remote sensing

MDPI

Article

A Spatial–Temporal Depth-Wise Residual Network for Crop Sub-Pixel Mapping from MODIS Images

Yuxian Wang [1,†], Yuan Fang [2,†], Wenlong Zhong [1], Rongming Zhuo [3], Junhuan Peng [1,*] and Linlin Xu [1,2]

[1] School of Land Science and Technology, China University of Geosciences, Beijing 100083, China
[2] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3] Aerospace Era Feihong Technology Co., Ltd., Beijing 100094, China
* Correspondence: 2006012076@cugb.edu.cn
† These authors contributed equally to this work.

Abstract: To address the problem caused by mixed pixels in MODIS images for high-resolution crop mapping, this paper presents a novel spatial–temporal deep learning-based approach for sub-pixel mapping (SPM) of different crop types within mixed pixels from MODIS images. High-resolution cropland data layer (CDL) data were used as ground references. The contributions of this paper are summarized as follows. First, we designed a novel spatial–temporal depth-wise residual network (ST-DRes) model that can simultaneously address both spatial and temporal data in MODIS images in efficient and effective manners for improving SPM accuracy. Second, we systematically compared different ST-DRes architecture variations with fine-tuned parameters for identifying and utilizing the best neural network architecture and hyperparameters. We also compared the proposed method with several classical SPM methods and state-of-the-art (SOTA) deep learning approaches. Third, we evaluated feature importance by comparing model performances with inputs of different satellite-derived metrics and different combinations of reflectance bands in MODIS. Last, we conducted spatial and temporal transfer experiments to evaluate model generalization abilities across different regions and years. Our experiments show that the ST-DRes outperforms the other classical SPM methods and SOTA backbone-based methods, particularly in fragmented categories, with the mean intersection over union (mIoU) of 0.8639 and overall accuracy (OA) of 0.8894 in Sherman County. Experiments in the datasets of transfer areas and transfer years also demonstrate better spatial–temporal generalization capabilities of the proposed method.

Keywords: MODIS; crop classification; sub-pixel mapping; spatial–temporal feature learning; deep learning; residual neural network; depth-wise convolutional neural network; pixel shuffle; generalization capability

Citation: Wang, Y.; Fang, Y.; Zhong, W.; Zhuo, R.; Peng, J.; Xu, L. A Spatial–Temporal Depth-Wise Residual Network for Crop Sub-Pixel Mapping from MODIS Images. *Remote Sens.* **2022**, *14*, 5605. https://doi.org/10.3390/rs14215605

Academic Editor: Javier J Cancela

Received: 14 September 2022
Accepted: 1 November 2022
Published: 7 November 2022

1. Introduction

Accurate information on crop types and their spatial distributions are essential for various agricultural applications, including cropland dynamic monitoring [1], crop yield estimation [2–4], disaster assessment and management [5,6], policy implementation [7–9], and crop insurance [10,11]. Remotely sensed observations have been widely used for crop mapping due to their extensive coverage with various spatial resolutions and scales [12,13]. Research indicates that coarse-resolution satellite images have great potential for crop mapping in large areas [14–16]. Compared with Landsat and Sentinel-2 images with relatively high resolutions of 10 m to 30 m [17], moderate-resolution imaging spectroradiometer (MODIS) images can better capture informative temporal patterns for better discriminating crop types. MODIS data are also more appropriate for large-area mapping due to the larger spatial coverage [18]. Therefore, MODIS has been widely used for crop mapping to support various applications [8,19–22].

However, one critical shortage of MODIS for crop mapping is caused by the low spatial resolution of MODIS, which leads to a large amount of "mixed" pixels in areas with

small and diverse crop parcels. Most existing MODIS-based crop mapping approaches have failed to address the mixed-pixel problem and have mainly focused on pixel-level segmentation methods without considering the crop type heterogeneity within mixed pixels [21–24]. The use of sub-pixel mapping (SPM) approaches for high-resolution crop mapping from MODIS images is highly insufficient [25]. To alleviate the problem of mixed pixels, instead of estimating the discrete class membership of mixed pixels, some researchers have estimated the continuous fractional coverage of crops in mixed pixels using spectral unmixing-based approaches. For example, the linear mixture model and neural networks were investigated to obtain the crop area proportions in Belgium [26]. Moreover, the 8-day composite MODIS product and agricultural statistics were used to generate sub-pixel crop-type maps using random forest (RF) crop fractions in Heilongjiang Province [14]. However, these unmixing-based approaches can only estimate the fractional existence of crop types in mixed pixels, and cannot localize subpixels of certain crop types in mixed pixels to improve the spatial resolution of derived crop maps. Some researchers directly obtained the super-resolution (SR) map by SPM methods or interpolation methods to address the mixed problem [27–29]. Atkinson [30] initially introduced SPM using spatial dependence to retrieve the appropriate spatial location for soft classification. Subsequently, more SPM methods have been proposed, i.e., each hybrid pixel is decomposed into several sub-pixels according to the given scaling factor. For example, the SPM algorithm based on sub-pixel spatial attraction models (SA-SMP) [31] and a simple pixel-swapping SPM method (PS-SMP) [32] were researched for sub-pixel target mapping. In addition, the new sub-pixel method based on radial basis function (RBF) interpolation was proposed for land SPM [33]. However, these SPM methods tend to be driven by other less relevant datasets, such as the ImageNet dataset and CLIC dataset, or small image samples. Specifically, these studies mainly focus on the exploration of methods, and the application is not comprehensive enough. It is critical to develop an advanced machine-learning approach that is tailored and designed to address the challenges in multi-temporal MODIS sub-pixel mapping with strong fragmentary particle patterns.

Efficient sub-pixel crop mapping from MODIS images relies on efficient feature extraction approaches for addressing the crop signature ambiguity problem in realistic cultivated land [1]. Spectral similarities among different crop types and dissimilarities within a crop type pose critical challenges for distinguishing crop types [34]. In recent years, deep learning (DL) has been widely used to learn discriminative features in the classification of multispectral or hyperspectral images [7,35], including some SPM applications. For instance, an efficient sub-pixel convolutional neural network (ESPCN) was proposed to generate high-resolution maps [36], which learned an array of upscaling filters. It was demonstrated that end-to-end neural network approaches can avoid complex feature engineering, and automatically learning the robust and discriminative features from remote sensing images [1,19,37–41].

However, these research studies were mainly conducted on high-resolution (e.g., GaoFen-2) or medium-resolution (Landsat and Sentinel-2) images. MODIS-based methods for crop mapping are mainly traditional statistical methods or machine learning methods [8,20–22,24,42–50], and the use of DL methods is insufficiently studied [19]. Designing DL approaches for MODIS SPM is especially important, not only because DL has demonstrated strong spectral feature learning capability, but also because DL can efficiently capture the spatial correlation effect in the image that is critical for SPM.

Exploring different vegetation indices (VIs) as classifier input for crop classification is critical to evaluate and understand the importance of different input features [34,38]. With the development of satellite time series images, many studies have explored different VIs to monitor crop growing and produce crop-type maps [51,52]. Different VIs show different crop phonological characters contained in multi-temporal remotely sensed data, and provide valuable information on the seasonal development of crops [51,53]. Among many vegetation indices, the Normalized Vegetation Difference Index (NDVI) is the most common for crop monitoring [19,21,22,51]. For example, ref. [34] selected

high-confidence pixels in the cropland data layer (CDL) and corresponding 30-m 15-day composited NDVI time series of different lengths as training samples to train the random forest (RF) classification model. Moreover, the eight-day, 500 m MODIS NDVI products were used to test the feasibility of crop unmixing in the U.S. Midwest [52]. In addition to NDVI, the enhanced vegetation index (EVI), which uses blue, red, and near-infrared reflectance from multi-temporal images, has proved more practical and less susceptible than NDVI to biases resulting from cloud and haze contamination for monitoring crop growth [39,43,45,47,54–56]. In addition to these vegetation indices, raw reflectance bands are efficient features for crop mapping. Evaluating the importance of these VIs and reflectance bands in the MODIS SPM mask is an important task for knowing the roles of different features in a DL-based MODIS SPM context.

This paper presents a spatial–temporal DL-based approach for crop SPM using MODIS images. We used 250 m MODIS products and the 50 m resampling cropland data layer (CDL) as input datasets and ground references, respectively. For a methods comparison, we used half of the MODIS data in 2017 from Sherman County in Kansas, US, to train models and applied these models to crop SPM for the other half dataset. Experiments on the data in Thomas and Gray counties in 2017 and in Sherman, Thomas, and McPherson counties in 2018 were conducted to validate the spatial and temporal generalization of different models. We conducted these experiments on five major classes (i.e., corn, winter wheat, sorghum, grass/pasture, fallow/idle cropland) in study areas. The contributions of this paper are summarized as follows. (1) A MODIS SPM spatial–temporal depth-wise residual network (ST-DRes) is designed to efficiently learn both spatial and temporal information for enhancing the classification of sub-pixel images. (2) The proposed method is compared with various classical SPM methods (PS-SPM, SA-SPM, RBF) and other state-of-the-art (SOTA) methods (ESPCN, UNet, Swin Transformer, T-DRes (temporal only) and S-DRes (spatial only)). These methods can be called SOTA methods in our research because few of them have been used in crop SPM, while most applications use traditional or simple machine learning methods [14,57,58]. (3) Model performance variations among different data combinations of satellite-derived metrics and reflectance bands from MODIS images for large-scale crop SPM were evaluated. (4) The spatial and temporal generalization capabilities of different DL approaches are compared.

