

Article

Identification of Soybean Planting Areas Combining Fused Gaofen-1 Image Data and U-Net Model

Sijia Zhang ¹, Xuyang Ban ², Tian Xiao ², Linsheng Huang ², Jinling Zhao ^{2,*} , Wenjiang Huang ³ 
and Dong Liang ^{2,*}

¹ School of Internet, Anhui University, Hefei 230039, China

² National Engineering Research Center for Analysis and Application of Agro-Ecological Big Data, Anhui University, Hefei 230601, China

³ Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

* Correspondence: zhaojl@ahu.edu.cn (J.Z.); dliang@ahu.edu.cn (D.L.)

Abstract: It is of great significance to accurately identify soybean planting areas for ensuring agricultural and industrial production. High-resolution satellite remotely sensed imagery has greatly facilitated the effective extraction of soybean planting areas but novel methods are required to further improve the identification accuracy. Two typical planting areas of Linhu Town and Baili Town in Northern Anhui Province, China, were selected to explore the accurate extraction method. The 10 m multispectral and 2 m panchromatic Gaofen-1 (GF-1) image data were first fused to produce training, test, and validation data sets after the min–max standardization and data augmentation. The deep learning U-Net model was then adopted to perform the accurate extraction of soybean planting areas. Two vital influencing factors on the accuracies of the U-Net model, including cropping size and training epoch, were compared and discussed. Specifically, three cropping sizes of 128×128 , 256×256 , and 512×512 px, and 20, 40, 60, 80, and 100 training epochs were compared to optimally determine the values of the two parameters. To verify the extraction effect of the U-Net model, comparison experiments were also conducted based on the SegNet and DeepLabv3+. The results show that U-Net achieves the highest *Accuracy* of 92.31% with a Mean Intersection over Union (*mIoU*) of 81.35%, which is higher than SegNet with an improvement of nearly 4% in *Accuracy* and 10% on *mIoU*. In addition, the *mIoU* has been also improved by 8.89% compared with DeepLabv3+. This study provides an effective and easily operated approach to accurately derive soybean planting areas from satellite images.

Keywords: remote sensing; data augmentation; parameter optimization; planting area extraction; deep learning



Citation: Zhang, S.; Ban, X.; Xiao, T.; Huang, L.; Zhao, J.; Huang, W.; Liang, D. Identification of Soybean Planting Areas Combining Fused Gaofen-1 Image Data and U-Net Model. *Agronomy* **2023**, *13*, 863. <https://doi.org/10.3390/agronomy13030863>

Academic Editor: Camilla Dibari

Received: 17 January 2023

Revised: 7 March 2023

Accepted: 13 March 2023

Published: 15 March 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

Soybean (*Glycine max* (L.) Merr.) is one of the most important oil-bearing crops around the world. It is also one of China's major food crops, which can be used to provide valuable oil and protein constituents for both humans and livestock. The largest production areas are in China's three northeastern provinces [1]. To make a decision on the cultivation and trade of soybean, it is highly important to figure out the planting areas and spatial distribution. Traditional methods mainly rely on manual measurement and statistical sampling to achieve statistical data, which are time-consuming, susceptible to subjective judgment, labor-intensive, etc. The advancement of earth-observing techniques has greatly improved the monitoring and extraction of crop planting and growth information, especially at a large spatial scale. Remote sensing (RS) technology can provide spatial, spectral, and temporal information of soybean, with macroscopic and dynamic characteristics. When RS technology is applied to the monitoring of soybean, the specific properties and features can be derived from various sensors.

With the development of satellite RS technology, RS images have gradually become a main data source for extracting crop planting information. RS technology has been widely used with soybean crops including estimating the planting areas [2], yield estimation [3], growth monitoring [4], detection and classification of diseases and insect pests [5], etc. It is obvious that a precise understanding of soybean planting areas and their geographical distribution is the prerequisite for various applications. In most previous studies, single-source RS images have been adopted to extract soybean planting areas, mainly including the MODerate Resolution Imaging Spectroradiometer (MODIS) series of satellites of Gaofen, Landsat, and Sentinel. Chang et al. [6] applied a 500-meter time-sequential composite MODIS to estimate corn and soybean areas for the dominant production areas of the USA by taking advantage of low spatial and high temporal resolution MODIS data. Huang et al. [7] identified the corn and soybean cropping areas using the random forest (RF) classifier and multi-temporal 16-meter-resolution GF-1 Wide Field of View (WV) imagery. Zhong et al. [8] developed an innovative phenology-based classification method to map corn and soybean via over 100 Landsat TM and ETM+ images. Multiple sets of input variables and RF classifiers were jointly used to achieve accuracies higher than 88%. Zhu et al. [9] integrated multi-temporal Sentinel-1/2 microwave and optical multispectral data to map the spatial distribution of soybean through a stepwise hierarchical extraction strategy. The RF proved to be superior to a Back-Propagation Neural Network (BPNN) and Support Vector Machine (SVM). It can be found that multitemporal features of satellite imagery are mainly used to identify soybean information. The overall accuracies are generally lower than 90%.

In recent years, the rapid development of Unmanned Aerial Vehicles (UAV) has provided higher-resolution remote sensing imagery for soybean monitoring. Randelović et al. [10] adopted the vegetation indices (VIs) derived from three-channel UAV images of Red, Green, and Blue (RGB) bands to predict soybean plant density. In addition to commonly used machine learning algorithms, some Convolutional Neural Networks (CNN)-based methods have been also used in accurately extracting soybean planting areas. CNN consist of three layers, a convolutional layer, a max pooling layer, and a fully connected layer, which can greatly improve the classification accuracy of single or multiple objects by learning deep features. Habibi et al. [11] used the You Only Look Once version 3 (YOLOv3) object detection algorithm to accurately measure actual soybean plant density, showing higher accuracy than the partial least squares and RF methods. Yang et al. [12] collected RGB and multispectral images using a quad-rotor UAV and employed the U-Net model to improve the soybean recognition accuracy. The results show that the accuracy of the U-Net was the best when compared with DeepLabv3+, RF, and SVM. In comparison, with centimeter-level spatial resolution UAV imagery, the spatial resolution has been improved from a hundred-meter to sub-meter resolution for spaceborne satellites. The improvement in spatial resolution has facilitated the use of deep learning algorithms to improve the extraction accuracies of crop planting areas [13].

