International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102451

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Transferable deep learning model based on the phenological matching principle for mapping crop extent

Shuang Ge [a,b], Jinshui Zhang [a,b,c,1,*], Yaozhong Pan [a,b,c], Zhi Yang [a,b], Shuang Zhu [d]

[a] State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China
[b] Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[c] Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China
[d] Beijing Polytechnic College, Beijing 100875, China

ARTICLE INFO

Keywords:
Crop extent mapping
Deep convolutional neural networks
Model generalization
Landsat
Cropland Data Layer

ABSTRACT

Accurate and timely crop mapping is essential for global food security assessments; however, conventional crop mapping models are usually applicable to specific spatial or temporal scales, i.e. "one-time, one-place" model. Moreover, the extensive application of a trained model to other regions is challenging when sufficient ground-truth samples used for training process are unavailable. This study exploited Cropland Data Layer (CDL) and Landsat data for Arkansas, United States (US), to train a U-Net model and then extensively tested the generalization ability of the model in the Corn Belt and California in the US, and even further tested in Liaoning, China, on a transcontinental scale. Two temporal images were generated by compositing the median values of images obtained during two crop growing time windows representing the sowing (from March to May) and vigorous growth periods (from June to August). In order to ensure the consistency of the data distribution between the target areas (testing areas) and the training area, we shifted the time windows of the target areas to match that of the training area following the phenological matching principle. Then we can composite the target data (testing data) according to the time windows matched in the target areas. The results showed a satisfactory accuracy. The average optimal *Overall Accuracy* of crop mapping in all the target areas exceeded 87%. The average optimal *F1-score* of corn and rice was 0.79. Finally, we compared the generalization performance of U-Net and Random Forest (RF) classifiers. The results showed that U-Net performed better in all the target areas in the US while RF performed better in target areas where the plots were smaller. The procedure and strategy developed will facilitate the realization of high-performance and automated global model transfer.

1. Introduction

Accurate and timely global crop mapping is essential for food security assessments (Wang et al., 2019a,b). With increases in available satellite remote sensing data, the use of remote sensing data combined with machine learning algorithms has become a mainstream strategy for the large-scale crop mapping paradigm (Dong et al., 2016; Azzari and Lobell, 2017; Cai et al., 2018; Jin et al., 2019; Johnson, 2019; You and Dong, 2020). Most conventional supervised classification methods rely on subjective manual feature selection, such as vegetation index calculation and spatial characteristic extraction. However, these features are not necessarily suitable for distinguishing ground objects and may be difficult to generalize (Zhong et al., 2019). In the past decade, deep learning has evolved rapidly by big data and powerful parallel computing. The deep learning model can automatically learn various representative features from training data without manually designing features (LeCun et al., 2015). Convolutional neural networks (CNNs) are one of the most popular network architectures in the field of deep learning (Li et al., 2019). With the ability to automatically learn multilevel representative deep features, CNNs have a strong generalization performance and have been widely used for object detection and semantic segmentation tasks in remote sensing images (Ding et al., 2018; Li et al., 2019; Martins et al., 2020; Qi et al., 2020; Segal-Rozenhaimer et al., 2020; Zhong et al., 2020).

Deep learning models need to be driven by a large number of training samples (Zhong et al., 2019). The acquisition of ground-truth samples

* Corresponding author at: State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China.
E-mail address: zhangjs@bnu.edu.cn (J. Zhang).
[1] Present address: Beijing Normal University No. 19, XinJieKouWai St., HaiDian District, Beijing 100,875, China.

https://doi.org/10.1016/j.jag.2021.102451
Received 4 May 2021; Received in revised form 5 July 2021; Accepted 15 July 2021
Available online 27 July 2021

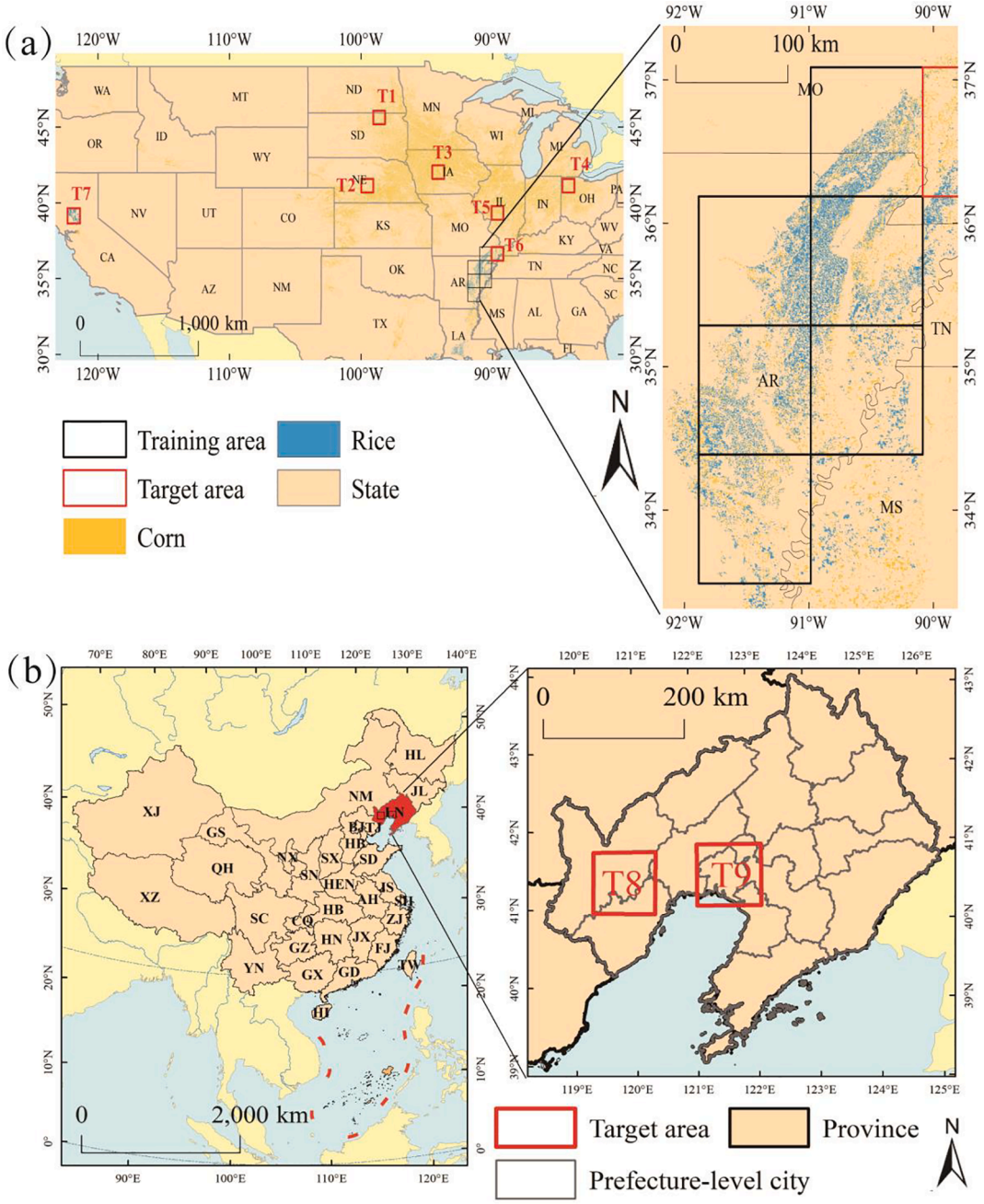

Fig. 1. Locations of the study areas. The training area is in Arkansas, US, and includes 6 grids with a size of 90 km × 90 km. The target areas consist of 9 grids with a size of 90 km × 90 km, of which T1 to T7 are located in the US (a) and T8 to T9 are located in Liaoning, China (b).

for model training is time-consuming and laborious, and some areas cannot be accessed due to various factors, including political or natural factors (Song et al., 2017; Xu et al., 2020). As such, crop classification models with robust spatiotemporal generalization performance are desired. These models can be trained in areas with sufficient and reliable samples and then applied to areas where samples are scarce or difficult to obtain. A number of studies have attempted to exploit the crop-specific land cover maps as ground-truth data for model training and spatial transfer experiments. Some studies have proved that Cropland Data Layer (CDL) was reliable as ground-truth for crop mapping (Cai et al., 2018; Wang et al., 2019a,b; Zhong et al., 2019; Hao et al., 2020). The CDL is published every year by the United States Department of Agriculture (USDA) and provides high-precision spatial distribution information for noncrop types and more than one hundred crop types, including several major crops, i.e., corn, soybean and rice. It has a spatial resolution of 30 m since 2008. The CDL is currently the most accurate national crop type map that can be accessed and downloaded freely (Cai et al., 2018). It has high overall accuracies for major crop types, especially corn and soybeans with an accuracy of over 95%, thus providing high reliability (Boryan et al., 2011).

Existing studies on model training and generalization based on CDL mostly adopted fixed-length time series data. Xu et al. (2020) applied CDL data as ground-truth to perform spatial and temporal transfer experiments in several subregions of the US Corn Belt and used a Long Short-Term Memory (LSTM)-based model and the Landsat Analysis Ready Data (ARD) time series product. The results showed that the model had a certain spatiotemporal transferable ability but presented a significant drop in performance in the case of missing data; moreover, the proposed method does not consider changes in crop growth periods in different regions. These two limitations in their study could hinder the

application of the proposed model in large-scale regions, where the crop growth period changes significantly due to various factors, including differences in latitude, and sufficient time series data may be difficult to obtain for all areas.

