



Article

Extracting Citrus-Growing Regions by Multiscale UNet Using Sentinel-2 Satellite Imagery

Yong Li, Wenjing Liu ^{*}, Ying Ge, Sai Yuan, Tingxuan Zhang and Xiuhui Liu

School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China; liyong@hhu.edu.cn (Y.L.); geying@hhu.edu.cn (Y.G.); yuansaihhu@hhu.edu.cn (S.Y.); zhangtingxuan@hhu.edu.cn (T.Z.); liuxiuhui1108@hhu.edu.cn (X.L.)

* Correspondence: lwwjing@hhu.edu.cn

Abstract: Citrus is an important commercial crop in many areas. The management and planning of citrus growing can be supported by timely and efficient monitoring of citrus-growing regions. Their complex planting structure and the weather are likely to cause problems for extracting citrus-growing regions from remote sensing images. To accurately extract citrus-growing regions, deep learning is employed, because it has a strong feature representation ability and can obtain rich semantic information. A novel model for extracting citrus-growing regions by UNet that incorporates an image pyramid structure is proposed on the basis of the Sentinel-2 satellite imagery. A pyramid-structured encoder, a decoder, and multiscale skip connections are the three main components of the model. Additionally, atrous spatial pyramid pooling is used to prevent information loss and improve the ability to learn spatial features. The experimental results show that the proposed model has the best performance, with the precision, the intersection over union, the recall, and the F1-score reaching 88.96%, 73.22%, 80.55%, and 84.54%, respectively. The extracted citrus-growing regions have regular boundaries and complete parcels. Furthermore, the proposed model has greater overall accuracy, kappa, producer accuracy, and user accuracy than the object-oriented random forest algorithm that is widely applied in various fields. Overall, the proposed method shows a better generalization ability, higher robustness, greater accuracy, and less fragmented extraction results. This research can support the rapid and accurate mapping of large-scale citrus-growing regions.

Keywords: Sentinel-2 satellite remote sensing; deep learning; extracting citrus-growing regions; UNet; image pyramid; atrous spatial pyramid pooling



Citation: Li, Y.; Liu, W.; Ge, Y.; Yuan, S.; Zhang, T.; Liu, X. Extracting Citrus-Growing Regions by Multiscale UNet Using Sentinel-2 Satellite Imagery. *Remote Sens.* **2024**, *16*, 36. <https://doi.org/10.3390/rs16010036>

Academic Editor: Jochem Verrelst

Received: 30 October 2023

Revised: 17 December 2023

Accepted: 18 December 2023

Published: 21 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/>).

1. Introduction

Citrus is an important commercial crop that plays a crucial role in boosting the local economy [1]. Rapid monitoring of the distribution and changes in citrus-growing regions is significant for the management and development of citrus production. However, a traditional field survey requires considerable time and workforce. Effective and periodic monitoring of the vast citrus-growing regions is a challenge. The technology of satellite remote sensing provides support for rapidly collecting detailed surface information on a large scale, which can be employed to monitor and analyze the growth of citrus and predict citrus production.

Considering the ability of remote sensing to classify different land covers, current research relies on spectral information obtained from satellite or drone imagery to identify some crops [2]. Some research detected citrus and other crop trees from UAV images using a CNN algorithm, with the aim of obtaining localized and refined results [3]. But for large-scale mapping, satellite imagery is more accessible than drones. Wei et al. [4] identified maize, rice, and soybean using random forest (RF) algorithms and Sentinel-2 time series data. Several studies have utilized the time series SAVI and OSAVI to improve crop mapping [5]. Constructing crop-sensitive time series vegetation indices can amplify

differences to facilitate the identification tasks [6]. However, the identification process is usually difficult because of the complexity of agricultural farming patterns and the limited amount of remote sensing data, due to overcast weather [7–9]. The effective use of spatial features is a promising idea for crop classification in such challenging areas.

Object-oriented methods have been widely applied to classify different land covers. Given that cropland appears in images as regular geometric shapes with specific textures, object-oriented methods analyze “objects in space” rather than “pixels in space”, which can learn the spatial information of croplands and suppress salt-and-pepper noise to some extent [10,11]. Luo et al. [12] verified the feasibility of the object-oriented RF algorithm on Google Earth Engine using temporal Sentinel-1 images. However, the object-oriented approach essentially does not consider the contextual semantic information contained in images and ignores the high-level features among objects. The object units obtained via the object-oriented approach are often inconsistent with the morphology that people expect for the actual target features.

Deep learning methods have the capability of acquiring the contextual information of each pixel to enhance performance and reduce noise, with great robustness in complex and various situations. Du et al. [13] employed Cropland Data Layer and Landsat time series images to train UNet for extracting Arkansas rice and confirmed that UNet outperformed RF in most cases. Gadiraju et al. [14] presented a deep learning scheme with multimodal inputs, such as spectral, spatial, and climatic information, to distinguish crop types, thereby reducing the prediction error by 60%. Bian et al. [15] designed CACPU-Net for crop mapping based on a Sentinel-2 autumn remote sensing image, adding an attention module and a difficulty-focused module to focus on the extraction of difficult regions for optimizing the final results.

Currently, buildings, water bodies, and some field crops remain the main targets for deep learning-based remote sensing semantic segmentation, while little research has been conducted on orchards. Orchards are often distributed in hilly areas with foggy and cloudy climates, which pose difficulties for providing time series data for deep learning models [16]. The deep learning semantic segmentation networks enable the extraction of fruit trees, such as banana and kiwifruit, from UAV images [17,18]. For large-scale areas, it is undoubtedly costly to collect UAV images. Some studies have achieved the mapping of poplar, willow, and palm trees using Sentinel-2 satellite images. However, it is challenging to extract citrus orchards due to their complex planting structures with other neighboring crops [19,20]. How to extract citrus orchards by making the best use of the structure and spectral information of planting regions under the influence of complex planting structures and the limited amount of remote sensing data affected by cloudy and rainy weather is the focus of this study.

For accurately mapping the distribution of citrus-growing regions, a method of deep learning is proposed for acquiring citrus orchards using Sentinel-2 remote sensing images. The UNet model is improved by combining the idea of an image pyramid based on the selected spectral bands and indices. In addition, the multiscale spatial information of citrus-growing regions is derived by atrous spatial pyramid pooling (ASPP). The proposed method is implemented on images of different areas and times and compared with other models. The evaluation results demonstrate that the model improves the extraction accuracy and the generalization ability and has the potential for citrus-growing region extraction.

The major contents of the rest of the paper are as follows: firstly, the study area and datasets are introduced in Section 2; next, the involved methodology and principles are described in Section 3; then, the experiments and the experimental results are shown in Section 4; and finally, the discussion and conclusion are given in Sections 5 and 6.

2. Study Area and Datasets

2.1. Study Area

The main area for collecting training samples is located in Xinping County, Yuxi City, Yunnan Province, China. The areas for acquiring test samples are situated in Binchuan and Longling counties in Yunnan Province, China. Xinping is located in the southwestern part of central Yunnan Province, belonging to the temperate climate zone. The terrain is mainly mountainous, with the altitude ranging from 422 m to 3165.9 m. Influenced by the altitude, three types of climates are formed in the area: the high-temperature zone in the river valley, the warm-temperature zone in the mid-mountains, and the low-temperature zone in the high mountains. Xinping has an annual precipitation of 869 mm, and the rainy season is concentrated from April to October, with a total of 2838.7 h of sunshine. Longling County is situated along the valley of the Nu River, and Binchuan County is located in the valley of the south bank of the Jinsha River. The study area has good conditions, such as sufficient light, abundant precipitation, an obvious three-dimensional climate, and typical mountainous agricultural characteristics, for citrus cultivation.

The agricultural cropping structure in the study area is complex, with the extensive cultivation of corn, rice, potatoes, soybeans, wheat, and other crops, in addition to citrus. Figure 1a–c show overview maps of the study area. Figure 1d,e show Google Earth images covering the citrus-growing regions and other croplands in the Xinping and Longling counties, and their locations are marked with yellow dots in Figure 1a,b.

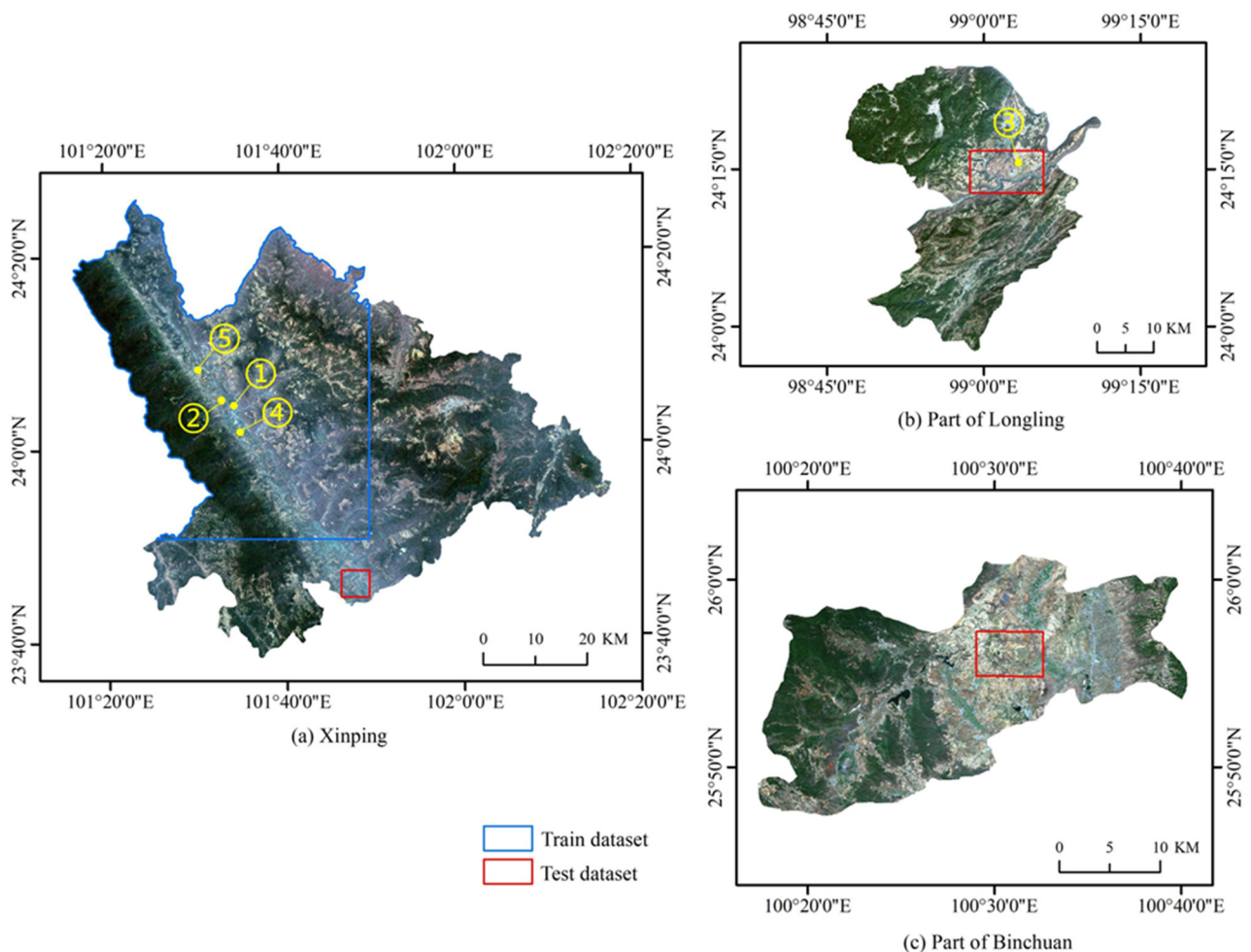


Figure 1. Cont.