2. Materials

2.1. Study Areas

In this study, we selected four crop-dominant counties as our study areas in Kansas, US: Sherman, Thomas, Gray, and McPherson, with areas of 2391 km^2, 2784 km^2, 2407 km^2 and 2334 km^2 (Figure 1). These study areas contain different main crops, and there are certain distances (both longitude and latitude) that can reflect the influences of different spatial complexities in the experiments [59]. We chose the five major crop classes: corn, sorghum, winter wheat, fallow/idle cropland, and grass/pasture. Table 1 shows the area, the number of pixels, and the proportion of each class in these areas in 2017. Other crop types, such as soybeans, alfalfa, oats, etc., which account for a small proportion of these two croplands, were not considered.

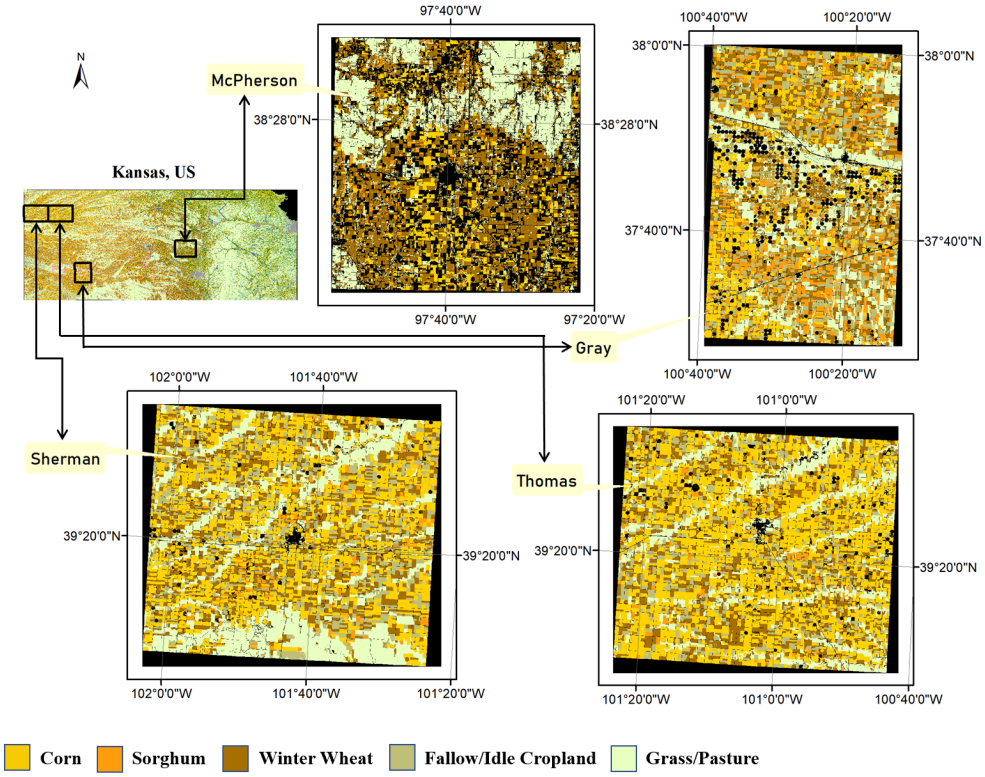

Figure 1. Reference data in study areas of Sherman, Thomas, Gray, and McPherson in Kansas, US.

Table 1. Each class is expressed as an area (km^2), the number of pixels, and the proportion in Sherman, Thomas, Gray, and McPherson in 2017, respectively.

Statistics	County	Corn	Sorghum	Winter Wheat	Fallow	Grass
Pixel Count	Sherman	684,100	174,898	640,994	616,584	758,504
	Thomas	566,438	175,689	544,659	491,250	1,119,019
	McPherson	222,513	64,398	627,134	460	697,496
	Gray	492,152	378,682	487,001	389,280	532,036
Area (km^2)	Sherman	615.71	157.41	576.91	554.94	682.68
	Thomas	509.81	158.13	490.21	442.14	1007.15
	McPherson	200.27	57.96	564.44	0.41	627.77
	Gray	442.95	340.83	438.32	350.36	478.85
Proportion	Sherman	22.51%	5.75%	21.09%	20.29%	24.96%
	Thomas	36.18%	5.68%	18.31%	17.61%	15.88%
	McPherson	8.58%	2.48%	24.19%	0.02%	26.90%
	Gray	19.67%	15.14%	19.47%	15.56%	21.27%

2.2. Remote Sensing Data

We used 250 m MODIS surface reflectance products in the sinusoidal projection. Time series of a collection of six data streams from Terra (MOD13Q1) satellite instruments for the CONUS are downloaded from the Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (LAADS-DAAC) in the period 1 January 2017 to 31 December 2017 for training and 1 January 2018 to 31 December 2018 for testing [8,19,60]. Each year has 23 images (days 1–363, at 16-day intervals). The MODIS images are re-projected to the Albers conical equal area projection to match the label data. In addition, the combinations of four optical bands (BRNM) including blue, red, near-infrared, and mid-infrared, and two additional vegetation indices (NDVI and EVI) at each observation date are used as the input variables for the classification models. Figure 2 shows the NDVI

curves of different crops in Sherman County and Thomas County, in which 60 pixels were selected for each crop as reference examples.

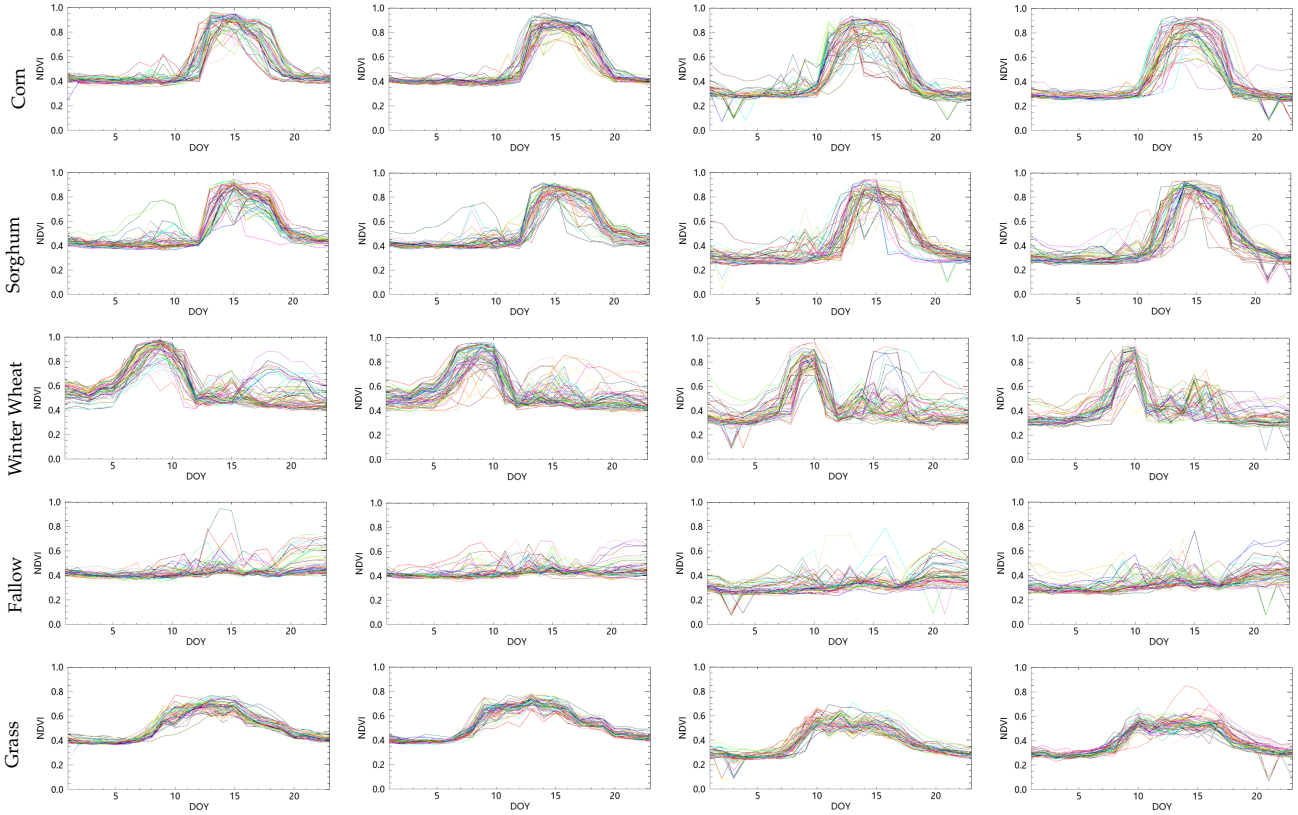

Figure 2. MODIS NDVI time series. Different colored lines represent the NDVI curves of different selected pixels in each subfigure. (**a**) NDVI time series of five crop classes at 60 pixels in Sherman County in 2017; (**b**) NDVI time series of five crops at 60 pixels in Thomas County in 2017; (**c**) NDVI time series of five crops at 60 pixels in Sherman County in 2018; (**d**) NDVI time series of five crops at 60 pixels in the Thomas County in 2018. Although these pixels were extracted from relatively pure pixels in the study areas, the time series curves still demonstrate the strong growth pattern variabilities in all crop types, especially in winter wheat. Moreover, the growth trends of different regions and different years for the same crop also have inevitable differences. The corn and soybean, which are both summer crops, have confusing phenological periods. These inner-class variabilities, inter-class similarities, and spatial/temporal discrepancies impose big challenges on crop type separation and spatial and temporal generalization capabilities of classifiers.