The critical periods of crop growth are valuable for crop identification. The characteristics of one crop in these critical periods are specific and significantly different from those of other crops (Xin et al., 2002; Dong et al., 2015; Hao et al., 2018), which represents the basis of many remote sensing efforts for crop classification. For example, previous studies found that paddy rice can be efficiently identified during the irrigation period (Dong et al., 2016). You and Dong (2020) combined Sentinel-1 and Sentinel-2 data to explore the earliest recognizable date of crops and found that corn and soybeans can be distinguished at the heading/flowering stage. Kirkegaard et al. (2018) found that the flowering or anthesis period was a key period for yield information of oil seed rape. Even if the same crop is distributed in different areas around the world, its phenological period will remain stable and measurable (Hao et al., 2020), which lays the foundation for determining the key period of remote sensing images for crop classification. It is worth noting that these critical periods do not necessarily have the same day of year (DOY) across different areas. In large-scale areas, due to differences in crop varieties, topographic features, soil conditions, climatic conditions, and farming practices, growth patterns could vary for the same crop and could be similar for different crops. Therefore, crop confusion easily occurs when crop models are transferred in large-scale areas (Lobell and Azzari, 2017).

There are currently two methods of solving the above problem. The first focuses on modifying the classifier to improve the transfer performance of a model in local areas (Diakogiannis et al., 2020; Xu et al., 2020; Zhang et al., 2020). However, according to No-Free-Lunch Theorem, finding a universal classifier is difficult (Ho and Pepyne, 2002) because the distribution of features between the training area and target area is rarely consistent. These modified classifiers are usually targeted at a specific region or specific date (Zhong et al., 2014; Dong et al., 2016), and their performances in other regions of the world have not been well evaluated (Zhang et al., 2019a,b,c). The second is to use training samples in multiple years (Konduri et al., 2020) or increase the training sample volume in the target areas (Hamrouni et al., 2021). However, samples from different years or different regions are often difficult to be obtained.

It is worth noting that few studies have focused on the problems that the generalization ability of crop classification models is limited by the inconsistency between the target data and the training data caused by changes in crop phenology. In this study, the existing methods to improve the model generalization ability by modifying classifiers or increasing training sample volume were abandoned. Instead, we fundamentally solved the inconsistency between the target data and the training data due to changes in crop phenology by adjusting the distribution of target data, so as to improve the generalization ability of the model in a novel way. The proposed method is conducive to the globalization of crop classification models. Two objectives of this study are as follows.

1. Propose an image selection strategy to train a deep learning classification model with CDL data from the US. The model was then applied to the Corn Belt and California in the US and Liaoning in China, and the applicability of the model was analyzed.
2. Analyze the impacts of the time span of image selection on the model's transferability performance and develop a principle for image determination in the model application.

2. Data and methods

2.1. Study areas

In this study, we focused on corn and rice, which are commonly planted in China and the US. We selected Arkansas, US, where both corn and rice are grown, as the training area. Some other areas where corn and rice are grown were selected as test areas, including the Corn Belt and California in the US and Liaoning in China. We tested the generalization performance of the model in these regions with completely different climate characteristics (Fig. 1). Table 1 shows the sowing and harvesting dates of crops in these areas.

Table 1
Crop phenological information in different study areas.

Type	Region	Dominated crop types	Selected crop type	Planting date	Harvest date
Training area	Arkansas	Rice, soybeans, corn and cotton	Rice	Last week of March to early June	Mid-August to the end of October or early November
			Corn	Early March to the end of March	August and September
Target area	Corn belt	Soybeans and corn	Corn	Start in late April to May	September to early December
	The Sacramento Valley	Rice, almonds, walnuts, plums and peaches	Rice	Around the beginning of May	Late September to late November
	Liaoning	Corn and rice	Corn	April to May	Late September to early October
			Rice	April to early June	October

Training area: Arkansas is located in the south-central US and has a humid subtropical climate. As the largest rice-producing state, rice production in Arkansas accounts for more than 40% of the total rice production in the US (Carroll et al., 2020). Most of the rice is planted through direct seeds and delayed floods. In 2016, Arkansas' corn planted area was 760,000 acres.

Target area 1: The Corn Belt is located in the midwestern US and represents the main area for the production of corn and soybean (Zhang et al., 2019a,b,c). This region has a humid continental climate with hot summers and cold winters. In this area, rain is the main water source for upland crops, such as corn and soybeans, planted in a single season with a corn-soybean rotation pattern. The rest of the land cover is dominated by nonarable land.

Target area 2: California is the second largest rice-producing state in the US, and most rice in this region is grown in the Sacramento Valley (Wang et al., 2019a,b). The climate in this area is Mediterranean and characterized by low precipitation, low humidity and long sunshine duration. Due to the lack of precipitation during the crop growth period, manual irrigation is required to ensure normal crop growth. Rice planting in this area involves sowing pregerminated rice seeds into still water by plane (Brodt et al., 2014).

Target area 3: Liaoning Province, which is located in Northeast China, is one of the country's most important corn bases. The corn and rice acreages are the first and second largest in the province, respectively. The region is in the transition zone between semihumid and semiarid climates. Corn in this area is irrigated by rainwater and planted in a single season. The rice is artificially irrigated starting in the sowing period.

The model trained using samples from Arkansas will be transferred to the Corn Belt and California in the US and Liaoning in China for corn and rice mapping to test its spatial generalization performance. Six grids with a size of 90 × 90 km covering rice- and corn-growing areas in

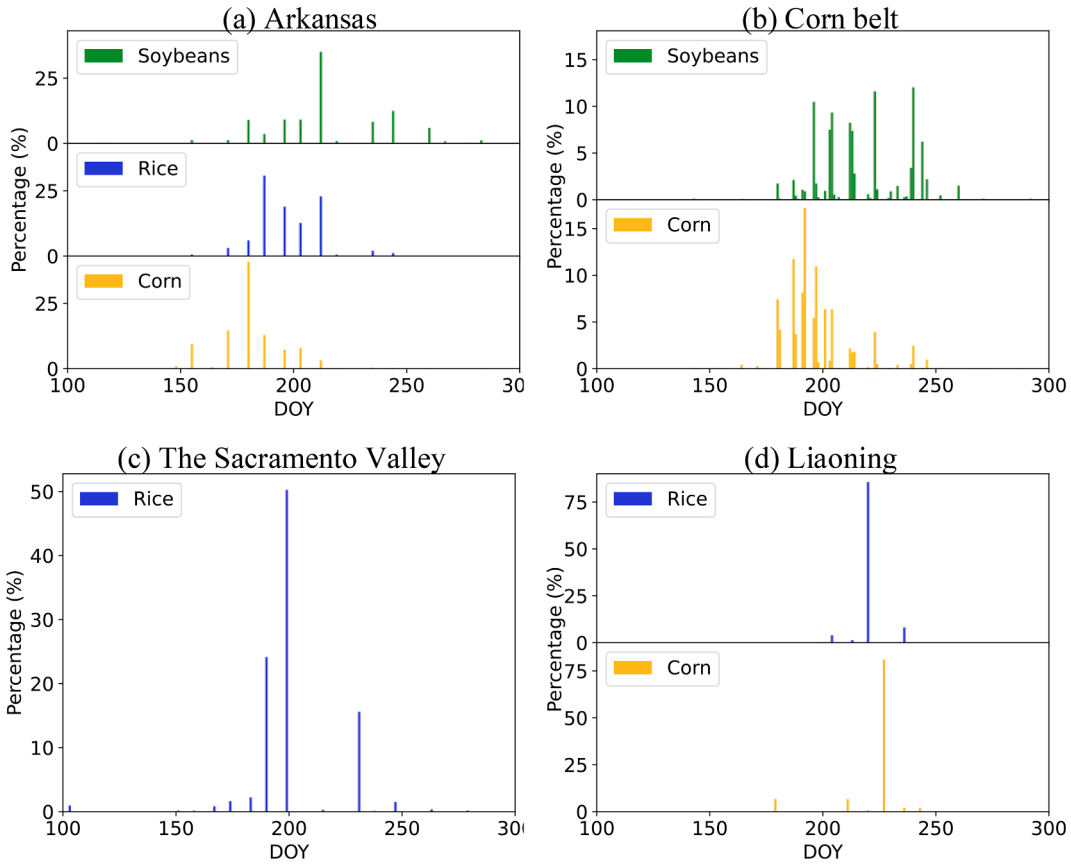

Fig. 2. Percentage of crop pixels that reached the maximum NDVI value on a certain DOY in the study areas. Colored bars indicate representative crop types of corn, rice and soybeans.

Arkansas were selected for training data organization. Nine 90 × 90 km grids in the US and Liaoning marked as T1 to T9 were selected for testing, with T1 to T6 located in the Corn Belt, T7 located in the rice planting area in California, and T8 and T9 located in Liaoning.

2.2. Data

In this study, data preprocessing was mainly performed using the Google Earth Engine (GEE) (Gorelick et al., 2017). A number of studies have proved that CDL was reliable as ground-truth for crop mapping (Cai et al., 2018; Hao et al., 2020; Xu et al., 2020). Therefore, we used the CDL integrated in the GEE for 2018 to label the training data. Classes other than corn and rice were labeled as *others*. To produce a similar detailed crop map as the CDL, the Landsat 8 Top of Atmosphere (TOA) dataset with a spatial resolution of 30 m, which was the same as the CDL dataset, was used. We chose the TOA for two reasons: atmospheric correction is not a prerequisite for remote sensing image classification (Song et al., 2001; d'Andrimont et al., 2020) and the Surface Reflectance (SR) dataset processed using the GEE includes less data compared with the TOA dataset (Teluguntla et al., 2018) and thus would result in incomplete coverage of the study area.

The removal of clouds and associated shadows is a necessary step for data preprocessing. According to the quality assessment band of Landsat data, pixels with high confidence for snow, cirrus clouds, cloud shadows and clouds have been removed (Roy et al., 2014). Six bands commonly used in crop mapping, namely, Blue, Green, Red, Near Infrared (NIR), Shortwave Infrared 1 (SWIR-1) and Shortwave Infrared 2 (SWIR-2) were selected for crop mapping (Cai et al., 2018; Ajadi et al., 2021).

