software (version 5.3.1), an image visualization environment. The preprocessing workflow for the multispectral images involved radiometric calibration, atmospheric correction, and

orthorectification. On the other hand, the panchromatic image underwent radiometric calibration and orthorectification. Radiometric calibration facilitates the mapping of digital values in remote sensing images to actual radiometric measurements, establishing a precise relationship between pixel values and ground reflectance or radiance [39]. Atmospheric correction is employed to remove the effects of atmospheric scattering and absorption during data transmission, thereby restoring the accurate reflectance information from the Earth's surface. Orthorectification corrects image distortions caused by variations in terrain, ensuring pixel alignment with their corresponding ground locations. Finally, the NNDiffuse pan-sharpening fusion method was applied to merge the panchromatic and multispectral images [40], resulting in a multispectral remote sensing image with a spatial resolution of 1 m. These preprocessing steps play a critical role in ensuring the provision of accurate and reliable information for subsequent analyses, particularly in precise image classification and feature extraction.

2.2. Methods

2.2.1. Improved DeepLabV3+ Network Modeling

During the process of visual interpretation, the identification of citrus orchards on satellite remote sensing images poses several challenges. These challenges arise from the difficulty of distinguishing citrus orchards from other land objects, primarily due to the presence of complex semantic information and rich detailed content in the images. To overcome these challenges, it is essential for the segmentation network to possess outstanding capabilities in extracting fine-grained details. In conventional segmentation networks, such as the original DeepLabV3+ network, they are commonly employed for handling complex and diverse datasets that involve various object features and large amounts of data. The backbone of these networks tends to be intricate, enabling them to learn more intricate patterns of features. However, for citrus recognition tasks, these networks may exhibit unnecessary drawbacks, including substantial computational costs and training difficulties. To address these issues, our study incorporates the lightweight MobileNetV2 backbone network into the encoder-decoder structure of DeepLabV3+ as the encoder component for feature extraction. The MobileNetV2 network possesses relatively fewer parameters, introduces more direct connections, is easier to train, and converges faster than the Xception structure. Furthermore, we introduce dilated convolutions to expand the receptive field of convolutional layers, leading to improved capture of the semantic information in the images and significantly enhanced computational speed. As a result, MobileNetV2 proves to be a suitable model for efficiently identifying citrus in remote sensing images. Additionally, we augment the feature input by increasing the number of channels based on the growth characteristics of citrus. This adjustment is motivated by the fact that during the rapid fruit swelling period in October, the normalized difference vegetation index (NDVI) values of citrus significantly decrease compared to other crops. By incorporating additional channels in the feature input, our model gains the ability to better capture and utilize this crucial feature, consequently enhancing the accuracy of citrus recognition. Such customization targeting citrus-specific features improves the model's adaptability and performance in citrus growth and characteristic analysis, thereby strengthening its capacity for recognizing citrus orchards. The introduction of the channel attention module (CBAM) plays a pivotal role in our model as it enables a more focused extraction of key information and accurate feature representation of citrus [41]. This module enhances the model's grasp of detailed information within the images while also capturing the spatial correlation of pixels in larger areas, facilitating more precise recognition of citrus within different regions. Consequently, this improvement greatly enhances the performance of our model in identifying citrus orchards of varying sizes, enabling it to comprehensively cover a wide range of citrus orchard types and improve its overall performance and applicability. Figure 2 illustrates the network model diagram utilized in our study.

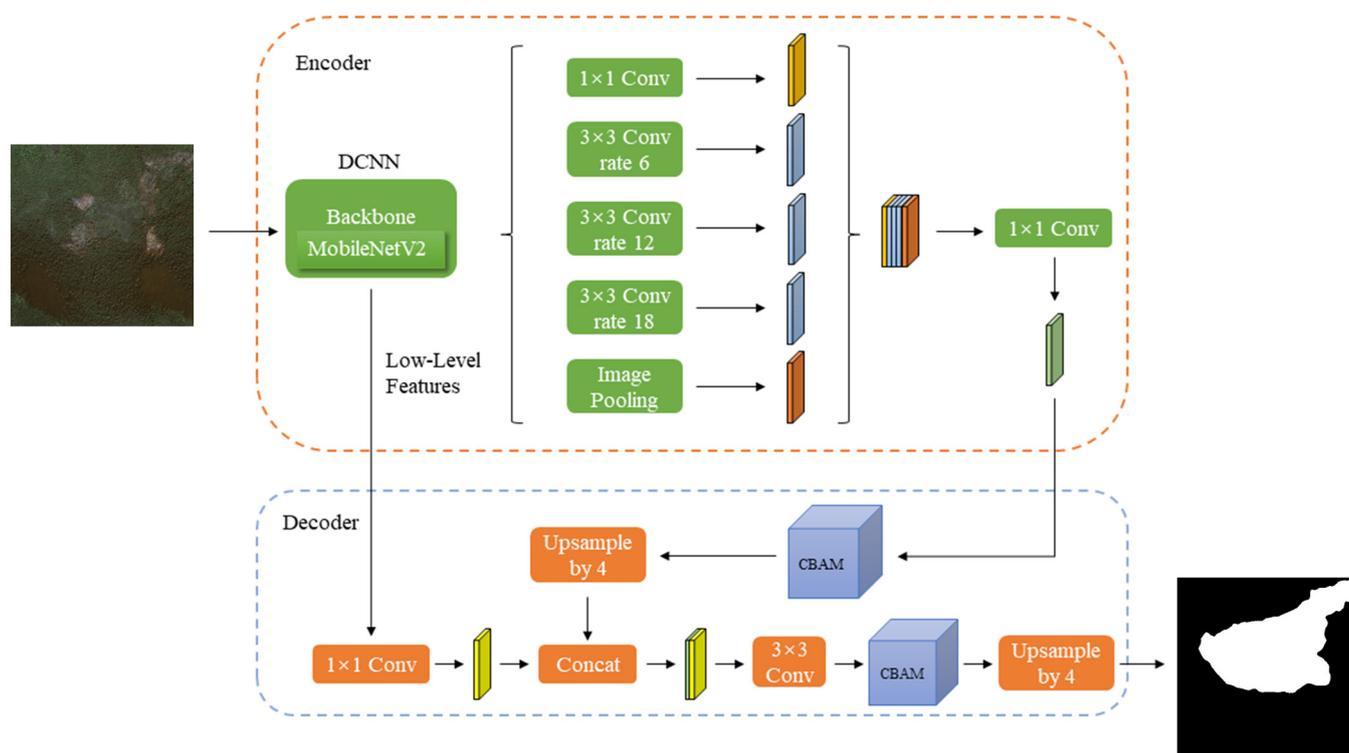


Figure 2. Structure of improved DeepLabV3+ model.

