



Article

A Lightweight Winter Wheat Planting Area Extraction Model Based on Improved DeepLabv3+ and CBAM

Yao Zhang ¹, Hong Wang ^{1,2,*}, Jiahao Liu ¹ , Xili Zhao ¹, Yuting Lu ¹, Tengfei Qu ² , Haozhe Tian ¹, Jingru Su ¹, Dingsheng Luo ¹ and Yalei Yang ²

¹ College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830046, China; zhangyao@stu.xju.edu.cn (Y.Z.); 107552103460@stu.xju.edu.cn (J.L.); 107552103482@stu.xju.edu.cn (X.Z.); 107552101101@stu.xju.edu.cn (Y.L.); 107552103458@stu.xju.edu.cn (H.T.); 107552203704@stu.xju.edu.cn (J.S.); 107552203697@stu.xju.edu.cn (D.L.)

² Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; qutengfei@mail.bnu.edu.cn (T.Q.); yangyalei@mail.bnu.edu.cn (Y.Y.)

* Correspondence: wanghong@bnu.edu.cn

Abstract: This paper focuses on the problems of inaccurate extraction of winter wheat edges from high-resolution images, misclassification and omission due to intraclass differences as well as the large number of network parameters and long training time of existing classical semantic segmentation models. This paper proposes a lightweight winter wheat planting area extraction model that combines the DeepLabv3+ model and a dual-attention mechanism. The model uses the lightweight network MobileNetv2 to replace the backbone network Xception of DeepLabv3+ to reduce the number of parameters and improve the training speed. It also introduces the lightweight Convolutional Block Attention Module (CBAM) dual-attention mechanism to extract winter wheat feature information more accurately and efficiently. Finally, the model is used to complete the dataset creation, model training, winter wheat plantation extraction, and accuracy evaluation. The results show that the improved lightweight DeepLabv3+ model in this paper has high reliability in the recognition extraction of winter wheat, and its recognition results of OA, mPA, and mIoU reach 95.28%, 94.40%, and 89.79%, respectively, which are 1.52%, 1.51%, and 2.99% higher than those for the original DeepLabv3+ model. Meanwhile, the model's recognition accuracy was much higher than that of the three classical semantic segmentation models of UNet, ResUNet and PSPNet. The improved lightweight DeepLabv3+ also has far fewer model parameters and training time than the other four models. The model has been tested in other regions, and the results show that it has good generalization ability. The model in general ensures the extraction accuracy while significantly reducing the number of parameters and satisfying the timeliness, which can achieve the fast and accurate extraction of winter wheat planting sites and has good application prospects.

Keywords: deep learning; lightweight; DeepLabv3+; attention mechanism; winter wheat recognition



Citation: Zhang, Y.; Wang, H.; Liu, J.; Zhao, X.; Lu, Y.; Qu, T.; Tian, H.; Su, J.; Luo, D.; Yang, Y. A Lightweight Winter Wheat Planting Area Extraction Model Based on Improved DeepLabv3+ and CBAM. *Remote Sens.* **2023**, *15*, 4156. <https://doi.org/10.3390/rs15174156>

Academic Editors: Enrico Corrado Borgogno Mondino, Filippo Sarvia, Samuele De Petris and Tommaso Orusa

Received: 22 June 2023

Revised: 12 August 2023

Accepted: 19 August 2023

Published: 24 August 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

With socio-economic development and rapid population growth, as well as the impact of global climate change and urban expansion, China's sustainable agricultural development is facing huge challenges [1]. Socio-economic development has significantly increased the cost of food production, increasing the risk of food supply shortage [2]. Rapid population growth has led to an increase in the demand for food, and the imbalance between food production and population size is becoming more and more obvious [3]. Global climate change can reduce the food production by destroying the environment in which food grows [4]. The destruction of arable land and environmental pollution caused by industrialization and urban expansion have reduced the space for sustainable agricultural development [5]. The issue of food supply and security has become one of the key issues to which China currently attaches importance [6]. Wheat is one of the world's main sources of

food crops [7], providing the most calories and protein [8], and its yield and quality are related to food security [9]. China is a major global wheat producer [10], accounting for approximately 17.98% of the world's total wheat production [11], with an annual output of 131.4 million tons [12]. Wheat is the third largest crop in China after rice and corn [13]. Furthermore, China's winter wheat production accounts for 95% of the total wheat production in the country [14]. China's winter wheat planting area is about 226 million hm^2 [15], and it plays an important role in guaranteeing food security [16]. Therefore, the timely, efficient, and accurate acquisition of winter wheat planting area information and spatial distribution is crucial for government management departments to formulate agricultural and food policies, optimize land resources, adjust agricultural planting structures, and ensure national food security [16,17]. It also has important applications in crop census, growth monitoring, yield prediction, and natural disaster monitoring and assessment [18–20]. The accurate and rapid extraction of winter wheat planting areas is an important prerequisite for the refined management of winter wheat.

The traditional way to obtain crop planting information is usually through on-site investigation and statistical data, and then, the survey data and statistical reports are reported from the grassroots level to obtain nationwide data [21]. These data take a significant amount of time to collect and process and are not available in a timely manner, as data confidentiality delays the release of planting information [21]. This method not only wastes time and manpower resources but also has low efficiency and a long cycle, is unable to provide an accurate spatial distribution of crop categories, and has a certain lag in time, making it difficult to meet the demand for the real-time control of agricultural information [19,22].

With the rapid development of remote sensing technology, the use of remote sensing for the large-scale monitoring of agricultural cultivation information is widely used [23]. Obtaining the crop planting area and distribution status through remote sensing methods has the advantages of fast information extraction, wide monitoring range, strong timeliness, and reliable results [24]. At the same time, remote sensing greatly reduces the investment of time, manpower, material resources, and financial resources and can quickly and accurately extract crop planting areas and obtain spatial distribution information [25,26]. Remote sensing technology has become one of the main methods for extracting crop planting information, which provides good technical support for crop planting information extraction and has good application and development prospects [27]. For example, Qi et al. [14] completed 10 m winter wheat mapping in Shandong Province, China, using Sentinel-2 imagery and coarse-resolution maps. Du et al. [28] used WorldView-2 high-resolution images to realize the classification and area extraction of smallholder crops based on the deep semantic segmentation method. In the process of extracting crop planting areas from medium- and low-resolution remote sensing data, due to the low spatial resolution, it is usually only applicable to large fields with high mechanization, and the extraction accuracy cannot be guaranteed for smaller plots [20]. However, in most parts of China, crop planting is still in the traditional agricultural stage, with small crop cultivation plots and the alternating planting of multiple crops [29]. The process of extracting crop planting areas using mid- to low-spatial-resolution remote sensing images is highly susceptible to mixed pixel problems, resulting in significantly poor recognition performance and low accuracy [14,28]. Because the spatial resolution of medium- and low-resolution data is generally larger than the size of a small farm field, a pixel may contain multiple land classes and cannot accurately extract the area of farmland [29].

At present, high-resolution remote sensing data are widely used in the fields of crop recognition and the extraction of planting areas [30,31]. The Gaofen-2 (GF-2) satellite is the first civilian optical remote sensing satellite developed by China with a spatial resolution better than 1 m [32]. The GF-2 remote sensing satellite has a sub-meter-level spatial resolution, and the images contain more abundant land information [32]. GF-2 high spatial resolution images can identify small fields and can more accurately reflect crop information at the field scale of smallholder farms, providing a new high-resolution remote sensing

data source for remote sensing crop monitoring [33]. However, with images of high spatial resolution, the spectral resolution is low, and the spectral statistical characteristics are not as stable as those of images with medium and low spatial resolution [34]. At the same time, the spectral differences between the same ground objects will increase, the spectral differences between different ground objects will decrease, the spectral overlap is serious, and the phenomenon of “different spectra of the same objects” and “same spectra of foreign objects” is common [34,35]. Lower interclass spectral variability and higher intraclass spectral variability in high-resolution images lead to ambiguity in classification [36]. On the other hand, high-resolution images provide more detailed descriptions and depictions of small land features, and the detailed information of features is amplified, which will increase the amount of data dramatically, resulting in a sharp increase in the complexity of crop feature extraction [37]. Traditional supervised and unsupervised classification methods are applied to feature extraction in high-resolution remote sensing images, but only low-level features such as color, shape, and texture can be extracted from the images, while higher-level semantic features cannot be extracted, and the discriminative ability of the extracted features is poor, often resulting in unsatisfactory classification results [37–40]. In order to obtain higher-level semantic features from high-resolution remote sensing images, scholars have conducted in-depth research on the remote sensing recognition and classification of various crops using methods such as deep learning image semantic segmentation [28,41]. Simultaneously applying GF-2 images to identify and extract crop planting information is of great significance for promoting the application of domestically produced satellites in China and promoting the development of remote sensing technology in China.

Deep learning, as a branch of artificial intelligence, provides new ideas for accurate crop classification. The crop remote sensing classification feature extraction process based on deep learning requires no or only a small amount of human involvement to extract the essential, abstract features of the target with advantages such as computational power oriented to big data processing [42,43]. The convolutional neural network (CNN) is a widely used deep learning network model in the field of computer vision with outstanding performance in image classification, target detection, and semantic segmentation tasks [44]. In addition, CNN also shows obvious advantages in processing remote sensing image data and has continuously achieved remarkable results in the classification and recognition of remote sensing images in recent years [45]. Crop recognition extraction based on deep learning has become an important field of research in the area of agricultural remote sensing as a result of improvements in computer processing power, and numerous CNN-based techniques have been successfully used in the study of crop recognition and classification [46]. Deep learning has the benefit over conventional algorithms in that multi-layer neural networks can automatically extract useful characteristics. Particularly, the CNN model extracts high-level semantic information from images in addition to local detailed features [47]. CNN consists of multiple nonlinear mapping layers, which improve crop recognition accuracy by mining spatial correlations between target pixels and performing combinatorial analysis to obtain high-dimensional features of the target [33]. For example, Sun et al. [48] realized the intelligent extraction of winter wheat planting areas in complex agricultural landscapes based on CNN. Chen et al. [19] extracted the spatial distribution of crops using GF-2 images based on CNN. The DeepLabv3+ semantic segmentation model is a more typical and high-accuracy network model in the field of semantic segmentation, which is applied to image segmentation in various complex scenes [28,49,50]. However, the DeepLabV3+ model also has some shortcomings. First, the large number of layers and the large number of parameters in its feature extraction network, Xception, lead to its high computation and resource consumption, making it difficult for the model to meet the demand of large-scale and real-time detection [51]. Second, the spatial dimension of the input data gradually decreases during feature extraction by the encoder, causing valuable information to be lost, and detailed restoration is unable to be accomplished during decoding, which eventually leads to a relatively low accuracy of target edge recognition [51].

