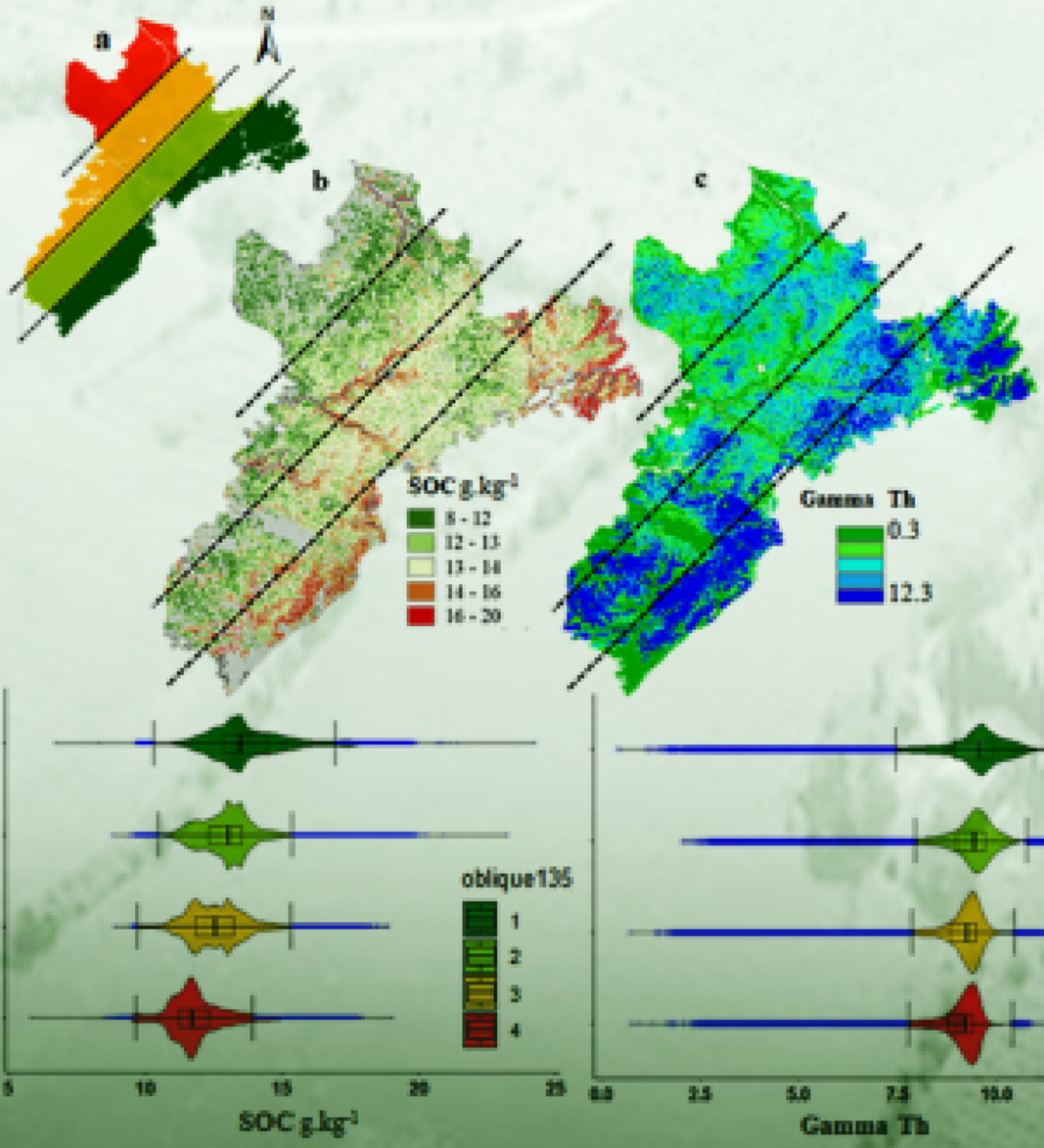
remote sensing

IMPACT FACTOR 5.349

CITESCORE 7.4



Sentinel-1/2 Bare Soil Temporal Mosaics for Soil Organic Carbon Content Mapping

Volume 15 · Issue 9 | May (I) 2023

MDPI

mdpi.com/journal/remotesensing
ISSN 2072-4292

remote sensing

IMPACT FACTOR 5.0

CITESCORE 7.9

Article

Cropland Extraction in Southern China from Very High-Resolution Images Based on Deep Learning

Dehua Xie, Han Xu, Xiliu Xiong, Min Liu, Haoran Hu, Mengsen Xiong and Luo Liu

Special Issue

Monitoring Agricultural Land-Use Change and Land-Use Intensity II

Edited by

Dr. Luo Liu, Dr. Yuanwei Qin, Dr. Bingwen Qiu, Dr. Qiangyi Yu and Dr. Zhi Qiao

MDPI

https://doi.org/10.3390/rs15092231

remote sensing

MDPI

Article

Cropland Extraction in Southern China from Very High-Resolution Images Based on Deep Learning

Dehua Xie [1], Han Xu [1], Xiliu Xiong [2], Min Liu [1], Haoran Hu [1], Mengsen Xiong [1] and Luo Liu [1,*]

[1] Guangdong Provincial Key Laboratory of Land Use and Consolidation, South China Agricultural University, Guangzhou 510642, China; huazai_heng_7@stu.scau.edu.cn (D.X.); 20202058003@stu.scau.edu.cn (H.X.); lm0803@stu.scau.edu.cn (M.L.); huhaoran@stu.scau.edu.cn (H.H.); 18538067289@stu.scau.edu.cn (M.X.)
[2] Institute of Ecological Environment Protection, Guangxi Eco-Engineering Vocational and Technical College, Liuzhou 545004, China; xiongxl@stu.scau.edu.cn
* Correspondence: liuluo@scau.edu.cn

Abstract: Accurate cropland information is crucial for the assessment of food security and the formulation of effective agricultural policies. Extracting cropland from remote sensing imagery is challenging due to spectral diversity and mixed pixels. Recent advances in remote sensing technology have facilitated the availability of very high-resolution (VHR) remote sensing images that provide detailed ground information. However, VHR cropland extraction in southern China is difficult because of the high heterogeneity and fragmentation of cropland and the insufficient observations of VHR sensors. To address these challenges, we proposed a deep learning-based method for automated high-resolution cropland extraction. The method used an improved HRRS-U-Net model to accurately identify the extent of cropland and explicitly locate field boundaries. The HRRS-U-Net maintained high-resolution details throughout the network to generate precise cropland boundaries. Additionally, the residual learning (RL) and the channel attention mechanism (CAM) were introduced to extract deeper discriminative representations. The proposed method was evaluated over four city-wide study areas (Qingyuan, Yangjiang, Guangzhou, and Shantou) with a diverse range of agricultural systems, using GaoFen-2 (GF-2) images. The cropland extraction results for the study areas had an overall accuracy (OA) ranging from 97.00% to 98.33%, with F1 scores (F1) of 0.830–0.940 and Kappa coefficients (Kappa) of 0.814–0.929. The OA was 97.85%, F1 was 0.915, and Kappa was 0.901 over all study areas. Moreover, our proposed method demonstrated advantages compared to machine learning methods (e.g., RF) and previous semantic segmentation models, such as U-Net, U-Net++, U-Net3+, and MPSPNet. The results demonstrated the generalization ability and reliability of the proposed method for cropland extraction in southern China using VHR remote images.

Keywords: cropland extraction; very high-resolution; GF-2; deep learning; southern China

Citation: Xie, D.; Xu, H.; Xiong, X.; Liu, M.; Hu, H.; Xiong, M.; Liu, L. Cropland Extraction in Southern China from Very High-Resolution Images Based on Deep Learning. *Remote Sens.* **2023**, *15*, 2231. https://doi.org/10.3390/rs15092231

Academic Editor: Jochem Verrelst

Received: 29 March 2023
Revised: 17 April 2023
Accepted: 20 April 2023
Published: 23 April 2023

1. Introduction

Cropland encompasses all agricultural land, including permanently cultivated land, newly cultivated land, fallow land, and grassland-farming rotating land [1]. It provides the largest share of the global food supply, constituting 90% of food calories and 80% of protein and fats [2]. Accurate information regarding the extent and location of cropland is a fundamental data requirement for various agricultural applications, such as field area estimation, crop yield prediction, and understanding of the spatiotemporal patterns of cropland changes [3,4]. These applications have essential roles in the assessment of food security and formulation of effective agricultural policies [5]. In Guangdong, the available cropland has continuously decreased because of rapid urbanization in recent decades [6,7]. Between 2010 and 2020, the population increased by 21.04%, whereas the cropland area decreased by 25.02%, resulting in a 38.05% decrease in cropland area per capita. Currently, land surveys in Guangdong still heavily rely on manual field surveys and

visual interpretation. These approaches are subjective, time-consuming, and labor-intensive over large areas. Thus, there is an urgent need to develop an automatic and efficient method for cropland extraction.

Remotely sensed observations provide accurate and cost-effective solutions for agricultural land-use mapping and monitoring [8] and have long been efficient tools for the identification and assessment of cropland extent and distribution at local, regional, and global scales [9–11]. For a long time, land-use mapping has generally relied on satellite data with a high temporal frequency and coarse spatial resolution, particularly over large areas [12], for example, the Global Land Cover 2000 (GLC2000) 1000-m resolution dataset [13], Moderate Resolution Imaging Spectroradiometer Land Cover (MODISLC) 500-m dataset [14,15], Copernicus Global Land Service (CGLS) Land Cover 100-m dataset [16], Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) [17], China land cover dataset (CLCD) at 30-m spatial resolution [18], and Dynamic World, Near Real-time Global 10-m Land Use Land (Dynamic World) [19]. However, the cropland field size is very small (<0.04 ha) in most areas of southern China [1]. The available datasets have low (250 m–1.5 km), moderate (40.0–249.9 m), and medium (10.0–39.9 m) spatial resolution [20] and hence are not sufficiently fine to accurately delineate cropland consisting of small parcels of land with fragmented distributions [21]. Therefore, these datasets generally have low accuracies and vast inconsistencies [22,23], resulting in the poor assessment of global and local food security scenarios [24].

Fortunately, advances in remote sensing technology have facilitated the availability of high-resolution (5.0–9.9 m) (e.g., RapidEye, GF-1) and very high-resolution (VHR) (<5.0 m) [20] very high-resolution (VHR) satellite images (e.g., GeoEye-1, QuickBird-2, WorldView-1, and GF-2), which provide more detailed ground information to resolve fuzzy boundaries and small cropland parcels [25]. Specifically, the GF-2 satellite is the first civil optical remote sensing satellite with better than 1-meter spatial resolution developed independently by China. Although the available data from this satellite are new, many studies have investigated its performance in cropland classification [25–29]. However, the abundance of information in VHR images has caused high variation between the same classes and subtle variation between the different classes [30]. Gardens, grassland, and bare land have spectral and textural signatures that are similar to the signatures of cropland. In southern China, most agricultural landscapes are dominated by smallholder farms that constitute a mosaic of small fields with a diverse range of crops [1]. Small fields are often interspersed with other land uses, leading to unclear boundaries between cropland and other land uses. Furthermore, the high spatiotemporal heterogeneity of cropland contributes to classification challenges. Burning, manual land clearing and preparation, heterogeneous management practices and labor inputs, and the presence of shade trees and shelters all result in diverse crop types and cover densities [31]. Additionally, some croplands adopt an intercropping practice, which both increases the variation within the fields and hinders the delineation of field boundaries. In addition to the challenges presented by the high intra-class spectral variance between and within fields, the rapid temporal dynamics create obstacles for accurate cropland extraction. The spectral reflectance of crops varies throughout the growing season because of distinct biological characteristics [32]. Moreover, additional management practices (e.g., field preparation and fallow periods before seeding and after harvesting) can increase the intra-class variance of the spectral-temporal signal [33]. Thus, robustness and transferability are essential for cropland extraction methods in southern China.