We used two images composited during two time windows representing the sowing period and the vigorous growth period for crop mapping. The composition of clear-sky images per period was conducted following the principle of the median image composition, which has been widely used in large-scale crop mapping (Jin et al., 2019; Johnson, 2019; You and Dong, 2020). This study considered the sowing dates of corn and rice in Arkansas and composited the images on these dates using the median composition process to obtain the sowing period image. An analysis of the crop phenology showed that a time window of 64 days could completely cover the sowing period in the representative study areas (Johnson, 2019). Hence, the time window selected for the sowing period was DOY 80–144 in Arkansas.

In this paper, we assumed that the images obtained in the training and target areas presented consistent phenological stage so that the model could be effectively applied to target areas. We combined the CDL data in T1-T7 and the ground survey data in T8-T9 to obtain the DOY when Normalized Difference Vegetation Index (NDVI) for the target pixels of corn and rice reached the maximum value (Fig. 2). Since vigorous growth tends to last for a period, the 64-day period after the maximum number of corn pixels in Arkansas reached the highest NDVI value (DOY 180) was considered the period of growth in full bloom (DOY 180–244). Rice and soybeans in Arkansas also reached the vigorous growth period within this time period (Fig. 2). In this time period, the vigorous growth period image was composited through median composition. Finally, we used a total of 12 bands in the two images (the sowing period image and the vigorous growth period image) as model input.

2.3. Methods

To address the problems regarding phenological variation across large areas faced by model transfer in different regions, this study proposed a strategy for filtering images in different target regions and explored the feasibility of extensively generalizing the classification

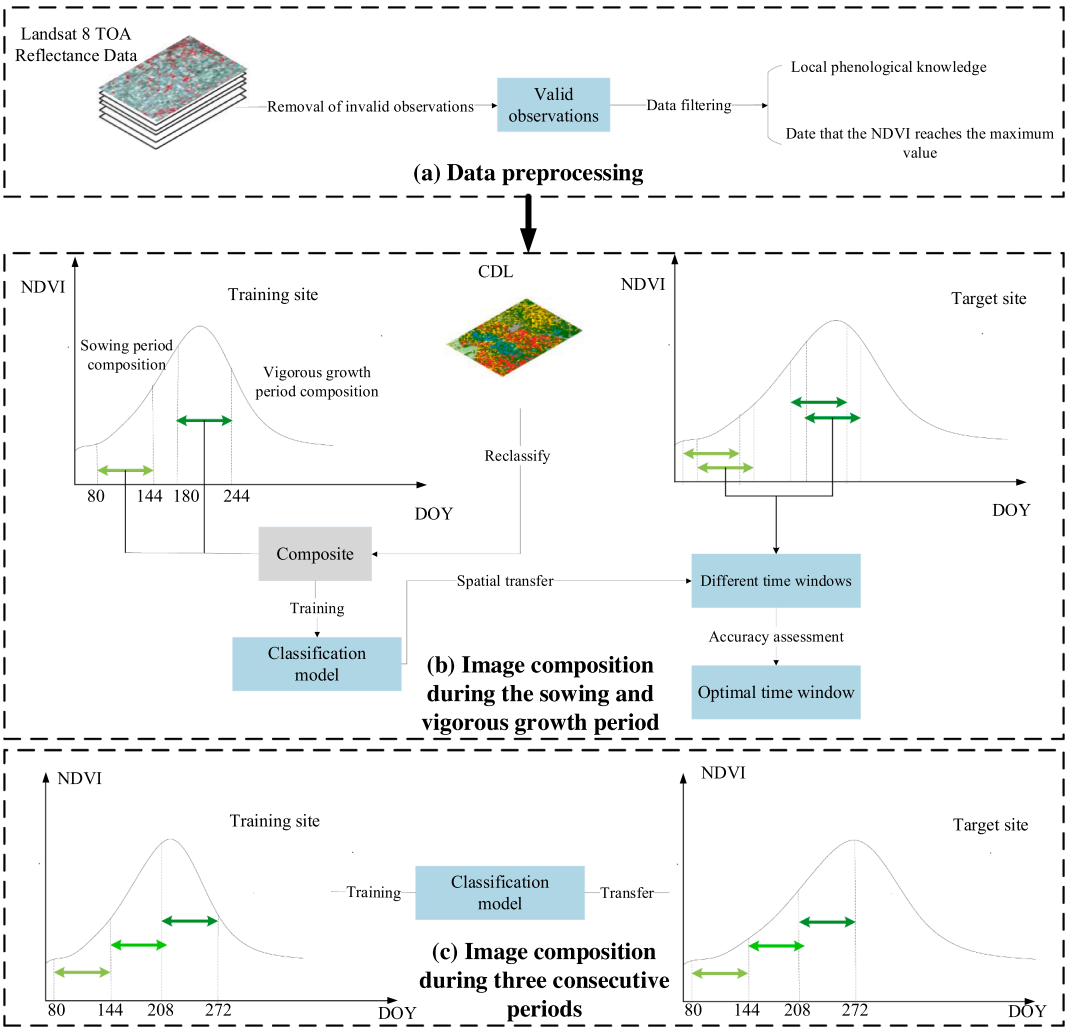

Fig. 3. Flow chart of this study includes the following: (a) data processing; (b) time window selection during the sowing and vigorous growth periods; and (c) image composition process during three consecutive periods.

model to other regions. For each crop, we extracted images for the target area and the training area so that the crop in the images were in the same critical growing period, which can address the spectral variability of a crop in different regions.

In this study, CDL data was used as ground-truth label for the two Landsat images composited in the sowing and vigorous growth periods in Arkansas and for model training. The model was then transferred to several target areas (T1-T9). In each target area, to analyze the impact of different time window selections during the sowing period and vigorous growth period on the performance of model transfer, the time windows during these two periods slid during the crop growth period. Images in target areas were generated by compositing the median values of images obtained during these time windows. The accuracy assessment was carried out by comparing the CDL in T1-T7 and the ground-truth data in T8-T9 with the model transfer results. To analyze the impact of phenological variations on model transfer performance in different regions, we used three images composited during three consecutive periods, which basically covered the growth period of the crops, for model training in the training area and then used images on the corresponding dates for model transfer in the target areas. Finally, the Random Forest (RF) classifier as a conventional classifier was compared with the U-Net model. Fig. 3 shows the flowchart in this study to briefly demonstrate the process.

Table 2
Time window information for image composition in the study areas. The training data were composited through a fixed time window. In the target areas, 3 and 6 time windows were obtained by shifting the time windows during the sowing and vigorous growth periods forward and backward with a step of 16 days, respectively.

Type	Region	Selected crop type	First phase (DOY)	Second phase (DOY)
Training area	Arkansas	Rice and corn	80–144	180–244
Target area	Corn belt	Corn	64–128	164–228
				180–244
	Sacramento Valley	Rice	80–144	196–260
				212–276
	Liaoning	Rice and corn	96–160	228–292
				244–308

2.3.1. Time window filtering for target regions

We analyzed the impact of changes in time windows during the sowing and vigorous growth periods in target areas on the performance of model transfer. Considering that the temporal resolution of Landsat data is 16 days, we shifted the time windows during the two periods forward and backward with a step of 16 days based on the time windows for the training area (DOY 80–144 and DOY 180–244) during the crop growing periods in target areas (Fig. 3(b)). A total of 3 time windows

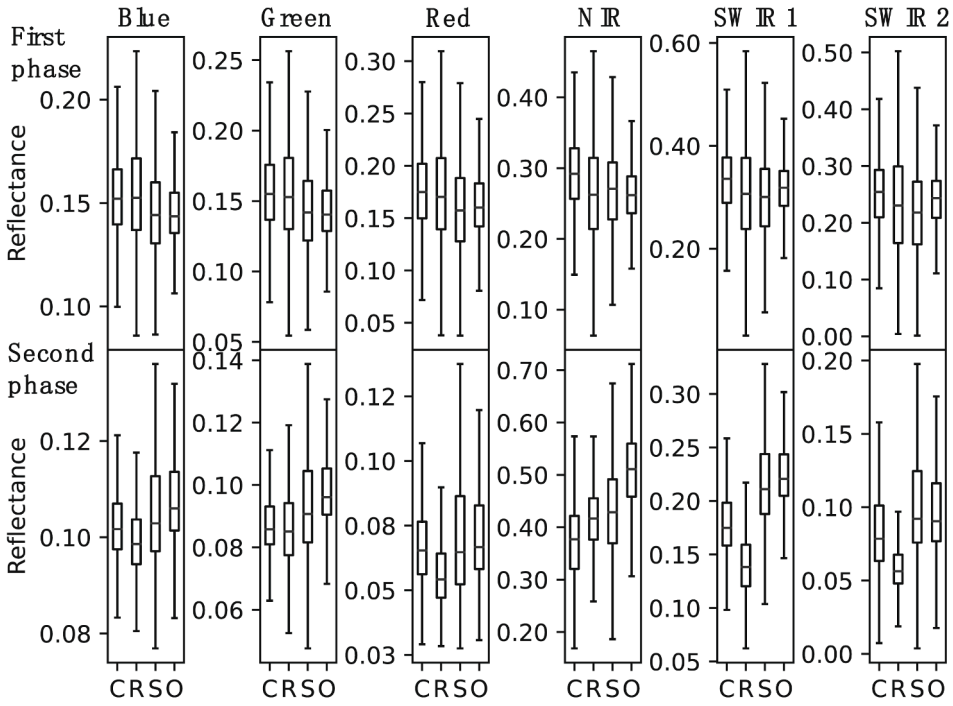

Fig. 4. Distribution of spectral reflectance in each band of Landsat data composited in the training area during the sowing period (first phase) and the vigorous growth period (second phase). C, R, S and O represent corn, rice, soybeans and cotton, respectively.

during the sowing period (DOY 64–128, DOY 80–144 and DOY 96–160) and 6 time windows during the vigorous growth period (DOY 164–228, DOY 180–244, DOY 196–260, DOY 212–276, DOY 228–292 and DOY 244–308) were obtained. A total of 18 sowing-vigorous growth combinations can be generated to represent the phenological variations in target areas. See Table 2 for the specific time windows for image selection in the training and target regions.

