

Article

Rice Mapping in Training Sample Shortage Regions Using a Deep Semantic Segmentation Model Trained on Pseudo-Labels

Pengliang Wei ^{1,2,3} , Ran Huang ⁴ , Tao Lin ⁵ and Jingfeng Huang ^{1,2,3,*} 

¹ Key Laboratory of Environment Remediation and Ecological Health, Ministry of Education, College of Environmental Resource Sciences, Zhejiang University, Hangzhou 310058, China; weipengliang@zju.edu.cn

² Institute of Applied Remote Sensing and Information Technology, Zhejiang University, Hangzhou 310058, China

³ Key Laboratory of Agricultural Remote Sensing and Information Systems, Hangzhou 310058, China

⁴ School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China; ran_huang@hdu.edu.cn

⁵ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; lintao1@zju.edu.cn

* Correspondence: hjf@zju.edu.cn; Tel.: +86-139-5717-1636

Abstract: A deep semantic segmentation model-based method can achieve state-of-the-art accuracy and high computational efficiency in large-scale crop mapping. However, the model cannot be widely used in actual large-scale crop mapping applications, mainly because the annotation of ground truth data for deep semantic segmentation model training is time-consuming. At the operational level, it is extremely difficult to obtain a large amount of ground reference data by photointerpretation for the model training. Consequently, in order to solve this problem, this study introduces a workflow that aims to extract rice distribution information in training sample shortage regions, using a deep semantic segmentation model (i.e., U-Net) trained on pseudo-labels. Based on the time series Sentinel-1 images, Cropland Data Layer (CDL) and U-Net model, the optimal multi-temporal datasets for rice mapping were summarized, using the global search method. Then, based on the optimal multi-temporal datasets, the proposed workflow (a combination of K-Means and random forest) was directly used to extract the rice-distribution information of Jiangsu (i.e., the K-RF pseudo-labels). For comparison, the optimal well-trained U-Net model acquired from Arkansas (i.e., the transfer model) was also transferred to Jiangsu to extract local rice-distribution information (i.e., the TF pseudo-labels). Finally, the pseudo-labels with high confidences generated from the two methods were further used to retrain the U-Net models, which were suitable for rice mapping in Jiangsu. For different rice planting pattern regions of Jiangsu, the final results showed that, compared with the U-Net model trained on the TF pseudo-labels, the rice area extraction errors of pseudo-labels could be further reduced by using the U-Net model trained on the K-RF pseudo-labels. In addition, compared with the existing rule-based rice mapping methods, the U-Net model trained on the K-RF pseudo-labels could robustly extract the spatial distribution information of rice. Generally, this study could provide new options for applying a deep semantic segmentation model to training sample shortage regions.

Keywords: Sentinel-1; rice mapping; K-Means algorithm; random forest classifier; pseudo-label; U-Net



Citation: Wei, P.; Huang, R.; Lin, T.; Huang, J. Rice Mapping in Training Sample Shortage Regions Using a Deep Semantic Segmentation Model Trained on Pseudo-Labels. *Remote Sens.* **2022**, *14*, 328. <https://doi.org/10.3390/rs14020328>

Academic Editors: Mi Wang and Jianlai Chen

Received: 1 December 2021

Accepted: 10 January 2022

Published: 11 January 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

For the remote sensing community, traditional machine learning algorithms have been well adapted to the application of crop classification in recent years [1–4], especially for supervised classifiers trained on pixel point samples. For instance, random forest methods (RF), support vector machines (SVM), and so on [5–8].

Recently, deep learning-based methods have achieved outstanding performances in computer vision tasks [9,10], and have attracted increasing attention in the remote sensing

community [11–14]. Among various deep learning network structures, deep semantic segmentation networks (i.e., the same as deep convolutional neural networks, DCNNs) are being expanded to be the new models for crop mapping based on remote sensing images [15–17]. Compared with traditional machine learning algorithms (e.g., SVM, RF and so on), they explore and learn global spatial features and temporal signatures in multi-temporal images, in an end-to-end manner, to improve crop classification accuracy and calculation efficiency [18]. It is well known that wide and varied training sample collection (i.e., reference data) is critical to train a supervised classifier with robust performance for large-scale crop mapping. Currently, for supervised classifier training, the annotation of its reference data is mainly based on a combination of ground surveys and high-resolution optical images for photointerpretation [19–21]. However, different from independent pixel point sample sets used by traditional supervised classifiers, training sample sets of the DCNN model are composed of a large number of annotated image sets [22–24]. This means that all pixel points in a continuous space need to be annotated to construct samples required for DCNN model training; the process is very time-consuming, at an operational level, and it is unfeasible to construct a large number of ground reference image sets through visual interpretation [25], limiting its application in large-scale crop mapping. In fact, when the coverage of a study area is relatively large, all supervised classification algorithms suffer this similar problem.

Nowadays, for supervised classification algorithms, in order to solve the major bottleneck of large-scale crop mapping in training sample shortage regions, two main categories of method have been proposed: (1) a method based on a combination of phenological characteristics and machine learning models [26–28]; (2) a transfer model-based method [29–31]. The former is achieved by using crop phenological characteristics to construct samples required for machine learning model training. The latter is achieved by training a transfer model based on existing public datasets. Although these methods can avoid the time-consuming problem of sample set annotation, there are still some limitations and disadvantages, as described below.

First, for the combination of phenological characteristics and machine learning models, only crop phenological rule-based experience can be used to construct samples required for machine learning model training. Since these phenological characteristics mainly rely on temporal signals, only those pixels with obvious target crop characteristics (e.g., rice transplanting periods) can be extracted in a complexity planting environment. For the traditional machine learning model, the samples can be a collection of independent pixel points, and a supervised classifier for mining target crops with unobvious features can be fitted by using these pixel points [27,28]. However, for DCNN model training, as mentioned above, there are higher requirements for sample annotation training. Consequently, if the data products generated by phenological characteristics are directly used as samples for DCNN training (i.e., direct combination of phenological characteristics and DCNN), the model will not be able to fully learn the potential characteristics of target crop, and it will cause target crop omission or commission extraction.

Second, for the transfer model-based method, in a certain spatial range, the transfer model can show robust performance. However, when the target region is far away from reference region where the model is trained, performance of the transfer model will be largely impaired [29]. The main reason for this is that phenological characteristics of the same crops between regions with different climatic zones are significantly different [29]. In addition, for model transfer, most studies only focus on the multi-temporal datasets that gradually increase from sowing to mature periods [29–32], ignoring the multi-temporal datasets that contain critical growth periods shared by the same crop in different regions (i.e., only the middle parts of the whole crop-growth stage). Consequently, a certain number of training samples are still essential for crop mapping in target regions.

Therefore, in order to fully explore the impact of multi-temporal datasets corresponding to various growth stages on crop mapping, and efficiently construct reliable training samples required for DCNN model training, rice was taken as an example in this research.

First, based on public reference datasets (i.e., Cropland Data Layer, CDL) and DCNN model (i.e., U-Net), a global search was performed on different multi-temporal Sentinel-1 datasets (i.e., datasets corresponding to different rice-growth stages) to directly explore the optimal multi-temporal Sentinel-1 datasets for rice mapping (i.e., the optimal rice recognition period). Second, in the training sample shortage regions, based on multi-temporal datasets acquired from the optimal rice recognition period, a two-step workflow was proposed to construct reliable rice and non-rice samples (i.e., pseudo-labels) for DCNN model training, that was achieved by combining K-means [33] and random forest [34] (i.e., K-RF). Step 1 involved homogeneous sample construction based on the K-Means algorithm; step 2 involved the final fine sample construction based on the random forest classifier (i.e., the samples required for DCNN model training). Effectiveness of the proposed workflow was validated by comparing the differences between pseudo-labels generated by the K-RF, and pseudo-labels generated by the optimal transfer model. Besides, advantages of DCNN trained on the K-RF pseudo-labels were validated by comparing the differences between the results of the model trained on pseudo-labels, and the results of typical rule-based methods.

The following three research questions (RQ) were explored in this study:

- (RQ1) Based on synthetic aperture radar (SAR) images, what was the optimal multi-temporal datasets for rice mapping?
- (RQ2) In the absence of training samples, could the proposed workflow (i.e., the combination of K-means and RF, K-RF) effectively extract rice distribution information as pseudo-labels?
- (RQ3) How was the performance of U-Net model trained on the pseudo-labels?

2. Study Area and Datasets

2.1. Study Area

Jiangsu Province ($116^{\circ}20'20.8''\text{E}\sim 121^{\circ}51'4.6''\text{E}$, $31^{\circ}42'27.7''\text{N}\sim 34^{\circ}37'40.9''\text{N}$) is located at the eastern coastal region of mainland China; it is one of the main rice production regions in China. Throughout the year, its daily average maximum temperature in summer is 25.9°C , its daily average minimum temperature in winter is 3°C , and its average annual rainfall is 977 mm. Jiangsu Province is bounded by the Huaihe River, its southern and northern regions are a subtropical monsoon climate and a temperate monsoon climate, respectively. There is only single-season rice in Jiangsu Province. The location of the target region in Jiangsu is shown in Figure 1a. The specific rice-growth stages in Jiangsu are shown in Figure 1c.

For comparison, a transfer model was used to overcome the problem of sample shortage in Jiangsu Province, China (i.e., the target region). Therefore, it is necessary to use a rice distribution map with high-confidence to construct a reliable transfer model. Consequently, we selected the main rice production regions in the United States as the reference region to train the transfer model, namely Arkansas River Basin ($89^{\circ}50'46''\text{W}\sim 91^{\circ}17'39''\text{W}$, $33^{\circ}4'22''\text{N}\sim 36^{\circ}58'10''\text{N}$), which has a similar rice growth cycle to that of the target region. We chose this mainly because reference data from the United States is public and valid (i.e., Cropland Data Layer, CDL). The daily average maximum temperature of the Arkansas River Basin in summer is 34.2°C , the daily average minimum temperature in winter is minus 3°C , and the average annual rainfall is 1220 mm. The specific location of the reference region in Arkansas River Basin, and the corresponding rice-growth stages, are shown in Figure 1.