2.3. Cropland Layer Dataset

The cropland data layer (CDL) product is a georeferenced raster-formatted dataset with the crop-specific land cover map at 30–56 m resolution since 2008, produced by the United States Department of Agriculture (USDA) National Agricultural Statistics Service in the Albers Equal Area Conic projection, which provides more than 100 land covers and crop type categories. We download CDL products from the CropScape website portal (https://nassgeodata.gmu.edu/CropScape/) used as the ground truth for SPM [12,52]. The classification accuracy of the major crop-specific land cover categories is generally 85% to 95%.

Several research studies have used CDL as the crop type reference data for crop classification [8,21,34,61]. For instance, Song et al. [41] used the CDL to compute the average soybean and corn cultivation intensity from 2012 to 2016. Moreover, CDL maps for Kansas available for 2006–2014 were used in early-season large-area mapping of winter

crops [21]. In this article, we selected five crop classes while other crops and non-crop covers were masked out from the label map [5].

3. Method and Experiments

3.1. Methodology

The MODIS multi-temporal image was denoted by $X = \{x_i | i \in N\}$, where the ith pixel is denoted by x_i. Suppose a coarse pixel x_i can be divided into $r \times r$ subpixels, with r being the upscaling factor. Then, the coarse resolution MODIS image X corresponds with a fine-resolution label map $Y = \{y_j | j \in N \times r^2\}$, where y_j is the label of subpixels. The goal of crop SPM is to estimate Y given X, which can be achieved by solving a maximum a posterior (MAP) problem.

$$\hat{Y} = \max_{Y} P(Y|X, \Theta) \tag{1}$$

where $P(Y|X, \Theta)$ is the posterior probability of Y given X, and Θ is the model parameters.

In this paper, $P(Y|X, \Theta)$ is implemented by the proposed ST-DRes model, which can effectively extract the discriminative spatial–temporal information from MODIS images by combining the advantages of some advanced neural network architectures, i.e., full convolutional network (FCN) [62], MobileNet [63], and ResNet [64].

$$\begin{aligned} P(Y|X, \Theta) = SoftMax(Upsample(Conv_{3\times3}(SpatRes(Conv_{1\times1} \\ (TempRes(Conv_{1\times1}(X)))))) \end{aligned} \tag{2}$$

where $SpatRes$ and $TempRes$ are implemented respectively by spatial and temporal residual modules, $Conv_{M\times M}$ is the convolution layer with kernel size being M. The $Upsample$ denotes a layer with spatial interpolation operation to upscale from the size of X to the size of Y. The $Softmax$ layer outputs soft class labels, which is also the posterior probability of Y.

The overall architecture of ST-DRes is expressed as shown in the top section of Figure 3, which is framed in a green dashed box. The red dotted box of Figure 3 shows $SpatRes$ and $TempRes$ blocks using depth-wise convolution and point-wise convolution respectively, so that the number of parameters in the network is greatly reduced and the network is more compact.

The modules $SpatRes$ and $TempRes$ are used to learn efficient spatial–temporal features in X, based on which the $Upsample$ (pixelshuffle) layer is added to obtain high-resolution label maps. There are two options available in the $Upsample$ layer, i.e., spatial interpolation layer and pixelshuffle layer, both of which can improve the feature resolution.

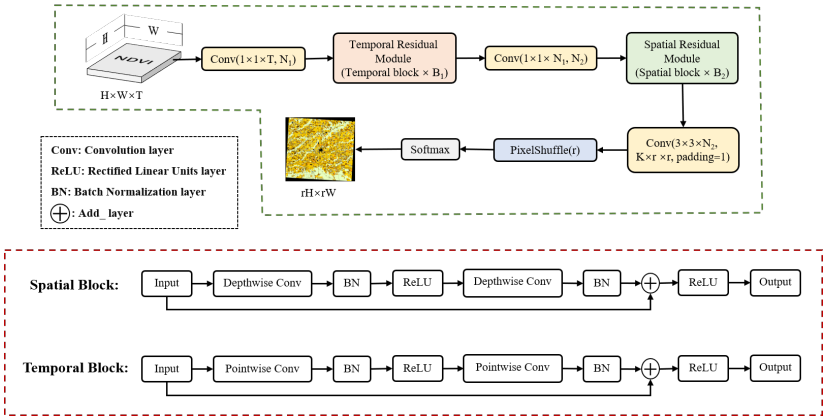

Figure 3. The framework of the ST-DRes model. As shown in the figure, the framework involves three parts. The top section (in the green dashed box) is the overall architecture of ST-DRes, while the red dashed box shows $SpatRes$ and $TempRes$ blocks using depth-wise convolution and point-wise convolution, respectively. The black dashed box gives the full names of some layers, which are denoted by acronyms in the figure.

3.2. Experiment Settings

Several classical SPM methods, i.e., PS-SPM, SA-SPM, and RBF algorithms, and several SOTA DL approaches, i.e., ESPCN [36], UNet [65], and Swin Transformer [66] models were adopted to compare with the proposed method. Since UNet and Swin Transformer methods are not originally designed for SPM, to enable comparison with our approach, we improve the two methods by adding an upsampling layer with the bilinear operation before the softmax layer. ESPCN mainly consists of convolution layers as the backbone unit. Moreover, the UNet structure contains four basic blocks, i.e., four compression (encoding) parts and four expansion (decoding) parts. Swin Transformer uses self-attention-based architecture, which has an excellent performance in many mainstream image processing tasks. In addition, variants of the proposed model, spatial DRes (S-DRes), and temporal DRes (T-DRes) are compared to do ablation studies of the proposed approach to demonstrate the improvement and benefits of using both spatial and temporal modules.

The learning rate is set to 0.0001 and the Adam optimizer is used in all experiments. The model trains a total of 1000 epochs. The multi-temporal MODIS image is divided into small multi-temporal blocks of size 64 by 64 for training and testing. Moreover, experiments are carried out on the workstation with an NVIDIA GeForce RTX 2080 Ti GPU in the PyTorch toolbox.

3.3. Model Validation

For model comparison experiments, we focused on Sherman County in 2017, in which 50% of the data is used for training and the rest for testing. We compute the accuracy and F1 score of each class, overall accuracy (OA), mean intersection over union (mIoU) and Kappa coefficient of the test set to evaluate the performance of all classifiers.

In the other three experiments, which explore the effects of different upsampling layers, different vegetation indices as input data, and different combinations of spatial and temporal modules of the model for SPM, we calculate not only the test accuracy but also the prediction accuracy in using all data of Sherman County.

For the model generalization comparison experiments, we fix 90% of the Sherman County data in 2017 as training samples, the remaining 10% as validation samples, and the transfer data as the testing samples. For example, the Thomas data in 2017 were used as the testing data set in the spatial generalization experiment, and the Sherman data in 2018 were used as the testing data set in the temporal generalization experiment. Because OA can similarly reflect the experimental results about the Kappa coefficient, we do not make redundant evaluations here.

4. Results

4.1. Methods Comparison

The results and predicted sub-pixel maps of the Sherman in 2017 achieved by different methods are demonstrated in Table 2 and Figure 4, respectively. We convert the SPM results into the RGB color display, which is the same as the CDL image in order to visualize the prediction more intuitively. Specifically, the proposed ST-DRes model has the highest OA (88.94%) among all traditional SPM methods and different backbone-based DL models. The mIoU (0.8639) is much higher than other classifiers (PS-SMP: 0.2765, RBF: 0.2714, SA-SMP: 0.2957, ESPCN: 0.7292, UNet: 0.7701, Swin Transformer: 0.8369). Figure 4 shows the whole SPM results and a small part of the maps (320×320) for more detailed analysis. The ST-DRes method achieves good classification results, while there are many misclassification pixels in traditional SPM methods, and ESPCN and UNet networks produce excessively smooth maps.

Table 2. The accuracy, F1 score of each class, OA, mIoU, and Kappa coefficients of the test sets of different methods (the highest accuracy in each column is in bold).