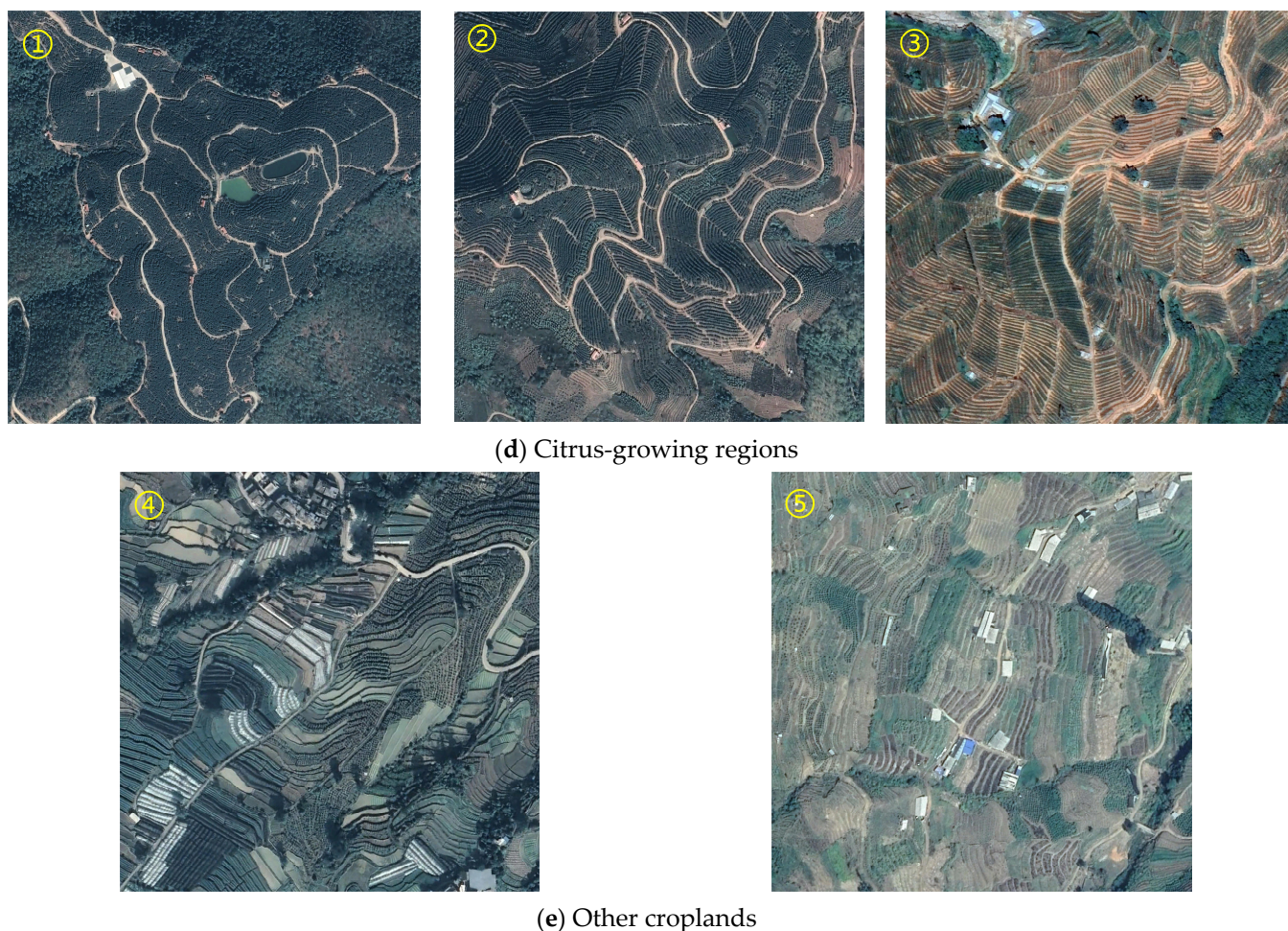


Figure 1. Geographical map of the study and test areas. (a) Xinping County. (b) Part of Longling County. (c) Part of Binchuan County. (d) Citrus-growing regions from different citrus orchards in Xinping and Longling counties, which are green to dark green with a uniform texture. The gaps between the citrus trees are large or small, regularly arranged in blocks, and separated by paths. (e) Other croplands from Xinping County, which are tawny or light green in color and are generally regular rectangular plots with a uniform texture and few gaps. Some crops have a similar texture to citrus trees and are well dispersed.

2.2. Study Data

2.2.1. Satellite Imagery

The Sentinel-2 satellite imagery is obtained from the website of the European Space Agency. The growth period of citrus is from February to October. The weather from April to October in the study area is influenced by the rainy season, and the images acquired during this period are heavily cloud obscured, which affects the extraction of citrus-growing regions. Therefore, the following data are selected under the consideration of images with little or no cloud cover: (1) Phase 3 on 30 March 2020 (Xinping); (2) Phase 2 on 13 March 2021 (Longling area); (3) Phase 1 on 15 March 2020 (Binchuan area). Sentinel-2 images have 13 spectral bands with three different spatial resolutions (10, 20, and 60 m), including three red-edge bands. These red-edge bands are very useful to monitor crop health and identify vegetation cover [21]. To obtain the true reflectance of the surface, this study uses the Sen2Cor plug-in to process the Level 1C images into Level 2A images and resample all bands to 10 m. Because the study area is characterized by large topographic relief, radiometric inconsistencies occur when imaging shaded and sunny slopes. The Teillet

model in the topographic correction extension module of ENVI 5.3 is adopted to perform topographic corrections.

2.2.2. Dataset Construction

The labeling data in this study are obtained from the field survey of citrus-growing regions and visual interpretation results from Google Maps. Figure 1a–c illustrate the image and data division. Images are clipped into patches of 256×256 size with 50% overlap to minimize any boundary discontinuities. Given that a considerable background in the study area is likely to hinder the model in learning positive categories, most negative samples are eliminated, leaving the number of positive and negative samples at approximately 1:1. Then, the data are augmented through geometric transformations to obtain a collection of 1372 image blocks. These image blocks are randomly assigned to the training and validation sets at a ratio of 8:2.

3. Methods

3.1. Calculating Spectral Indices

Difficulties and errors in identifying citrus orchards often occur in areas covered by vegetation, and the construction of vegetation indices can utilize spectral differences to expand the distinction between vegetation types. On the basis of the growth characteristics of citrus and existing research [22,23], nine spectral indices and five red-edge indices are calculated using nine bands (B2–B8, B11 and B12) of Sentinel-2. The spectral indices include RVI, DVI, EVI, NDVI, GNDVI, GCVI, SAVI, OSAVI, and MSAVI. The vegetation red-edge channel is unique to Sentinel-2 imagery and facilitates the differentiation of vegetation species. The introduction of the red-edge index can take full advantage of the red-edge bands for distinguishing various crops effectively [24]. To assess the performance of the red-edge data for extracting citrus orchards, NDre and NDVire are derived using three red-edge bands.

3.2. Selecting Spectral Indices

The introduction of excessive indices can lead to model redundancy and affect the results. In this study, the 14 citrus-related vegetation indices mentioned in Section 3.1 are considered, and the highest-scoring vegetation indices are selected to be input into the subsequent neural network. The Relief F algorithm [25] is a filtered feature preselection algorithm, meaning that feature selection is performed before training the learner and the feature selection process is independent of the learner. This algorithm makes it easy to obtain the optimal citrus feature space before model training. Relief F is an extension of the Relief algorithm, which works well for multiclassification issues [26]. The algorithm has high robustness and is appropriate for handling noisy and partial data [27]. The main types in this study area are forest, farmland, grassland, citrus orchards, and bare land. A test sample R is randomly selected from the sample set, and k nearest neighbors of sample R are searched. From the remaining four groups of samples of the different types, k nearest neighbor samples of the different types are taken. The weight of each feature is determined by comparing the distances between sample features of the same and different types. A feature F whose weight increases when the distance of different category samples is larger than the distance of the same category samples, indicating that the feature is meaningful for classification, is considered. After T iterations, the average weight of each feature is obtained as the final weight [28]. The weights are calculated as follows:

$$\omega(F_i) = \omega(F_i) - \frac{1}{T \cdot k} \sum_{h \in H} |R_i - h_i| + \frac{1}{T \cdot k} \sum_{m \in M} |R_i - m_i|, \quad (1)$$

where $\omega(F_i)$ denotes the weight assigned to the i th feature, and k is the amount of nearest samples. $\sum_{h \in H} |R_i - h_i|$ is the distance sum of sample R and the selected k nearest samples of the same types in terms of the i th feature. $\sum_{m \in M} |R_i - m_i|$ denotes the distance sum of sample R and the k nearest samples of different types in terms of the i th feature.

Additionally, the correlation between these 14 vegetation indices is evaluated using the Pearson correlation coefficient [29]. The correlation coefficients are calculated as follows:

$$r_{\text{pearson}} = \frac{\sum(x_i - X)(y_i - Y)}{\sqrt{\sum(x_i - X)^2 \cdot \sum(y_i - Y)^2}}, \quad (2)$$

where r_{pearson} means the Pearson correlation coefficient, and i represents the number of the sample point. x and y represent two different features, X denotes the mean of x , and Y denotes the mean of y .

3.3. Improving the UNet Model by Incorporating the Image Pyramid

For complex planting structures, semantic segmentation networks are more capable of considering neighboring pixels and aggregating contextual information than other methods, such as the object-oriented method. How to utilize the limited data is critical when only a few images are available.

The UNet model has relatively high accuracy with a small sample size, and it has skip connections that combine deep and shallow features [30]. During feature extraction, simple image attributes, such as color, borders, and other elements, are captured by the shallow structure. Deep structures with a large valid receptive field (VRF) and added convolutional operations are able to capture deep semantic information within the image. This study makes improvement on the basis of the UNet model using the pretrained three-layer Residual Networks 50 (ResNet50) on the ImageNet dataset as the encoder. The proposed architecture is depicted in Figure 2.