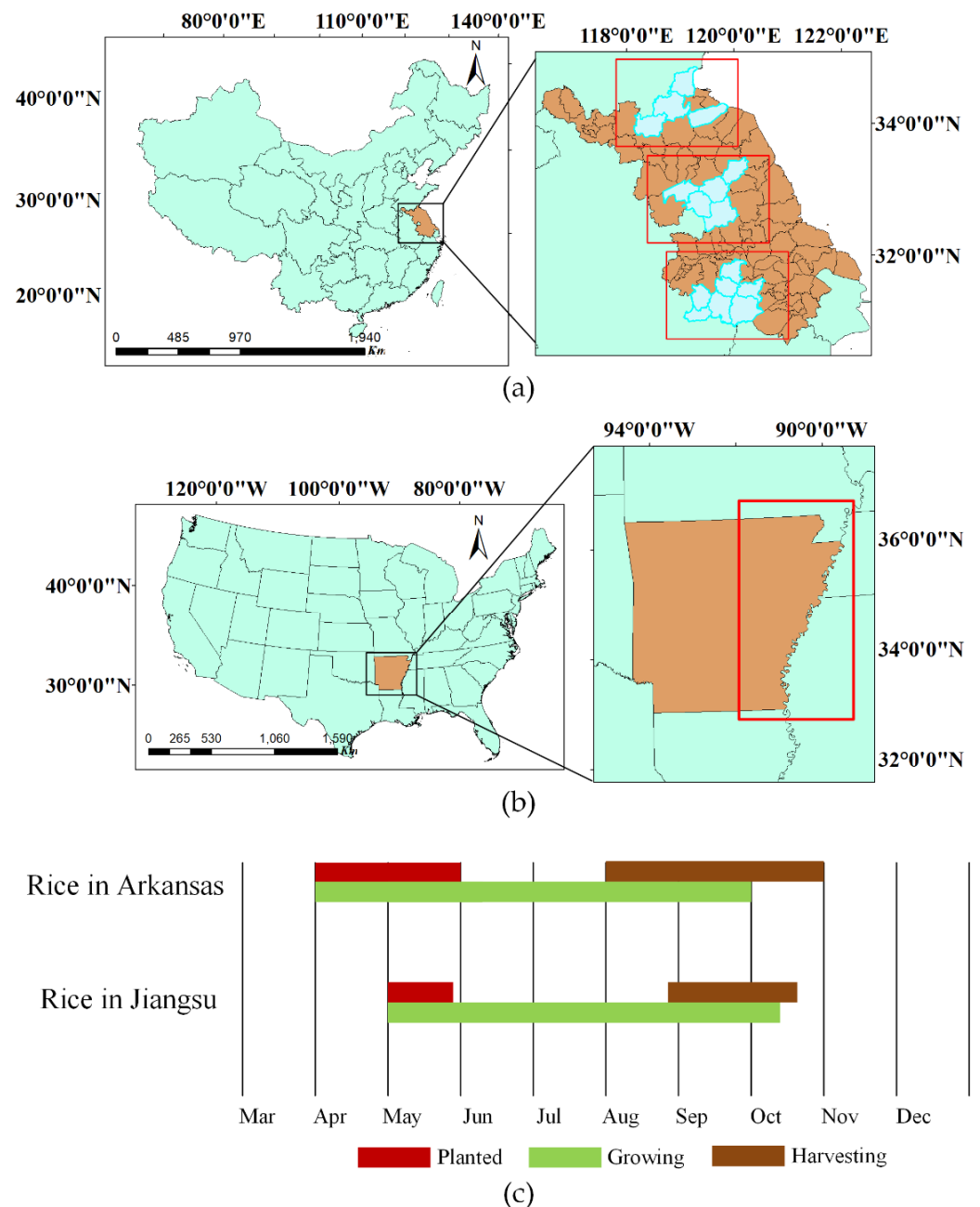


Figure 1. Location of the target and reference regions. (a) Location of the target region in Jiangsu. (b) Location of the reference region in Arkansas River Basin. (c) Rice calendar of the target and reference regions. The red boxes represent scenes covered by the selected Sentinel-1 images. The light blue polygonal areas represent the counties used to verify the extracted rice planting area.

2.2. Sentinel-1 Datasets

For the target region (i.e., the region in Jiangsu Province, China), according to the rice growth cycle of the target region, the Sentinel-1 images covering the three parts of the target region in Figure 1a were selected and downloaded from (<https://search.asf.alaska.edu/> (accessed on 15 April 2021)), and the specific data acquisition dates are listed in Table 1. For the reference region (i.e., the region in the Arkansas River Basin), in order to construct the optimal transfer model, multi-temporal datasets acquired from different years (i.e., 2017, 2018 and 2019) were used to enable the model to overcome differences of the rice growth cycle caused by different planting times between different years. Consequently, the three years of Sentinel-1 images that covered the whole rice-growth stages of the reference

region (Figure 1c) were selected and downloaded from the same website, and the specific data acquisition dates are listed in Table 1. Furthermore, all these Level-1 ground range detected (GRD) products were preprocessed in SNAP software, using the same steps with reference [18], and the spatial resolution of all datasets was processed to 30 m.

Table 1. Data acquisition dates (Month/date).

Jiangsu (i.e., the Target Region)		Arkansas River Basin (i.e., the Reference Region)	
2019	2017	2018	2019
4/23	4/4	3/30	4/6
5/5	4/28	4/23	4/30
5/17	5/22	5/17	5/24
5/29	6/3	5/29	6/5
6/10	6/15	6/10	6/17
6/22	6/27	6/22	6/29
7/4	7/9	7/4	7/11
7/16	7/21	7/16	7/23
7/28	8/2	7/28	8/4
8/9	8/14	8/9	8/16
8/21	8/26	8/21	8/28
9/2	9/7	9/2	9/9
9/26	9/19	9/14	9/21

2.3. Reference Datasets

For the target region located in China, the rice planting area of each county derived from Sentinel-1 images was compared with that of statistical data to verify the reliability of rice mapping results, since coverage of county in the target region is relatively small in the scene of Sentinel-1 image. The statistical data of rice planting area was collected from prefectural statistical yearbook of the target region in 2019, it recorded the rice planting area information of each county. Although there are some problems with statistical data, it is the only available open-source data that can reflect the rice planting situation of China. Moreover, some researchers have already used the crop planting area information recorded in the statistical yearbook to verify the reliability of the crop mapping results [35–38]. Consequently, we selected a total of 16 counties of the target region for the accuracy assessment of the rice mapping results (i.e., the counties located in northern, central and southern Jiangsu, which are covered by the blue polygons in Figure 1a).

For the transfer model construction, the reference datasets were downloaded from the Cropland data layer (CDL) (<https://nassgeodata.gmu.edu/CropScape/> (accessed on 20 April 2021)). The CDL is an annual publicly available land cover classification map provided by the U.S. Department of Agriculture (USDA) [39]. The 30 m CDL map was derived from NASS June Agriculture Survey, National Land Cover Dataset (NLCD), Common Land Unit (CLU), and imagery from several satellites (e.g., Landsat 5/7/8, Sentinel-2, and so on). More than 100 crop type categories are contained in the CDL, and it has a high accuracy for major crop types [40]. Moreover, in recent years, the reliability of the CDL has been proven by some scholars, who have used the datasets to carry out research on crop mapping [41–43].

3. Methodology

3.1. Global Search of Optimal Multi-Temporal Datasets for Rice Mapping

Based on a DCNN model and multi-temporal Sentinel-1 datasets with different time series lengths from the reference region (i.e., Arkansas River Basin), global search was performed to find the optimal multi-temporal datasets, which were suitable for local rice mapping and model transfer between different regions. Specifically, each observation time point was taken as the start observation date, and sequence length of the input observation date gradually increased until the last observation date was included (i.e., 13th observation

date). Take, as an example, the first observation date that was used as the start date, as shown in Figure 2; this process can also be used when other observation dates were used as start dates. Consequently, the optimal rice recognition period for local rice mapping could be found through this type of method. Additionally, the multi-temporal datasets containing the intersection of the rice growth cycle between different regions were screened, to be used in reducing the negative impact of model transfer caused by the difference in growth cycles between the two regions.

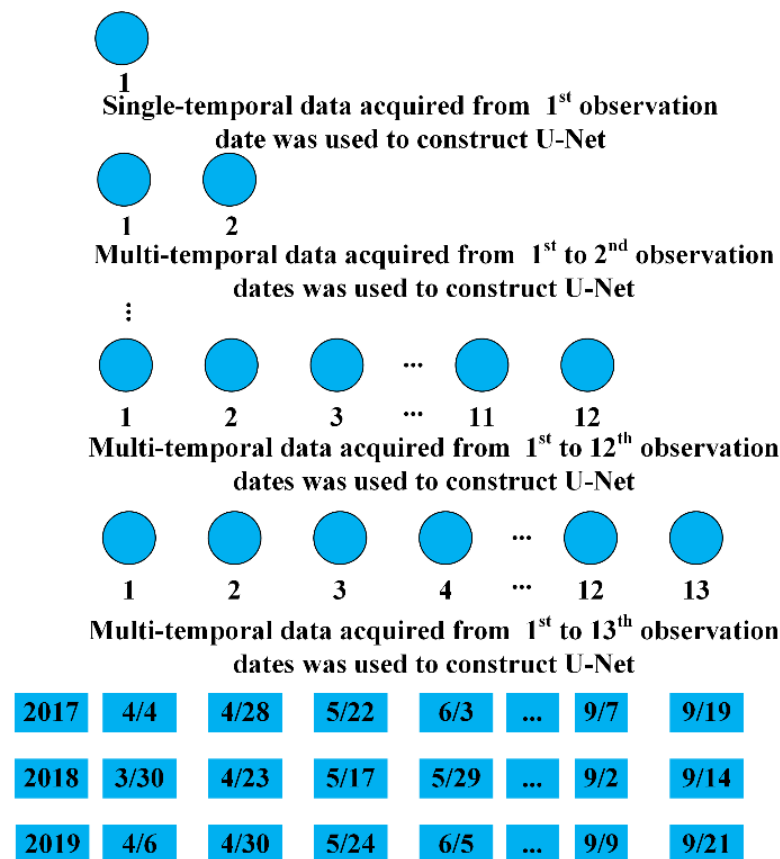


Figure 2. Based on the reference region, global search of optimal multi-temporal datasets for rice mapping. Take, as an example, the first observation date that was used as the start date; this process could also be used when other observation dates were used as the start dates.

As for DCNN model selection, among various DCNN model structures, the U-Net has been widely applied in the remote sensing community in recent years [18,22,23], since it has achieved outstanding performance and needed fewer samples [44]. Hence, the U-Net model was used to find the optimal multi-temporal datasets for local rice mapping to perform the model transfer.

3.2. Optimal Time Series Analysis for Rice Mapping

(RQ1) The datasets acquired from 2017, 2018, and 2019 in the reference region were used to construct the U-Net model. The specific model structure and parameter setting were the same as those in [18]. In order to avoid overlap between training, validation and test datasets, based on the CDL datasets covering the reference region (i.e., about 412 km × 142 km, ≈58,504 km²), the corresponding Sentinel-1 images were partitioned into a set of small non-overlapping images (i.e., a sliding window with the size of 256 × 256 pixels was moved across the entire image, with intervals of both 256 pixels in rows and columns, as shown in Figure 3), and a total of 2649 256 × 256 pixel non-overlapping images were generated for the three-year Sentinel-1 images. Then, 70% (1854)

of the image blocks were randomly selected from each year as the training datasets, 10% (264) from each year as the validation datasets, and 20% (531) from each year as the test datasets. The images in test datasets were only used for the final evaluation, which means that they were not involved in training or tuning. For the accuracy assessment of the results generated by the model, the scores, including the user's accuracy, producer's accuracy, Kappa coefficient, and F1 score, were calculated from a pixel-by-pixel comparison between the mapping results and CDL maps. Please refer to [45] for details about the mathematical expression of these parameters. The specific evaluation results of the model trained on different multi-temporal datasets in the reference region are listed in Figure 4.