2.2.2. Backbone Feature Extraction Network

MobileNet is a lightweight deep neural network that was introduced by Google in 2016 to cater to embedded devices like mobile phones. MobileNetV2 is an upgraded version of MobileNetV1 that maintains simplicity, eliminates the need for special operators, and improves accuracy [42]. The key enhancement in MobileNetV2 is the introduction of a new activation function called ReLU6 [43], which restricts the maximum output value to 6. The purpose behind this design is to ensure high numerical resolution even in scenarios with low precision. In traditional network architectures, convolutions with ReLU activation functions are commonly used for normalization during the process of feature extraction. However, employing ReLU activation functions in low-dimensional spaces can result in the loss of valuable information. In order to overcome this challenge, MobileNetV2 adopts a linear bottleneck structure where the ReLU activation function is replaced by a linear function. This substitution minimizes the loss of crucial network information. Consequently, this innovation enhances the efficiency and accuracy of the network, especially in resource-limited environments and on embedded devices. As a result, MobileNetV2 is able to adapt better to various application scenarios and achieve exceptional performance in lightweight deep learning tasks. The structural parameters of MobileNetV2 utilized in this experiment are presented in Table 2. In the table, t denotes the expansion factor, c indicates the depth of the output feature matrix, n signifies the number of internal widening layers in the bottleneck structure, and s refers to the stride [44].

Table 2. MobileNetV2 structure parameter table.

Input	Operator	t	c	n	s
$224^2 \times 3$	Conv2d	–	32	1	2
$112^2 \times 32$	Bottleneck	1	16	1	1
$112^2 \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$14^2 \times 96$	Bottleneck	6	96	3	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv2d 1×1	–	1280	1	1
$7^2 \times 1280$	Avgpool 7×7	–	–	1	–
$1^2 \times 1 \times k$	Conv2d 1×1	–	k	–	–

2.2.3. Convolutional Block Attention Module

The application of attention mechanisms in image processing is aimed at acquiring the contextual information present in the image in order to capture relevancy and help the model prioritize important regions while disregarding irrelevant information. These mechanisms include the channel attention module and the spatial attention module, which, respectively, emphasize the significance of feature channels and spatial regions within the image. The channel attention module, known as CAM, aids in identifying feature channels critical for a specific task by analyzing the relationships between different channels and optimizing the allocation of feature maps, thereby enhancing model performance. On the other hand, the spatial attention module, referred to as SAM, focuses on determining the importance of pixel regions within the image. This module facilitates a better understanding of local regions, particularly for accurately extracting edge features. The CBAM introduces these two attention mechanisms and employs a sequential structure of channel and spatial attention, considering the analysis dimensions of channels and spatial scopes. This enables neural networks to process image features meticulously, while attending to information at various scales. Furthermore, the CBAM is known for its lightweight nature and seamless integration into different neural networks, thereby enhancing the versatility and performance of the models. In conclusion, the utilization of attention mechanisms allows the model to concentrate on crucial information, leading to more precise and efficient extraction of edge features, thereby enhancing model performance and improving the understanding of the image.

The CBAM structure [43], illustrated in Figure 3, comprises the channel attention module and the spatial attention module. Within the channel attention module, pooling is applied to the input feature map to acquire weight information for each channel. Subsequently, this weight information is propagated to the spatial attention module. In turn, the spatial attention module employs both maximum and average pooling operations on the feature values of each specific point throughout all channels on the input feature map. These operations effectively capture features at various scales. The weights for each feature point on the input feature map are then derived using the identical procedures as those employed within the channel attention module. Lastly, the obtained weights are employed to perform weighted convolutions on the original input feature map, leading to the generation of deep features that integrate multi-scale contextual information. In summary, the CBAM structure, through its channel attention and spatial attention modules, enables multi-scale feature extraction on the input feature map. This enhancement results in a refined comprehension of image content and improved analytical and recognition capabilities for the model.