In summary, we aim to effectively solve the problems of inaccurate extraction of winter wheat edges from high-resolution images, including misclassification and omission due to intraclass differences, as well as the large number of network parameters and long training time of existing classical semantic segmentation models. This paper proposes a lightweight winter wheat plantation extraction model based on DeepLabv3+ and an attention mechanism. The main contributions of this study are as follows. (1) We use the lightweight network MobileNetv2 to replace the backbone network, Xception, of DeepLabv3+, thus reducing the number of parameters and improving the training speed. (2) We introduce the lightweight Convolutional Block Attention Module (CBAM) attention mechanism to filter the background information, which makes the model pay more attention to the key information and extract the winter wheat feature information more accurately and efficiently. (3) We rely on the improved model to achieve the fast and accurate extraction of winter wheat planting sites from GF-2 remote sensing images, which provides reference for other crop extraction research.

2. Data and Methods

2.1. Study Area

In this study, the eastern part of Yuncheng Basin in Shanxi Province, located at longitude $110^{\circ}15'$ – $112^{\circ}04'E$ and latitude $34^{\circ}35'$ – $35^{\circ}49'N$, was the study area (Figure 1). Figure 1a shows the location of Shanxi Province in China; Figure 1b shows the location of the study area in Shanxi Province; and Figure 1c shows the location map of the study area. The study area is about 35 km wide from east to west and 57 km long from north to south, with a total area of about 1111 km². The Yuncheng Basin is bounded by the Emei Ridge to the north, the Yellow River to the west, and the Zhongtiao Mountains to the east and south, including the entire Sushui River Basin. Due to its location within the Zhongtiao Mountains pre-mountain fault zone, it is a strongly subsiding basin and belongs to one of the three important components of the Weihe River fault subsidence basin belt. The interior of the basin is mostly river–lake phase accumulation except for a flood fan group around the foot of the Zhongtiao Mountains. The study area has a warm–temperate continental monsoon climate, with precipitation occurring mainly in summer, with an average annual precipitation of 537 mm and an average annual temperature of 13 °C [52]. Except for some low-lying areas with salinization, most of the Yuncheng Basin has fertile soil and sufficient rainfall, and it has the longest frost-free period in the province. With a long history of agricultural development and utilization, winter wheat and other grain crops are widely grown in the basin, which is one of the major grain-producing areas in Shanxi Province [53].

2.2. Data

2.2.1. Remote Sensing Data

In this study, Gaofen-2 (GF-2) images were used as the data source of remote sensing images to meet the requirement of high spatial resolution. The parameters of GF-2 satellite data are shown in Table 1. GF-2 images consisted of one panchromatic band and four multispectral bands of red, green, blue, and near infrared, with 1 m spatial resolution in the panchromatic band and 4 m spatial resolution in the multispectral band [19]. GF-2 image data with no clouds on 19 April 2021 within the study area were collected and downloaded, and the entire study area was covered with 6 views of GF-2 images. At this time, the winter wheat in the study area was in the jointing–booting stage and growing vigorously, and it was in a different growth period from other crops; the spectral differences between crops were large, so it was easy to identify and extract winter wheat by remote sensing.

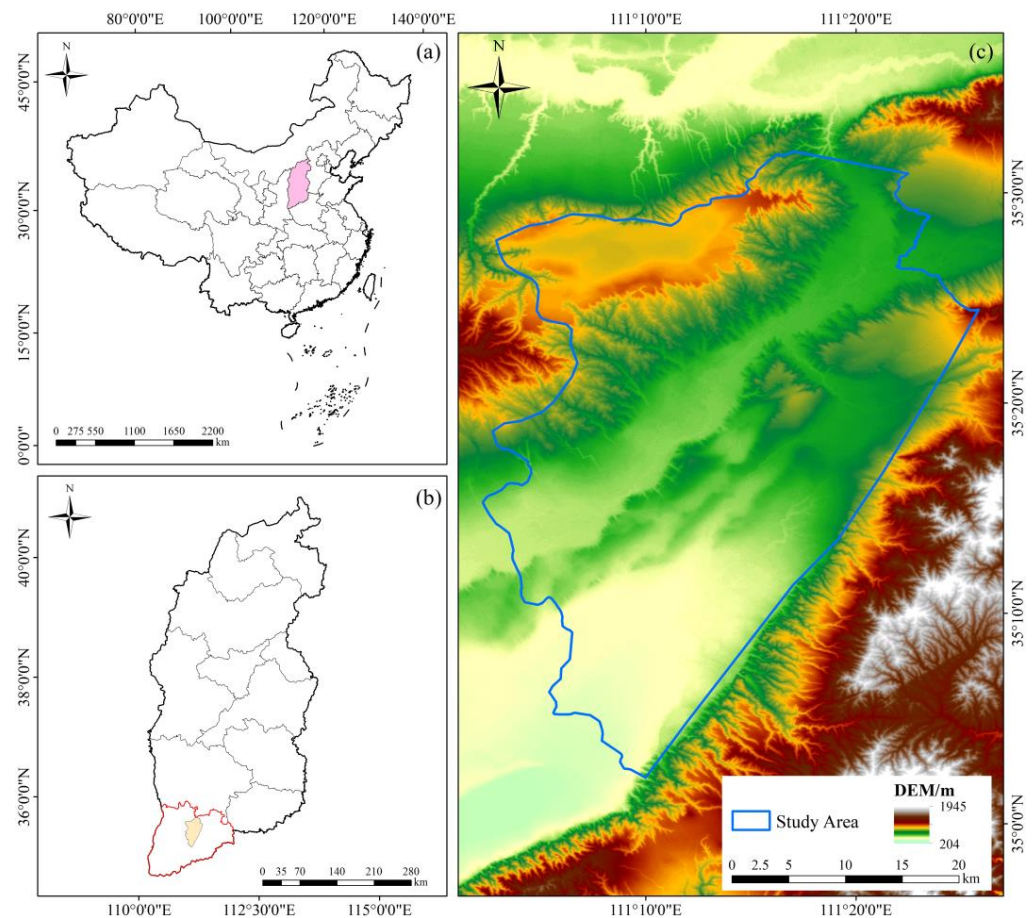


Figure 1. The map of location of study area: (a) the location of Shanxi Province in China; (b) the location of the study area in Shanxi Province; (c) the location map of the study area.

Table 1. Parameters of GF-2 satellite data.

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

The GF-2 image data in the study area were pre-processed using Environment for Visualizing Images (ENVI, version 5.3.1) software, and the pre-processing workflow is shown in Figure 2. The pre-processing process of multispectral images included radiometric calibration, atmospheric correction, and orthorectification, and the pre-processing process of panchromatic images included radiometric calibration and orthorectification. Radiometric calibration is used to map the digital values of the remote sensing image to the actual radiometric measure so that an accurate relationship is established between each pixel value in the image and the ground reflectance or radiance [54]. Atmospheric correction is used to remove the effects of atmospheric scattering and absorption during data transmission to restore the true reflectance information of the ground surface [54]. Orthorectification corrects image distortions caused by changes in terrain so that the image pixels correspond

to their actual positions on the ground [54]. Finally, the NNDiffuse Pan Sharpening fusion method was used to fuse the panchromatic images with the multispectral images to obtain multispectral remote sensing images with a spatial resolution of 1 m.

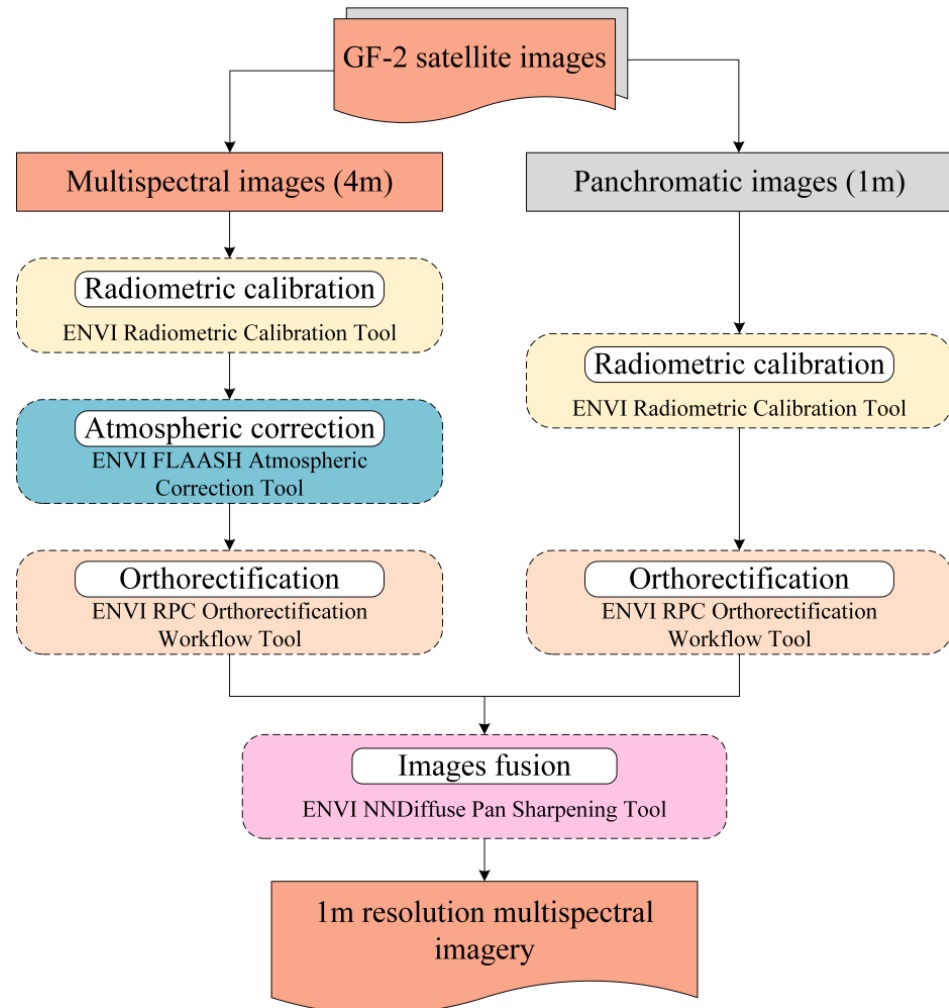


Figure 2. GF-2 image pre-processing workflow.