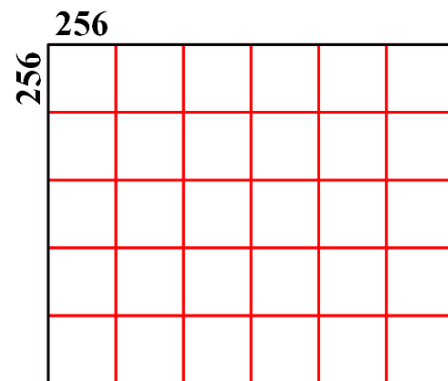


Figure 3. Image partition example for an image. The original valid image is indicated by the black box. According to the red grid, the image is partitioned into a set of small non-overlapping 256×256 pixel images.

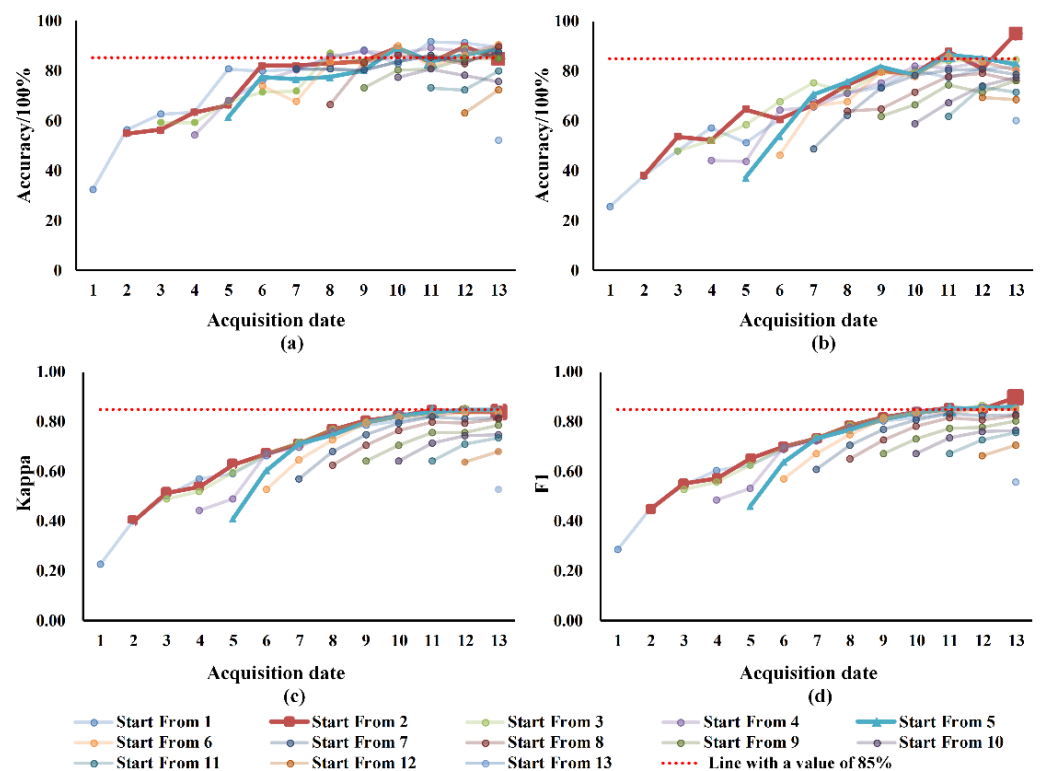


Figure 4. Over time in the reference region, the result accuracy assessments of the U-Net trained on different multi-temporal datasets; (a–d) are the classification PA, UA, Kappa, and F1 score over time in the reference region, respectively. Each point reflects the classification results of the model using temporal input till corresponding time steps.

As reported in Figure 4, for local rice mapping in the reference region, when observation dates contained in multi-temporal datasets were started before the tillering period (i.e., observation dates before the 6th point of Figure 4) and ended in the mature period (i.e., observation dates after the 10th point of Figure 4), the corresponding extraction results of the U-Net could basically reach a stable state (i.e., both kappa and F1 were close to 0.85). This means that multi-temporal datasets that corresponded to the middle and late rice-growth stages were more important for rice mapping, and only the datasets containing middle and late rice-growth stages (i.e., tillering to maturity) could also achieve acceptable mapping results.

Specifically, the optimal rice mapping result could be achieved (i.e., result with the highest F1) when the multi-temporal datasets acquired from the 2nd observation date to the 13th observation date were used as input for the U-Net model; that is, when the result of the last point of the crimson curve in Figure 4 (i.e., the end of April to the end of September that corresponds to the rice sowing period to the mature period of the reference region), and its rice producer's accuracy, user's accuracy, kappa and F1 were 84.63%, 94.98%, 0.84, and 0.90, respectively. In addition, the mapping result comparable to the optimal result could be achieved (i.e., both kappa and F1 were close to 0.85), when the datasets acquired from the 5th observation date to the 11th observation date were used as input of the U-Net model; that is, the result of the 7th point of blue curve in Figure 4 (i.e., the middle of June to the end of August that corresponds to the rice tillering period to the mature period of the reference region). Its rice producer's accuracy, user's accuracy, kappa and F1 were 83.71%, 86.24%, 0.84, and 0.85, respectively. Moreover, among the several acceptable results based on different multi-temporal datasets (i.e., both the kappa and F1 were close to 0.85), its time series length was the shortest. Theoretically, when phenological characteristics for the same crops between target and reference regions were significantly different (e.g., growth cycle), compared with the multi-temporal datasets containing a full rice-growth cycle, it is more reliable to use the datasets that only contain key growth stages of rice for model transfer, which will be validated in Section 4.

3.3. The Process of K-RF

Based on multi-temporal Sentinel-1 images covering the whole rice-growth stages of the target region (i.e., the optimal rice recognition period for local rice mapping, for which the corresponding data acquisition dates were from 5 May to 26 September 2019, as listed in Table 1), a pseudo-label generation workflow was constructed by combining the K-Means algorithm and RF classifier (i.e., K-RF). The workflow was composed of two parts: (1) homogeneous pseudo-label construction based on the K-Means algorithm; (2) final fine pseudo-label construction based on the RF classifier (i.e., the samples required for U-Net model training). The specific process is shown in Figure 5.

The details of the process and specific parameter settings are described as follows.

- (1) Masks for water bodies, buildings and parts of dryland vegetation. It is well known that the backscattering temporal signal of a water body always keeps a lower value, and therefore the threshold was set to $VH_{\max} \leq -20$ after experimental comparison. Similarly, buildings could be masked out by using a higher minimum value of backscattering coefficient, since their backscattering temporal signal always keeps a higher value. In addition, the minimum backscattering characteristics of rice are close to that of water bodies, due to the existence of rice transplanting periods (i.e., the specular scattering of water is dominant). During the rice-growth cycle, for the part of dryland vegetation, its backscattering coefficient is always higher than that of a water body, since there are no obvious water bodies in the growing environment (i.e., volume scattering is dominant). Consequently, during the whole rice-growth stages, an appropriate minimum value of the backscattering coefficient could be used to mask out the building and part of dryland vegetation at the same time. After experimental comparison, the threshold was set to $VH_{\min} \geq -17$.

- (2) The difference calculation of pixel points between adjacent acquisition time images. Due to a larger fluctuation of rice in backscattering characteristics, caused by the flooding phenomenon (i.e., before and after rice transplanting period), the backscattering coefficient of the previous observation time was subtracted from the next observation time, to highlight the relative changes of backscattering characteristics of rice during its growth cycle. The specific mathematical expression is formulated as follows:

$$V_{\text{relative}} = [V_2 - V_1, V_3 - V_2, \dots, V_n - V_{n-1}] \quad (1)$$

where V_{relative} represents the vector of relative change of backscattering characteristics; V_n represents the original backscattering characteristics at the n th observation time.

- (3) Clustering for the time series difference images. The time series difference images generated in step 2 were clustered into four classes (the value was determined after testing three, four and five) using the K-means algorithm based on Euclidean distance. Then, the standard deviation of each centroid vector was calculated. The centroid vector with the largest standard deviation was defined as the centroid of rice; the centroid vector with the smallest standard deviation was defined as the centroid of water and buildings (i.e., non-vegetation); the rest were defined as the centroid of non-rice vegetation.
- (4) Homogeneous sample construction. Based on current samples (the samples generated the first time were collected from step 3), 10,000 non-vegetation pixels were randomly selected from the current samples as non-vegetation training samples. For the construction of rice and non-rice vegetation training samples, on the basis of a homogeneous window, when all pixels in the window were classified as rice (or non-rice vegetation), the center pixel point of the window was selected as the rice (or non-rice vegetation) training point. After experimental comparison, the homogeneous window size was set to 11×11 .
- (5) Loop-based training. The RF classifier was trained based on the homogeneous samples generated in step 4, and then the well-trained RF classifier could be used to further extract the potential rice points with unobvious features as the current samples. The overlap between the current samples extracted by RF and the previous samples was calculated. If it was higher than the set threshold, then the sample construction process could jump out of the loop, otherwise, based on samples extracted by the current RF, the 4th to 5th steps were repeated. After experimental comparison, the threshold of the overlap was set to 0.9.
- (6) Final fine training sample generation. After jumping out of the loop, the last produced result was taken as the final fine training sample (i.e., pseudo-label) required for U-Net model training.

The pseudo-labels generated by different methods (i.e., the K-RF and transfer models) in the target region were compared, and further used as training samples to construct a U-Net model, which was more suitable for rice mapping in the target region. The process of automatically constructing samples (i.e., K-RF) could save a lot of manpower and time compared with visual interpretation. This could provide an important foundation for the practical application of the DCNN model.

3.4. Evaluation Parameters for Quantitative Analysis

For rice mapping results of the target region, based on each county, the root mean square error (RMSE) and relative root mean square error (RRMSE) between the rice area derived from Sentinel-1 images and the statistical data were calculated to verify the reliability of the rice mapping results [36,46]. The specific formulations are expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{S}_i - S_i)^2} \quad (2)$$

$$\text{RRMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(\frac{\hat{S}_i - S_i}{\hat{S}_i} \right)^2} \quad (3)$$

where m represents the number of counties; \hat{S}_i and S_i are the rice areas of the i th county recorded in the statistical yearbook and rice areas derived from Sentinel-1 images, respectively.

