



Article

Integrating Optical and SAR Time Series Images for Unsupervised Domain Adaptive Crop Mapping

Luwei Feng^{1,2}, Dawei Gui^{3,4}, Shanshan Han^{3,4}, Tianqi Qiu^{3,4}  and Yumiao Wang^{5,*} 

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; lwfeng@whu.edu.cn

² Key Laboratory of Digital Mapping and Land Information Application, Ministry of Natural Resources, Wuhan 430079, China

³ Guangzhou Urban Planning & Design Survey Research Institute Co., Ltd., Guangzhou 510060, China; guidawei@gzpi.com.cn (D.G.); hanshanshan@gzpi.com.cn (S.H.); qtq@whu.edu.cn (T.Q.)

⁴ Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China

⁵ Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China

* Correspondence: wymfrank@whu.edu.cn; Tel.: +86-135-6749-9442

Abstract: Accurate crop mapping is crucial for ensuring food security. Recently, many studies have developed diverse crop mapping models based on deep learning. However, these models generally rely on a large amount of labeled crop samples to investigate the intricate relationship between the crop types of the samples and the corresponding remote sensing features. Moreover, their efficacy is often compromised when applied to other areas owing to the disparities between source and target data. To address this issue, a new multi-modal deep adaptation crop classification network (MDACCN) was proposed in this study. Specifically, MDACCN synergistically exploits time series optical and SAR images using a middle fusion strategy to achieve good classification capacity. Additionally, local maximum mean discrepancy (LMMD) is embedded into the model to measure and decrease domain discrepancies between source and target domains. As a result, a well-trained model in a source domain can still maintain satisfactory accuracy when applied to a target domain. In the training process, MDACCN incorporates the labeled samples from a source domain and unlabeled samples from a target domain. When it comes to the inference process, only unlabeled samples of the target domain are required. To assess the validity of the proposed model, Arkansas State in the United States was chosen as the source domain, and Heilongjiang Province in China was selected as the target domain. Supervised deep learning and traditional machine learning models were chosen as comparison models. The results indicated that the MDACCN achieved inspiring performance in the target domain, surpassing other models with overall accuracy, Kappa, and a macro-averaged F1 score of 0.878, 0.810, and 0.746, respectively. In addition, the crop-type maps produced by the MDACCN exhibited greater consistency with the reference maps. Moreover, the integration of optical and SAR features exhibited a substantial improvement of the model in the target domain compared with using single-modal features. This study indicated the considerable potential of combining multi-modal remote sensing data and an unsupervised domain adaptive approach to provide reliable crop distribution information in areas where labeled samples are missing.

Keywords: agriculture monitoring; remote sensing; machine learning; spatial analysis; multi-modal



Citation: Feng, L.; Gui, D.; Han, S.; Qiu, T.; Wang, Y. Integrating Optical and SAR Time Series Images for Unsupervised Domain Adaptive Crop Mapping. *Remote Sens.* **2024**, *16*, 1464. <https://doi.org/10.3390/rs16081464>

Academic Editor: João Catalão Fernandes

Received: 28 February 2024

Revised: 16 April 2024

Accepted: 18 April 2024

Published: 20 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the growing population, extreme weather events, and regional conflicts, food security has become a hot issue in countries and regions around the world. Precise mapping of crops plays a critical role in grasping the agricultural land utilization and monitoring crop growth, enabling decision-makers to formulate effective agricultural policies for ensuring food security [1,2]. Additionally, accurate planting information can assist in agricultural

insurance, land leasing, and farmland management [3,4]. Hence, it is of tremendous significance to conduct precise crop-type mapping in large-scale areas.

Remote sensing is commonly used in large-scale crop mapping owing to its benefits of rapid data acquisition, wide coverage, and cost-effectiveness. Optical remote sensing images contain spectral information that is highly correlated with crop physiology and morphological characteristics, and they have become the primary important data source for crop mapping [5,6]. In particular, time series optical images are superior to single optical images because they can capture differences in physiological characteristics between crops within a growing season [7,8]. Synthetic Aperture Radar (SAR) is another important remote sensing data source, which is capable of all-weather imaging using microwaves with longer wavelengths than optical sensors. It can provide rich information on crop structure and moisture content, and it has been extensively applied in crop classification [1,9]. For example, time series SAR imagery has been successfully utilized to detect changes in plant canopy structures to identify rice [10]. Given the large differences between SAR and optical imagery, the combined use of these two modalities is able to provide highly complementary information [11–14] and thus exhibits significant potential in crop mapping [15–17].

The essence of remote sensing-based crop classification lies in establishing the relationship between crop types and their corresponding remote sensing features. Machine learning approaches have the advantage of handling complex and huge data, and they have been widely used in exploring optical and SAR time series images for crop mapping [18–21]. For example, Random Forest (RF) exhibited good performance in extracting important features from the two-modal data for crop classification [22,23]. Recently, deep learning methods, which involve deeper neural networks and more intricate architectures, have led to significant breakthroughs in agricultural remote sensing [24]. Many studies have demonstrated the superiority of deep learning models over traditional machine learning approaches in crop mapping [12,25,26]. The deep learning models proposed to synergistically utilize multi-modal data can be broadly categorized into three groups according to the fusion strategies. (1) Early fusion: the original data of the two modalities are directly interpolated and stacked into a single sequence and then processed by a network [27,28]. (2) Middle fusion: the high-level features of the two modalities are extracted independently and then concatenated for classification by a network [29,30]. (3) Decision fusion: the two modalities are processed independently by two similar networks and the classification decisions are averaged or weighted to obtain the final result [31]. However, these fusion strategies have advantages and disadvantages. For example, early fusion is the simplest fusion form, but it usually requires interpolation to concatenate optical and SAR images [32], which may add additional computation and give rise to information redundancy. Decision fusion can obtain more reliable performance by incorporating two classification results [31], but it requires more parameters and is prohibitively costly for complex computation [32]. Middle fusion is able to obtain important information from multi-modal data and reduce data redundancy [33]. Ignoring the complex structure, several studies have shown that middle fusion methods are preferable to the other fusion methods in synergistically utilizing multi-modal data [2,32,34].

Although deep learning models have exhibited good performance in crop mapping using optical and SAR data, most models are supervised models that need abundant labeled samples for model training. However, collecting ground samples can be a labor-intensive and time-consuming task. Transfer learning aims to leverage the knowledge acquired from a previous task to improve the learning performance of a related task, which can reduce or avoid data-collecting efforts. Applying models trained in regions with ample labeled samples to other regions directly has been regarded as a straightforward transfer approach [35]. Specifically, the United States (U.S.) is often chosen as a domain with abundant samples owing to its open source and extensively utilized Cropland Data Layer (CDL). For instance, Hao et al. (2020) collected high-confidence samples from the CDL and corresponding time series images to construct an RF model and then used the model to classify crops in three target regions [36]. However, meteorology-induced phenological

variations can lead to differences in spectral information for a specific crop across different domains. This disparity between the data distributions of the source and target domains is commonly referred to as domain shift, which leads to a performance decline when applying a trained model from the source domain to the target domain [3,37]. Fine-tuning is regarded as a commonly used transfer learning technique that aims to mitigate the problem of domain shift. It utilizes a small number of labeled samples from the target domain to fine-tune the entire or specific parts of a trained model, allowing it to adapt to a new task. Several studies have adopted fine-tuning in crop classification and achieved good results [38–40]. However, overfitting can occur when using a small dataset for fine-tuning. Compared with the transfer approaches mentioned above, unsupervised domain adaptation (UDA) methods migrate knowledge obtained from a source domain, which has a substantial number of labeled samples, to a target domain that only has unlabeled data. UDA methods have been successfully applied to many fields, such as computer vision and signal recognition [41–43]. Very recently, UDA has been explored in crop mapping [44,45]. For example, Wang et al. (2022) designed a UDA model for crop classification using remote sensing data and achieved promising classification accuracies [46]. However, they focused on field-level crop mapping using only optical images. Considering the effectiveness of UDA methods in transfer learning, it is necessary to further explore their potential for crop mapping in large areas using multi-modal remote sensing data.

This study designed a multi-modal deep adaptation crop classification network (MDACCN). The proposed model aims to solve the problem of missing labeled samples in the target domain based on UDA, and it improves the accuracy of crop classification by synergistically utilizing optical and SAR images. Arkansas State in the U.S. and Heilongjiang (HLJ) Province in China were selected as the source domain and target domain, respectively. Experiments were designed to verify the effect of MDACCN by comparing it with two supervised models. The impacts of different combinations of multi-modal data and different fusion schemes were also evaluated. In addition, the constructed models were employed to conduct crop mapping in different regions, and the superior performance of MDACCN was further elucidated. The main contributions of this study are summarized as follows. (1) For the first time, we combined the UDA and multi-modal remote sensing images for unsupervised crop mapping. (2) We designed a middle fusion framework with attention modules to exploit time series optical and SAR data, which exhibited better performance than those of early fusion and decision fusion. (3) The MDACCN achieved inspiring results in actual crop mapping, showing that it has great potential to provide reliable crop distribution information in areas where labeled samples are lacking.