In addition, the transfer performance of a model was analyzed when the training data and target data were composited on the same date to analyze the impact of phenological variations on model transfer performance in different regions. We used three images composited during three consecutive periods to retrain the model (Fig. 3(c)). These periods started on DOY 80 and ended on DOY 272, and the time windows were 64 days. Hence, three time windows were obtained: DOY 80–144, DOY 144–208 and DOY 208–272. Images in each time window were reduced to a single image following the principle of median composition. Hence, a total of three images were obtained. These images in Arkansas were used to retrain the model. The transfer performance of the retained model was evaluated for T1-T9 using images composited on the same date.

2.3.2. *Model building*

This study adopted the CNNs model for crop mapping. CNNs combine convolutional layers and pooling layers to directly extract features from the original image. It can process input data at multiple context levels and have strong generalization capabilities compared with conventional machine learning algorithms (Zhu et al., 2017; Zhang et al., 2020). Specifically, the CNN used in this study was based on the U-Net architecture, which was originally used for biomedical image segmentation. We chose U-Net because it could learn from very little data and achieve state-of-the-art results on the semantic segmentation benchmark data set (Ronneberger et al., 2015), which could help solve the problem faced by large-scale crop mapping in terms of the difficulty obtaining sufficient ground samples. U-Net has been widely used in remote sensing classification tasks (Wieland et al., 2019; Zhang et al., 2019a,b,c; Pan et al., 2020; Wei et al., 2021). The U-Net architecture consists of two parts: encoder and decoder (Fig. S1). The encoder has four submodules, each of which contains two convolutional layers. After each submodule, a downsampling layer is implemented by max pool. The decoder contains four submodules, and the resolution of the feature is sequentially increased by upsampling until it is consistent with the resolution of the input image. The network also uses a skip connection to connect the upsampling result with the output of the submodule with the same resolution in the encoder as the input of the next submodule in the decoder.

We coupled CDL and Landsat 8 data from 2018 to create a training and validation dataset. In Arkansas, we covered the areas where most rice and corn were planted through 6 grids with a size of 90 × 90 km. Then, the data were cropped in the grids into tiles of 256 × 256 pixels in size with a 25% overlap to reduce the memory requirement during the training process (Wieland et al., 2019). Finally, we divided the processed data into training (80%) and validation data (20%), which had 2864 and 713 samples, respectively.

During the model training process, we used Adam to optimize the update weights (Zhang et al., 2020) and set the initial learning rate to 0.00001 and batch size to 10. We trained the model until it convergences. ImageNet pretrained model parameters were used to speed up training. We expanded the convolution filters in the first layer of U-Net to make them have more than 3 channels. The RGB channels of these filters will be initialized with pre-trained weights, while the new channels will start from scratch and need to be learned.

As one of the most popular machine learning algorithms, RF has been widely used in applications related to remote sensing image classification (Teluguntla et al., 2018; Wang et al., 2019a,b; Wieland et al., 2019; You et al., 2021). It was adopted for comparison with U-Net in this study. We used the grid-search method based on 5-fold cross-validation using training data to find the optimal hyperparameters of the RF classifier from the given parameter set (Hamrouni et al., 2021), including the number of trees (n_estimator), the maximum depth of the tree (max_depth) and the maximum number of features considered during division (max_features).

Table 3
Composition time windows of the image in the target area when the model trained in Arkansas achieved optimal performance.

Region	Selected crop type	Optimal time window for the best *Overall Accuracy*	
		First phase (DOY)	Second phase (DOY)
T1	Corn	96–160	164–228
T2	Corn	96–160	196–260
T3	Corn	96–160	196–260
T4	Corn	96–160	196–260
T5	Corn	96–160	180–244
T6	Corn Rice	64–128	180–244
T7	Rice	64–128	212–276
T8	Corn	80–144	180–244
T9	Rice	96–160	212–276

2.3.3. Accuracy assessment

A model trained and validated in Arkansas was used to classify the data in each target area (T1-T9) for 18 combinations. *Overall Accuracy, Producer's Accuracy, User's Accuracy and F1-score (F1)* were used to evaluate the accuracy of the model transfer results for each area.

$$Overall\ \ Accuracy = \frac{\sum_{i=1}^{k} X_{ii}}{N} \tag{1}$$

$$Producer's\ \ Accuracy\ \ (PA) = \frac{X_{ii}}{N_i} \tag{2}$$

$$User's\ \ Accuracy\ \ (UA) = \frac{X_{ii}}{n_i} \tag{3}$$

$$F1 = 2 \times \frac{PA \times UA}{PA + UA} \tag{4}$$

where X_{ii} represents the number of pixels of class i that matches the spatial position in the model transfer result and the ground-truth data. k represents the total number of class. N represents the total number of pixels in ground-truth data. N_i represents the total number of pixels of class i in ground-truth data.n_i represents the total number of pixels of class i in the result by the model.

3. Results

3.1. Characteristics of training data

The reflectance of all training data is shown in Fig. 4. As rice was sown with delayed flooding in Arkansas, its reflectance during the sowing period was similar to that of corn, soybeans and cotton for all bands. However, the features in this period were helpful for distinguishing fallow land from other vegetation types. In the vigorous growth period image composited on DOY 180–244, corn, rice and soybeans showed diverse reflectance in the Shortwave Infrared (SWIR) band. The spectral characteristics of cotton and soybeans are similar. This finding indicated that the period of vigorous growth was a key phase for distinguishing corn and rice. Diagnostic features for these three crops could be extracted from the SWIR band in this period.

3.2. Impacts of time window variations on model transferability

The model trained in Arkansas was used to infer the data composited by 18 different time window combinations in individual target areas. The *Overall Accuracy, Producer's Accuracy, User's Accuracy* and *F1-score* analysis of the classification results is shown in Table A1 and Figs. S2–S5. In each target area, there was a set of determined time windows during the sowing and vigorous growth periods that made the model achieve the best performance. The average optimal *Overall Accuracy* was 87.64%, of which T8 contributed the lowest of 74.13%. Other target areas except T8 were higher than 82%, and T7 had the highest of 96.32%. The average optimal *F1-score* of corn and rice was 0.79, of which T8 contributed the lowest of 0.54. Other target areas

Table 4
Detailed accuracy indices of the U-Net and RF results: *Overall Accuracy* (OA), *Producer's Accuracy* (PA), *User's Accuracy* (UA) and *F1-score* (F1). "U-Net2" and "RF" were trained by the same dataset and tested by the images composited in the two optimal time windows in each target area. "U-Net3" was trained and tested by three images obtained during three consecutive periods on the same date in different regions.

Classifier			Corn			Rice			Other		
	Region	OA (%)	F1	PA (%)	UA (%)	F1	PA (%)	UA (%)	F1	PA (%)	UA (%)
U-Net2	T1	84.22	0.79	81.81	76.39				0.90	96.31	85.24
	T2	92.67	0.80	74.85	85.67				0.96	96.97	94.11
	T3	82.84	0.78	72.78	83.65				0.86	89.95	82.39
	T4	89.73	0.80	79.52	79.73				0.93	93.12	93.24
	T5	89.70	0.83	77.61	88.77				0.93	95.38	90.08
	T6	93.67	0.75	71.45	79.64	0.85	87.13	83.65	0.96	97.02	95.87
	T7	96.32				0.92	90.49	94.22	0.98	98.19	96.99
	T8	74.13	0.53	87.83	38.31				0.82	71.36	96.58
	T9	85.54				0.85	93.89	78.42	0.89	82.08	97.23
U-Net3	T1	78.00	0.04	1.99	81.17				0.93	99.68	87.43
	T2	80.56	0.02	0.83	87.13				0.91	99.80	83.54
	T3	59.88	0.09	4.58	76.68				0.88	98.93	79.84
	T4	77.72	0.22	12.71	86.76				0.91	99.31	84.13
	T5	79.07	0.56	40.82	87.36				0.93	97.02	89.55
	T6	94.60	0.92	89.18	95.31	0.97	97.74	96.31	0.96	96.70	95.18
	T7	94.31				0.91	91.33	90.88	0.95	95.03	95.25
	T8	83.15	0.58	88.61	43.21				0.86	77.73	97.13
	T9	80.05				0.84	94.87	75.94	0.77	73.91	79.34
RF	T1	80.32	0.22	12.34	92.45				0.89	99.71	79.95
	T2	91.35	0.72	57.63	96.50				0.95	99.49	90.68
	T3	61.94	0.17	9.11	89.42				0.75	99.23	60.73
	T4	86.50	0.66	50.89	91.91				0.92	98.33	85.81
	T5	76.95	0.45	29.46	94.85				0.85	99.24	74.99
	T6	90.07	0.43	28.30	89.44	0.80	78.01	81.00	0.95	99.12	90.52
	T7	92.30				0.82	71.92	95.33	0.95	98.84	91.64
	T8	85.13	0.79	71.47	87.79				0.91	93.56	89.10
	T9	86.01				0.86	92.34	80.75	0.90	83.38	96.65