Method	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU	Kappa
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1			
PS-SMP	0.6052	0.4828	0.1762	0.0995	0.3024	0.3765	0.3925	0.4066	0.2348	0.2934	0.3650	0.2765	0.3362
RBF	0.2676	0.3267	0.0310	0.0287	0.5368	0.3876	0.4275	0.4547	0.3753	0.4309	0.3766	0.2714	0.3401
SA-SMP	0.6576	0.5248	0.1670	0.0943	0.3194	0.3976	0.4359	0.4513	0.2450	0.3062	0.3924	0.2957	0.3649
ESPCN	0.8315	0.7759	0.3726	0.4413	0.7821	0.7727	0.7483	0.7657	0.8467	0.8625	0.7827	0.7292	0.7588
UNet	0.8273	0.8042	0.5458	0.5884	0.8191	0.7914	0.7692	0.7782	0.8598	0.8820	0.8070	0.7701	0.7942
Swin Transformer	0.9045	0.8660	0.5776	0.6775	0.8543	0.8552	0.8531	0.8498	0.9139	0.9231	0.8680	0.8369	0.8443
ST-DRes	**0.9185**	**0.8867**	**0.6405**	**0.7357**	**0.8946**	**0.8802**	**0.8659**	**0.8744**	**0.9267**	**0.9346**	**0.8894**	**0.8639**	**0.8684**

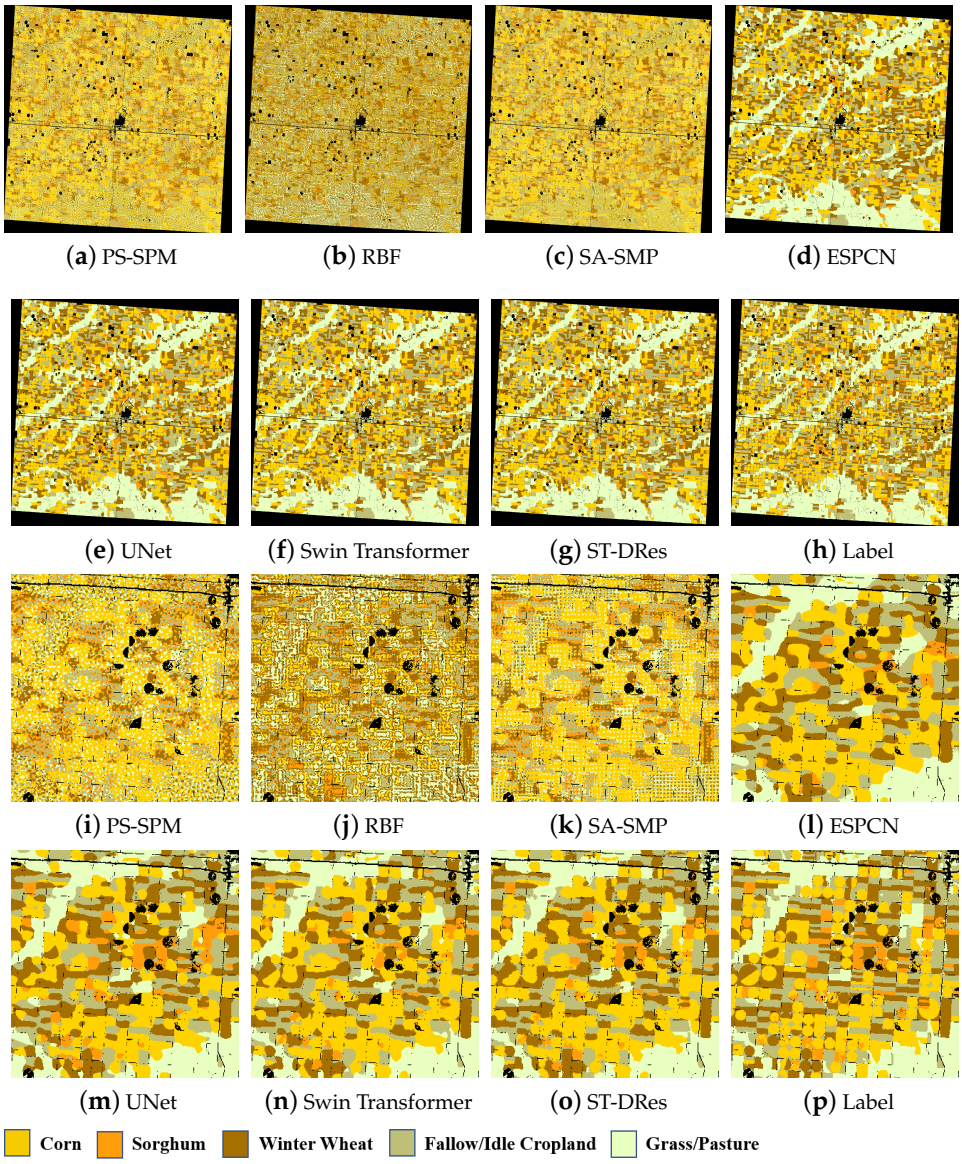

Figure 4. Predicted sub-pixel maps of different methods in Sherman Country in 2017 achieved by different methods. (**a**–**g**) show predictions for the entire Sherman County by different methods and (**h**) is the label map, (**i**–**o**) show SPM results for a small example area in Sherman County, and (**p**) is the corresponding label. Obviously, there are many misclassification pixels in the traditional SPM methods, and ESPCN and UNet networks produce excessively smooth maps. Moreover, the ST-DRes method achieves a good classification result, while Swin Transformer also shows a good result, inferior to ST-DRes only in some details and small classes, such as sorghum.

4.2. Upsampling Methods

This section focuses on validating the effectiveness of different upsample approaches, i.e., the pixelshuffle layer, and different spatial interpolation layers (architectures in Figure 5). We set the upscaling scale r as 5 in all experiments. Four interpolation methods lead to very different SPM performances. Then, the final softmax layer calculates the likelihood of each class. In Table 3, the pixelshuffle method has the highest test accuracy (87.84%) among all upsampling methods. As can be observed from Figure 6, the nearest and area upsampling layers show the zigzag SPM maps, while bilinear, bicubic, and pixelshuffle layers display more fine-grained parcels and boundary details. In particular, pixelshuffle is closer to the ground truth on the fragmentary and small crop categories.

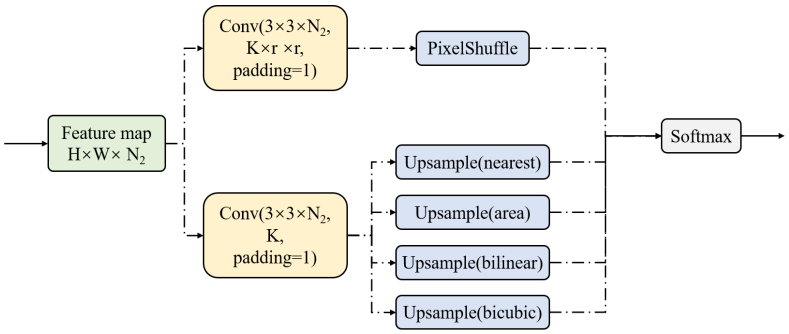

Figure 5. The optional upsampling architectures at the end of the network.

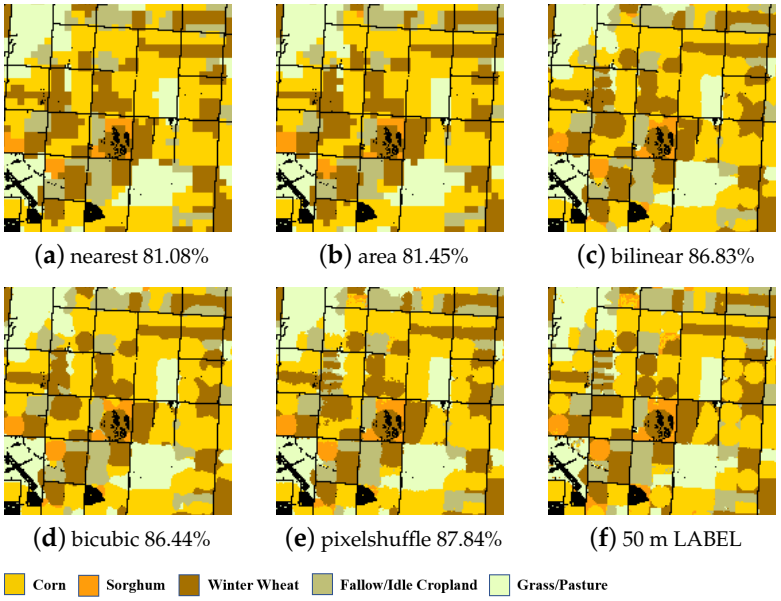

Figure 6. The subfigures (**a**–**e**) represent the sub-pixel mapping of five different upsampling methods (nearest, area, bilinear, bicubic and pixelshuffle) implemented on a small portion of the Sherman dataset, including the test accuracy they achieved. And (**f**) is the corresponding label. It shows that the pixelshuffle layer is more capable of generating sub-pixel maps even in fragmentary parcels and limited crop categories.

