

Article

A Transferable Learning Classification Model and Carbon Sequestration Estimation of Crops in Farmland Ecosystem

Lijun Wang ^{1,2,3}, Yang Bai ^{1,2,3,*} , Jiayao Wang ^{1,2,4}, Fen Qin ^{1,2,4,5}, Chun Liu ^{1,2,3}, Zheng Zhou ^{1,2,3} and Xiaohao Jiao ^{2,3}

- ¹ Henan Industrial Technology Academy of Spatio-Temporal Big Data, Henan University, Kaifeng 475004, China
- ² College of Geography and Environmental Science, Henan University, Kaifeng 475004, China
- ³ Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng 475004, China
- ⁴ Henan Technology Innovation Center of Spatial-Temporal Big Data, Henan University, Kaifeng 475004, China
- ⁵ Key Research Institute of Yellow River Civilization and Sustainable Development, Ministry of Education, Henan University, Kaifeng 475004, China
- * Correspondence: daisy.baiyang@henu.edu.cn

Abstract: Frequent agricultural activities in farmland ecosystems bring challenges to crop information extraction from remote sensing (RS) imagery. The accurate spatiotemporal information of crops serves for regional decision support and ecological assessment, such as disaster monitoring and carbon sequestration. Most traditional machine learning algorithms are not appropriate for prediction classification due to the lack of historical ground samples and poor model transfer capabilities. Therefore, a transferable learning model including spatiotemporal capability was developed based on the UNet++ model by integrating feature fusion and upsampling of small samples for Sentinel-2A imagery. Classification experiments were conducted for 10 categories from 2019 to 2021 in Xinxiang City, Henan Province. The feature fusion and upsampling methods improved the performance of the UNet++ model, showing lower joint loss and higher mean intersection over union (mIoU) values. Compared with the UNet, DeepLab V3+, and the pyramid scene parsing network (PSPNet), the improved UNet++ model exhibits the best performance, with a joint loss of 0.432 and a mIoU of 0.871. Moreover, the overall accuracy and macro F_1 values of prediction classification results based on the UNet++ model are higher than 83% and 58%, respectively. Based on the reclassification rules, about 3.48% of the farmland was damaged in 2021 due to continuous precipitation. The carbon sequestration of five crops (including corn, peanuts, soybean, rice, and other crops) is estimated, with a total carbon sequestration of 2460.56, 2549.16, and 1814.07 thousand tons in 2019, 2020, and 2021, respectively. The classification accuracy indicates that the improved model exhibits a better feature extraction and transferable learning capability in complex agricultural areas. This study provides a strategy for RS semantic segmentation and carbon sequestration estimation of crops based on a deep learning network.

Keywords: transferable learning; crop segmentation and classification; Sentinel-2A; UNet++; carbon sequestration



Citation: Wang, L.; Bai, Y.; Wang, J.; Qin, F.; Liu, C.; Zhou, Z.; Jiao, X. A Transferable Learning Classification Model and Carbon Sequestration Estimation of Crops in Farmland Ecosystem. *Remote Sens.* **2022**, *14*, 5216. <https://doi.org/10.3390/rs14205216>

Academic Editors: Stanisław Lewiński and Wojciech Drzewiecki

Received: 24 August 2022

Accepted: 16 October 2022

Published: 18 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

As a part of the terrestrial ecosystem, the farmland ecosystem plays an important role in global climate change, food production, and ecological assessment [1]. As an increasing number of satellite observations become available, there is an urgent need to obtain accurate crop maps from the massive images and assess the significance of each crop type in farmland ecosystems. In recent years, several studies have been carried out on crop monitoring and carbon sequestration in different ecosystems. However, these studies mainly focused on the forest ecosystem and grassland ecosystem [2–4]. Due to the strong human

factors, crops in farmland have seasonal and complex characteristics, leading to more challenges in crop automatic classification and carbon sequestration estimation. Therefore, a method must be established to extract accurate crop information using remote sensing (RS) technology in order to estimate crop area, disaster loss, and carbon sequestration.

Farmland vegetation is an essential part of the carbon cycle of the terrestrial ecosystem, and the crops play different roles in carbon sequestration [5]. As a developing agricultural country with a large population, cropland accounts for about 12.5% of the total area of China. It is necessary to pay more attention to monitoring the crop planting area and its impact on the carbon cycle. The emerging satellite technologies with diverse sensors provide reliable large-scale land use and crop observation [6]. In particular, the Sentinel-2A (S-2A) satellite developed by the European Space Agency provides 13-band optical imagery with a high spatial resolution of 10 m. The S-2A images have been widely used in global agricultural monitoring, such as crop yield estimation, disaster monitoring, and environmental assessment. In addition, the Google Earth Engine (GEE) provides an interactive platform for feature acquisition and geospatial algorithm development for satellite imagery [7,8]. Several studies show that it is convenient to obtain multiple feature images from GEE, which are essential for crop classification in complex agricultural areas [7,9]. Therefore, combining spectral bands, derived features (including vegetable indices and texture features), and crop phenological information, has become an important method to overcome the problems of within-class diversity and between-class similarity and to obtain high-accuracy crop information.

Due to the seasonality of crop planting, the acquisition of ground samples and imagery often has to be completed in a short time. On the one hand, it is difficult to obtain time-series optical images during the crop growth period because of climatic factors. On the other hand, traditional machine learning methods, such as support vector machine (SVM), random forest (RF), and object-oriented classification methods, rely on a large number of high-quality ground sample data sets [10]. However, these sample data sets would lose their effectiveness due to crop seasonality limitations across years. Furthermore, the pixel-based classification results struggle to meet the precise application requirements due to the “salt and pepper phenomenon” [8]. For the object-oriented classification method, it is difficult to determine the optimal segmentation scale due to the image resolution, crop complexity, and field fragmentation, resulting in over-segmentation and under-segmentation problems [11]. Despite the success of traditional machine learning in earth science, the above problems and limitations have hampered the model performance and classification accuracy [12]. To alleviate the dependence and workload related to ground investigations, it is essential to use the transferable learning capability of classification models and to complete the automatic extraction of crop information from RS images based on few label data sets.

Deep learning has become a semantic segmentation and target detection tool for solving many challenging problems in computer vision [9,12]. As data-driven deep neural networks, deep learning networks usually outperformed shallow classifiers due to the hierarchical feature transformation capability from multi-feature RS imagery. The 2012 ImageNet classification challenge and massive label data sets have been crucial in demonstrating the effectiveness of deep convolutional neural networks (CNNs) [13]. However, it is difficult to obtain a large number of spatiotemporal crop data sets for RS images. The end-to-end transferable learning capacity reduced manual feature engineering and improved model performance. Many studies have shown that CNNs and fully convolutional networks (FCNs) are important network architecture for RS image and natural scene semantic segmentation [14,15]. Furthermore, FCNs introduced the encoder–decoder paradigm, which transforms the feature map size to the original image size and overcomes the limitations of losing detailed information in the down-sampling and the fixed size of the input data set [16]. Therefore, the development of FCN facilitates the transferable model performance and high-accuracy crop mapping in the large-scale region across years.

To obtain high-accuracy crop maps, current studies mainly focus on feature fusion and FCN-based semantic segmentation methods. Li et al. [17] proposed an improved

deep learning method to classify crops of corn and soybean from RS time-series images and achieved the classification results with a Kappa coefficient of 0.79 and an overall accuracy of 0.86. Giannopoulos et al. [18] extended a deep learning model based on UNet architecture to extract information from Landsat-8 images and achieved higher accuracy than low-order deep learning models. Yang et al. [19] used different semantic segmentation models, including temporal feature-based segmentation model, long short-term memory (LSTM), and UNet, to accomplish rice mapping based on time-series SAR images. Wang et al. [20] proposed a two-stage model that fused DeepLab V3+ and UNet for cucumber leaf disease severity classification with a segmentation accuracy of 93.27%. The above-mentioned studies have achieved high classification accuracy and provided a reference for image segmentation and classification of major crops using remote sensing images. However, applying the semantic segmentation models of the studies in complex agricultural areas with imbalanced samples still needs further testing. Moreover, Zhou et al. [21] proved that the UNet++ model had more advantages in feature map-generating strategy and image semantic segmentation with few training sample data sets. Wang et al. [9] used the improved UNet++ architecture to classify 10 categories from Sentinel-2 imagery (including 17 bands of spectral, vegetation indices, and texture features) in 3 years, and indicated that the UNet++ achieved higher segmentation and classification accuracy than UNet and DeepLab V3+. These studies provided methodological support for feature fusion and agricultural information extraction of Sentinel-2 imagery. However, time series prediction classification results of recent research still need to be further applied and analyzed in farmland ecosystems, such as in disaster assessment and crop carbon sequestration. Therefore, it is essential to establish a regional and transferable-learning model with few training data sets for RS crop classification and carbon sequestration estimation during the critical crop growth stage.