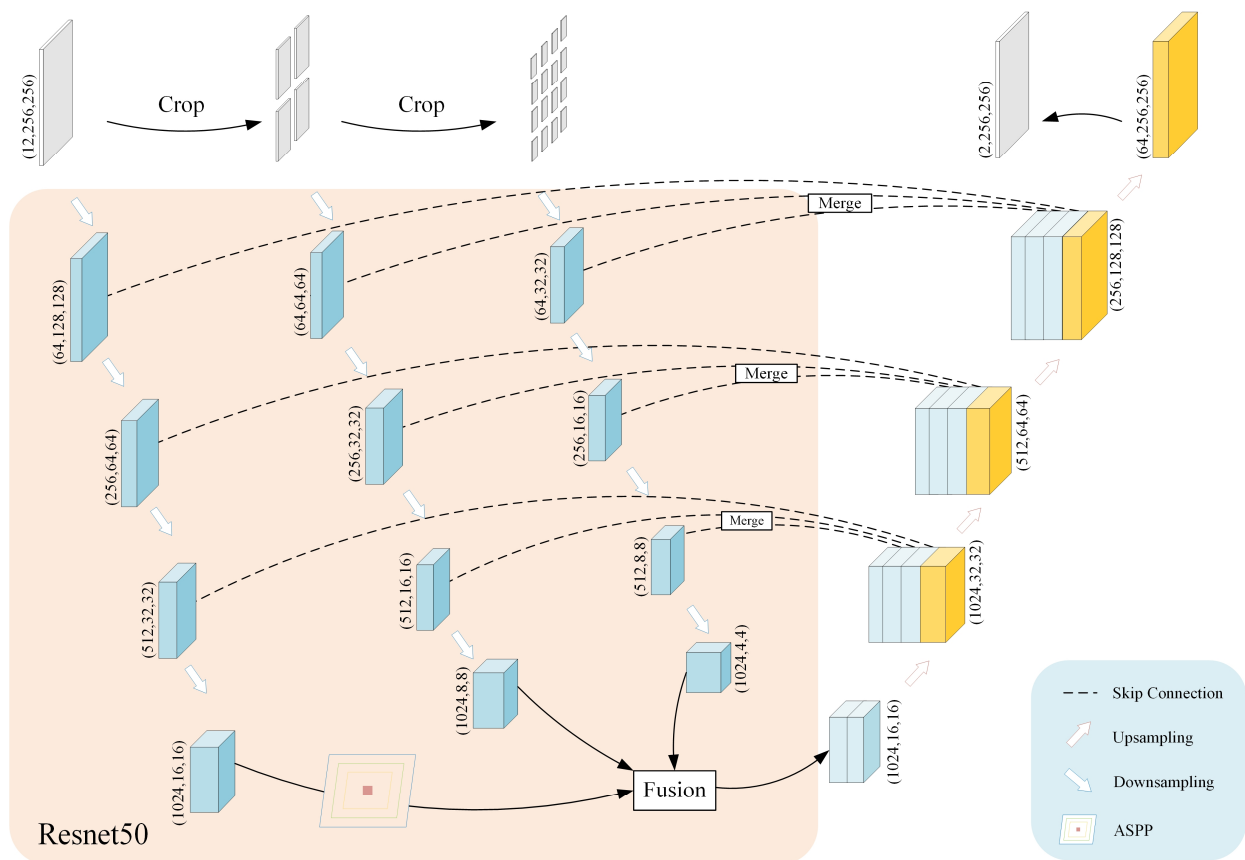


Figure 2. Overview of the proposed network. The yellow rectangles represent the output of the up-sampling, connected with the skip connection.

In the trunk part, the image blocks with the selected features are used as inputs to the network. The 256×256 VRF happens to completely cover the citrus-growing area, allowing the model to learn planting structure information from the global perspective, such as the small amount of wild forest vegetation and scattered houses inhabited by irrigators that are usually distributed around citrus orchards. ASPP is added after the three-layer ResNet50 to capture contextual information on multiple scales.

The ASPP module is derived from the DeepLab model. A pyramid structure is constructed by pooling and using a number of parallel atrous convolutions with varying dilation rates [31,32]. This module can enlarge the VRF and obtain multiscale information [33].

Inspired by the image pyramid and the multiscale training architecture of Ding et al. [34], two segmentation branches are designed in the second part of the model to obtain detailed information from the cropped image blocks for supplementation. This module is only suitable for learning the texture and spectral information of small image blocks. ASPP is not added to lower the training cost. First, the image blocks are segmented into 128×128 and 64×64 sizes and fed into two encoders separately. Then, to integrate the features, this study designs a feature fusion module after the encoders, as depicted in Figure 3. The general idea of the feature fusion module is to supplement the loss of detail in the trunk parts with information from the branches. To facilitate data transmission, these branch feature graphs are recovered to their original 256×256 size, and then combined via pixel addition for information enhancement. To avoid an excessive data volume, the outputs of the trunk and branches are downscaled using $\text{Conv}1 \times 1$. Finally, the information of the branch and the trunk is integrated via channels concatenation. In addition, the segmentation branch uses the skip connections and restores the image size at the decoder to fuse the multiscale information. The proposed architecture is able to learn features from multiple levels for accurate semantic segmentation.

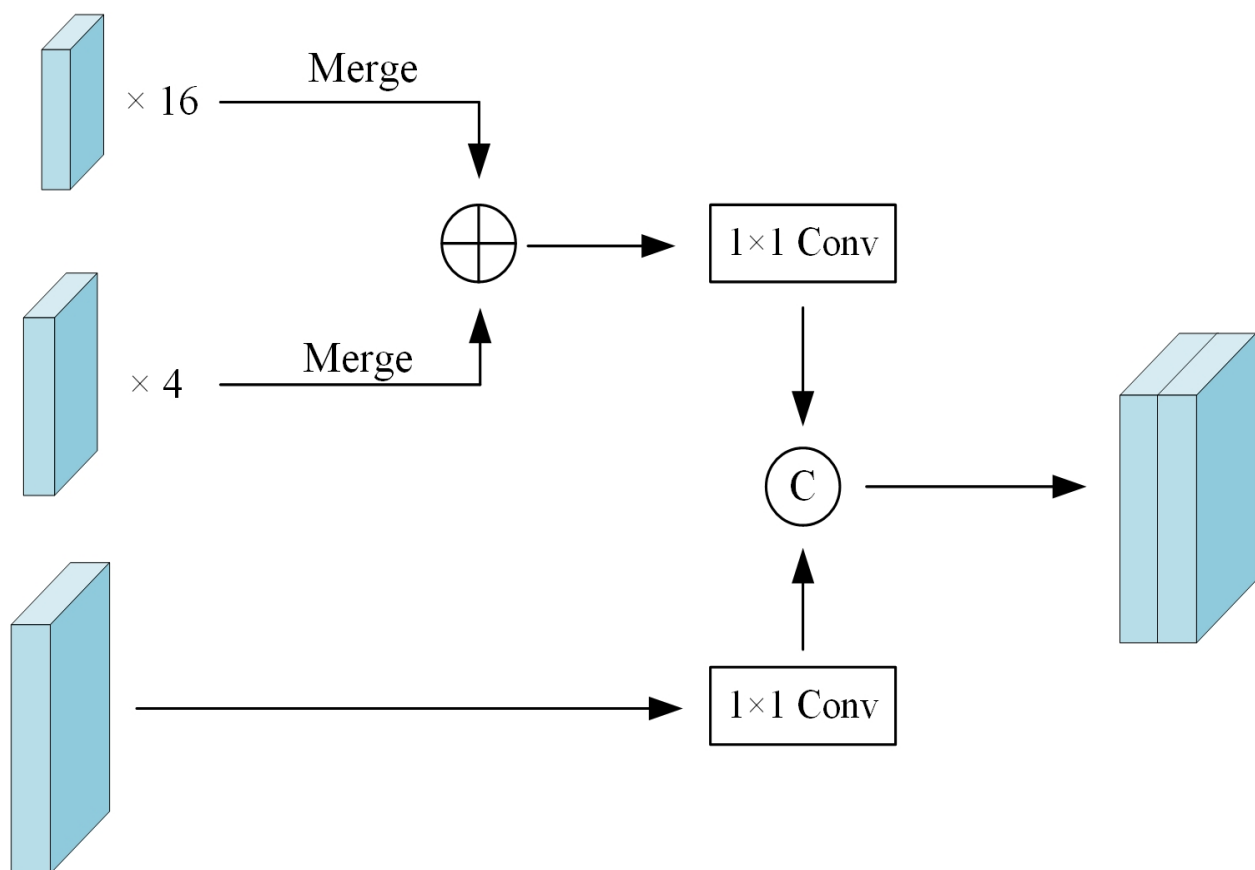


Figure 3. Feature fusion block (“ \oplus ” denotes pixel-wise sum, circle c denotes concatenate).

3.4. Model Setting and Compiling

3.4.1. Loss Function

Adding weights to the cross-entropy loss is expected to deal with the data imbalance of different categories. In consideration of the imbalance of the dataset, the weighted cross-entropy loss is employed to assign higher weights to citrus-growing regions with fewer appearances. The loss expression is the following:

$$L = -\sum_{n=1}^N w_n y_n \log(p_n), \quad (3)$$

where p_n is the vector containing the predicted probabilities of category n , y_n is a one-hot vector with elements having only 1 and 0, and N is the number of categories.

The frequency of occurrence of class pixels is inversely proportional to the weights of each category. As suggested in [35], the class weights are assigned as follows:

$$w_n = \frac{1}{\ln(1.02 + \beta_n)}, \quad (4)$$

where w_n is the weight of category n , and β_n is the frequency of occurrence of category n .

3.4.2. Building the Model

In this study, the models are performed using Python 3.7.13 and PyTorch 1.11.0 on a computer with Intel(R) Core(TM) I7-12700H CPU @ 4.70 GHz, 16.0 GB RAM, and NVIDIA GeForce RTX 3060 GPU, with 6.0 GB memory. Because the Adam optimizer performs well with remote sensing data [36], it is used in this study for model training [37]. Based on the tuning results in Figure 4 and GPU memory, the batch size is 4, and the initial learning rate is 1×10^{-4} . According to loss convergence, the number of training epochs is set to 80. In addition, this study tests various learning rate schedulers before setting a schedule and searches for the optimal regulation frequency between 10 and 30 epochs. Lastly, a scheduler that reduces the learning rate to one-tenth of the current every 25 epochs is set up.

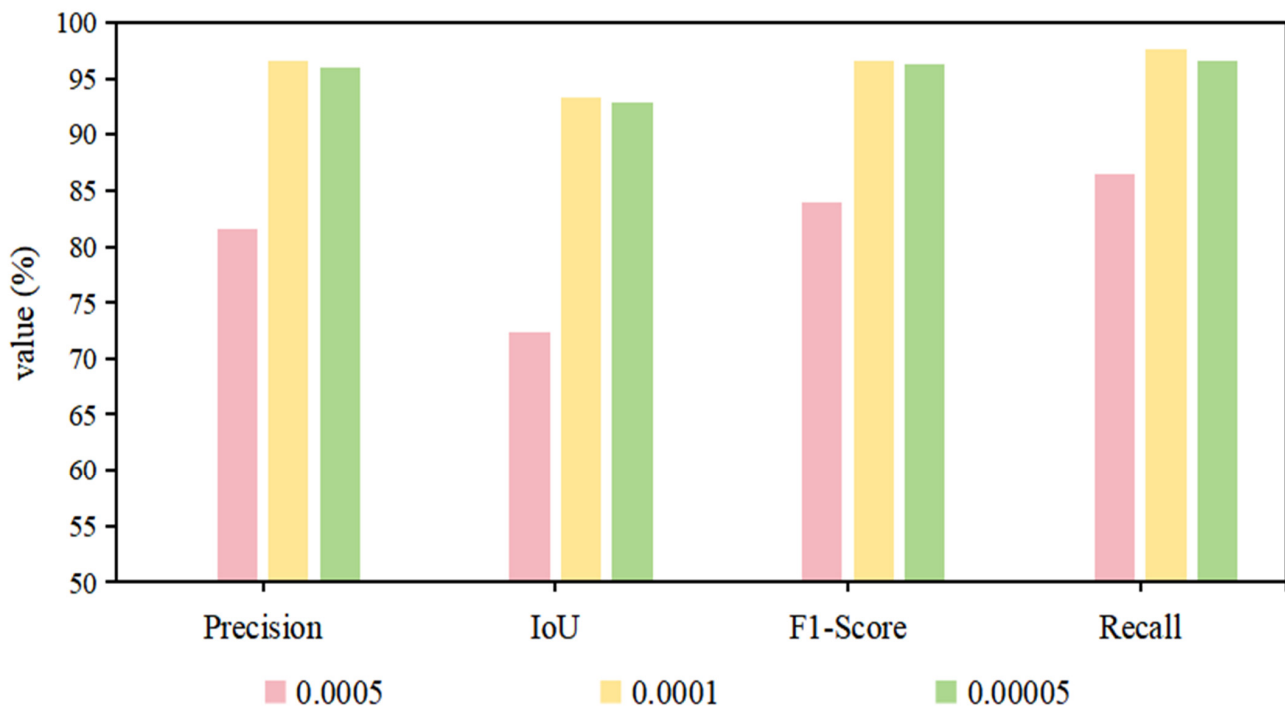


Figure 4. Selection of learning rates.