Table 3. The train accuracy, train mIoU, test accuracy, test mIoU, predict accuracy, and predict mIoU of the dataset from four kinds of spatial interpolation methods and pixelshuffle layer at the end of the network (the highest accuracy in each column is in bold format.

Upsample	Train Acc	Train mIoU	Test Acc	Test mIoU	Predict Acc	Predict mIoU
nearest	0.8411	0.8295	0.8108	0.7726	0.7999	0.7550
area	0.8421	0.8310	0.8145	0.7772	0.8033	0.7570
bilinear	0.9323	0.9231	0.8683	0.8345	0.8521	0.8122
bicubic	0.9272	0.9184	0.8644	0.8319	0.8459	0.8062
pixelshuffle	**0.9566**	**0.9508**	**0.8784**	**0.8498**	**0.8671**	**0.8337**

4.3. Evaluation of Spatial and Temporal Modules

We compare the proposed approach ST-DRes with two variants S-DRes and T-DRes by just using spatial and temporal information. As shown in Table 4, the ST-DRes architecture produces improved results compared to S-DRes and T-DRes, with a significantly higher test accuracy at 88.94%, with the best S-DRes at 87.79%, and the best T-DRes at 82.70%. It demonstrates that using both temporal and spatial data in MODIS can lead to more desirable results. In Figure 7, ST-DRes outperforms both spatial-only and temporal-only models, especially in the highlighted red box area in Figure 7.

Table 4. Experimental results of the test dataset from the different temporal channels (N_1), the number of temporal blocks (B_1), spatial channel (N_2), and the number of temporal blocks (B_2) from the model architecture (Figure 3). T-DRes indicates that the model uses only temporal blocks. S-DRes indicates that the model uses only spatial blocks. ST-DRes indicates that the model uses both temporal and spatial blocks.

Method	S Channel	S Block	Train Acc	Train mIoU	Test Acc	Test mIoU
S-DRes	512	1	0.8876	0.8748	0.8471	0.8153
S-DRes	512	2	0.9287	0.9191	0.8642	0.8336
S-DRes	512	3	0.9440	0.9357	0.8667	0.8361
S-DRes	512	4	0.9531	0.9464	**0.8670**	**0.8382**
S-DRes	512	5	0.9558	0.9497	0.8663	0.8352
S-DRes	512	6	**0.9574**	**0.9513**	0.8630	0.8326
S-DRes	64	4	0.8725	0.8593	0.8327	0.7954
S-DRes	128	4	0.8978	0.8858	0.8420	0.8080
S-DRes	256	4	0.9271	0.9165	0.8565	0.8234
S-DRes	512	4	0.9531	0.9464	0.8670	0.8382
S-DRes	1024	4	**0.9765**	**0.9734**	**0.8779**	**0.8494**
Method	**T Channel**	**T Block**	**Train Acc**	**Train mIoU**	**Test Acc**	**Test mIoU**
T-DRes	512	1	0.7819	0.7553	0.7640	0.7217
T-DRes	512	2	0.8319	0.8156	0.7852	0.7479
T-DRes	512	3	0.8858	0.8702	0.8084	0.7723
T-DRes	512	4	0.8997	0.8857	0.8154	0.7804
T-DRes	512	5	**0.9229**	**0.9103**	**0.8263**	**0.7917**
T-DRes	512	6	0.9129	0.8999	0.8196	0.7860
T-DRes	64	5	0.8307	0.8172	0.7808	0.7404
T-DRes	128	5	0.8843	0.8701	0.8013	0.7659
T-DRes	256	5	**0.9325**	**0.9205**	**0.8270**	**0.7938**
T-DRes	512	5	0.9229	0.9103	0.8263	0.7917
T-DRes	1024	5	0.7642	0.7428	0.7550	0.7092
Method	**S Channel**	**S Block**	**T Channel**	**T Block**	**Test Acc**	**Test mIoU**
ST-DRes	**1024**	**4**	**256**	**5**	**0.8894**	**0.8639**

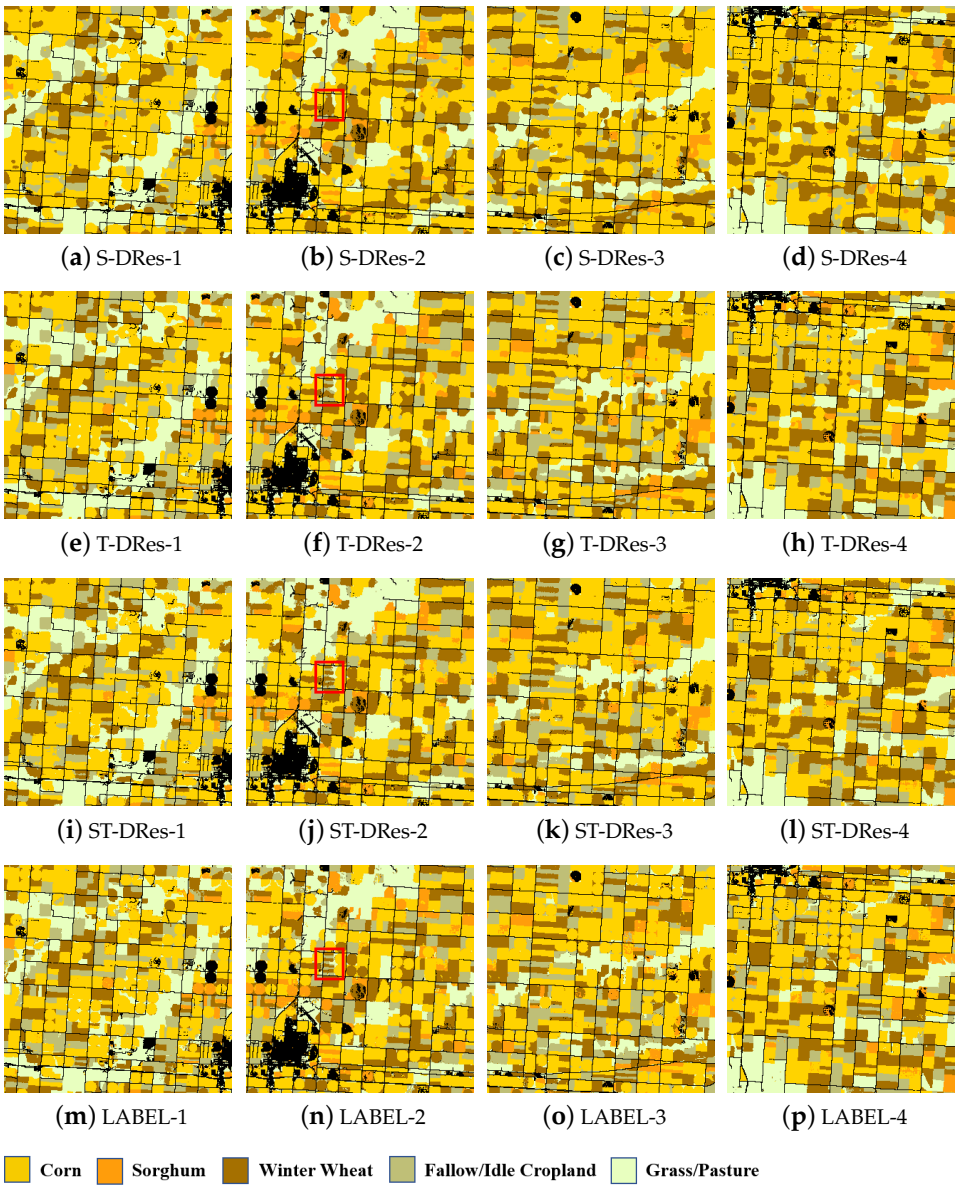

Figure 7. Four small sub-pixel example maps (64 × 64) in the Sherman dataset are displayed by using different temporal and spatial information. (**a**,**e**,**i**) represent the results generated by the three methods (S-DRes,T-DRes, ST-DRes) on example map-1, while (**m**) is the corresponding label. Similarly, (**b**,**f**,**j**) represent the results generated by the three methods on example map-2, while (**n**) is the corresponding label; (**c**,**g**,**k**) represent the results generated by the three methods on example map-3, while (**o**) is the corresponding label; (**d**,**h**,**l**) represent the results generated by the three methods on example map-4, while (**p**) is the corresponding label. It shows that T-DRes is more capable of generating sub-pixel maps that are closest to the ground truth than S-DRes architecture. ST-DRes outperformed the other two models, which is especially noticeable in the highlighted red box area.

4.4. Vegetation Index Selection

We evaluated the importance of different combinations of satellite-derived metrics and reflectance bands for crop SPM. We compare the use of NDVI, EVI, and BRNM as inputs for the proposed methods (results in Table 5). The results show that EVI and BRNM perform similarly to NDVI, while the accuracy of BRNM is slightly better. All test accuracies can achieve 0.87 ∼ 0.88, demonstrating the proposed DL approach can extract discriminative features from all input data combinations.