Chinese agricultural land has experienced dramatic changes in crop area, cropping system, and planting structure optimization in recent decades [22]. These changes can substantially affect crop carbon sequestration in the farmland. The widespread and scattered smallholders in China have a profound impact on agricultural production in response to climate change. Tang et al. [23] evaluated the contribution and estimation error of the farmland ecosystem to carbon sequestration in China's terrestrial ecosystem. The uncertainties of carbon sequestration estimation can be reduced by collecting the finer crop area and statistical data in farmland ecosystem. In terms of agro-ecosystem services, vegetation indices and time-series production were used to evaluate and predict the impact of current vegetation cover on the farmland ecosystem [24,25], or to quantify the spatiotemporal changes, including the assessment of carbon sources/sinks and soil erosion [26]. Zhang et al. [27] provided a novel perspective on LCC (land cover change)-induced gross primary production (GPP) changes and concluded that the LCC-induced reduction in GPP was partially offset by increases in cropland using the GEE platform. Wang et al. [28] showed that the changes in crop planting area can substantially affect greenhouse gas emissions, and that farmland in China was a carbon source due to a large amount of CH₄ emission in paddy lands. Such a capability to collect timely and high-accuracy crop information from multispectral imagery using the transferable-learning model and to produce accurate crop maps is crucial to assess the farmland ecosystem services.

In this study, we pay more attention to the automatic classification of crop information from S-2A imagery in complex agricultural areas. As an agricultural region in China, Xinxiang City of Henan Province has a representative topography and diversified crop types. This study aims to (1) evaluate the influence of feature-selection schemes on the UNet++ model; (2) evaluate the performance of different deep learning models; (3) compare prediction classification accuracies of different models based on overall accuracy, user's accuracy, producer's accuracy, and F₁ scores; and (4) complete crop mapping and carbon sequestration estimation in farmland ecosystem. Overall, this work aims to offer an improved deep learning procedure as applied to RS in crop monitoring and carbon sequestration estimation.

2. Study Sites and Data

2.1. Study Sites

This study focused on crop mapping in Xinxiang City of Henan Province, China (Figure 1), with a total area of about 8251.08 km². The mountains are mainly distributed in the northern part. The crop types in the plain area are diverse, mainly including corn, soybean, peanut, rice, other crops (OTH, including vegetables, tobacco, and Chinese herbal medicine), non-cultivated land (NCL), and greenhouses (GH). Other categories include urban (URB, including residential areas, roads, and other construction), forest land (FL, including nursery forest planted on farmland, green areas around cities and villages, and mountain areas), and water (WAT, including the rivers, ditches, ponds, and inundated areas). The crops are sown in early June and harvested in mid-September, with a growth period of about 110 d. The mean annual precipitation of the study area is about 581 mm and concentrated from mid-July to late August. However, the precipitation in 2021 reaches 1377.6 mm, which limits the acquisition of time-series optical satellite images. Four sites (A, B, C, and D) in 2019 and 2020 were chosen as the source of training data sets for the deep learning model, each covering $\sim 20 \times 20$ km. In addition, five sample plots (S1–S5), six validation sample plots (VS1–VS6), and some validation sample points are selected and investigated for image labeling and classification accuracy evaluation.

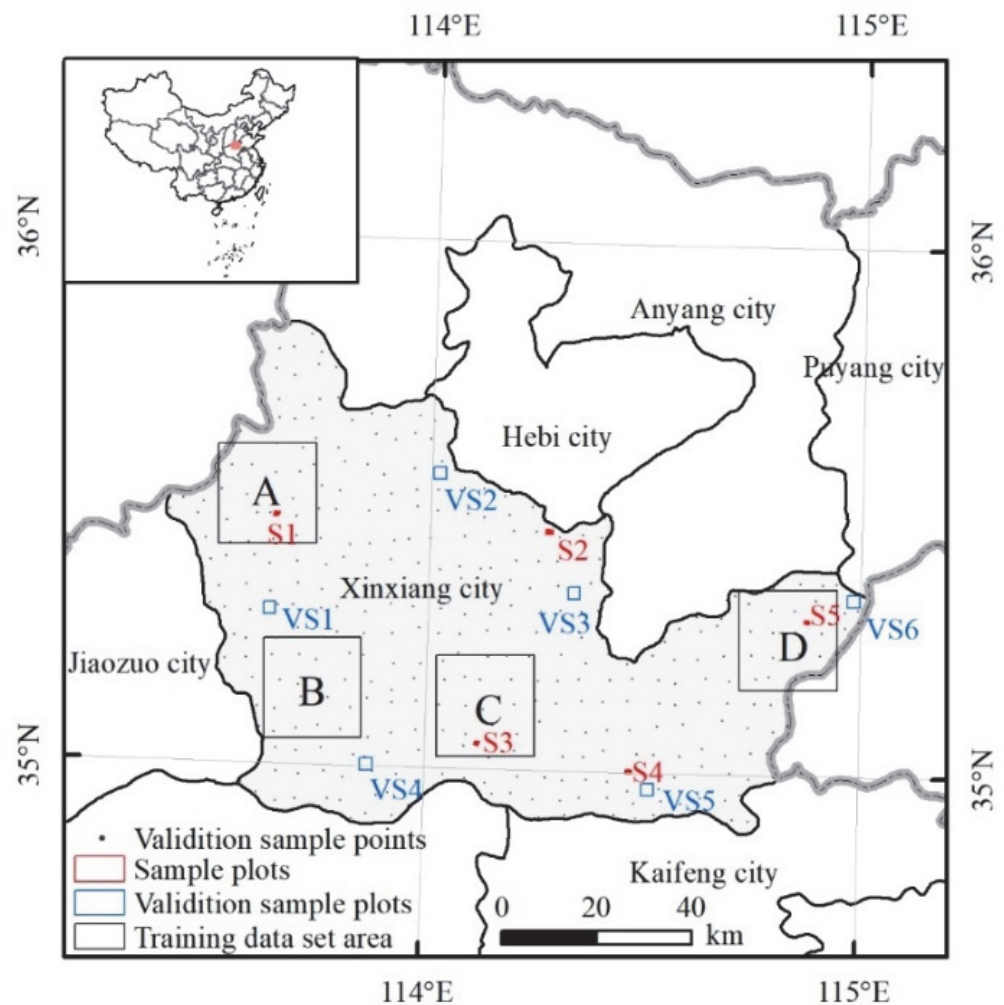


Figure 1. Geography of Xinxiang City on the northern plain of Henan Province, China.

2.2. Sentinel-2A Data

Due to the low vegetation cover and soil spectral effects in the early stage, high quality and cloudless images with similar dates were selected in the late growth stage in early September of 2019, 2020, and 2021. On 4 September 2019, 5 September 2020, and 9 September 2021, we collected five satellite images covering the whole area, with orbital numbers of T49SGU, T50SGV, T50SKD, T50SKE, and T50SLE, respectively. Sentinel Level-2A atmospheric bottom reflectance data follow the U.S. Military Grid Reference System (US-MGRS) to set the satellite orbit number (such as T50SGV) and have been preprocessed in GEE [8]. Hence, we can quickly calculate and compose S-2A images with different features at 10 m resolution, as listed in Table 1. The vegetation indices and texture features were referred to the supplementary materials of Chen et al. [29] and Wang et al. [14], respectively. The texture features were derived based on the *NNIR* band.

Table 1. S-2A schemes with different features.

Scheme	Features	Feature Variables
1	Spectral bands	Blue (B), green (G), red (R), red edge 1 (RE1), red edge 2 (RE2), red edge 3 (RE3), near-infrared (NIR), narrow NIR (NNIR), shortwave infrared 1 (SWIR1), shortwave infrared 2 (SWIR2)
2	Spectral bands, vegetation indices, and texture	B, G, R, RE1, RE2, RE3, NIR, NNIR, SWIR1, SWIR2, modified normalized difference water index (MNDWI), normalized difference build-up index (NDBI), normalized difference vegetation index (NDVI), red-edge NDVI (RENDVI), sum average (SAVG), correlation (CORR), dissimilarity (DISS)

2.3. Sample Data

The sample data were collected using a handheld global positioning system every year, and an interpretation mark library was established as a reference for image labeling. We investigated five sample plots ($\sim 1 \text{ km}^2$) with a total area of 4.41 km^2 and collected 509, 532, and 695 sample points in 2019, 2020, and 2021, respectively. The sample plots were shown in Figure 2a–d. Then, the training data sets and validation sample data sets were accomplished based on the interpretation mark library. As noted by Luo et al. [30], the selection of representative samples is vital for classification accuracy evaluation. Therefore, six validation sample plots (256×256 pixels) were selected each year to evaluate the prediction classification accuracy. In addition, we randomly selected 100 sample points (Figure 1) each year to verify the two data sets based on a ground survey. The classification verification based on sample plots can reduce the impact of the location uncertainty of sample points on the accuracy of evaluation results. The results revealed 13 misclassifications, including NCL and GH, soybean and other crops, GH and urban, and peanuts and other crops (such as alfalfa). The re-correction training data sets shown in Figure 2e–h satisfy the label quality and the training requirements. The validation sample plots were shown in Figure 2i–l.