2.2.2. Ground Survey Data

The collection of ground survey data helps to produce more accurate training samples for the classification process as well as to verify the subsequent winter wheat identification results and planting area extraction accuracy [33]. The data of winter wheat ground survey samples in the study area were collected in the field on 26 April 2021, and the coordinate location information of the winter wheat field survey sample points were accurately recorded using a GNSS receiver during the field survey. The winter wheat at the measured sample points were photographed and kept, which helped to grasp the characteristics of winter wheat more accurately when making training labels later. The samples were collected to meet the relevant accuracy requirements and were required to cover the whole study area uniformly. A total of 78 winter wheat sample points were obtained from the ground survey, and the distribution of the ground sample points is shown in Figure 3.

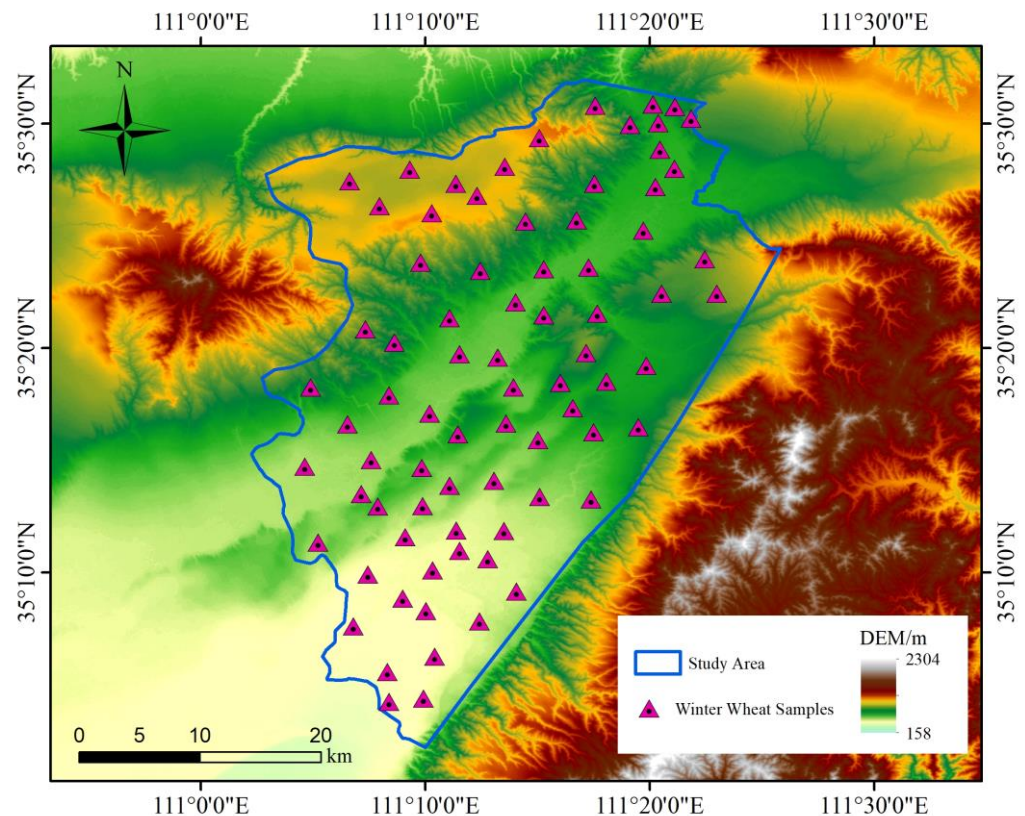


Figure 3. Spatial distribution map of sample points for ground survey of winter wheat.

2.2.3. Dataset Production

The pre-processed GF-2 images of the study area were combined with ground survey data and visual interpretation, and the pre-processed GF-2 images were manually labeled using LabelMe software (version 3.16.7, <http://labelme.csail.mit.edu/Release3.0/>, accessed on 16 March 2022) to produce the dataset. The winter wheat planting areas were labeled as winter wheat and the other areas were labeled as background. Labeled images were created and cropped to 600×600 pixels, as shown in Figure 4. A final dataset consisting of 400 original images and corresponding labels is formed, and the dataset is randomly partitioned by 8:2, of which 80% of the dataset was used for training and 20% was used for verification of recognition accuracy.

2.3. Research Methodology

2.3.1. DeepLabv3+ Model

The DeepLabv3+ model is a semantic segmentation model that was proposed in 2018 with the model structure shown in Figure 5 [55], which is a more typical network structure with higher accuracy in the field of semantic segmentation [56]. DeepLabv3+ is proposed as a further improvement on the DeepLabv3 model, and their differences mainly lie in the introduction of the encoder–decoder structure in DeepLabv3+, the different dilation rates in the ASPP module, and the differences in the backbone networks used [51]. DeepLabv3+ uses Xception [57] as the backbone network and combines it with the atrous spatial pyramid pooling (ASPP) module to form the encoder [49]. The encoder is used to extract deep feature information from the input image; it usually consists of multiple convolutional layers to learn different abstract features of the image, gradually increasing the extraction of image features from shallow to deep layers [32]. The encoder gradually reduces the size of the feature map by layer-by-layer convolution and pooling operations while increasing the number of channels of the features to achieve the extraction and compression of image features [33]. The ASPP module is designed to capture image multi-

scale contextual information using dilated convolution with different dilation rates for the better understanding of objects of different sizes and scales in an image [58]. DeepLabv3+ has improved the ASPP module by using a larger dilation rate and adding a larger receptive field, which further improves the performance of image segmentation. In the decoder part, the fused depth feature map is up-sampled four times using bilinear interpolation; then, it is spliced and fused with the low-level feature map extracted by Xception; and finally, the final segmentation result is obtained with the same size as the original image by 3×3 convolution and 4-fold up-sampling using bilinear interpolation [47].

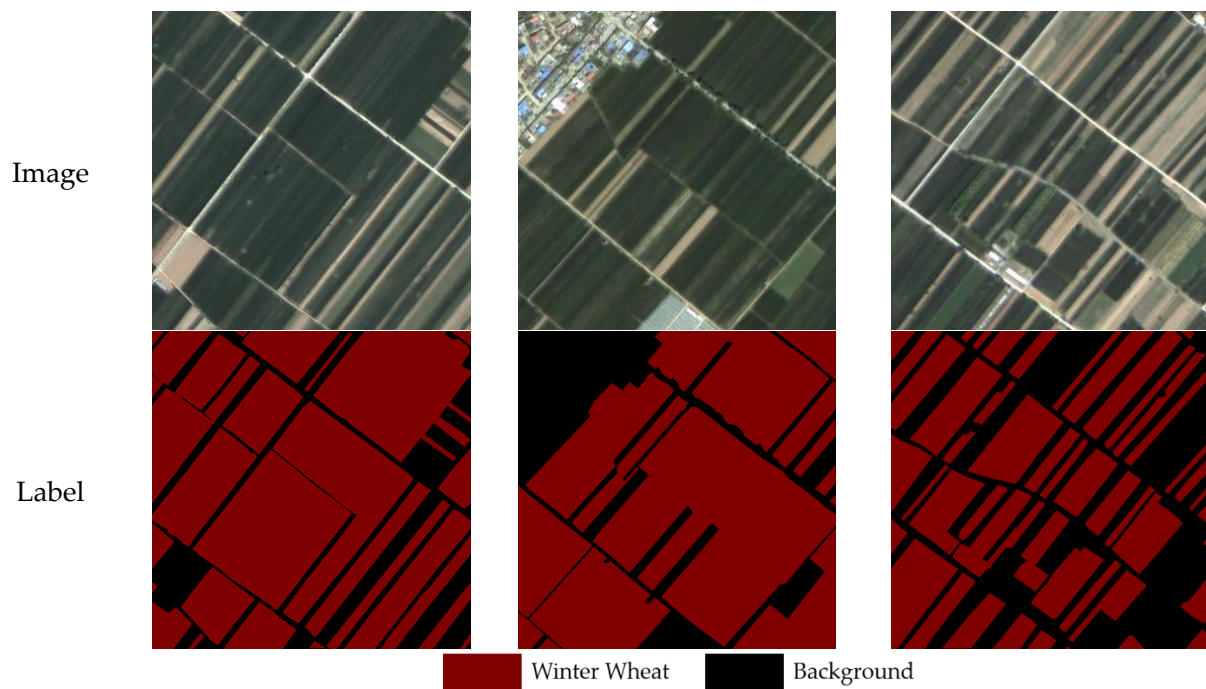


Figure 4. Examples of images and labels for the dataset.