There are several influencing factors on the extraction of soybean planting areas, such as RS images, classifiers, planting structure, and the area of the study space, etc. Soybean and corn are the two crops that are generally misclassified during their growing seasons. Both crops show similar spectral and textural features at the initial growth stages but the indicative features can be explored in their middle and late growth seasons [14]. It is highly necessary to find the discriminative features between soybean and corn using remote-sensing technology. The selection of imaging time for RS images is essential for accurately identifying soybean planting areas. In addition to the RS images, classifiers are also factors affecting accuracy. For example, the Simple Non-Iterative Clustering segmentation method and the Continuous Naive Bayes classifier were used to map the soybeans and corn in Paraná state, Brazil, with a minimum global accuracy of 90% [15]. Xu et al. [16] developed a deep learning approach, named DeepCropMapping (DCM), to dynamically map corn and soybean. The DCM model significantly outperformed the Transformer, RF, and Multilayer Perceptron (MLP) methods. To improve the monitoring and classification performance,

multitemporal and multispectral RS data are generally input into the classifiers [17]. It is inevitable that high computing power is required to generate a large number of training samples for obtaining reliable and accurate performance. To increase work efficiency in practice, it is significant to map the soybean in a relatively short time. In this study, a 2-meter-resolution fused GF-1 image with the appropriate imaging time was used to identify soybean planting areas via the CNN-based U-Net model. The model is composed of a contracting path and an expansive path, in which the U-shaped architecture and skip connections are the outstanding features [18]. It is simple, efficient, easy to understand, and customizable.

Our highlights for this study are: (1) Fusing the 8-meter multispectral and 2-meter panchromatic GF-1 satellite images to assist in the production of high-quality training samples. (2) Using the intelligible and certified U-Net model to identify soybean planting areas. (3) Comparing and discussing the two crucial parameters, image cropping size and training epoch, which greatly affect the accuracy of the U-Net model to find out the optimal values. The main objective of this study was to optimally determine a U-Net model for accurately extracting soybean planting areas, with the best cropping sizes and training epochs based on high-resolution GF-1 fused imagery. An additional objective was to validate the accuracy of the model by comparing the SegNet and DeepLabv3+. The two networks are mainstream CNN architectures for image segmentation, which have been usually adopted as comparison models.

2. Materials and Methods

2.1. Study Area

The study area was located in Guoyang County, spanning from $33^{\circ}27'$ to $33^{\circ}47'$ N and $115^{\circ}53'$ to $116^{\circ}33'$ E, Bozhou City, Northern Anhui Province, China (Figure 1). The soybean planting area reached 71,086.67 ha in 2022. Linhu Town and Biaoli Town were selected as the study areas, which are important soybean planting areas in Guoyang County. The two towns have flat terrain and four distinct seasons, with an annual rainfall of about 800 mm. A warm-temperate semi-humid monsoon climate and sufficient sunlight are beneficial to the growth of crops such as soybean, wheat, maize, etc.

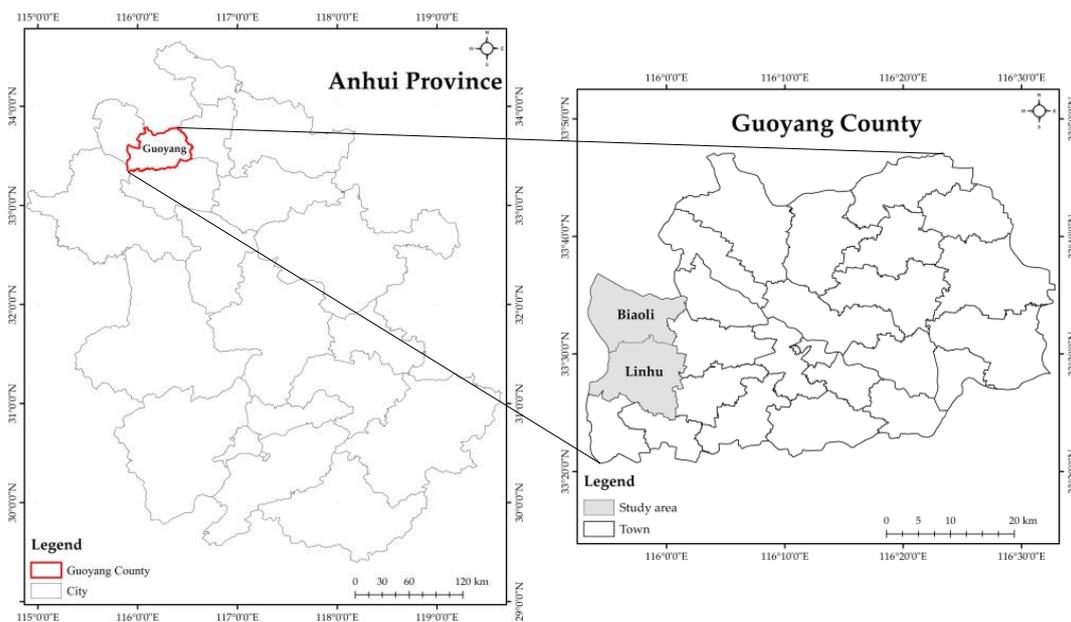


Figure 1. Geographic locations of Biaoli Town and Linhu Town, Guoyang County, Bozhou City, Anhui Province, China.

2.2. Growth Stages of Soybean

Growth stages of soybeans are divided into vegetative growth stages and reproductive growth stages. There are distinctive spectral and spatial characteristics for soybeans at different stages. In addition, other crops (e.g., corn, sorghum, cotton) may also show similar characteristics to soybeans, which causes trouble in accurately identifying soybean planting areas. It is highly important to figure out the phenological stages for selecting appropriate remote-sensing images [4]. When applying remote-sensing technology to the classification of soybeans, it is vital for finding out the optimal imagery. In this study area, soybeans are generally sown in mid-to-late June and harvested in late September or early October (Table 1).

Table 1. Primary soybean growth stages in the study area.

Month	June		July		August			September			October		
Phenological stage	Mid-Sowing	Late	Early	Mid-Third node	Late	Early	Mid-Blooming	Late	Early	Mid-	Late	Early	Mid-
		Seedling emergence			Side branch				Podding		Maturity		

2.3. Data Sources and Preprocessing

The GF-1 optical satellite was launched on 26 April 2013, which is the first satellite of the China High-resolution Earth Observation System (CHEOS). The satellite has broken through key technologies of high spatial resolution, combination of multispectral and wide coverage, etc., which was widely used in various fields (e.g., modern agriculture, disaster prevention and reduction, resource and environment monitoring) [19–21]. According to Table 1 and the optimal temporal selection for identifying soybeans [15], GF-1 satellite images (Table 2) with 2-meter multispectral and 8-meter panchromatic spectral bands were acquired on 18 August 2019 at blooming and podding stages of soybean. The in situ experiments were also simultaneously carried out to collect the ground truth data. More specifically, the samples were randomly selected in a large soybean field and positioned using a sub-meter GeoXH2008 handheld GPS receiver (Trimble, Westminster, CO, USA). The positioned samples were then overlaid on the GF-1 imagery to select training datasets. The radiometric correction and orthorectification were first carried out, and then the NNDiffuse Pan Sharpening algorithm, proposed by the Rochester Institute of Technology (RIT), was used to generate a 2-meter resolution fused image in ENVI (The Environment for Visualizing Images, Exelis Visual Information Solutions, Inc., Broomfield, CO, USA) 5.3 software.