3.5. Evaluation Metrics

As mentioned in the above section, the training images are split into training and validation data in this study. The test data derived from three different areas are used to precisely assess the different models in practice. The evaluation metrics of precision, recall, F1-score, and intersection over union (IoU) are employed, as shown in Formulas (5)–(8). These assessment metrics are widely utilized in semantic segmentation tasks. The F1-score, which accurately displays the model’s performance on unbalanced datasets, is the harmonic mean of precision and recall [38].

$$\text{Precision} = \frac{TP}{TP + FP'} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN'} \quad (6)$$

$$\text{F1 - Score} = \frac{2 \times TP}{2 \times TP + FN + FP'} \quad (7)$$

$$\text{IoU} = \frac{TP}{TP + FN + FP'} \quad (8)$$

where TP denotes the amount of pixels correctly classified as citrus-growing regions, FP is the amount of background pixels wrongly judged as citrus-growing regions, and FN is the amount of citrus pixels misclassified as the background.

4. Results

4.1. Selecting Spectral Features

On the basis of the Relief F algorithm, a total of 1200 sample points are randomly selected from the Sentinel-2 image for spectral index selection. The area of each sample is consistent with the area of one pixel of the image, which is 10 m × 10 m. The samples include the five categories of forest, farmland, bareland, grassland, and citrus orchard, derived from ground truth data and a visual interpretation of Google Maps, with exactly 240 points in each type. The results of the importance evaluation are shown in Figure 5. From the 14 vegetation indices, the 3 indices with the highest importance are selected, namely, RVI, Ndre2, and NDVI. The local images can be seen in Figure 6. Figure 7 shows the results of the correlation analysis for all features, with the three most important vegetation indices having relatively low correlations between themselves. The final selected features and the corresponding formulas are shown in Table 1.

Table 1. Selected features and corresponding formulas.

Feature	Calculation Formula
Bands	B2, B3, B4, B5, B6, B7, B8, B11, B12
Ratio vegetation index (RVI)	$RVI = B8/B4$
Normalized difference red edge index 2 (NDre2)	$NDre2 = (B7 - B5)/(B7 + B5)$
Normalized difference vegetation index (NDVI)	$NDVI = (B8 - B4)/(B8 + B4)$

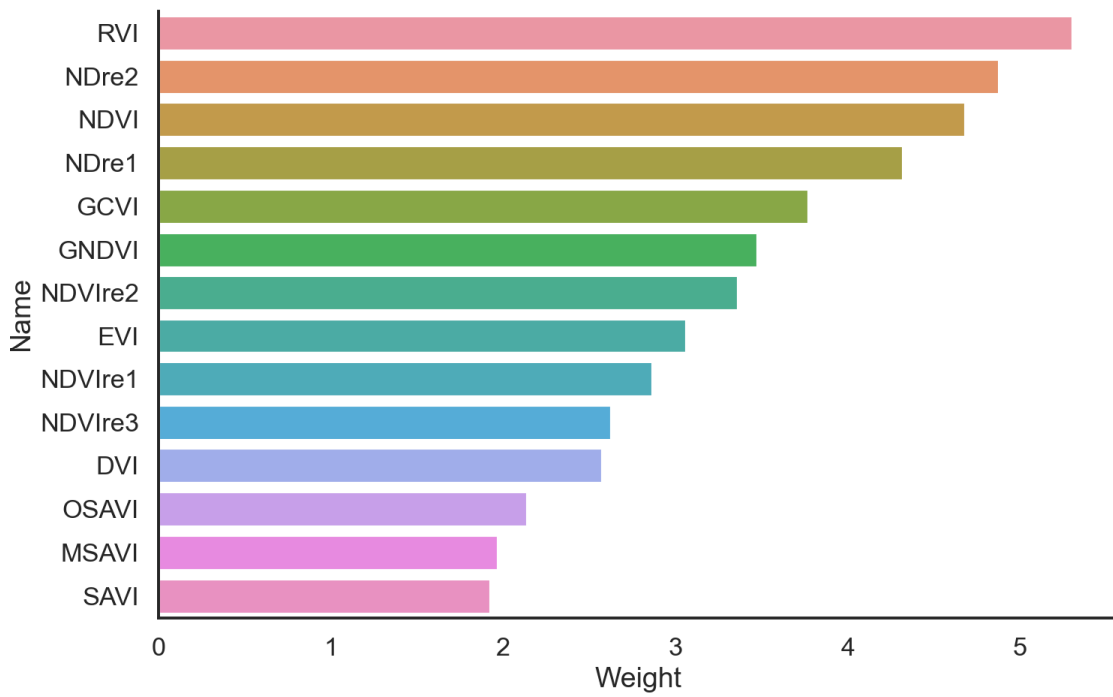
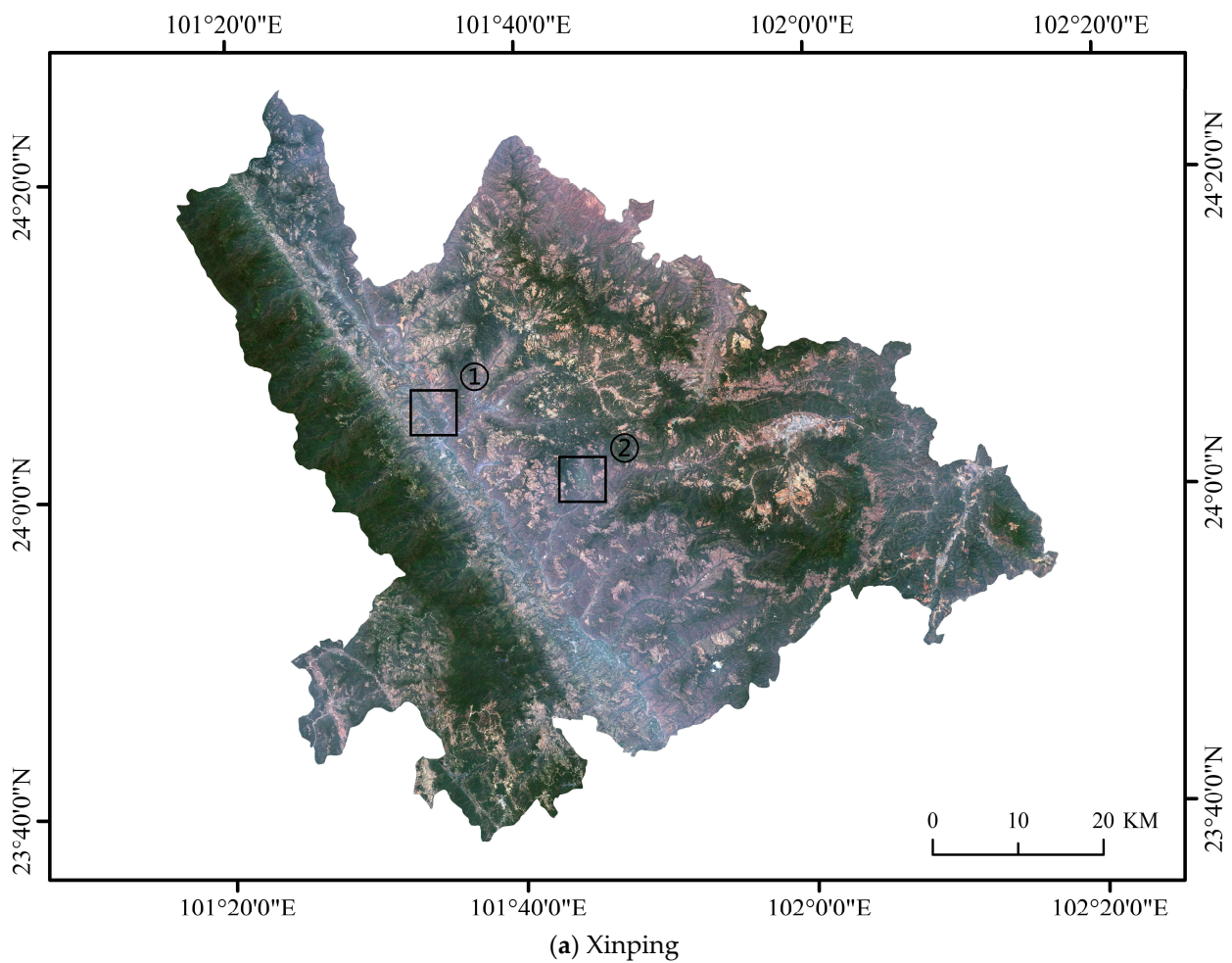


Figure 5. Weights of features.



(a) Xinpingshan

Figure 6. Cont.

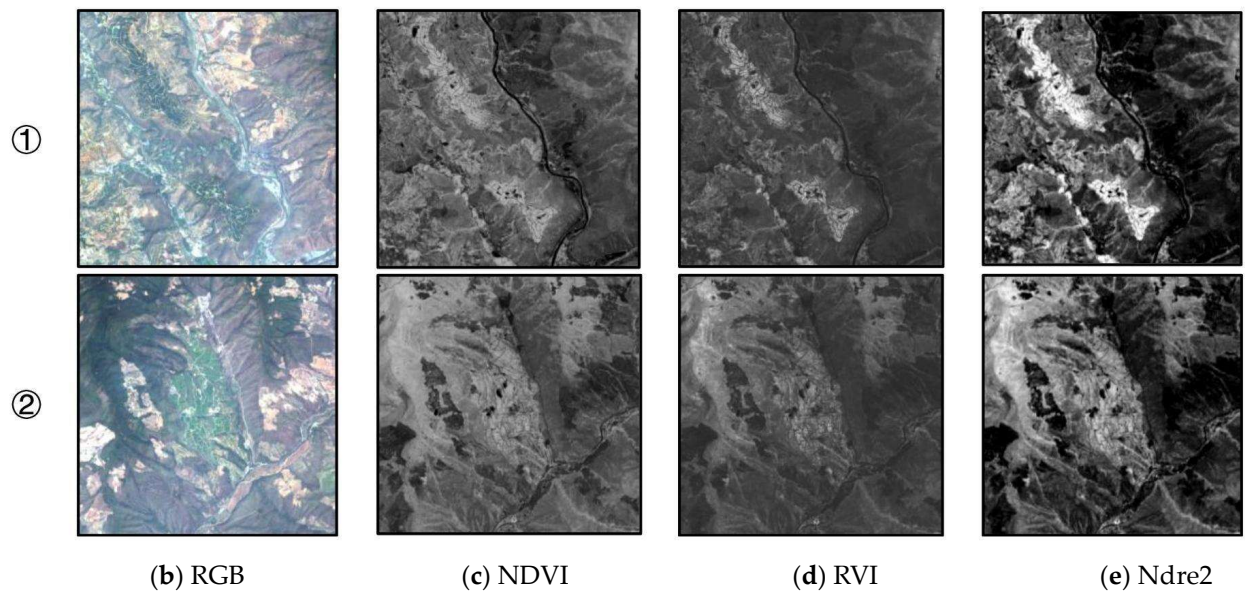


Figure 6. Local visualization of selected vegetation indices. The locations of the image blocks are shown by the black boxed lines.