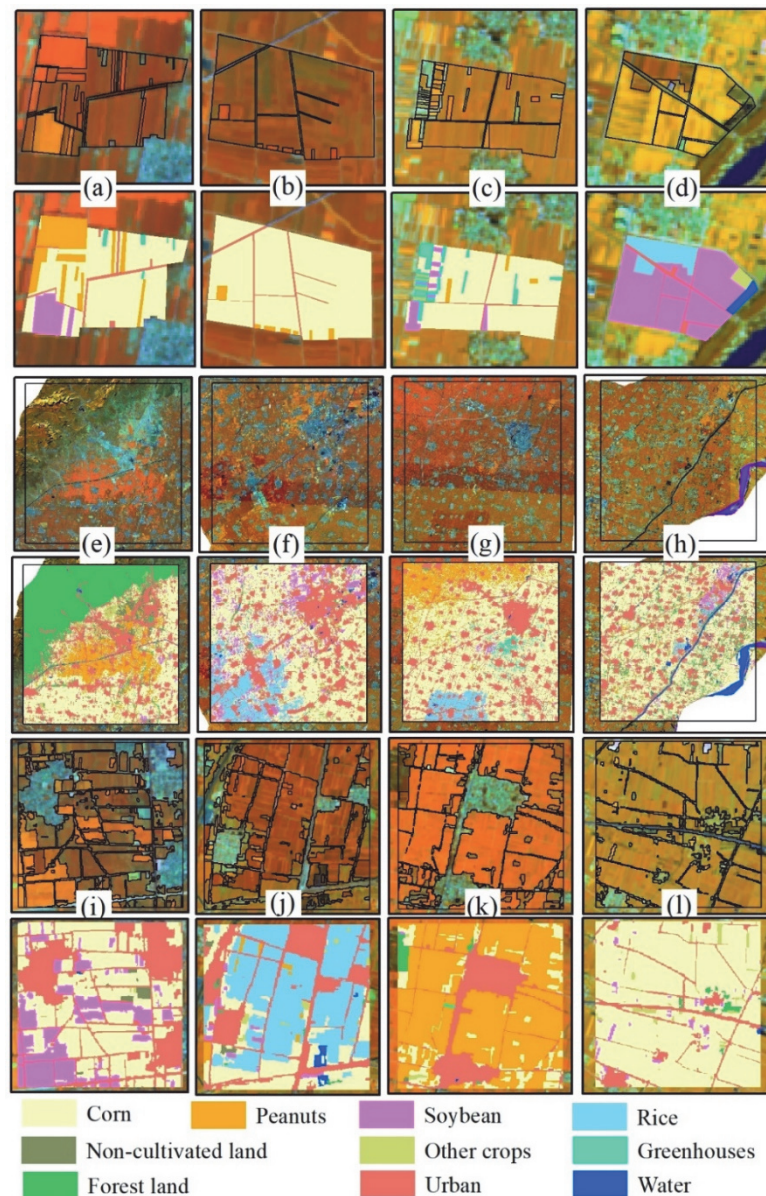


Figure 2. Sample plots, training data set, and validation sample plots. (a) Sample plot 1 (S1) in 2019; (b) sample plot 2 (S2) in 2020; (c) sample plot 4 (S4) in 2021; (d) sample plot 5 (S5) in 2021; (e) training data set of Site A in 2019; (f) training data set of Site B in 2019; (g) training data set of Site C in 2020; (h) training data set of Site D in 2020; (i) validation sample plot (VS1) in 2019; (j) validation sample plot (VS4) in 2020; (k) validation sample plot (VS3) in 2021; (l) validation sample plot (VS6) in 2021.

The pixel numbers of training data sets and validation data sets were 43,373,026 and 1,180,755, respectively. The pixel proportions of corn, peanuts, soybean, rice, NCL, other crops, GH, forest land, urban, and water in the training data set were 44.44%, 7.68%, 2.79%, 3.34%, 0.69%, 3.45%, 8.48%, 27.64%, 1.16%, and 0.33%, respectively. Figure 2 shows that corn and peanuts are the main crops, rice is relatively concentrated, and other crops are scattered. Based on the conclusions of Wang et al. [9] and Chen et al. [29], the training data sets of sites A, B, C, and D were clipped to 256×256 pixels using the regular-grid clip method with a zero-repetition rate. The number of clipped images (in *tif* format) and labels (convert from *tif* format to *png* format) used for training was 704 pairs.

3. Methodology

To evaluate the capability of the transferable learning model and the accuracy of crop segmentation and classification, the experiments in this study were designed based on the UNet++ architecture. The workflow of the experiments included image and sample data acquisition and preprocessing, model training and assessment, prediction result accuracy evaluation, and crop mapping and carbon sequestration. The modeling process was performed on an Ubuntu workstation with 32 GB RAM and one NVIDIA Tesla T4 graphics card (16 GB RAM). The training and testing process was implemented on the Python platform using the PyTorch and GDAL packages.

3.1. UNet++ Model and Assessment Indicators

The UNet++ architecture was modified from Zhou et al. [21] and shown in Figure 3. The UNet++ used the intermediate convolutional blocks and the skip connections between blocks to extract image features and transforms the height and width of the feature map to the size of the input image. Compared with the UNet model, the re-designed skip connection could effectively retrieve object details and acquire the shallow and deep features, thus, improving the image segmentation accuracy [9,21]. The backbone pre-trained on the ImageNet data set provided parameters to enhance the model performance and accelerate the convergence on the target task, such as object detection, semantic segmentation, and fine-grained image recognition, especially in the early training stage [31]. In this study, we used Timm-RegNetY-320 in the *segmentation_models_pytorch* library as the backbone for model training. The RegNetY was proposed by Radosavovic et al. [32] and has been verified to outperform other models, such as the DenseNet, Visual Geometry Group, and residual networks.

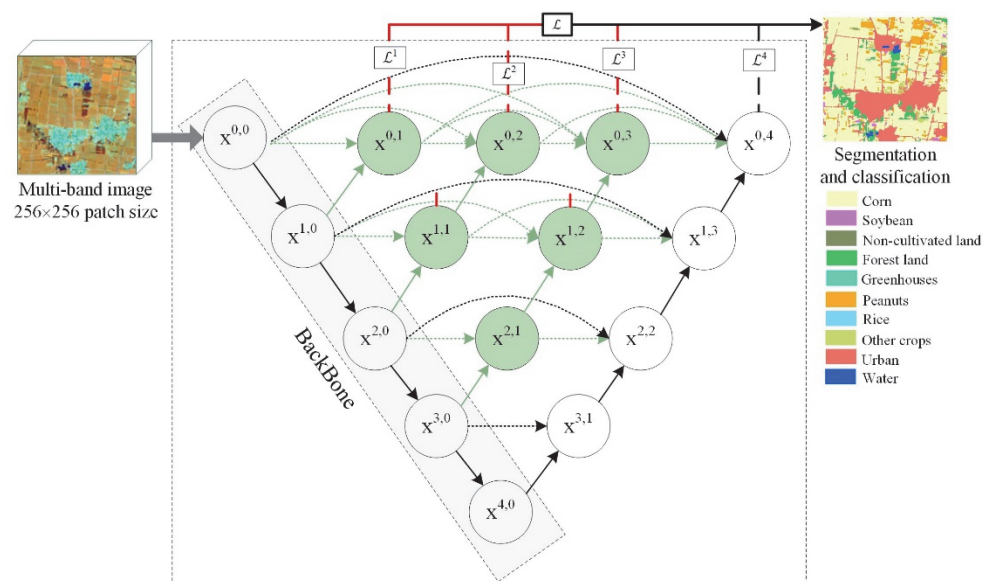


Figure 3. UNet++ architecture used for crop segmentation and classification.

Moreover, three-fold upsampling of small samples (including non-cultivated land and water) was used to alleviate the class-imbalanced problem and improve the model performance according to the pixel proportions in the training data set. Resulting from the small number of training-set pairs, all the pairs were used as training data sets and validation data sets with different linear stretching methods. The enhancement strategy of training data included horizontal flip, vertical flip, diagonal flip, and 5% linear stretch; random 0.8%, 1%, or 2% linear stretching was applied to the validation data. In addition, to accelerate the model convergence and avoid local minima, the cosine annealing learning rate schedule with restarts was adopted to replace the fixed constant learning rate in the

training process [33]. The initial learning rate, weight decay, number of epochs for the first restart (T_0), and value controlling the speed of the learning rate (T_{mult}) were set to 0.0001, 0.001, 2, and 2, respectively.

The model performance was evaluated by using the joint loss function and mean intersection over union (mIoU). The joint loss function involved the label smoothing cross entropy loss function and dice coefficient loss, each with a weight of 0.5. For details of both loss functions, please refer to the website (<https://github.com/BloodAxe/pytorch-toolbelt> (accessed on 5 June 2022)). The smoothing factor of the label smoothing cross entropy loss function was set to 0.1 to enhance the model generalization ability. Dice loss was applied in image segmentation to weaken the effect of imbalanced class problems. The mIoU was used in multi-class tasks to assess the average accuracy between the ground truth and classification of each category.