Table 5. The results of different satellite-derived metrics and combinations of reflectance bands for crops SPM.

Input	Train Acc	Train mIoU	Test Acc	Test mIoU	Predict Acc	Predict mIoU
NDVI	0.9840	0.9817	**0.8893**	**0.8638**	0.8790	**0.8524**
EVI	0.9753	0.9808	0.8744	0.8552	0.8630	0.8436
BRNM	**0.9885**	**0.9826**	0.8842	0.8582	**0.8806**	0.8522

4.5. Generalizability Analysis

The Thomas-2017 and Gray-2017 sections in Table 6 show the accuracy, F1 score of each class, OA, and mean mIoU of the test sets of different methods in spatial generalization experiments across different regions, while the Sherman-2018, Thomas-2018, and McPherson-2018 sections in Table 6 show corresponding results in temporal generalization experiments across different years.

Table 6. The accuracy, F1 score of each class, mIoU, OA, and mean mIoU of the test sets of different methods (the highest accuracy in each column is in bold). The table is divided into five sections. The "Thomas-2017" and "Gray-2017" sections represent the results of training in Sherman County in 2017 and testing in Thomas County and Gray County in 2017, respectively. The "Sherman-2018", "Thomas-2018", and "McPherson-2018" sections represent the results of training in Sherman County in 2017 and testing in Sherman County, Thomas County, and McPherson County in 2018, respectively.

Thomas-2017

Method	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1		
ESPCN	0.7741	0.7594	0.2841	0.3194	0.6706	**0.6589**	0.6001	0.6498	0.7453	0.6952	0.6924	0.6165
UNet	0.6310	0.7112	**0.4338**	0.3767	**0.6794**	0.6401	**0.7532**	0.6606	0.7195	0.7119	0.6669	0.6201
Swin Transformer	0.7561	0.7608	0.3663	**0.4015**	0.6662	0.6490	0.6502	0.6607	0.7460	0.7179	0.6992	0.6380
ST-DRes	**0.7831**	**0.7727**	0.3286	0.3986	0.6597	0.6577	0.6569	**0.6640**	**0.7588**	**0.7251**	**0.7096**	**0.6436**

Gray-2017

Method	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1		
ESPCN	0.5014	0.4166	0.1593	0.2259	0.5859	**0.5611**	0.5055	0.4926	**0.6461**	0.6050	0.5132	0.4602
UNet	0.4752	0.3934	0.2111	0.2886	0.5686	0.5411	**0.5703**	0.4922	0.6179	0.5939	0.4949	0.4618
Swin Transformer	**0.5828**	0.4795	0.2211	0.2955	0.5684	0.5387	0.5422	0.4830	0.5707	0.6188	0.5251	0.4831
ST-DRes	0.5462	**0.4856**	**0.2329**	**0.3177**	**0.6050**	0.5479	0.5396	**0.4990**	0.6173	**0.6363**	**0.5418**	**0.4973**

Sherman-2018

Method	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1		
ESPCN	0.6743	0.6597	0.0399	0.0643	0.5045	0.6264	0.6891	**0.7139**	**0.8566**	0.6530	0.6589	0.5435
UNet	**0.6783**	**0.6926**	**0.2997**	**0.2928**	**0.7052**	0.6951	0.7879	0.6766	0.6379	0.6875	0.6831	0.6089
Swin Transformer	0.6382	0.6669	0.0700	0.1133	0.6711	0.5730	**0.8051**	0.6176	0.3826	0.4978	0.6001	0.4937
ST-DRes	0.6685	0.6891	0.0518	0.0802	0.6621	**0.6962**	0.6564	0.6425	0.8393	**0.7182**	**0.6862**	**0.5653**

Thomas-2018

Method	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1		
ESPCN	**0.7857**	0.7116	0.0377	0.0534	0.4248	0.5425	0.5264	**0.5664**	0.6813	0.5438	0.6131	0.4835
UNet	0.5608	0.6473	**0.3409**	**0.2987**	**0.6318**	**0.6058**	0.7388	0.5422	0.4359	0.4438	0.5664	**0.5076**
Swin Transformer	0.7617	0.7158	0.0262	0.0446	0.4640	0.5127	**0.7898**	0.4960	0.2017	0.3101	0.5686	0.4159
ST-DRes	0.7561	**0.7214**	0.0306	0.0498	0.5583	0.6221	0.4690	0.5063	**0.7271**	**0.5866**	**0.6335**	0.4972

Table 6. *Cont.*

Method	McPherson-2018											
	Corn		Sorghum		Winter Wheat		Fallow		Grass		OA	mIoU
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1		
ESPCN	0.3833	**0.3389**	0.2338	0.2186	0.1765	0.2802	0.1752	0.0006	**0.6803**	0.6172	0.4631	0.2911
UNet	**0.4248**	0.2942	0.0000	0.0000	0.5355	0.6220	**0.1910**	0.0007	0.6415	**0.6372**	0.5610	0.3108
Swin Transformer	0.0937	0.1604	**0.4518**	**0.3625**	0.8594	0.7004	0.0125	0.0040	0.4192	0.4688	0.6271	0.3392
ST-DRes	0.1473	0.2323	0.3217	0.3364	**0.8341**	**0.7409**	0.0111	**0.0055**	0.6071	0.5901	**0.6872**	**0.3810**

4.5.1. Spatial Generalizability

Overall, the spatial generalization results of the proposed method ST-DRes are better than the other methods, achieving the best OA (0.7096) and the best mIoU (0.6436) in nearby Thomas County, and the best OA of 0.5418 and the best mIoU (0.4973) in Gray County, when the MODIS NDVI data of Sherman County of the same year are used as training data (from the Table 6 "Thomas-2017" and "Gray-2017"). The results also reflect that the more distant the study area is predicted by the trained model, the worse the generalization ability of the model, and other methods have shown similar trends, which are illustrated in both the OA and CAs.

To compare and analyze the predicted SPM results of different models, intuitively, we select a small part of each transfer area to visualize, as shown in Figure 8. The selected small regions are mainly concentrated in the inner area of these counties, and the size is all 320×320, which includes the main crops we choose as far as possible. We can see that the proposed ST-DRes has better performances in details and boundaries from Figure 8, which is the same as the accuracy results in Table 6. The SPM maps on ESPCN and UNet show overly smooth parcels and many misclassified classes, especially in the predicted map of ESPCN.

4.5.2. Temporal Generalizability

The bottom three parts of Table 6 and the right three columns of Figure 8 show the accuracy results and SPM maps for the three testing sub-regions in 2018 (Sherman, Thomas, and McPherson, respectively). The ST-DRes model also shows the optimal performance (OA is 0.6862, 0.6335, and 0.6872 in three regions, respectively). The "Thomas-2018" part of Table 6 shows that the better accuracies of UNet for small classes, such as sorghum (ESPCN: 0.0377; UNet:0.3409; Swin Transformer: 0.0262; ST-DRes: 0.0306), but it can be seen from Figure 8 that there are most spatial misclassifications on corn and soybeans, and the relatively high accuracy of sorghum is also due to the fact that UNet assigns other classes into the sorghum class. Thus, class accuracies such as this are not representative.