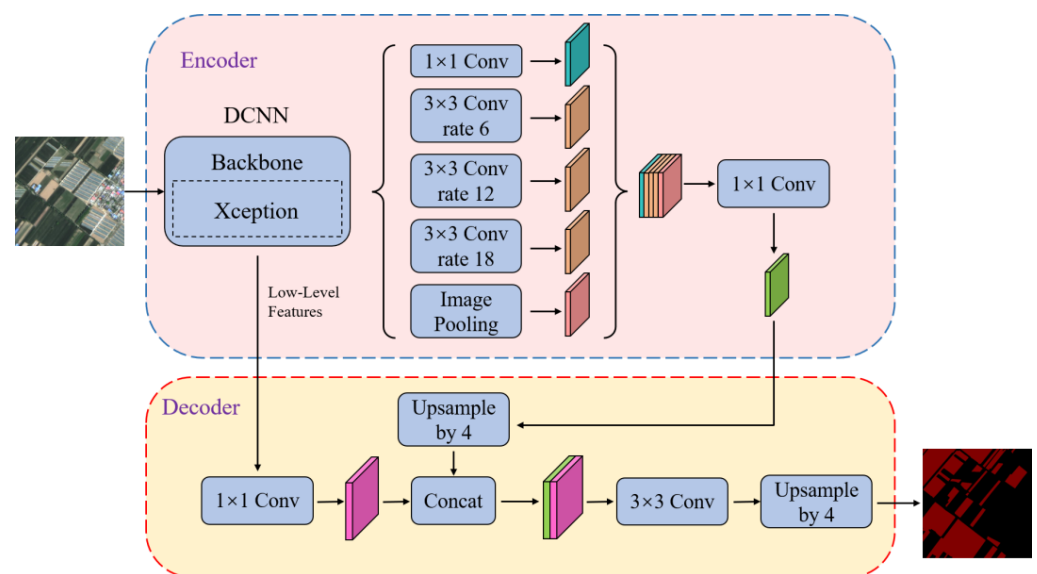


Figure 5. Structure of DeepLabv3+ model.

2.3.2. Replacing the Backbone Network

The traditional DeepLabv3+ semantic segmentation model uses Xception as the backbone network [49]. In contrast, this study used the lightweight improved MobileNetv2 network as the backbone network of the semantic segmentation model. MobileNetv2 is a lightweight network model proposed by Google [59], which introduces the inverted residual module and linear bottleneck layer based on the use of deep separable convolution, which greatly reduces the number of model parameters and thus makes the network converge faster [60].

The MobileNetv2 network does not use activation functions in the low-dimensional convolutional layers, while ReLU6 activation functions are used in the other layers to prevent the nonlinear layers from corrupting too much feature information [59]. This feature extraction network first obtains features of the same dimension by 3×3 deep convolution and ReLU activation function; then, it obtains the reduced dimensional features by 1×1 convolution and ReLU6 processing, and finally, it performs dimensionality enhancement using 1×1 convolution [51]. The inverted residual module is primarily utilized to improve the network's capacity for feature extraction and the efficient transmission of multi-layer feature information. For this module, first, 1×1 convolution is used to up-dimension; then, 3×3 depth convolution is used to extract features; and finally, 1×1 convolution is used to downscale the features to obtain feature information [59].

2.3.3. Attention Mechanism

The attention mechanism learns the contextual information of the image and captures the internal relevance of the image; the core idea is for the model to focus on critical regional information and ignore irrelevant information [47]. The attention mechanism is made up of the channel attention module and the spatial attention module, where the channel attention module concentrates on which features of the image are more critical and the spatial attention module concentrates on which region features are more critical [58]. In order to accurately extract winter wheat planting regions, the extraction of winter wheat edge features is made more accurate and efficient with the attention mechanism. In this study, a Convolutional Block Attention Module (CBAM) combining the above two modules was added to the semantic segmentation network [61]. This module focuses not only on the scale of each channel but also on the scale of each pixel point, which can be adaptively optimized according to the features of the input image [61]. In addition, a major advantage of the CBAM is that it is lightweight and can be seamlessly integrated into any neural network for plug-and-play [62].

The CBAM structure is shown in Figure 6 [61], where the input feature maps in the channel attention module are subjected to a pooling operation to obtain the weights of each channel of the input feature maps and apply them to the spatial attention module. Subsequently, the spatial attention module takes the maximum value and the average value on each channel of each feature point; after that, the weights of each feature point of the input feature map are obtained by the same operation as the channel attention; and finally, the weights are multiplied with the original input feature map and convolved to obtain the deep features containing multi-scale contextual information [33].

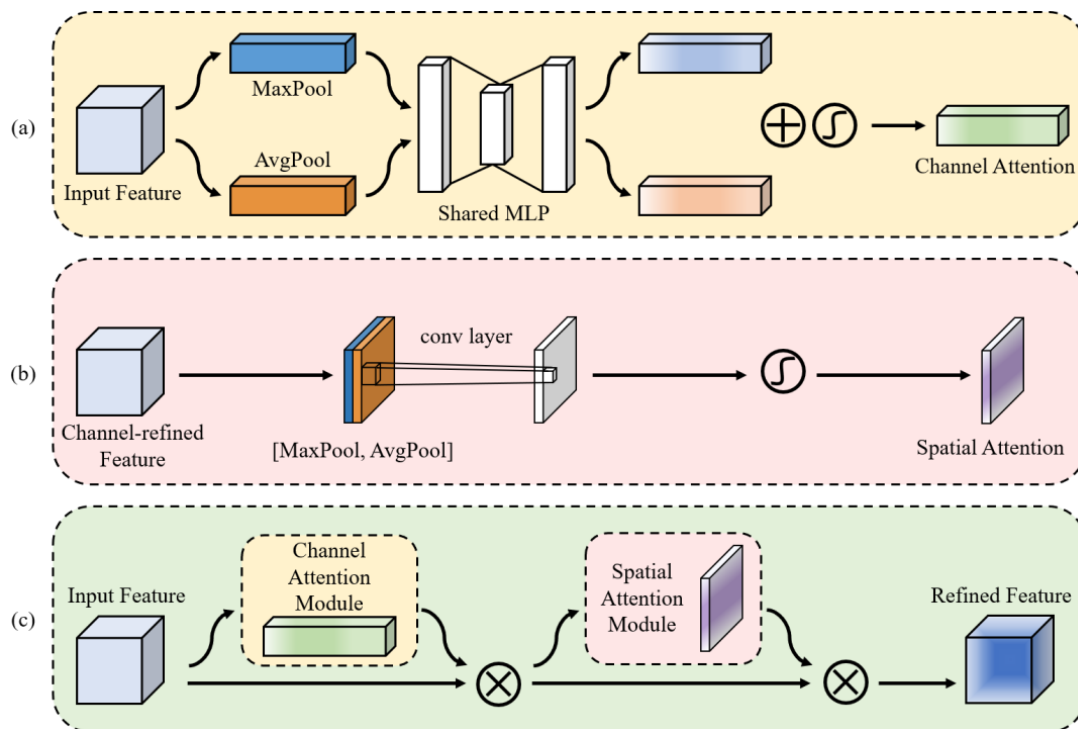


Figure 6. Structure of CBAM: (a) Channel attention module; (b) spatial attention module; (c) CBAM.

2.3.4. Model Improvement

Winter wheat mostly presents a faceted structure on satellite remote sensing images, with simple semantic information but rich detailed information, which puts high demands on the detail extraction ability of a segmentation network. The original DeepLabv3+ network and other classical segmentation networks face more complex and diverse datasets with more types of object features and larger data volumes, so the feature extraction part of the backbone network used is more complex, enabling it to learn more complex feature patterns [49]. The Xception structure of the backbone network used in the original DeepLabv3+, its complex structure and large number of parameters are suitable for segmentation tasks with more object types, but at the same time, it brings disadvantages such as huge computation volume and difficulty in training, which are not fully suitable for winter wheat recognition tasks [51]. For winter wheat features on satellite remote sensing images, this study applied the lightweight MobileNetv2 backbone network to the encoder–decoder structure of DeepLabv3+ as an encoder to extract features. The MobileNetv2 network has fewer parameters, and more direct connections are added in the network, having fewer parameters, easier training, and faster convergence compared with Xception [51]. In addition, the receptive field of convolution is increased with the addition of dilated convolution, which can well combine the semantic information of the image, and it substantially improves the operation speed, so it is suitable for the winter wheat recognition task from remote sensing images [33]. The introduction of the CBAM makes the model pay more attention to the key information and can extract the winter wheat feature information map more accurately and efficiently, and the detailed information of the image and the spatial correlation of pixels in a larger area are used to more accurately recognize in most areas and improve the recognition effect of winter wheat plantation with different area sizes [33]. Figure 7 shows the network model diagram of the method in this study.

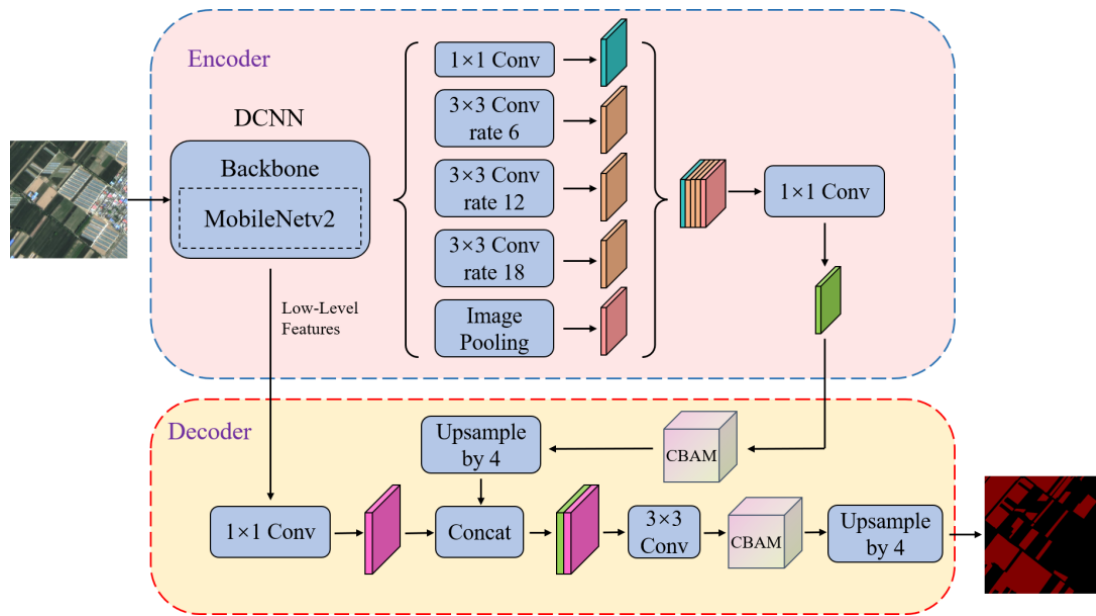


Figure 7. Structure of improved lightweight DeepLabv3+ model.