3.2. Baseline Classification Models and Evaluation Indicators

We selected UNet, DeepLab V3+ [34], and the pyramid scene parsing network (PSPNet) [35] to compare the model performance and prediction classification accuracy. The UNet architecture appeared to be more conveniently modified and achieved satisfactory segmentation results based on few training samples. The DeepLab V3+ extended by FCN showcased the importance of atrous spatial pyramid pooling for the image segmentation tasks and we added a decoder module to refine the boundary details and improve the segmentation accuracy. The PSPNet used a spatial pyramid pooling module to harvest different sub-region representations for pixel-level prediction. Here, we use the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), F_1 score, and macro F_1 (Equation (1)) to evaluate the prediction classification accuracies. The macro F_1 is the average value of F_1 for each category, and is calculated as follows:

$$macro F_1 = \frac{1}{N} \sum_{i=1}^N 2 \times \frac{UA_i \times PA_i}{UA_i + PA_i} \quad (1)$$

where N denotes the number of categories, and UA_i and PA_i are the user's accuracy and producer's accuracy of the i^{th} category, respectively.

3.3. Estimation of Crop Carbon Sequestration

The carbon sequestration of the farmland ecosystem mainly includes crop carbon sequestration and soil carbon sequestration. In this study, the carbon sequestration of crops refers to the amount of fixed C (carbon) in the total biomass by photosynthesis [26]. Due to the category limitations in complex agricultural areas, corn, soybean, peanut, rice, and other crops were selected to estimate carbon sequestration. Here, the results of other crops were calculated based on the statistical data of vegetables. Building on previous carbon sequestration works [23,36,37], the estimation of crop carbon sequestration in this study refers to the carbon of economic yield (C_e , Equation (2)), straw (C_s , Equation (2)), and root (C_r , Equation (2)) [36]. The function of crop carbon sequestration estimation is shown in Equation (3). The estimation parameters are listed in Table 2 [37]. The image segmentation and classification results provide the crop area, and the average yield of each crop comes from the Henan Provincial People's Government (<http://www.henan.gov.cn/> (accessed on 23 May 2022)) and the Henan Province Bureau of Statistics (<https://tjj.henan.gov.cn/> (accessed on 5 June 2022)). Currently, it is challenging to obtain actual yield data of different crops due to the small-holder farming mode in China. Furthermore, accurate yield estimates for some crops, such as peanuts, remain elusive. Previous studies usually used statistical yearbook data to estimate crop carbon sequestration. However, due to the difference between the classified area of crops in remote sensing images and the statistical yearbook, the average crop yield was used to reduce crop carbon sequestration estimation errors. Affected by the lag of the statistical yearbook, some crop yield values,

including soybean and rice in 2021, were obtained by averaging over the last four years. Equations (2) and (3) are as follows:

$$C_e = B \times Y \times A \times (1 - w_i), C_s = C_e / L - C_e, C_r = (C_e + C_s) \times R \quad (2)$$

$$C_T = \sum_i C_i = \sum_i \frac{B_i \times Y_i \times A_i \times (1 - w_i) \times (1 + R_i)}{L_i} \quad (3)$$

where C_T denotes the total carbon sequestration of crops in farmland ecosystem, and C_i is the carbon sequestration of the i^{th} category. For the i^{th} category, B_i is the carbon absorption rate, Y_i is the average yield, A_i is the classification area, w_i is the water content, R_i is the root-shoot ratio, and L_i is the economic coefficient, i.e., the ratio of economic output to biological output.

Table 2. Estimation parameters of crop carbon sequestration.

Type	Carbon Content (B_i)/%	Water Content (w_i)/%	Root-Shoot Ratio (R_i)	Economic Coefficient (L_i)
Corn	0.470	0.13	0.170	0.438
Peanuts	0.450	0.10	0.200	0.556
Soybean	0.450	0.13	0.130	0.425
Rice	0.410	0.12	0.125	0.489
Other crops	0.450	0.90	0.250	0.830

4. Results

4.1. Performance Evaluation of the UNet++ Model

Figure 4 shows the model training results based on the UNet++ architecture for both schemes and upsampling methods. The training process was stopped when the joint loss and mIoU values were not improved relative to the previous 10 consecutive results. The training data set was increased from 704 to 2573 pairs by upsampling of NCL and water. The total epochs and training time of both schemes were 72 and 72, and 445.27 and 450.09 min, respectively. After approximately the 33nd epoch, the result curves for both schemes became smoother and more stable. The optimal results were achieved in the 62nd epoch, with a joint loss and mIoU of 0.434 and 0.869, and 0.432 and 0.871, respectively. The model performance with upsampling between both schemes was very small, and the differences of joint loss values and mIoU values were only 0.002.

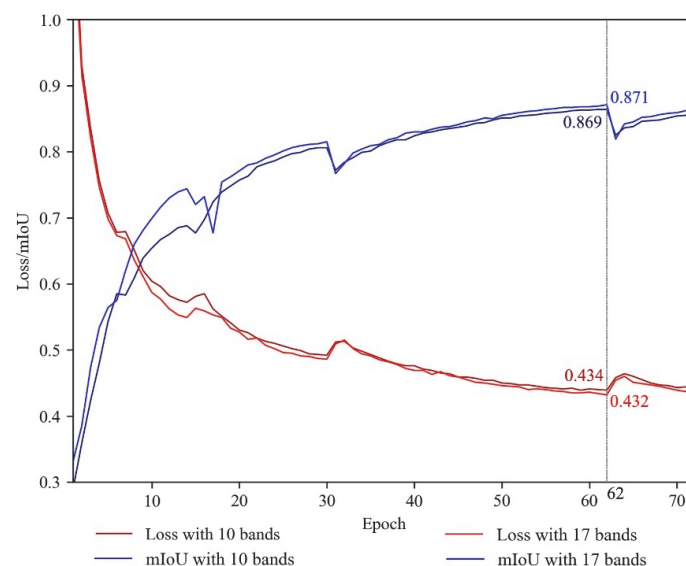


Figure 4. Training parameter results of the two schemes with upsampling of small samples.

In addition, we completed the training models of the two schemes without upsampling to further analyze the influence on model performance. The total epoch and total time of schemes 1 and 2 were 136 and 72, and 279.39 and 151.90 min, respectively. At the 126th and 62nd epochs, the models obtained the optimal joint loss and mIoU values of 0.492 and 0.795, and 0.489 and 0.798, respectively. The training times without upsampling are shorter than those with upsampling because of the unchanged number of training data sets. However, the total epoch of scheme 1 (Table 1) without upsampling was significantly higher than that with upsampling. The mIoU values of the models without upsampling were about 7% lower than those with upsampling. Therefore, the upsampling of small samples can effectively alleviate the class-imbalanced problem and improve the model performance. The UNet++ models with feature fusion and upsampling yielded higher mIoU and lower joint loss values, indicating the applicability in complex crop-classification tasks.

4.2. Evaluation of Baseline Models and Prediction Classification Accuracy

Based on the same training data set, backbone, scheme 2 (Table 1), and upsampling, Table 3 shows the training parameter results of the UNet, DeepLab V3+, and PSPNet models. It can be seen that the total epoch of these models is approximately the same as that of the UNet++ model, while the optimal performance results are lower than that of the UNet++ model. The joint loss values of UNet, DeepLab V3+, and PSPNet are 0.067, 0.089, and 0.156 higher than that of the UNet++ model, respectively, whereas the mIoU values of these models are 0.075, 0.094, and 0.163 lower than that of UNet++ model, respectively. The training time of UNet and PSPNet is shorter than that of UNet++, but the UNet++ model yielded a better performance. The advantages of the modified UNet++ model were evaluated mainly from the model performance and the prediction classification accuracy in this study. Table 3 shows that the modified UNet++ model achieves lower joint loss and higher mIoU values compared with other deep learning models. In addition, the overall accuracy and F_1 score of the prediction results were evaluated to demonstrate the advantages of the modified UNet++ model. Tables 3 and 4 show that, based on the feature selection and upsampling of small samples, the UNet, Deeplab V3+, and PSPNet models have higher mIoU values and higher prediction accuracy than those of the UNet++ model without feature fusion or upsampling methods. The PSPNet can complete the model training in the shortest time. However, its prediction classification accuracy is the lowest compared with other models. Therefore, the improved UNet++ model yields a better generalization capability and spatiotemporal transfer performance than the baseline models, according to the model performance results and prediction classification accuracy.

Table 3. Training parameter results of different models.

Model	Total Epoch	Loss	mIoU	Total Time	Average Time
UNet	72	0.499	0.796	331.43	4.60
DeepLab V3+	72	0.521	0.777	527.65	7.33
PSPNet	70	0.588	0.708	162.51	2.32

To further clarify the transferable learning capability and the influence on the prediction classification accuracy, the OA and macro F_1 of different schemes with or without upsampling were compared based on the validation sample plots and points. The prediction classification accuracies were listed in Table 4. With upsampling of small samples, the OA and macro F_1 values of the UNet++ model were higher than those without upsampling. Based on the UNet++ model, the OA and macro F_1 differences of scheme 1 with upsampling in 2019, 2020, and 2021 were 3.88% and 2.17%, 2.16% and 4.11%, and 2.96% and 0.66% higher than those without upsampling, respectively, whereas those values for scheme 2 were 7.59% and 7.88%, 6.80% and 6.06%, and 5.51% and 0.09% higher than those without upsampling, respectively. In addition to the macro F_1 in 2021, the classification accuracy differences UNet++-based between schemes 1 and 2 without upsampling were lower than those with upsampling.