2. Materials

2.1. Study Area

Both the U.S. and China are crucial agricultural producers on a global scale [47]. The U.S. conducts agricultural surveys annually to obtain sufficient crop samples and then combines satellite remote sensing data and machine learning models to publish a crop-type map covering the continental U.S. [48]. However, there is no official crop-type map for China. Therefore, this study explored the capabilities of the proposed model for crop mapping in China using labeled samples from the U.S. Specifically, corn, soybeans, and rice were chosen, as they are the main crops commonly cultivated in both countries and have approximate phenological periods that pose challenges for classification. Arkansas in the U.S. (Figure 1a) was selected as the source domain and Heilongjiang (HLJ) in China (Figure 1b) as the target domain because these regions widely cultivate the three selected crops. HLJ is the biggest crop-growing province in China [49]. It has annual precipitation of 380–600 mm and an average annual temperature of 6.01 °C [50]. Arkansas (AR) is a state in the southern U.S., located in the middle of the Mississippi River. It has hot, humid summers, whereas the winters are dry and cool. Annual precipitation throughout the state averages between 1140 and 1292 mm. The phenological calendar of the three crops is

shown in Figure 2, and the entire growing period for most crops is mainly from April to October.

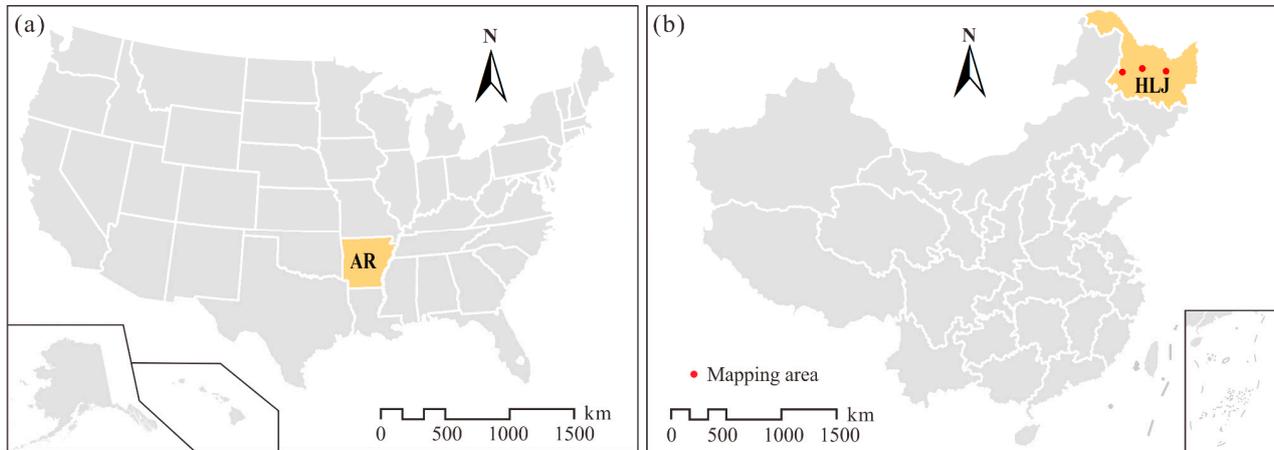


Figure 1. Study areas (a) Arkansas State and (b) HLJ Province.

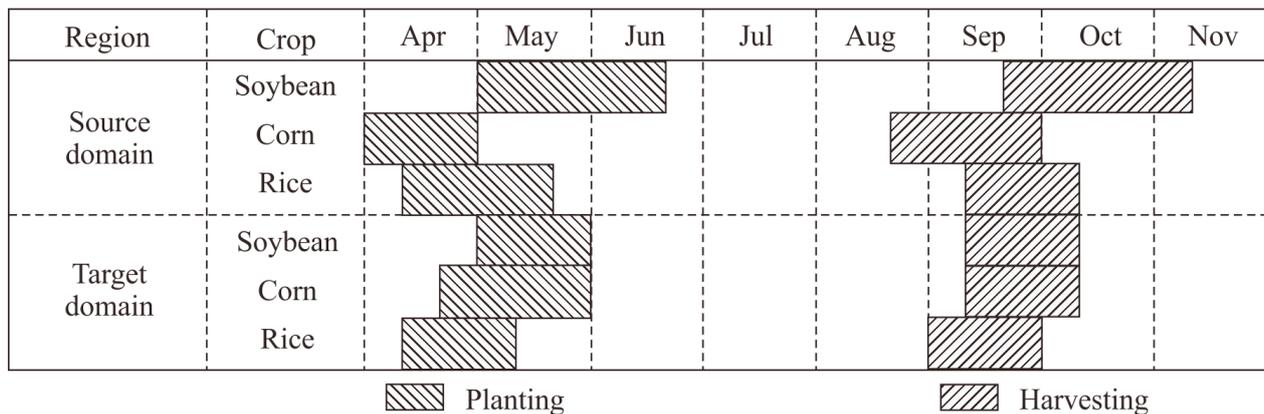


Figure 2. Crop planting and harvesting date.

2.2. Data Acquisition and Preprocessing

2.2.1. Ground Truth

We conducted a field survey in HLJ in 2019 to collect crop samples using a mobile GPS device. Each crop sample contains the crop type and geographic coordinates. Besides corn, soybean, and rice, other categories are grouped as “others”. A total of 1764 labeled samples were collected, including 458 corn, 874 soybean, 370 rice, and 62 “others”. These samples were primarily used to assess the transferability performance of the models. Owing to the lack of field samples in Arkansas, we employed CDL as an alternative to ground truth data. CDL is a high-quality land cover map that focuses on specific crops, providing a spatial resolution of 30 m, annual updates, and nationwide coverage; therefore, it has gained extensive usage as reference data in the field of crop classification [37,48,51,52]. It also offers a confidence layer that indicates the predicted confidence level for each pixel. We set 95% confidence to filter the CDL map from 2019 to improve the quality of sampling. Additionally, the European Space Agency (ESA) WorldCover 2020 map was used to mask the non-crop land. Finally, we randomly selected 6000 labeled samples for each crop type in Arkansas, totaling 24,000 samples for model training and testing.

2.2.2. Remote Sensing Images

Sentinel-2 and Sentinel-1 are two key missions in a series of satellite missions initiated by the ESA to support global environmental monitoring and resource management.

Sentinel-1 mission consists of a pair of polar-orbiting satellites that provide high-resolution, all-weather imaging of the Earth's surface. It includes C-band imaging operating in four modes with different resolutions (up to 5 m). Sentinel-2 is a multi-spectral imaging system that provides high-resolution imaging of the Earth's surface with 13 bands. It also comprises two satellites, each with a spatial resolution of up to 10 m. These two satellites can be combined to provide full coverage of the surface of the earth every five-day interval.

This study selected Sentinel-1 IW images and Sentinel-2 Level-2A surface reflectance images (bottom-of-atmosphere) as SAR and optical data sources, respectively. Sentinel-1 images were processed to produce a calibrated, ortho-corrected product, and Sentinel-2 images were atmospherically corrected. Ten bands of Sentinel-2 were selected as spectral features, including three visible bands, one Near-Infrared (NIR) band, four Red-edge bands, and two short-wave infrared (SWIR) bands. Two bands of Sentinel-1 were used as SAR features, including VH and VV bands.

2.2.3. Data Preprocessing

The remote sensing observations for the crop samples were collected between April and October according to the crop calendars of the three main crops. For Sentinel-2 images, 10-day composites were produced by taking the median of the remaining observations after cloud removal [53,54]. Further, linear interpolation and the Savitzky–Golay algorithm were employed to fill gaps and smooth outliers in the resulting time series images, respectively [55]. The above processes were also conducted for Sentinel-1 images, but the median values were replaced by the average values to composite the images [56]. In the end, a total of 240 features were incorporated into the modeling process, consisting of 12 features with 20 temporal observations. All the necessary steps for data collection and preprocessing were executed using the Google Earth Engine [57].

3. Methodology

3.1. Multi-Modal Deep Adaptation Crop Classification Network (MDACCN)

The framework of the MDACCN is shown in Figure 3. To take advantage of the multi-modal time series data, two branches are designed to process optical and SAR features respectively. In the training process, the MDACCN incorporates two input data streams, one from a labeled source domain and one from an unlabeled target domain. The total loss is computed using forward propagation, and the model parameters are optimized through a gradient descent-based optimization process. In the inference process, the trained network is fed with only the unlabeled target domain samples to obtain class probabilities. The class of the target domain sample is then determined by selecting the class with the highest probability.

Since the middle fusion scheme is employed to design MDACCN, an optical feature extractor and a SAR feature extractor are constructed to process optical features and SAR features, respectively. The two feature extractors share the same structure, and both use the modified ResNet-18 as the extractor to obtain key information from the satellite images of the two modalities. ResNet-18 is a popular ResNet variant that has 18 convolutional layers and consists of five stages [58]. The first stage contains a single convolutional layer followed by a max pooling layer. The four subsequent stages are each composed of two residual blocks with a varying number of filters. A residual block is made up of two convolutional layers and a shortcut connection. These shortcuts can be classified into two types, namely, identity shortcuts and projection shortcuts, depending on whether the input and output have the same dimensions. The five stages are selected to construct the feature extractor in MDACCN. In addition, we embedded the Convolutional Block Attention Module (CBAM) into each residual block to enhance the extractor's focus on important features while suppressing irrelevant features [59]. To perform the convolution operations, the optical and SAR time series features are transformed into matrix form along the time and feature dimensions. Since the size of the SAR matrix is 2×20 , which is too small to operate the reductions among the stages, an interpolation operator is adopted to expand

the size of SAR features to the same as that of the optical features (10×20). The optical matrix and SAR matrix are processed separately by their corresponding feature extractors. The resulting features are then globally averaged and concatenated together before being fed into the classifier for the final output. The classifier in MDACCN is designed as a multilayer perceptron containing two fully connected layers.