Previous studies have explored various methods to identify cropland information. These methods can generally be divided into three categories. First, some studies used statistical methods, such as K-nearest neighbors [34], support vector machine [35], decision trees [36], and random forest (RF) [37]. These methods extract low-level features (e.g., spectra, texture, and geometry) and use minimal semantic information [38]. However, because of the highly variable spectral signature of cropland and the spectral confusion in VHR images, low-level features are less effective, resulting in difficulties in transferring the

method both temporally and spatially [39,40]. Second, some methods used time-series data for cropland classification using threshold-based algorithms. For example, Dong et al. [41] and Guo et al. [42] presented phenology-based approaches to map the paddy rice planting area and cropping intensity. Despite achieving good performance, these methods were susceptible to uncertainty and thus required numerous high-quality images. Unfortunately, VHR sensors make insufficient observations because of the long revisit intervals and poor observational conditions (cloudy and frequent precipitation) in southern China [43]. The third category is the latest popular deep learning-based method, which can hierarchically learn more representative and discriminative features than existing techniques [44,45]. As the most effective deep learning architecture for semantic segmentation, convolutional neural networks (CNNs) [46] can autonomously extract contextual associations and learn abstract features in the image without reliance on manually designed features [47,48]. For example, Liu et al. [29] used U-Net to identify cropland and effectively reduced the salt-and-pepper phenomenon of conventional methods. Zhang et al. [25] utilized the modified PSPNet (MPSPNet) to extract cropland from different agricultural systems at a large scale, demonstrating excellent generalizability and transferability.

However, these methods are hindered by the loss of spatial information and the inefficient utilization of spectral information. CNN-based models use down-sampling operations to capture long-term dependency information, but they lose important spatial details [49]. Although the generated abstract feature maps are subsequently up-sampled to the original resolution in the decoder, boundary information cannot be explicitly determined because the up-sampled representations are essentially pseudo-high-resolution [50]. This property severely reduces the classification accuracy of fragmented small fields, which are common in most areas of southern China. Additionally, CNNs can learn the spatial dependency of neighboring pixels, but they have struggled to capture correlations between adjacent spectra [51]. Considering the substantial intra-class variability and extra-class similarity of cropland in VHR images, the extracted results are subject to many commission and omission errors without sufficient exploitation of spectral information.

Many efforts have been made to improve the performances of CNN-based models. For example, the High-Resolution Network (HRNet) [52] maintains the high-resolution details and accordingly learns semantically strong and spatially precise representations. However, the massive additional convolutional modules in HRNet both substantially increase the computational cost and present gradient problems. To address the degradation problem of deep networks, He et al. [53] proposed the Deep Residual Network (ResNet), which introduced a residual learning framework. Recently, attention mechanisms have attracted increasing interest in the deep learning community. These mechanisms simulate the manner in which humans understand and perceive images, and they can facilitate the rapid and accurate acquisition of essential features [54]. For example, Hu et al. [55] developed the Squeeze-and-Excitation Network (SENet), which uses the channel attention mechanism (CAM) to adaptively recalibrate the weight of each channel, thereby increasing sensitivity to representative features. The CAM can mitigate the effect of small differences between classes and large differences within classes [56] and has been demonstrated to be an effective method to improve CNN performance [57].

In this study, we developed a deep learning-based method for accurate cropland extraction in southern China from VHR images using an improved CNN model. The improved model was built on the end-to-end U-Net architecture [58] and fully exploited state-of-the-art algorithms, including HRNet, ResNet, and SENet, and thus it was named HRRS-U-Net. HRRS-U-Net uses parallel convolutional module flows to overcome spatial information loss and constructs residual squeeze and excitation blocks (RS-Blocks) to learn discriminative representations. We selected four study areas across Guangdong Province to evaluate the performance of HRRS-U-Net using GF-2 images.

The main contributions of this study are listed as follows: (1) we developed a robust and transferable deep learning-based method to extract highly spatiotemporal heterogeneous croplands in southern China; (2) we have comprehensively investigated the impact

of RL and CAM on model performance; (3) to evaluate the superiority of our model, we compared it with RF machine learning algorithm and other popular semantic segmentation models.

2. Methodology

Figure 1 illustrates the workflow of this study, which involves several stages: GF-2 image preparation and pre-processing, sample datasets production, deep learning model development, cropland extraction and accuracy assessment, and comparison of different modules and methods.

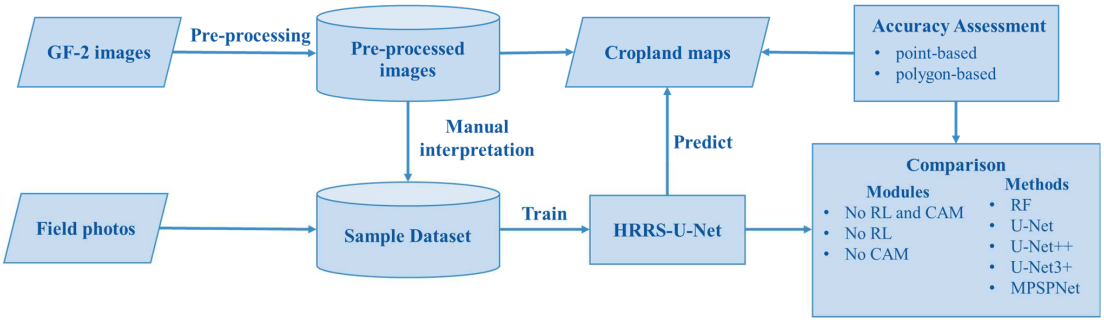


Figure 1. The workflow diagram of this study.

2.1. Study Area

Guangdong Province is located in the southern part of mainland China (20°13′−25°31′N, 109°39′−117°19′E), with a land area of 179,800 km^2. The region has a tropical/subtropical temperate and monsoon climate, with mostly cloudy and rainy weather year-round. The topography is primarily mountainous and hilly, and the cropland is generally fragmented, consisting of multiple small fields. Because of its complex topography and heterogeneous landscapes, the spectral characteristics and spatial patterns of cropland are highly variable throughout the province. Therefore, four representative areas were selected as study areas, which included different landscapes, cropping systems, and environmental conditions. Figure 2 shows the location and topography of the study areas. The general characteristics of the four study areas are described below.

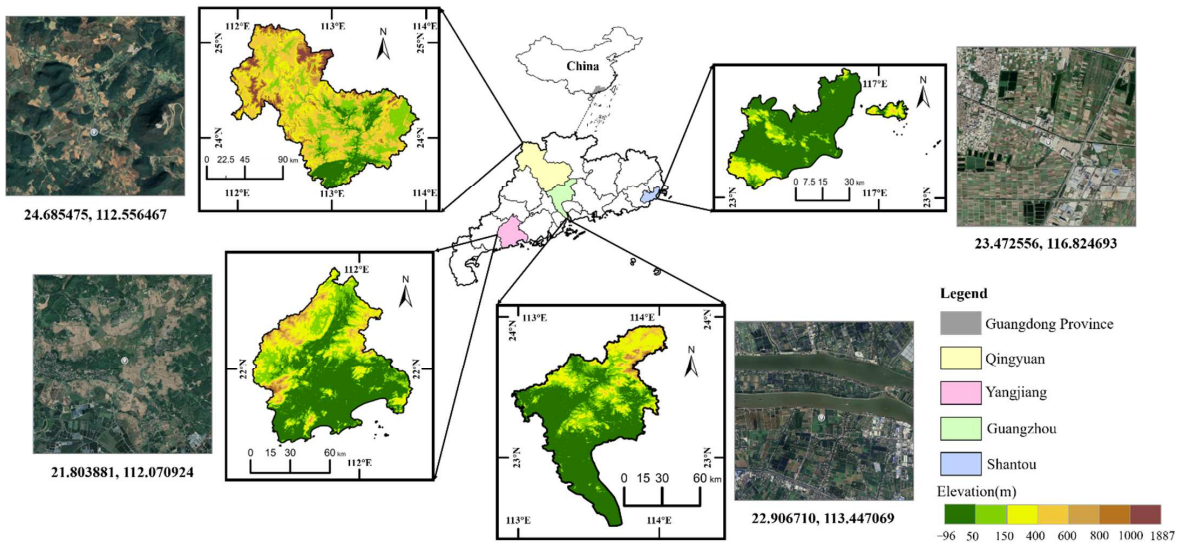


Figure 2. The location and topography of study areas.

Study area 1: Qingyuan is located in northern Guangdong (23°26′−25°11′N, 111°55′−113°55′E) and has a total land area of 19,036 km^2. The topography mainly constitutes mountains and hills, which cover 70.4% of the total area. The cropland area is 2681 km^2 (14.08%), and a significant portion is fragmented. The main crops include early and late rice, peanut, and corn. Early rice is planted in mid to late May and harvested in mid to late July, while late rice is planted before the lunar autumn and harvested in November. Spring peanut is sown in mid to late February and harvested in mid to late July, while summer peanut is sown in late May and early June, then harvested in September and October. The phenological calendar of corn is from late October or early November until the following March.

Study area 2: Yangjiang is located on the southwestern coast of Guangdong (21°28′−22°41′N, 111°16′−112°21′E). It has a land area of 7955 km^2, of which 1491 km^2 (18.74%) is cropland. This area has a mixed topography: 25.6% of the area is hilly, 42.0% is mountainous, and 21.8% constitutes plains. The area borders the South China Sea to the south and has large mudflat regions. The main crops include early and late rice and peanuts.

Study area 3: Guangzhou, the capital of Guangdong Province, is located at 22°26′–23°56′N, 112°57′–114°03′E. It covers 7434 km^2, of which cropland is only 924 km^2 (10.76%). The landscape is complex, including mountains, hills, alluvial plains, and mudflats. The main crop types include early rice, late rice, and sugarcane. Sugarcane is sown in spring from the end of January to mid-March and harvested from May to mid-July, while in autumn, it is sown from the end of August to the end of September and harvested in mid-December.

Study area 4: Shantou, located in the southeast of South China (23°02′–23°38′N, 116°14′–117°19′E), covers 2199 km^2, and the area of cropland is 368 km^2 (16.73%). The topography of Shantou is flat, and the landform is mainly delta alluvial plains, which account for 63.62% of the area. This area has a high level of urbanization, and the ratio of built-up area is one of the highest in South China. The major crop types include early and late rice.

2.2. Dataset

2.2.1. Data Sources and Pre-Processing

We collected available high-quality GF-2 satellite images with slight cloud coverage (<5%) to cover the whole study area from the Guangdong Data and Application Center of High-Resolution Earth Observation System (http://gdgf.gd.gov.cn/, accessed on 5 September 2022). Successfully launched on 19 August 2014, the GF-2 satellite is equipped with two high-resolution 1-m panchromatic and 4-m multispectral cameras that exhibit sub-meter spatial resolution. The detailed specifications of the GF-2 satellite are presented in Table 1. In total, images of 118 scenes were collected, which were acquired from 2018 to 2020. Due to insufficient observations of GF-2, a small portion of the study areas remains uncovered. Figure 3 shows the spatial distribution of the GF-2 images used in the study.

Table 1. The detailed specifications of the GF-2 satellite.