2.3.5. Evaluation Metrics

To validate the performance of the improved winter wheat recognition model, six accuracy evaluation metrics, Precision, Recall, Intersection over Union (IoU), mean pixel accuracy (mPA), mean Intersection over Union (mIoU), and overall accuracy (OA), were used to evaluate the recognition accuracy of the model.

Precision represents the proportion of actual winter wheat in pixels predicted by the model as winter wheat. The calculation formula is as follows [62]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Recall represents the proportion of pixels correctly predicted as winter wheat to all actual winter wheat pixels. The calculation formula is as follows [62]:

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

IoU is the ratio between the intersection area and the union area of the predicted and real regions. The calculation formula is as follows [33]:

$$\text{IoU} = \frac{TP}{FN + FP + TP} \quad (3)$$

The mPA is the average of the proportion of correctly categorized pixels for all categories. The calculation formula is as follows [63]:

$$\text{mPA} = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FN} \quad (4)$$

The mIoU is the average of the IoU of all categories. The calculation formula is as follows [9]:

$$\text{mIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (5)$$

OA is the ratio of the number of all correctly predicted pixels to the total number of pixels in the sample. The calculation formula is as follows [64]:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

where k is the number of identified species except background, which is 1 in this study; TP is True Positive; TN is True Negative; FP is False Positive; and FN is False Negative. TP and TN represent the number of pixels correctly predicted as “winter wheat” or “non-winter wheat”, respectively, FP represents the number of “non-winter wheat” pixels incorrectly judged as “winter wheat”, and FN represents the number of “winter wheat” pixels incorrectly judged as “non-winter wheat”.

3. Results

3.1. Model Training

In this study, deep learning was run in the Windows 10 operating system; the computer was equipped with one NVIDIA GeForce RTX 3070Ti GPU (8 G memory), CUDA version 11.6; the Pytorch1.11.0 deep learning framework was used to build the network; and the software environment was Anaconda (python3.8). The initial learning rate of the model was 0.01, the cosine annealing decay attenuated the learning rate, and the SGD optimizer was used with a batch size of 2 and an epoch number of 300.

3.2. Model Training Results

The training results of the improved lightweight winter wheat recognition model in this study are shown in Figure 8.

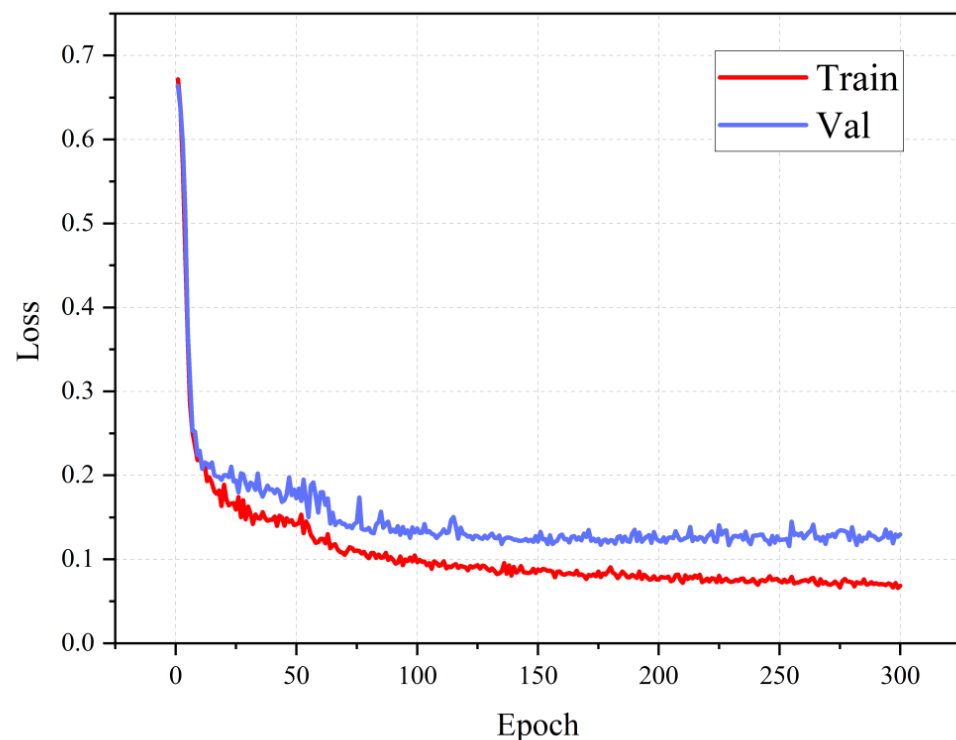


Figure 8. Loss curves of the training set and validation set. Train indicates training set; Val indicates validation set.

As can be seen from Figure 8, the loss function value gradually decreases with the increase in training times after the training starts, and the loss function values of the training set and the validation set decrease to 0.076 and 0.122, respectively, by the 200th training, and they gradually stabilize, with the lowest loss function value of 0.066 in the training set at the 299th training and the lowest loss function value of 0.115 in the validation set at the 254th training. Therefore, with the increase in training times, the loss values of both the training set and test set decrease, indicating that the model attains better loss function convergence, which will help to improve the wheat recognition accuracy.

3.3. Model Recognition Accuracy

In order to verify the effectiveness of the improved lightweight model for winter wheat planting site recognition, the improved lightweight DeepLabv3+ model built on the MobileNetv2 backbone network was compared with four classical semantic segmentation models, UNet, ResUNet, PSPNet, and the original Deeplabv3+ model, under the condition that other training parameters were guaranteed to be the same. The improved lightweight DeepLabv3+ model has the highest recognition accuracy among the five models, and the results are shown in Table 2.

Table 2. Comparison of winter wheat recognition results of different models.

Models	IoU	Winter Wheat Recall	Precision	mIoU	mPA	OA
UNet	80.78%	88.94%	89.80%	85.58%	92.05%	93.15%
ResUNet	85.37%	91.86%	92.36%	89.06%	94.11%	94.91%
PSPNet	80.67%	88.24%	90.38%	85.54%	91.87%	93.16%
DeepLabv3+	82.43%	90.44%	90.30%	86.80%	92.89%	93.76%
Improved DeepLabv3+	86.30%	91.93%	93.37%	89.79%	94.40%	95.28%

Table 2 shows the accuracy results of the improved lightweight DeepLabv3+ model and other classical semantic segmentation models for winter wheat planting area recognition. As can be seen from Table 2, the mPA and OA of all five models are above 90%, and all five models achieved good results for winter wheat recognition extraction, indicating that deep learning has a very good performance for winter wheat planting area recognition extraction. Through the six evaluation indexes selected, the recognition accuracy of the improved lightweight DeepLabv3+ model for winter wheat was higher than the other four models. The overall accuracy, mPA, and mIoU of the improved lightweight DeepLabv3+ model reached 95.28%, 94.40%, and 89.79%, respectively, which are 2.13%, 2.35%, and 4.21% better than the UNet model, 0.37%, 0.29%, and 0.73% better than the ResUNet model, 2.12%, 2.53%, and 4.25% better than the PSPNet model, and 1.52%, 1.51%, and 2.99% better than the original DeepLabv3+ model, respectively. The improved lightweight DeepLabv3+ model achieved IoU, Recall, and Precision of 86.30%, 91.93%, and 93.37%, respectively, for winter wheat recognition results, which were 5.52%, 2.99%, and 3.57% better than the UNet model, 0.93%, 0.07%, and 1.01% better than the ResUNet model, 5.63%, 3.69%, and 2.99% better than the PSPNet model, and 3.9%, 1.49%, and 3.07% better than the original DeepLabv3+ model, respectively. This shows that the improved lightweight DeepLabv3+ model used in this study has better results for winter wheat recognition extraction.

Figure 9a,b shows the prediction results in complex scenarios and in small fields; Figure 9c shows the prediction results under the influence of other green vegetation when the interclass variability is small; and Figure 9d shows the prediction results when the intraclass variability is large in winter wheat.