Table 4. Training parameter results from different schemes and models. Without upsampling and upsampling are abbreviated as WUS and US, respectively.

Year	Indicator	Scheme 1		Scheme 2				
		WUS	US	WUS	US	US		
		UNet++		UNet		DeepLabV3+	PSPNet	
2019	OA	79.46	83.34	80.59	88.18	86.02	80.26	71.66
	Macro F ₁	50.51	52.68	50.66	58.54	55.11	54.73	47.85
2020	OA	76.15	78.31	80.76	87.56	80.07	81.12	77.99
	Macro F ₁	47.03	51.14	51.29	58.71	54.52	54.73	49.86
2021	OA	76.91	79.87	78.01	83.52	74.76	75.85	67.42
	Macro F ₁	58.43	59.09	59.58	59.67	53.21	52.16	47.39

With upsampling of small samples, the prediction classification accuracies of PSPNet were all lower than those of UNet++, UNet, and DeepLab V3+. However, the classification accuracies of UNet and DeepLab V3+ indicated that both models had different advantages in crop classification tasks in complex agricultural planting areas. Therefore, UNet++ is an effective model for classifying multiple categories in complex agricultural areas with respect to UNet, DeepLab V3+, and PSPNet. We fed the training data set of four sites from 2019 to 2020 into different deep learning models with or without sampling and predicted the classification results of the study area in 2019, 2020, and 2021. The prediction accuracies without training data sets demonstrated the better transferable learning performance and generalization of the UNet++ model.

4.3. Reclassification Rules and Crop Mapping

Based on the overall indicators, Table 4 showed that scheme 2 of the UNet++ model yields the optimal classification results. Compared with the results in 2019 and 2020 in Table 4, the OA and macro F₁ values of the prediction classification results in 2021 were almost lower than those in the previous two years. In addition to the lack of training data sets, the rainstorm and continuous precipitation since 20 July 2021 have caused serious damage to farmland. Compared with the images of 2019 and 2020 in Figure 5, the extreme precipitation led to notable water stains and ponding on the farmland until 9 September 2021. Figure 5c shows that the images of the disaster area were misclassified as other crops, urban areas, and water. Thus, it is essential to establish classification rules to further process the misclassification results in 2021. The reclassification rules processed in ArcMap software are as follows:

- (1) We manually drew the approximate boundary of the disaster area shown in Figure 6c, and then converted the three-year classification results into the vector format;
- (2) The disaster area was spatially intersected with the 2021 classification results (including other crops, cities, and water) and the result was named as *2021_disaster_intersect_area*;
- (3) The *2021_disaster_intersect_area* data was spatially intersected with the 2019 and 2020 classification results (including corn, peanuts, soybean, rice, NCL, other crops, and greenhouses), respectively, and the results were named as *2019_disaster_intersect_area* and *2020_disaster_intersect_area*, respectively;
- (4) We spatially intersected *2021_disaster_intersect_area* and *2019_disaster_intersect_area*, and *2021_disaster_intersect_area* and *2020_disaster_intersect_area*, respectively, and the results were named as *2019_2021_disaster_intersect_area* and *2020_2021_disaster_intersect_area*, respectively. Finally, after operating by merge tool, attribute field assignment, dissolve tool, and manual editing, the disaster area result was shown in Figure 5d. The land use and crop mapping of each year recorrected by using the road data of Xinxiang City were shown in Figure 6.

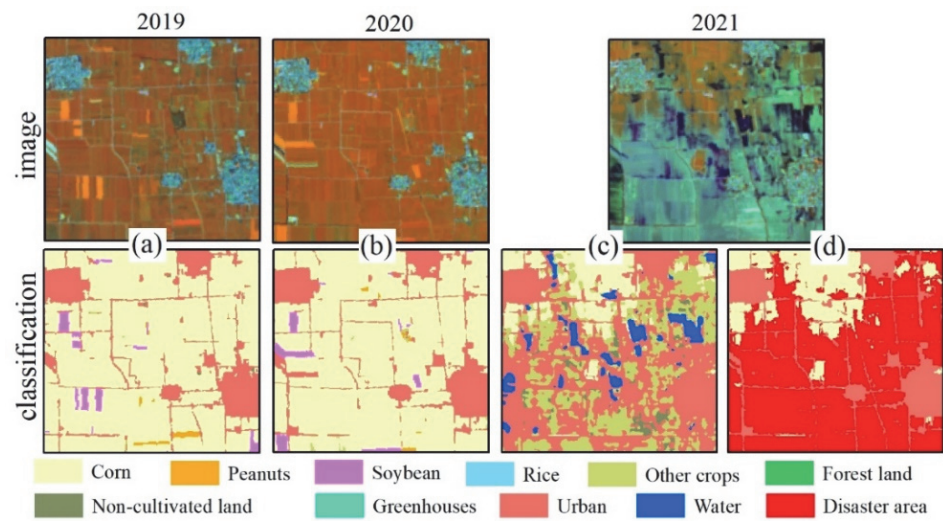


Figure 5. Local segmentation and classification results in different years. (a) Classification result in 2019; (b) classification result in 2020; (c,d) classification and reclassification results in 2021.

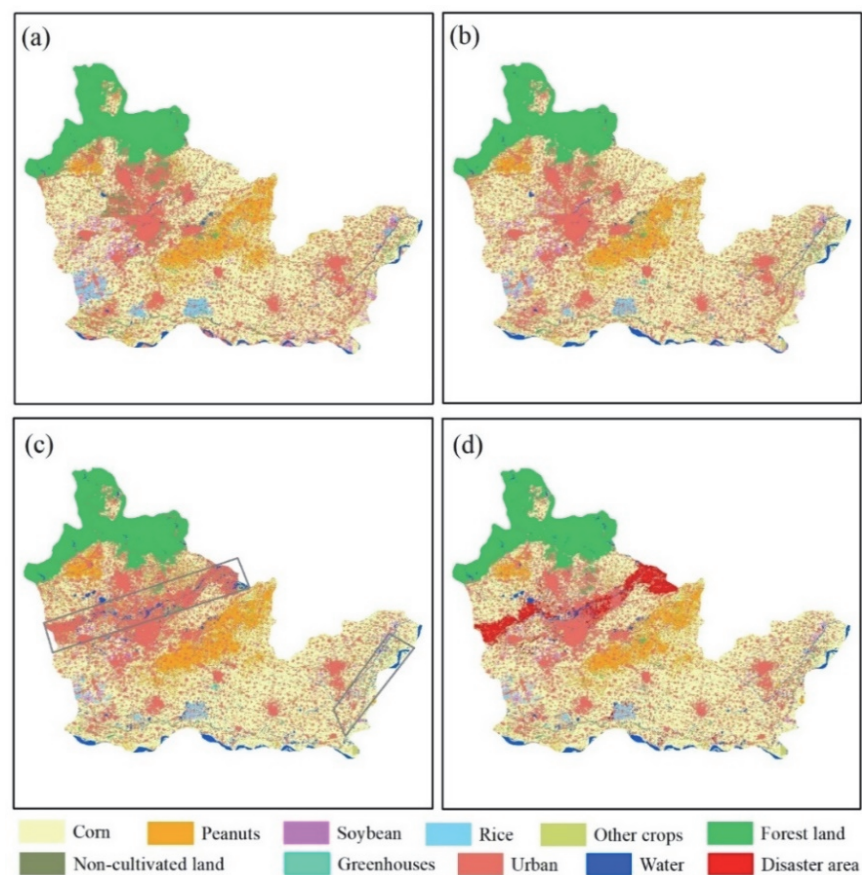


Figure 6. Segmentation and classification results of scheme 2 based on UNet++ model in 2019, 2020, and 2021. (a) Classification result in 2019; (b) classification result in 2020; (c) classification result in 2021; (d) reclassification result in 2021.

Figure 6d shows that the disaster area in 2021 was mainly distributed in the central part (both banks of the Wei River) and the eastern part (along the Yellow River) of Xinxiang City, with the central part being the most severely damaged. Based on the validation sample plots and points, the PA, UA, and F_1 score values of the classification results in

2019 and 2020 and the reclassification result in 2021 were shown in Table 5. The major crops, such as corn, peanuts, soybean, and rice, showed higher classification accuracy than minor crops. Except for the F_1 score of soybean in 2021, the F_1 scores of corn, peanuts, soybean, and rice were higher than 80%. Due to the few validation samples of NCL, OTH, and GH, their classification accuracies were lower. Wang et al. [8] pointed out that the factors, such as within-class diversity, between-class similarity, and mixed pixels, remain major challenges in crop and land use classification, especially for minor crops. Therefore, it is necessary to further reduce the misclassification and omission of small categories by increasing the number of samples or adopting synthetic minority over-sampling technique methods. Since the carbon sequestration was calculated for five categories in the farmland ecosystem, the accuracies of forest land, urban areas, and water were not further analyzed in this study.

Table 5. PA, UA, and F_1 score from scheme 2 and the UNet++ model. The abbreviation NaN represents not-a-number due to the zero denominator.