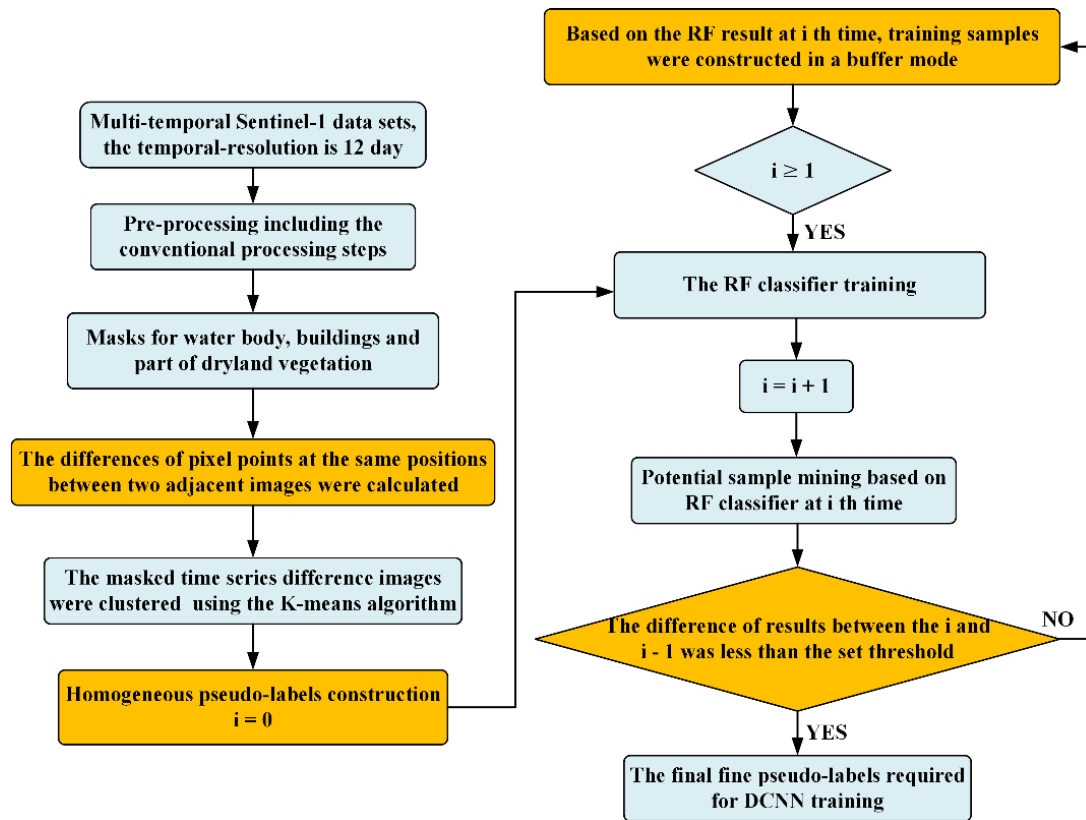


Figure 5. The pseudo-labels generation process based on K–RF.

In order to further evaluate the reliability of the pseudo-labels, the Intersection over Union (IoU) was introduced to calculate the differences between the pseudo-labels and the extraction results (i.e., the rice mapping results of DCNN trained on the pseudo-labels) [23], the specific formulation is expressed as follows:

$$\text{IoU} = \frac{A_{\text{pseudo}} \cap A_{\text{predict}}}{N} \quad (4)$$

where N represents the pixel number of a label image; A_{pseudo} represents the pseudo-labels, A_{predict} represents the extraction result of the DCNN model trained on the pseudo-labels, \cap represents the pixel number statistics with the same category and the same position between the two results.

4. Experimental Results and Discussion

4.1. Pseudo-Labels Generated by Different Methods

Pseudo-labels generated by the K–RF and by the optimal transfer model were compared to construct the optimal pseudo-labels required for DCNN model training in the target region. Among them, the transfer model was directly applied to the target region (i.e., Jiangsu) after being well trained in the reference region (i.e., the Arkansas River Basin). Then, the rice planting area collected from the prefectural statistical yearbook was used to validate the reliability of the pseudo-labels.

4.1.1. Pseudo-Labels Generated by the K–RF

(RQ2) Following the steps described in Section 3.3, the rice pseudo-label construction was carried out in the three parts of the target region with different rice planting patterns to verify the robustness of the K–RF (i.e., northern and central Jiangsu where rice planting is dense, and southern Jiangsu where rice planting is sparse). The specific clustering results generated by the K-Means algorithm are shown in first row of Figure 6. Pseudo-labels generated by the K–RF method are shown in the third row of Figure 6. The difference between the clustering results and results of the K–RF are shown in second row of Figure 6. In addition, the final pseudo-labels generated by the K–RF method in the northern, central, and southern target regions were summarized as rice and non-rice, as shown in the fourth row of Figure 6. The specific rice area of each county contained in the pseudo-labels are shown in Figure 7 and Table 2.

Table 2. Rice area extraction results in different parts of the target region based on different methods.

Region	County Name	S1/Saa	S1/10 ³ Hectares	S2/10 ³ Hectares	S3/10 ³ Hectares	S4/10 ³ Hectares
Northern part	Ganyu	18%	27.76	7.41	17.82	26.21
	Donghai	31%	65.07	34.31	58.81	66.26
	Xinyi	17%	28.10	14.72	21.72	30.38
	Guanyun	31%	48.16	48.97	64.79	46.97
	RMSE	/	/	19.62	10.66	1.61
	RRMSE	/	/	49.70%	27.76%	5.16%
Central part	Jinhu	27%	38.29	30.58	38.43	36.73
	Gaoyou	27%	53.48	46.03	45.57	46.48
	Baoying	36%	54.34	52.75	59.71	53.44
	Hongze	22%	31.66	31.62	34.76	39.09
	Jianhu	40%	47.08	37.97	44.09	48.87
	RMSE	/	/	6.33	4.68	4.70
Southern part	RRMSE	/	/	14.02%	9.52%	12.30%
	Gaochun	16%	6.04	8.17	6.56	6.12
	Lishui	20%	17.10	10.51	7.87	9.24
	Liyang	12%	31.37	22.42	21.09	31.47
	Yixing	4%	24.98	20.49	20.09	23.44
	Wujin	13%	4.40	17.95	4.65	1.11
	Jintan	28%	12.73	26.03	13.93	10.33
	Danyang	16%	29.60	30.34	24.67	28.62
	RMSE	/	/	8.52	5.86	3.41
	RRMSE	/	/	125.14%	26.30%	34.04%
The three parts	RMSETotal	/	/	11.85	7.09	3.56
	RRMSETotal	/	/	86.78%	22.88%	23.68%

Note: S1, rice planting area in statistical yearbook; Saa, county-level administrative district area; S2, rice area extraction of pseudo-labels generated by K-Means; S3, rice extraction results of pseudo-labels generated by K–RF method; S4, the optimal rice extraction results of the transfer model.

From Figure 6 it can be seen that, based on the time series difference images, the K-Means algorithm could cluster the rice pixel points. However, there were still some errors in the clustering results, which were caused by other non-rice pixels with the temporal signals similar to rice. Consequently, a supervised RF classifier trained on homogeneous sample sets (i.e., the K–RF) was used to extract the potential rice pixel points in Sentinel-1 images, and suppress the misclassification in the results of K-Means algorithm; as we expected, based on the K–RF classifier, categories of some points in the results of the K-Means algorithm were reassigned, as reported in second row of Figure 6. Moreover, for the results of the regions with different rice planting patterns, from the rice area extraction results shown in Figure 7 and Table 2, it could be also verified that some incorrectly clustered pixel points had been re-classified and corrected by the K–RF classifier, since the $RMSE_{Total}$ ($RRMSE_{Total}$) was improved from 11.58×10^3 hectares to 7.09×10^3 hectares (86.78% to

22.88%). Among the results of the three parts of the target region, the improvement of result in the southern part was the most obvious, and the corresponding RRMSE was improved from 125.14% to 26.36%.

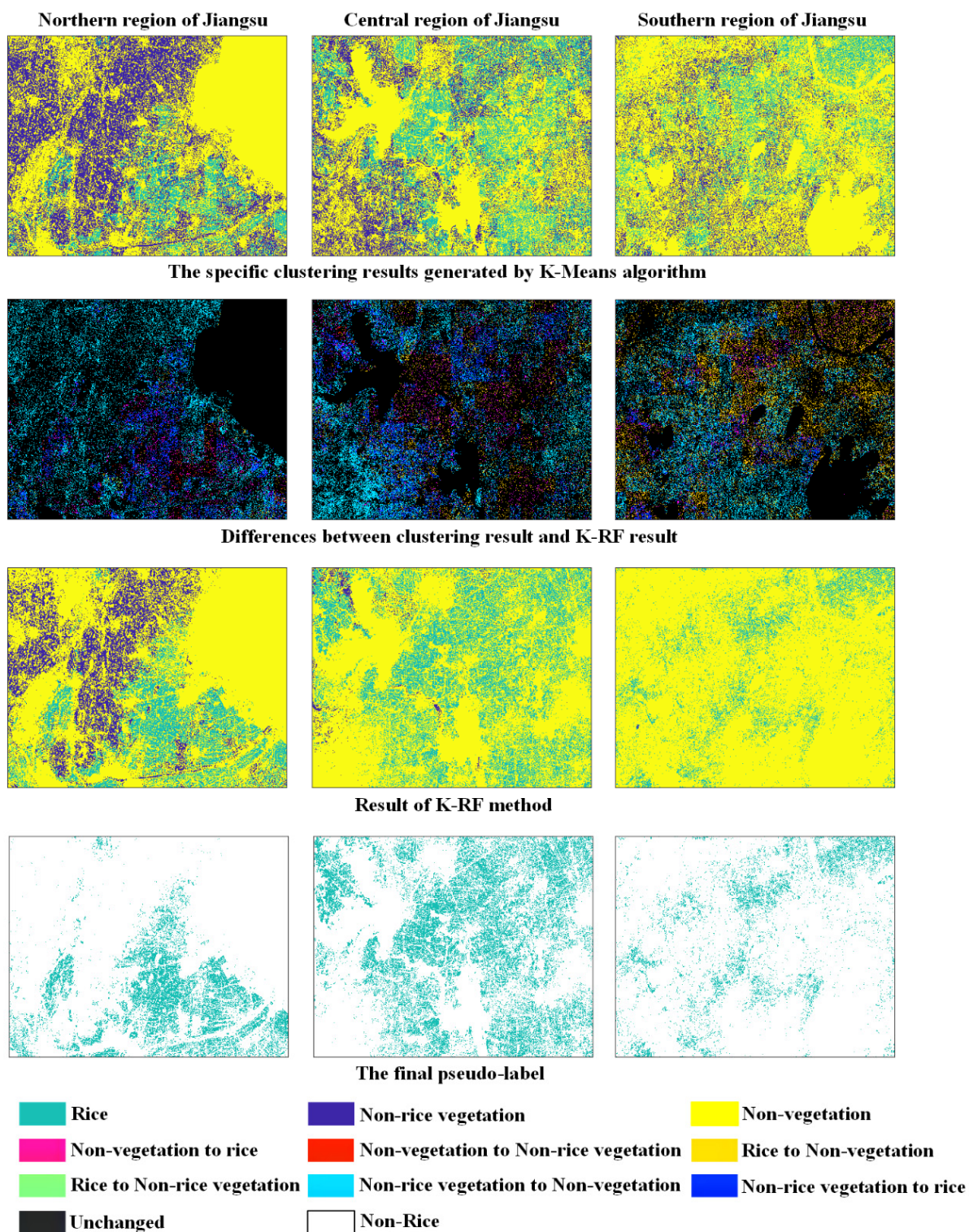


Figure 6. Pseudo-label extraction in the different parts of the target region. The first row is the results of the K-Means algorithm. The third row is the results of the K-RF. The second row is the variations between the points in the same spatial positions as the first and third rows. The fourth row contains the final pseudo-labels generated by the K-RF, which were summarized as rice and non-rice. The first to third columns represent the northern, central and southern target regions, respectively.