Orbital Type	Orbital Altitude	Coverage Cycle	Revisit Cycle	Swath Width	Band						
					Spectral Range (μm)					Spatial Resolution (m)	
					MSS				PAN	MSS	PAN
					Blue	Green	Red	Infrared			
Sun-synchronous	631 km	69 days	5 days	45 km	0.45–0.52	0.52–0.59	0.63–0.69	0.77–0.89	0.45–0.90	4	1

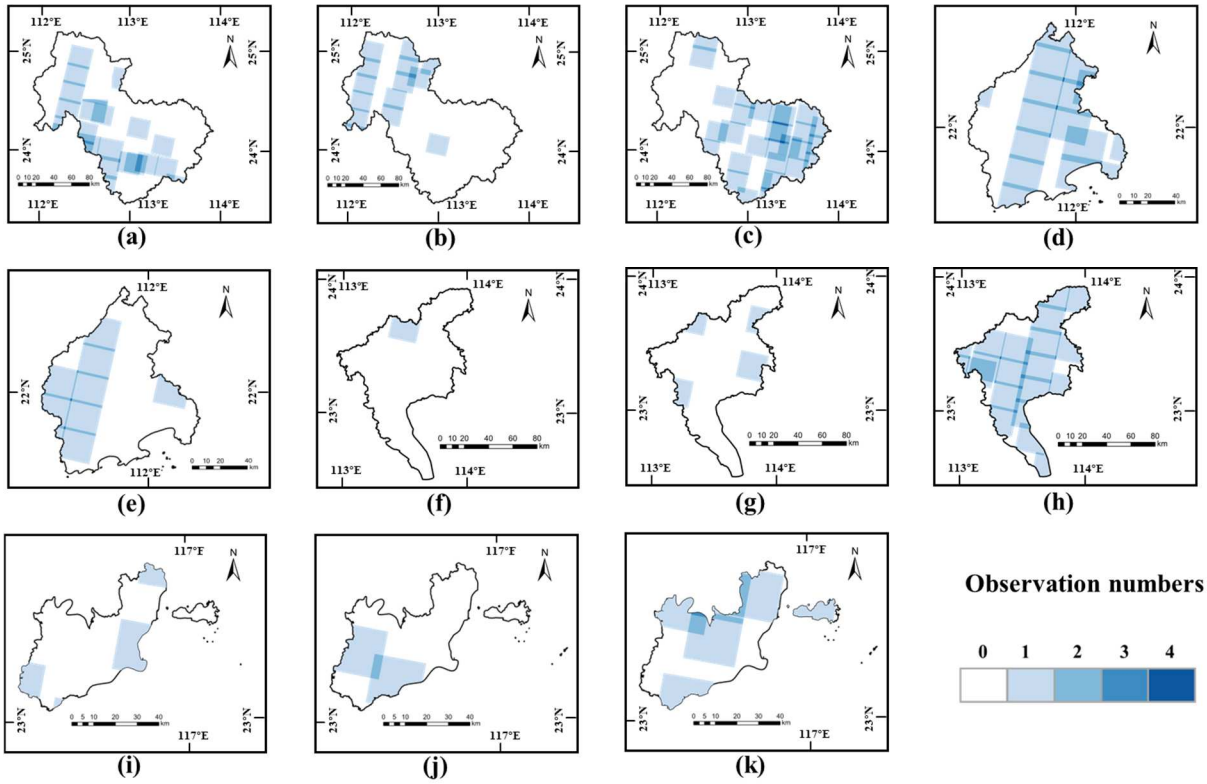


Figure 3. Spatial distribution of the observation numbers of each season in all study areas. (**a**–**c**) Spring, summer, and winter in Qingyuan, (**d**,**e**) Spring and winter in Yangjiang, (**f**–**h**) Spring, summer, and winter in Guangzhou, (**i**–**k**) Spring, autumn, and winter in Shantou.

To correct the distortions, we pre-processed the collected remote-sensing images. Orthorectification was performed to remove geometric distortions, followed by radiometric calibration and quick atmospheric correction(QUAC) [59] to remove scattering and absorption effects from the atmosphere. The images were then processed as composites with blue, green, red, and near-infrared bands at VHR by fusing multispectral scanner system (MSS) images and the corresponding panchromatic (PAN) images based on Pan-sharpening [60]. Eventually, the composites were resampled to 1-m spatial resolution.

2.2.2. Sample Dataset

To ensure that the deep learning model could effectively learn the spectral characteristics and spatial distribution of cropland in different agricultural landscapes, five scenes of GF-2 images across the study area were used to produce the sample dataset for model training. Binary labeled images were generated by manual delineation of cropland boundaries from GF-2 images, and the labeled dataset was validated and revised based on field surveys. Notably, because of insufficient observations, the labeled dataset did not cover all seasons. CNNs use context to identify targets, and objects at the edges of the image may lack complete contextual information, leading to incorrect classification [61]. Therefore, the GF-2 images and corresponding labeled images were split into 256 × 256-pixel patch sizes through a sliding window with a stride of 128 pixels in each direction. Additionally, the dataset was expanded using data augmentation algorithms, including random rotation and flip (horizontal and vertical). Finally, there were 3,343,076 sample patches, of which 80% were randomly allocated for training, and the remaining 20% were allocated for validation.

2.3. Deep Learning Model

In this study, we developed an improved HRRS-U-Net for cropland extraction. Figure 4 shows the architecture of the HRRS-U-Net, which was built on the U-Net frame-

work and utilized the three recently developed HRNet, ResNet, and SENet. Compared to the normal U-Net, HRRS-U-Net uses parallel convolutional module streams and multi-resolution fusion operations to produce precise boundaries and constructs RS-Block to learn discriminative representations. The above improvements appropriately increased the network parameters; therefore, deep supervision [62] was introduced to efficiently train the parameters and prevent overfitting. The following subsections describe the details of each improvement.

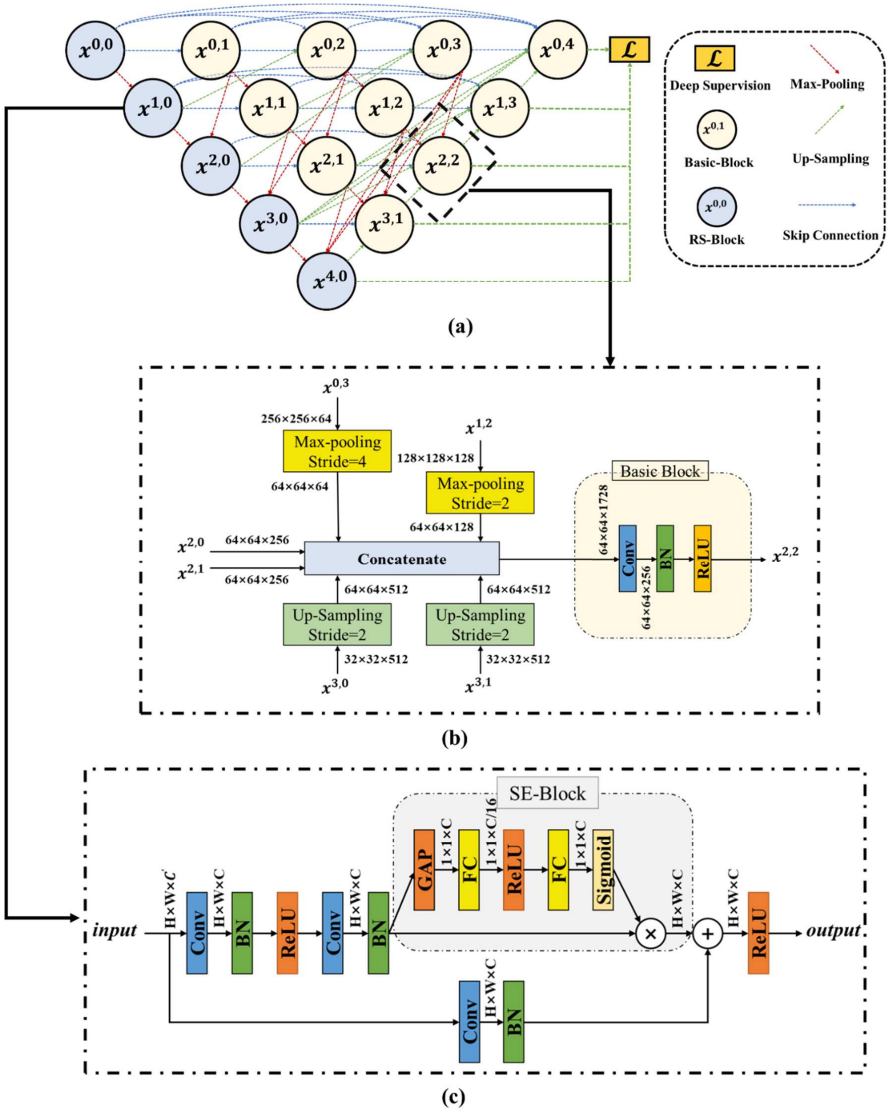


Figure 4. (**a**) The general structure of the HRRS-U-Net. $x^{i,j}$ denotes the jth block of the ith level. (**b**) The specific details of multi-resolution fusion, taking $x^{2,2}$ as an example. (**c**) The detailed composition diagram of RS-Block, where the gray box represents SE-Block.

2.3.1. Parallel Convolutional Module Streams

Inspired by HRNet, we adopted the strategy of maintaining high-resolution representations throughout the network to generate precise cropland boundaries. In HRRS-U-Net, the feature maps are maintained at their original level by using repeated convolutional modules between the encoder and decoder. As shown in Figure 4a, convolutional modules with the same resolution are connected in series, and convolutional module streams with different resolutions are connected in parallel. In the same resolution stream, the shallow features in each layer are delivered to the subsequent layers via skip connections to maintain the details. For example, the previous $x^{0,1}$ was passed to the subsequent $x^{0,2}$,

$x^{0,3}$, and $x^{0,4}$. Feature maps generated by convolutional modules share the same resolution and number of channels within the same convolutional module stream. Specifically, from $x^{0,j}$ to $x^{4,j}$, the scale size was 256, 128, 64, 32, and 16, while the channel numbers were 64, 128, 256, 512, and 1024, respectively.

2.3.2. Multi-Resolution Fusion

Multi-resolution fusion was repeatedly used to mutually enrich different-resolution feature maps to obtain stronger semantic representations with precise locations. Low-level feature maps with high-resolution representations had fine location information, and high-level feature maps with large receptive fields had rich semantic information [63]. Therefore, the combination of different level features is an effective approach to improving model performance. Multi-resolution fusion aggregates feature maps that are generated from previous basic blocks with different resolutions. Figure 4b demonstrates the implementation details of multi-resolution fusion using $x^{2,2}$ as an example. Different resolution feature maps have distinct scales; therefore, the feature maps undergoing aggregation required transformation to ensure that they were consistent with the corresponding basic blocks. Specifically, the high-resolution features ($x^{0,3}$, $x^{1,2}$) were performed by max-pooling operations, low-resolution features ($x^{3,0}$, $x^{3,1}$) were performed by up-sampling operations with specified strides, and some resolution features ($x^{2,0}$, $x^{2,1}$) were directly copied. After alignment, all feature maps were aggregated into the high-dimensional feature map. Considering that high-dimensional data require substantial computational consumption, the generated features had reduced dimensions in subsequent basic blocks, which included a sequential 3×3 convolutional (Conv) layer, batch normalization (BN) layer, and rectified linear unit (ReLU).

2.3.3. RS-Block

The RS-block with embedded RL and CAM was used in encoded layers to extract deep discriminative and representative features. CNNs are not effective in the modeling of high dependencies between spectra in VHR images, which potentially hinders the generalization and robustness of the model. Thus, the extraction of more representative and discriminative features is a crucial procedure for cropland identification. The detailed composition diagram of the RS-Block is shown in Figure 4c. The RS-Block consists of three parts: two basic blocks, the identity shortcut connection, and the SE-Block. The identity shortcut connection is a branch that delivers input features to be summed and merged with the features, which are processed by the SE-Block using a 1×1 Conv layer and subsequent ReLU. The SE-Block is a CAM module that selectively emphasizes informative channels and suppresses noise and irrelevant information. In SE-Block, the input feature maps are first passed through a global average pooling (GAP) operation over the feature maps, which produces a vector of channel-wise statistics. These statistics are then passed through two fully connected (FC) layers, ReLU and Sigmoid activation functions, and these operations learn to model channel dependencies and generate a channel-wise attention map. Specifically, the first layer reduces the dimensionality of the input vector, while the second layer produces a set of channel-wise scaling factors. Finally, the attention map is applied to the original feature maps using element-wise multiplication, effectively amplifying the important channels and suppressing the less important ones.