Year	Indicator	Corn	Peanuts	Soybean	Rice	NCL	OTH	GH	FL	Urban	Water
2019	PA	0.94	0.95	0.71	0.97	0.21	0.28	0.00	0.54	0.76	0.57
	UA	0.89	0.95	0.90	0.95	0.43	0.10	0.00	0.80	0.76	0.31
	F_1 score	0.91	0.95	0.80	0.96	0.29	0.15	0.00	0.64	0.76	0.40
	UA	0.86	0.95	0.82	0.92	0.08	0.27	NAN	0.76	0.88	0.54
2020	PA	0.90	0.79	0.84	0.93	0.59	0.19	NAN	0.91	0.85	0.21
	F_1 score	0.88	0.86	0.83	0.93	0.14	0.22	0.00	0.83	0.87	0.31
	PA	0.98	0.93	0.83	0.83	0.40	0.15	0.00	0.70	0.76	0.39
2021	UA	0.91	0.89	0.75	0.89	0.44	0.17	NAN	0.60	0.71	0.69
	F_1 score	0.95	0.91	0.79	0.86	0.42	0.16	0.00	0.65	0.73	0.50

4.4. Estimation of Crop Area and Carbon Sequestration

Based on the classification results, Table 6 shows the estimation results of the crop area and carbon sequestration. As a major crop, the area proportion of corn changed little in recent years. The areas of soybean, rice, and NCL decreased year by year. In 2021, the disaster area of farmland accounted for about 3.48% of the total study area, and the crops were mainly corn. Except for the forest land, urban areas, and water, the proportion of farmland in 2019, 2020, and 2021 was 51.44%, 51.64%, and 51.56%, respectively. Here, the crops in the disaster area were not involved in the carbon sequestration results. For corn, peanuts, soybean, rice, and other crops, the total carbon sequestration in 2019, 2020, and 2021 was about 2460.56, 2549.16, and 1814.07 thousand tons, respectively. In 2021, the crop yield was reduced in a large area due to the precipitation disasters, resulting in a significant decrease in crop carbon sequestration.

Table 6. Estimation of crop area and carbon sequestration. Area proportion and carbon sequestration are abbreviated as AP and CS. The units of area and CS area % and thousand tons.

Year	Indicator	Corn	Peanuts	Soybean	Rice	NCL	OTH	GH	FL	Urban	Water
2019	AP	35.45	7.82	2.19	1.49	1.23	3.17	0.09	14.16	33.14	1.27
	CS	1997.85	248.13	62.98	64.19		87.41				
2020	AP	36.84	6.79	2.15	0.88	0.56	4.31	0.11	13.53	33.18	1.65
	CS	2114.99	218.43	55.54	40.44		119.76				
2021	AP	35.15	7.23	1.35	0.69	0.16	3.33	0.16	12.93	32.88	2.63
	CS	1442.01	217.09	35.69	32.90		86.38				

To verify the estimation results of crop carbon sequestration, we compared the results to several existing references and methods. The results of Wang et al. [38] showed that the carbon sequestration of corn, peanut, soybean, and rice in Xinxiang City in 2019 and 2020 were 2286.63 and 2427.81, 324.40 and 332.12, 59.00 and 43.75, and 116.22 and 76.87 thousand

tons, respectively. These estimated results were higher than this assessment owing to the differences in crop area and yield in the statistical data. In addition, Tan et al. [39] used the net carbon sink method to assess the carbon sequestration of cropland ecosystems in Henan Province. The carbon absorption results of corn, peanut, soybean, and rice in Xinxiang City in 2019 and 2020 were 1923.1 and 2041.9, 349.5 and 357.9, 53.9 and 40.0, and 100.8 and 66.6 thousand tons, respectively. The estimated results of corn and soybean were lower than those of this study, while those of peanut and rice were the opposite. Besides crop yield errors, dry weight ratios used in the method parameters also caused differences in the estimation of crop carbon sequestration. Therefore, it is necessary to further evaluate and analyze the estimation results of different models in combination with the ground survey data.

5. Discussion

5.1. Evaluation of Model Parameters and Classification Accuracy

Figure 4 and Table 3 show the influence of feature fusion and upsampling on the deep learning model performance with a 256×256 patch size. With upsampling of small samples, the UNet++ model performance of the two schemes showed little difference. However, the UNet++ performance outperformed the UNet, DeepLab V3+, and PSPNet with the same training data set and model parameters. Moreover, the segmentation and classification results based on UNet++ also proved its capability for correctly identifying multiple categories in complex agricultural areas concerning UNet, DeepLab V3+, and PSPNet. As concluded by Zhou et al. [21], Wang et al. [9], and Giannopoulos et al. [18], the major advantage of the UNet++ model is the multilevel full-resolution feature map-generating strategy for multi-feature RS imagery. In addition, Table 3 showed that the performance of the UNet model of scheme 2 is better than the DeepLab V3+ model. However, the prediction classification accuracy of UNet was not always higher than that of DeepLab V3+. Therefore, when the different deep learning models have similar performances, the prediction classification accuracy should be further evaluated.

Combined with Figure 6 and Table 5, the prediction accuracies without training data sets demonstrated the generalization of the deep learning models in ground object monitoring, especially the UNet++ model. In addition, the cosine annealing learning rate schedule with restarts could further accelerate the model convergence and improve the performance of deep learning networks. Overall, the image segmentation and classification based on UNet++ was an effective method for classifying the multiple crops referring to UNet, DeepLab V3+, and PSPNet. Due to the influence of continuous precipitation, the time-series images were difficult to collect and the OA in 2021 was lower than 85% based on single imagery. In current studies, the time-series images and recurrent neural network (RNN), such as LSTM and transformer, tended to improve the classification accuracy [40,41]. Zhong et al. [40] developed a deep learning model based on a one-dimensional convolutional network for economic crops using Landsat time-series data, with the highest OA being 85.54% and F_1 score of 0.73. Teimouri et al. [42] classified 14 major classes of crops based on a convolutional LSTM network, and the OA of 8 major crops was more than 84%. Wang et al. [41] proposed a novel architecture named the coupled CNN and transformer network and achieved 72.97% mIoU of the Barley Remote Sensing Dataset. However, the time-series images and deep learning methods would increase the labeling workload related to ground investigations and manual editing. The experimental results of this study demonstrated that the proposed methods can improve the model performance and yield state-of-the-art segmentation and classification results.

The model transferability in this study is mainly shown in the following two points. First, the focus of the current study was placed on the training data set of the four sites in Xinxiang City, and the training model can be transferred to the segmentation and classification of crops and land use to a more considerable spatial extent. Second, without a training data set and prior knowledge in 2021, the OA and macro F_1 values of prediction classification based on UNet++ model were higher than 83% and 59%, respectively, which proved the transferable learning capability of the deep learning model across the years.

Therefore, the modified UNet++ model has better spatiotemporal transfer learning ability in the study area.

5.2. Comparison of Local Results and Error Analysis

Table 5 shows that the accuracy is lower for minor categories, such as non-cultivated land, other crops, greenhouses, and water. Due to the influence of image spatial resolution, there are a large number of mixed pixels around urban and rural areas. In addition, the non-cultivated greenhouses are often scattered and small, which often makes them difficult to effectively distinguish from the field paths. Figures 5 and 7 show the main misclassified land use and crop types, including corn misclassified as rice and forest land (seedling forest in farmland), peanuts misclassified as corn and soybean, and field roads misclassified as crops. In addition, as a pixel-based classification method, the UNet++ model cannot completely avoid but can effectively alleviate the “salt and pepper phenomenon” in the classification. In addition, the representative error of samples and training data sets influences the land use and crop classification accuracy. The representative error depends on the sample size, image resolution, label production, and the difference in the entire categories over the study area [9]. The misclassification of disaster areas in 2021 was mainly caused by the representative error of the samples, especially the difference in the entire categories. This misclassification result means that the more representative sample data sets should be added soon in order to improve the generalization and transferable learning capability of deep learning models.

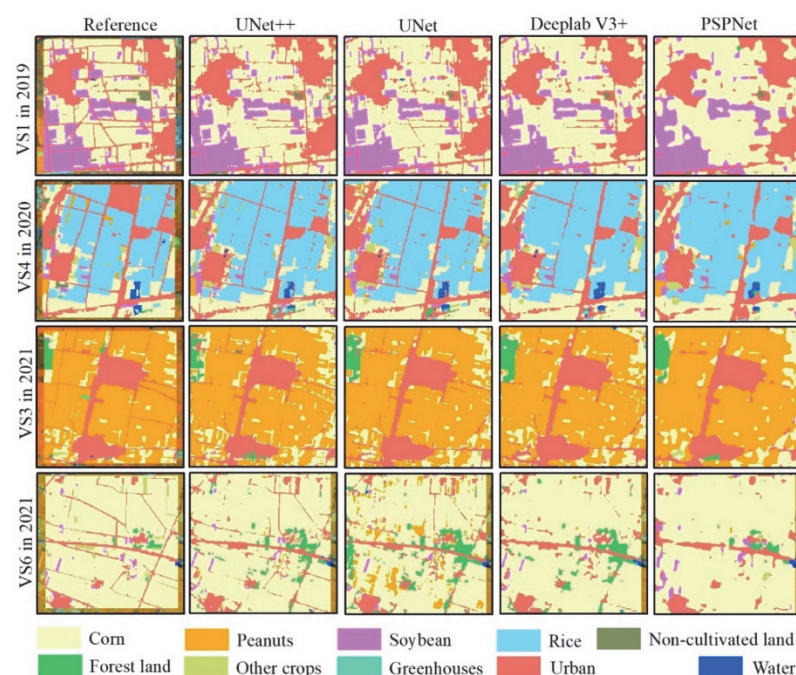


Figure 7. Validation sample plots and local-scale prediction classification results of Scheme 2 in 2019, 2020, and 2021.