Table 2. Technical parameters of payloads for GF-1 satellite.

Band Number	Band Name	Spectral Range (μm)	Spatial Resolution (m)	Revisit Cycle	Swath (km)
P	Panchromatic	0.45–0.90	2		
B1	Blue	0.45–0.52			60
B2	Green	0.52–0.59		4 days	(Two
B3	Red	0.63–0.69	8		cameras)
B4	NIR	0.77–0.89			

2.4. Dataset Production

(1) Image cropping

For a satellite remote sensing image, the size and data volume are much larger than a picture photographed by a handheld camera. When a complete image is directly input into a classifier, especially for a CNN-based method, the computing capacity is generally incapable of supporting the complex and interactive algorithms for a computer. In addition, the target objects of soybeans are randomly distributed in a remote-sensing image. Consequently, the

image must be cropped into small tiles to accelerate the training and classification. As a kind of data preprocessing, cropping can produce new data by cropping the central pixels of an image. Considering the town-based images in this study, three cropping sizes of 128×128 , 256×256 , and 512×512 px images were obtained and compared (Figure 2). It can be found that the 256×256 px images were the most suitable size as the training, validation, and test data sets.

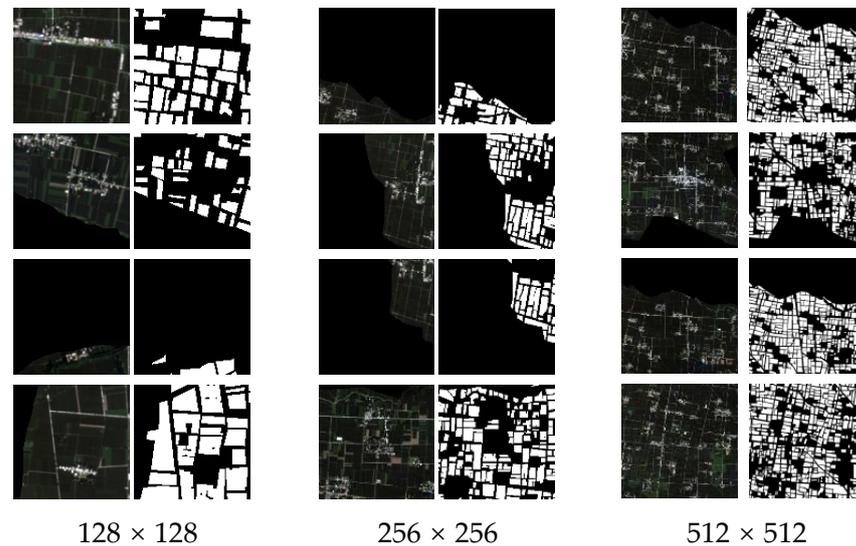


Figure 2. Comparison of three image cropping sizes of 128×128 , 256×256 , and 512×512 px.

(2) Training and test datasets

Two folders were created to deposit training and test datasets, which were named train and test. The original GF-1 images and datasets used in the experiment were also deposited in both folders. For the training folder, the cropped images, two folders of original images named “src” and labeled samples named “label” were placed in the folder. For the test folder, the folder named “prediction” was also created. When producing the datasets, the cropped images were in one-to-one correspondence with corresponding labels. Each image dataset was named using the natural numbers with the format *.png. Similarly, the labeled samples were also named in accordance with the same rules (Figure 3).

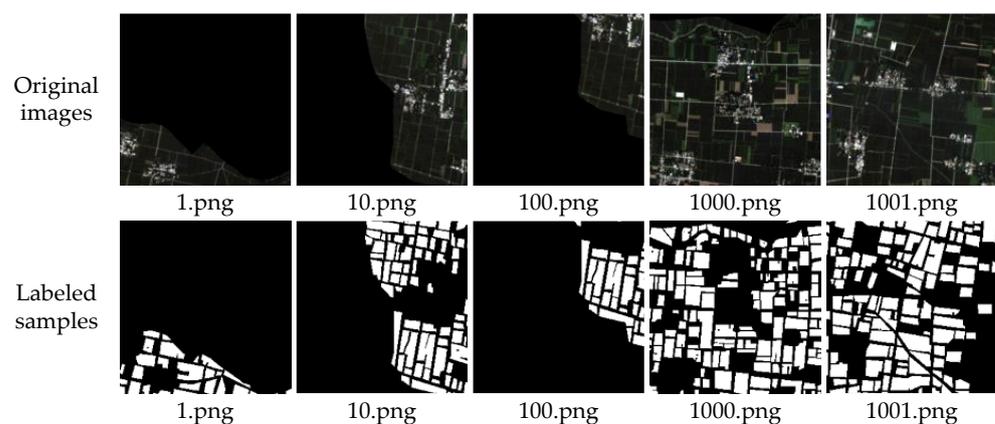


Figure 3. Original images and corresponding labeled samples.

(3) Data normalization

Normalization is an important procedure for optimizing neural networks, which can effectively reduce the influence among different datasets and improve the interactive speed

of training models [22]. In this study, the min–max normalization was adopted to perform the data normalization, which can perform the linear transformations of original data [23]. The pixel values of an image in the range (0, 255) were normalized to (0, 1). The specific formula is shown in Equation (1).

$$x_{\text{new}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

where x_{new} is the normalized value; x is the old value, x_{max} is the maximum value of original pixel; and x_{min} is the minimum value of original pixel.

(4) Data augmentation

Insufficient training datasets will cause several severe problems such as unbalanced samples, over-fitting, or poor generalization ability in neural networks [24]. It is highly necessary to carry out the data augmentation because we manually produce the training samples. In addition, they are random and irregular for the spatial distribution of soybean planting areas. To fully train the U-Net model, various samples are required to be derived from data augmentation. Specifically, the scale transformation, horizontal and vertical flip, rotation, etc., were used to obtain more training datasets.

2.5. U-Net Model

The U-Net model was first proposed by Ronneberger et al. based on the fully convolutional networks (FCN) [25]. More specifically, convolutional layers and pooling layers are used to extract features and transposed convolutional layers are adopted to revert image sizes. The model was originally applied to the semantic segmentation of medical images. Afterward, it was widely used in crop detection and classification [26–28], due to the advantages of the encoder-decoder structure and skipping networks. Previous studies have shown that U-Net model has strong feature extraction abilities and good segmentation effects even if the sample size is small. As shown in Figure 4, the structure is composed of contracting path (encoder) and expansive path (decoder). In the down-sampling process, every two convolutional layers form a convolution block and there are in total five convolution blocks. In each up-sampling process, the convolution feature maps to the two convolution layers are reduced, whose number is from the encoding path. In the process of feature extraction, the size of a remote sensing image will be reduced when passing through a pooling layer every time. For the U-Net network structure, feature extraction and up-sampling are contacted to form a U-shaped structure.