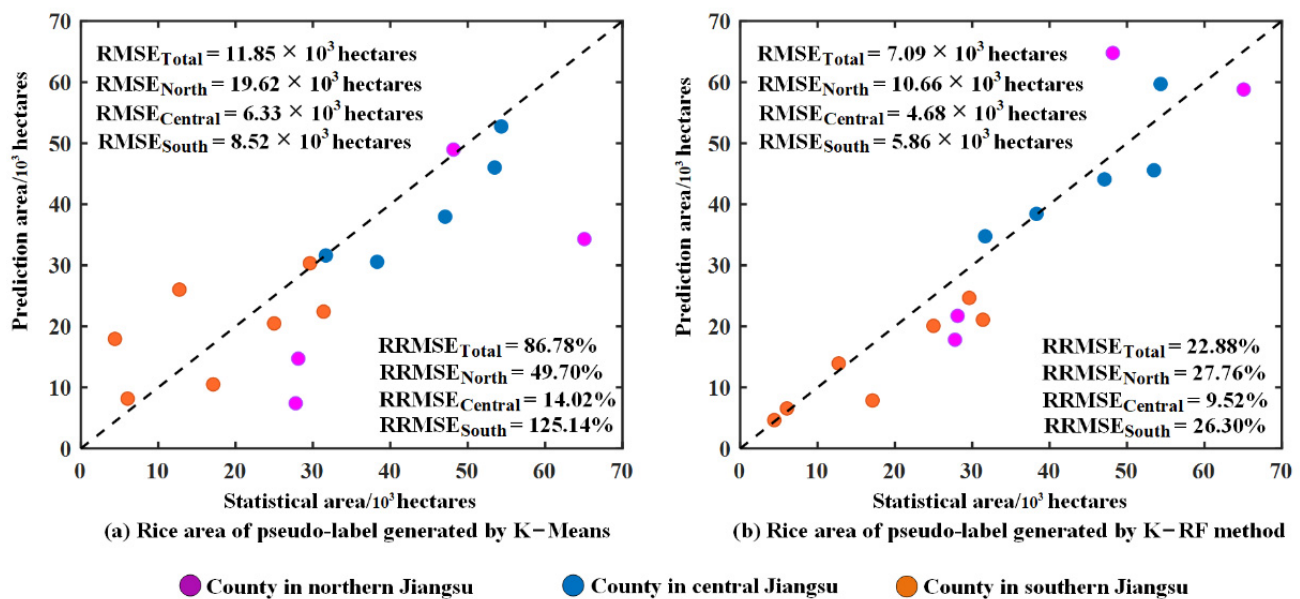


Figure 7. The county-level comparison of rice area between statistics data and pseudo-labels; (a) and (b) are the rice area of pseudo-labels generated by K-means algorithm and K-RF method, respectively.

In general, the K-RF could effectively extract rice-distribution information in the target region with different rice planting patterns (i.e., dense and sparse planting patterns).

4.1.2. Pseudo-Labels Generated by the Transfer Model

(RQ2) Hao et al. [29] have verified that transfer model could be used to extract similar crops in different regions. Consequently, pseudo-labels generated by the K-RF were compared with the optimal pseudo-labels generated by transfer models to further verify the effectiveness of the K-RF. In order to obtain the optimal pseudo-labels generated by transfer model, based on different multi-temporal datasets in the target regions, all of the well-trained transfer models, which are presented in Section 3.2, were directly used to extract rice-distribution information in the three parts of the target region. Accuracy assessments of the pseudo-labels generated by different transfer models are shown in Figure 8. The specific rice area extraction results of each county contained in the optimal pseudo-labels are listed in Table 2.

As reported in Figure 8, in the target region, for the pseudo-labels generated by different transfer models, the reliability of the rice pseudo-labels in the northern part of the target region were always higher than those of the central part. Specifically, in the northern part of the target region, for most accuracies of the pseudo-labels generated by transfer models, when the observation dates contained in multi-temporal datasets were started from the sowing period (i.e., observation dates between the first and fourth points of Figure 8a) and ended in the heading or mature period (i.e., observation dates between 8th and 13th point of Figure 8a), the RMSE values could basically reach a stable state (i.e., the RMSE was close to 5×10^3 hectares). However, in the central and southern parts of the target region, for most accuracies of the pseudo-labels generated by the transfer model, RMSE values were better when observation dates contained in multi-temporal datasets were started from the tillering or heading period (i.e., observation dates between 6th and 8th points of Figure 8b,c) and ended in the mature period (i.e., the 11th and 12th observation points of Figure 8b,c).

Consequently, in the following contents, we only compared the optimal results of the transfer model with the results of K-RF to analyze the differences of the pseudo-labels generated by different methods.

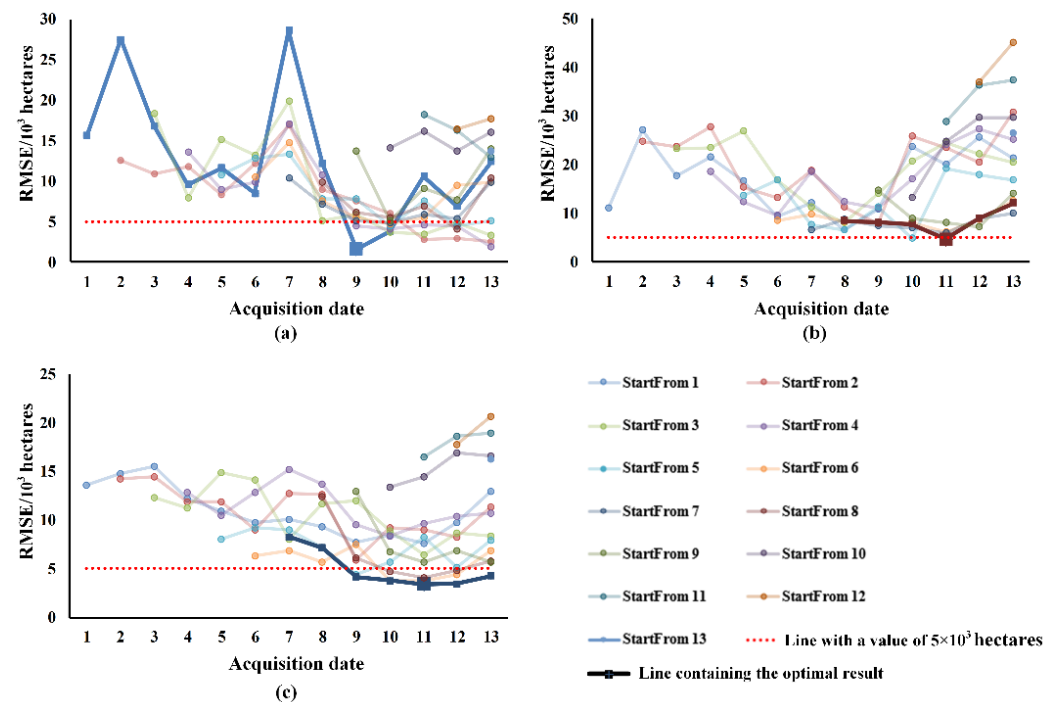


Figure 8. Based on different multi-temporal datasets, RMSE of pseudo-labels in the different parts of the target region directly extracted by the well-trained transfer model; (a–c) are RMSE values of pseudo-label in the northern, central, and southern target region, respectively. Each point reflects the classification results of the transfer model using temporal input to the corresponding time steps.

4.1.3. Comparison between the Two Types of Pseudo Labels

As reported in Table 2, in the northern part of the target region, for accuracy of the optimal pseudo-labels generated by the transfer model (i.e., the TF pseudo-labels), the RMSE (RRMSE) was 1.61×10^3 hectares (5.16%). The multi-temporal datasets in the target region acquired from the first observation date to the ninth observation date were used as input of the transfer model; that is, the result of ninth point of the blue curve in Figure 8a, which was improved by 9.05×10^3 hectares (22.60%) compared with that of the pseudo-labels generated by the K-RF (i.e., the K-RF pseudo-labels). In the central part of the target region, for accuracy of the optimal TF pseudo-labels, its RMSE was 4.7×10^3 hectares (i.e., the result of fourth point of the crimson curve in Figure 8b), which was slightly lower than that of the K-RF pseudo-labels. The difference was amplified in RRMSE, which was decreased by 2.78% compared with that of the K-RF pseudo-labels. In the southern part of the target region, for accuracy of the optimal TF pseudo-labels, its RMSE was 3.41×10^3 hectares (i.e., the result of fifth point of the navy-blue curve in Figure 8c), which was improved by 2.45×10^3 hectares compared with that of the K-RF pseudo-labels. However, its RRMSE was decreased by 7.74% compared with that of the K-RF pseudo-labels. In addition, local differences between the two types of pseudo-labels are shown in Figure 9, from which it can be seen that differences between the two types of pseudo-labels were mainly reflected in the boundaries between different fields.

4.2. Rice Mapping Results of U-Net Trained on Different Pseudo-Labels

(RQ3) The time saved in annotation of training samples could provide an important foundation for the practical application of the DCNN model. Consequently, in order to verify feasibility of using the pseudo-labels to construct a DCNN model, the two types of pseudo-labels (i.e., the TF pseudo-labels and K-RF pseudo-labels) were used to further retrain the U-Net model. Moreover, the rice area statistics were used to validate reliabilities of the results generated by U-Net trained on the pseudo-labels.