The factors, such as planting patterns of small farmers, diversified crop types, and topography and climate difference terrain in large-scale regions, limit the applicability and parameters of the transferable learning models [7]. Therefore, the image classification in a large-scale region should be divided into different subregions according to Tobler’s First Law. In this way, the representativeness error of samples and the fine-tuning parameters of the model can limit the error propagation within each subregion and improve the model performance and prediction accuracy [9,12,15]. Therefore, in addition to the feature fusion, training set enhancement, and upsampling of the small samples, the geographical zones

should be involved to reduce the representative error and improve the prediction accuracy at a large scale.

5.3. Analysis of Crop Carbon Sequestration

Figure 8 shows the estimation results of crop carbon sequestration per unit area in 2019, 2020, and 2021. Corn and rice have a stronger carbon sequestration capacity per unit area compared to peanuts, soybean, and other crops. The carbon sequestration capacity of corn, peanuts, and other crops decreased due to the continuous precipitation and reduced production. In this study, we used the average yield in the statistical yearbook to estimate the results. However, it is unable to display the difference in crop carbon sequestration in spatial distribution. As it is affected by farming patterns, it is often difficult to obtain the measured yield of different crops. Combining the RS time-series vegetation indices models to obtain a more accurate crop yield can further eliminate the estimation error of carbon sequestration [4]. Moreover, the carbon sequestration model in this study does not fully consider the physiological parameters of crops or the performance of different models, leading to some errors in the estimation results. Soil carbon sequestration, as an important component of the farmland ecosystem, has not been evaluated. The farmland ecosystem is an important part of the carbon source/sink, and an increasing number of studies show that some measures should be adopted to improve the carbon sequestration capacity, such as by increasing the proportion of the tertiary industry, optimizing crop structure, and increasing green areas [22,37]. In a word, to accurately assess crop carbon sequestration and promote the process of carbon neutrality, we need to collect more detailed data and build a rigorous mechanism model simultaneously.

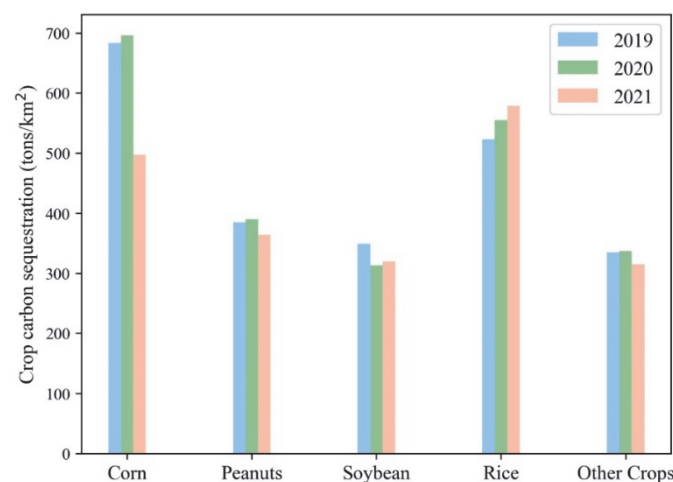


Figure 8. Estimation results of crop carbon sequestration per unit area in 2019, 2020, and 2021.

5.4. Limitations and Future Work

Many deep learning approaches based on domain knowledge and expertise have been designed to segment and classify the land use and crops from RS imagery [29,34]. The variance in geographical categories and meteorological conditions constrained the generalizability and transferability of semantic segmentation methods, especially in large-scale crop mapping applications [9]. In the classification task, many factors, such as differences in terrain, crop types and phenology, and imagery style would affect the transfer learning performance of the training model. According to Tobler's first law of geography, these limitations lead to crop segmentation and classification models that can only be applied within a certain geographical area. Wang et al. [9] noted that it was necessary to construct geographical zones to reduce the representative error on the transferable learning model of crop classification on a large scale. In addition, a satisfactory result can be achieved by using histogram matching, or histogram specification, which was utilized to adjust the characteristics from the different dates and imaging conditions. Therefore, the

actual transfer of a deep learning model to a larger spatiotemporal scale may be challenging because of different factors, such as crop types, phenological period, and upsampling of small samples.

The experimental results showed that the upsampling of small samples improved the model performance. However, the upsampling times of different categories should be further analyzed to alleviate the imbalance problem, such as the lower accuracy of greenhouses. Furthermore, feature selection was based on several studies rather than the feature sensitivity analysis or importance evaluation. The representative samples caused by extreme weather or other disasters should be increased to improve the generalization and robustness capability of the model. Moreover, the proposed model also had shortcomings on a larger scale. The prediction results of the UNet++ model with feature fusion and upsampling methods in the rice-growing area of Xinyang City (approximately 4° latitude difference from Site C) were not satisfactory. Therefore, combined with the digital elevation model and crop phenological data, it is necessary to divide the large region into different subregions with relatively consistent crop planting conditions to avoid representativeness error propagation and improve the model performance and classification accuracy.

6. Conclusions

An end-to-end transferable learning model based on the UNet++ architecture is proposed for crop classification in complex agricultural areas. First, based on the feature fusion and upsampling of small samples, the UNet++ model yields the best performance and classification accuracy when compared to UNet, DeepLab V3+, and PSPNet, with a lower joint loss value and higher mIoU value of 0.432 and 0.871, respectively. The OA and macro F_1 values of the UNet++ model based on the spatiotemporal transfer experiment are higher than 83% and 58%, respectively. Subsequently, according to the three-year time series classification results and reclassification rules, the disaster area in 2021 accounted for about 3.48% of the total area and was concentrated in the middle part of the study area. Finally, the total carbon sequestration of the five target crops in 2019, 2020, and 2021 was estimated by integrating statistical data, with the values of 2460.56, 2549.16, and 1814.07 thousand tons, respectively. These results can provide data and method support for damage assessment and carbon sequestration assessment in farmland ecosystems. The prediction classification accuracy without prior knowledge available from ground samples proves that the improved UNet++ model can provide better spatiotemporal transfer learning capability with respect to the baseline models. The transferable learning model is suitable for automatically extracting regional crop information from multi-feature imagery to produce near-real-time crop maps. The experimental conclusion can provide a method reference to RS segmentation and classification, and offer inspiration for understanding the crop carbon sequestration in farmland ecosystems.

Author Contributions: Conceptualization, L.W. and F.Q.; methodology, L.W.; investigation, Y.B. and L.W.; supervision, J.W. and F.Q.; writing—original draft, L.W.; writing—review and editing, L.W. and Z.Z.; resources, C.L.; validation, Y.B. and X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China [grant number U21A2014]; National Science and Technology Platform Construction [grant number 2005DKA32300]; Major Research Projects of the Ministry of Education [grant number 16JJD770019]; The Open Program of Collaborative Innovation Center of Geo-Information Technology for Smart Central Plains Henan Province [grant number G202006]; Key Laboratory of Geospatial Technology for Middle and Lower Yellow River Regions (Henan University), Ministry of Education [grant number GTYR202203]; China Postdoctoral Science Foundation [grant number 2018M64099]; Key Scientific Research Projects in Colleges and Universities of Henan Province [grant number 21A420001]; and 2022 Henan College Student's Innovation and Entrepreneurship Training Program [grant number 202210475034].

Data Availability Statement: The code is shared at https://github.com/AgriRS/SummerCrop_DeepLearning.