2.4. Model Training

To evaluate the performance of the neural network, the loss function was used to measure the fitness between the predicted and the ground truth values. For remote sensing images, there is an imbalance in the proportion of cropland and non-cropland, which is more pronounced in mountainous and urban areas. In the commonly used binary cross-entropy (BCE) loss, the weights of the different categories are equal. This can result in a biased learning direction of the model. Dice loss [64], which is more inclined to extract the foreground, is appropriate for cases of sample imbalance. However, Dice loss exhibits

gradient instability problems that lead to suboptimal convergence [65]. To overcome these issues, we utilized a composite function known as BCE-Dice loss, which benefits from the stability of BCE loss and robustness of dice loss. The calculation formula of BCE-Dice loss can be expressed as follows:

$$\text{Dice} = -\frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \frac{2\text{p}_\text{i}\text{g}_\text{i}}{\text{p}_\text{i} + \text{g}_\text{i}} \tag{1}$$

$$\mathcal{L}_{\text{Dice}} = 1 - \text{Dice} \tag{2}$$

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} (g_i \log p_i + (1 - g_i) \log(1 - p_i)) \tag{3}$$

$$\mathcal{L}_{BCE-Dice} = \mathcal{L}_{Dice} + \mathcal{L}_{BCE} \tag{4}$$

where Dice is the Dice coefficient; $\mathcal{L}_{\text{Dice}}$, $\mathcal{L}_{\text{BCE}}$, and $\mathcal{L}_{\text{BCE}-\text{Dice}}$ represent Dice loss, BCE loss, and BCE-Dice loss, respectively; p_i is the predicted probability value of the ith pixel; g_i is the ground truth value of the ith pixel, and N is the total number of pixels in the image.

To combat the overfitting during model training and ensure sufficient generalization performance, we adopted L2 regularization [66]. This approach ensured that the weights were small but avoided reducing them to zero. The calculation formula of the L2 regularization can be expressed as follows:

$$\mathcal{L}'(\omega) = \mathcal{L}(\omega) + \lambda \, \|\omega\|_2^2 \tag{5}$$

where, $\mathcal{L}(\omega)$ represents the original function; $\mathcal{L}'(\omega)$ represents the function after regularization; $\|\omega\|_2^2$ represents the squared constraint of the L2 norm, and λ represents a constant, which we set to 0.01.

We used Adam [67] as the algorithm for gradient descent optimization. The initial learning rate was 0.0001, and the learning rate decay strategy was adopted to decrease the original value 0.9 times every 3000 steps. The training batch size was set to 8, the maximum number of training epochs was set to 50, and the early-stopping strategy was designed to stop training after 10 epochs without performance improvement. Finally, He initialization [68] was adopted to initialize model weights.

2.5. Accuracy Assessment

In this study, the accuracy of the classification results was assessed using two different assessment methods: point-based and polygon-based assessment.

In the point-based assessment approach, stratified random sampling design and pixel-by-pixel validation were used to assess the accuracies of our maps and the other cropland maps. To reduce the standard error of the producer accuracy (PA) and OA [69], sample points were assigned to cropland and non-cropland based on their areal proportions. We randomly selected 300 sample points in each of the four study areas according to the areal proportion of land use types for validation, including 180 samples for cropland and 1020 samples for non-cropland. The distribution of sample points is shown in Figure 5. We visually interpreted the land use for all sample points by combining field survey results. Finally, using cropland extraction results and sample data, we calculated confusion matrixes to evaluate the accuracy.

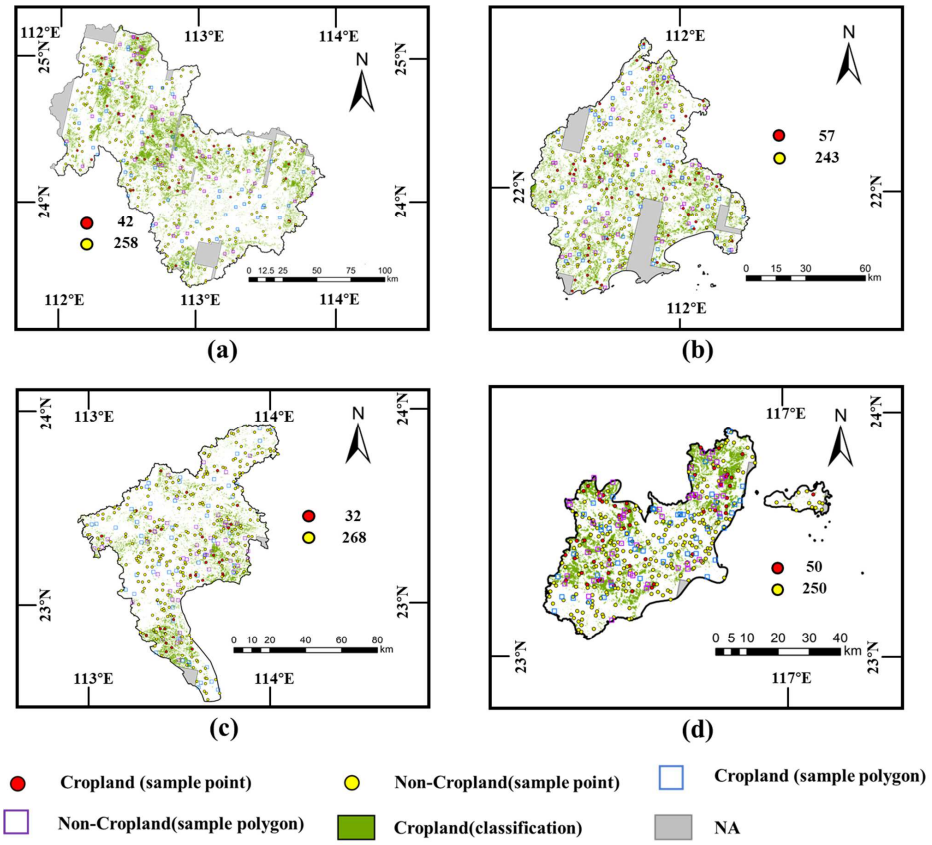


Figure 5. Spatial distributions of sample points and polygons of two strata (cropland and non-cropland), (**a**) Qingyuan, (**b**) Yangjiang, (**c**) Guangzhou, and (**d**) Shantou.

We adopted the following evaluation indicators: user accuracy (UA), PA, OA, F1 of cropland, and Kappa. UA is the probability that a value predicted to be in a specific class really is that class, and PA is the probability that a value in a particular class was classified correctly. UA and PA measure the completeness and precision of decisions, respectively; F1 is a balanced metric between the two. Additionally, Kappa was used to estimate the consistencies of prediction and ground truth; it represents the ratio of classification to the reduction of errors generated by completely random classification. The formula for each evaluation indicator was as follows:

$$UA_C = \frac{TP}{TP + FP}, \; UA_N = \frac{TN}{TN + FN} \tag{6}$$

$$PA_C = \frac{TP}{TP + FN}, \; PA_N = \frac{TN}{TN + FP} \tag{7}$$

$$OA = \frac{TP + TN}{T} \tag{8}$$

$$F1 = 2 \times \frac{UA_C \times PA_C}{UA_C + PA_C} \tag{9}$$

$$p_e = \frac{(TP + FP) \times (TP + FN) + (TN + FP) \times (TN + FN)}{(TP + FP + TN + FN)^2} \tag{10}$$

$$Kappa = \frac{OA - p_e}{1 - p_e} \tag{11}$$

where TP is true positive (i,e., correctly identified croplands), TN is true negative (i,e., correctly identified non-croplands); FN is false negative (i.e., true cropland omitted by the method), and FP is false positive (i.e., not cropland but erroneously detected as cropland). Additionally, UA_C and UA_N denote the UA of cropland and non-cropland, respectively; PA_C and PA_N denote the PA of cropland and non-cropland, respectively; and p_e represents the hypothetical probability of chance agreement.

Although point-based accuracy metrics are commonly used to evaluate the accuracy of image classifications, they do not provide information about the thematic representation quality of land objects [70]. On the other hand, polygon-based approaches take into account the thematic and geometric properties of map units [71]. Therefore, we additionally used a polygon-based evaluation to determine the thematic accuracy of the object representation. This method is used to evaluate the segmentation accuracy by simply converting the classification results into vector data and extracting the intersection with the reference polygon [70]. We employed the spatial point sampling approach for polygon sampling [72]. We first randomly generated a set of sample points within the bounding box of the polygon and then selected the mapped polygon in which the sample points fall. We extracted and visually interpreted 100 polygons (50 polygons for cropland and 50 polygons for non-cropland) in each of the four study areas. The distribution of sample polygons is shown in Figure 5. This evaluation performs an area-based ratio calculation between the intersection area of the result and the area of the reference polygon. The formula for the intersection rate was as follows:

$$IntRate = \frac{Area_{Int}}{Area_{Ref}} \tag{12}$$

where $IntRate$ refers to intersection rate, $Area_{Int}$ refers to the area of the intersection area, and $Area_{Ref}$ refers to the area of the reference polygon.

The rate takes values from 0 to 1. After calculating the ratio, the mean value of the rate (Mean) and standard deviation of the rate (Std) metrics are provided for each category.

3. Results

3.1. Ablation Experiment Results of the RS-Block

As mentioned in Section 2.3.2, we constructed the RS-Block embedding RL and CAM to improve the accuracy of cropland extraction. To investigate the effectiveness of the RS-Block in terms of improving model performance, ablation experiments were conducted. As shown in Figure 6, the RS-Block was detached in the experiment, resulting in conditions that excluded only RL, only CAM, or both RL and CAM.

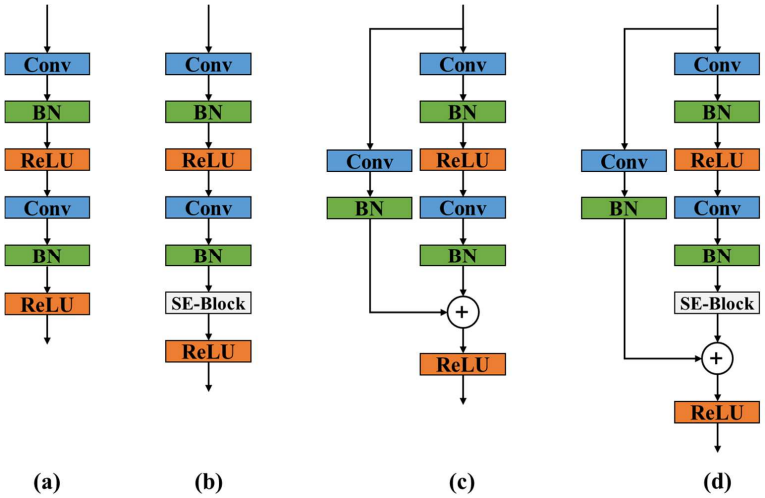


Figure 6. Ablation experiment on RS-Block: (**a**) No RL and CAM, (**b**) No RL, (**c**) No CAM, and (**d**) complete RS-Block.

Figure 7 shows detailed comparisons of the extraction results of different modules. In each row, the GF-2 images of regions and the corresponding segmentation results of different modules are presented. Table 2 presents the comparative accuracies of the cropland results for different modules. Because the backbone is the same, the delineation results of different modules were generally consistent. However, there were differences in the delineation of small fields and dense boundaries. When RL or CAM was absent, the segmented results became coarse, and tiny ridges were neglected. In particular, the model without RL produced classification maps with more missed and misclassified pixels, particularly in backgrounds with gardens and forests. The addition of RL significantly reduced these errors, resulting in cropland UA, cropland PA, and OA improvements of 4.80%, 0.85%, and 0.69%, respectively; F1 and Kappa increased by 3.0% and 3.5%, compared with the non-application of RS-Blocks. However, the performance of CAM alone was mediocre, with no significant improvement and even a decrease in some accuracies. This was because deeper models usually become stuck on the gradient problem; thus, it is impractical to perform parameter updates. In the presence of RL, CAM could further improve accuracy; the cropland UA, PA, OA, F1, and Kappa increased by 1.37%, 0.69%, 0.34%, 1.3%, and 1.6%, respectively, compared with the use of RL alone. The complete RS-Block clearly surpassed the other modules and had the highest accuracy. Compared with the simultaneous absence of RL and CAM, the complete RS-Block significantly improved model performance, with 6.85% higher cropland UA, 1.54% higher cropland PA, and 1.04% higher OA, as well as 4.3% and 5.1% higher F1 and Kappa, respectively. Furthermore, the integration of RL and CAM can enhance the accuracy of object representation, resulting in an increase of 2.30% and 1.58% in the Mean values of cropland and non-cropland, respectively. Moreover, the complete RS-Block significantly improves transferability and robustness, as demonstrated by a reduction of 17.32% and 27.56% in the Std values of cropland and non-cropland, respectively.