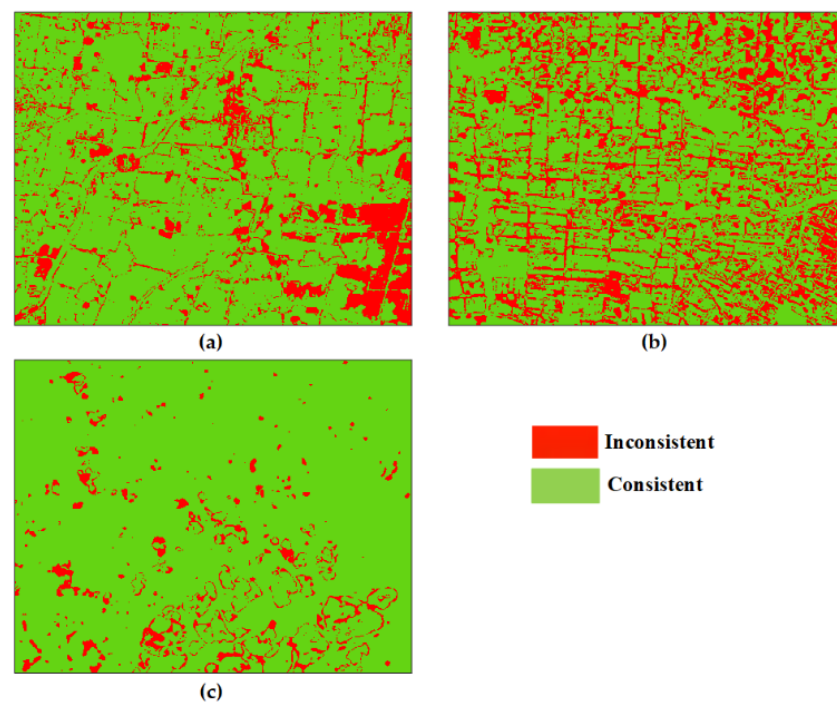


Figure 9. The local differences between the K-RF and optimal TF pseudo-labels, (a–c), are differences between the two types of pseudo-labels in the northern, central, and southern parts of the target region.

Similarly, based on the pseudo-label sets, the training image sets were made using the method presented in Section 3.2; this resulted in $1152 \times 256 \times 256$ pixel small non-overlapping images for Sentinel-1 images, covering the three parts of the target region: 70% (806) of the image blocks were randomly selected as the training datasets, 10% (115) as the validation datasets, and 20% (231) as the test datasets. The specific network structure and parameter settings were the same as described in Section 3.2. IoU values (IoUs) between the pseudo-labels and extraction results (i.e., rice mapping results of U-Net trained on pseudo-labels) were calculated, as listed in Table 3. In addition, the comparison of the rice area extraction results generated by the U-Net models trained on different pseudo-labels are shown in Table 3 and Figure 10.

Table 3. Consistency of pseudo-labels and extraction results.

Pseudo-Label Generation Method	Datasets	IoU/%	RMSE/ 10^3 Hectares	RRMSE
K-RF (K-RF pseudo-labels)	Training	94.46	/	/
	Validation	93.79	/	/
	Test	95.02	/	/
	Northern	96.73	8.80	25.34%
	Central	93.16	4.25	8.39%
	Southern	95.20	4.23	29.82%
Transfer model (TF pseudo-labels)	Training	92.41	/	/
	Validation	92.20	/	/
	Test	93.25	/	/
	Northern	94.15	4.21	14.06%
	Central	86.86	5.64	13.51%
	Southern	96.66	3.74	35.87%

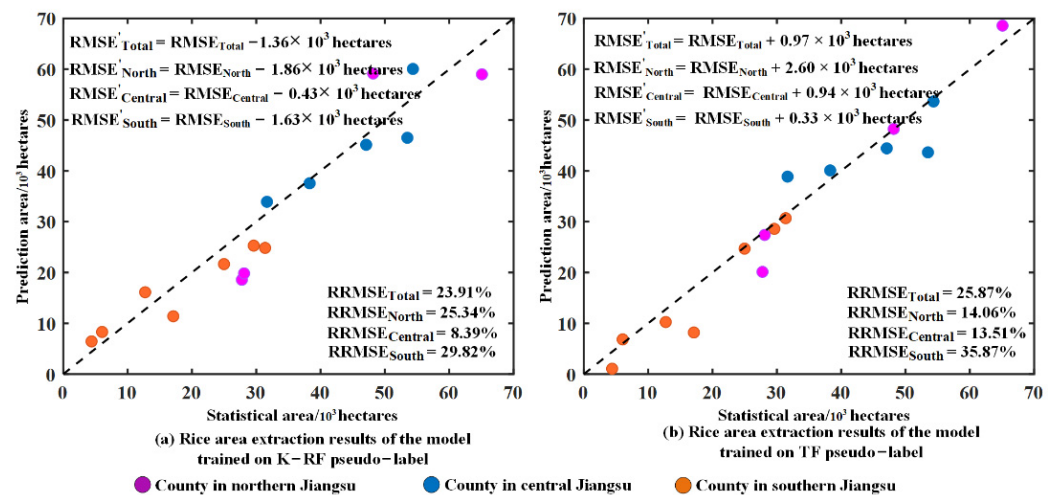


Figure 10. The comparison of rice areas between the statistics data and the extraction results of the model trained on pseudo-labels; (a,b) are the rice area of extraction results of the model trained on the K-RF and TF pseudo-labels. RMSE represents the rice area extraction errors of the U-Net trained on the pseudo-labels, RMSE represents the rice area extraction errors of the pseudo-labels.

For U-Net models trained on the two types of pseudo-labels, it can be seen from Table 3 that both models had achieved ideal IoUs between the extraction results and pseudo-labels. This means that both models were not over fitted. Specifically, for U-Net trained on the K-RF pseudo-labels, IoUs of the three datasets were from 93.79% to 95.02%, and IoUs of the three parts were from 93.16% to 96.73%. The errors of these IoUs mainly came from two parts: (1) the model trained on pseudo-labels could further correct the false information in the pseudo-labels while reducing the errors of extracted rice area (as shown in Figure 10), and thus the difference between the extraction result and pseudo-labels was increased; (2) errors were caused by incorrectly marked points during the training process. For the U-Net model trained on the TF pseudo-labels, IoUs of the three datasets were from 92.20% to 93.25%, which were slightly lower than those of U-Net trained on the K-RF pseudo-labels. The fluctuation ranges of IoUs corresponded to the three parts that were from 86.86% to 96.66%, which were significantly higher than those of the U-Net trained on the K-RF pseudo-labels. The reasons for the errors of IoUs were the same as those of U-Net trained on the K-RF pseudo-labels. More importantly, compared with the rice area extraction errors of the pseudo-labels, the extracted area errors of the U-Net trained on the TF pseudo-labels were increased for all of these three parts of the target region. On the contrary, the U-Net trained on the K-RF pseudo-labels reduced these errors, as reported in Figure 10. In addition, except for the northern part of the target region, RRMSE values of the rice extraction results of U-Net trained on the TF pseudo-labels in other parts were worse than those of U-Net trained on the K-RF pseudo-labels.

In summary, from the experiment that U-Net trained on different pseudo-labels, it could be concluded that the K-RF was more reliable for pseudo-label construction than the transfer model.

4.3. Comparison of Rice Mapping Performance with Other Rice Mapping Methods

(RQ3) Table 4 and Figure 11 report the comparison between the U-Net trained on the K-RF pseudo-labels and the two-rule-based rice mapping methods (i.e., the methods proposed by Nguyen et al. [47] and the method proposed by Zhan et al. [48]). For the three parts of the target region with different rice planting patterns (i.e., sparse and dense planting patterns), the RMSE (RRMSE) of the U-Net trained on the K-RF pseudo-labels was always the smallest, compared with that of the other two methods. For the method proposed by Nguyen et al. [47], RMSE values of the three parts of the target region were from 13.86×10^3 hectares to 37.52×10^3 hectares, which were higher by 5.06×10^3 hectares,

33.27×10^3 hectares, and 15.22×10^3 hectares, compared with those of the U-Net trained on the K–RF pseudo-labels. For the method proposed by Zhan et al. [48], RMSE values of the three parts of the target region were from 9.66×10^3 hectares to 33.91×10^3 hectares, which were higher by 25.11×10^3 hectares, 5.41×10^3 hectares, and 16.45×10^3 hectares compared with those of the U-Net trained on the K–RF pseudo-labels. Advantages of U-Net trained on the K–RF pseudo-labels were more obvious in the RRMSE indexes for the three parts of the target region. Similarly, these differences were significantly obvious in the extracted rice spatial distribution information, as shown in Figure 11.

Table 4. The comparison with rule-based rice mapping methods.

Method	Region	RMSE/ 10^3 Hectares	RRMSE
Nguyen et al. [47]	Northern	13.86	31.97%
Zhan et al. [48]		33.91	68.57%
U-Net trained on the K–RF pseudo-labels		8.80	25.34%
Nguyen et al. [47]	Central	37.52	80.14%
Zhan et al. [48]		9.66	26.31%
U-Net trained on the K–RF pseudo-labels		4.25	8.39%
Nguyen et al. [47]	Southern	19.45	90.65%
Zhan et al. [48]		20.68	257.34%
U-Net trained on the K–RF pseudo-labels		4.23	29.82%

In conclusion, at the rice area extraction level, compared with the two-rule-based methods, the U-Net trained on the K–RF pseudo-labels could robustly extract rice spatial distribution information in the regions with different rice planting patterns.

4.4. Discussion

Through the analysis of temporal signals based on backscattering or spectral characteristics [47,49], it could be concluded that the transplanting period was essential to discriminate rice from other land covers (i.e., the mixing of volume scattering from rice and the specular reflection from water). However, only considering the temporal signals of backscattering characteristics would ignore the potential features contained in different multi-temporal dataset combinations (e.g., spatial characteristics). Consequently, from the perspective of mapping results, instead of just taking sowing period as the start date and gradually increasing to the maturity period [29,30], this study explored the influence of different multi-temporal datasets (i.e., data that covered each growth stage of rice) on rice mapping more comprehensively through the global search method. This could be used to find the optimal multi-temporal datasets for local rice mapping and the model transfer between different regions. Then, under the condition that no training sample sets existed in the target region (i.e., manually annotated training sample sets), based on the optimal multi-temporal datasets, this study combined the K-Means and RF classifier (i.e., K–RF) to construct reliable pseudo-labels, and compared it with the optimal pseudo-labels generated by the transfer model. Furthermore, in the different parts of the target region (i.e., training sample shortage regions in Jiangsu Province, China), this study combined the K–RF pseudo-labels and the DCNN model (i.e., U-Net) to extract rice spatial distribution information in the target region, instead of the traditional machine learning algorithm.

For the process of constructing the K–RF pseudo-labels in the different parts of the target region, errors in the initial clustering results of the K-Means algorithm could be effectively suppressed by using the K–RF based method, especially for the southern part of the target region (as shown in Figure 7 and Table 2). The main reason was that compared with the northern and central parts of the target region, the southern part is a region where rice is sparsely planted. In the initial clustering results of the southern part, there were more independent and scattered pixel points. Based on the homogeneous window, while collecting pure samples, it could also avoid collecting independent and wrong sample sets.

Therefore, more scattered error information in the initial clustering results of the southern part could be further corrected by the K–RF classifier trained on the pure samples.