Acknowledgments: We sincerely thank the anonymous reviewers for their constructive comments and insightful suggestions that greatly improved the quality of this manuscript.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Guo, L.; Sun, X.R.; Fu, P.; Shi, T.Z.; Dang, L.N.; Chen, Y.Y.; Linderman, M.; Zhang, G.L.; Zhang, Y.; Jiang, Q.H.; et al. Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. *Geoderma* **2021**, *398*, 115118. [\[CrossRef\]](#)
- Anderegg, W.R.L.; Trugman, A.T.; Badgley, G.; Anderson, C.M.; Bartuska, A.; Ciais, P.; Cullenward, D.; Field, C.B.; Freeman, J.; Goetz, S.J.; et al. Climate-driven risks to the climate mitigation potential of forests. *Science* **2020**, *368*, eaaz7005. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chapungu, L.; Nhamo, L.; Gatti, R.C. Estimating biomass of savanna grasslands as a proxy of carbon stock using multispectral remote sensing. *Remote Sens. Appl. Soc. Environ.* **2020**, *17*, 100275. [\[CrossRef\]](#)
- Zhao, J.F.; Liu, D.S.; Cao, Y.; Zhang, L.J.; Peng, H.W.; Wang, K.L.; Xie, H.F.; Wang, C.Z. An integrated remote sensing and model approach for assessing forest carbon fluxes in China. *Sci. Total Environ.* **2022**, *811*, 152480. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li Johansson, E.; Brogaard, S.; Brodin, L. Envisioning sustainable carbon sequestration in Swedish farmland. *Environ. Sci. Policy* **2022**, *135*, 16–25. [\[CrossRef\]](#)
- Xu, J.F.; Yang, J.; Xiong, X.G.; Li, H.F.; Huang, J.F.; Ting, K.C.; Ying, Y.B.; Lin, T. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [\[CrossRef\]](#)
- Liu, X.K.; Zhai, H.; Shen, Y.L.; Lou, B.K.; Jiang, C.M.; Li, T.Q.; Hussain, S.B.; Shen, G.L. Large-scale crop mapping from multisource remote sensing images in Google Earth Engine. *IEEE J.-Stars* **2020**, *13*, 414–427. [\[CrossRef\]](#)
- Wang, L.J.; Wang, J.Y.; Qin, F. Feature fusion approach for temporal land use mapping in complex agricultural areas. *Remote Sens.* **2021**, *13*, 2517. [\[CrossRef\]](#)
- Wang, L.J.; Wang, J.Y.; Zhang, X.W.; Wang, L.G.; Qin, F. Deep segmentation and classification of complex crops using multi-feature satellite imagery. *Comput. Electron. Agric.* **2022**, *200*, 107249. [\[CrossRef\]](#)
- Camps-Valls, G.; Bioucas-Dias, J.; Crawford, M. A special issue on advances in machine learning for remote sensing and geosciences. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 5–7. [\[CrossRef\]](#)
- Watts, J.D.; Lawrence, R.L.; Miller, P.R.; Montagne, C. Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. *Remote Sens. Environ.* **2009**, *113*, 1843–1852. [\[CrossRef\]](#)
- Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
- Morid, M.A.; Borjali, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Comput. Biol. Med.* **2021**, *128*, 104115. [\[CrossRef\]](#)
- Wang, L.J.; Wang, J.Y.; Liu, Z.Z.; Zhu, J.; Qin, F. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. *Crop J.* **2022**, *10*, 1435–1451. [\[CrossRef\]](#)
- Yu, R.G.; Fu, X.Z.; Jiang, H.; Wang, C.H.; Li, X.W.; Zhao, M.K.; Ying, X.; Shen, H.Q. Remote sensing image segmentation by combining feature enhanced with fully convolutional network. *Lect. Notes Comput. Sci.* **2018**, *11301*, 406–415.
- Ma, L.; Liu, Y.; Zhang, X.L.; Ye, Y.X.; Yin, G.F.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *Isprs J. Photogramm.* **2019**, *152*, 166–177. [\[CrossRef\]](#)
- Li, J.T.; Shen, Y.L.; Yang, C. An adversarial generative network for crop classification from remote sensing timeseries images. *Remote Sens.* **2021**, *13*, 65. [\[CrossRef\]](#)
- Giannopoulos, M.; Tsagkatakis, G.; Tsakalides, P. 4D U-Nets for multi-temporal remote sensing data classification. *Remote Sens.* **2022**, *14*, 634. [\[CrossRef\]](#)
- Yang, L.B.; Huang, R.; Huang, J.F.; Lin, T.; Wang, L.M.; Mijiti, R.; Wei, P.L.; Tang, C.; Shao, J.; Li, Q.Z.; et al. Semantic segmentation based on temporal features: Learning of temporal-spatial information from time-series SAR images for paddy rice mapping. *IEEE Trans. Geosci. Remote* **2022**, *60*, 4403216. [\[CrossRef\]](#)
- Wang, C.S.; Du, P.F.; Wu, H.R.; Li, J.X.; Zhao, C.J.; Zhu, H.J. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Comput. Electron. Agric.* **2021**, *189*, 106373. [\[CrossRef\]](#)
- Zhou, Z.W.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J.M. UNet plus plus: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Dlmia 2018*; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 3–11.
- Zhang, A.X.; Deng, R.R. Spatial-temporal evolution and influencing factors of net carbon sink efficiency in Chinese cities under the background of carbon neutrality. *J. Clean. Prod.* **2022**, *365*, 132547. [\[CrossRef\]](#)
- Tang, X.L.; Zhao, X.; Bai, Y.F.; Tang, Z.Y.; Wang, W.T.; Zhao, Y.C.; Wan, H.W.; Xie, Z.Q.; Shi, X.Z.; Wu, B.F.; et al. Carbon pools in China's terrestrial ecosystems: New estimates based on an intensive field survey. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4021–4026. [\[CrossRef\]](#)

24. Pineux, N.; Lisein, J.; Swerts, G.; Biielders, C.L.; Lejeune, P.; Colinet, G.; Degre, A. Can DEM time series produced by UAV be used to quantify diffuse erosion in an agricultural watershed? *Geomorphology* **2017**, *280*, 122–136. [[CrossRef](#)]
25. Feyisa, G.L.; Palao, L.K.; Nelson, A.; Gumma, M.K.; Paliwal, A.; Win, K.T.; Nge, K.H.; Johnson, D.E. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Comput. Electron. Agric.* **2020**, *175*, 105595. [[CrossRef](#)]
26. Chen, R.; Zhang, R.Y.; Han, H.Y.; Jiang, Z.D. Is farmers' agricultural production a carbon sink or source?—Variable system boundary and household survey data. *J. Clean. Prod.* **2020**, *266*, 122108. [[CrossRef](#)]
27. Zhang, Y.L.; Song, C.H.; Hwang, T.; Novick, K.; Coulston, J.W.; Vose, J.; Dannenberg, M.P.; Hakkenberg, C.R.; Mao, J.F.; Woodcock, C.E. Land cover change-induced decline in terrestrial gross primary production over the conterminous United States from 2001 to 2016. *Agric. For. Meteorol.* **2021**, *308*, 108609. [[CrossRef](#)]
28. Wang, Y.C.; Tao, F.L.; Yin, L.C.; Chen, Y. Spatiotemporal changes in greenhouse gas emissions and soil organic carbon sequestration for major cropping systems across China and their drivers over the past two decades. *Sci. Total Environ.* **2022**, *833*, 155087. [[CrossRef](#)]
29. Chen, Z.Y.; Li, D.L.; Fan, W.T.; Guan, H.Y.; Wang, C.; Li, J. Self-attention in reconstruction bias U-Net for semantic segmentation of building rooftops in optical remote sensing images. *Remote Sens.* **2021**, *13*, 2524. [[CrossRef](#)]
30. Luo, B.H.; Yang, J.; Song, S.L.; Shi, S.; Gong, W.; Wang, A.; Du, L. Target classification of similar spatial characteristics in complex urban areas by using multispectral LiDAR. *Remote Sens.* **2022**, *14*, 238. [[CrossRef](#)]
31. Ienco, D.; Interdonato, R.; Gaetano, R.; Minh, D.H.T. Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. *Isprs J. Photogramm.* **2019**, *158*, 11–22. [[CrossRef](#)]
32. Radosavovic, I.; Johnson, J.; Xie, S.N.; Lo, W.Y.; Dollar, P. On network design spaces for visual recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 1882–1890.
33. Jung, H.; Choo, C. SGDR: A simple GPS-based disrupt-tolerant routing for vehicular networks. In Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 18–20 October 2017; pp. 1013–1016.
34. Chen, L.C.E.; Zhu, Y.K.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Volume 11211, pp. 833–851.
35. Zhao, H.S.; Shi, J.P.; Qi, X.J.; Wang, X.G.; Jia, J.Y. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
36. Shi, L.G.; Fan, S.C.; Kong, F.L.; Chen, F. Preliminary study on the carbon efficiency of main crops production in North China Plain. *Acta Agron. Sin.* **2011**, *37*, 1485–1490. [[CrossRef](#)]
37. Zhang, P.Y.; He, J.J.; Pang, B.; Lu, C.P.; Qin, M.Z.; Lu, Q.C. Temporal and spatial differences in carbon footprint in farmland ecosystem: A case study of Henan Province, China. *Chin. J. Appl. Ecol.* **2017**, *28*, 3050–3060.
38. Wang, L.; Liu, Y.Y.; Zhang, Y.H.; Dong, S.H. Spatial and temporal distribution of carbon source/sink and decomposition of influencing factors in farmland ecosystem in Henan Province. *Acta Sci. Circumstantiae* **2022**, 1–13. [[CrossRef](#)]
39. Tan, M.Q.; Cui, Y.P.; Ma, X.Z.; Liu, P.; Fan, L.; Lu, Y.Y.; Wen, W.; Chen, Z. Study on carbon sequestration estimation of cropland ecosystem in Henan Province. *J. Ecol. Rural. Environ.* **2022**, *9*, 1–14.
40. Zhong, L.H.; Hu, L.N.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
41. Wang, H.; Chen, X.Z.; Zhang, T.X.; Xu, Z.Y.; Li, J.Y. CCTNet: Coupled CNN and Transformer Network for crop segmentation of remote sensing images. *Remote Sens.* **2022**, *14*, 1956. [[CrossRef](#)]
42. Teimouri, N.; Dyrmann, M.; Jorgensen, R.N. A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sens.* **2019**, *11*, 990. [[CrossRef](#)]