Table 2. Comparison of extraction accuracy with different modules and different methods (CL: cropland, Non-CL: non-cropland).

Scenario	Model	Point-Based							Polygon-Based			
		UA (%)		PA (%)		OA (%)	F1	Kappa	Mean		Std	
		CL	Non-CL	CL	Non-CL				CL	Non-CL	CL	Non-CL
Comparison of modules	No RL and CAM	81.11	99.31	95.42	96.75	96.58	0.877	0.857	0.871	0.951	0.179	0.127
	No RL	82.22	99.22	94.87	96.94	96.67	0.881	0.862	0.865	0.943	0.161	0.104
	No CAM	85.00	99.41	96.23	97.41	97.25	0.903	0.887	0.877	0.960	0.165	0.097
Comparison of methods	RF	26.11	96.67	58.02	88.11	86.08	0.360	0.295	0.525	0.647	0.217	0.164
	U-Net	75.56	99.02	93.15	95.83	95.50	0.834	0.809	0.813	0.913	0.196	0.132
	U-Net++	80.56	99.22	94.77	96.66	96.42	0.871	0.850	0.833	0.925	0.153	0.118
	U-Net3+	80.00	98.43	90.00	96.54	95.67	0.847	0.822	0.807	0.891	0.173	0.122
	MPSPNet	82.78	99.51	96.75	97.04	97.00	0.892	0.875	0.862	0.953	0.158	0.114
Our Model		86.67	99.51	96.89	97.69	97.58	0.915	0.901	0.891	0.966	0.148	0.092

3.2. Comparison of HRRS-U-Net with Other Methods

To evaluate the classification performance of the proposed HRRS-U-Net model, we conducted a comprehensive comparison with traditional machine learning method (RF) and other deep semantic segmentation algorithms (U-Net, U-Net++ [73], U-Net3+ [74], and MPSPNet). The RF classifier can successfully handle high data dimensionality and multicollinearity, being both fast and immune to data noise and overfitting [75,76]. U-Net++ and U-Net3+ are built on U-Net and reduce the information gap by increasing the skip connection between the encoder and decoder. MPSPNet exploits global contextual information by aggregating the contexts of different regions to make the final prediction more reliable [77].

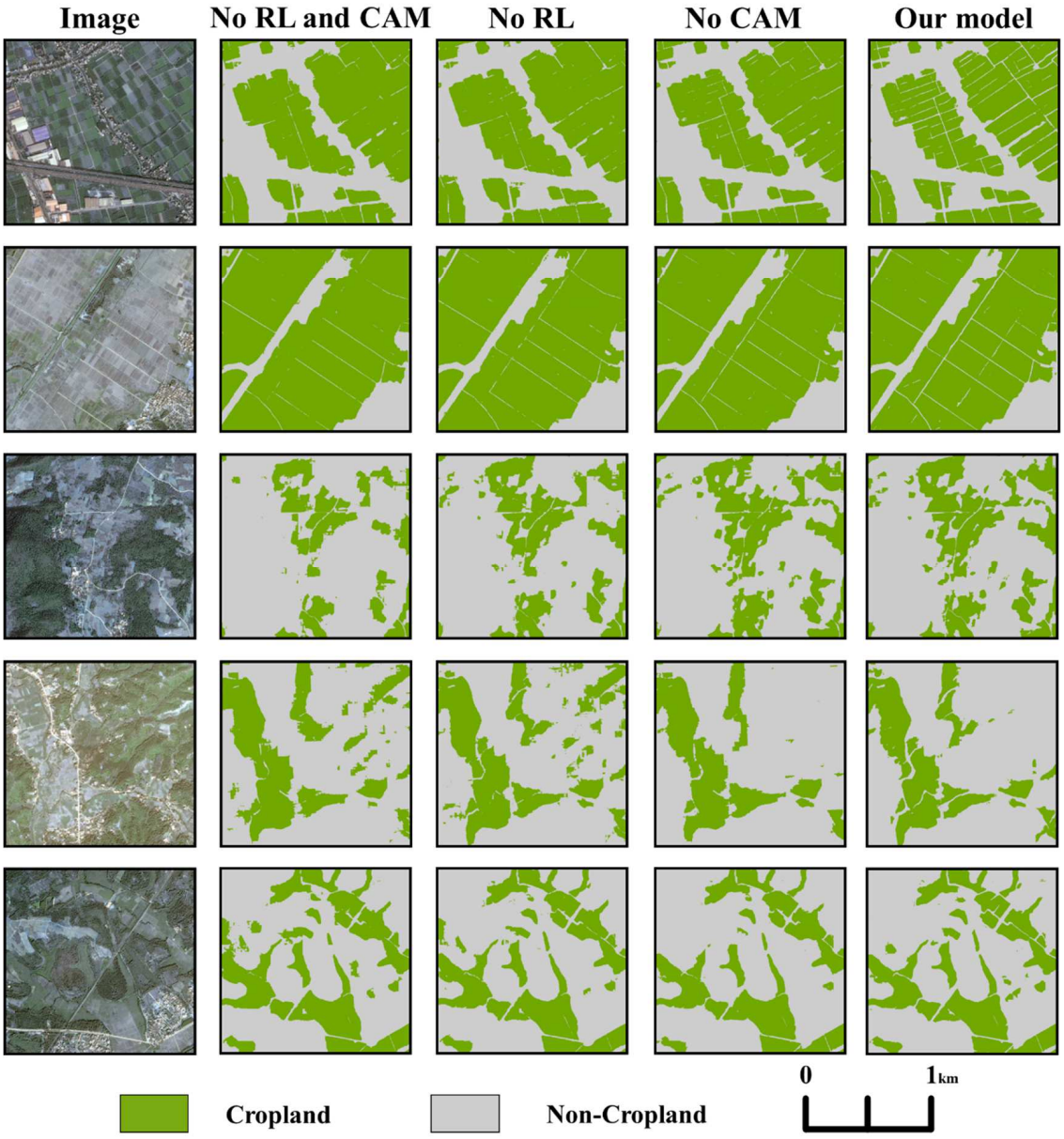


Figure 7. Cropland extraction results of the comparison of different modules: No RL and CAM, No RL, No CAM, and our model (RS-Block).

Figure 8 presents a detailed comparison of the different models, including the representative zones for each study area. Table 2 summarizes the cropland extraction accuracies of the different methods. The cropland maps of RF showed salt-and-pepper noise, with undesirable visual effects and the most misclassifications. When compared to the deep learning approach, the RF method exhibited a reduction in overall accuracy (OA), ranging from 9.86% to 13.36%. Additionally, the F1 and Kappa of the RF were only 39.34% to 43.17% and 32.74% to 36.46% of those achieved by semantic segmentation algorithms, respectively. The results of the polygon-based accuracy evaluation indicated that the object representation accuracy of the RF method was the lowest, as reflected in its lowest Mean values and significantly higher Std values. The poor performance of the RF method confirms the limited ability of traditional machine learning methods to learn complex features and patterns in VHR remote sensing images.

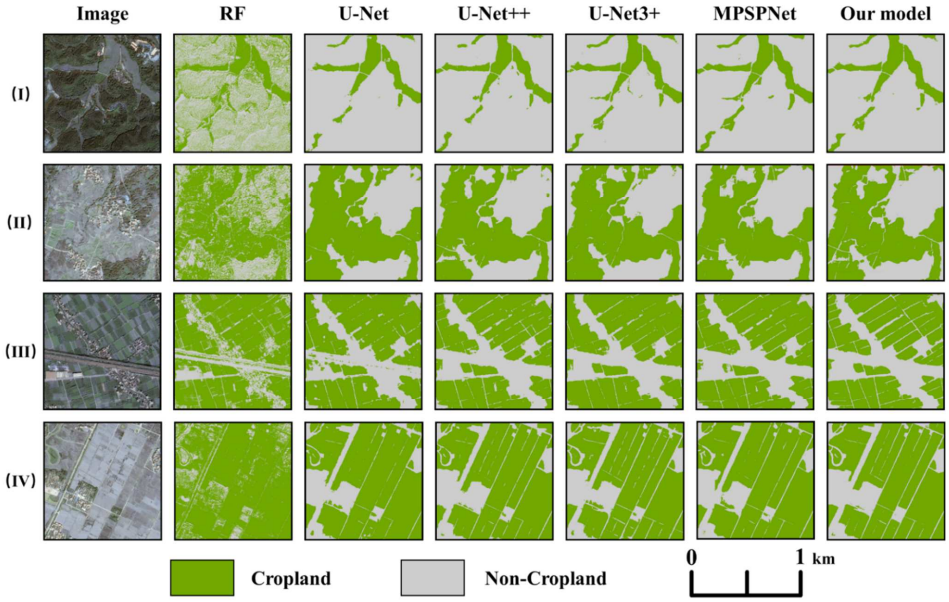


Figure 8. Cropland extraction results of the comparison of different methods: RF, U-Net, U-Net++, U-Net3+, MPSPNet, and our model (HRRS-U-Net). (**I**–**IV**) Typical sub-region results of Qingyuan, Yangjiang, Guangzhou, and Shantou, respectively.

Among the semantic segmentation models evaluated, HRRS-U-Net exhibited the best agreement with the actual cropland boundaries, particularly in fragmented landscapes. Furthermore, HRRS-U-Net achieved the highest accuracy, surpassing the other models by 0.58% to 2.08% in OA, 2.3% to 8.1% in F1, and 2.6% to 9.2% in Kappa. Moreover, the cropland maps of HRRS-U-Net provide the highest Mean values (cropland: 0.891, non-cropland: 0.966) and the lowest Std values (cropland: 0.148, non-cropland: 0.092), confirming the excellent boundary delineation ability and generalization of the model. The cropland results of U-Net significantly deviated from the ground truth and demonstrated the worst performance. U-Net++ and U-Net3+ exhibited improved performance, with OA increasing by 0.17% to 0.92%, F1 value growing by 1.3% to 3.7%, and Kappa value rising by 1.3% to 4.2% and cropland Std value reducing by 7.58% to 10.61%, compared to U-Net. Additionally, U-Net++ and U-Net3+ produced more visually favorable results with more precise boundaries and less noise. Notably, U-Net3+ had many omission errors in mountainous areas and around built-up areas, with a lower cropland PA than the other models, including U-Net. MPSPNet achieved comparatively good performance with an OA of 97.00%, F1 of 0.915, and Kappa of 0.901. The cropland results of MPSPNet were visually consistent with the results of HRRS-U-Net, but the boundary localization and robustness were inferior. The Mean value of cropland for MPSPNet is 3.36% lower, while the Std value of cropland is 6.33% higher compared to HRRS-U-Net.