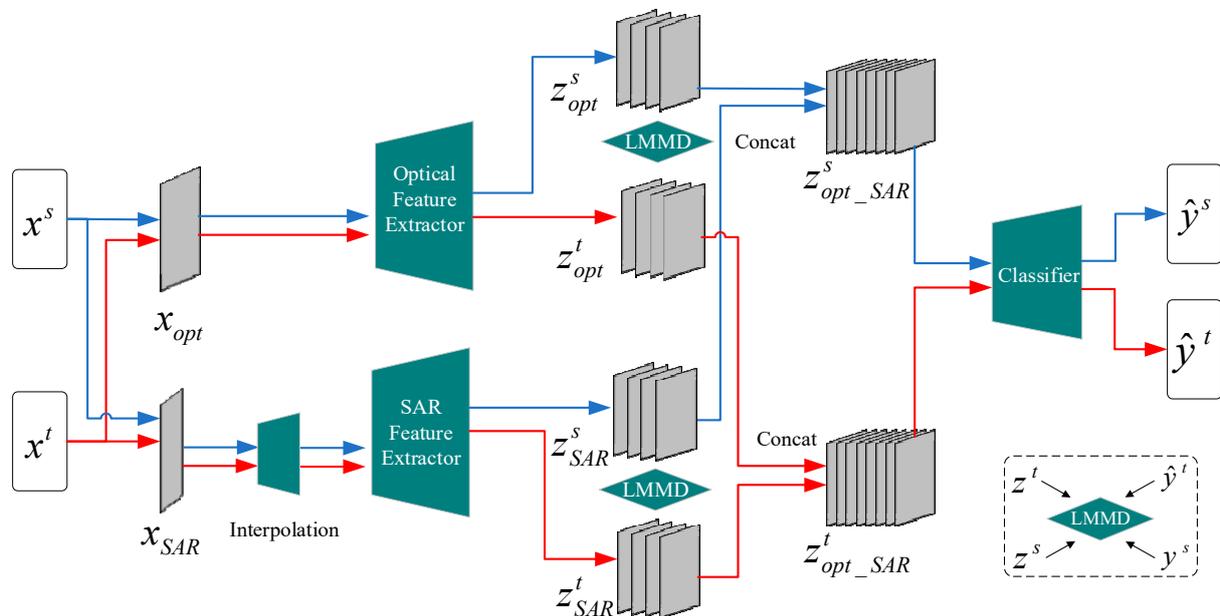


Figure 3. Architecture of MDACCN. x^s and x^t denote the source domain instances s and target domain instances t , and opt and SAR denote the optical and SAR features, respectively. LMMD needs four inputs: the true label y^s , the predicted label y^t , and the activated intermediate features z^s and z^t .

Domain adaptation is the part that distinguishes MDACCN from traditional supervised deep learning models. In this study, $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ with n_s labeled samples is utilized to represent the distribution of the source domain, and $\mathcal{D}_t = \{(x_i^t)\}_{i=1}^{n_t}$ with n_t unlabeled samples is employed to represent the distribution of the target domain. Specifically, x_i indicates the features of a sample i , and y_i means the corresponding type. Generally, \mathcal{D}_s and \mathcal{D}_t are different due to the varied crop phenology across a large area. Therefore, models trained using labeled source domain samples often perform poorly when applied to the target domain directly. To solve this issue, local maximum mean discrepancy (LMMD) [60] was adopted in the MDACCN to measure the shift between the two domains and employ gradient descent to minimize the shift. Specifically, LMMD is an improved extension of MMD that has been adopted in many models for domain adaptation [61,62]. MMD focuses on measuring the global distribution of the two domains but fails to consider the correlations between subdomains within the same category across different domains. LMMD is designed for measuring the distinction between the distributions of related subdomains in both the source domain and target domain, resulting in more precise alignment (Figure 4). The formula of LMMD can be described as follows:

$$d_{\mathcal{H}}(p, q) \triangleq \mathbf{E}_c \left\| \mathbf{E}_{p^{(c)}}[\phi(x^s)] - \mathbf{E}_{q^{(c)}}[\phi(x^t)] \right\|_{\mathcal{H}}^2 \quad (1)$$

where \mathcal{H} is a reproducing kernel Hilbert space equipped with a kernel k . The k is defined as $k(x^s, x^t) = \langle \phi(x^s), \phi(x^t) \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product between two vectors. p and q represent the data distributions of \mathcal{D}_s and \mathcal{D}_t , $p^{(c)}$ and $q^{(c)}$ represent the data distributions of $\mathcal{D}_s^{(c)}$ and $\mathcal{D}_t^{(c)}$, respectively, where c represents the crop type. \mathbf{E} is the mathematical expectation. x^s and x^t are the instances in \mathcal{D}_s and \mathcal{D}_t , respectively. The function ϕ is a feature map that projects the initial samples into the Hilbert space.

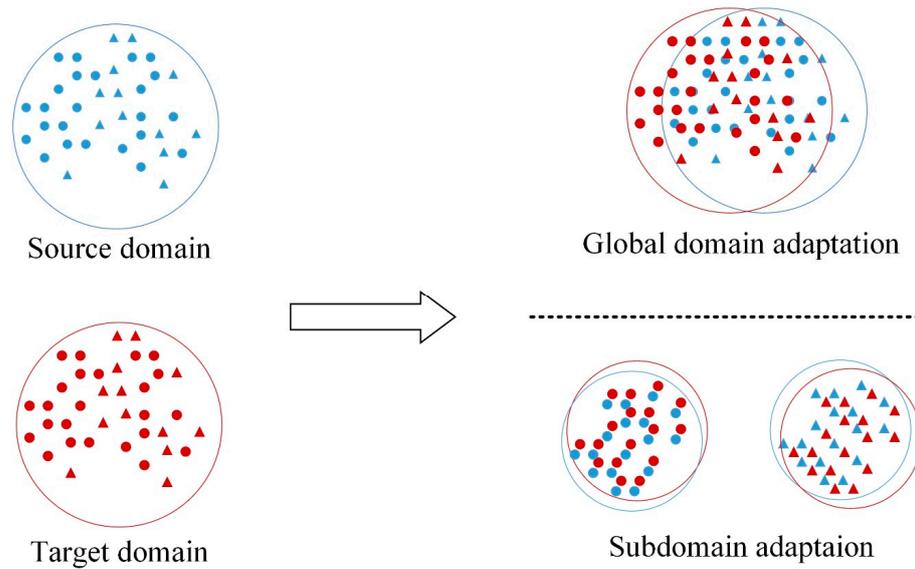


Figure 4. Global domain adaptation and subdomain adaptation (Blue and red represent the source domain and target domain, respectively. Circles and squares represent different categories).

On the assumption that each sample is assigned to a class based on a weight parameter ω_s , an unbiased estimator of Equation (1) can be calculated as follows:

$$d_{\mathcal{H}}(p, q) = \frac{1}{C} \sum_{c=1}^C \left\| \sum_{x_i^s \in D_s} \omega_i^{sc} \phi(x_i^s) - \sum_{x_j^t \in D_t} \omega_j^{tc} \phi(x_j^t) \right\|_{\mathcal{H}}^2 \quad (2)$$

where ω_i^{sc} and ω_j^{tc} are the weights of x_i^s and x_j^t for class c , respectively. ω_i^c of the sample x_i can be computed as follows:

$$\omega_i^c = \frac{y_{ic}}{\sum_{(x_j, y_j) \in D} y_{jc}} \quad (3)$$

where y_{ic} is the c th entry of the y_i .

Nonetheless, in unsupervised adaptation, where the labeled data is lacking in the target domain, it is impossible to calculate ω_j^{tc} directly because y_j^t is unavailable. We observe that the output \hat{y}_i represents a probability distribution that indicates the likelihood of assigning x_i to each of the crop classes. Therefore, we can use \hat{y}_j^t as the substitution for y_j^t to calculate ω_j^{tc} for each target sample. Given activations in the feature extractor layer l as $\{z_i^{sl}\}_{i=1}^{n_s}$ and $\{z_j^{tl}\}_{j=1}^{n_t}$, Equation (2) can be computed as follows:

$$\hat{d}_l(p, q) = \frac{1}{C} \sum_{c=1}^C \left[\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \omega_i^{sc} \omega_j^{sc} k(z_i^{sl}, z_j^{sl}) + \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \omega_i^{tc} \omega_j^{tc} k(z_i^{tl}, z_j^{tl}) - 2 \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \omega_i^{sc} \omega_j^{tc} k(z_i^{sl}, z_j^{tl}) \right] \quad (4)$$

Finally, the total loss of MDACCN contains the domain adaptation loss and classification loss:

$$loss_{total} = \frac{1}{n_s} \sum_{i=1}^{n_s} J(f(x_i^s), y_i^s) + \lambda (\hat{d}_l^{opt}(p, q) + \hat{d}_l^{SAR}(p, q)) \quad (5)$$

where J represents the cross-entropy loss. \hat{d}_l^{opt} and \hat{d}_l^{SAR} are the domain adaptation loss of optical and SAR features, respectively. $\lambda > 0$ is a trade-off parameter.

3.2. Experimental Setting

We carried out a comprehensive comparison to evaluate the performance of the proposed model. First, the accuracies of the MDACCN with different modal inputs were

compared to verify the importance of the multi-modal data. In particular, the corresponding branch in the model was frozen when a particular modal datum was not inputted. Second, the early fusion scheme and decision fusion scheme were used to reconstruct MDACCN to explore the influence of the fusion scheme on model performance. The structures of early- and decision-fusion-based models are described in Figure 5. Finally, the well-trained models were used to generate the predicted crop maps in the target domain, and the maps were compared with the reference maps to assess the mapping performance of the models.