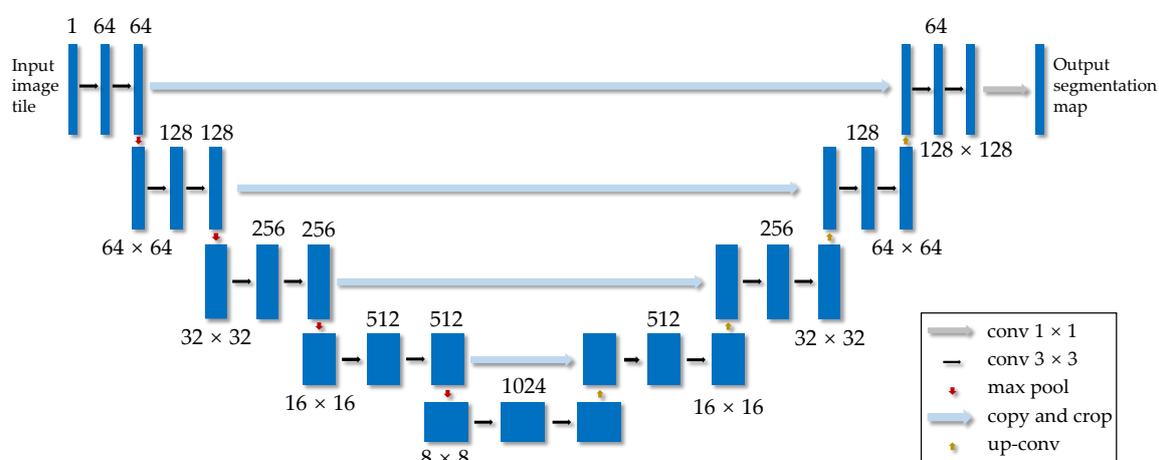


Figure 4. Structure diagram of the U-Net model used in this study.

2.6. Evaluation Metrics

The confusion matrix-based *Accuracy*, *Recall*, *Precision*, *F1-score (F1)*, *Intersection over Union (IoU)* and *Mean Intersection over Union (MIoU)* were selected as the evaluation metrics to assess accuracy (Table 3) [29,30]. *Accuracy* is generally used to evaluate the overall accuracy of the model. *Recall* is used to evaluate the classification effect of test set. *F1* is used to comprehensively evaluate the classification performance. *IoU* is used to calculate the coincidence degree between target and prediction areas. *MIoU* is the evaluation index to assess the prediction results of a network. The closer its value approaches 1, the better the segmentation effect of the network is.

Table 3. Evaluation metrics for assessing accuracy.

Metric	Formula	Variable Explanation
<i>Accuracy</i>	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	<i>TP</i> (true positives) are the number of pixels correctly classified as soybean planting areas; <i>TN</i> (true negatives) represent the number of background pixels that are predicted as background pixels; <i>FP</i> (false positives) represent the number of pixels that are background pixels but misclassified as soybean planting areas; and <i>FN</i> (false negatives) are the number of pixels that are soybean planting areas but misclassified as background pixels.
<i>Recall</i>	$Recall = \frac{TP}{TP+FN}$	
<i>Precision</i>	$Precision = \frac{TP}{TP+FP}$	
<i>F1-score</i>	$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$	
<i>IoU</i>	$IoU = \frac{P_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - P_{ii}}$	<i>k</i> is the number of categories; <i>P_{ii}</i> is the number of correctly identified pixels of category <i>i</i> ; <i>p_{ij}</i> is the number of pixels that are category <i>i</i> but predicted as the category <i>j</i> ; and <i>p_{ji}</i> is the number of pixels that are category <i>j</i> but predicted as the category <i>i</i> .
<i>MIoU</i>	$MIoU = \frac{1}{k} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - P_{ii}}$	

3. Results and Discussion

3.1. Model Training

The curves of loss and accuracy based on the U-Net model are shown in Figure 5. As seen in Figure 5a, the loss value gradually decreases with continuous training. When the training epochs are 20 and 40, there is a little fluctuation but it maintains about 0.0004 when the number of training epochs reaches 50. As shown in Figure 5b, the training accuracy steadily increases with the progression of training. When the epoch reaches 40, the *Accuracy* exceeds 99% and reaches 99.51% as the epoch increases to 60. The *Accuracy* finally reaches 99.69%, indicating that the model has high classification accuracy and achieves a good training performance. Sixty epochs can be a good choice, after comprehensively considering the training accuracy and speed. In addition, during the training process, it is also important to adjust the model parameters. The learning rate and batch size are the two crucial parameters [31]. The learning rate greatly influences the minimum final convergence of the model. In this study, we used Adam to adjust it from the default value and its value was finally set to 0.0001. Batch size affects the model's performance, large values lead to a poor generalization ability but small values affect the model's convergence. It was finally set to four through a debugging process. The epoch of 100 was used to train the model.

We also mapped the training results for epochs 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. It is found that the prediction result is extremely poor for one epoch. Some samples have no corners and many areas are not predicted and some pixels are not soybean areas, showing that the model is not well-trained at the beginning. With the increase in epochs, the prediction result gradually shows a better performance. When it reaches 60, the classification has already shown a good result, in which the training result reaches a high coincidence with the labeled samples.

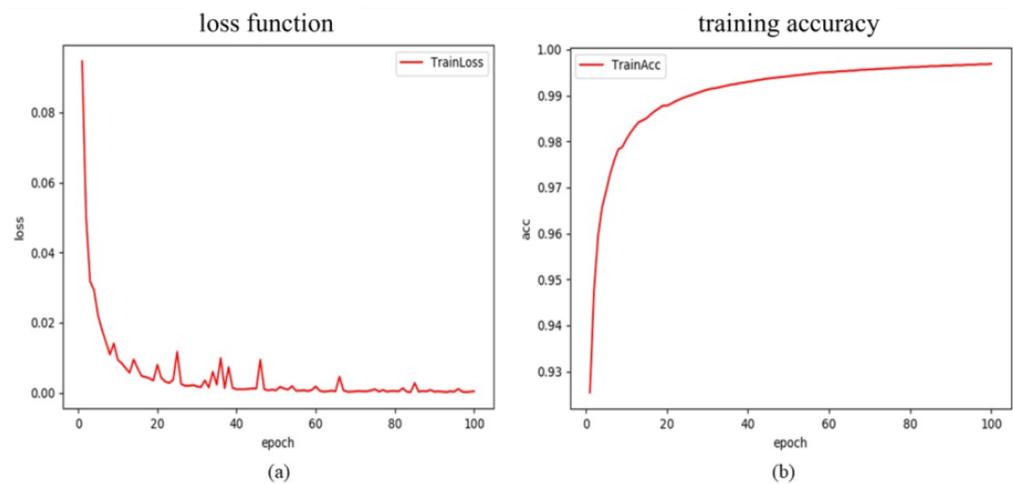


Figure 5. Curves of (a) loss and (b) accuracy.