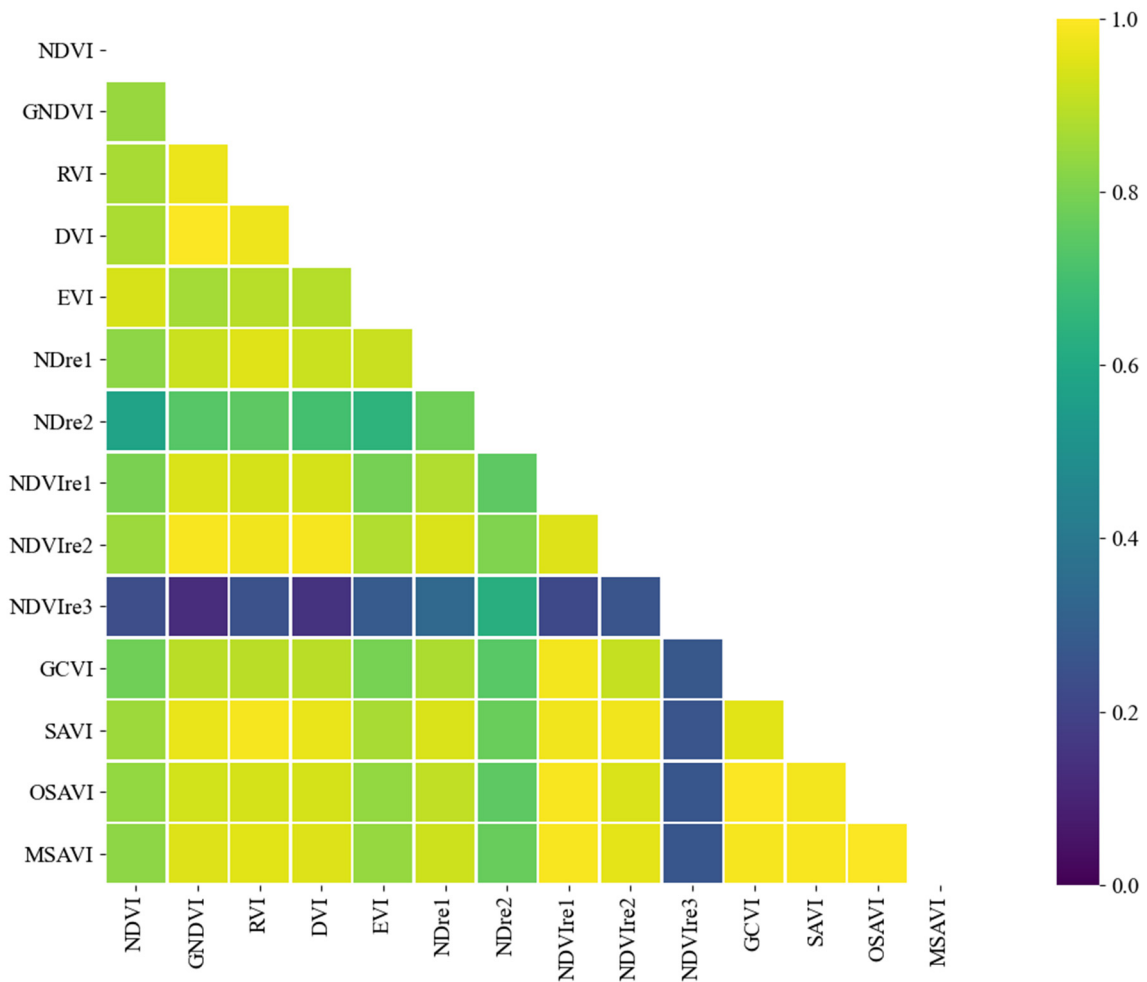


Figure 7. Pearson's correlation of features.

4.2. Comparison of Models

To assess the effectiveness and accuracy of the method used in this research, several classical semantic segmentation models are employed for comparison, namely, UNet [39], PSPNet [40], DeepLabv3 [32], DeepLabv3+ [41], and MANet [42]. The model is trained with the parameters mentioned in Section 3.4, using Resnet50 as the backbone of the semantic segmentation model. Table 2 shows the results of the citrus-growing regions extracted by different models.

Table 2. Comparison of different model results.

	Precision	IoU	F1-Score	Recall
PSPNet	90.31%	59.43%	74.56%	63.48%
MANet	86.52%	64.69%	78.75%	72.25%
DeepLabv3	89.81%	65.29%	79.01%	70.52%
DeepLabv3+	89.97%	62.73%	77.10%	67.45%
UNet	85.74%	68.43%	81.26%	77.22%
Ours	88.96%	73.22%	84.54%	80.55%

Note: Bold numbers indicate the highest values in each column.

The proposed method has the highest performance with an IoU of 73.22%, an F1-score of 84.54%, and a recall of 80.55%. Compared with UNet, the proposed model enhances the precision by 3.22%, the IoU by 4.79%, the F1-score by 3.28%, and the recall by 3.33%. In terms of IoU and recall, the proposed model outweighs the other five models by more than 3%. Although PSPNet receives the highest precision, the others have the lowest. Regarding the F1-score, PSPNet, MANet, DeepLabv3, and DeepLabv3+ do not reach 80%.

To visualize and analyze the extraction results of citrus-growing regions, the results of four image blocks from the test set are compared with the true-color images, false-color images, and ground-truth labels. The locations of the image blocks and a comparison of the extraction results of each model are shown in Figure 8, where the white area represents the citrus-growing regions, and the black color is the background. Figure 8 indicates that all models basically extract the citrus-growing regions with different degrees of commission and omission errors. Among all the models, the proposed method yields predictions that are closest to the real labels, improving the degree of the fragmentation of the boundary recognized by the UNet model, with a smaller error. By contrast, PSPNet and DeepLab V3 have the worst results in recognizing the boundaries of the citrus-growing regions; they cannot accurately identify the irregular boundaries, with some details lost. DeepLab V3+ improves the results on the boundary compared with the previous two models but has more serious commission and omission errors. The performance of MANet on the boundaries is similar to that of DeepLab V3+ with the same problem of commission and omission errors. Overall, the proposed method can identify the boundary most accurately and can fully utilize the spatial information to reduce the generation of commission and omission errors to improve the extraction accuracy.

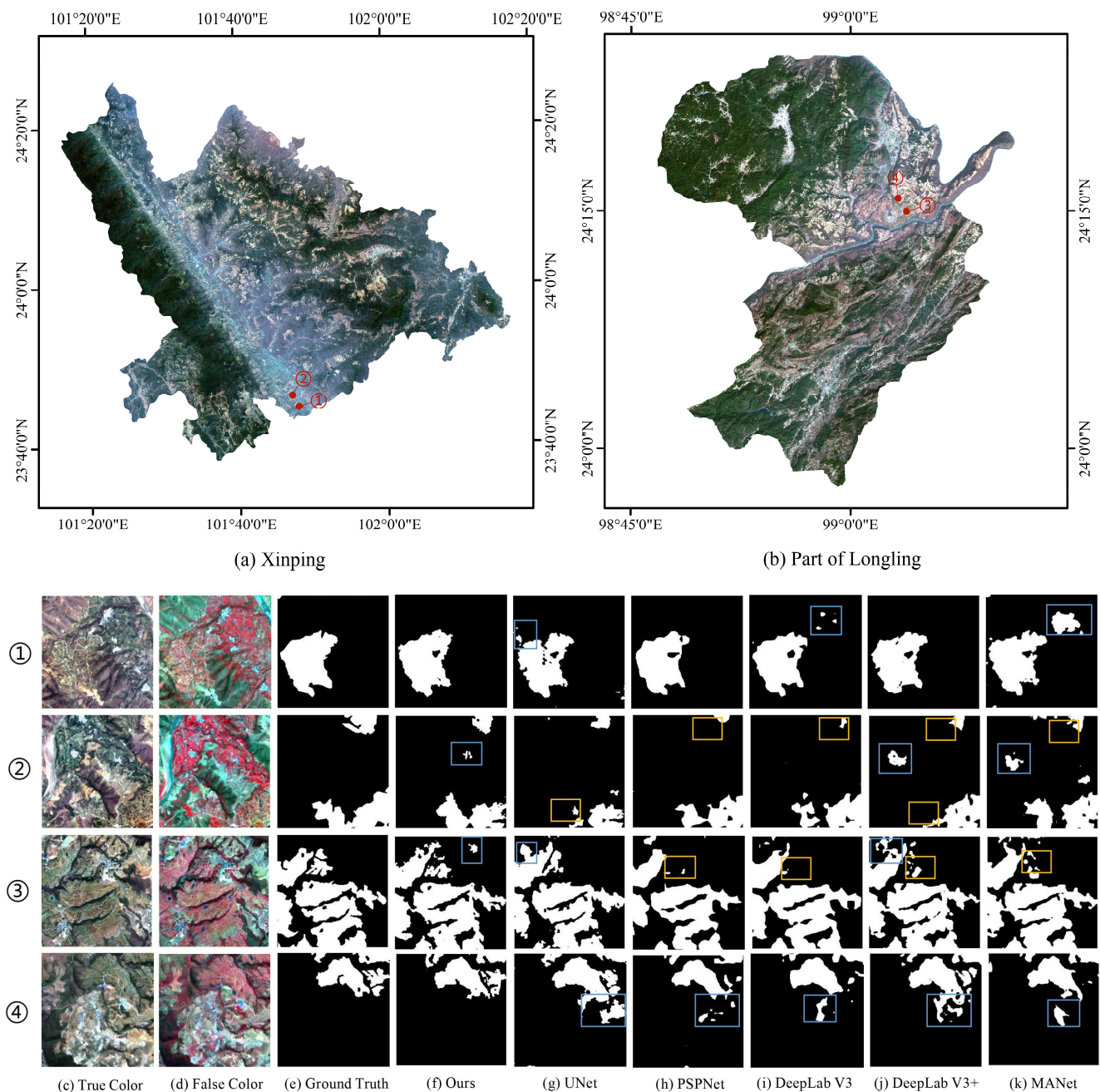


Figure 8. Overview map and comparison of extraction results of each model. The locations of the image blocks are shown by the red dots, the commission errors are shown by the blue boxed lines, while the omission errors are shown by the yellow boxed lines.

4.3. Ablation Experiments

In order to comprehensively evaluate the performance of the model, this study designs ablation experiments regarding the structure of model, the introduction of different modules and loss functions. Similarly, the results of the experiments are evaluated through the metrics of the test set.