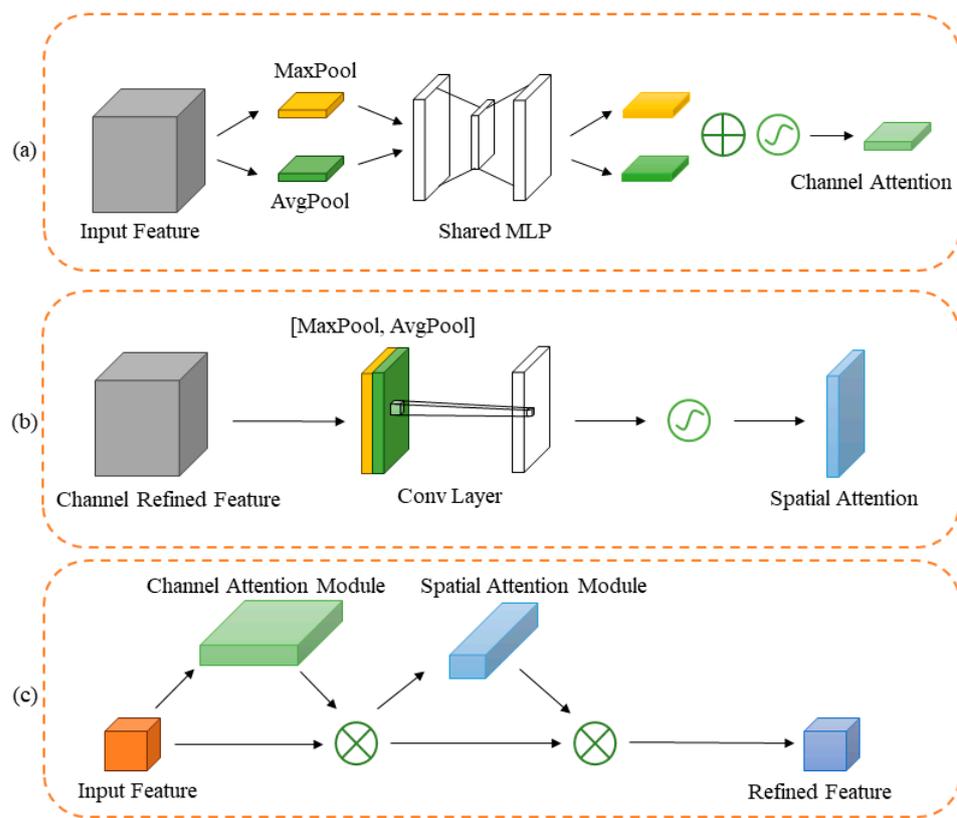


Figure 3. Structure of CBAM: (a) Channel attention module; (b) Spatial attention module; (c) CBAM.

2.2.4. Evaluation Metrics

To assess the performance of the enhanced citrus recognition model, multiple validation metrics were utilized in this study, namely overall accuracy (OA), recall intersection over union (IoU), F1-score, mean intersection over union (mIoU), and mean pixel accuracy (mPA). These metrics are derived from accuracy measurement techniques that rely on the confusion matrix. The confusion matrix tabulates pixel values in the images according to the true and predicted classes, providing insights into the model's performance on a given dataset. The rows and columns of the matrix represent the true values and predicted values, respectively (Table 3).

Table 3. Confusion matrix (TP represents the pixels in which citrus was correctly identified; FP represents pixels where citrus was incorrectly identified as non-citrus; TN represents the pixels where non-citrus was correctly identified; FN represents pixels where non-citrus was incorrectly identified as citrus).

Confusion Matrix	Citrus	Non-Citrus
Citrus	TP	FN
Non-Citrus	FP	TN

The calculation formulas for each parameter and intermediate variable are as follows: OA represents the overall accuracy of predicting citrus. Recall represents the proportion of correctly predicted citrus pixels to all actual citrus pixels. Precision represents the proportion of actual citrus pixels within the pixels predicted as citrus by the model. F1-score is the harmonic mean of precision and recall. IoU (intersection over union) is the ratio of the intersection area between the predicted region and the actual region to the union

area. mIoU (mean intersection over union) is the average IoU for all classes. mPA (mean pixel accuracy) is the average proportion of correctly classified pixels among all classes.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (5)$$

$$mIoU = \frac{1}{N + 1} \sum_{i=0}^N IoU \quad (6)$$

$$mPA = \frac{1}{k + 1} \sum_{i=0}^k \frac{TP}{TP + FN} \quad (7)$$

2.2.5. Dataset Production

High-resolution remote sensing imagery from the Gaofen-2 (GF-2) satellite was utilized in this study to analyze the research area. The image data was combined with ground survey data and manual visual interpretation to create a dataset for training deep learning models. The annotation process involved labeling citrus orchard areas as “citrus” and other land cover types as “background” using ArcGIS software (version 10.8). The annotated images were then cropped into patches of size 512×512 , resulting in a total of 2403 training samples. The dataset was randomly divided into an 80:20 ratio for training and validation, respectively, to ensure enough data for training and assessing recognition accuracy. To combat the issues of overfitting and limited training sample size, data augmentation techniques were implemented to enhance the model’s generalization capability. These steps aimed to provide an adequate number of training samples to accurately identify and classify citrus using deep learning models. This process contributes to the model’s robustness and performance in diverse circumstances.

2.2.6. Experimental Setting

The experiments in this study were carried out on the Windows 11 operating system. The computer utilized an NVIDIA GeForce RTX2060Super GPU with 8G memory, running on CUDA version 12.2. The deep learning network was constructed using PyTorch version 1.12.1. The software environment was established using Anaconda (Python 3.9). The initial learning rate of the model was set to 0.01, with the learning rate decayed using cosine annealing. The SGD optimizer was employed with a batch size of five, and a total of 300 epochs were executed.

3. Results and Analyses

3.1. Model Training Results

To validate the effectiveness of the improved DeepLabV3+ model in citrus recognition, this study compared it with three classical semantic segmentation models (DeepLabV3+, UNet, and PSPNet) while keeping other training parameters consistent. The results demonstrated that the improved DeepLabV3+ model achieved the highest recognition accuracy among the four models. Table 4 presents the accuracy results of the improved DeepLabV3+ model and other classical semantic segmentation models in citrus recognition tasks. From Table 4, it can be observed that the overall accuracy (OA) of all models exceeded 80%, indicating the strong performance of deep learning in citrus recognition. Considering

the six evaluation metrics collectively, the improved DeepLabV3+ model significantly outperformed the other three models in terms of citrus recognition accuracy. As shown in Figure 4, the improved DeepLabV3+ model achieved an overall accuracy (OA) of 96.23%, a mean intersection over union (mIoU) of 83.79%, and a mean pixel accuracy (mPA) of 85.40%. Compared to the original model, these figures improved by 11.16%, 14.88%, and 14.98% respectively. Compared to the UNet model, they improved by 15.81%, 18.21%, and 17.07% respectively. Compared to PSPNet, they improved by 14.91%, 19.56%, and 16.34%, respectively. These results indicate that the improved DeepLabV3+ model excels in citrus extraction tasks in terms of performance.