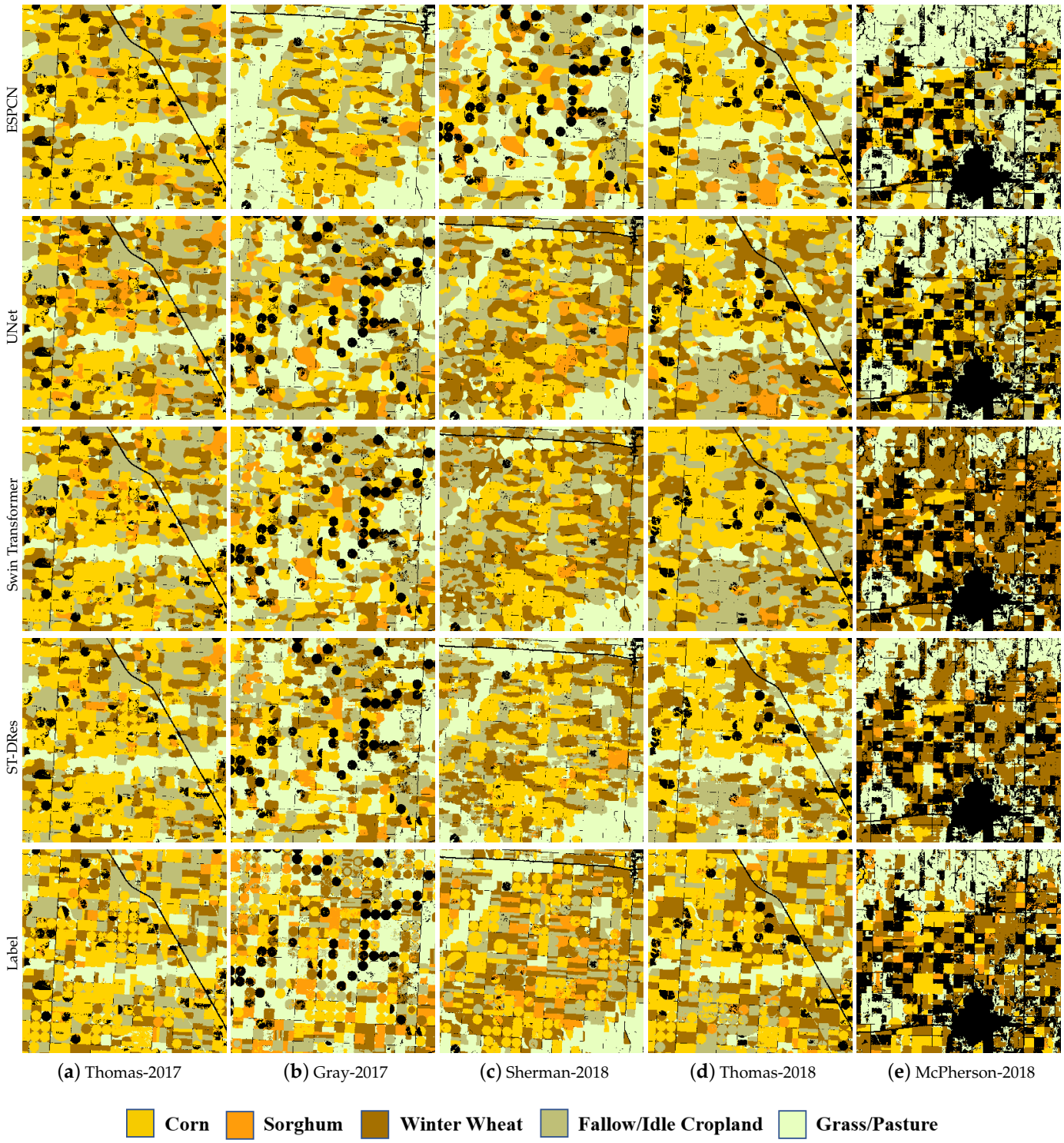

Figure 8. SPM results of small examples (320 × 320) in different spatial and temporal generalization experiments of different methods are displayed. The rows of "(**a**) Thomas-2017" and "(**b**) Gray-2017" represent the results of training in Sherman County in 2017 and testing in Thomas County and Gray County in 2017, respectively. The rows of "(**c**) Sherman-2018", "(**d**) Thomas-2018", and "(**e**) McPherson-2018" represent the results of training in Sherman County in 2017 and testing in Sherman County, Thomas Count, and McPherson County in 2018, respectively.

5. Discussion

5.1. SPM Methods Analysis

According to the results in Table 2 and Figure 4, the traditional SPM methods (PA-SPM, RBF and SA-SPM) are not effective in the application of large-scale crop SPM. Most of them assign soft classes for sub-pixels within the pixel based on the polygons or class proportions [31–33]. These methods only consider spatial information, and it is difficult for SPM in complex crop research because of the differences between inner-classes and the similarities of inter-classes. Moreover, these traditional SPM methods can be defined as unsupervised mapping. They only use spatial correlation to obtain soft labels, while DL models (ESPCN, UNet, Swin Transformer, and ST-DRes) need training data to train the models so that DL models can understand the data more efficiently. It is unfair to compare the simple SPM algorithms with DL models briefly. DL models are more effective for large-area crop SPMs with a small number of training samples, and they can learn more implicit features rather than the shallow features in the data. Although the recent Swin Transformer model serves as a general-purpose backbone for computer vision, it does not seem to be the best for the crop time series sub-pixel classification experiment.

5.2. Time Series Analysis

In order to better understand the changes in the characteristics of crop growth, we analyzed the time series curves. Figure 2 shows that although these points are extracted from relatively purer pixels in the study areas, the time series curves still demonstrate crop signature variabilities, especially for winter wheat. Corn and soybean have a confusing phenological period [15,34,67], which are both harvested in October. These inner-class variabilities, inter-class similarities, and spatial/temporal discrepancies impose big challenges on crop type separation and the spatial and temporal generalization capability of classifiers.

In addition, the results in Table 5 show that different data inputs have slight differences in the experimental results. The purpose of evaluating input features is not only to explore the best feature for crop SPM applications but also to provide a reference for feature selection for crop SPM using MODIS time series images.

5.3. Uncertainty of Model Generalization Ability

Although the generalization experiments in different study areas and different years show good potential for identifying the crop sub-pixel classes in Table 6 and Figure 8, there is still much misclassification due to the differences in crop phenology and crop calendar [34]. The temporal transfer experiments, which are different from the spatial transfer experiments, show more uncertainties. For example, the generalization result (OA of ST-DRes is 0.5418) in Gray County, which is further from the training Sherman County, is worse than the generalization result (OA of ST-DRes is 0.7096) of neighboring Thomas County, which is consistent with the larger spatial complexity in more distant areas [59]. However, in the temporal transfer experiments between the years, there is more uncertainty. Therefore, although the ST-DRes model shows great potential in SPM application, the generalization ability is still limited. The SPM results depend not only on the solution of the image but also on the temporal and spatial complexity of regions [59]. We cannot directly transfer the general model to other diverse data, and it needs to refer to more auxiliary data and consider the different characteristics of agricultural regions, such as climate conditions and crop planting heterogeneity.

6. Conclusions

In this study, MODIS time series data were used to identify sub-pixel crop types using a new DL SPM model, i.e., ST-DRes. Different numerical measures (i.e., class accuracy, F1 score, OA, mIoU, and Kappa coefficient) showed that the proposed ST-DRes method can outperform the traditional SPM algorithms and SOTA DL approaches, demonstrating that the proposed architecture can efficiently learn spatial–temporal discriminative information for enhancing crop SPM from MODIS images. Specifically, we achieved the following

conclusions. First, the ST-DRes method achieved the best performance for the crop SPM application, even when the training samples are small. Second, compared with different spatial interpolation upsampling layers, using the pixelshuffle layer can better improve feature resolution. Third, different combinations of reflection bands and VIs as model inputs achieved similar crop SPM performances, demonstrating that the proposed model is relatively insensitive to feature inputs and has excellent feature learning abilities from different input data. Fourth, the proposed method showed good generalization than other methods, but there were still some uncertainties in transfer experiments. These results indicate that the proposed method can realize crop SPM well in practical applications, and can provide effective decision support for yield prediction and agricultural management using the remote sensing images of coarse resolution in the future. Our future work will concentrate on investigating the efficient techniques of SPM for transfer learning and the application of other SOTA DL methods for crop mapping.

Author Contributions: Conceptualization, Y.W. and L.X.; data collection and data processing, Y.W. and R.Z.; methodology, Y.F.; formal analysis, Y.F.; experiments, Y.W. and W.Z.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W. and L.X.; visualization, Y.W.; supervision, J.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China under Grant 42074004 and the Ministry of Natural Resources of the People's Republic of China under Grant 0733-20180876/1.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the reviewers and associate editor for their valuable comments and suggestions to improve the quality of the paper.

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

References

1. 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]
2. Arvor, D.; Jonathan, M.; Meirelles, M.S.P.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* **2011**, *32*, 7847–7871. [CrossRef]
3. Kussul, N.; Skakun, S.; Shelestov, A.; Lavreniuk, M.; Yailymov, B.; Kussul, O. Regional scale crop mapping using multi-temporal satellite imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 45. [CrossRef]
4. Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
5. Konduri, V.S.; Kumar, J.; Hargrove, W.W.; Hoffman, F.M.; Ganguly, A.R. Mapping crops within the growing season across the United States. *Remote Sens. Environ.* **2020**, *251*, 112048. [CrossRef]
6. Waldner, F.; Fritz, S.; Di Gregorio, A.; Defourny, P. Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps. *Remote Sens.* **2015**, *7*, 7959–7986. [CrossRef]
7. Li, J.; Shen, Y.; Yang, C. An Adversarial Generative Network for Crop Classification from Remote Sensing Timeseries Images. *Remote Sens.* **2021**, *13*, 65. [CrossRef]
8. Yang, Y.; Tao, B.; Ren, W.; Zourarakis, D.P.; Masri, B.E.; Sun, Z.; Tian, Q. An improved approach considering intraclass variability for mapping winter wheat using multitemporal MODIS EVI images. *Remote Sens.* **2019**, *11*, 1191. [CrossRef]
9. Lobell, D.B. The use of satellite data for crop yield gap analysis. *Field Crop. Res.* **2013**, *143*, 56–64. [CrossRef]
10. Hamidi, M.; Safari, A.; Homayouni, S. An auto-encoder based classifier for crop mapping from multitemporal multispectral imagery. *Int. J. Remote Sens.* **2021**, *42*, 986–1016. [CrossRef]
11. Whelen, T.; Siqueira, P. Use of time-series L-band UAVSAR data for the classification of agricultural fields in the San Joaquin Valley. *Remote Sens. Environ.* **2017**, *193*, 216–224. [CrossRef]
12. 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]
13. Azzari, G.; Lobell, D. Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. *Remote Sens. Environ.* **2017**, *202*, 64–74. [CrossRef]
14. 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]