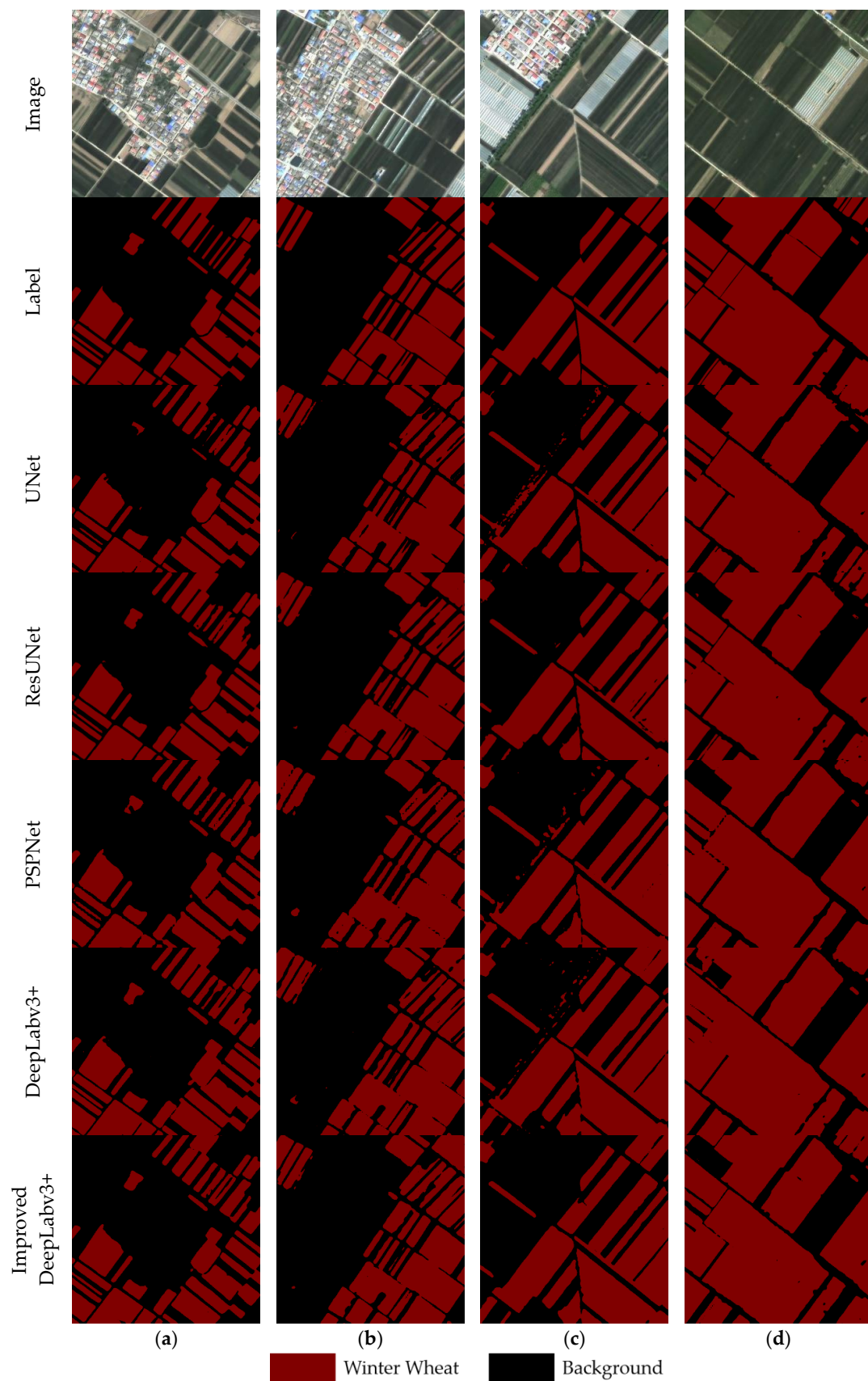


Figure 9. The results of winter wheat identification with different models ((a,b) for complex scenarios and small fields, (c) for less interclass variation, and (d) for large intraclass variation).

It can be seen from Figure 9 that all five network models successfully identified and extracted large areas of winter wheat plantations, but the improved lightweight DeepLabv3+ model identified winter wheat better than the other models. The recognition results of UNet, PSPNet, and the DeepLabv3+ original model are more messy, there are more obvious problems of misclassification and omission, and the prediction result images have fragmented misclassification results in some places, which will predict the features with similar spectral information to winter wheat in the images as winter wheat. The recognized winter wheat planting land has relatively rounded edges, and some large areas of winter wheat planting land are connected together. The missing edge information of winter wheat planting sites is more serious, and these models cannot handle winter wheat planting sites with different areas and spatial dispersion effectively. ResUNet predicted poorly for some of the larger intraclass variability in winter wheat. The overall recognition effect of the improved lightweight DeepLabv3+ model is better. By replacing the backbone network for feature extraction and adding the CBAM, the model pays more attention to the key feature information. The noise phenomenon in the recognition results is significantly reduced, the integrity of the plots is enhanced, the overall recognition effect is improved, the edge information of the winter wheat plantation is extracted more accurately, and the phenomenon of misclassification and omission of small winter wheat plots is alleviated.

4. Discussion

4.1. Model Evaluation

Deep learning semantic segmentation models have a definite advantage over traditional classification methods in dealing with the high-resolution image classification of complex scenes [49]. This study optimized and improved the DeepLabv3+ semantic segmentation model for the problems of many network parameters, long training time, and poor convergence [51]. The backbone network in the DeepLabv3+ semantic segmentation model is replaced with a lightweight MobileNetv2 network, and the CBAM attention mechanism module is introduced to synthesize the training accuracy and efficiency of the model. A lightweight winter wheat classification extraction model based on GF-2 remote sensing images is established to provide reference for the deep learning planting information extraction of high-resolution remote sensing images. The results show that the improved lightweight model in this study outperforms the original model and other classical semantic segmentation models for winter wheat recognition (Table 2, Figure 9). This is due to the fact that winter wheat mostly presents a faceted structure on remote sensing images with simple semantic information, and with a small number of datasets, the deeper and more complex model will produce overfitting, resulting in a decrease in recognition accuracy. The improved lightweight model can focus on the local information of the recognition target, and the attention mechanism enables the network to pay more attention to the pixels of the finely fractured winter wheat plantation, thus filtering out other interfering information. These improvements improved the recognition efficiency and accuracy of the model, which makes the recognition of winter wheat plantations more complete. At the same time, there are some shortcomings in this paper, and the results of extracting complex finely fractured plots are not satisfactory (Figure 10). The presence of other interfering elements in complex finely fractured plots brings greater difficulties to winter wheat recognition.

To further verify the generalization ability of the improved model, the improved model was tested in Linfen City, Shanxi Province, China. The test results are shown in Figure 11. The accuracy of the improved model for winter wheat extraction in Linfen City is comparable to that of the study area, indicating that the model has strong generalization ability and is worth further promoting its use.

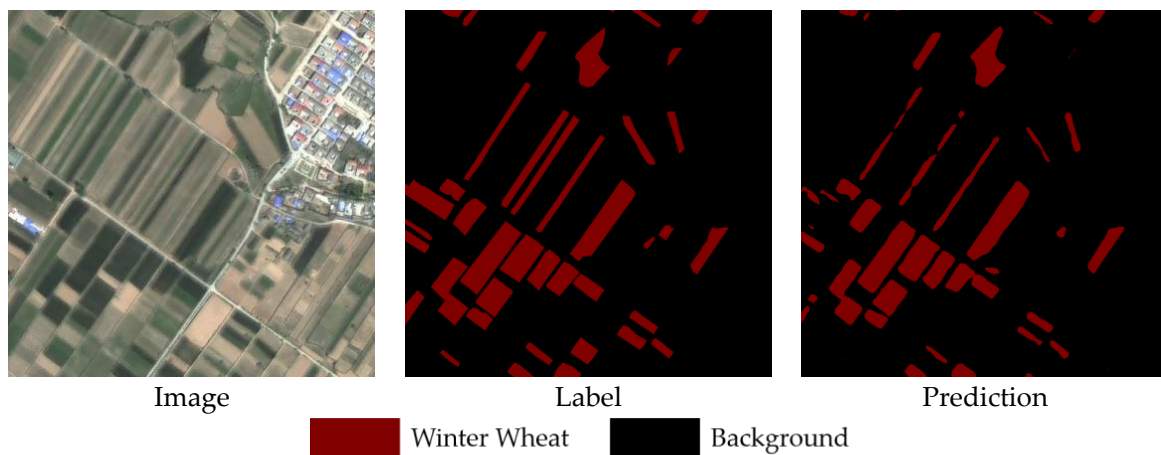


Figure 10. The recognition results of the model for finely fractured plots.

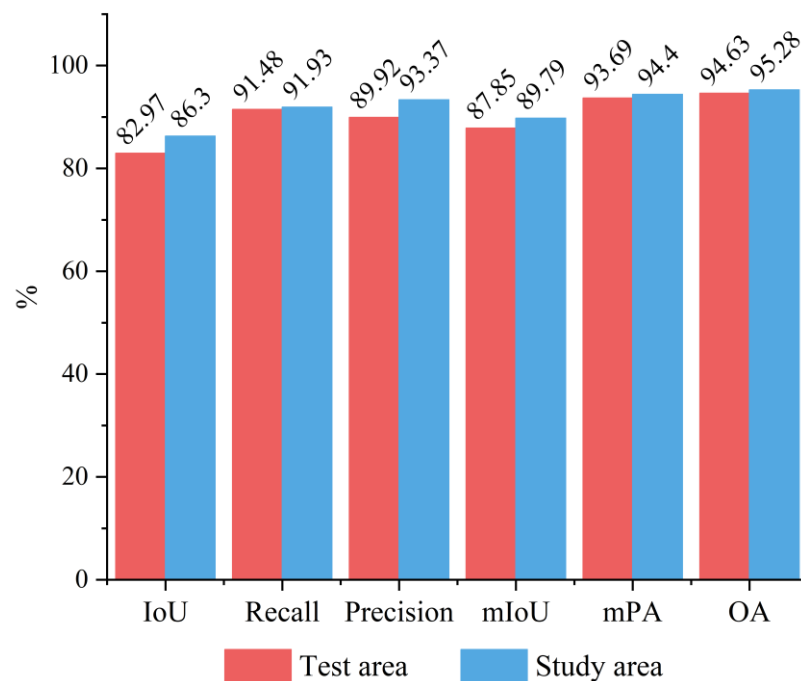


Figure 11. Test results of the model in other regions.

4.2. Model Comparison

From the recognition results in Figure 9, it can be seen that the UNet and PSPNet models have relatively poor extraction results for winter wheat recognition, with many misclassifications and voids. The original DeepLabv3+ semantic segmentation model uses dilated convolution and a multi-scale strategy, which obviously increases the receptive field and can recognize most of the winter wheat plantations, but it can only recognize the winter wheat plantations with small reflectance differences in the image. When the reflectance differences within the plantation increase, the phenomenon of holes will appear, and there are many errors in predicting the edges of small roads and planting areas in the plot. Other crops with spectral characteristics similar to winter wheat planting areas will be predicted as scattered winter wheat planting areas. The improved lightweight DeepLabv3+ model outperforms the original DeepLabv3+ model and obtains better recognition results. Winter wheat mostly presents a faceted structure on remote sensing images with simple semantic information, and with a small number of datasets, the deeper and more complex DeepLabv3+ model will produce overfitting, resulting in a decrease in recognition accuracy.

Therefore, a simple backbone network is needed to extract winter wheat features from a small number of datasets. By replacing the lightweight backbone network and introducing the CBAM, the model focuses more on important information and the pixels of the fine winter wheat plantation so as to filter out the other interfering information and improve the recognition efficiency and accuracy of the model. This greatly reduces the occurrence of holes and minor road prediction errors in the prediction results. And it uses the detailed information of the image and the spatial correlation of pixels in a larger area to more accurately recognize results in most areas, improve the recognition of fine-grained winter wheat planting sites with different area sizes, and improve the effect of handling winter wheat planting sites with large differences in spectral information features, as shown in Figure 9. From the experimental results, it can be seen that there are still some shortcomings in the improved model for winter wheat planting land extraction. In a follow-up study, we will continue to adjust the combination of the hole rate and backbone network parameters in ASPP so that the model can achieve a better extraction effect on the fine-grained winter wheat plantation. Furthermore, we consider the idea of boundary loss and propose new methods that can improve the accuracy of boundary extraction. We consider adding other structures and mechanisms to the model so as to strengthen the robustness of the network and the extraction effect and more effectively improve the extraction of blurring and hole phenomena.