Table 4. Comparison of extraction accuracy of various models for citrus.

Models	IoU	Recall	OA	F1-Score	mIoU	mPA
Improved DeepLabV3+	0.8078	0.8894	0.9623	0.9583	0.8379	0.8540
DeepLabV3+	0.7046	0.8125	0.8507	0.8478	0.6891	0.7042
UNet	0.6839	0.7923	0.8042	0.8312	0.6558	0.6833
PSPNet	0.6902	0.8087	0.8132	0.8377	0.6423	0.6906

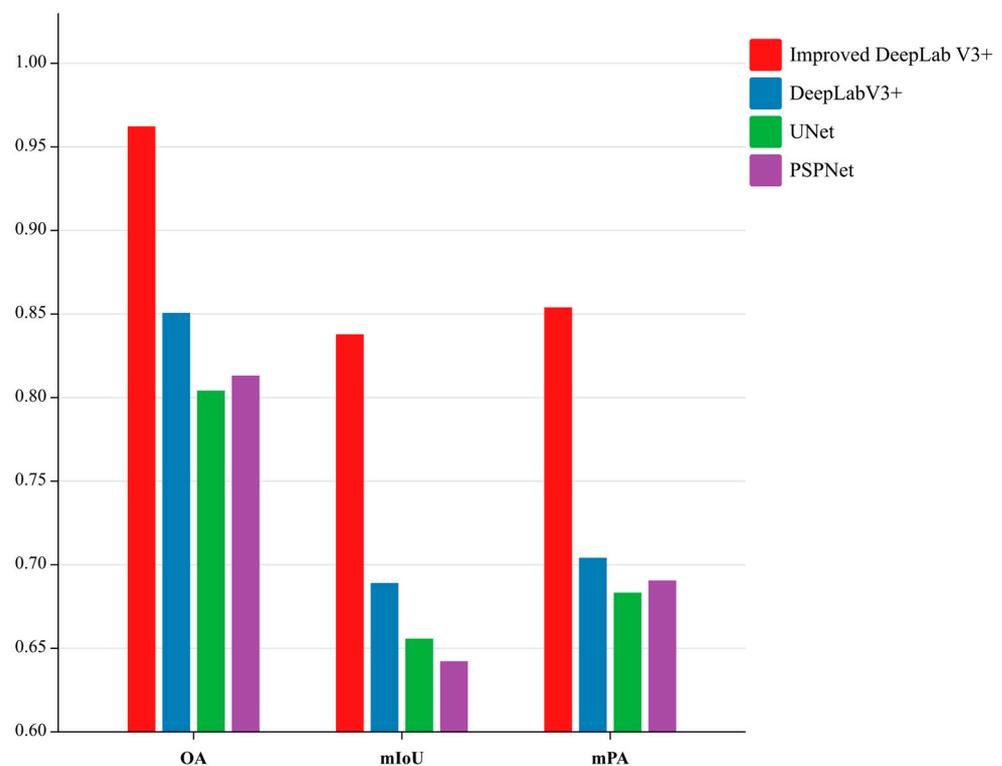


Figure 4. Comparison of extraction accuracy of various models for citrus.

From Figure 5, it is evident that all four network models successfully identified and extracted large-scale citrus orchards. However, the improved DeepLabV3+ model demonstrated noticeable superiority in citrus recognition. The original DeepLabV3+ model, UNet, and PSPNet yielded relatively poor results, with evident misclassification and omission problems. These models produced predicted images with fragmented misclassified regions, erroneously identifying spectral information resembling citrus vegetation as actual citrus. Moreover, they struggled to effectively handle scattered areas of citrus vegetation. In terms of citrus recognition, the improved DeepLabV3+ model outperformed the other models significantly. This improvement involved replacing the backbone network for feature extraction and integrating the CBAM to enhance the model's attention to crucial features. Consequently, the occurrence of noise reduced noticeably in the results, the integrity of vegetation areas improved, and overall recognition performance was enhanced. The improved

DeepLabV3+ model achieved more accurate extraction of edge information from citrus vegetation, thereby alleviating misclassification and omission issues. These results indicate that through enhancing the backbone network and integrating the CBAM, the improved DeepLabV3+ model exhibits superior performance in citrus recognition tasks. It enhances its ability to handle scattered citrus vegetation, consequently reducing misclassification and omission problems and improving overall recognition effectiveness.