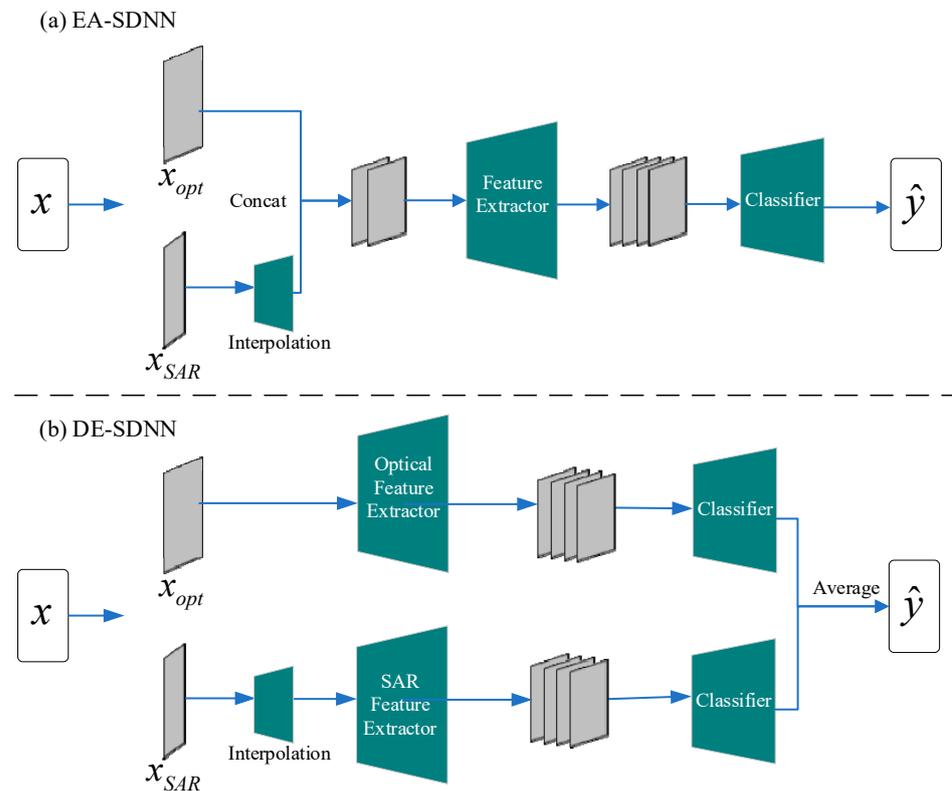


Figure 5. Architectures of the (a) early-fusion-based model and (b) decision-fusion-based model.

The 24,000 labeled samples from the source domain were randomly split into three subsets: 60% for training, 20% for validation, and the rest for testing. An equal number of unlabeled samples was randomly selected from the target domain and added to the training samples from the source domain to train the MDACCN. The 1764 labeled samples from the target domain were used as the testing data to measure the final transferability performance in the target domain. The number of training epochs for the MDACCN was set to 200. According to the result of the validation, the model configuration was determined as follows: a batch size of 256, a learning rate of 0.001, an SGD optimizer, and a trade-off parameter λ set to 0.5. RF and a supervised neural network (SDNN) were selected for comparison. Specifically, the RF was set to have 400 trees, with each having a maximum depth of 20. To ensure a fair comparison, SDNN was designed with a similar architecture to MDACCN, but without the domain adaptation loss component. Experiments were conducted on a computational platform comprising an Intel i7-12700kf CPU, 64 GB RAM, an RTX 3090 GPU, and 2 TB storage. Pytorch 1.10 was used for developing the deep learning models, and Scikit-learn 1.2.1 was employed for implementing the machine learning models. Moreover, model performance was assessed based on confusion matrixes, overall accuracy (OA), Cohen's kappa coefficient (Kappa), and macro-averaged F1 score (F1).

4. Results

4.1. Comparison of Single-Modal Data and Multi-Modal Data

The testing accuracies of the models using single-modal data and multi-modal data were compared, and the results are demonstrated in Table 1. Specifically, the single-modal data include time series features of S1 data (S1) and time series features of S2 data (S2), and the multi-modal data are a combination of S1 and S2 data (S1S2). In the source domain, the models trained using S1 features had significantly lower accuracy compared with the models trained using S2 features, which is consistent with previous studies [14,15]. The RF and the SDNN trained using S1S2 features outperformed the corresponding models trained using S2 features, and the SDNN trained using S1S2 features obtained the best performance among all models, with an OA of 0.957. It is worth noting that the MDACCN has a slightly inferior performance compared with the SDNN in the source domain. This is because the SDNN just focuses on the samples of the source domain, while the MDACCN uses domain adaption to close the distribution of high-level features between the source and target domains to improve the performance in the target domain. In the target domain, it is observed that all models had lower accuracies due to the domain shift. The accuracy degradation of the MDACCN was not as significant as the RF and SDNN. This may be because the MDACCN reduced the domain discrepancy between the source and target domains in the training process, so the good performance of the source domain can be migrated to the target domain. Furthermore, models trained using S1S2 features showed better performance than those trained using only S1 or S2 features. The MDACCN model trained using S1S2 features obtained the highest performance in the target domain, with an OA of 0.878.

Table 1. Classification performance of models using different modal data.

Model	Modal	Source Domain			Target Domain		
		OA	F1	Kappa	OA	F1	Kappa
RF	S1	0.825	0.823	0.767	0.645	0.463	0.396
	S2	0.940	0.939	0.919	0.626	0.520	0.476
	S1S2	0.946	0.946	0.928	0.730	0.602	0.587
SDNN	S1	0.821	0.820	0.761	0.605	0.440	0.343
	S2	0.956	0.955	0.941	0.756	0.601	0.633
	S1S2	0.957	0.957	0.943	0.731	0.594	0.577
MDACCN	S1	0.774	0.769	0.698	0.618	0.478	0.397
	S2	0.946	0.946	0.928	0.843	0.710	0.751
	S1S2	0.938	0.938	0.917	0.878	0.746	0.810

Confusion matrixes were further computed to show the distribution of predictions for each crop type using the testing data. In the source domain (Figure 6), the SDNN trained using S2 features exhibited higher accuracy in almost every class of crop compared with other models. By using S2 or S1S2 features, RF and MDACCN were also capable of accurately identifying the crops, only slightly inferior to the SDNN. However, all models trained using S1 features showed lower accuracy for each crop except rice. In the target domain (Figure 7), all models exhibited a decline in accuracy for each class. The MDACCN trained using S1S2 features was the only model that performed well in recognizing soybean, corn, and rice, with an OA exceeding 0.8, indicating the effectiveness of multi-modal data in transfer learning. It should be noted that all models struggled to accurately identify the “others” category. Although MDACCN had the largest number of samples correctly classified in the “others” category, its accuracy was still below 0.5.

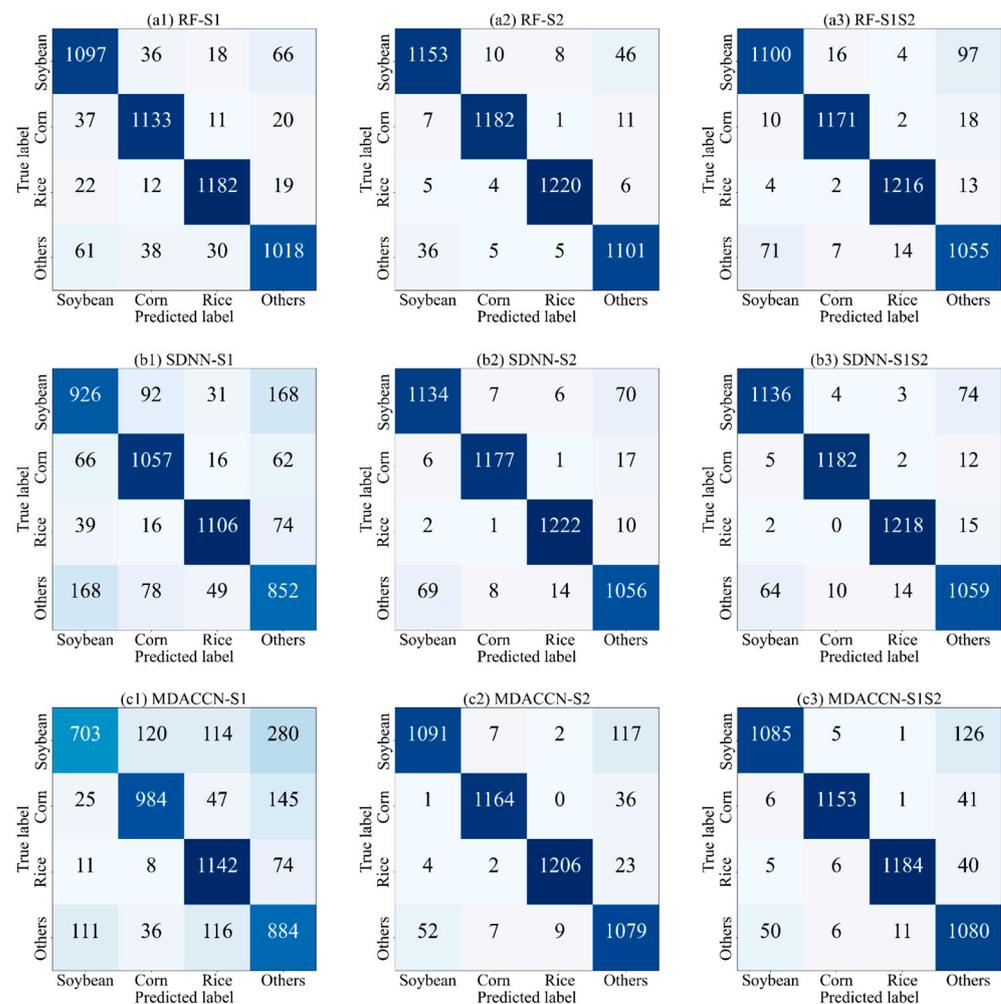


Figure 6. Confusion matrices of (a) RF, (b) SDNN, and (c) MDACCN with (1) S1 features, (2) S2 features, and (3) S1S2 features using the testing data from Arkansas.