The final results of the parameter sizes and training times for the five models are shown in Table 3.

Table 3. The parameter size and training time of different models.

Models	Model Parameters (MB)	Training Time (h)
UNet	94.97	3.39
ResUNet	175.72	2.56
PSPNet	178.51	5.16
DeepLabv3+	209.70	3.45
Improved DeepLabv3+	22.47	2.05

From Table 3, it can be seen that the improved DeepLabv3+ model is the one with the least parameters among the five models. The model parameter size is 22.47 MB, which is much smaller than the other four models, and the parameter size is 10.72% of the original DeepLabv3+ model, 12.59% of PSPNet, 12.79% of ResUNet, and 23.66% of UNet. The training time of the improved DeepLabv3+ model is 2.05 h, which is the shortest training time among the five models: it is 3.11 h less than PSPNet, 1.4 h less than the original DeepLabv3+ model, 1.34 h less than UNet, and 0.51 h less than ResUNet. Therefore, the improved lightweight model based on the DeepLabv3+ model in this study improves the image segmentation speed and reduces the model parameters while ensuring no reduction in accuracy and aiming to achieve a fast and accurate extraction of winter wheat.

4.3. Future Work

As deep learning techniques are data-driven techniques, for the GF-2 satellite images relied on in this study, the image features acquired within a certain period of time have high similarity and similar feature distributions, which can be obtained using deep neural network learning, and therefore, the method in this study has good adaptive performance. As for the satellite remote sensing images on different sources, it is necessary to add this type of data to the training set and fine-tune the network training parameters. The improved winter wheat recognition model in this study has room for further improvement: for example, further improving the training accuracy and using additional multi-source remote sensing images for migration learning to enhance generalization ability. The next step of this study will be to use multispectral remote sensing information to further solve the problem of winter wheat recognition and classification in mixed cropping areas, as well as subsequent analysis and processing, and to integrate it into actual agricultural production.

With the increasing abundance of high-resolution remote sensing images and the increasing information demand of modern agricultural cultivation, it is necessary to strengthen the ground survey in the future in order to obtain typical features and crop interpretation marks. Establishing localized crop sample datasets can help solve the problem of restricted sample extraction for deep learning crop classification and promote research on the application of deep learning in the remote sensing monitoring of agricultural crops. At the same time, using the rich geometric structure and texture features of high-spatial-resolution remote sensing images, the deep learning method, which is good at feature learning, will be used for farmland plot extraction. Then, taking the plot as the basic unit, multi-source high-resolution remote sensing data will be applied to further realize the precise classification extraction of major crops oriented to the plot scale. Ultimately, it will provide information services for the accurate census of crop planting information, precise management of agricultural production and crop planting structure adjustment, etc.

5. Conclusions

In this study, a lightweight winter wheat extraction model is proposed based on the DeepLabv3+ semantic segmentation model and CBAM attention mechanism module. The model uses the lightweight MobileNetv2 as the backbone network and introduces the lightweight CBAM module, which effectively solves the problems of inaccurate extraction of winter wheat edges from high-resolution images, misclassification and omission due to intraclass differences, as well as the large number of network parameters and long training time of the existing classical semantic segmentation models. The OA, mPA, and mIoU of this model for winter wheat extraction results in the study area reached 95.28, 94.40, and 89.79, respectively, which are better than the other four comparison models, and the model parameters and training time are also the least. Meanwhile, the stronger generalization ability of the model was also verified in other regions. In conclusion, the model guarantees the extraction accuracy while reducing the number of parameters and training time, and it has strong generalization ability, which is worthy of further promotion and use.

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

Funding: This research was funded by the National Key Research and Development Program of China (2022YFF1303404), the Key Science and Technology Project of Inner Mongolia (2021ZD0011) and the Key Science and Technology Project of Inner Mongolia (2021ZD0015).

Data Availability Statement: Not applicable.

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

References

1. Li, S.; Gong, Q.; Yang, S. A Sustainable, Regional Agricultural Development Measurement System Based on Dissipative Structure Theory and the Entropy Weight Method: A Case Study in Chengdu, China. *Sustainability* **2019**, *11*, 5313. [\[CrossRef\]](#)
2. Huang, J.; Yang, G. Understanding Recent Challenges and New Food Policy in China. *Glob. Food Secur.* **2017**, *12*, 119–126. [\[CrossRef\]](#)
3. Liu, Y.; Zhou, Y. Reflections on China's Food Security and Land Use Policy under Rapid Urbanization. *Land Use Policy* **2021**, *109*, 105699. [\[CrossRef\]](#)
4. Norse, D.; Ju, X. Environmental Costs of China's Food Security. *Agric. Ecosyst. Environ.* **2015**, *209*, 5–14. [\[CrossRef\]](#)
5. Zhang, H.; Zhang, J.; Song, J. Analysis of the Threshold Effect of Agricultural Industrial Agglomeration and Industrial Structure Upgrading on Sustainable Agricultural Development in China. *J. Clean. Prod.* **2022**, *341*, 130818. [\[CrossRef\]](#)
6. Deng, Y.; Zeng, F. Sustainable Path of Food Security in China under the Background of Green Agricultural Development. *Sustainability* **2023**, *15*, 2538. [\[CrossRef\]](#)

7. Zhao, L.; Wang, C.; Wang, T.; Liu, J.; Qiao, Q.; Yang, Y.; Hu, P.; Zhang, L.; Zhao, S.; Chen, D.; et al. Identification of the Candidate Gene Controlling Tiller Angle in Common Wheat through Genome-Wide Association Study and Linkage Analysis. *Crop J.* **2023**, *11*, 870–877. [[CrossRef](#)]
8. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating Satellite and Climate Data to Predict Wheat Yield in Australia Using Machine Learning Approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
9. Yang, B.; Zhu, Y.; Zhou, S. Accurate Wheat Lodging Extraction from Multi-Channel UAV Images Using a Lightweight Network Model. *Sensors* **2021**, *21*, 6826. [[CrossRef](#)]
10. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* **2020**, *12*, 236. [[CrossRef](#)]
11. Zhang, L.; Wang, F.; Song, H.; Zhang, T.; Wang, D.; Xia, H.; Zhai, S.; Liu, Y.; Wang, T.; Wang, Y.; et al. Effects of Projected Climate Change on Winter Wheat Yield in Henan, China. *J. Clean. Prod.* **2022**, *379*, 134734. [[CrossRef](#)]
12. Huang, Y.; Wang, F.; Su, Y.; Yu, M.; Shen, A.; He, X.; Gao, J. Risk Assessment of Waterlogging in Major Winter Wheat-Producing Areas in China in the Last 20 Years. *Sustainability* **2022**, *14*, 14072. [[CrossRef](#)]
13. Sun, J.-S.; Zhou, G.-S.; Sui, X.-H. Climatic Suitability of the Distribution of the Winter Wheat Cultivation Zone in China. *Eur. J. Agron.* **2012**, *43*, 77–86. [[CrossRef](#)]
14. Qi, X.; Wang, Y.; Peng, J.; Zhang, L.; Yuan, W.; Qi, X. The 10-Meter Winter Wheat Mapping in Shandong Province Using Sentinel-2 Data and Coarse Resolution Maps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9760–9774. [[CrossRef](#)]
15. Liu, J.; Wang, L.; Yang, F.; Yao, B.; Yang, L. National-scale Mapping of Winter Wheat in China Using GF-1 Imagery. *Chin. Agric. Sci. Bull.* **2019**, *35*, 155–164. [[CrossRef](#)]
16. Dong, Q.; Chen, X.; Chen, J.; Zhang, C.; Liu, L.; Cao, X.; Zang, Y.; Zhu, X.; Cui, X. Mapping Winter Wheat in North China Using Sentinel 2A/B Data: A Method Based on Phenology-Time Weighted Dynamic Time Warping. *Remote Sens.* **2020**, *12*, 1274. [[CrossRef](#)]
17. Ren, S.; Guo, B.; Wu, X.; Zhang, L.; Ji, M.; Wang, J. Winter Wheat Planted Area Monitoring and Yield Modeling Using MODIS Data in the Huang-Huai-Hai Plain, China. *Comput. Electron. Agric.* **2021**, *182*, 106049. [[CrossRef](#)]
18. Cai, Y.; Guan, K.; Peng, J.; Wang, S.; Seifert, C.; Wardlow, B.; Li, Z. A High-Performance and in-Season Classification System of Field-Level Crop Types Using Time-Series Landsat Data and a Machine Learning Approach. *Remote Sens. Environ.* **2018**, *210*, 35–47. [[CrossRef](#)]
19. Chen, Y.; Zhang, C.; Wang, S.; Li, J.; Li, F.; Yang, X.; Wang, Y.; Yin, L. Extracting Crop Spatial Distribution from Gaofen 2 Imagery Using a Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 2917. [[CrossRef](#)]
20. Dong, J.; Fu, Y.; Wang, J.; Tian, H.; Fu, S.; Niu, Z.; Han, W.; Zheng, Y.; Huang, J.; Yuan, W. Early-Season Mapping of Winter Wheat in China Based on Landsat and Sentinel Images. *Earth Syst. Sci. Data* **2020**, *12*, 3081–3095. [[CrossRef](#)]
21. Wu, X.; Xiao, X.; Steiner, J.; Yang, Z.; Qin, Y.; Wang, J. Spatiotemporal Changes of Winter Wheat Planted and Harvested Areas, Photosynthesis and Grain Production in the Contiguous United States from 2008–2018. *Remote Sens.* **2021**, *13*, 1735. [[CrossRef](#)]
22. Wang, S.; Azzari, G.; Lobell, D.B. Crop Type Mapping without Field-Level Labels: Random Forest Transfer and Unsupervised Clustering Techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [[CrossRef](#)]
23. Ashourloo, D.; Nematollahi, H.; Huete, A.; Aghighi, H.; Azadbakht, M.; Shahrabi, H.S.; Goodarzashti, S. A New Phenology-Based Method for Mapping Wheat and Barley Using Time-Series of Sentinel-2 Images. *Remote Sens. Environ.* **2022**, *280*, 113206. [[CrossRef](#)]
24. Li, W.; Zhang, H.; Li, W.; Ma, T. Extraction of Winter Wheat Planting Area Based on Multi-Scale Fusion. *Remote Sens.* **2023**, *15*, 164. [[CrossRef](#)]
25. Blaes, X.; Vanhalle, L.; Defourny, P. Efficiency of Crop Identification Based on Optical and SAR Image Time Series. *Remote Sens. Environ.* **2005**, *96*, 352–365. [[CrossRef](#)]
26. Siachalou, S.; Mallinis, G.; Tsakiri-Strati, M. A Hidden Markov Models Approach for Crop Classification: Linking Crop Phenology to Time Series of Multi-Sensor Remote Sensing Data. *Remote Sens.* **2015**, *7*, 3633–3650. [[CrossRef](#)]
27. Seifi Majdar, R.; Ghassemian, H. A Probabilistic SVM Approach for Hyperspectral Image Classification Using Spectral and Texture Features. *Int. J. Remote Sens.* **2017**, *38*, 4265–4284. [[CrossRef](#)]
28. Du, Z.; Yang, J.; Ou, C.; Zhang, T. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* **2019**, *11*, 888. [[CrossRef](#)]
29. Zhao, J.; Zhong, Y.; Hu, X.; Wei, L.; Zhang, L. A Robust Spectral-Spatial Approach to Identifying Heterogeneous Crops Using Remote Sensing Imagery with High Spectral and Spatial Resolutions. *Remote Sens. Environ.* **2020**, *239*, 111605. [[CrossRef](#)]
30. Turker, M.; Ozdarici, A. Field-Based Crop Classification Using SPOT4, SPOT5, IKONOS and QuickBird Imagery for Agricultural Areas: A Comparison Study. *Int. J. Remote Sens.* **2011**, *32*, 9735–9768. [[CrossRef](#)]
31. Vogels, M.F.A.; de Jong, S.M.; Sterk, G.; Addink, E.A. Mapping Irrigated Agriculture in Complex Landscapes Using SPOT6 Imagery and Object-Based Image Analysis—A Case Study in the Central Rift Valley, Ethiopia-. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *75*, 118–129. [[CrossRef](#)]
32. Zhou, K.; Zhang, Z.; Liu, L.; Miao, R.; Yang, Y.; Ren, T.; Yue, M. Research on UNet Winter Wheat Identification Method Based on GF-2. *Remote Sens.* **2023**, *15*, 3094. [[CrossRef](#)]