3.3. Results of Cropland Extraction

We conducted accuracy assessments of cropland results for the four study areas. Figure 9 shows the confusion matrixes, and Table 3 summarizes the results of accuracy assessments in each study area. The OA was 97.85%, F1 was 0.915, Kappa was 0.901, cropland Mean was 0.891, and cropland Std was 0.148 over all study areas. The cropland category had UA and PA values of 86.67% and 96.89%, respectively, whereas the non-cropland category had UA and PA values of 99.51% and 97.69%, respectively. The accuracy of cropland results differed among study areas, with an OA of 97.00–98.33%, an F1 of 0.830–0.940, and a Kappa of 0.814–0.929. All cropland results were reasonably accurate, indicating that our cropland extraction method was generally reliable. Qingyuan (OA: 98.33%, F1: 0.938, Kappa: 0.929) and Shantou (OA: 98.00%, F1: 0.940, Kappa: 0.928) had the high-

est accuracies, whereas Yangjiang (OA: 97.00%, F1: 0.916, Kappa: 0.898) and Guangzhou (OA: 97.00%, F1: 0.830, Kappa: 0.814) had the lowest accuracies. Generally, the accuracy of cropland extraction was higher in single-topography areas than in mixed-topography areas. Specifically, cropland UAs significantly differed across the four study areas. Shantou had the highest cropland UA of 92.16%, followed by Qingyuan, with a slightly lower UA of 90.48%. Yangjiang exhibited a comparatively low cropland UA of 85.97%, with some mudflats misclassified as cropland. Guangzhou had the lowest cropland UA at 73.33%, with many commission errors mainly concentrated in artificial greenbelts. The PAs and Means of cropland in all four study areas exceeded 95.00% and 0.850, respectively, indicating that HRRS-U-Net detected the most cropland areas. The results indicated that the Std values for cropland exhibited variability across the study areas, with the smallest value observed in Shantou (0.118) and the largest in Guangzhou (0.175). However, the overall Std values were observed to be satisfactory, which demonstrated the reliability and stability of our model when applied to different regions.

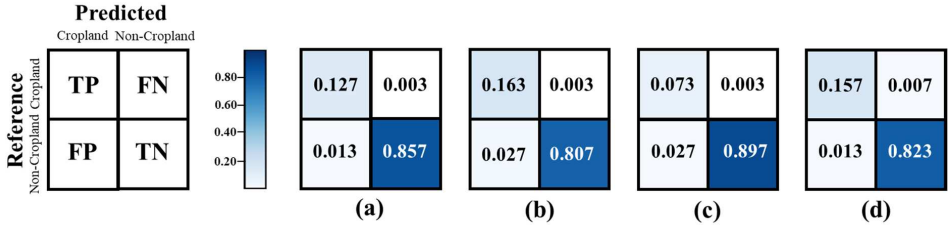


Figure 9. Normalized binary confusion matrixes of cropland mapping results in four study areas: (**a**) Qingyuan, (**b**) Yangjiang, (**c**) Guangzhou, and (**d**) Shantou.

Table 3. Extraction accuracy of the four study areas using the proposed method (CL: cropland, Non-CL: non-cropland).

Region	Point-Based							Polygon-Based			
	UA (%)		PA (%)		OA (%)	F1	Kappa	Mean		Std	
	CL	Non-CL	CL	Non-CL				CL	Non-CL	CL	Non-CL
Qingyuan	90.48	99.61	97.44	98.47	98.33	0.938	0.929	0.895	0.958	0.154	0.057
Yangjiang	85.97	99.59	98.00	96.80	97.00	0.916	0.898	0.918	0.963	0.136	0.045
Guangzhou	73.33	99.63	95.65	97.11	97.00	0.830	0.814	0.862	0.962	0.175	0.043
Shantou	92.16	99.20	95.92	98.41	98.00	0.940	0.928	0.888	0.980	0.118	0.036
Total	86.67	99.51	96.89	97.69	97.58	0.915	0.901	0.891	0.966	0.148	0.092

Figure 10 shows the cropland extraction results for each of the four study areas; (a) shows the cropland results for cities, and (b) contains six typical sub-regions with magnified details. These sub-regions spanned different seasons and had different landscapes, cropping systems, and environmental conditions. The spatial pattern of the cropland results from HRRS-U-Net was visually consistent with climate factors and topographical characteristics. Qingyuan and Yangjiang are dominated by mountainous and hilly areas; in these areas, cropland is characterized by small field sizes and fragmented distribution of land parcels. In the results, small fields were sensitively recognized, and irregular boundaries were explicitly determined. In Guangzhou, most cropland is concentrated in the southern plains and interspersed with built-up land. Similar to Guangzhou, cropland in Shantou are distributed in clusters and interspersed with other land uses. In the cropland maps, although there were many grassland and bare land areas, which had spectral, textural, and shape characteristics similar to the characteristics of cropland, most were effectively identified and filtered. In particular, dense field ridges in cropland fields were clearly distinguished and delineated. Additionally, although the training dataset did not span all seasons, different periods of cropland were accurately recognized. In summary, HRRS-U-Net accurately identified the cropland extent and explicitly located the boundaries in different periods and locations.

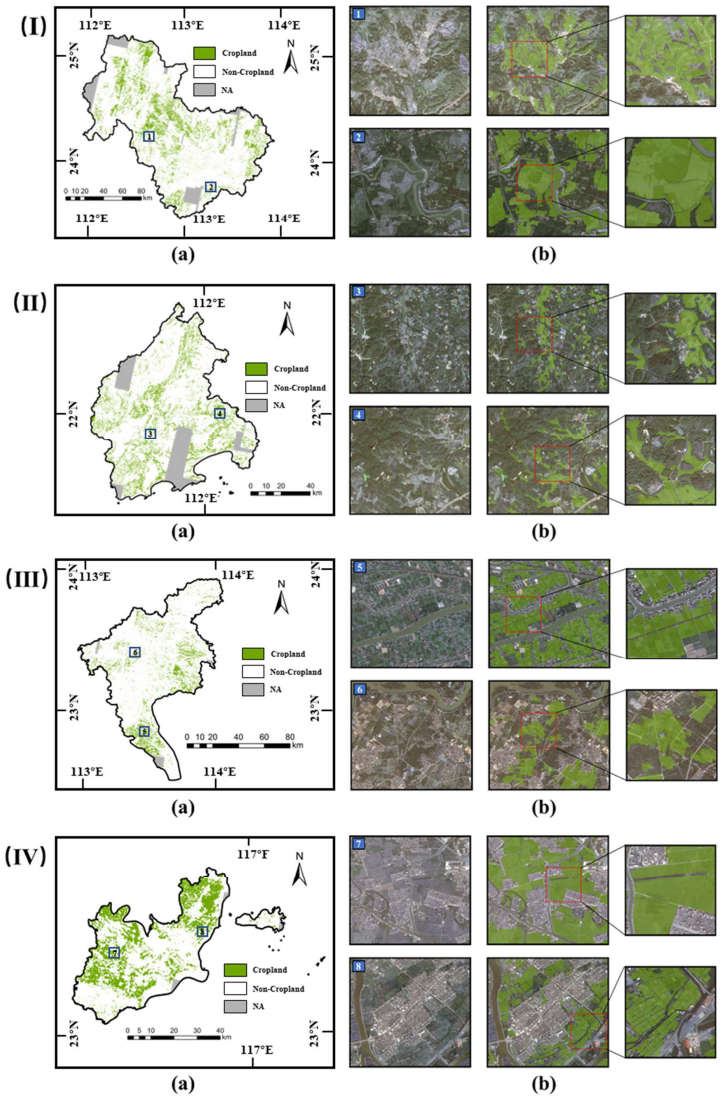


Figure 10. Cropland mapping results for (**I**) Qingyuan, (**II**) Yangjiang, (**III**) Guangzhou, and (**IV**) Shantou. (**a**) Cropland maps covering study areas. (**b**) Six typical sub-regions are marked on the maps with magnified detail views.

4. Discussion

4.1. Maintaining High-Resolution Representation to Improve Boundary Delineation

Previous studies of cropland identification mainly relied on medium- and low-resolution images, which resulted in data products that were unable to distinguish individual fields. However, because agricultural statistics are generally field-based, precise field boundary delineation is critical for cropland extraction [78]. Therefore, our study utilized VHR images, which provided fine-grained spectral and spatial information regarding the ground. However, down-sampling in CNN-based models loses high-resolution, detailed information that is not fully recoverable. Strategies to solve this issue can be classified into four categories: (1) refining boundary prediction through post-processing [79]; (2) improving the loss function to focus more on the boundary [80,81]; (3) aggregating low-level local features into high-level semantic features to enhance boundary delineation (e.g., U-Net, U-Net++, U-Net3+, and MPSPNet); and (4) maintaining high-resolution representations throughout the network (e.g., HRNet).

Strategies (1) and (2) require long inferential time and high computational costs and are difficult to implement in large-area mapping applications. Strategy (3) usually relies on skip connections, which are sometimes ineffective and may even negatively affect segmentation

performance. MPSPNet, U-Net++, and U-Net3+ reduce the information gap between the encoder and decoder by increasing skip connections between blocks at different levels. However, these models still partially rely on up-sampling to recover information, which is insufficient for a complete reconstruction of local information. Additionally, some skip connections are ineffective because of incompatible features in the encoder and decoder [82]. As mentioned in Section 3.3, U-Net3+ has high omission errors and demonstrated a lower cropland PA compared with U-Net.

We adopted strategy (4) to improve the accuracy of the cropland boundary and obtained the most competitive cropland maps with complete internal and continuous boundaries. Even in fragmented landscape areas, the irregular boundaries of small fields were precisely located and completely delineated. To understand the effectiveness of model improvements, we visualized shallow and deep feature maps within the model. Figure 11 presents the feature maps resulting from the process of cropland extraction, where (d–h) are feature maps generated by the module in the highest resolution streams. From $x^{0,0}$ to $x^{0,4}$, fine details were adequately preserved after feature extraction. Shallow feature maps (closer to the inputs) and deep feature maps (after massive stacked transformations) both had fine visual characteristics and accurate local representation. These findings indicate that HRRS-U-Net effectively solved the problem of location information loss and improved boundary delineation.

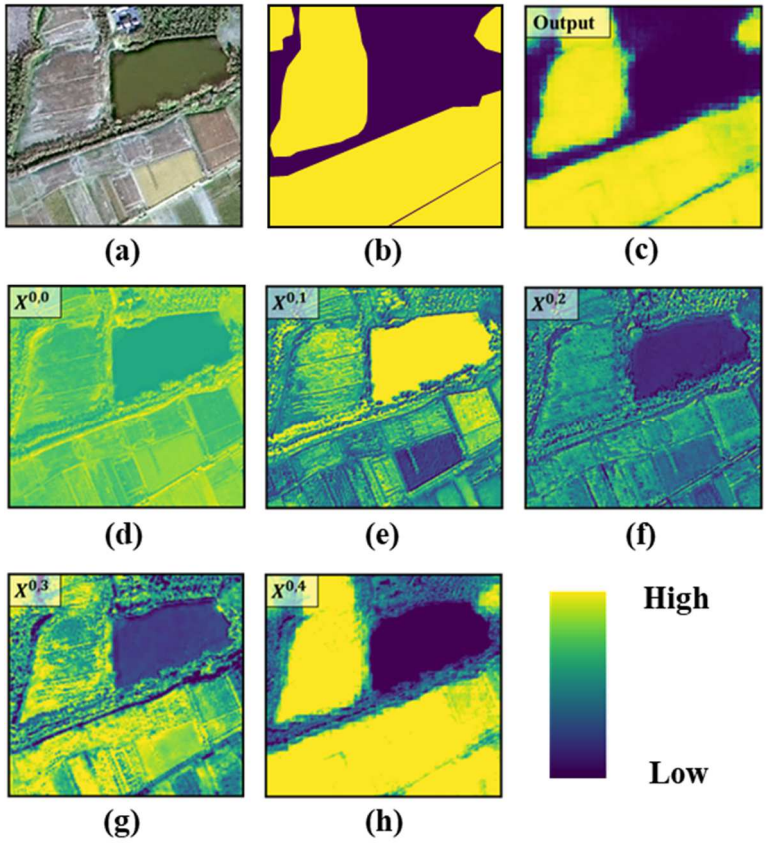


Figure 11. Visualized feature maps within the model: (**a**) original image, (**b**) reference label, (**c**) classification map, (**d**–**h**) feature maps in $x^{0,0}$, $x^{0,1}$, $x^{0,2}$, $x^{0,3}$, $x^{0,4}$.