4.2. Comparison of Different Fusion Schemes

This experiment aimed to validate the effects of different fusion schemes on classification performance using the testing data. Specifically, the MDACCN and SDNN were compared using early, middle, and decision fusion schemes. The performance of different fusion schemes in both the source and target domains is presented in Table 2. It is obvious that the middle fusion outperformed the other two schemes significantly in terms of each model and each domain. For example, the middle-fusion-based MDACCN had a Kappa that was over 0.1 higher than that of the early- and decision-fusion-based MDACCN. The early- and decision-fusion-based models performed similarly in the source domain, but the decision-fusion-based model showed significant superiority over the early-fusion-based model in the target domain. This may be because both middle-fusion- and decision-fusion-based models have two branches that process SAR and optical features respectively, allowing them to better learn the domain invariant representations of the crops and gain an advantage in transfer applications. Compared with decision fusion, middle fusion integrates the features of different modalities at the middle layers of the model, enabling more effective modal integration and interaction in the early stages of feature extraction. This allows the model to capture more discriminative and complementary features.

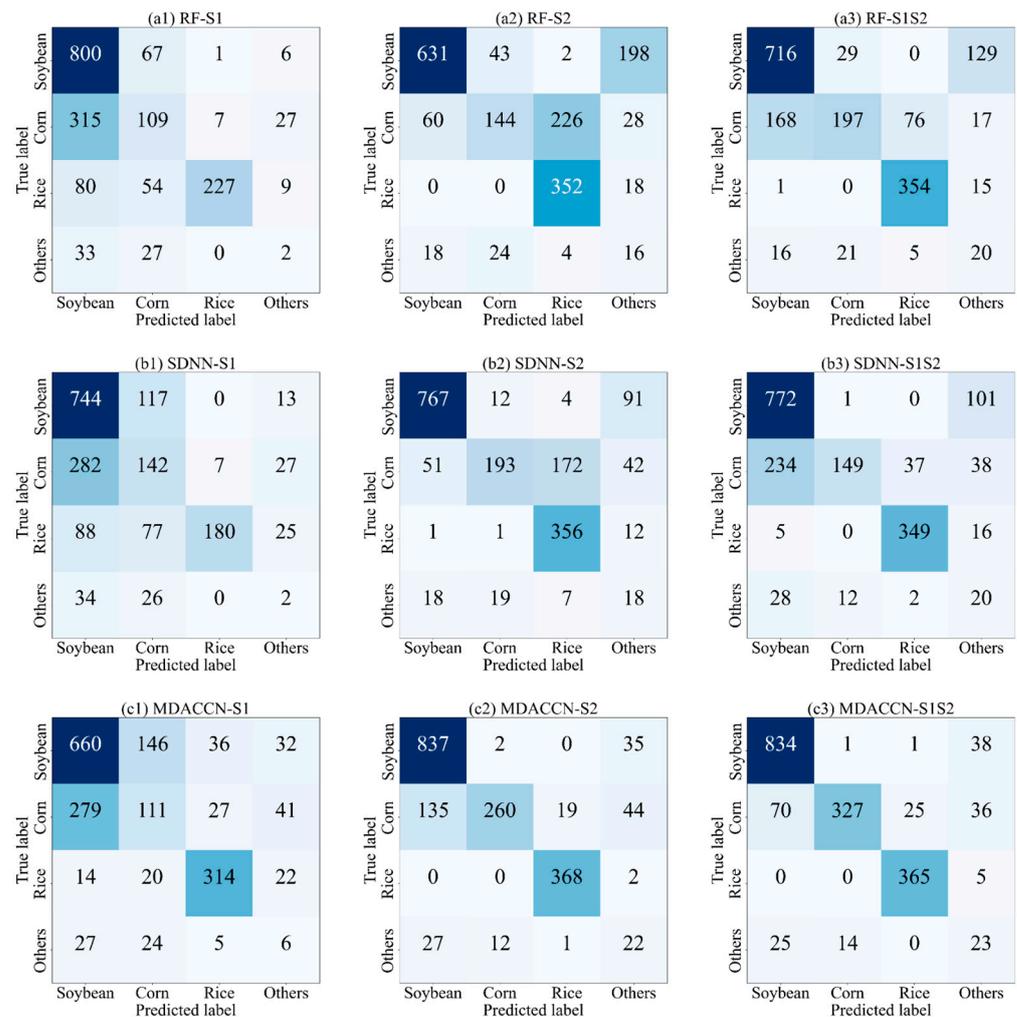


Figure 7. Confusion matrices of (a) RF, (b) SDNN, and (c) MDACCN with (1) S1 features, (2) S2 features, and (3) S1S2 features using the testing data from HLJ.

Table 2. Classification performance of models using different fusion schemes.

Model	Fusion Scheme	Source Domain			Target Domain		
		OA	F1	Kappa	OA	F1	Kappa
MDACCN	Early	0.904	0.903	0.872	0.768	0.591	0.627
	Middle	0.938	0.938	0.917	0.878	0.746	0.810
	Decision	0.890	0.889	0.853	0.832	0.665	0.730
SDNN	Early	0.924	0.923	0.899	0.694	0.509	0.490
	Middle	0.957	0.957	0.943	0.731	0.594	0.577
	Decision	0.945	0.944	0.927	0.701	0.527	0.506

The classification results of the models using three fusion strategies for each crop in the target domain are depicted in Figure 8. It is evident that the models with the middle fusion scheme exhibited superior identification accuracy for each crop compared with other schemes. The decision-fusion-based models demonstrated a clear advantage over early-fusion-based models in corn and “others” identification. Owing to the domain adaptation, the MDACCN showed superior performance over the SDNN in identifying crops in the target domain with each fusion scheme. It is worth noting that the accuracy of the model was still relatively low for corn and “others”. According to Figure 7, the model tended to misclassify the corn as soybean, which may be attributed to the similar phenological

characteristics of the two crops. As for the “others”, the other crop types were grouped as “others” except the main crops. The variation of the “others” between the source and target domains induces low classification accuracy. However, among the models with different fusion strategies, the middle-fusion-based MDACCN proposed in this study achieved the best and most balanced classification results.

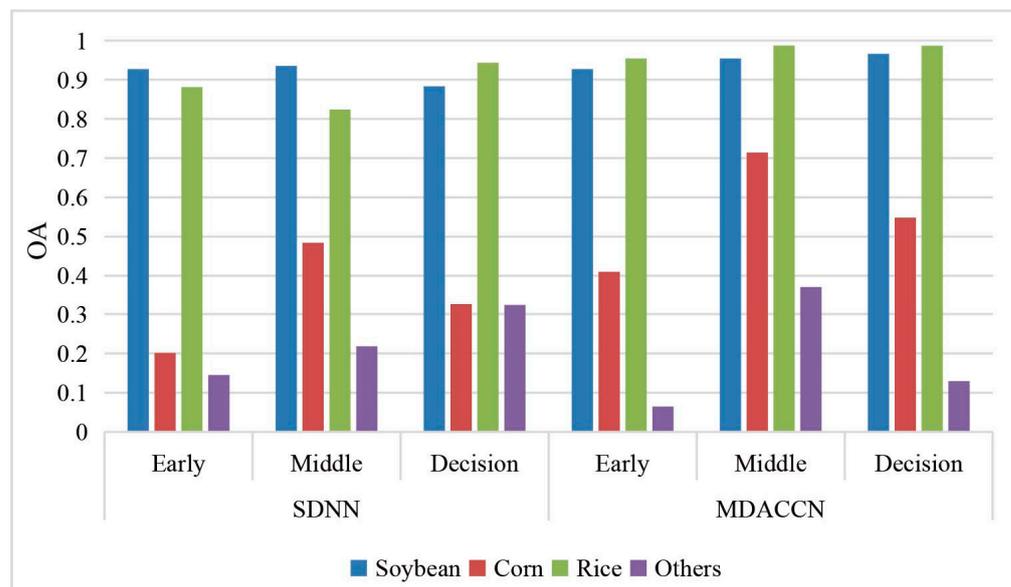


Figure 8. Classification accuracy of the models for each crop with different fusion schemes using the testing data from HLJ Province.

4.3. Crop Mapping Performance

To explore the crop mapping performance in practice, the middle-fusion-based deep learning models were applied to produce the crop-type maps for the three mapping areas (Figure 1b) in the target domain. The center point coordinates of mapping areas (1)–(3) are (43.844N, 124.649E), (46.891N, 129.971E), and (47.720N, 127.208E), respectively. The area of each mapping area is about 100 km². Since there is no official crop distribution map in China, the 2019 crop map in Northeast China provided by You et al. (2021) [55] was utilized as a reference map, which was proved to have high accuracy with an OA of 0.87.

The reference maps and the generated maps of the mapping areas in HLJ Province are shown in Figure 9. The majority of crops in mapping area 1 are corn and rice, primarily distributed in the central and northern regions. The RF and SDNN misclassified some corn as rice in the southern region. In addition, the generated crop maps exhibited pronounced salt-and-pepper noise, with blurry boundaries between crop fields. MDACCN produced a reliable crop map that showed better consistency with the reference map. Rice cultivation is predominant in the southeastern part of mapping area 2, while corn and soybean are widely distributed in other parts. All models successfully captured the general trend in the spatial distribution of crops. However, RF and SDNN models misidentified a considerable portion of corn planting areas as soybean. In mapping area 3, soybean, corn, and rice are widely distributed, among which soybean is the main crop. Although the RF and SDNN were able to accurately map the distribution of rice, they failed to correctly identify corn and soybean. In contrast to RF and SDNN, the crop map generated using MDACCN showed a higher level of consistency with the reference map.