33. Liu, J.; Wang, H.; Zhang, Y.; Zhao, X.; Qu, T.; Tian, H.; Lu, Y.; Su, J.; Luo, D.; Yang, Y. A Spatial Distribution Extraction Method for Winter Wheat Based on Improved U-Net. *Remote Sens.* **2023**, *15*, 3711. [[CrossRef](#)]
34. Song, D.; Zhang, C.; Yang, X.; Li, F.; Han, Y.; Gao, S.; Dong, H. Extracting Winter Wheat Spatial Distribution Information from GF-2 Image. *Natl. Remote Sens. Bull.* **2020**, *24*, 596–608. [[CrossRef](#)]
35. Liu, D.; Han, L.; Han, X. High Spatial Resolution Remote Sensing Image Classification Based on Deep Learning. *Acta Opt. Sin.* **2016**, *36*, 0428001. [[CrossRef](#)]
36. Debats, S.R.; Luo, D.; Estes, L.D.; Fuchs, T.J.; Caylor, K.K. A Generalized Computer Vision Approach to Mapping Crop Fields in Heterogeneous Agricultural Landscapes. *Remote Sens. Environ.* **2016**, *179*, 210–221. [[CrossRef](#)]
37. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-Cover Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models. *Remote Sens. Environ.* **2020**, *237*, 111322. [[CrossRef](#)]
38. Li, D.; Zhang, L.; Xia, G. Automatic Analysis and Mining of Remote Sensing Big Data. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 1211–1216. [[CrossRef](#)]
39. Scott, G.J.; England, M.R.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [[CrossRef](#)]
40. Zhang, B.; Wang, C.; Shen, Y.; Liu, Y. Fully Connected Conditional Random Fields for High-Resolution Remote Sensing Land Use/Land Cover Classification with Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 1889. [[CrossRef](#)]
41. Lv, X.; Ming, D.; Chen, Y.; Wang, M. Very High Resolution Remote Sensing Image Classification with SEEDS-CNN and Scale Effect Analysis for Superpixel CNN Classification. *Int. J. Remote Sens.* **2019**, *40*, 506–531. [[CrossRef](#)]
42. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [[CrossRef](#)]
43. Zhong, L.; Hu, L.; Zhou, H. Deep Learning Based Multi-Temporal Crop Classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
44. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]
45. Zhong, L.; Hu, L.; Zhou, H.; Tao, X. Deep Learning Based Winter Wheat Mapping Using Statistical Data as Ground References in Kansas and Northern Texas, US. *Remote Sens. Environ.* **2019**, *233*, 111411. [[CrossRef](#)]
46. Wang, H.; Chang, W.; Yao, Y.; Yao, Z.; Zhao, Y.; Li, S.; Liu, Z.; Zhang, X. Cropformer: A New Generalized Deep Learning Classification Approach for Multi-Scenario Crop Classification. *Front. Plant Sci.* **2023**, *14*, 1130659. [[CrossRef](#)]
47. Chu, X.; Yao, X.; Duan, H.; Chen, C.; Li, J.; Pang, W. Glacier Extraction Based on High-Spatial-Resolution Remote-Sensing Images Using a Deep-Learning Approach with Attention Mechanism. *Cryosphere* **2022**, *16*, 4273–4289. [[CrossRef](#)]
48. Sun, H.; Wang, B.; Wu, Y.; Yang, H. Deep Learning Method Based on Spectral Characteristic Rein-Forcement for the Extraction of Winter Wheat Planting Area in Complex Agricultural Landscapes. *Remote Sens.* **2023**, *15*, 1301. [[CrossRef](#)]
49. Zhang, D.; Ding, Y.; Chen, P.; Zhang, X.; Pan, Z.; Liang, D. Automatic Extraction of Wheat Lodging Area Based on Transfer Learning Method and Deeplabv3+ Network. *Comput. Electron. Agric.* **2020**, *179*, 105845. [[CrossRef](#)]
50. Huang, L.; Wu, X.; Peng, Q.; Yu, X. Depth Semantic Segmentation of Tobacco Planting Areas from Unmanned Aerial Vehicle Remote Sensing Images in Plateau Mountains. *J. Spectrosc.* **2021**, *2021*, 6687799. [[CrossRef](#)]
51. Mo, L.; Fan, Y.; Wang, G.; Yi, X.; Wu, X.; Wu, P. DeepMDSCBA: An Improved Semantic Segmentation Model Based on DeepLabV3+ for Apple Images. *Foods* **2022**, *11*, 3999. [[CrossRef](#)] [[PubMed](#)]
52. Zhang, C.; Luo, S.; Zhao, W.; Wang, Y.; Zhang, Q.; Qu, C.; Liu, X.; Wen, X. Impacts of Meteorological Factors, VOCs Emissions and Inter-Regional Transport on Summer Ozone Pollution in Yuncheng. *Atmosphere* **2021**, *12*, 1661. [[CrossRef](#)]
53. He, P.; Wang, J.; Cao, C.; Xu, L.; Liu, Z.; Bi, R. Yield estimation of summer maize in yuncheng basin based on fusion of multi-source remote sensing data. *Chin. J. Agric. Resour. Reg. Plan.* **2023**, *44*, 213–221.
54. Kuang, X.; Guo, J.; Bai, J.; Geng, H.; Wang, H. Crop-Planting Area Prediction from Multi-Source Gaofen Satellite Images Using a Novel Deep Learning Model: A Case Study of Yangling District. *Remote Sens.* **2023**, *15*, 3792. [[CrossRef](#)]
55. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [[CrossRef](#)]
56. Liu, C.; Feng, Q.; Liu, J.; Wang, Y.; Shi, T.; Li, Y.; Gong, J.; Zhao, H. Urban Green Plastic Cover Extraction and Spatial Pattern Changes in Jinan City Based on DeepLabv3+ Semantic Segmentation Model. *Natl. Remote Sens. Bull.* **2022**, *26*, 2518–2530. [[CrossRef](#)]
57. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [[CrossRef](#)]
58. Liu, J.; Zhang, Y.; Liu, C.; Liu, X. Monitoring Impervious Surface Area Dynamics in Urban Areas Using Sentinel-2 Data and Improved Deeplabv3+ Model: A Case Study of Jinan City, China. *Remote Sens.* **2023**, *15*, 1976. [[CrossRef](#)]
59. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. [[CrossRef](#)]
60. Li, W.; Liu, K. Confidence-Aware Object Detection Based on MobileNetv2 for Autonomous Driving. *Sensors* **2021**, *21*, 2380. [[CrossRef](#)]

61. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[CrossRef](#)]
62. Ma, R.; Wang, J.; Zhao, W.; Guo, H.; Dai, D.; Yun, Y.; Li, L.; Hao, F.; Bai, J.; Ma, D. Identification of Maize Seed Varieties Using MobileNetV2 with Improved Attention Mechanism CBAM. *Agriculture* **2023**, *13*, 11. [[CrossRef](#)]
63. Chen, L.; Tan, S.; Pan, Z.; Xing, J.; Yuan, Z.; Xing, X.; Zhang, P. A New Framework for Automatic Airports Extraction from SAR Images Using Multi-Level Dual Attention Mechanism. *Remote Sens.* **2020**, *12*, 560. [[CrossRef](#)]
64. Tang, Z.; Sun, Y.; Wan, G.; Zhang, K.; Shi, H.; Zhao, Y.; Chen, S.; Zhang, X. Winter Wheat Lodging Area Extraction Using Deep Learning with GaoFen-2 Satellite Imagery. *Remote Sens.* **2022**, *14*, 4887. [[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.