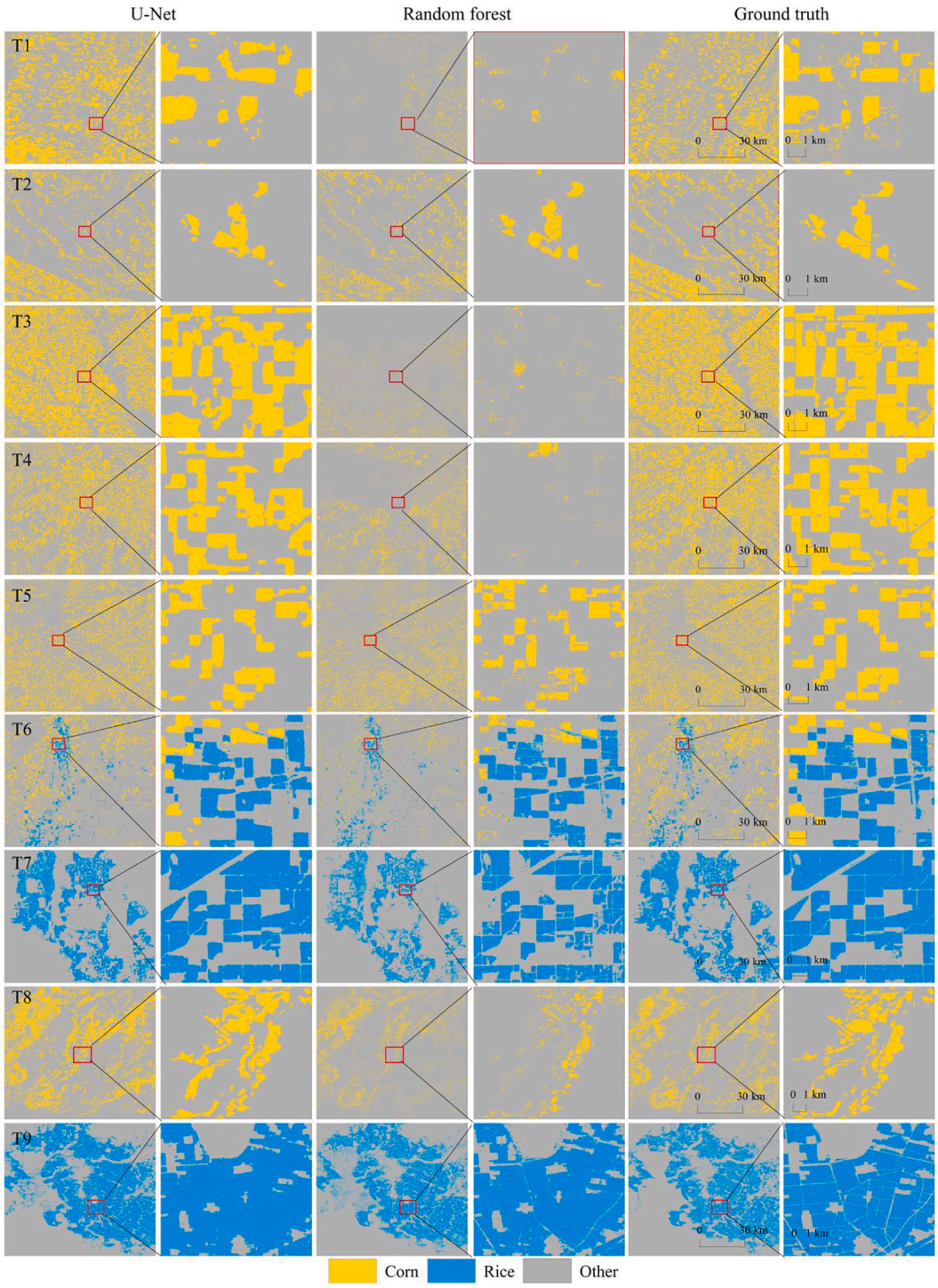

Fig. 5. Comparison between the U-Net and RF transfer results at different scales. These two classifiers were tested by the data composited by the optimal time windows in target areas. The ground-truth labels are shown on the right.

except T8 were higher than 0.76, and T7 had the highest of 0.92. At T4, when the time window during the sowing period was DOY 64–128, the composited image only covered a small portion of the corn planting area, resulting in low *Overall Accuracy* and *F1-score*.

Since the *Overall Accuracy* directly indicated the ratio of correctly classified pixels to the total number of pixels, the time windows with the best *Overall Accuracy* in each target area were determined as the optimal windows. The optimal time windows for T1-T9 are shown in Table 3. With the optimal time windows, the classification results for each region are shown in Fig. 5. The spatial distribution of the corresponding ground-truth labels is also shown. The classification results for most areas were visually accurate. There was a "salt-and-pepper effect" in CDL corresponding to T1, and corn showed a fragmented distribution. The use of U-Net led to fewer isolated pixels and more continuous and homogeneous crop type in each plot. This is because U-Net can comprehensively consider spatial information at different scales through sequence operations, including convolutional operations and down-sampling operations, and learn features directly from data, which can solve the above problem (Wieland et al., 2019). T8 was located in Liaoning, China, and compared with T1-T7, it had smaller plots and a

Table 5
The accuracy comparison between the crop map published by You et al. (2021) and our results.

Crop maps			Corn			Rice			Other		
	Region	OA (%)	F1	PA (%)	UA (%)	F1	PA (%)	UA (%)	F1	PA (%)	UA (%)
You et al. (2021)	T8	87.76	0.87	89.67	84.85				0.90	86.36	94.18
	T9	88.04				0.87	88.45	84.78	0.92	87.48	95.96
Our results by RF	T8	85.13	0.79	71.47	87.79				0.91	93.56	89.10
	T9	86.01				0.86	92.34	80.75	0.90	83.38	96.65

more complex landscape. The classification results of this study in T8 had a certain degree of misclassification relative to the ground-truth data. Its optimal *Overall Accuracy* only reached 74.13%.

3.3. Two vs three consecutive periods

Table 4 and Fig. S6 shows a comparison of the model performance obtained using two optimal time windows during the sowing and vigorous growth periods and three consecutive periods with strictly corresponding dates in different areas. Except for T6 and T8, the model trained with two images composited during two periods had better performance than the model trained with three images composited during three consecutive periods, and the *Overall Accuracy* of the former was 6.83% higher than the latter on average. At T3, the *Overall Accuracy* of the former was 22.96% higher than that of the latter, and the accuracy gap was the largest among all the target areas (from 82.84% to 59.88%).

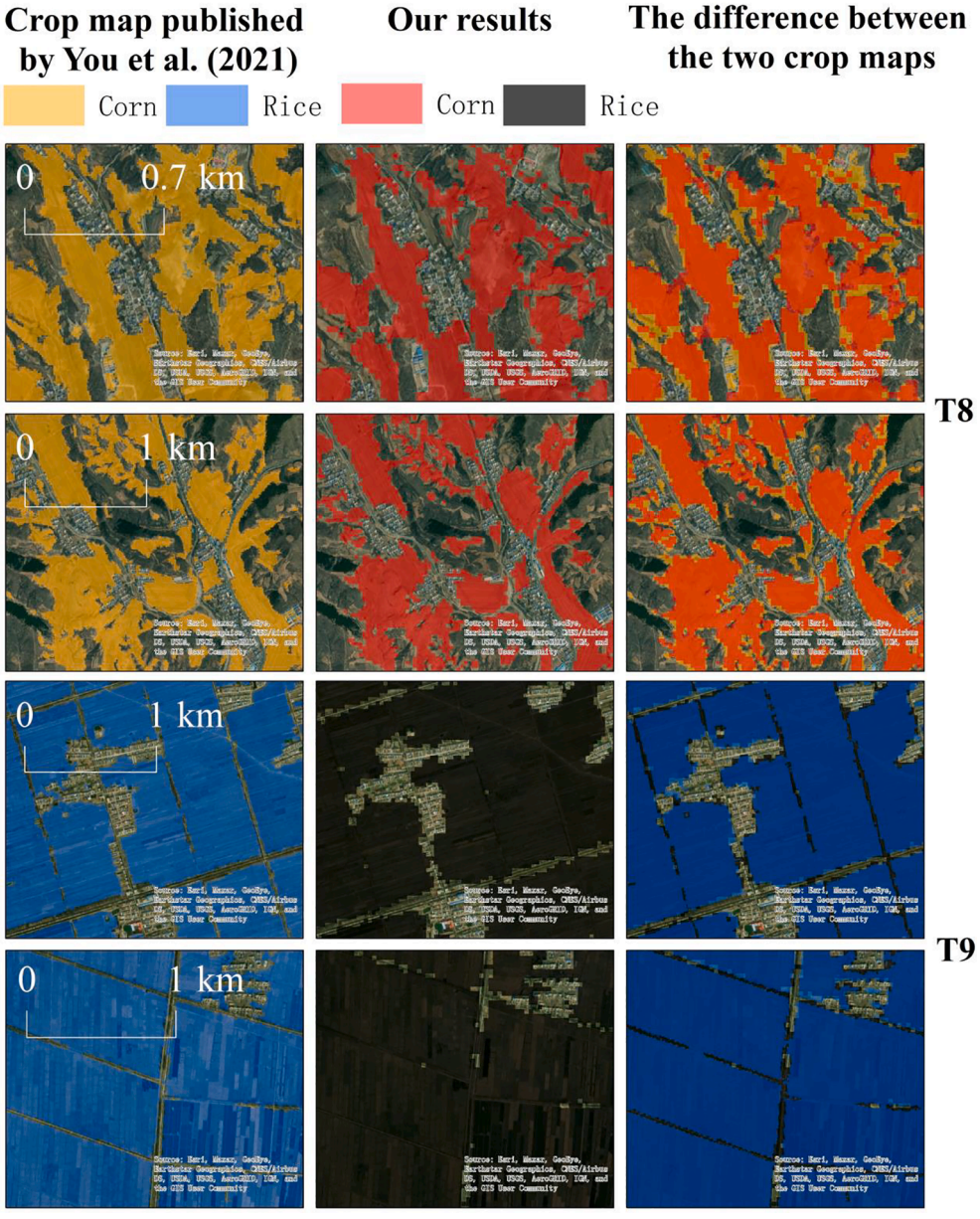

Fig. 6. Comparison between the crop map published by You et al. (2021) and our results. On the right side of the figure, we put the former on top of the latter to highlight the difference between them.