3.2. Influence of Cropping Size on Prediction Accuracy

In order to explore the optimal cropping size, three sizes of 128×128 , 256×256 , and 512×512 px were adopted to comparatively analyze its influence on accuracy. A total of 143 test images for the 128×128 px, 42 for the 256×256 px, and 12 for the 512×512 px were cropped. Figures 6–8 compare the original images, labeled samples, and predicted images for the three cropping sizes, respectively. In order to assure comparability, the network parameters of the U-Net model were the same in the comparison experiments, namely, the batch size was set to four and the epochs were set to 100. As shown in Table 4, all the evaluation metrics are the highest for the cropping size of 256×256 px. There are great differences in accuracy among the three cropping sizes, indicating that cropping sizes have a nonnegligible influence on the U-Net model [32].

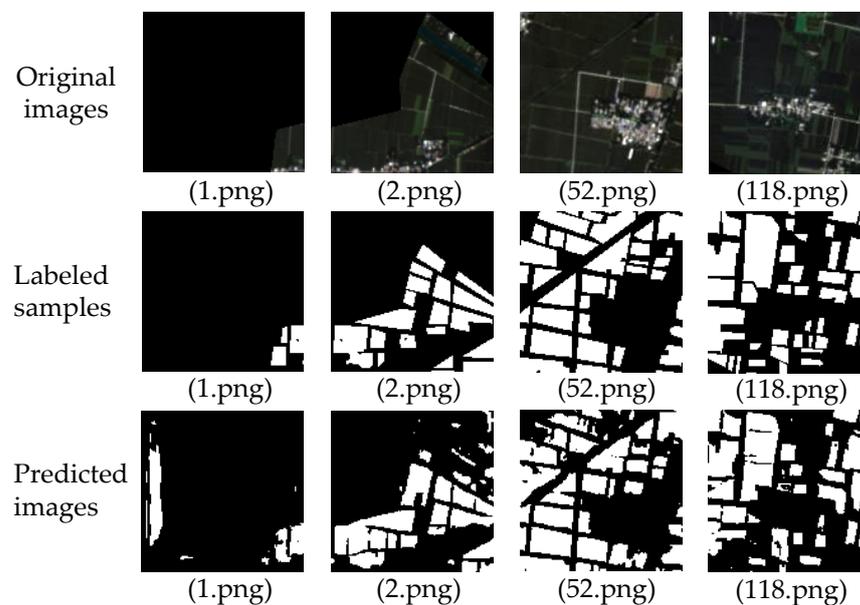


Figure 6. The test results under the 128×128 px cropping size.

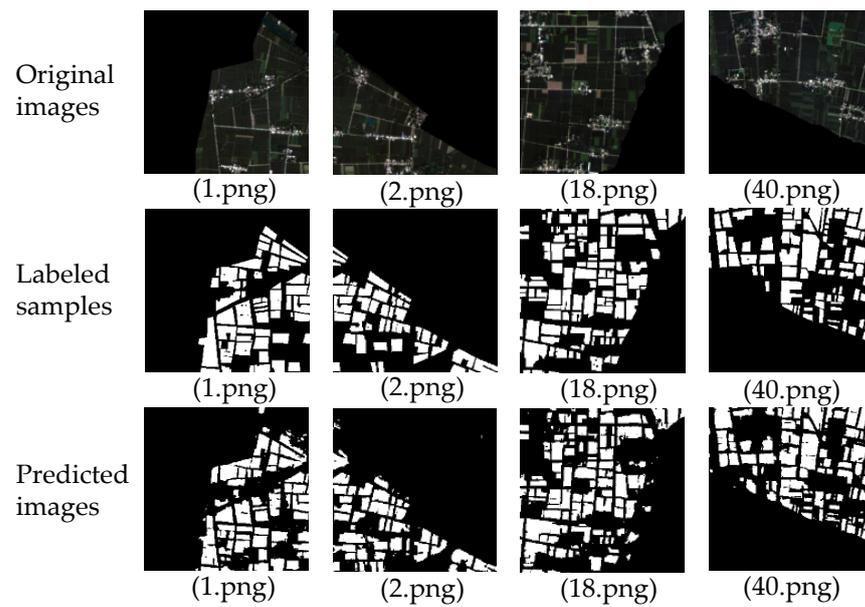


Figure 7. The test results under the 256×256 px cropping size.

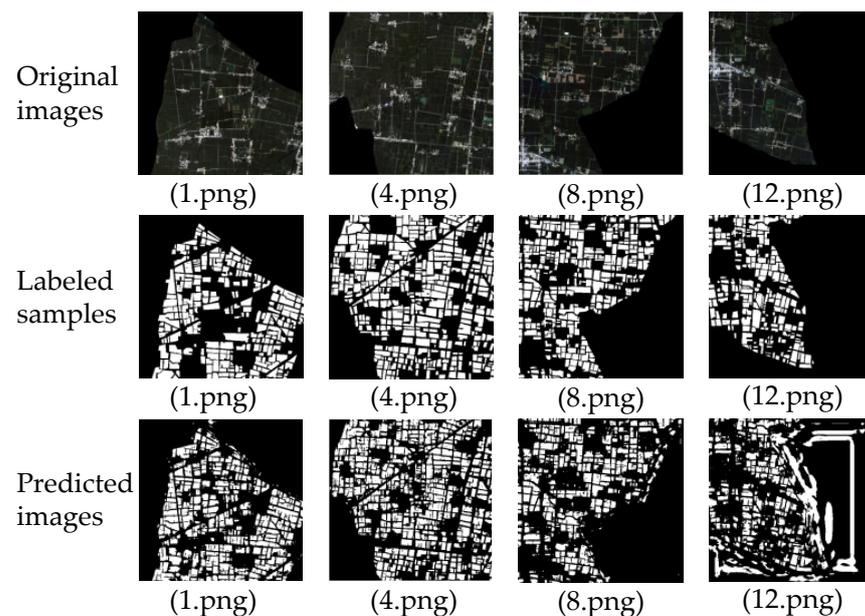


Figure 8. The test results under the 512×512 px cropping size.

Table 4. Comparisons of accuracy under different cropping sizes.

Cropping Size	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)	IoU (%)	MIoU (%)
128×128	88.75	75.85	80.73	78.21	64.22	75.06
256×256	92.31	85.43	82.52	83.95	72.34	81.35
512×512	77.46	53.13	71.82	61.08	43.96	58.29

Figure 6 shows some test results under the 128×128 px cropping size. The overall prediction effect is fairly good, however, the edges and corners in some images are not accurately predicted. For the predicted image of 1.png, a large area is classified as soybean planting areas but they are not really soybeans according to the labeled image. It can be found that more pixels are misclassified as soybean planting areas. For the predicted image

of 2.png, some small areas are identified as soybean planting areas, with an *Accuracy* of 88.75% and *MIoU* of 75.06%. As a whole, the integrity and completeness are worse for the four images.