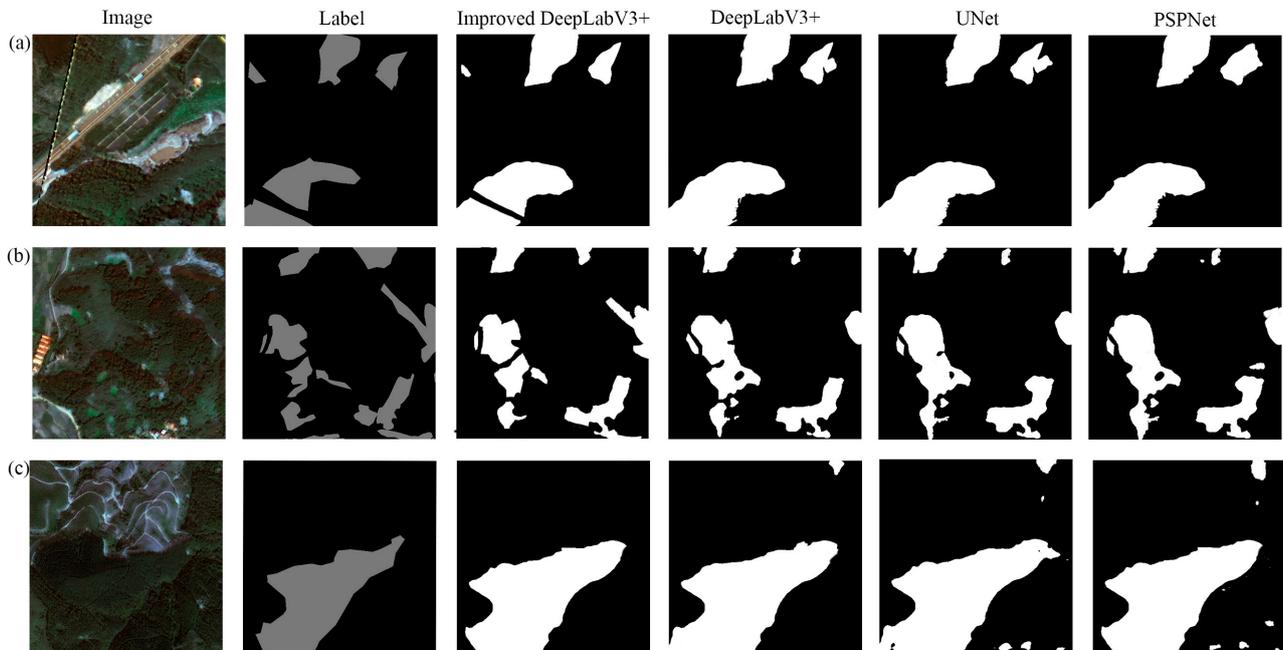


Figure 5. Citrus extraction results using four different models, where the black area is the background area, the gray is the citrus sample labeled area, and the white is the citrus area extracted by the models. Among the three special plots selected, plot (a) contains roads and water, plot (b) contains complex and fragmentary citrus planting areas, and plot (c) contains concentrated citrus planting areas.

3.2. Ablation Experiment

To assess the efficacy of incorporating attention mechanism modules into both the backbone feature extraction network and the decoder component, four distinct experimental approaches were formulated and subsequently compared in this study. The findings of these experiments are presented in Table 5.

Table 5. Ablation experiment results.

Scheme	OA (%)	mIoU (%)	mPA (%)	Training Time (h)
1	85.07	68.91	70.42	9.23
2	88.39	71.33	72.10	4.51
3	92.34	78.57	79.83	4.69
4	96.23	83.79	85.40	4.75

Scheme 1 utilizes the traditional DeepLabV3+ network architecture with Xception as the backbone network. In Scheme 2, MobileNetV2 replaces the backbone network from Scheme 1. Scheme 3 further extends Scheme 2 by incorporating the CBAM into the decoder part. Lastly, Scheme 4 enhances Scheme 2 by integrating an attention mechanism into the decoder part of the backbone feature extraction network.

The results of the ablation experiments highlight the advantages of Scheme 2's model over the traditional DeepLabV3+ model, with fewer parameters and reduced training time, yet still achieving commendable performance in citrus extraction tasks. These findings support the rationale for replacing the backbone feature extraction network, thereby reducing

model complexity. Comparatively, Scheme 3 and Scheme 4 yield improved performance, exhibiting higher overall accuracy (OA) and mean intersection over union (mIoU). Notably, Scheme 4 demonstrates the most significant enhancement, emphasizing the effectiveness of integrating attention mechanisms in both the encoder and decoder for enhancing citrus recognition accuracy. Consequently, we can conclude that by employing the lightweight MobileNetV2 network to replace the backbone network and by incorporating attention mechanisms in the encoder and decoder, the performance of citrus recognition tasks can be substantially improved.

3.3. Migrability of the Segmentation Model

The majority of validation data used for remote sensing image segmentation originate from the same study area as the training data, which does not effectively demonstrate the robustness of the segmentation model. Although numerous segmentation models achieve high accuracy on their validation data, their performance diminishes when applied to diverse regions. This decline is primarily attributed to insufficient training samples or limited model generalization capabilities. In order to validate the transferability of the enhanced model, tests were conducted in Nanning City, located in Guangxi Province. The experiment also used the GF-2 satellite image as the remote sensing data source on 3 October 2021, and a consistent preprocessing process was applied to the image. The outcomes of these tests are presented in Figure 6. The improved model demonstrates a comparable level of accuracy in citrus extraction within Nanning City, indicating its considerable transferability and suggesting its potential for broader application and promotion.

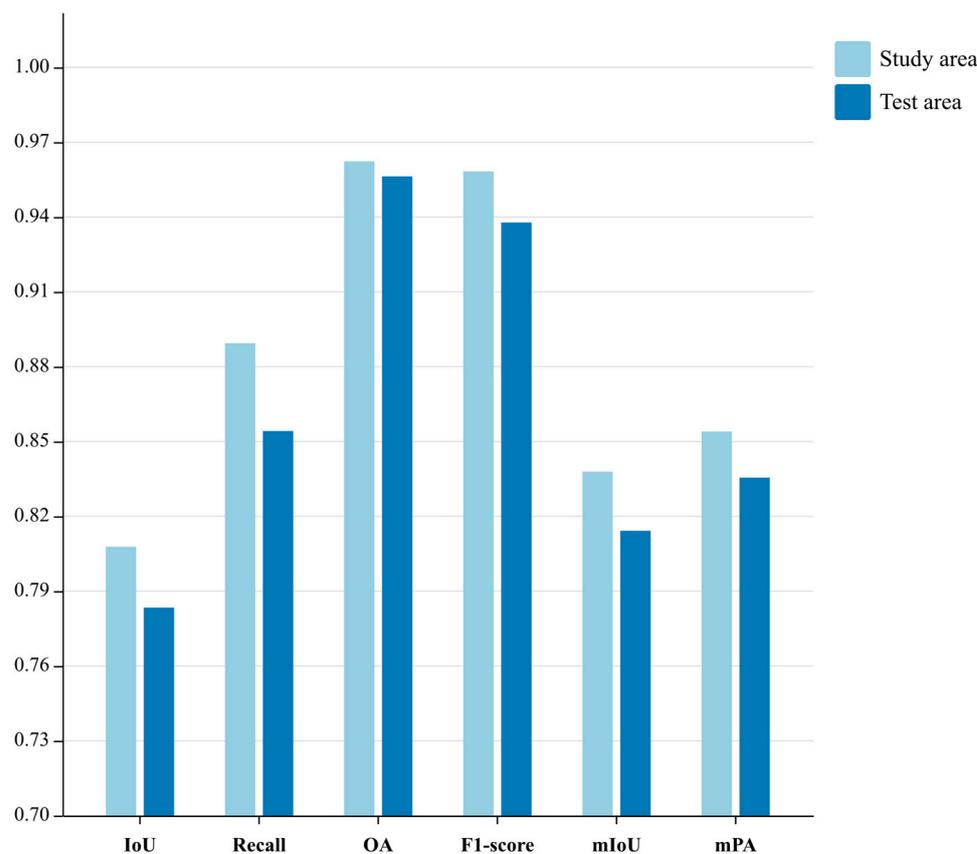


Figure 6. Results of model testing in Nanning City.