4.3.1. Effectiveness for the Image Pyramid Structure and ASPP

In this study, ablation experiments are conducted using UNet as a benchmark to validate the effectiveness of the image pyramid structure and the ASPP module in enhancing the extraction of citrus-growing regions. Table 3 exhibits the results of the experiments,

from which the pyramid structure and ASPP module improve the extraction accuracy. Overall, the addition of the image pyramid structure improves the precision by 4.35%, the F1-score by 1.06% and the IoU by 1.53%, but the recall is not as good as that of the UNet. The IoU, the F1-score, and the recall of citrus-growing region extraction are improved by adding the ASPP module to the UNet and the pyramid UNet. The extraction of citrus-growing regions is also optimized via the addition of the image pyramid structure and the ASPP module to the model.

Table 3. Results of ablation experiments.

	Precision	IoU	F1-Score	Recall
UNet	85.74%	68.43%	81.26%	77.22%
UNet_ASPP	90.09%	72.92%	84.34%	79.28%
Pyramid UNet	91.74%	69.96%	82.32%	74.66%
Pyramid UNet_ASPP (ours)	88.93%	73.22%	84.54%	80.55%

Note: Bold numbers indicate the highest values in each column.

4.3.2. Comparison of Different Modules

The comparative results of basic receptive field block (BasicRFB) [43], simple spatial pyramid pooling fusion (SimSPPF) [44] and ASPP are exhibited in Table 4. The BasicRFB is derived from RFBNet, which is a combination of the multi-branch convolution and atrous convolution. The multi-branch convolution is mainly derived from GoogleNet [45] and the atrous convolution from the idea of ASPP. The SimSPPF module is initially introduced in YOLOv6, utilized to simplify the complexity of the model and reduce the inference time [46]. Based on the results, the image pyramid UNet with ASPP obtains the highest IoU, F1-score and recall, which are 73.22%, 84.54%, and 80.55%, respectively. Although the image pyramid UNet with SimSPPF has the greatest precision of 91.06%, the other three metrics are the lowest of these three models. Even though ASPP has the largest number of parameters among the three modules, which would increase the burden of model computation, ASPP is chosen to be added to the network in this study in order to obtain the most accurate citrus-growing regions extraction results.

Table 4. Results of module ablation experiments.

	Precision	IoU	F1-Score	Recall
Pyramid UNet BasicRFB	89.13%	66.22%	79.68%	72.04%
Pyramid UNet SimSPPF	91.06%	61.30%	76.01%	65.23%
Pyramid UNet_ASPP (ours)	88.96%	73.22%	84.54%	80.55%

Note: Bold numbers indicate the highest values in each column.

4.3.3. Comparison of Loss

During the study, different loss functions are selected for ablation experiments, including cross-entropy loss, weighted cross-entropy loss, dice loss [47], and IoU loss [48]. Table 5 shows the evaluation results, in which the three metrics received the highest scores when using weighted cross-entropy as a loss function. Although, dice loss obtains the highest accuracy in precision, several other metrics are inferior to weighted cross-entropy loss. Thus, the weighted cross-entropy loss is suitable for the purpose of this study.

Table 5. Results of loss ablation experiments.

	Precision	IoU	F1-Score	Recall
Cross-entropy Loss	89.51%	67.32%	80.47%	73.09%
Dice Loss	92.27%	59.67%	74.74%	62.81%
IoU Loss	92.01%	67.37%	80.51%	71.55%
Weighted Cross-entropy Loss	88.96%	73.22%	84.54%	80.55%

Note: Bold numbers indicate the highest values in each column.

4.4. Analysis of Results from Various Regions

The citrus-growing regions in the three study areas are extracted using the proposed model, as shown in Figure 9. Owing to the multiscale feature capture structure, the location and contours of citrus-growing regions in the study areas can be extracted accurately by the proposed model. Figure 9a shows the extraction results for the whole county of Xinping. Citrus-growing regions are mainly distributed on both sides of the river valley, with good extraction results. Citrus-growing regions with a small area and scattered distribution exist in the eastern county. The regional images in Figure 9b,c are the results of testing the model. The extraction of the test area in Longling County (Figure 9b) shows that the citrus-growing regions in this study area are concentrated in the river valley area, with only a few scattered distributions throughout the map. The extraction results in Figure 9c of the Binchuan test area demonstrate that more citrus-growing areas exist in the region, mainly in the east-central part of the country, with the area of citrus plots gradually decreasing to both sides. The proposed model obtains more reliable and complete extraction results on the regions that are not involved in the training model, which indicates that the model has a good generalization ability for various regions.

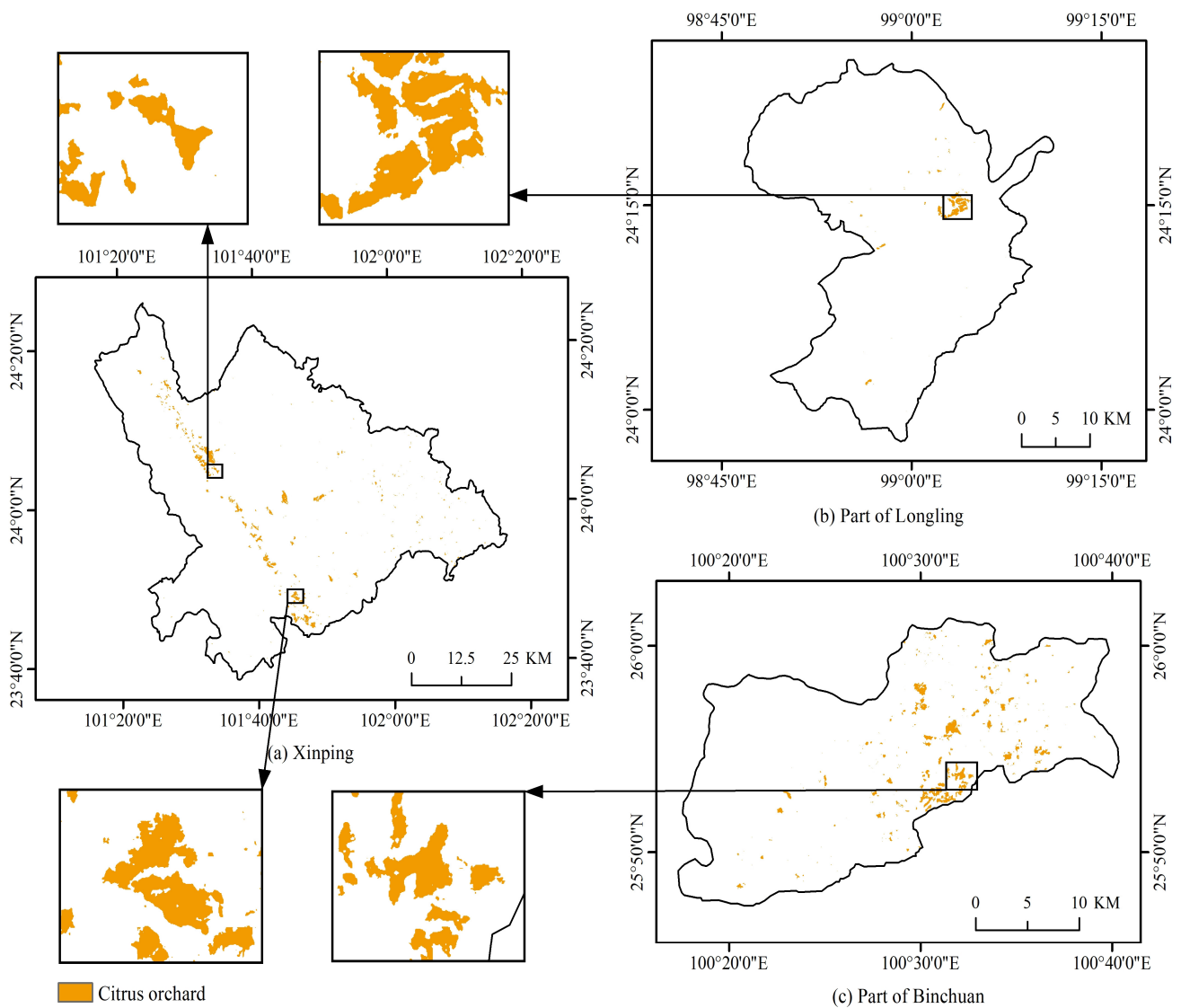


Figure 9. Extraction results for citrus-growing regions from different areas by the proposed model.

4.5. Comparison with the Method of Object-Oriented RF

To verify the advantages of deep learning models for contextual information utilization, the object-oriented RF algorithm is chosen to compare with the proposed model. The training samples for the object-oriented RF are all from Xiping County imagery, and no further samples are added from the test dataset in Longling and Binchuan. According to the field survey and the interpretation on Google Maps, 2372 background and 530 citrus samples are generated. The used spectral bands and indices of the object-oriented RF are consistent with those of the proposed model. The proposed method outperforms the object-oriented RF method, as shown in Table 6. It achieves greater than 85% for all evaluation metrics, whereas the kappa and UA of the object-oriented RF cannot reach 50%.

Table 6. Results of the object-oriented RF and the proposed method. UA is the user accuracy, PA is the producer accuracy, and OA is the overall accuracy.

	Ours	Object-Oriented RF
OA	97.28%	87.04%
Kappa	0.9051	0.4649
PA	97.27%	78.31%
UA	87.55%	40.19%

Note: Bold numbers indicate the highest values in each row.

Figure 10 shows the classification results of the object-oriented RF for citrus-growing regions. The distribution trend in Figure 10a is roughly the same as that in Figure 9a, and the citrus plots are uniformly dispersed in the river valley area. Compared with those in Figure 9b,c, the detected citrus-growing regions in Figure 10b,c are significantly fewer and not clustered, with smaller plot sizes. Local zoomed-in views of the extraction results are also shown in Figures 9 and 10. The proposed model obtains more complete results in the local areas in Figure 9a, whereas the object-oriented method has more mixed results in Figure 10a, making the specific boundaries of the planting regions difficult to identify. The citrus orchard in Figure 9b is fully identified, whereas the citrus orchard in Figure 10b is only partially extracted. For the region in Figure 10c, the object-oriented method cannot recognize the citrus-growing regions at all. The object-oriented classification is a process of segmentation followed by classification, and the objects with irregularly structured features are difficult to segment [49].