In comparison with Figure 6, the *Accuracy* reaches 92.31% under the 256×256 px cropping size (Figure 7). The *MIoU* reaches 81.35% and the training accuracy of the model reaches 99.69%. It is obvious that there are only subtle differences between predicted images and labeled samples. The primary reason is that there are textural differences in some corners, which causes the model not to achieve the ideal effect. More than 80% of *MIoU* indicates that the overall performance achieves the desired results.

When the cropping size reaches 512×512 px, there are large misclassification areas (Figure 8), especially for the 12.png. More pixels are incorrectly predicted, resulting in an *Accuracy* of less than 80%. The *MIoU* is only 58.29%, showing that the training effect is not satisfactory.

3.3. Influence of Training Epochs on Prediction Accuracies

When training the datasets, the predicted map of each epoch was saved. It can be found that with the increase in training epochs, the training effect becomes better. An optimal model can be achieved when the accuracy reached its peak. As shown in Figure 9, the test results are optimal when the training epoch is 60, and the predicted soybean planting areas are closest to the labeled samples. When the training epoch is less than 60, the test results show more or less misclassified pixels to a varying degree, especially for the 24.png test image. A few background areas are misinterpreted as soybean planting areas. This phenomenon shows that insufficient training will lead to underfitting. When the training epoch reaches 80 and 100, the misclassification phenomena are slightly reduced compared with the epoch of 60, however, more predicted pixels are produced. Combined with Figure 5, when the training epoch reaches 60, the accuracy and loss reach the most stable state. Excessive training does not increase the training effect and the epoch of 60 was selected to train and test the model.

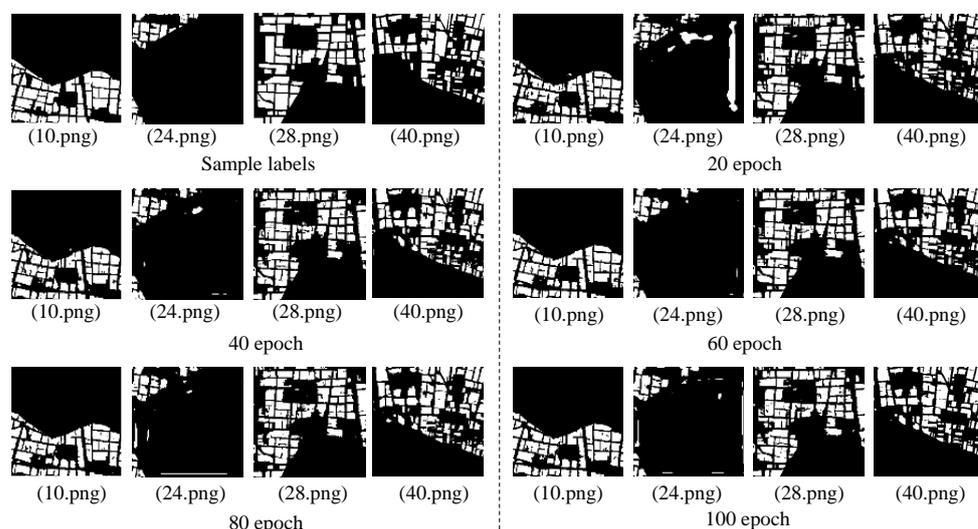


Figure 9. Visual comparison of extracted soybeans at different training epochs.

Figure 10 shows that the *Accuracy*, *Recall*, *F1*, *IoU*, and *MIoU* show a first increase and then a decrease in trend with the increase in training epochs. When the epochs reach 60, the prediction accuracies are the highest with an *MIoU* of 81.36%, showing that the prediction achieves good performance.

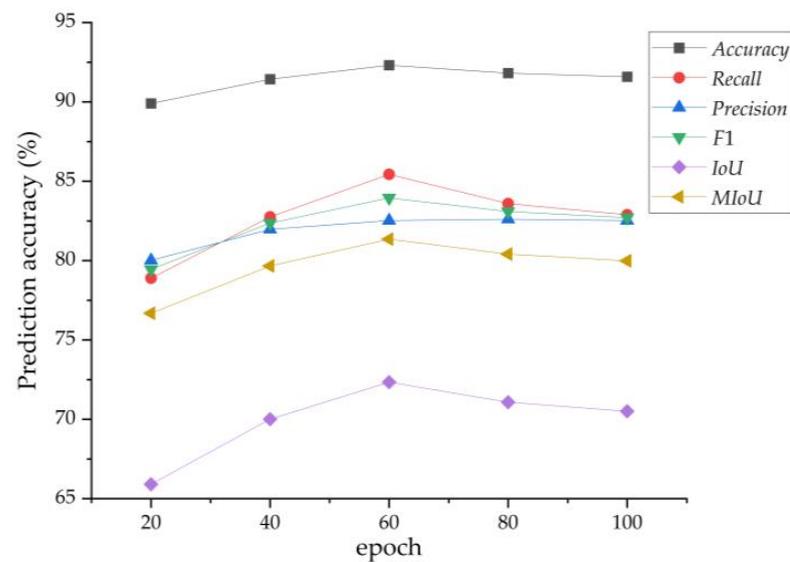


Figure 10. Comparison of prediction accuracies at different training epochs.

3.4. Comparison of Prediction Accuracies among Different Models

SegNet [33] and DeepLabv3+ [34] deep learning network models were selected as the comparative experiments to verify the U-Net model. The training epoch was set to 100, the batch size was set to 4, and the cropping size was 256×256 px for the three models. Four typical images of soybean planting areas were selected to compare the three models, and their original images and corresponding labels are shown in Figure 11.

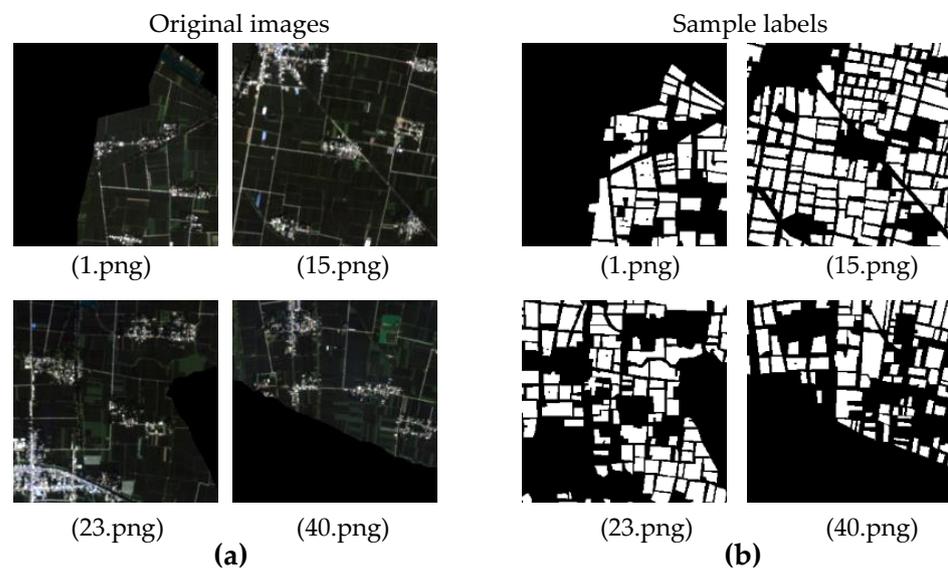


Figure 11. (a) Original sample images and (b) labeled samples.