At T6 and T8, the *Overall Accuracy* of the latter reached 94.60% and 83.15%, respectively, which were improved by 0.93% and 9.02% compared with the former, respectively. Except for T6 and T8, the crops' *F1-score* of the former were also higher than that of the latter (Table 4).

This finding indicated that the crop phenology in the training area and the target area were often inconsistent; therefore, the same crops showed different spectral characteristics on the images on the same date in different regions. Since the length of the growth cycle of the same crop in different regions may be different, the use of data composited during three consecutive periods or time series data cannot make the target data and the training data equally distributed.

3.4. U-Net vs random forest

U-Net and RF are two different machine learning algorithms. Here, we attempted to verify their individual advantages (Table 4 and Fig. S7). The *Overall Accuracy* and *F1-score* by U-Net were higher than that of RF in most target areas except for T8 and T9. The RF-derived results had more isolated pixels (Fig. 5), and the *Producer's Accuracy* of corn was low (9.11–71.47%), indicating that corn had obvious omissions. RF or other classification methods based on pixels as a unit made predictions only according to the spectral feature of a single pixel, resulting in a "salt-and-pepper effect" (Diakogiannis et al., 2020). The *Producer's Accuracy* of rice varied from 71.92% to 92.34%, which was higher than that of corn, indicating that the omissions in corn were more obvious than those in rice. This also showed that by selecting the optimal time window to composite images in sowing and vigorous growth periods, rice in different regions had more similar characteristics than corn in the selected 12 bands of the two images. Therefore, the RF classifier could also achieve satisfactory transfer performance in different rice growing areas. For corn, the CNN had to learn deeper features from the original 12 spectral bands to support high-precision recognition of corn in different regions.

In T9, RF was able to better identify roads while U-Net classified some roads as rice (Fig. 5), resulting in a slightly lower *User's Accuracy* of rice for U-Net (the former was 80.75%, and the latter was 78.42%). In T8, where the parcels were relatively fragmented, the *User's Accuracy* of corn in the U-Net results dropped to 38.31%, which was the lowest among the target areas where corn was planted. This finding indicated that there were clear commissions of corn in the U-Net results in T8. It is worth noting that the plot size in T8 was significantly smaller than that in T1-T7 located in the US. In T8, the average plot size was approximately 200 m, but in T1 to T7, it was approximately 1000 m. Although the plots in T9 had an average size of 600 m, which were larger than those in T8, their shapes were obviously different from those in T1-T7. Some studies on U-Net also showed that U-Net performance was significantly decreased when the shape, size or structure of the target data were changed relative to the training data (Ibtehaz and Rahman, 2020). Therefore, although this study used the moving the time window method to ensure that the spectral values of the target data maintained the same distribution as the training data, differences in the shape and size of the objects in different target areas still existed. These differences were the main reasons for the limited generalization performance of U-Net.

4. Discussion

4.1. Comparison with current crop type maps

You et al. (2021) published the first 10-m maps of major crops, including rice, corn and soybean in Northeast China for 2017–2019. The crop maps were generated by time series Sentinel-2 data, field survey samples in Northeast China and RF classifiers. It was reported that the *Overall Accuracy* of its crop map for 2018 was 0.81. We compared its crop map for 2018 with our results by RF in T8 and T9 in Northeast China. The accuracy of its map was analyzed based on the ground-truth

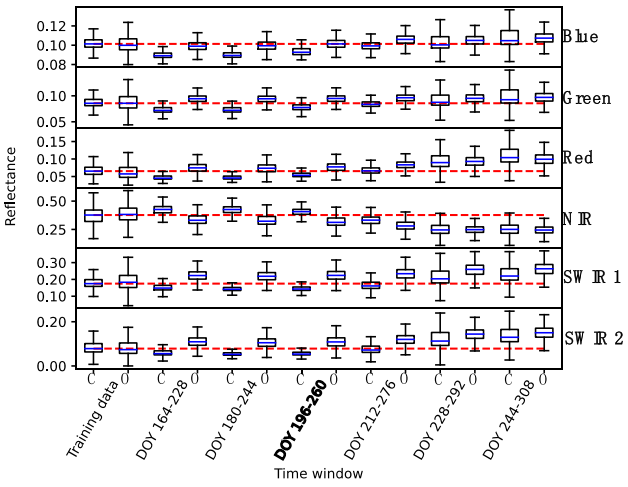

Fig. 7. Reflectance distribution of each band of the images composited by different time windows during the vigorous growth period, taking T2 as an example. The distribution of training data composited during DOY 180–244 is shown as a comparison on the left. C and O represent corn and others, respectively. The red dashed line indicates the median reflectance of corn pixels of the training data. Bold font indicates the optimal time window for T2.

in T8 and T9 in this study. As shown in Table 5, the crop map published by You et al. (2021) exhibited slightly higher *Overall Accuracies* in T8 and T9 than our results. The *Overall Accuracy* of the former was 2% higher than the latter on average. The corn's *F1-score* of the former in T8 was 0.08 higher than that of the latter (from 0.87 to 0.79), while the difference in rice was small (from 0.87 to 0.86).

The visual differences between the two crop maps are shown in Fig. 6. Our results produced similar distribution patterns with the crop map published by You et al. (2021), but obvious omissions in corn at the edge of plots in T8. This led to low *Producer's Accuracy* (71.47%) of corn in our results. T8 was located in a mountainous area and the plots in it were small. Our crop map was generated by Landsat data with a spatial resolution of 30-m, so some details at the edge of plots can't be captured. The crop map published by You et al. (2021) was generated by Sentinel-2 data with a 10-m resolution and can capture small objects such as roads in T9 (Fig. 6). In the production process of this crop map, You et al. (2021) used 21,431 samples in 2018 in Northeast China for model training and testing, and their classification process was complicated, including the process of feature selection, RF classifiers training in each agro-climate region, then identifying cropland, and identifying crops within the range of cropland. It is worth noting that we did not use samples in T8 and T9 for model training. Although the accuracies of our results were slightly lower than the former, the results also exhibited the crop distribution accurately, and our classification process was more concise by training only one model. This showed the advantages of the proposed method in this study in large-scale crop mapping.

4.2. Comparison with current crop mapping methods based on transfer learning

Some approaches based on transfer learning have recently been proposed to improve the spatial generalization ability of crop classification models. Some of these methods need to use samples in the target area to fine-tune the pre-trained model (Nowakowski et al., 2021) or use the crop statistics in the target area to correct the classifier (Kluger et al., 2021) to help the model achieve better generalization performance. However, these methods were not suitable for target areas with limited crop samples or inaccurate statistics. This study started from the perspective of adjusting the distribution of target data according to the crop phenology, so that the target data can be adapted to the classifier

Table A1

Overall Accuracy, Producer's Accuracy, User's Accuracy and F1-score (F1) analysis between the model results and the ground-truth. Two images composited during the sowing (DOY-s) and vigorous growth periods (DOY-v) were used for training and testing. Images to be tested changed along with variations in the time windows during the sowing and vigorous growth periods in target areas.

Overall Accuracy (%):

		DOY-v					
Region	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
T1	**64–128**	82.87	81.13	79.86	63.98	66.30	70.55
	80–144	82.88	81.21	80.01	64.31	66.51	70.64
	96–160	84.22	82.67	81.28	66.00	68.75	73.30
T2	**64–128**	86.78	87.85	81.11	69.34	60.75	47.54
	80–144	87.34	88.39	81.93	70.74	62.26	49.22
	96–160	90.57	92.24	92.67	89.02	84.98	81.69
T3	**64–128**	70.58	80.29	81.17	81.20	73.12	63.54
	80–144	70.84	80.65	81.65	81.65	73.35	63.93
	96–160	73.43	82.17	82.84	82.67	73.80	65.41
T4	**64–128**	45.05	46.44	46.83	46.18	37.38	31.53
	80–144	85.97	88.69	89.12	87.82	71.94	60.55
	96–160	87.07	89.43	89.73	88.37	71.73	61.75
T5	**64–128**	82.38	83.64	88.14	74.84	69.45	64.69
	80–144	84.54	85.70	88.94	74.39	69.12	64.77
	96–160	88.68	89.70	89.31	70.98	66.20	62.85
T6	**64–128**	92.90	93.67	88.27	83.28	72.22	68.50
	80–144	93.19	93.54	86.52	79.45	65.33	60.55
	96–160	93.22	93.55	86.48	79.25	64.92	59.43
T7	**64–128**	90.97	93.74	95.74	96.32	94.71	89.18
	80–144	91.07	93.78	95.69	96.22	94.63	89.20
	96–160	91.35	94.49	95.74	95.89	94.17	89.07
T8	**64–128**	73.04	73.99	70.17	68.52	61.44	60.18
	80–144	73.01	74.07	70.30	68.74	61.77	60.54
	96–160	73.95	74.13	70.49	69.76	64.54	64.18
T9	**64–128**	82.93	83.84	83.46	85.40	81.51	70.84
	80–144	82.89	83.79	83.43	85.37	81.51	70.81
	96–160	83.03	83.99	83.64	85.54	81.33	70.68