4.2. Extracting Representative Features to Generalize Highly Spatio-Temporal Heterogeneous Cropland

In southern China, cropland has complex characteristics in both spatial and temporal domains. The VHR images further increase cropland extraction difficulty because of substantial intra-class variations and subtle extra-class similarity. The traditional statistical methods (e.g., RF) find it difficult to capture complex spatial and spectral relationships in

VHR images. Previous studies using deep learning-based methods have demonstrated unreliable performance in heterogeneous and fragmented landscape areas. U-Net, U-Net++, U-Net3+, and MPSPNet produced high commission errors (17.22–24.44%) and omission errors (3.25–10.00%), and their results estimated croplands with comparatively large deviations. To enhance the generalization and robustness of cropland extraction, HRRS-U-Net introduced RL and CAM to sufficiently extract more representative and discriminative features. With fewer commission and omission errors (13.33% and 3.11%, respectively), HRRS-U-Net was able to identify cropland accurately in different periods and locations.

We revealed the process of cropland extraction through the feature maps in Figure 11. The feature map (c) in $x^{0,\,0}$ was the shallowest; it was also insensitive to the boundary between cropland and non-cropland. The deeper feature map (e) in $x^{0,\,1}$ strongly responded to non-cropland, which facilitated efficient identification and further filtering of other objects. However, the responses were inconsistent among cropland fields, indicating that representative features had not been completely extracted at this stage. The subsequent feature map (f) in $x^{0,\,2}$ merged fields based on the shallow feature maps; it initially delineated the cropland and non-cropland extent. Concurrently, the response of non-cropland began to be suppressed, and cropland received increased attention. With the exploration of more general representations, the feature maps (g) in $x^{0,\,3}$ and (h) in $x^{0,\,4}$ significantly reduced the response variation between cropland fields. Moreover, cropland was continuously enhanced, whereas non-cropland was consistently suppressed. In the final classification map (c), the estimated extent of cropland was close to the actual extent. HRRS-U-Net effectively addressed the challenge of identifying highly heterogeneous croplands and achieved superior cropland extraction performance in southern China.

4.3. Uncertainty Analysis

Despite the successful implementation of our proposed method, there remain some areas for improvement. First, accurate recognition of cropland in hillside areas remains challenging. Because of ecological restoration and farming constraints, most cropland on the slopes of lowland hills has been abandoned, resulting in less intensive farming practices and indistinct boundaries. Additionally, shadows created by terrain and trees reduce image information, leading to false boundaries and incomplete patches in classified results. Second, HRRS-U-Net misidentified some non-cropland objects, including artificial shrubs and mudflats. Many artificially planted greenbelts with shrubs were classified as cropland in the resultant maps because the pattern, shape, texture, and spectra of the shrubs were similar to those characteristics in arable crops. Moreover, the interspersed distribution of shrubs and crops along roads hindered the identification of shrubs in the greenbelt, thereby affecting the accuracy of cropland extraction. Because of its high degree of urbanization with large areas of artificial greenery, HRRS-U-Net produced significantly higher commission errors (26.67%) in Guangzhou than in other study areas. Additionally, coastal mudflats were easily confused with post-harvest irrigated cropland, leading to lower-than-average cropland UA (85.97%) in Yangjiang. Finally, we acknowledged the uncertainty of the sample datasets, which may contain some incorrect or imprecise boundaries because of the difficulty involved in the visual interpretation of cropland from VHR images. The effects of trees, buildings, and their shadows caused some field boundaries to be indistinct or vaguely visible, which appeared plausible in the images. Despite supplementation with in situ photos, it was difficult to determine boundaries explicitly. Additionally, the GF-2 images used in this study had fewer images available at some locations, which may have hindered diverse label production, particularly among different periods.

4.4. Implications and Future Work

The HRRS-U-Net performed well in terms of extracting cropland using VHR remote sensing images, particularly in heterogeneous and fragmented areas. Because of rapid economic development and urban expansion, cropland areas in southern China have

been substantially reduced in recent years. Accurate and detailed information regarding cropland serves as the foundation for various agricultural operating applications, including cropland area estimation, dynamic cropland monitoring, and grain productivity prediction. These applications are essential for agricultural resource monitoring and the formulation of effective policies to guarantee food security.

It is observed that the accuracy was unbalanced in single topography areas (e.g., Qingyuan) and mixed topography areas (e.g., Guangzhou). In future work, we plan to separate single and mixed topography areas and use individual models for training and cropland classification. Additionally, HRRS-U-Net will be generalized to extract cropland at larger scales (e.g., the national scale). To ensure reliable performance, we will augment the existing dataset with more representative samples, considering the various characteristics of cropland in other areas. Larger scale applications inevitably require multi-source image datasets because of the limited observations from GF-2 sensors. Thus, we plan to incorporate additional sensor data, such as GaoFen-1 and QuickBird, to increase the temporal resolution of the available images and obtain more diverse samples to improve the generalization ability of the model. In summary, our future work will use multi-source image datasets and separate single and mixed topography areas to facilitate larger-scale cropland extraction.

5. Conclusions

Due to the significant heterogeneity and fragmentation of cropland in southern China, traditional cropland classification methods are generally ineffective. To address this issue, we developed a deep learning-based method for extracting cropland from VHR remote sensing images. Our approach uses an improved HRRS-U-Net model, which overcomes limitations in localization accuracy loss and spectral information exploitation exhibited by existing models. The HRRS-U-Net maintains high-resolution details to improve boundary delineation and introduces RL and CAM to extract representative features. We evaluated our method using GF-2 images in four cities (Qingyuan, Yangjiang, Guangzhou, and Shantou) across Guangdong Province and obtained resultant maps with an OA of 97.58%, an F1 of 0.915, a Kappa of 0.901, and a Mean of 0.891. Our method outperformed existing methods in accurately identifying the extent and precisely locating boundaries. Despite the high spatiotemporal heterogeneity of cropland and the limited sample dataset, our method demonstrated strong generalization and transportability. With support for multi-source image datasets, our method has significant potential for large-scale VHR cropland extraction, particularly in fragmented and heterogeneous landscape areas.

Author Contributions: Conceptualization, D.X. and L.L.; methodology, D.X., L.L. and H.X.; software, D.X. and M.X.; validation, D.X. and L.L.; formal analysis, D.X., H.X. and M.L.; investigation, D.X. and X.X.; resources, D.X., L.L. and X.X.; data curation, D.X. and L.L.; writing—original draft preparation, D.X.; writing—review and editing, D.X., H.X. and L.L.; visualization, D.X. and H.H.; supervision, L.L.; project administration, L.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (U1901601).

Data Availability Statement: The cropland maps in this study can be accessed from the corresponding author upon request.

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

References

1. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping Cropping Intensity in China Using Time Series Landsat and Sentinel-2 Images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624. [CrossRef]
2. Viana, C.M.; Freire, D.; Abrantes, P.; Rocha, J.; Pereira, P. Agricultural Land Systems Importance for Supporting Food Security and Sustainable Development Goals: A Systematic Review. *Sci. Total Environ.* **2022**, *806*, 150718. [CrossRef] [PubMed]

3. Di, Y.; Zhang, G.; You, N.; Yang, T.; Zhang, Q.; Liu, R.; Doughty, R.B.; Zhang, Y. Mapping Croplands in the Granary of the Tibetan Plateau Using All Available Landsat Imagery, A Phenology-Based Approach, and Google Earth Engine. *Remote Sens.* **2021**, *13*, 2289. [CrossRef]
4. Waldner, F.; Bellemans, N.; Hochman, Z.; Newby, T.; de Abelleyra, D.; Verón, S.R.; Bartalev, S.; Lavreniuk, M.; Kussul, N.; Maire, G.L.; et al. Roadside Collection of Training Data for Cropland Mapping Is Viable When Environmental and Management Gradients Are Surveyed. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 82–93. [CrossRef]
5. Liu, J.; Liu, M.; Tian, H.; Zhuang, D.; Zhang, Z.; Zhang, W.; Tang, X.; Deng, X. Spatial and Temporal Patterns of China's Cropland during 1990–2000: An Analysis Based on Landsat TM Data. *Remote Sens. Environ.* **2005**, *98*, 442–456. [CrossRef]
6. Wang, X.; Yan, F.; Su, F. Impacts of Urbanization on the Ecosystem Services in the Guangdong-Hong Kong-Macao Greater Bay Area, China. *Remote Sens.* **2020**, *12*, 3269. [CrossRef]
7. Liu, L.; Xu, X.; Liu, J.; Chen, X.; Ning, J. Impact of Farmland Changes on Production Potential in China during 1990–2010. *J. Geogr. Sci.* **2015**, *25*, 19–34. [CrossRef]
8. Potapov, P.; Turubanova, S.; Hansen, M.C.; Tyukavina, A.; Zalles, V.; Khan, A.; Song, X.-P.; Pickens, A.; Shen, Q.; Cortez, J. Global Maps of Cropland Extent and Change Show Accelerated Cropland Expansion in the Twenty-First Century. *Nat. Food* **2022**, *3*, 19–28. [CrossRef]
9. Fritz, S.; See, L.; McCallum, I.; You, L.; Bun, A.; Moltchanova, E.; Duerauer, M.; Albrecht, F.; Schill, C.; Perger, C.; et al. Mapping Global Cropland and Field Size. *Glob. Change Biol.* **2015**, *21*, 1980–1992. [CrossRef]
10. Hao, P.; Löw, F.; Biradar, C. Annual Cropland Mapping Using Reference Landsat Time Series—A Case Study in Central Asia. *Remote Sens.* **2018**, *10*, 2057. [CrossRef]
11. Oliphant, A.J.; Thenkabail, P.S.; Teluguntla, P.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K. Mapping Cropland Extent of Southeast and Northeast Asia Using Multi-Year Time-Series Landsat 30-m Data Using a Random Forest Classifier on the Google Earth Engine Cloud. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *81*, 110–124. [CrossRef]
12. Htitiou, A.; Boudhar, A.; Chehbouni, A.; Benabdelouahab, T. National-Scale Cropland Mapping Based on Phenological Metrics, Environmental Covariates, and Machine Learning on Google Earth Engine. *Remote Sens.* **2021**, *13*, 4378. [CrossRef]
13. Bartholomé, E.; Belward, A.S. GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [CrossRef]
14. Friedl, M.; Sulla-Menashe, D. *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006*; NASA: Washington, DC, USA, 2019.
15. Friedl, M.; Gray, J.; Sulla-Menashe, D. *MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006*; NASA: Washington, DC, USA, 2019.
16. Buchhorn, M.; Lesiv, M.; Tsendbazar, N.-E.; Herold, M.; Bertels, L.; Smets, B. Copernicus Global Land Cover Layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [CrossRef]
17. Yu, L.; Wang, J.; Clinton, N.; Xin, Q.; Zhong, L.; Chen, Y.; Gong, P. FROM-GC: 30 m Global Cropland Extent Derived through Multisource Data Integration. *Int. J. Digit. Earth* **2013**, *6*, 521–533. [CrossRef]
18. Yang, J.; Huang, X. The 30 m Annual Land Cover Dataset and Its Dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]
19. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
20. Panda, S.S.; Rao, M.N.; Thenkabail, P.; Fitzerald, J.E. Remote Sensing Systems—Platforms and Sensors: Aerial, Satellite, UAV, Optical, Radar, and LiDAR. In *Remotely Sensed Data Characterization, Classification, and Accuracies*; CRC Press: Boca Raton, FL, USA, 2015; ISBN 978-0-429-08939-8.
21. Zhang, H.; Liu, M.; Wang, Y.; Shang, J.; Liu, X.; Li, B.; Song, A.; Li, Q. Automated Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using Recurrent Residual U-Net. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102557. [CrossRef]
22. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy Assessment of Seven Global Land Cover Datasets over China. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 156–173. [CrossRef]
23. Zhang, C.; Dong, J.; Ge, Q. Quantifying the Accuracies of Six 30-m Cropland Datasets over China: A Comparison and Evaluation Analysis. *Comput. Electron. Agric.* **2022**, *197*, 106946. [CrossRef]
24. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine. *Remote Sens.* **2017**, *9*, 1065. [CrossRef]
25. Zhang, D.; Pan, Y.; Zhang, J.; Hu, T.; Zhao, J.; Li, N.; Chen, Q. A Generalized Approach Based on Convolutional Neural Networks for Large Area Cropland Mapping at Very High Resolution. *Remote Sens. Environ.* **2020**, *247*, 111912. [CrossRef]
26. Lu, R.; Wang, N.; Zhang, Y.; Lin, Y.; Wu, W.; Shi, Z. Extraction of Agricultural Fields via DASFNet with Dual Attention Mechanism and Multi-Scale Feature Fusion in South Xinjiang, China. *Remote Sens.* **2022**, *14*, 2253. [CrossRef]
27. Xu, L.; Ming, D.; Zhou, W.; Bao, H.; Chen, Y.; Ling, X. Farmland Extraction from High Spatial Resolution Remote Sensing Images Based on Stratified Scale Pre-Estimation. *Remote Sens.* **2019**, *11*, 108. [CrossRef]