As shown in Figure 12, the overall spatial distributions of soybean planting areas are similar for the U-Net, SegNet, and DeepLabv3+ models but there are actually significant differences between them. It is obvious that the most similar test results to the labeled samples are from the U-Net model. There are also some additional pixels that are misclassified soybean planting areas for U-Net and SegNet models but the extraction effect of U-Net is better. More soybean planting areas are not fully identified for the SegNet model. The relatively regular shaped soybean fields cannot be embodied. When analyzing the predicted results from DeepLabv3+ model, it can be found that the boundaries of most extracted areas are smoothed and do not show the angular shapes of soybean planting

areas. In general, the *Accuracy* of the U-Net model reaches 92% and is improved by 4% than the SegNet model. The *MIoU* reaches 81.35%, 71.31%, and 72.46%, respectively, for the U-Net, SegNet, and DeepLabv3+, which is, respectively, improved by 10% and 9% for U-Net compared with SegNet and DeepLabv3+.

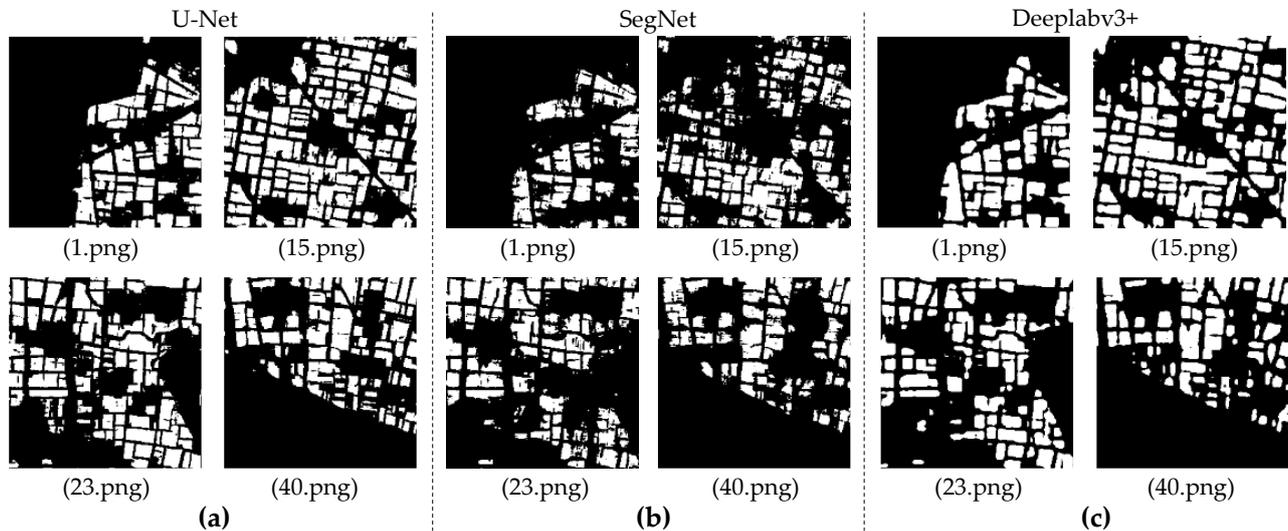


Figure 12. Comparison of identified soybean planting areas using the three models: (a) U-Net, (b) SegNet, (c) DeepLabv3+.

4. Conclusions

It is necessary to detect images with good resolution in an advanced phase of the vegetative cycle. The 2-meter-resolution fused GF-1 imagery and blooming and podding stages are selected to accurately identify soybean planting areas. In addition, considering the parameter optimization of the U-Net model, different cropping sizes, training epochs, and comparative models are adopted to explore the influence of the U-Net model on the identification accuracy. The best U-Net model is obtained by comparing three cropping sizes of 128×128 , 256×256 , and 512×512 px and 20, 40, 60, 80, and 100 training epochs. The comparative analysis shows that the extraction effect is optimal when the cropping size is 256×256 px, with an *Accuracy* of 92.31% and *MIoU* of 81.35%. Afterward, the training epoch was set to 100 and every 20 epochs were extracted as comparisons. When the number of training epochs reaches 60, the prediction result is optimal. After comparing the SegNet and DeepLabv3+ models, the extraction effect of the U-Net model proved to be the best. The cropping images were used to compare the extraction results rather than a complete administrative division. When the separate identified images were merged into an administrative division, there were some problems to be solved, such as seams and boundaries among different soybean fields. This study can provide a methodological reference for extracting planting areas of soybean or other crops. In a future study, sub-meter remote sensing imagery can be used to further validate our model. In addition, improved U-Net models by adding an attention mechanism, residual module, multi-scale features, etc., can be adopted to further improve the identification accuracy of soybean planting areas.

Author Contributions: J.Z. and D.L. conceived and designed the experiments; S.Z. and T.X. performed the experiments; S.Z., X.B. and L.H. analyzed the data; J.Z., W.H. and D.L. wrote and proofread the paper. 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 (2019YFE0115200), the Natural Science Foundation of Anhui Province (2008085MF184), Science and Technology Major Project of Anhui Province (202003a06020016), and the Excellent Scientific Research and Innovation Team (2022AH010005).

Data Availability Statement: Not applicable.