To more accurately evaluate the reliability of the maps produced by the models, we used the labeled samples collected in the mapping areas and calculated the accuracy based on their true and predicted classes (Table 3). The generated map from the RF in mapping area 1 failed to identify the “others” samples. The crop map produced by the SDNN in mapping area 2 failed to assign the correct category to the corn samples. Among these

models, only MDACCN generated reliable crop maps in which most samples had the correct type, proving that the proposed model has good transfer performance.

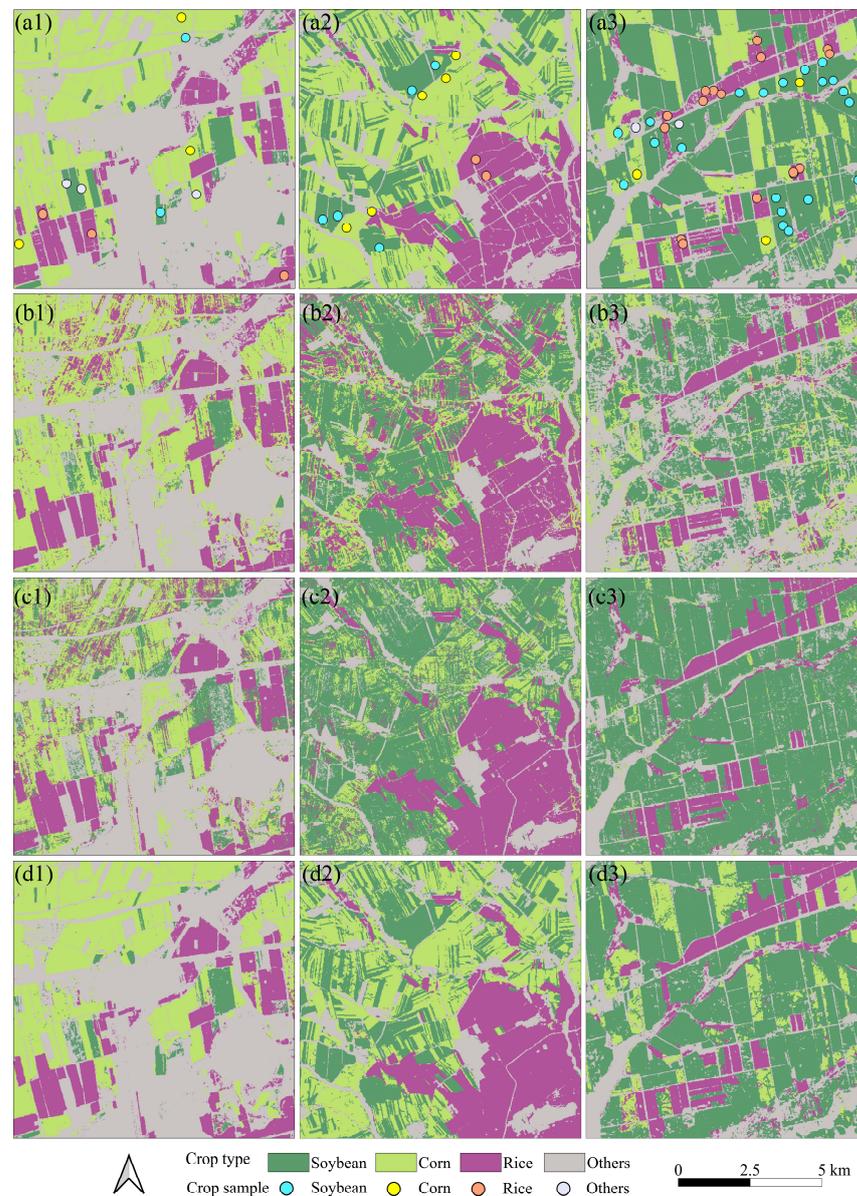


Figure 9. Crop maps from (a) reference maps, (b) RF, (c) SDNN, and (d) MDACCN of the mapping areas (1–3) in HLJ Province.

Table 3. Accuracy of sample recognition on the generated maps. (“-” means that the sample size of the current category is 0, so the accuracy cannot be calculated).

Mapping Area	Model	Soybean	Corn	Rice	Others
1	RF	1	0.667	0.667	0
	SDNN	0.333	1	0.667	0.333
	MDACCN	1	1	1	0.333
2	RF	1	0.6	1	-
	SDNN	1	0.4	1	-
	MDACCN	1	1	1	-
3	RF	0.762	0.667	1	0.941
	SDNN	1	0.333	1	0.750
	MDACCN	1	0.667	1	1

5. Discussion

5.1. Interpretation of the Domain Adaptation in MDACCN

The fundamental distinction between the MDACCN and conventional supervised models is domain adaptation. To assess the impact of the domain adaptation of the MDACCN, t-SNE [63], a widely used dimensionality reduction approach, was employed to visualize and compare the feature distributions before and after the domain adaptation from the source domain and the target domain. Specifically, the distributions of original features inputted to the model and the extracted features from the domain adaptation in the MDACCN were selected for visualization. As MDACCN was constructed with two branches to process optical and SAR features separately, we visualized the changes in the feature distribution for each modal datum individually.

Figure 10 shows the distribution changes of SAR features on each crop using the testing data of the source and target domains. The clear separation of original crop features from the two domains indicates the presence of the distribution shift. This discrepancy is likely a result of variations in agricultural management practices and natural conditions between the domains. This separation may lead to the phenomenon that the same type of crops in two domains are misclassified as different crops, thereby resulting in a decrease in performance when applying the well-trained models to new domains directly. In contrast, the distributions of the extracted features from the source domain and target domain are much closer. This change can be attributed to the domain adaptation, which is designed to alleviate the differences between the feature distributions of the two domains. When the data distributions of the two domains are similar, the model that performs well in the source domain can theoretically achieve good accuracy in the target domain as well, which aligns with the results in Section 4.1. A similar phenomenon is also observed in the distribution of optical features (Figure 11). However, the change in optical features was relatively weaker compared with those of SAR features. This could be due to the greater quantity of features present in optical data, which poses a challenge for domain alignment. Overall, these visualization figures provide intuitive evidence of MDACCN's superiority in mitigating domain shift and offer an explanation for the high crop classification accuracy achieved by the model in the target domain.

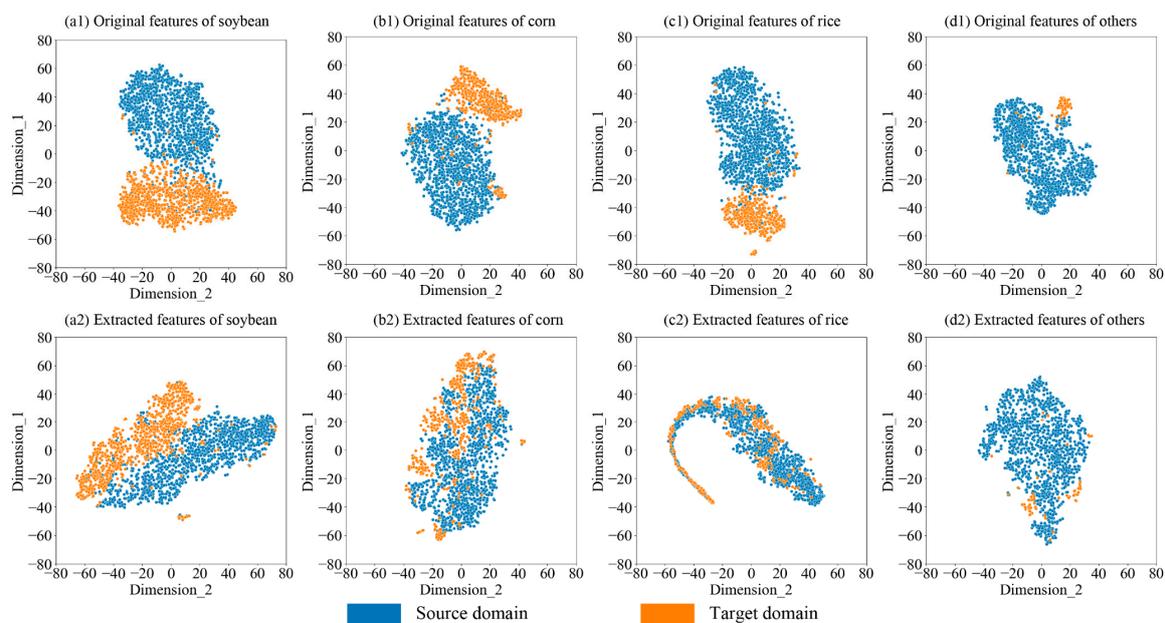


Figure 10. Distributions of (1) original SAR features and (2) extracted SAR features of (a) soybean, (b) corn, (c) rice, and (d) “others” using the testing data of the source and target domains.