4. Discussion

4.1. Model Evaluation

This study aims to optimize and improve the DeepLabV3+ semantic segmentation model to address the challenges posed by complex scenes in high-resolution image clas-

sification. Traditional classification methods perform inadequately in such scenarios. We focused on addressing various issues, including the presence of numerous network parameters, long training time, and poor convergence. To overcome these challenges, we implemented improvements to the DeepLabV3+ semantic segmentation model. Firstly, we replaced the backbone network of the original model with a lightweight MobileNetV2 network, thereby reducing the model's complexity. Secondly, we introduced the CBAM attention mechanism module to enhance the model's ability to grasp semantic information. Many studies have been conducted on improving recognition accuracy through the incorporation of attention mechanisms into models. For instance, Wang et al. utilized the CBAM to enhance the discriminative capability of ecological environment elements in the Yangtze River source area [45]. Similarly, Liu et al. successfully resolved the problem of unclear edges and rough contours in winter wheat extraction by incorporating the CBAM [46]. These enhancements resulted in substantial improvements in training accuracy and efficiency. Furthermore, we developed a citrus classification extraction model based on GF-2 remote sensing images, which serves as a potent tool for extracting tree species information from high-resolution remote sensing images. Compared to other classification methods, Liang et al. achieved a spatial recognition of orange orchards in the study area by constructing multiple spectral vegetation indices, with an overall accuracy of 82.75% [38]. Our research findings demonstrate that the enhanced DeepLabV3+ model outperforms both the original model and other classification methods in citrus recognition. However, further improvements are necessary to enhance extraction accuracy. The complexity of citrus semantic information in remote sensing images, coupled with our relatively small dataset, necessitates caution in employing deeper and more complex models to prevent overfitting, which could diminish recognition accuracy. The upgraded DeepLabV3+ model exhibits the capability to focus on local information related to the target object, while the CBAM attention mechanism enables the network to prioritize pixels in fragmented citrus plantations and suppress other sources of interference. These enhancements enhance the efficiency and accuracy of the model, resulting in more comprehensive citrus recognition. Nonetheless, we still face substantial challenges in accurately recognizing citrus in complex and fragmented plots, as these areas may contain interfering elements that hinder accurate classification.

4.2. Future Prospects

Due to the data-driven nature of deep learning technology, the GF-2 satellite images utilized in this study exhibit high similarity and comparable feature distributions in the acquired image characteristics over a specific time period. These characteristics can be effectively captured through deep neural networks, indicating the favorable adaptability of the proposed methodology. It is essential to incorporate satellite remote sensing images from various sources into the training set and fine-tune the network training parameters. The enhanced citrus recognition model in this study can be further improved in several aspects. For instance, measures can be taken to enhance training accuracy and leverage a broader range of multi-source remote sensing images for transfer learning, thereby strengthening the model's generalization abilities. Additionally, this research intends to address the identification, classification, and subsequent analysis of complex and fragmented citrus plantations by exploiting multispectral remote sensing information and integrating it with practical agricultural practices. With the increasing availability of high-resolution remote sensing images and the growing demand for information in modern agricultural cultivation, conducting comprehensive ground surveys to obtain representative features and crop interpretation indicators becomes imperative. Establishing localized crop sample datasets contributes to overcoming the limitations associated with sample extraction in deep learning-based crop classification, consequently promoting the application and advancement of deep learning in crop remote sensing monitoring. Moreover, by harnessing the rich geometric structures and texture features present in high-resolution remote sensing images, deep learning methods excel at feature learning for extracting specific tree species

plots. Subsequently, by utilizing multi-source high-resolution remote sensing data, precise classification and extraction of major crops at the plot scale can be achieved, leading an accurate agricultural census, precision management of agricultural production, and adjustments in crop planting structures.

5. Conclusions

This study introduces a lightweight citrus extraction model based on the DeepLabV3+ semantic segmentation model and the CBAM attention mechanism module. The model adopts the lightweight MobileNetV2 as its backbone network and incorporates the CBAM module. By considering the growth characteristics of citrus, the model includes additional channels in the feature input. This approach effectively resolves the issues present in existing classical semantic segmentation models, such as inaccurate citrus edge extraction, misclassification caused by intra-class differences, and the challenges associated with numerous network parameters and long training time. In the study area, the proposed model achieves superior results in terms of overall accuracy (OA), mean pixel accuracy (mPA), and mean intersection over union (mIoU) for citrus extraction, achieving percentages of 96.23%, 83.79%, and 85.40%, respectively. Moreover, it outperforms other comparative models while minimizing the number of model parameters and training time. Nanning City in Guangxi Province has affirmed the robust generalization ability of this model. To summarize, this model ensures accurate extraction, reduces the number of parameters and training time, and exhibits strong generalization ability. Consequently, it is highly recommended for further promotion and application.