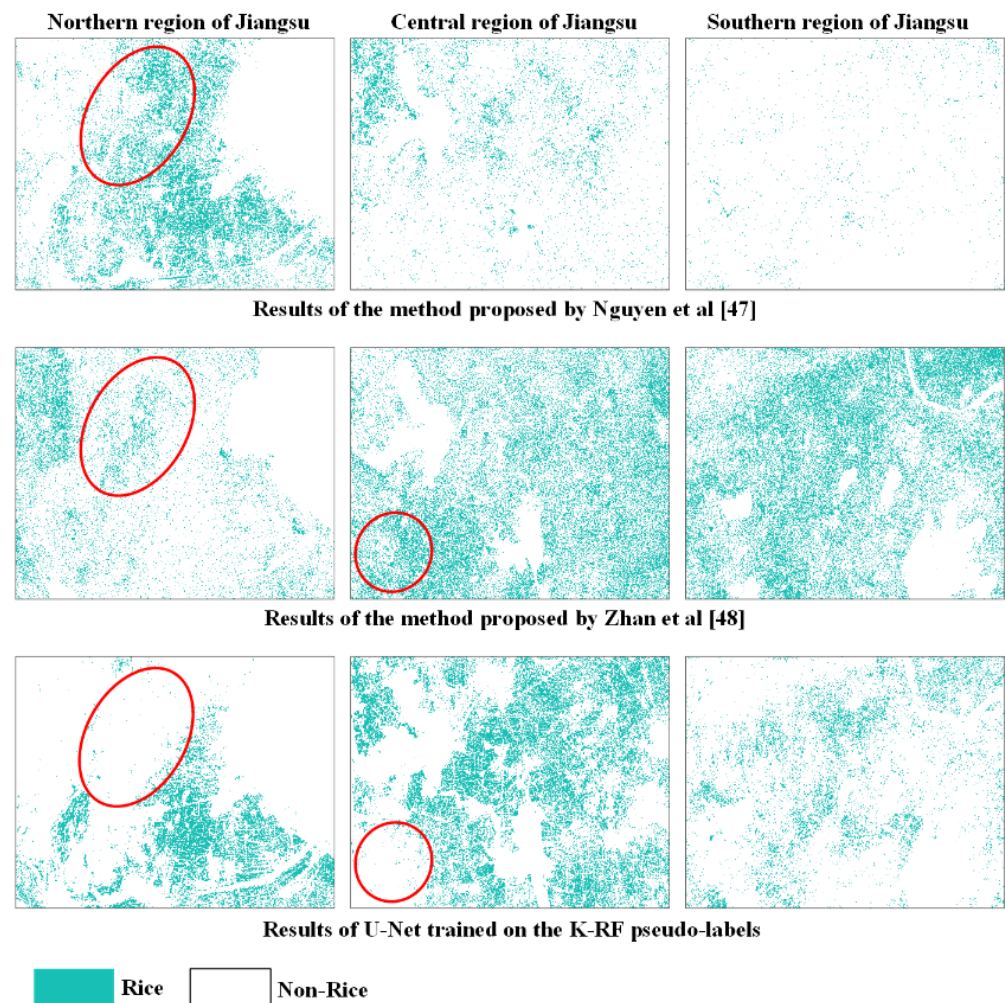


Figure 11. Extraction results in the different parts of the target region based on different methods. The first row to third row represent the rice extraction results using the method based on Nguyen et al. [47], and Zhan et al. [48], and the U-Net model trained on the K–RF pseudo-labels, respectively. The first to third columns represent the rice extraction results in the northern, central and southern parts of the target region, respectively.

For the process of constructing the TF pseudo-labels, there was an interesting phenomenon; when latitudes between the target and reference regions were significantly different, the accuracy of the pseudo-labels generated by the transfer model trained on multi-temporal datasets that only contained middle to late rice-growth stages were higher than those of the pseudo-labels generated by transfer model trained on multi-temporal datasets that contained the whole rice-growth stages (i.e., results of the central and southern parts of the target region). The main reason for this phenomenon was that, compared with the northern part of the target region, latitudes of the central and southern parts were significantly different from the latitude of the reference region (as shown in Figure 1), which further caused differences of phenological characteristics for the same crops between different regions (e.g., growth cycle). Consequently, the transfer ability of the model had been weakened in the central and southern parts of the target region when the multi-temporal datasets that contained the whole growth stages were used.

For the comparison between the K–RF pseudo-labels and TF pseudo-labels at the level of rice planting area verification, the results showed that both the optimal transfer model

and the K–RF could effectively extract rice distribution information in the target region as pseudo-labels. Moreover, compared with the TF pseudo-labels, the K–RF pseudo-labels had obvious advantages in the RRMSE index, especially for the southern part of the target region. The main reason was that the relative error of rice area extraction results based on K–RF in rice cultivation in the sparsest counties (i.e., Wujin) was much smaller than that of the transfer model, whereas the relative errors of other counties between the two methods were similar to each other. Therefore, RRMSE of the K–RF pseudo-labels was much better than that of the optimal TF pseudo-labels in the southern part of the target region.

For the U-Net model trained on different pseudo-labels, the results showed that, compared with the rice area extraction errors of the pseudo-labels, the U-Net trained on the TF pseudo-labels increased these errors, whereas they were reduced by using the U-Net trained on the K–RF pseudo-labels (as shown in Figure 10). The main reason for this phenomenon is that even if a transfer model was constructed based on the multi-temporal datasets suitable for transfer between different regions, the model would be inevitably affected by phenological differences between different regions. Therefore, although the accuracy of the TF pseudo-labels in the evaluated counties was higher than that of the K–RF pseudo-labels, more rice misrecognition may exist in these regions that were outside of the assessed counties for the TF pseudo-labels. Consequently, when the TF pseudo-labels were used to train the U-Net model, rice and other land covers would be mixed, and the errors would be further amplified. The process of constructing K–RF pseudo-labels using local data would not be affected by crop phenological differences. Under the condition where training samples were effective, compared with the traditional machine learning model (i.e., RF), the U-Net model learnt global spatial features more easily, including the temporal signature of the rice in the multi-temporal Sentinel-1 datasets, which could be used to further suppress errors in the pseudo-labels.

Finally, for comparison between the U-Net trained on the K–RF pseudo-labels and the two-rule-based rice mapping methods, from the perspective of RMSE and RRMSE, the method proposed by Nguyen et al. [47] and the method proposed by Zhan et al. [48] could extract rice distribution information in the target region. However, from the perspective of extracted rice spatial distribution, there were actually a large number of rice over-identifications in the results of the two methods. As reported in Figure 11, for the northern part of the target region, according to actual experience and statistical year-books, some counties of Shandong Province that are close to northern Jiangsu only have a small number of rice planting areas. For instance, the total rice area in Junan and Linshu (i.e., the region covered by the red circle in the first column of Figure 11) was only 0.23×10^3 hectares in 2019 (i.e., as reported in <http://tjj.linyi.gov.cn/info/1061/9883.htm> (accessed on 20 October 2021)). However, the corresponding results extracted by using the method proposed by Nguyen et al. [47] and the method proposed by Zhan et al. [48] were significantly higher than this value. For the central part of the target region, based on the method proposed by Zhan et al. [48], the result showed that dense rice areas were extracted in the bottom right corner of the central part. In fact, this region is not suitable for rice cultivation, since it is located in the mountainous district (i.e., the region covered by the red circle in the second column of Figure 11). Additionally, the plain in the central part of the target region belongs to an area where rice is intensively planted, and the rice spatial distribution of this location should have a strong regional concentration, such as the results of the U-Net trained on the K–RF pseudo-labels. However, the corresponding results, extracted by using the method proposed by Nguyen et al. [47] and the method proposed by Zhan et al. [48], were shown as irregular scattered rice pixel points. For the southern part of the target region, compared with the northern and central parts of the target region, the rice planting area in the southern part was relatively small. However, the results showed that dense rice planting areas were extracted in this region by using the method proposed by Zhan et al. [48]. On the contrary, the result extracted by using the method proposed by Nguyen et al. [47] went to the other extreme: almost no rice planting information in this region could be extracted.

Although this study could provide new options for applying the deep semantic segmentation model to training sample shortage regions, there were still some drawbacks of the proposed workflow.

Since county-level coverage of the target region in the Sentinel-1 image is small, this study mainly used county-level rice area statistics to verify the reliability of the rice mapping results. However, when the unit to be verified covers a large area in the remote sensing image, the area verification-based method may be inferior to the ground survey verification-based method.

In order to avoid isolated noise points being selected as samples, this study used the homogenization window to construct homogeneous samples in the first part of the K–RF. However, if rice points in the initial clustering results are very sparse, this will result in only a small number of rice points, or even no points being collected in the sample sets.

For the construction of pseudo-labels, since the backscattering temporal signals and rice flooding phenomenon has contributed to rice mapping [48,50,51], this study directly used the difference of backscattering characteristics between adjacent observation points to highlight the relative changes of rice backscattering characteristics during the growth cycle. However, if more observation time points are missing, the span of the backscattering coefficients corresponding to different crops at adjacent observation points will be close. Consequently, this may cause confusion between rice and non-rice extracted by the K–RF, when discontinuous multi-temporal datasets are used.

5. Conclusions

This study proposed a workflow that could apply a DCNN model to training sample shortage regions. Based on the optimal multi-temporal Sentinel-1 images summarized by using the global search method, pseudo-labels were collected from rice mapping results of the K–RF (i.e., the combination of K-Means and RF classifier). As a comparison, the optimal well-trained transfer model constructed from the reference region was used to generate pseudo-labels. Moreover, the two types of pseudo-labels were used for U-Net model-training to extract rice-planting information in the three parts of the target region (i.e., the training sample shortage regions of Jiangsu Province). Rice area statistics were used to validate the effectiveness of DCNN model trained on the pseudo-labels, and a comparison with a rule-based method was used to validate the advantages of the DCNN model trained on the pseudo-labels. The main conclusions are described as follows:

- (1) Based on the CDL data and the multi-temporal Sentinel-1 images, through global search, it can be concluded that multi-temporal datasets acquired from the sowing period to the mature period were the optimal datasets for local rice mapping, when the rice mapping accuracy, user accuracy, kappa and F1 were 84.63%, 94.98%, 0.84, and 0.90, respectively;
- (2) When the latitudes between the target and reference regions were significantly different, compared with multi-temporal datasets containing the whole rice-growth stages, the datasets that only contained middle to late growth stages shared by rice in different regions were more suitable for the transfer model;
- (3) For the U-Net model trained on pseudo-labels, compared with the U-Net trained on the TF pseudo-labels, the model trained on the K–RF pseudo-labels could further correct the errors contained in the pseudo-labels;
- (4) Compared with the two-rule-based methods, whether from the perspective of extracted rice area error or the perspective of extracted rice spatial distribution information, the U-Net trained on the K–RF pseudo-labels had obvious advantages.

In future work, new features—which are more stable to describe rice phenological characteristics in different regions and the differences between rice and other land covers—need to be constructed, so that enough high-confidence pseudo-labels can be easily collected, and the process of applying the DCNN model to actual large-scale rice mapping can be simplified.