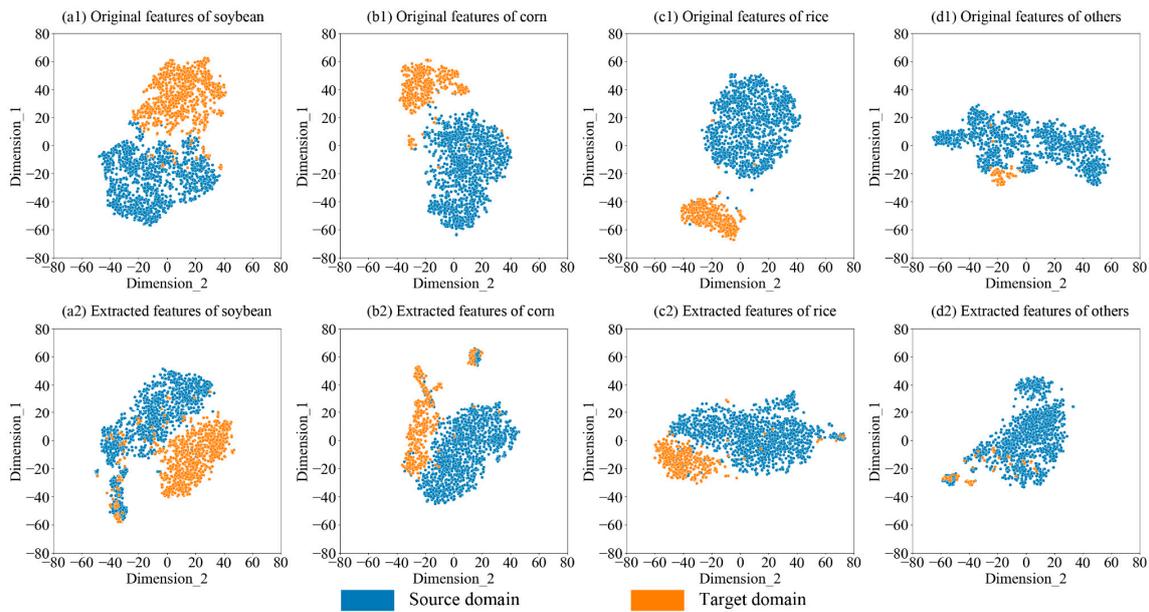


Figure 11. Distributions of (1) original optical features and (2) extracted optical features of (a) soybean, (b) corn, (c) rice, and (d) “others” using the testing data of the source and target domains.

5.2. Impact of Multi-Modal Data

The proposed model MDACCN was designed to fuse the optical features and SAR features for crop mapping. We explored the impact of S1, S2, and the combination of S1 and S2 data on model performance. The results discovered that the lowest accuracy was obtained when only SAR data was used, which is consistent with the results of previous studies [14,15,64]. The previous studies also show that the combination of optical and SAR features resulted in higher classification accuracy compared with using single-modal data. For instance, Van Tricht et al. (2018) employed time series S1 and S2 data to classify eight crop types in Belgium, and their findings revealed that the inclusion of SAR enhanced the overall accuracy from 0.76 to 0.82 [14]. However, the results in the source domain indicated that combining the two modal features did not improve classification accuracy significantly compared with using only optical data. With the additional SAR features, the OA of RF and SDNN only increased by 0.006 and 0.001, while the OA of MDACCN decreased by 0.008. One possible reason is that the classification task in this study is much easier than that in Belgium. Here there are only four crop types, and time series optical features can supply sufficient information to obtain satisfactory performance (OA > 0.94).

In contrast, the results in the target domain showed the benefits of combining multi-modal data. Using optical and SAR features, the OA of the MDACCN was 0.26 better than using only SAR data and 0.035 higher than using only optical data. The discrepancy in the results between the two domains may be due to the domain adaption in the MDACCN. Two key factors influence the accuracy in the target domain. First, the model must perform well in the source domain. Second, the discrepancy between the source and the target domains should be minimized. Although the performances of using multi-modal features and using single optical features were similar in the source domain, multi-modal features can better describe the data distribution of a specific crop and help the LMMD to measure and alleviate the domain discrepancy between the two domains, leading to improved performance in the target domain. In particular, the middle fusion strategy outperformed the early and late fusion strategies significantly in the target domain. The early fusion strategy concatenates the two-modal data first and then applies the LMMD to align the features, while the middle fusion strategy can enhance the feature alignment by aligning the modal data individually and thus improve the classification performance in the target domain. The late fusion strategy individually processes the two-modal data and just

averages the two classification results to obtain the final result. The middle fusion strategy can leverage the complementary information from the two-modal data at the processes of feature extraction and feature learning, leading to superior performance in the target domain. However, Sections 4.1 and 4.2 demonstrated that all models cannot accurately identify the “others” type in the target domain. The reason for the phenomenon may be that the “others” type contains various crops which increases the difficulty of classification.

5.3. Limitations and Future Work

While the proposed model achieved good results, there are still several limitations that require further attention and resolution in the future, especially for global applications. Given the variation in crop types and environmental conditions across different regions, the model cannot be directly applied globally. It is necessary to divide the global data into several target domains, each with similar major crop types and environmental conditions. The proposed approach can then be implemented in each target domain with the support of a source domain that shares the same main crop types to achieve accurate crop mapping. Furthermore, this study used all temporal features within the crop-growth period to construct the model. The large volume of the input not only increases the difficulty of data acquisition and processing but also makes it difficult to generate global-scale crop maps quickly. We will streamline the model input through feature engineering to improve the model efficiency while maintaining its performance for global applications.

6. Conclusions

The unavailability of abundant labeled samples is a major constraint in achieving accurate crop mapping. To address the problem, this study proposed a new unsupervised domain adaptive crop mapping model, MDACCN, utilizing labeled samples from the source domain alongside unlabeled samples from the target domain. The middle fusion strategy was applied to design the structure of the MDACCN for synergistically utilizing the time series optical and SAR data. It was found that the MDACCN significantly outperformed the SDNN and RF model in the target domain, obtaining 0.878 OA, 0.746 F1, and 0.810 Kappa. The proposed model also achieved the most reliable results in actual crop mapping. Compared with the single-modal data, fusing the optical and SAR data can enhance the model’s performance in the target domain. The visualization results of the t-SNE demonstrated that the MDACCN can narrow the distribution discrepancy of a specific crop between the domains, allowing the accurate classification capability in the source domain to be transferred to the target domain. This study designed a novel model to precisely map crops in areas lacking labeled samples, which could greatly benefit scientists and policymakers in managing agriculture production to ensure food security.

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

Funding: This research was funded by the National Natural Science Foundation of China (No. 42201354), the Zhejiang Provincial Natural Science Foundation of China (No. LQ22D010007), the Public Projects of Ningbo City (No. 2022S101, No. 2023S102), the Ningbo Natural Science Foundation (No. 2022J076), the Ningbo Science and Technology Innovation 2025 Major Special Project (No. 2021Z107, No. 2022Z032), the China Postdoctoral Science Foundation (2023M742679), and the Open Research Fund Program of Key Laboratory of Digital Mapping and Land Information Application, Ministry of Natural Resources (ZRZYBWD202306).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to thank the data provider ESA for the Sentinel-1 and Sentinel-2 images.

Conflicts of Interest: Authors Dawei Gui, Shanshan Han and Tianqi Qiu were employed by the Guangzhou Urban Planning & Design Survey Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Bargiel, D. A New Method for Crop Classification Combining Time Series of Radar Images and Crop Phenology Information. *Remote Sens. Environ.* **2017**, *198*, 369–383. [[CrossRef](#)]
- Yuan, Y.; Lin, L.; Zhou, Z.-G.; Jiang, H.; Liu, Q. Bridging Optical and SAR Satellite Image Time Series via Contrastive Feature Extraction for Crop Classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 222–232. [[CrossRef](#)]
- Wang, Y.; Feng, L.; Sun, W.; Zhang, Z.; Zhang, H.; Yang, G.; Meng, X. Exploring the Potential of Multi-Source Unsupervised Domain Adaptation in Crop Mapping Using Sentinel-2 Images. *GIScience Remote Sens.* **2022**, *59*, 2247–2265. [[CrossRef](#)]
- Guo, Y.; Jia, X.; Paull, D.; Benediktsson, J.A. Nomination-Favoured Opinion Pool for Optical-SAR-Synergistic Rice Mapping in Face of Weakened Flooding Signals. *ISPRS J. Photogramm. Remote Sens.* **2019**, *155*, 187–205. [[CrossRef](#)]
- Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
- Liu, C.; Chen, Z.; Shao, Y.; Chen, J.; Hasi, T.; Pan, H. Research Advances of SAR Remote Sensing for Agriculture Applications: A Review. *J. Integr. Agric.* **2019**, *18*, 506–525. [[CrossRef](#)]
- Ashourloo, D.; Shahrabadi, H.S.; Azadbakht, M.; Aghighi, H.; Nematollahi, H.; Alimohammadi, A.; Matkan, A.A. Automatic Canola Mapping Using Time Series of Sentinel 2 Images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 63–76. [[CrossRef](#)]
- Wang, Y.; Zhang, Z.; Feng, L.; Ma, Y.; Du, Q. A New Attention-Based CNN Approach for Crop Mapping Using Time Series Sentinel-2 Images. *Comput. Electron. Agric.* **2021**, *184*, 106090. [[CrossRef](#)]
- McNairn, H.; Kross, A.; Lapen, D.; Caves, R.; Shang, J. Early Season Monitoring of Corn and Soybeans with TerraSAR-X and RADARSAT-2. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 252–259. [[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](#)]
- Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.-F.; Ceschia, E. Understanding the Temporal Behavior of Crops Using Sentinel-1 and Sentinel-2-like Data for Agricultural Applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [[CrossRef](#)]
- Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
- Skakun, S.; Kussul, N.; Shelestov, A.Y.; Lavreniuk, M.; Kussul, O. Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3712–3719. [[CrossRef](#)]
- Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium. *Remote Sens.* **2018**, *10*, 1642. [[CrossRef](#)]
- Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of Crop Types and Crop Sequences with Combined Time Series of Sentinel-1, Sentinel-2 and Landsat 8 Data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [[CrossRef](#)]
- Pott, L.P.; Amado, T.J.C.; Schwalbert, R.A.; Corassa, G.M.; Ciampitti, I.A. Satellite-Based Data Fusion Crop Type Classification and Mapping in Rio Grande Do Sul, Brazil. *ISPRS J. Photogramm. Remote Sens.* **2021**, *176*, 196–210. [[CrossRef](#)]
- Orynbaikyzy, A.; Gessner, U.; Conrad, C. Crop Type Classification Using a Combination of Optical and Radar Remote Sensing Data: A Review. *Int. J. Remote Sens.* **2019**, *40*, 6553–6595. [[CrossRef](#)]
- Salehi, B.; Daneshfar, B.; Davidson, A.M. Accurate Crop-Type Classification Using Multi-Temporal Optical and Multi-Polarization SAR Data in an Object-Based Image Analysis Framework. *Int. J. Remote Sens.* **2017**, *38*, 4130–4155. [[CrossRef](#)]
- Tufail, R.; Ahmad, A.; Javed, M.A.; Ahmad, S.R. A Machine Learning Approach for Accurate Crop Type Mapping Using Combined SAR and Optical Time Series Data. *Adv. Space Res.* **2022**, *69*, 331–346. [[CrossRef](#)]
- Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K. Assessing the Suitability of Data from Sentinel-1A and 2A for Crop Classification. *GIScience Remote Sens.* **2017**, *54*, 918–938. [[CrossRef](#)]
- Cheng, Y.; Yu, L.; Cracknell, A.P.; Gong, P. Oil Palm Mapping Using Landsat and PALSAR: A Case Study in Malaysia. *Int. J. Remote Sens.* **2016**, *37*, 5431–5442. [[CrossRef](#)]
- Li, H.; Zhang, C.; Zhang, S.; Atkinson, P.M. Crop Classification from Full-Year Fully-Polarimetric L-Band UAVSAR Time-Series Using the Random Forest Algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *87*, 102032. [[CrossRef](#)]
- Onojeghuo, A.O.; Blackburn, G.A.; Wang, Q.; Atkinson, P.M.; Kindred, D.; Miao, Y. Mapping Paddy Rice Fields by Applying Machine Learning Algorithms to Multi-Temporal Sentinel-1A and Landsat Data. *Int. J. Remote Sens.* **2018**, *39*, 1042–1067. [[CrossRef](#)]
- Feng, L.; Wang, Y.; Zhang, Z.; Du, Q. Geographically and Temporally Weighted Neural Network for Winter Wheat Yield Prediction. *Remote Sens. Environ.* **2021**, *262*, 112514. [[CrossRef](#)]