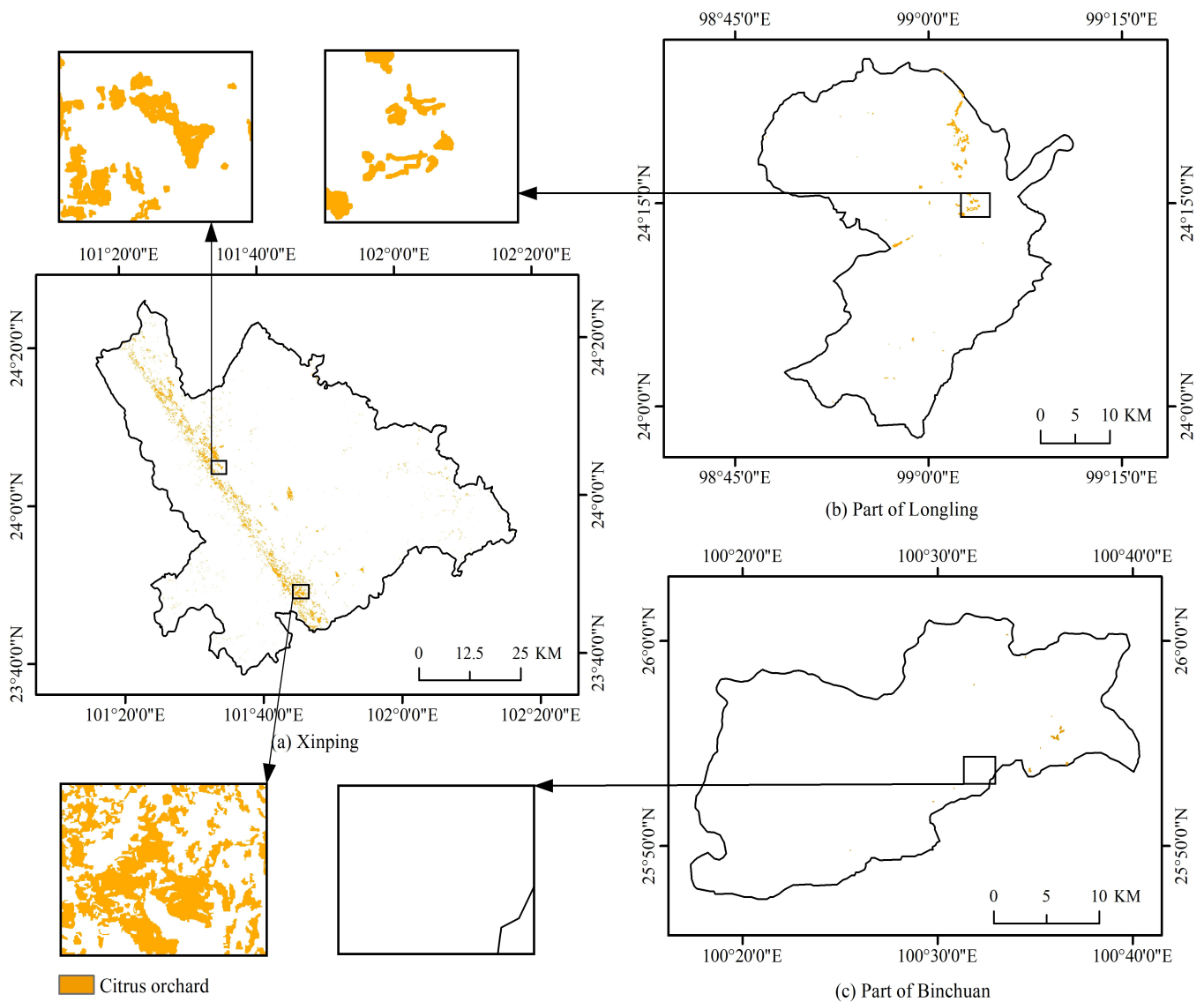


Figure 10. Extraction results for citrus-growing regions from different areas by the object-oriented RF.

5. Discussion

5.1. Validity of the Proposed Model

Deep learning has achieved significant results in various fields. However, specific modifications must be made to improve results to implement deep learning in different domains [15]. Currently, most deep learning-based semantic segmentation techniques (e.g., computer vision) utilize only the RGB bands of images. Remote sensing images contain rich spectral information beneficial for distinguishing crops, but further research is needed to utilize this information [19,50]. In this study, nine bands from Sentinel-2 satellite images are employed to create three vegetation indices, which are selected using the Relief F algorithm. As inspired by [34], the trunk part of the proposed model can acquire global information and maintain the integrity of the extracted region boundaries, and the branch part can retain and complement detailed local information, reducing the confusion and errors.

The study area consists of other vegetation and crops that are likely to pose difficulties for the extraction of citrus-growing regions. The citrus-growing regions generally have clear boundaries and a bigger area than other surrounding crops. The proposed method has the structure of detail learning that helps differentiate other crops. The idea of combining this model with the ASPP module is inspired by [51,52], and experiments confirm that this combination is effective. The comparison shown in Figure 8f,g demonstrates that the

proposed model can distinguish the borders of citrus-growing regions. Overall, a combination of modules improves the model to some extent. It is able to fulfill the requirement of providing a fast extraction of citrus-growing regions with optimal results.

5.2. Extraction Model Transferability

The proposed model has acceptable spatial transferability. When transferring the model to Longling for the extraction of citrus-growing regions without adding extra training data to fine-tune the model, it can be found in Figure 9b that the citrus-growing regions are intact with very little confusion. In fact, the citrus orchard in Longling and some citrus orchards in Xinping are planted by the same company, which may make the citrus orchards in the two regions have some similarity in planting structure. This result demonstrates that this model is able to learn the planting structure of citrus planting regions.

However, the results in the Binchuan test area are fragmented, which can be attributed to two possible reasons. First, the differences in the images acquired at different times affects the results. Second, the citrus orchards in Binchuan are located in the suburbs, and the images are influenced by the local climate and soil conditions [5]. The planting structure is related to the neighboring crops of the local citrus orchards, which is different from Xinping areas. As shown by the results in [53], the context information for each domain is variable and unique. Therefore, the model can be fine-tuned with data to have better transferability. The inclusion of multi-source data in the future is also expected to uncover correlation features between different regions.

5.3. Computational Costs of Extraction Models

The number of parameters of all the models in the experiments is shown in Figure 11. It can be seen that the parameters of the proposed model are relatively large with a higher training cost [54]. However, the proposed model uses pre-training parameters that can make the model converge faster. Since a lightweight high-accuracy model would be more practical in production [55], the model will be light-weighted with guaranteed accuracy to reduce training costs in future works.

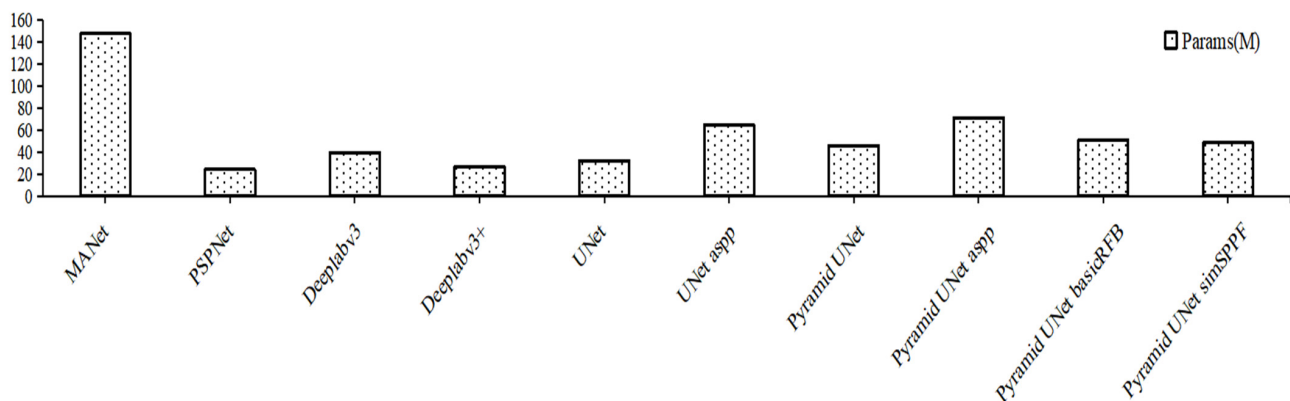


Figure 11. Number of parameters for all models.

6. Conclusions

This work focuses on the difficulty of extracting citrus-growing regions with few available multispectral images and irregular planting structures. An improved model for citrus-growing region extraction based on UNet, combining the image pyramid and the ASPP module, is proposed. The spectral bands and indices selected from Sentinel-2 satellite images are input into the model. This model consists of two parts, namely, the trunk and branch parts. The trunk part is employed to increase VRF and obtain multiscale spatial information, and the branch part can introduce detailed information to reduce the errors in extracting citrus-growing regions. To confirm the effectiveness and strength of the model, commonly used models, such as UNet, PSPNet, DeepLabv3, DeepLabv3+,

MANet, and object-oriented RF models, are employed for the comparison of extracting citrus-growing regions. By comparison with other classical semantic segmentation models, the proposed model obtains the optimal extraction accuracy with a precision of 88.96%, an IoU of 73.22%, a recall of 80.55%, and an F1-score of 84.54%, owing to the designed spatial learning structure. The extraction results in this study have more complete boundaries with fewer commission and omission errors. For various areas and images, this model has a generalization capability to obtain more complete results. Moreover, by comparison with the object-oriented RF, the object-oriented RF model is prone to make misclassifications in the training areas and perform poorly in test areas, whereas the proposed model can obtain smooth and accurate segmentation results in different areas. This result confirms that the proposed method is robust and effective for the automatic extraction of citrus-growing regions. However, this study has some problems that need further investigation. The results are impacted by the discrepancies between the images that are obtained at various times, as well as by the nearby plants that have spectral features and planting structures that are similar to citrus plants. Future studies will solve the problem of distinguishing similar crops from citrus orchards by trying more spectral combinations and adding multisource data, such as radar data, topographic data, or environmental information.

Author Contributions: W.L. conceived and designed the experiment; Y.L. revised and improved the manuscript; Y.G. supervised the research implementation; and S.Y., T.Z. and X.L. analyzed the data. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 41977394); the National Key Research and Development Program of China (Grant No. 2023YFE0207900); the Major Project of Science and Technology of Yunnan Province (Grant No. 202002AE090010).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, Z.; Zhang, L.; Zhao, J.-F.; Zhang, X.-K.; Wang, Y.; Li, T.-S.; Zhang, W.; Zhou, Y. New Geographic Distribution and Molecular Diversity of Citrus Chlorotic Dwarf-Associated Virus in China. *J. Integr. Agric.* **2022**, *21*, 293–298. [[CrossRef](#)]
2. Tang, P.; Du, P.; Xia, J.; Zhang, P.; Zhang, W. Channel Attention-Based Temporal Convolutional Network for Satellite Image Time Series Classification. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
3. Csillik, O.; Cherbini, J.; Johnson, R.; Lyons, A.; Kelly, M. Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* **2018**, *2*, 39. [[CrossRef](#)]
4. Wei, P.; Ye, H.; Qiao, S.; Liu, R.; Nie, C.; Zhang, B.; Song, L.; Huang, S. Early Crop Mapping Based on Sentinel-2 Time-Series Data and the Random Forest Algorithm. *Remote Sens.* **2023**, *15*, 3212. [[CrossRef](#)]
5. Morell-Monzó, S.; Sebastiá-Frasquet, M.-T.; Estornell, J.; Moltó, E. Detecting Abandoned Citrus Crops Using Sentinel-2 Time Series. *A Case Study in the Comunitat Valenciana. ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 54–66. [[CrossRef](#)]
6. Yang, Y.J.; Zhan, Y.L.; Tian, Q.J.; Wang, L.; Wang, P.Y.; Zhang, W.M. Winter Wheat Extraction Using Curvilinear Integra Of GF-1 NDVI Time Series. In Proceedings of the 36th IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3174–3177.
7. Zhang, M.; Li, Q.Z.; Wu, B.F. Investigating the Capability of Multi-Temporal Landsat Images for Crop Identification in High Farmland Fragmentation Regions. In Proceedings of the 1st International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Shanghai, China, 2–4 August 2012; pp. 26–29.
8. Di, W.; Zhou, Q.B.; Yan, S.; Chen, Z.X. Advances in Research on Crop Identification Using SAR. In Proceedings of the Fourth International Conference on Agro Geoinformatics, Istanbul, Turkey, 20–24 July 2015; pp. 312–317.
9. Zhang, R.; Tang, Z.; Luo, D.; Luo, H.; You, S.; Zhang, T. Combined Multi-Time Series SAR Imagery and InSAR Technology for Rice Identification in Cloudy Regions. *Appl. Sci.* **2021**, *11*, 6923. [[CrossRef](#)]
10. Chen, L.; Chien, T.-W.; Hsu, C.-S.; Tan, C.-H.; Hsu, H.-Y.; Kou, C.-H. Water Requirement for Irrigation of Complicated Agricultural Land by Using Classified Airborne Digital Sensor Images. *J. Indian Soc. Remote Sens.* **2019**, *47*, 1307–1314. [[CrossRef](#)]
11. Zhang, X.; Sun, Y.; Shang, K.; Zhang, L.; Wang, S. Crop Classification Based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4117–4128. [[CrossRef](#)]
12. Luo, C.; Qi, B.; Liu, H.; Guo, D.; Lu, L.; Fu, Q.; Shao, Y. Using Time Series Sentinel-1 Images for Object-Oriented Crop Classification in Google Earth Engine. *Remote Sens.* **2021**, *13*, 561. [[CrossRef](#)]
13. Du, M.; Huang, J.; Wei, P.; Yang, L.; Chai, D.; Peng, D.; Sha, J.; Sun, W.; Huang, R. Dynamic Mapping of Paddy Rice Using Multi-Temporal Landsat Data Based on a Deep Semantic Segmentation Model. *Agronomy* **2022**, *12*, 1583. [[CrossRef](#)]