***Producer's Accuracy (%)*:**

			DOY-v					
Region	**Crop type**	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
T1	**Corn**	**64–128**	75.86	42.07	70.50	53.22	48.07	28.32
		80–144	78.90	41.97	70.43	53.19	48.04	28.19
		96–160	81.81	51.69	75.09	55.37	52.15	31.88
T2	**Corn**	**64–128**	52.06	68.95	80.70	93.09	90.43	78.11
		80–144	52.79	68.93	80.01	92.83	90.14	77.66
		96–160	57.71	67.99	74.85	89.16	85.94	72.98
T3	**Corn**	**64–128**	35.86	63.17	67.36	71.06	78.66	57.27
		80–144	36.26	64.10	68.64	72.22	80.08	58.01
		96–160	42.55	69.32	72.78	74.91	82.39	58.99
T4	**Corn**	**64–128**	20.36	31.52	37.02	40.28	43.16	41.00
		80–144	53.20	71.37	78.45	85.55	89.69	87.08
		96–160	56.89	73.27	79.52	86.76	90.15	86.73
T5	**Corn**	**64–128**	50.10	54.60	75.06	67.94	45.86	30.47
		80–144	58.63	63.42	81.16	73.05	54.05	38.80
		96–160	73.17	77.61	87.05	72.79	61.34	42.90
T6	**Corn**	**64–128**	66.74	71.45	54.03	48.72	41.09	30.20
		80–144	73.07	77.85	63.48	60.10	55.15	43.28
		96–160	74.06	78.94	64.74	61.13	56.54	44.44
	Rice	**64–128**	66.49	87.13	52.96	34.33	12.05	1.69
		80–144	69.36	87.42	55.93	39.28	18.68	3.80
		96–160	69.35	87.41	55.93	39.28	18.74	3.82
T7	**Rice**	**64–128**	82.37	88.72	90.79	90.49	81.61	57.30
		80–144	82.39	88.66	90.57	90.07	81.28	57.38
		96–160	78.53	88.13	90.02	88.56	79.22	56.82
T8	**Corn**	**64–128**	88.32	86.42	89.87	87.62	61.98	35.86
		80–144	88.36	86.34	89.78	87.61	61.87	35.69
		96–160	89.23	87.83	90.94	88.98	62.09	34.74
T9	**Rice**	**64–128**	95.80	95.95	96.36	93.99	65.92	22.35

(*continued on next page*)

Table A1 (*continued*)

***Producer's Accuracy (%)*:**								
			DOY-v					
Region	**Crop type**	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
		80–144	95.73	95.80	96.29	94.02	65.95	22.32
		96–160	95.56	95.75	96.25	93.89	65.06	21.93
User's Accuracy (%):								
			DOY-v					
Region	**Crop type**	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
T1	**Corn**	**64–128**	77.90	60.80	53.50	31.53	32.48	31.71
		80–144	77.93	61.15	53.78	31.81	32.69	31.80
		96–160	76.39	63.43	55.81	33.77	35.93	37.93
T2	**Corn**	**64–128**	72.19	68.70	50.91	38.17	31.99	23.96
		80–144	74.69	70.67	52.32	39.32	32.86	24.54
		96–160	90.30	89.60	85.67	66.15	57.65	52.10
T3	**Corn**	**64–128**	83.87	85.45	84.01	81.20	64.37	56.17
		80–144	84.42	85.52	84.12	81.37	64.31	56.59
		96–160	86.31	84.81	83.65	81.69	64.35	58.46
T4	**Corn**	**64–128**	79.26	74.28	71.38	66.02	42.30	33.86
		80–144	85.26	81.26	78.57	71.51	46.81	37.54
		96–160	86.80	82.45	79.73	72.27	46.58	38.27
T5	**Corn**	**64–128**	90.74	90.40	86.19	69.07	52.51	42.66
		80–144	89.48	88.60	83.87	66.43	51.61	44.18
		96–160	89.64	88.77	81.07	60.61	47.75	42.05
T6	**Corn**	**64–128**	81.33	79.64	60.36	42.55	21.40	15.16
		80–144	76.80	74.78	50.25	34.44	19.43	14.58
		96–160	76.38	74.30	50.14	34.28	19.49	14.39
	Rice	**64–128**	75.66	83.65	85.94	85.38	79.99	69.25
		80–144	78.34	84.47	85.35	84.51	81.70	69.63
		96–160	78.90	84.53	85.32	84.36	81.39	69.63
T7	**Rice**	**64–128**	80.88	86.02	91.68	94.22	96.11	97.06
		80–144	81.23	86.21	91.69	94.20	96.10	97.07
		96–160	84.79	89.09	92.31	94.22	96.14	97.01
T8	**Corn**	**64–128**	37.30	38.02	34.98	33.42	24.50	17.22
		80–144	37.28	38.09	35.07	33.58	24.67	17.34
		96–160	38.27	38.31	35.36	34.55	26.45	19.07
T9	**Rice**	**64–128**	72.20	73.88	73.01	77.92	84.79	82.56
		80–144	72.20	73.90	73.04	77.91	84.86	82.63
		96–160	72.59	74.39	73.52	78.42	85.13	82.67
***F1-score*:**								
			DOY-v					
Region	**Crop type**	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
T1	**Corn**	**64–128**	0.77	0.50	0.61	0.40	0.39	0.30
		80–144	0.78	0.50	0.61	0.40	0.39	0.30
		96–160	0.79	0.57	0.64	0.42	0.43	0.35
T2	**Corn**	**64–128**	0.60	0.69	0.62	0.54	0.47	0.37
		80–144	0.62	0.70	0.63	0.55	0.48	0.37
		96–160	0.70	0.77	0.80	0.76	0.69	0.61
T3	**Corn**	**64–128**	0.50	0.73	0.75	0.76	0.71	0.57
		80–144	0.51	0.73	0.76	0.77	0.71	0.57
		96–160	0.57	0.76	0.78	0.78	0.72	0.59
T4	**Corn**	**64–128**	0.32	0.44	0.49	0.50	0.43	0.37
		80–144	0.66	0.76	0.79	0.78	0.62	0.52
		96–160	0.69	0.78	0.80	0.79	0.61	0.53
T5	**Corn**	**64–128**	0.65	0.68	0.80	0.68	0.49	0.36
		80–144	0.71	0.74	0.82	0.70	0.53	0.41
		96–160	0.81	0.83	0.84	0.66	0.54	0.42
T6	**Corn**	**64–128**	0.73	0.75	0.57	0.45	0.28	0.20
		80–144	0.75	0.76	0.56	0.44	0.29	0.22
		96–160	0.75	0.77	0.57	0.44	0.29	0.22
	Rice	**64–128**	0.71	0.85	0.66	0.49	0.21	0.03
		80–144	0.74	0.86	0.68	0.54	0.30	0.07
		96–160	0.74	0.86	0.68	0.54	0.30	0.07
T7	**Rice**	**64–128**	0.82	0.87	0.91	0.92	0.88	0.72

(*continued on next page*)

Table A1 (*continued*)

***F1-score*:**								
			DOY-v					
Region	**Crop type**	**DOY-s**	**164–228**	**180–244**	**196–260**	**212–276**	**228–292**	**244–308**
		80–144	0.82	0.87	0.91	0.92	0.88	0.72
		96–160	0.82	0.89	0.91	0.91	0.87	0.72
T8	**Corn**	**64–128**	0.52	0.53	0.50	0.48	0.35	0.23
		80–144	0.52	0.53	0.50	0.49	0.35	0.23
		96–160	0.54	0.53	0.51	0.50	0.37	0.25
T9	**Rice**	**64–128**	0.82	0.83	0.83	0.85	0.74	0.35
		80–144	0.82	0.83	0.83	0.85	0.74	0.35
		96–160	0.83	0.84	0.83	0.85	0.74	0.35

without adjusting the classifier. According to Fig. 7, the optimal window kept the distribution of target data similar to the training data and allowed the categories of the target data to be distinguished. This ensure that a model achieved satisfactory transfer performance in the target area (Zhong et al., 2019). This study did not adopt the time series data that was widely used in crop mapping (Zhang et al., 2012; Cai et al., 2018; Xu et al., 2020), but used two time windows to flexibly composite data. This framework reduced the requirement for remote sensing images and ground-truth samples, which was conducive to large-scale crop mapping.

4.3. Application prospects and limitations analysis

This research was based on freely available Landsat and CDL data for deep learning model training and then performed a large-scale spatial transfer experiment within and even across continents. By moving the time windows during the sowing and vigorous growth periods in different target areas, the average optimal *Overall Accuracy* and *F1-score* of the classification results in different target areas reached 87.64% and 0.79 respectively, achieving the goal of cross-continental-scale crop mapping by training only one model. In future applications, in order to realize the global transfer of models, different data composition time windows need to be adopted in different regions. The determination of these time windows needs to consider the phenology of local target crops and non-target crops at the same time. In this study, we used Landsat 8 and CDL data with a spatial resolution of 30 m. In the future, remote sensing data with higher spatial and temporal resolutions, such as Sentinel-2, could be tested to obtain more detailed crop maps and time windows.

This study aimed to adjust the data distribution in target areas to improve the transfer performance of a trained crop classification model. Compared with continuous time series data for crop mapping, the methods proposed in this study effectively resolved the adverse effects of phenological changes in the target areas on model transfer performance (Wang et al., 2019a,b; Hao et al., 2020; Konduri et al., 2020; Xu et al., 2020). For example, although the rice planting practices in Arkansas (irrigation after sowing) and California (irrigation before sowing) were different, the image filtering strategy proposed in this study enabled the model to recognize rice in different regions with high accuracy.

RF, as a conventional pixel-based classifier, produces "salt-and-pepper effect". Although the U-Net result did not have this problem, its performance dropped significantly in areas with fragmented plots. In future research, we could try to improve the architecture of U-Net and test other networks, such as by integrating methods on high-level and low-level feature fusion (Zhang et al., 2020) to allow the model to better overcome the influence of changes in the shape, size and structure of the ground objects in different areas.

5. Conclusions

To address issues associated with conventional model transfer, this research developed a method of ensuring that the data in the target area and the training area are identically distributed by moving the image composition time window in the target areas to improve the model transfer performance from the perspective of adjusting the data distribution. In this study, we used CDL and Landsat data for Arkansas to composite crop sowing and vigorous growth period data to train the U-Net model and then transferred the model to the Corn Belt and California in the US and Liaoning in China for corn and rice mapping. The results showed that the spectral values of the same crops in different regions were not consistent on the same date and data from the corresponding date could not been applied for model transfer in large-scale areas. By moving the time windows in the sowing and the vigorous growth period in the target area, the target data and training data presented increased consistency in distribution to realize effective model transfer and achieve the purpose of a global model. The model transfer scheme proposed in this paper provides an innovative concept for rapid global crop mapping.