Author Contributions: Conceptualization, H.L., J.Z. (Jia Zhang), J.W. and Z.F.; methodology, H.L. and J.Z. (Jia Zhang); software, H.L., J.Z. (Jia Zhang), J.Z. (Junping Zhang), X.S., Y.L. and S.L.; validation, H.L.; formal analysis, H.L., J.Z. (Jia Zhang), J.W. and Z.F.; investigation, H.L.; resources, H.L.; data curation, H.L.; writing—original draft preparation, H.L.; writing—review and editing, H.L., J.W., Z.F., N.X. and B.L.; visualization, H.L.; supervision, J.W. and Z.F.; project administration, J.W. and Z.F.; funding acquisition, J.W. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Foundation 55 on Beijing Forestry University (340/GK112301013), the Fundamental Research Funds for the Beijing Natural Science Foundation Program (grant number 8222069, 8222052) and the Natural Science Foundation of China (grant numbers 42071342, 42101473, 42171329).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We are grateful to the undergraduate students and staff of the Laboratory of Forest Management and “3S” technology, Beijing Forestry University.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, Y.; Wang, Y.; Wang, J.; Yuan, Y.; Zhang, Z. An ontology-based approach to integration of hilly citrus production knowledge. *Comput. Electron. Agric.* **2015**, *113*, 24–43. [[CrossRef](#)]
2. Li, X.; Li, Y.; Ai, J.; Shu, Z.; Xia, J.; Xia, Y. Semantic segmentation of UAV remote sensing images based on edge feature fusing and multi-level upsampling integrated with Deeplabv3. *PLoS ONE* **2023**, *18*, e0279097. [[CrossRef](#)] [[PubMed](#)]
3. Houssein, E.H.; Helmy, B.E.-d.; Oliva, D.; Elngar, A.A.; Shaban, H. A novel Black Widow Optimization algorithm for multilevel thresholding image segmentation. *Expert Syst. Appl.* **2021**, *167*, 114159. [[CrossRef](#)]
4. Qin, J.; Wang, C.; Qin, G. A Multilevel Image Thresholding Method Based on Subspace Elimination Optimization. *Math. Probl. Eng.* **2019**, *2019*, 6706590. [[CrossRef](#)]
5. Zhao, D.; Liu, L.; Yu, F.; Heidari, A.A.; Wang, M.; Liang, G.; Muhammad, K.; Chen, H. Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy. *Knowl. Based Syst.* **2021**, *216*, 106510. [[CrossRef](#)]
6. Martino, F.D.; Sessa, S. PSO image thresholding on images compressed via fuzzy transforms. *Inf. Sci.* **2020**, *506*, 308–324. [[CrossRef](#)]

7. Eusuff, M.M.; Lansley, K.E. Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm. *J. Water Resour. Plan. Manag.* **2003**, *129*, 210–225. [[CrossRef](#)]
8. Chen, Y.; Wang, M.; Heidari, A.A.; Shi, B.; Hu, Z.; Zhang, Q.; Chen, H.; Mafarja, M.; Turabieh, H. Multi-threshold image segmentation using a multi-strategy shuffled frog leaping algorithm. *Expert Syst. Appl.* **2022**, *194*, 116511. [[CrossRef](#)]
9. Guobin, C.; Sun, Z.; Zhang, L. Road Identification Algorithm for Remote Sensing Images Based on Wavelet Transform and Recursive Operator. *IEEE Access* **2020**, *8*, 141824–141837. [[CrossRef](#)]
10. Xu, D.; Zhao, Y.; Jiang, Y.; Zhang, C.; Sun, B.; He, X. Using Improved Edge Detection Method to Detect Mining-Induced Ground Fissures Identified by Unmanned Aerial Vehicle Remote Sensing. *Remote Sens.* **2021**, *13*, 3652. [[CrossRef](#)]
11. Chetia, R.; Boruah, S.M.B.; Sahu, P.P. Quantum image edge detection using improved Sobel mask based on NEQR. *Quantum Inf. Process.* **2021**, *20*, 21. [[CrossRef](#)]
12. Jan, A.; Parah, S.A.; Malik, B.A.; Rashid, M. Secure data transmission in IoTs based on CLoG edge detection. *Future Gener. Comput. Syst.* **2021**, *121*, 59–73. [[CrossRef](#)]
13. Roy, S.; Das, D.; Lal, S.; Kini, J. Novel edge detection method for nuclei segmentation of liver cancer histopathology images. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *14*, 479–496. [[CrossRef](#)]
14. Ali, I.; Rehman, A.U.; Khan, D.M.; Khan, Z.; Shafiq, M.; Choi, J.-G. Model Selection Using K-Means Clustering Algorithm for the Symmetrical Segmentation of Remote Sensing Datasets. *Symmetry* **2022**, *14*, 1149. [[CrossRef](#)]
15. Mahata, K.; Das, R.; Das, S.; Sarkar, A. Land Use Land Cover map segmentation using Remote Sensing: A Case study of Ajoy river watershed, India. *J. Intell. Syst.* **2020**, *30*, 273–286. [[CrossRef](#)]
16. Pastorino, M.; Montaldo, A.; Fronda, L.; Hedhli, I.; Moser, G.; Serpico, S.B.; Zerubia, J. Multisensor and Multiresolution Remote Sensing Image Classification through a Causal Hierarchical Markov Framework and Decision Tree Ensembles. *Remote Sens.* **2021**, *13*, 849. [[CrossRef](#)]
17. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
18. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Zhu, D.e.; Zheng, J.; Zhang, M.; Xing, L.; et al. Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *13*, 113–128. [[CrossRef](#)]
19. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
20. Razaque, A.; Ben Haj Frej, M.; Almi'ani, M.; Alotaibi, M.; Alotaibi, B. Improved Support Vector Machine Enabled Radial Basis Function and Linear Variants for Remote Sensing Image Classification. *Sensors* **2021**, *21*, 4431. [[CrossRef](#)]
21. Alex, K.; Ilya, S.; Geoffrey, E.H. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90.
22. Li, X.; Xu, F.; Xia, R.; Li, T.; Chen, Z.; Wang, X.; Xu, Z.; Lyu, X. Encoding Contextual Information by Interlacing Transformer and Convolution for Remote Sensing Imagery Semantic Segmentation. *Remote Sens.* **2022**, *14*, 4065. [[CrossRef](#)]
23. Yang, Q.; Shi, L.; Han, J.; Zha, Y.; Zhu, P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.* **2019**, *235*, 142–153. [[CrossRef](#)]
24. Su, Z.; Wang, Y.; Xu, Q.; Gao, R.; Kong, Q. LodgeNet: Improved rice lodging recognition using semantic segmentation of UAV high-resolution remote sensing images. *Comput. Electron. Agric.* **2022**, *196*, 4065. [[CrossRef](#)]
25. Zhou, H.; Zhang, J.; Lei, J.; Li, S.; TU, D. Image Semantic Segmentation Based on FCN-CRF Model. In Proceedings of the 2016 International Conference on Image, Vision and Computing, Palmerston North, New Zealand, 21–22 November 2016.
26. Li, X.; Xu, F.; Liu, F.; Lyu, X.; Tong, Y.; Xu, Z.; Zhou, J. A Synergistical Attention Model for Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5400916. [[CrossRef](#)]
27. Tian, L.; Zhong, X.; Chen, M.; Wang, P. Semantic Segmentation of Remote Sensing Image Based on GAN and FCN Network Model. *Sci. Program.* **2021**, *2021*, 9491376. [[CrossRef](#)]
28. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
29. Li, X.; Xu, F.; Liu, F.; Xia, R.; Tong, Y.; Li, L.; Xu, Z.; Lyu, X. Hybridizing Euclidean and Hyperbolic Similarities for Attentively Refining Representations in Semantic Segmentation of Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5003605. [[CrossRef](#)]
30. Wang, Z.; Peng, T.; Lu, Z. Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks. *Forests* **2022**, *13*, 1133. [[CrossRef](#)]
31. Wang, Z.; Wang, J.; Yang, K.; Wang, L.; Su, F.; Chen, X. Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. *Comput. Geosci.* **2022**, *158*, 104969. [[CrossRef](#)]
32. Du, S.; Du, S.; Liu, B.; Zhang, X. Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images. *Int. J. Digit. Earth* **2020**, *14*, 357–378. [[CrossRef](#)]
33. Gu, X.; Li, S.; Ren, S.; Zheng, H.; Fan, C.; Xu, H. Adaptive enhanced swin transformer with U-net for remote sensing image segmentation. *Comput. Electr. Eng.* **2022**, *102*, 108223. [[CrossRef](#)]