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

References

1. Liu, X.; Jin, J.; Wang, G.; Herbert, S.J. Soybean yield physiology and development of high-yielding practices in Northeast China. *Field Crop. Res.* **2008**, *105*, 157–171. [[CrossRef](#)]
2. da Silva Junior, C.A.; Leonel-Junior, A.H.S.; Rossi, F.S.; Correia Filho, W.L.F.; de Barros Santiago, D.; de Oliveira-Júnior, J.F.; Teodoro, P.E.; Lima, M.; Capristo-Silva, G.F. Mapping soybean planting area in midwest Brazil with remotely sensed images and phenology-based algorithm using the Google Earth Engine platform. *Comput. Electron. Agr.* **2020**, *169*, 105194. [[CrossRef](#)]
3. Monteiro, L.A.; Ramos, R.M.; Battisti, R.; Soares, J.R.; Oliveira, J.C.; Figueiredo, G.K.; Lamparelli, R.A.C.; Nendel, C.; Lana, M.A. Potential use of data-driven models to estimate and predict soybean yields at national scale in Brazil. *Int. J. Plant Prod.* **2022**, *16*, 691–703. [[CrossRef](#)]
4. Diao, C. Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages. *Remote Sens. Environ.* **2020**, *248*, 111960. [[CrossRef](#)]
5. Santos, L.B.; Bastos, L.M.; de Oliveira, M.F.; Soares, P.L.M.; Ciampitti, I.A.; da Silva, R.P. Identifying nematode damage on soybean through remote sensing and machine learning techniques. *Agronomy* **2022**, *12*, 2404. [[CrossRef](#)]
6. Chang, J.; Hansen, M.C.; Pittman, K.; Carroll, M.; DiMiceli, C. Corn and soybean mapping in the United States using MODIS time-series data sets. *Agron. J.* **2007**, *99*, 1654–1664. [[CrossRef](#)]
7. Huang, J.; Hou, Y.; Su, W.; Liu, J.; Zhu, D. Mapping corn and soybean cropped area with GF-1 WFV data. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 164–170.
8. Zhong, L.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 1–13. [[CrossRef](#)]
9. Zhu, M.; She, B.; Huang, L.; Zhang, D.; Xu, H.; Yang, X. Identification of soybean based on Sentinel-1/2 SAR and MSI imagery under a complex planting structure. *Ecol. Inform.* **2022**, *72*, 101825. [[CrossRef](#)]
10. Randelović, P.; Đorđević, V.; Milić, S.; Balešević-Tubić, S.; Petrović, K.; Miladinović, J.; Đukić, V. Prediction of soybean plant density using a machine learning model and vegetation indices extracted from RGB images taken with a UAV. *Agronomy* **2020**, *10*, 1108. [[CrossRef](#)]
11. Habibi, L.N.; Watanabe, T.; Matsui, T.; Tanaka, T.S. Machine learning techniques to predict soybean plant density using UAV and satellite-based remote sensing. *Remote Sens.* **2021**, *13*, 2548. [[CrossRef](#)]
12. Yang, Q.; She, B.; Huang, L.; Yang, Y.; Zhang, G.; Zhang, M.; Hong, Q.; Zhang, D. Extraction of soybean planting area based on feature fusion technology of multi-source low altitude unmanned aerial vehicle images. *Ecol. Inform.* **2022**, *70*, 101715. [[CrossRef](#)]
13. Zhao, J.; Wang, J.; Qian, H.; Zhan, Y.; Lei, Y. Extraction of winter-wheat planting areas using a combination of U-Net and CBAM. *Agronomy* **2022**, *12*, 2965. [[CrossRef](#)]
14. Shen, Y.; Li, Q.; Du, X.; Wang, H.; Zhang, Y. Indicative features for identifying corn and soybean using remote sensing imagery at middle and later growth season. *Natl. Remote Sens. Bull.* **2022**, *26*, 1410–1422.
15. Paludo, A.; Becker, W.R.; Richetti, J.; Silva, L.C.D.A.; Johann, J.A. Mapping summer soybean and corn with remote sensing on Google Earth Engine cloud computing in Parana state–Brazil. *Int. J. Digital Earth* **2020**, *13*, 1624–1636. [[CrossRef](#)]
16. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [[CrossRef](#)]
17. Seo, B.; Lee, J.; Lee, K.D.; Hong, S.; Kang, S. Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. *Field Crop. Res.* **2019**, *238*, 113–128. [[CrossRef](#)]
18. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land use land cover classification with U-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote Sens.* **2021**, *13*, 3600. [[CrossRef](#)]
19. Yao, Y.; Liang, S.; Fisher, J.B.; Zhang, Y.; Cheng, J.; Chen, J.; Jia, K.; Zhang, X.; Bei, X.; Shang, K.; et al. A novel NIR–red spectral domain evapotranspiration model from the Chinese GF-1 satellite: Application to the Huailai agricultural region of China. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4105–4119. [[CrossRef](#)]
20. Sun, W.; Tian, Y.; Mu, X.; Zhai, J.; Gao, P.; Zhao, G. Loess landslide inventory map based on GF-1 satellite imagery. *Remote Sens.* **2017**, *9*, 314. [[CrossRef](#)]
21. Li, J.; Chen, X.; Tian, L.; Huang, J.; Feng, L. Improved capabilities of the Chinese high-resolution remote sensing satellite GF-1 for monitoring suspended particulate matter (SPM) in inland waters: Radiometric and spatial considerations. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 145–156. [[CrossRef](#)]
22. Sola, J.; Sevilla, J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* **1997**, *44*, 1464–1468. [[CrossRef](#)]
23. Saranya, C.; Manikandan, G. A study on normalization techniques for privacy preserving data mining. *Int. J. Eng. Technol.* **2013**, *5*, 2701–2704.
24. Wambugu, N.; Chen, Y.; Xiao, Z.; Tan, K.; Wei, M.; Liu, X.; Li, J. Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review. *Int. J. Appl. Earth Observ. Geoinform.* **2021**, *105*, 102603. [[CrossRef](#)]

25. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
26. Freudenberg, M.; Nölke, N.; Agostini, A.; Urban, K.; Wörgötter, F.; Kleinn, C. Large scale palm tree detection in high resolution satellite images using U-Net. *Remote Sens.* **2019**, *11*, 312. [[CrossRef](#)]
27. Liu, G.; Bai, L.; Zhao, M.; Zang, H.; Zheng, G. Segmentation of wheat farmland with improved U-Net on drone images. *J. Appl. Remote Sens.* **2022**, *16*, 034511. [[CrossRef](#)]
28. Zhang, S.; Zhang, C. Modified U-Net for plant diseased leaf image segmentation. *Comput. Electron. Agric.* **2023**, *204*, 107511. [[CrossRef](#)]
29. Liu, X.; Liu, X.; Wang, Z.; Huang, G.; Shu, R. Classification of laser footprint based on random forest in mountainous area using GLAS full-waveform features. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2022**, *15*, 2284–2297. [[CrossRef](#)]
30. Behera, S.K.; Rath, A.K.; Sethy, P.K. Fruits yield estimation using Faster R-CNN with MIoU. *Multimed. Tools Appl.* **2021**, *80*, 19043–19056. [[CrossRef](#)]
31. Lee, S.; He, C.; Avestimehr, S. Achieving small-batch accuracy with large-batch scalability via Hessian-aware learning rate adjustment. *Neural Netw.* **2023**, *158*, 1–14. [[CrossRef](#)]
32. Dong, X.; Lei, Y.; Wang, T.; Thomas, M.; Tang, L.; Curran, W.J.; Liu, T.; Yang, X. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Med. Phys.* **2019**, *46*, 2157–2168. [[CrossRef](#)]
33. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
34. 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.

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.