14. Gadiraju, K.K.; Ramachandra, B.; Chen, Z.; Vatsavai, R.R. Multimodal Deep Learning Based Crop Classification Using Multispectral and Multitemporal Satellite Imagery. In Proceedings of the KDD '20: The 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, 23–27 August 2020; pp. 3234–3242.
15. Bian, Y.; Li, L.; Jing, W. CACPU-Net: Channel Attention U-Net Constrained by Point Features for Crop Type Mapping. *Front. Plant Sci.* **2023**, *13*, 1030595. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, T.; Hu, D.; Wu, C.; Liu, Y.; Yang, J.; Tang, K. Large-Scale Apple Orchard Mapping from Multi-Source Data Using the Semantic Segmentation Model with Image-to-Image Translation and Transfer Learning. *Comput. Electron. Agric.* **2023**, *213*, 108204. [[CrossRef](#)]
17. Clark, A.; McKechnie, J. Detecting Banana Plantations in the Wet Tropics, Australia, Using Aerial Photography and U-Net. *Appl. Sci.* **2020**, *10*, 2017. [[CrossRef](#)]
18. Niu, Z.; Deng, J.; Zhang, X.; Zhang, J.; Pan, S.; Mu, H. Identifying the Branch of Kiwifruit Based on Unmanned Aerial Vehicle (UAV) Images Using Deep Learning Method. *Sensors* **2021**, *21*, 4442. [[CrossRef](#)] [[PubMed](#)]
19. Li, X.; Tian, J.; Li, X.; Wang, L.; Gong, H.; Shi, C.; Nie, S.; Zhu, L.; Chen, B.; Pan, Y.; et al. Developing a Sub-Meter Phenological Spectral Feature for Mapping Poplars and Willows in Urban Environment. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 77–89. [[CrossRef](#)]
20. Culman, M.; Rodríguez, A.C.; Wegner, J.D.; Delalieux, S.; Somers, B. Deep Learning for Sub-Pixel Palm Tree Classification Using Spaceborne Sentinel-2 Imagery. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXIII, Online Only, Spain, 3–18 September 2021.
21. da Costa, L.B.; de Carvalho, O.L.F.; de Albuquerque, A.O.; Gomes, R.A.T.; Guimarães, R.F.; de Carvalho, O.A. Deep Semantic Segmentation for Detecting Eucalyptus Planted Forests in the Brazilian Territory Using Sentinel-2 Imagery. *Geocarto Int.* **2022**, *37*, 6538–6550. [[CrossRef](#)]
22. Xue, Z.; Qian, S. Two-Stream Translating LSTM Network for Mangroves Mapping Using Sentinel-2 Multivariate Time Series. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [[CrossRef](#)]
23. Li, Y.; Wu, T.; Ge, Y.; Xi, S.; Zhang, T.; Zhang, W. Semi-Supervised Cooperative Regression Model for Small Sample Estimation of Citrus Leaf Nitrogen Content with UAV Images. *Int. J. Remote Sens.* **2023**, *44*, 7237–7262. [[CrossRef](#)]
24. Otunga, C.; Odindi, J.; Mutanga, O.; Adjorlolo, C. Evaluating the Potential of the Red Edge Channel for C3 (*Festuca* spp.) Grass Discrimination Using Sentinel-2 and Rapid Eye Satellite Image Data. *Geocarto Int.* **2019**, *34*, 1123–1143. [[CrossRef](#)]
25. Kononenko, I. Estimating Attributes: Analysis and Extensions of RELIEF. In Proceedings of the Machine Learning: ECML-94, Catania, Italy, 6–8 April 1994; pp. 171–182.
26. Pan, Y.; Jiang, J.; Liu, Z.; Du, Y.; Xiong, K. Identification of Vegetation Under Natural Gas Leakage by Spectral Index Based on Feature Selection. *Int. J. Remote Sens.* **2022**, *43*, 3082–3105. [[CrossRef](#)]
27. Huang, W.; Guan, Q.; Luo, J.; Zhang, J.; Zhao, J.; Liang, D.; Huang, L.; Zhang, D. New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2516–2524. [[CrossRef](#)]
28. Xu, G.; Wang, Y.; Wang, L.; Soares, L.P.; Grohmann, C.H. Feature-Based Constraint Deep CNN Method for Mapping Rainfall-Induced Landslides in Remote Regions With Mountainous Terrain: An Application to Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2644–2659. [[CrossRef](#)]
29. Hu, X.; Huang, C.; Mei, H.; Zhang, H. Landslide Susceptibility Mapping Using an Ensemble Model of Bagging Scheme and Random Subspace-Based Naïve Bayes Tree in Zigui County of the Three Gorges Reservoir Area, China. *Bull. Eng. Geol. Environ.* **2021**, *80*, 5315–5329. [[CrossRef](#)]
30. Zeng, Z.; Fan, C.; Xiao, L.; Qu, X. DEA-UNet: A Dense-Edge-Attention UNet Architecture for Medical Image Segmentation. *J. Electron. Imaging* **2022**, *31*, 043032. [[CrossRef](#)]
31. Zhang, J.; Wang, S.; Qiu, J.; Pan, X.; Zou, J.; Duan, Y.; Pan, Z.; Li, Y. A fast X-shaped foreground segmentation network with CompactASPP. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104077. [[CrossRef](#)]
32. 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](#)]
33. Lu, X.Y.; Zhong, Y.F.; Zhao, J. Multi-Scale Enhanced Deep Network for Road Detection. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 3947–3950.
34. Ding, L.; Zhang, J.; Bruzzone, L. Semantic Segmentation of Large-Size VHR Remote Sensing Images Using a Two-Stage Multiscale Training Architecture. *IEEE Trans. Geosci. Remote* **2020**, *58*, 5367–5376. [[CrossRef](#)]
35. Abadal, S.; Salgueiro, L.; Marcello, J.; Vilaplana, V. A Dual Network for Super-Resolution and Semantic Segmentation of Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 4547. [[CrossRef](#)]
36. Wenger, R.; Puissant, A.; Weber, J.; Idoumghar, L.; Forestier, G. Multimodal and Multitemporal Land Use/Land Cover Semantic Segmentation on Sentinel-1 and Sentinel-2 Imagery: An Application on a MultiSenGE Dataset. *Remote Sens.* **2023**, *15*, 151. [[CrossRef](#)]
37. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. [[CrossRef](#)]
38. Takahashi, K.; Yamamoto, K.; Kuchiba, A.; Koyama, T. Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Appl. Intell.* **2022**, *52*, 4961–4972. [[CrossRef](#)] [[PubMed](#)]

39. Navab, N.; Hornegger, J.; Wells, W.M.; Frangi, A.F. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
40. Zhao, H.S.; Shi, J.P.; Qi, X.J.; Wang, X.G.; Jia, J.Y. Pyramid Scene Parsing Network. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
41. Chen, L.C.E.; Zhu, Y.K.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851.
42. Fan, T.; Wang, G.; Li, Y.; Wang, H. MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. *IEEE Access* **2020**, *8*, 179656–179665. [[CrossRef](#)]
43. Liu, S.; Huang, D. Receptive Field Block Net for Accurate and Fast Object Detection. In Proceedings of the European conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
44. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A Single-stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
45. Li, M.; Zhou, G.; Chen, A.; Li, L.; Hu, Y. Identification of tomato leaf diseases based on LMBRNet. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106195. [[CrossRef](#)]
46. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8. *Sensors* **2023**, *23*, 8361. [[CrossRef](#)] [[PubMed](#)]
47. Wang, L.; Wang, C.; Sun, Z.; Chen, S. An Improved Dice Loss for Pneumothorax Segmentation by Mining the Information of Negative Areas. *IEEE Access* **2020**, *8*, 167939–167949. [[CrossRef](#)]
48. Wu, S.; Yang, J.; Wang, X.; Li, X. IoU-Balanced Loss Functions for Single-Stage Object Detection. *Pattern Recognit. Lett.* **2022**, *156*, 96–103. [[CrossRef](#)]
49. Cheng, G.; Li, Z.; Yao, X.; Guo, L.; Wei, Z. Remote Sensing Image Scene Classification Using Bag of Convolutional Features. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1735–1739. [[CrossRef](#)]
50. Ulku, I.; Akagundüz, E.; Ghamisi, P. Deep Semantic Segmentation of Trees Using Multispectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7589–7604. [[CrossRef](#)]
51. Li, Q.; Jia, W.; Sun, M.; Hou, S.; Zheng, Y. A Novel Green Apple Segmentation Algorithm Based on Ensemble U-Net under Complex Orchard Environment. *Comput. Electron. Agric.* **2021**, *180*, 105900. [[CrossRef](#)]
52. Zuo, X.; Lin, H.; Wang, D.; Cui, Z. A Method of Crop Seedling Plant Segmentation on Edge Information Fusion Model. *IEEE Access* **2022**, *10*, 95281–95293. [[CrossRef](#)]
53. Zheng, J.; Yuan, S.; Wu, W.; Li, W.; Yu, L.; Fu, H.; Coomes, D. Surveying Coconut Trees Using High-Resolution Satellite Imagery in Remote Atolls of the Pacific Ocean. *Coord. Chem. Rev.* **2023**, *481*, 113485. [[CrossRef](#)]
54. Cai, X.; Jing, Q.; Peng, B.; Zhang, Y.; Wang, Y.; Tang, J. Automatic Traffic State Recognition Based on Video Features Extracted by an Autoencoder. *Math. Probl. Eng.* **2022**, *2022*, 2850111. [[CrossRef](#)]
55. Chen, S.; Yao, H.; Qiao, F.; Ma, Y.; Wu, Y.; Lu, J. Vehicles Driving Behavior Recognition Based on Transfer Learning. *Expert Syst. Appl.* **2023**, *213*, 119254. [[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.