Funding

This work was supported by the major project of National High Resolution Earth Observation System (The Civil Part) under Grant No. 20-Y30F10-9001–20/22.

CRediT authorship contribution statement

Shuang Ge: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Jinshui Zhang:** Formal analysis, Writing - review & editing. **Yaozhong Pan:** Resources, Supervision, Writing - review & editing. **Zhi Yang:** Resources, Data curation. **Shuang Zhu:** Supervision.

Declaration of Competing 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.

Appendix A

See Table A1.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102451.

References

Ajadi, O.A., Barr, J., Liang, S.-Z., Ferreira, R., Kumpatla, S.P., Patel, R., Swatantran, A., 2021. Large-scale crop type and crop area mapping across Brazil using synthetic

aperture radar and optical imagery. Int. J. Appl. Earth Obs. Geoinf. 97, 102294. https://doi.org/10.1016/j.jag.2020.102294.
Azzari, G., Lobell, D.B., 2017. Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. Remote Sens. Environ. 202, 64–74. https://doi.org/10.1016/j.rse.2017.05.025.
Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int. 26 (5), 341–358. https://doi.org/10.1080/10106049.2011.562309.
Brodt, S., Kendall, A., Mohammadi, Y., Arslan, A., Yuan, J., Lee, I.-S., Linquist, B., 2014. Life cycle greenhouse gas emissions in California rice production. Field Crops Res. 169, 89–98. https://doi.org/10.1016/j.fcr.2014.09.007.
Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47. https://doi.org/10.1016/j.rse.2018.02.045.
Carroll, S.R., Le, K.N., Moreno-García, B., Runkle, B.R.K., 2020. Simulating Soybean-Rice Rotation and Irrigation Strategies in Arkansas, USA Using APEX. Sustainability 12 (17), 6822.
d'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., van der Velde, M., 2020. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series. Remote Sens. Environ. 239, 111660. https://doi.org/10.1016/j.rse.2020.111660.
Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote Sens. 162, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013.
Ding, P., Zhang, Y., Deng, W.-J., Jia, P., Kuijper, A., 2018. A light and faster regional convolutional neural network for object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 141, 208–218. https://doi.org/10.1016/j.isprsjprs.2018.05.005.
Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L.i., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., Moore, B., 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. Remote Sens. Environ. 160, 99–113. https://doi.org/10.1016/j.rse.2015.01.004.
Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens. Environ. 185, 142–154. https://doi.org/10.1016/j.rse.2016.02.016.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.
Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C., Sheeren, D., 2021. From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. ISPRS J. Photogramm. Remote Sens. 171, 76–100. https://doi.org/10.1016/j.isprsjprs.2020.10.018.
Hao, P., Di, L., Zhang, C., Guo, L., 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. Sci. Total Environ. 733, 138869. https://doi.org/10.1016/j.scitotenv.2020.138869.
Hao, P., Tang, H., Chen, Z., Liu, Z., 2018. Early-season crop mapping using improved artificial immune network (IAIN) and Sentinel data. PeerJ 6. https://doi.org/10.7717/peerj.5431 e5431–e5431.
Ho, Y.C., Pepyne, D.L., 2002. Simple Explanation of the No-Free-Lunch Theorem and Its Implications. J. Optim. Theory Appl. 115 (3), 549–570. https://doi.org/10.1023/A:1021251113462.
Ibtehaz, N., Rahman, M.S., 2020. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 121, 74–87.
Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. Remote Sens. Environ. 228, 115–128. https://doi.org/10.1016/j.rse.2019.04.016.
Johnson, D.M., 2019. Using the Landsat archive to map crop cover history across the United States. Remote Sens. Environ. 232, 111286. https://doi.org/10.1016/j.rse.2019.111286.
Kirkegaard, J.A., Lilley, J.M., Brill, R.D., Ware, A.H., Walela, C.K., 2018. The critical period for yield and quality determination in canola (Brassica napus L.). Field Crops Res. 222, 180–188. https://doi.org/10.1016/j.fcr.2018.03.018.
Kluger, D.M., Wang, S., Lobell, D.B., 2021. Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. Remote Sens. Environ. 262, 112488. https://doi.org/10.1016/j.rse.2021.112488.
Konduri, V.S., Kumar, J., Hargrove, W.W., Hoffman, F.M., Ganguly, A.R., 2020. Mapping crops within the growing season across the United States. Remote Sens. Environ. 251, 112048. https://doi.org/10.1016/j.rse.2020.112048.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.
Li, Z., Shen, H., Cheng, Q., Liu, Y., You, S., He, Z., 2019. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS J. Photogramm. Remote Sens. 150, 197–212. https://doi.org/10.1016/j.isprsjprs.2019.02.017.
Lobell, D.B., Azzari, G., 2017. Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. Environ. Res. Lett. 12 (1), 014014. https://doi.org/10.1088/1748-9326/aa5371.
Martins, V.S., Kaleita, A.L., Gelder, B.K., da Silveira, H.L.F., Abe, C.A., 2020. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. ISPRS J. Photogramm. Remote Sens. 168, 56–73. https://doi.org/10.1016/j.isprsjprs.2020.08.004.
Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P.P., Garcia-Herranz, M., Kim, D.-H., 2021. Crop type mapping by using transfer learning. Int. J. Appl. Earth Obs. Geoinf. 98, 102313. https://doi.org/10.1016/j.jag.2021.102313.
Pan, Z., Xu, J., Guo, Y., Hu, Y., Wang, G., 2020. Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. Remote Sens. 12 (10), 1574. https://doi.org/10.3390/rs12101574.
Qi, X., Zhu, P., Wang, Y., Zhang, L., Peng, J., Wu, M., Chen, J., Zhao, X., Zang, N., Mathiopoulos, P.T., 2020. MLRSNet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding. ISPRS J. Photogramm. Remote Sens. 169, 337–350. https://doi.org/10.1016/j.isprsjprs.2020.09.020.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.
Roy, D.P., Wulder, M.A., Loveland, T.R., C.E., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001.
Segal-Rozenhaimer, M., Li, A., Das, K., Chirayath, V., 2020. Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN). Remote Sens. Environ. 237, 111446. https://doi.org/10.1016/j.rse.2019.111446.
Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2001. Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? Remote Sens. Environ. 75 (2), 230–244. https://doi.org/10.1016/S0034-4257(00)00169-3.
Song, X.-P., Potapov, P.V., Krylov, A., King, LeeAnn, Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote Sens. Environ. 190, 383–395. https://doi.org/10.1016/j.rse.2017.01.008.
Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J. Photogramm. Remote Sens. 144, 325–340. https://doi.org/10.1016/j.isprsjprs.2018.07.017.
Wang, R., Luo, Y., Chen, H., Yuan, Y., Bingner, R.L., Denton, D., Locke, M., Zhang, M., 2019a. Environmental fate and impact assessment of thiobencarb application in California rice fields using RICEWQ. Sci. Total Environ. 664, 669–682. https://doi.org/10.1016/j.scitotenv.2019.02.003.
Wang, S., Azzari, G., Lobell, D.B., 2019b. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317. https://doi.org/10.1016/j.rse.2018.12.026.
Wei, P., Chai, D., Lin, T., Tang, C., Du, M., Huang, J., 2021. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. ISPRS J. Photogramm. Remote Sens. 174, 198–214. https://doi.org/10.1016/j.isprsjprs.2021.02.011.
Wieland, M., Li, Y., Martinis, S., 2019. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. Remote Sens. Environ. 230, 111–203. https://doi.org/10.1016/j.rse.2019.05.022.
Xin, J., Yu, Z., van Leeuwen, L., Driessen, P.M., 2002. Mapping crop key phenological stages in the North China Plain using NOAA time series images. Int. J. Appl. Earth Obs. Geoinf. 4 (2), 109–117. https://doi.org/10.1016/S0303-2434(02)00007-7.
Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H., Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. Remote Sens. Environ. 247, 111946. https://doi.org/10.1016/j.rse.2020.111946.
You, N., Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. ISPRS J. Photogramm. Remote Sens. 161, 109–123. https://doi.org/10.1016/j.isprsjprs.2020.01.001.
You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y., Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. Sci. Data 8 (1). https://doi.org/10.1038/s41597-021-00827-9.
Zhang, C., Di, L., Lin, L.i., Guo, L., 2019a. Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps. Comput. Electron. Agric. 166, 104989. https://doi.org/10.1016/j.compag.2019.104989.
Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., Chen, Q., 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. Remote Sens. Environ. 247, 111912. https://doi.org/10.1016/j.rse.2020.111912.
Zhang, E., Liu, L., Huang, L., 2019b. Automatically delineating the calving front of Jakobshavn Isbræ from multitemporal TerraSAR-X images: a deep learning approach. The Cryosphere 13 (6), 1729–1741. https://doi.org/10.5194/tc-13-1729-2019.
Zhang, L., Liu, Z., Liu, D., Xiong, Q., Yang, N., Ren, T., Zhang, C., Zhang, X., Li, S., 2019c. Crop Mapping Based on Historical Samples and New Training Samples Generation in Heilongjiang Province, China. Sustainability 11 (18), 5052. https://doi.org/10.3390/su11185052.
Zhang, M., Li, Q., Wu, B., 2012. Investigating the capability of multi-temporal Landsat images for crop identification in high farmland fragmentation regions. In: First

International Conference on Agro- Geoinformatics (Agro-Geoinformatics). https://doi.org/10.1109/Agro-Geoinformatics.2012.6311604.
Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13. https://doi.org/10.1016/j.rse.2013.08.023.
Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. https://doi.org/10.1016/j.rse.2018.11.032.
Zhong, Y., Hu, X., Luo, C., Wang, X., Zhao, J.i., Zhang, L., 2020. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. Remote Sens. Environ. 250, 112012. https://doi.org/10.1016/j.rse.2020.112012.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307.