15. Zhong, L.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 1–13. [CrossRef]
16. Wardlow, B.D.; Egbert, S.L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116. [CrossRef]
17. Ozdogan, M. The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis. *Remote Sens. Environ.* **2010**, *114*, 1190–1204. [CrossRef]
18. Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolking, S.; Li, C.; Salas, W.; Moore III, B. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens. Environ.* **2005**, *95*, 480–492. [CrossRef]
19. Zhong, L.; Hu, L.; Zhou, H.; Tao, X. Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US. *Remote Sens. Environ.* **2019**, *233*, 111411. [CrossRef]
20. Li, L.; Friedl, M.A.; Xin, Q.; Gray, J.; Pan, Y.; Frolking, S. Mapping crop cycles in China using MODIS-EVI time series. *Remote Sens.* **2014**, *6*, 2473–2493. [CrossRef]
21. Skakun, S.; Franch, B.; Vermote, E.; Roger, J.C.; Becker-Reshef, I.; Justice, C.; Kussul, N. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sens. Environ.* **2017**, *195*, 244–258. [CrossRef]
22. Massey, R.; Sankey, T.T.; Congalton, R.G.; Yadav, K.; Thenkabail, P.S.; Ozdogan, M.; Meador, A.J.S. MODIS phenology-derived, multi-year distribution of conterminous US crop types. *Remote Sens. Environ.* **2017**, *198*, 490–503. [CrossRef]
23. Qiong, H.; Yaxiong, M.; Baodong, X.; Qian, S.; Huajun, T.; Wenbin, W. Estimating Sub-Pixel Soybean Fraction from Time-Series MODIS Data Using an Optimized Geographically Weighted Regression Model. *Remote Sens.* **2018**, *10*, 491.
24. Zhong, L.; Yu, L.; Li, X.; Hu, L.; Gong, P. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. *Sci. Rep.* **2016**, *6*, 1–14. [CrossRef] [PubMed]
25. Shao, Y.; Lunetta, R.S. Comparison of sub-pixel classification approaches for crop-specific mapping. In Proceedings of the 2009 17th International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–4.
26. Verbeiren, S.; Eerens, H.; Piccard, I.; Bauwens, I.; Van Orshoven, J. Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 486–497. [CrossRef]
27. Aplin, P.; Atkinson, P.M. Sub-pixel land cover mapping for per-field classification. *Int. J. Remote Sens.* **2001**, *22*, 2853–2858. [CrossRef]
28. Chen, Y.; Ge, Y.; Chen, Y.; Jin, Y.; An, R. Subpixel land cover mapping using multiscale spatial dependence. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5097–5106. [CrossRef]
29. Wang, Q.; Zhang, C.; Atkinson, P.M. Sub-pixel mapping with point constraints. *Remote Sens. Environ.* **2020**, *244*, 111817. [CrossRef]
30. Atkinson, P.M. Mapping sub-pixel boundaries from remotely sensed images. In *Innovations in GIS*; CRC Press: Boca Raton, FL, USA, 1997; pp. 184–202.
31. Mertens, K.C.; De Baets, B.; Verbeke, L.P.; De Wulf, R.R. A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. *Int. J. Remote Sens.* **2006**, *27*, 3293–3310. [CrossRef]
32. Atkinson, P.M. Sub-pixel target mapping from soft-classified, remotely sensed imagery. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 839–846. [CrossRef]
33. Wang, Q.; Shi, W.; Atkinson, P.M. Sub-pixel mapping of remote sensing images based on radial basis function interpolation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 1–15. [CrossRef]
34. Pengyu, H.; Liping, D.; Chen, Z.; Liying, G. Transfer learning for crop classification with Cropland Data Layer data (CDL) as training samples. *Sci. Total Environ.* **2020**, *733*, 138869.
35. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* **2018**, *10*, 75. [CrossRef]
36. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
37. Li, Z.; Chen, G.; Zhang, T. A CNN-Transformer Hybrid Approach for Crop Classification Using Multitemporal Multisensor Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 847–858. [CrossRef]
38. Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [CrossRef]
39. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2018**, *221*, 430–443. [CrossRef]
40. Ozgur Turkoglu, M.; D'Aronco, S.; Perich, G.; Liebisch, F.; Streit, C.; Schindler, K.; Wegner, J.D. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sens. Environ.* **2021**, *264*, 112603. [CrossRef]
41. Song, X.P.; Huang, W.; Hansen, M.C.; Potapov, P. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* **2021**, *3*, 100018. [CrossRef]
42. Hao, P.; Zhan, Y.; Wang, L.; Niu, Z.; Shakir, M. Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. *Remote Sens.* **2015**, *7*, 5347–5369. [CrossRef]

43. Liu, J.; Huffman, T.; Qian, B.; Shang, J.; Jing, Q. Crop yield estimation in the Canadian Prairies using Terra/MODIS-derived crop metrics. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2685–2697. [CrossRef]
44. Hao, P.; Wang, L.; Zhan, Y.; Wang, C.; Niu, Z.; Wu, M. Crop classification using crop knowledge of the previous-year: Case study in Southwest Kansas, USA. *Eur. J. Remote Sens.* **2016**, *49*, 1061–1077. [CrossRef]
45. Pan, Y.; Li, L.; Zhang, J.; Liang, S.; Zhu, X.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index. *Remote Sens. Environ.* **2012**, *119*, 232–242. [CrossRef]
46. Gusso, A.; Arvor, D.; Ricardo Ducati, J.; Veronez, M.R.; da Silveira, L.G. Assessing the MODIS crop detection algorithm for soybean crop area mapping and expansion in the Mato Grosso State, Brazil. *Sci. World J.* **2014**, *2014*, 863141. [CrossRef] [PubMed]
47. Nguyen-Thanh, S.; Chen, C.F.; Chen, C.R.; Huynh-Ngoc, D.; Chang, L.Y. A Phenology-Based Classification of Time-Series MODIS Data for Rice Crop Monitoring in Mekong Delta, Vietnam. *Remote Sens.* **2013**, *6*, 135.
48. Sakamoto, T.; Gitelson, A.A.; Arkebauer, T.J. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sens. Environ.* **2013**, *131*, 215–231. [CrossRef]
49. Mkhabela, M.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **2011**, *151*, 385–393. [CrossRef]
50. Qiu, B.; Li, W.; Tang, Z.; Chen, C.; Qi, W. Mapping paddy rice areas based on vegetation phenology and surface moisture conditions. *Ecol. Indic.* **2015**, *56*, 79–86. [CrossRef]
51. Onojeghuo, A.O.; Blackburn, G.A.; Wang, Q.; Atkinson, P.M.; Kindred, D.; Miao, Y. Rice crop phenology mapping at high spatial and temporal resolution using downscaled MODIS time-series. *GIScience Remote Sens.* **2018**, *55*, 659–677. [CrossRef]
52. Zhong, C.; Wang, C.; Wu, C. Modis-based fractional crop mapping in the US Midwest with spatially constrained phenological mixture analysis. *Remote Sens.* **2015**, *7*, 512–529. [CrossRef]
53. Liang, L.; Schwartz, M.D.; Fei, S. Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. *Remote Sens. Environ.* **2011**, *115*, 143–157. [CrossRef]
54. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
55. Galford, G.L.; Mustard, J.F.; Melillo, J.; Gendrin, A.; Cerri, C.C.; Cerri, C. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens. Environ.* **2008**, *112*, 576–587. [CrossRef]
56. Sakamoto, T.; Yokozawa, M.; Toritani, H.; Shibayama, M.; Ishitsuka, N.; Ohno, H. A crop phenology detection method using time-series MODIS data. *Remote Sens. Environ.* **2005**, *96*, 366–374. [CrossRef]
57. Yang, S.; Gu, L.; Li, X.; Jiang, T.; Ren, R. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote Sens.* **2020**, *12*, 3119. [CrossRef]
58. Dimitrov, P.; Dong, Q.; Eerens, H.; Gikov, A.; Filchev, L.; Roumenina, E.; Jelev, G. Sub-pixel crop type classification using PROBA-V 100 m NDVI time series and reference data from Sentinel-2 classifications. *Remote Sens.* **2019**, *11*, 1370. [CrossRef]
59. Papadimitriou, F. *Spatial Complexity: Theory, Mathematical Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2020.
60. Sun, H.; Xu, A.; Lin, H.; Zhang, L.; Mei, Y. Winter wheat mapping using temporal signatures of MODIS vegetation index data. *Int. J. Remote Sens.* **2012**, *33*, 5026–5042. [CrossRef]
61. Wang, C.; Zhong, C.; Yang, Z. Assessing bioenergy-driven agricultural land use change and biomass quantities in the US Midwest with MODIS time series. *J. Appl. Remote Sens.* **2014**, *8*, 085198. [CrossRef]
62. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
63. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
64. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
65. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
66. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
67. Lunetta, R.S.; Shao, Y.; Ediriwickrema, J.; Lyon, J.G. Monitoring agricultural cropping patterns across the Laurentian Great Lakes Basin using MODIS-NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 81–88. [CrossRef]