34. Hou, J.; Guo, Z.; Wu, Y.; Diao, W.; Xu, T. BSNet: Dynamic Hybrid Gradient Convolution Based Boundary-Sensitive Network for Remote Sensing Image Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5624022. [[CrossRef](#)]
35. Lin, Z.; Guo, W. Cotton Stand Counting from Unmanned Aerial System Imagery Using MobileNet and CenterNet Deep Learning Models. *Remote Sens.* **2021**, *13*, 2822. [[CrossRef](#)]
36. Li, L.; Li, H.; Peng, L.; Li, Y.; Zhou, Y.; Chai, F.; Mo, Z.; Chen, Z.; Mao, J.; Wang, W. Characterization of precipitation in the background of atmospheric pollutants reduction in Guilin: Temporal variation and source apportionment. *J. Environ. Sci.* **2020**, *98*, 1–13. [[CrossRef](#)] [[PubMed](#)]
37. Li, Y.; Wang, C.; Wright, A.; Liu, H.; Zhang, H.; Zong, Y. Combination of GF-2 high spatial resolution imagery and land surface factors for predicting soil salinity of muddy coasts. *Catena* **2021**, *202*, 105304. [[CrossRef](#)]
38. Liang, C.; Huang, Q.; Wang, S.; Wang, C.; Yu, Q.; Wu, W. Identification of citrus orchard under vegetation indexes using multi-temporal remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 168–176. (In Chinese) [[CrossRef](#)]
39. Kuang, X.; Guo, J.; Bai, J.; Geng, H.; Wang, H. Crop-Planting Area Prediction from Multi-Source Gaofen Satellite Images Using a Novel Deep Learning Model: A Case Study of Yangling District. *Remote Sens.* **2023**, *15*, 3792. [[CrossRef](#)]
40. Sun, W.; Chen, B.; Messinger, D.W. Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images. *Opt. Eng.* **2014**, *53*, 013107. [[CrossRef](#)]
41. Li, X.; Xu, F.; Lyu, X.; Gao, H.; Tong, Y.; Cai, S.; Li, S.; Liu, D. Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images. *Int. J. Remote Sens.* **2021**, *42*, 3583–3610. [[CrossRef](#)]
42. Mo, L.; Fan, Y.; Wang, G.; Yi, X.; Wu, X.; Wu, P. DeepMDSCBA: An Improved Semantic Segmentation Model Based on DeepLabV3+ for Apple Images. *Foods* **2022**, *11*, 3999. [[CrossRef](#)] [[PubMed](#)]
43. Ma, R.; Wang, J.; Zhao, W.; Guo, H.; Dai, D.; Yun, Y.; Li, L.; Hao, F.; Bai, J.; Ma, D. Identification of Maize Seed Varieties Using MobileNetV2 with Improved Attention Mechanism CBAM. *Agriculture* **2022**, *13*, 11. [[CrossRef](#)]
44. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
45. Wang, C.; Zhang, R.; Chang, L. A Study on the Dynamic Effects and Ecological Stress of Eco-Environment in the Headwaters of the Yangtze River Based on Improved DeepLab V3+ Network. *Remote Sens.* **2022**, *14*, 2225. [[CrossRef](#)]
46. Liu, J.; Wang, H.; Zhang, Y.; Zhao, X.; Qu, T.; Tian, H.; Lu, Y.; Su, J.; Luo, D.; Yang, Y. A Spatial Distribution Extraction Method for Winter Wheat Based on Improved U-Net. *Remote Sens.* **2023**, *15*, 3711. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.