28. Cai, Z.; Hu, Q.; Zhang, X.; Yang, J.; Wei, H.; He, Z.; Song, Q.; Wang, C.; Yin, G.; Xu, B. An Adaptive Image Segmentation Method with Automatic Selection of Optimal Scale for Extracting Cropland Parcels in Smallholder Farming Systems. *Remote Sens.* **2022**, *14*, 3067. [CrossRef]
29. Liu, Z.; Li, N.; Wang, L.; Zhu, J.; Qin, F. A Multi-Angle Comprehensive Solution Based on Deep Learning to Extract Cultivated Land Information from High-Resolution Remote Sensing Images. *Ecol. Indic.* **2022**, *141*, 108961. [CrossRef]
30. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [CrossRef]
31. Rufin, P.; Bey, A.; Picoli, M.; Meyfroidt, P. Large-Area Mapping of Active Cropland and Short-Term Fallows in Smallholder Landscapes Using PlanetScope Data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102937. [CrossRef]
32. Hu, Q.; Wu, W.; Song, Q.; Lu, M.; Chen, D.; Yu, Q.; Tang, H. How Do Temporal and Spectral Features Matter in Crop Classification in Heilongjiang Province, China? *J. Integr. Agric.* **2017**, *16*, 324–336. [CrossRef]
33. Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [CrossRef]
34. Samaniego, L.; Schulz, K. Supervised Classification of Agricultural Land Cover Using a Modified K-NN Technique (MNN) and Landsat Remote Sensing Imagery. *Remote Sens.* **2009**, *1*, 875–895. [CrossRef]
35. Waldner, F.; Canto, G.S.; Defourny, P. Automated Annual Cropland Mapping Using Knowledge-Based Temporal Features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 1–13. [CrossRef]
36. Lin, L.; Di, L.; Zhang, C.; Guo, L.; Di, Y.; Li, H.; Yang, A. Validation and Refinement of Cropland Data Layer Using a Spatial-Temporal Decision Tree Algorithm. *Sci. Data* **2022**, *9*, 63. [CrossRef] [PubMed]
37. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m Landsat-Derived Cropland Extent Product of Australia and China Using Random Forest Machine Learning Algorithm on Google Earth Engine Cloud Computing Platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [CrossRef]
38. Liu, R.; Tao, F.; Liu, X.; Na, J.; Leng, H.; Wu, J.; Zhou, T. RAANet: A Residual ASPP with Attention Framework for Semantic Segmentation of High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 3109. [CrossRef]
39. Wang, M.; Wang, J.; Cui, Y.; Liu, J.; Chen, L. Agricultural Field Boundary Delineation with Satellite Image Segmentation for High-Resolution Crop Mapping: A Case Study of Rice Paddy. *Agronomy* **2022**, *12*, 2342. [CrossRef]
40. Xiong, J.; Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Poehnelt, J.; Congalton, R.G.; Yadav, K.; Thau, D. Automated Cropland Mapping of Continental Africa Using Google Earth Engine Cloud Computing. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 225–244. [CrossRef]
41. Dong, J.; Xiao, X.; Menarguez, M.A.; Zhang, G.; Qin, Y.; Thau, D.; Biradar, C.; Moore, B., III. Mapping Paddy Rice Planting Area in Northeastern Asia with Landsat 8 Images, Phenology-Based Algorithm and Google Earth Engine. *Remote Sens. Environ.* **2016**, *185*, 142–154. [CrossRef]
42. Guo, Y.; Xia, H.; Pan, L.; Zhao, X.; Li, R.; Bian, X.; Wang, R.; Yu, C. Development of a New Phenology Algorithm for Fine Mapping of Cropping Intensity in Complex Planting Areas Using Sentinel-2 and Google Earth Engine. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 587. [CrossRef]
43. Zheng, J.; Liu, L.; Chen, H.; Gou, Y.; Che, Y.; Xu, H.; Li, Q. Characteristics of Warm Clouds and Precipitation in South China during the Pre-Flood Season Using Datasets from a Cloud Radar, a Ceilometer, and a Disdrometer. *Remote Sens.* **2019**, *11*, 3045. [CrossRef]
44. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
45. Kotaridis, I.; Lazaridou, M. Remote Sensing Image Segmentation Advances: A Meta-Analysis. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 309–322. [CrossRef]
46. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
47. Aminoff, E.M.; Baror, S.; Roginek, E.W.; Leeds, D.D. Contextual Associations Represented Both in Neural Networks and Human Behavior. *Sci. Rep.* **2022**, *12*, 5570. [CrossRef] [PubMed]
48. Qing, Y.; Liu, W. Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism. *Remote Sens.* **2021**, *13*, 335. [CrossRef]
49. Xu, W.; Deng, X.; Guo, S.; Chen, J.; Sun, L.; Zheng, X.; Xiong, Y.; Shen, Y.; Wang, X. High-Resolution U-Net: Preserving Image Details for Cultivated Land Extraction. *Sensors* **2020**, *20*, 4064. [CrossRef]
50. Shi, H.; Cao, G.; Zhang, Y.; Ge, Z.; Liu, Y.; Fu, P. H2A2Net: A Hybrid Convolution and Hybrid Resolution Network with Double Attention for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 4235. [CrossRef]
51. Mei, X.; Pan, E.; Ma, Y.; Dai, X.; Huang, J.; Fan, F.; Du, Q.; Zheng, H.; Ma, J. Spectral-Spatial Attention Networks for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 963. [CrossRef]
52. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef]
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

54. Tong, W.; Chen, W.; Han, W.; Li, X.; Wang, L. Channel-Attention-Based DenseNet Network for Remote Sensing Image Scene Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4121–4132. [CrossRef]
55. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
56. Guo, J.; Jia, N.; Bai, J. Transformer Based on Channel-Spatial Attention for Accurate Classification of Scenes in Remote Sensing Image. *Sci. Rep.* **2022**, *12*, 15473. [CrossRef] [PubMed]
57. Huang, G.; Zhu, J.; Li, J.; Wang, Z.; Cheng, L.; Liu, L.; Li, H.; Zhou, J. Channel-Attention U-Net: Channel Attention Mechanism for Semantic Segmentation of Esophagus and Esophageal Cancer. *IEEE Access* **2020**, *8*, 122798–122810. [CrossRef]
58. Ronneberger, O.; Fischer, P.; Brox, T. 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; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
59. Bernstein, L.S.; Adler-Golden, S.M.; Sundberg, R.L.; Levine, R.Y.; Perkins, T.C.; Berk, A.; Ratkowski, A.J.; Felde, G.; Hoke, M.L. A New Method for Atmospheric Correction and Aerosol Optical Property Retrieval for VIS-SWIR Multi- and Hyperspectral Imaging Sensors: QUAC (QUick Atmospheric Correction). In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing SymposiumIGARSS '05, Seoul, Republic of Korea, 29 July; 2005; Volume 5, pp. 3549–3552.
60. Zhang, Y. Problems in the Fusion of Commercial High-Resolution Satellites Images as Well as LANDSAT 7 Images and Initial Solutions. In Proceedings of the Proceedings of the ISPRS, CIG, and SDH Joint International Symposium on Geospatial Theory, Processing and Applications,, Ottawa, ON, Canada, 9–12 July 2002; pp. 9–12.
61. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
62. Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-Supervised Nets. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, San Diego, CA, USA, 9 May 2015; Lebanon, G., Vishwanathan, S.V.N., Eds.; PMLR: San Diego, CA, USA, 2015; Volume 38, pp. 562–570.
63. Li, S.; Wan, L.; Tang, L.; Zhang, Z. MFEAFN: Multi-Scale Feature Enhanced Adaptive Fusion Network for Image Semantic Segmentation. *PLoS ONE* **2022**, *17*, e0274249. [CrossRef] [PubMed]
64. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
65. Yeung, M.; Sala, E.; Schönlieb, C.-B.; Rundo, L. Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [CrossRef] [PubMed]
66. Cortes, C.; Mohri, M.; Rostamizadeh, A. L2 Regularization for Learning Kernels. *arXiv* **2012**, arXiv:1205.2653.
67. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
68. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) Las Condes, Chile, 11–18 December 2015; pp. 1026–1034. [CrossRef]
69. Olofsson, P.; Foody, G.; Herold, M.; Stehman, S.; Woodcock, C.; Wulder, M. Good Practices for Estimating Area and Assessing Accuracy Of Land Change. *Remote Sens. Environ.* **2013**, *148*, 42–57. [CrossRef]
70. Alganci, U. Dynamic Land Cover Mapping of Urbanized Cities with Landsat 8 Multi-Temporal Images: Comparative Evaluation of Classification Algorithms and Dimension Reduction Methods. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 139. [CrossRef]
71. Congalton, R.G.; Green, K. What Are the Thematic Map Classes to Be Assessed. In *Assessing the Accuracy of Remotely Sensed Data*; CRC Press: Boca Raton, FL, USA, 2008; ISBN 978-0-429-14397-7.
72. Ye, S.; Pontius, R.G., Jr.; Rakshit, R. A Review of Accuracy Assessment for Object-Based Image Analysis: From per-Pixel to per-Polygon Approaches. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 137–147. [CrossRef]
73. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef]
74. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
75. Xia, J.; Yokoya, N.; Adriano, B.; Kanemoto, K. National High-Resolution Cropland Classification of Japan with Agricultural Census Information and Multi-Temporal Multi-Modality Datasets. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103193. [CrossRef]
76. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
77. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
78. Turker, M.; Kok, E.H. Field-Based Sub-Boundary Extraction from Remote Sensing Imagery Using Perceptual Grouping. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 106–121. [CrossRef]
79. Shrestha, S.; Vanneschi, L. Improved Fully Convolutional Network with Conditional Random Fields for Building Extraction. *Remote Sens.* **2018**, *10*, 1135. [CrossRef]

80. Zhang, Y.; Li, W.; Gong, W.; Wang, Z.; Sun, J. An Improved Boundary-Aware Perceptual Loss for Building Extraction from VHR Images. *Remote Sens.* **2020**, *12*, 1195. [CrossRef]
81. Wang, C.; Qiu, X.; Huan, H.; Wang, S.; Zhang, Y.; Chen, X.; He, W. Earthquake-Damaged Buildings Detection in Very High-Resolution Remote Sensing Images Based on Object Context and Boundary Enhanced Loss. *Remote Sens.* **2021**, *13*, 3119. [CrossRef]
82. Wang, H.; Cao, P.; Wang, J.; Zaiane, O.R. UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. *Proc. Conf. AAAI Artif. Intell.* **2022**, *36*, 2441–2449. [CrossRef]