25. Rußwurm, M.; Körner, M. Self-Attention for Raw Optical Satellite Time Series Classification. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 421–435. [[CrossRef](#)]
26. Zhong, L.; Hu, L.; Zhou, H. Deep Learning Based Multi-Temporal Crop Classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
27. He, W.; Yokoya, N. Multi-Temporal Sentinel-1 and-2 Data Fusion for Optical Image Simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [[CrossRef](#)]
28. Ghamisi, P.; Höfle, B.; Zhu, X.X. Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3011–3024. [[CrossRef](#)]
29. Chen, Y.; Li, C.; Ghamisi, P.; Jia, X.; Gu, Y. Deep Fusion of Remote Sensing Data for Accurate Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1253–1257. [[CrossRef](#)]
30. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 937–949. [[CrossRef](#)]
31. Audebert, N.; Le Saux, B.; Lefèvre, S. Beyond RGB: Very High Resolution Urban Remote Sensing with Multimodal Deep Networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 20–32. [[CrossRef](#)]
32. Sainte Fare Garnot, V.; Landrieu, L.; Chehata, N. Multi-Modal Temporal Attention Models for Crop Mapping from Satellite Time Series. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 294–305. [[CrossRef](#)]
33. Cai, Y.; Li, X.; Zhang, M.; Lin, H. Mapping Wetland Using the Object-Based Stacked Generalization Method Based on Multi-Temporal Optical and SAR Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102164. [[CrossRef](#)]
34. Farahnakian, F.; Heikkonen, J. Deep Learning Based Multi-Modal Fusion Architectures for Maritime Vessel Detection. *Remote Sens.* **2020**, *12*, 2509. [[CrossRef](#)]
35. Hu, Y.; Zeng, H.; Tian, F.; Zhang, M.; Wu, B.; Gilliams, S.; Li, S.; Li, Y.; Lu, Y.; Yang, H. An Interannual Transfer Learning Approach for Crop Classification in the Hetao Irrigation District, China. *Remote Sens.* **2022**, *14*, 1208. [[CrossRef](#)]
36. 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](#)]
37. 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](#)]
38. Nowakowski, A.; Mrzigił, J.; Spiller, D.; Bonifacio, R.; Ferrari, I.; Mathieu, P.P.; Garcia-Herranz, M.; Kim, D.-H. Crop Type Mapping by Using Transfer Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *98*, 102313. [[CrossRef](#)]
39. Chew, R.; Rineer, J.; Beach, R.; O’Neil, M.; Ujeneza, N.; Lapidus, D.; Miano, T.; Hegarty-Craver, M.; Polly, J.; Temple, D.S. Deep Neural Networks and Transfer Learning for Food Crop Identification in UAV Images. *Drones* **2020**, *4*, 7. [[CrossRef](#)]
40. Suh, H.K.; IJsselmuiden, J.; Hofstee, J.W.; van Henten, E.J. Transfer Learning for the Classification of Sugar Beet and Volunteer Potato under Field Conditions. *Biosyst. Eng.* **2018**, *174*, 50–65. [[CrossRef](#)]
41. Han, T.; Liu, C.; Yang, W.; Jiang, D. A Novel Adversarial Learning Framework in Deep Convolutional Neural Network for Intelligent Diagnosis of Mechanical Faults. *Knowl.-Based Syst.* **2019**, *165*, 474–487. [[CrossRef](#)]
42. Wang, Q.; Rao, W.; Sun, S.; Xie, L.; Chng, E.S.; Li, H. Unsupervised Domain Adaptation via Domain Adversarial Training for Speaker Recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4889–4893.
43. Li, Z.; Togo, R.; Ogawa, T.; Haseyama, M. Learning Intra-Domain Style-Invariant Representation for Unsupervised Domain Adaptation of Semantic Segmentation. *Pattern Recognit.* **2022**, *132*, 108911. [[CrossRef](#)]
44. Bejiga, M.B.; Melgani, F.; Beraldini, P. Domain Adversarial Neural Networks for Large-Scale Land Cover Classification. *Remote Sens.* **2019**, *11*, 1153. [[CrossRef](#)]
45. Kwak, G.-H.; Park, N.-W. Unsupervised Domain Adaptation with Adversarial Self-Training for Crop Classification Using Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4639. [[CrossRef](#)]
46. Wang, Z.; Zhang, H.; He, W.; Zhang, L. Cross-Phenological-Region Crop Mapping Framework Using Sentinel-2 Time Series Imagery: A New Perspective for Winter Crops in China. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 200–215. [[CrossRef](#)]
47. Veeck, G.; Veeck, A.; Yu, H. Challenges of Agriculture and Food Systems Issues in China and the United States. *Geogr. Sustain.* **2020**, *1*, 109–117. [[CrossRef](#)]
48. 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](#)]
49. Wang, G. *China Rural Statistical Yearbook 2022*; China Statistical Press: Beijing, China, 2022.
50. Hu, Q.; Sulla-Menashe, D.; Xu, B.; Yin, H.; Tang, H.; Yang, P.; Wu, W. A Phenology-Based Spectral and Temporal Feature Selection Method for Crop Mapping from Satellite Time Series. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 218–229. [[CrossRef](#)]
51. Cai, Y.; Guan, K.; Peng, J.; Wang, S.; Seifert, C.; Wardlow, B.; Li, Z. A High-Performance and in-Season Classification System of Field-Level Crop Types Using Time-Series Landsat Data and a Machine Learning Approach. *Remote Sens. Environ.* **2018**, *210*, 35–47. [[CrossRef](#)]
52. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean Mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [[CrossRef](#)]

53. Naboureh, A.; Li, A.; Bian, J.; Lei, G.; Amani, M. A Hybrid Data Balancing Method for Classification of Imbalanced Training Data within Google Earth Engine: Case Studies from Mountainous Regions. *Remote Sens.* **2020**, *12*, 3301. [[CrossRef](#)]
54. Oreopoulos, L.; Wilson, M.J.; Várnai, T. Implementation on Landsat Data of a Simple Cloud-Mask Algorithm Developed for MODIS Land Bands. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 597–601. [[CrossRef](#)]
55. 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](#)] [[PubMed](#)]
56. Ghorbanian, A.; Kakooei, M.; Amani, M.; Mahdavi, S.; Mohammadzadeh, A.; Hasanlou, M. Improved Land Cover Map of Iran Using Sentinel Imagery within Google Earth Engine and a Novel Automatic Workflow for Land Cover Classification Using Migrated Training Samples. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 276–288. [[CrossRef](#)]
57. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
58. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
59. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
60. Zhu, Y.; Zhuang, F.; Wang, J.; Ke, G.; Chen, J.; Bian, J.; Xiong, H.; He, Q. Deep Subdomain Adaptation Network for Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 1713–1722. [[CrossRef](#)] [[PubMed](#)]
61. Zhu, Y.; Zhuang, F.; Wang, D. Aligning Domain-Specific Distribution and Classifier for Cross-Domain Classification from Multiple Sources. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 5989–5996. [[CrossRef](#)]
62. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **2011**, *22*, 199–210. [[CrossRef](#)]
63. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
64. Adrian, J.; Sagan, V.; Maimaitijiang, M. Sentinel SAR-Optical Fusion for Crop Type Mapping Using Deep Learning and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 215–235. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.