Author Contributions: Conceptualization and resources, J.H.; methodology and software, P.W.; validation, P.W. and R.H.; writing—original draft preparation, P.W.; writing—review and editing, R.H., T.L. and J.H.; supervision, R.H., T.L. and J.H.; project administration, J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 42171314). It was also funded by Erasmus+Programme of the European Union (No. 598838-EPP-1-2018-EL-EPPKA2-CBHE-JP).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the European Space Agency (ESA) for providing the multi-temporal Sentinel-1 datasets. Many thanks are also given to the USDA-NASS for providing the Cropland Data Layer.

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

References

- Sharma, A.; Liu, X.; Yang, X.; Shi, D. A patch-based convolutional neural network for remote sensing image classification. *Neural Netw.* **2017**, *95*, 19–28. [\[CrossRef\]](#)
- Tian, F.; Wu, B.; Zeng, H.; Zhang, X.; Xu, J. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth Engine Cloud Platform. *Remote Sens.* **2019**, *11*, 629. [\[CrossRef\]](#)
- Phan, A.; Ha, D.N.; Man, C.D.; Nguyen, T.T.; Bui, H.Q.; Nguyen, T.T. Rapid Assessment of Flood Inundation and Damaged Rice Area in Red River Delta from Sentinel 1A Imagery. *Remote Sens.* **2019**, *11*, 2034. [\[CrossRef\]](#)
- Xu, J.; Yang, J.; Xiong, X.; Li, H.; Huang, J.; Ting, K.C.; Ying, Y.; Lin, T. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [\[CrossRef\]](#)
- Lasko, K.; Vadrevu, K.P.; Tran, V.T.; Justice, C. Mapping Double and Single Crop Paddy Rice With Sentinel-1A at Varying Spatial Scales and Polarizations in Hanoi, Vietnam. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 498–512. [\[CrossRef\]](#) [\[PubMed\]](#)
- Singha, M.; Dong, J.; Zhang, G.; Xiao, X. High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Sci. Data* **2019**, *6*, 24–31. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, X.; Zhai, H.; Shen, Y.; Lou, B.; Jiang, C.; Li, T.; Hussain, S.B.; Shen, G. Large-Scale Crop Mapping From Multisource Remote Sensing Images in Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 414–427. [\[CrossRef\]](#)
- Mansaray, L.R.; Wang, F.; Huang, J.; Yang, L.; Kanu, A.S. Accuracies of support vector machine and random forest in rice mapping with Sentinel-1A, Landsat-8 and Sentinel-2A datasets. *Geocarto Int.* **2020**, *35*, 1088–1108. [\[CrossRef\]](#)
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [\[CrossRef\]](#)
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
- Chai, D.; Newsam, S.; Zhang, H.K.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316. [\[CrossRef\]](#)
- Qiu, C.; Schmitt, M.; Geiß, C.; Chen, T.-H.K.; Zhu, X.X. A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 152–170. [\[CrossRef\]](#)
- Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [\[CrossRef\]](#)
- Thorp, K.R.; Drajat, D. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia. *Remote Sens. Environ.* **2021**, *265*, 112679. [\[CrossRef\]](#)
- Parente, L.; Taquary, E.; Silva, A.P.; Souza, C.M.; Ferreira, L.G. Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data. *Remote Sens.* **2019**, *11*, 2881. [\[CrossRef\]](#)
- Gargiulo, M.; Dell’Aglia, D.A.G.; Iodice, A.; Riccio, D.; Ruello, G. Integration of Sentinel-1 and Sentinel-2 Data for Land Cover Mapping Using W-Net. *Sensors* **2020**, *20*, 2969. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, D.; Pan, Y.; Zhang, J.; Hu, T.; Zhao, J.; Li, N.; Chen, Q. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sens. Environ.* **2020**, *247*, 111912. [\[CrossRef\]](#)
- Wei, P.; Chai, D.; Lin, T.; Tang, C.; Du, M.; Huang, J. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 198–214. [\[CrossRef\]](#)
- Ni, R.; Tian, J.; Li, X.; Yin, D.; Li, J.; Gong, H.; Zhang, J.; Zhu, L.; Wu, D. An enhanced pixel-based phenological feature for accurate paddy rice mapping with Sentinel-2 imagery in Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 282–296. [\[CrossRef\]](#)

20. You, N.; Dong, J.; Huang, J.; Du, G.; Zhang, G.; He, Y.; Yang, T.; Di, Y.; Xiao, X. The 10-m crop type maps in Northeast China during 2017–2019. *Sci. Data* **2021**, *8*, 41. [\[CrossRef\]](#)
21. Jiang, X.; Fang, S.; Huang, X.; Liu, Y.; Guo, L. Rice Mapping and Growth Monitoring Based on Time Series GF-6 Images and Red-Edge Bands. *Remote Sens.* **2021**, *13*, 579. [\[CrossRef\]](#)
22. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sens.* **2019**, *11*, 68. [\[CrossRef\]](#)
23. Pan, Z.; Xu, J.; Guo, Y.; Hu, Y.; Wang, G. Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. *Remote Sens.* **2020**, *12*, 1574. [\[CrossRef\]](#)
24. Pang, J.; Zhang, R.; Yu, B.; Liao, M.; Lv, J.; Xie, L.; Li, S.; Zhan, J. Pixel-level rice planting information monitoring in Fujin City based on time-series SAR imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102551. [\[CrossRef\]](#)
25. Paris, C.; Bruzzone, L. A Novel Approach to the Unsupervised Extraction of Reliable Training Samples From Thematic Products. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1930–1948. [\[CrossRef\]](#)
26. Zhu, A.-X.; Zhao, F.-H.; Pan, H.-B.; Liu, J.-Z. Mapping Rice Paddy Distribution Using Remote Sensing by Coupling Deep Learning with Phenological Characteristics. *Remote Sens.* **2021**, *13*, 1360. [\[CrossRef\]](#)
27. Zhang, C.; Zhang, H.; Zhang, L. Spatial domain bridge transfer: An automated paddy rice mapping method with no training data required and decreased image inputs for the large cloudy area. *Comput. Electron. Agric.* **2021**, *181*, 105978. [\[CrossRef\]](#)
28. Yang, G.; Yu, W.; Yao, X.; Zheng, H.; Cao, Q.; Zhu, Y.; Cao, W.; Cheng, T. AGTOC: A novel approach to winter wheat mapping by automatic generation of training samples and one-class classification on Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102446. [\[CrossRef\]](#)
29. Hao, P.; Di, L.; Zhang, C.; Guo, L. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Sci. Total Environ.* **2020**, *733*, 138869. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Lin, T. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [\[CrossRef\]](#)
31. Ge, S.; Zhang, J.; Pan, Y.; Yang, Z.; Zhu, S. Transferable deep learning model based on the phenological matching principle for mapping crop extent. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102451. [\[CrossRef\]](#)
32. Zhang, W.; Liu, H.; Wu, W.; Zhan, L.; Wei, J. Mapping rice paddy based on machine learning with Sentinel-2 multi-temporal data: Model comparison and transferability. *Remote Sens.* **2020**, *12*, 1620. [\[CrossRef\]](#)
33. Arthur, D.; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, PA, USA, 2007; pp. 1027–1035. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.360.7427&rep=rep1&type=pdf> (accessed on 1 September 2020).
34. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
35. Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolking, S.; Li, C.; Salas, W.; Moore, B. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens. Environ.* **2005**, *95*, 480–492. [\[CrossRef\]](#)
36. Liu, L.; Huang, J.; Xiong, Q.; Zhang, H.; Song, P.; Huang, Y.; Dou, Y.; Wang, X. Optimal MODIS data processing for accurate multi-year paddy rice area mapping in China. *GISci Remote Sens.* **2020**, *57*, 687–703. [\[CrossRef\]](#)
37. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624. [\[CrossRef\]](#)
38. Hu, Q.; Yin, H.; Friedl, M.A.; You, L.; Li, Z.; Tang, H.; Wu, W. Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates. *Remote Sens. Environ.* **2021**, *258*, 112365. [\[CrossRef\]](#)
39. USDA National Agricultural Statistics Service. Cropland Data Layer. 2020. Available online: https://www.nass.usda.gov/Research_and_Science/ (accessed on 24 January 2020).
40. Boryan, C.; Yang, Z.; Mueller, R.; Craig, M. Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* **2011**, *26*, 341–358. [\[CrossRef\]](#)
41. Ashourloo, D.; Shahrabadi, H.S.; Azadbakht, M.; Rad, A.M.; Aghighi, H.; Radiom, S. A novel method for automatic potato mapping using time series of Sentinel-2 images. *Comput. Electron. Agric.* **2020**, *175*, 105583. [\[CrossRef\]](#)
42. Yaramasu, R.; Bandaru, V.; Pnvr, K. Pre-season crop type mapping using deep neural networks. *Comput. Electron. Agric.* **2020**, *176*, 105664. [\[CrossRef\]](#)
43. Sun, L.; Gao, F.; Xie, D.; Anderson, M.; Chen, R.; Yang, Y.; Yang, Y.; Chen, Z. Reconstructing daily 30 m NDVI over complex agricultural landscapes using a crop reference curve approach. *Remote Sens. Environ.* **2021**, *253*, 112156. [\[CrossRef\]](#)
44. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. pp. 234–241. Available online: <https://arxiv.org/pdf/1505.04597.pdf> (accessed on 20 October 2020).
45. Chai, D.; Newsam, S.; Huang, J. Aerial image semantic segmentation using DCNN predicted distance maps. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 309–322. [\[CrossRef\]](#)
46. Samani, N.; Gohari-Moghadam, M.; Safavi, A.A. A simple neural network model for the determination of aquifer parameters. *J. Hydrol.* **2007**, *340*, 1–11. [\[CrossRef\]](#)
47. Nguyen, D.B.; Gruber, A.; Wagner, W. Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data. *Remote Sens. Lett.* **2016**, *7*, 1209–1218. [\[CrossRef\]](#)

-
48. Zhan, P.; Zhu, W.; Li, N. An automated rice mapping method based on flooding signals in synthetic aperture radar time series. *Remote Sens. Environ.* **2021**, *252*, 112112. [[CrossRef](#)]
 49. Dong, J.; Xiao, X.; Kou, W.; Qin, Y.; Zhang, G.; Li, L.; Jin, C.; Zhou, Y.; Wang, J.; Biradar, C.; et al. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. *Remote Sens. Environ.* **2015**, *160*, 99–113. [[CrossRef](#)]
 50. Bazzi, H.; Baghdadi, N.; Hajj, E.; Mohammad, Z.; Mehrez, M.; Tong, D.H. Mapping Paddy Rice Using Sentinel-1 SAR Time Series in Camargue, France. *Remote Sens.* **2019**, *11*, 887. [[CrossRef](#)]
 51. Inoue, S.; Ito, A.; Yonezawa, C. Mapping Paddy Fields in Japan by Using a Sentinel-1 SAR Time Series Supplemented by Sentinel-2 Images on Google Earth Engine. *Remote Sens.* **2020**, *12*, 1622. [[CrossRef](#)]